

**Romualdo Alves Pereira Júnior**

# **Uma Proposta de Arquitetura Genética da Informação**

Brasília - DF  
Dezembro/2012



**Romualdo Alves Pereira Júnior**

## **Uma Proposta de Arquitetura Genética da Informação**

Tese para o Programa de Pós-Graduação em  
Ciência da Informação como requisito parcial  
para obtenção do grau de doutor.

**Orientador: Mamede Lima-Marques**

Universidade de Brasília – UnB  
Faculdade de Ciência da Informação - FCI  
Programa de Pós-Graduação em Ciência da Informação - PPGCInf

Brasília - DF  
Dezembro/2012

---

Romualdo Alves Pereira Júnior

Uma Proposta de Arquitetura Genética da Informação/ Romualdo Alves Pereira Júnior. – Brasília - DF, Dezembro/2012

240 p. : il. (algumas color.) ; 30 cm.

Orientador: Mamede Lima-Marques

Tese (Doutorado) – Universidade de Brasília – UnB

Faculdade de Ciência da Informação - FCI

Programa de Pós-Graduação em Ciência da Informação - PPGCInf, Dezembro/2012.

Bibliografia: p. 169–187.

1. Arquitetura da Informação. 2. Fenomenologia Genética. I. Lima-Marques, Mamede. II. Universidade de Brasília. III. Faculdade de Ciência da Informação. IV. Título

CDU 02:141:005.7

---

**Romualdo Alves Pereira Júnior**  
**Uma Proposta de Arquitetura Genética da Informação**

Tese para o Programa de Pós-Graduação em  
Ciência da Informação como requisito parcial  
para obtenção do grau de doutor.

Tese aprovada. Brasília - DF, 14 de dezembro de 2012:

---

**Mamede Lima-Marques**  
Orientador, Presidente (UnB/PPGCIInf)

---

**Dr. Manoel Fernando da Mota Tenorio**  
Membro Externo (USC)

---

**Dr. Jackson Max Furtunato Maia**  
Membro Externo (INPE)

---

**Prof. Dr. André Porto Ancona Lopez**  
Membro Interno (UnB/PPGCIInf)

---

**Prof. Dr. Cláudio Gottschalg Duque**  
Membro Interno (UnB/PPGCIInf)

---

**Dr. Gustavo Vasconcellos Cavalcante**  
Membro Interno (UnB/PPGCIInf)

Brasília - DF  
Dezembro/2012



*À minha querida sobrinha, **Cristiane** (in memoriam),  
pelo seu vigor, alegria, perseverança  
e tantas outras importantes lições de vida que nos deixou.*





# Agradecimentos

*“O conhecimento ensoberbece, mas o amor edifica.”*

— 1 Coríntios 8:1.

Com o passar dos anos, eu pude constatar que “precisamos aprender a contar os nossos dias para que alcancemos coração sábio” (nas palavras de Moisés, que havia sido ensinado em toda a ciência da nação mais próspera de sua época, o Egito). E um dos aspectos da sabedoria é o reconhecimento da necessidade dos vínculos. No entanto, há vínculos que são fortes e outros que são fracos, dependendo do fator de ligação. De minha parte, estou certo de que o amor vincula perfeitamente todas as coisas. Assim, este é um pequeno espaço de reconhecimento e de declaração de que não podemos ir muito longe sozinhos, pois precisamos uns dos outros. Daí, os meus sinceros agradecimentos aos que fizeram parte desses vínculos:

Ao amigo, professor e orientador, Mamede Lima-Marques, pela inesquecível e arrebatadora primeira aula, que desencadeou em uma nova e desafiadora maneira de desenvolver o meu tema de tese. A ele sou grato por sua gentileza e disposição incondicionais; pelo direcionamento de olhos que vão além; pelas exigências de rigor da forma e conteúdo; e pela oportunidade de participar de seu seletto grupo de pesquisa.

Ao amigo Tenorio, pelo constante encorajamento e cumplicidade desde o início e, mesmo antes, em outros projetos de pesquisa. A ele sou grato pelas brilhantes ideias e revelações, sempre acolhidas com muito carinho; e pela comunhão na Palavra.

Ao amigo Jackson Maia, por sua disposição em aproximar-se e intervir em momentos cruciais, que invariavelmente resultava em longas e prazerosas conversas; pelas oportunas refutações e questionamentos, em sua declarada postura de quasi-ceticismo; por ensinar-me a fugir dos adjetivismos nas argumentações, por todas as ótimas dicas de leitura e pela minuciosa revisão crítica da tese apresentada para a defesa.

Aos demais membros da banca examinadora, pelo interesse no tema, pelo árduo trabalho de ler e avaliar a minha tese, com sugestões valiosas para a versão final do documento: André Porto, Cláudio Duque e Gustavo Cavalcante.

Aos professores e professoras da Faculdade de Ciência da Informação: Sofia, Emir, Marisa, Suzana e Sely; do Departamento de Letras: Maria Coroa; e da Comunicação: Martino.

Ao amigo de longas datas, Anderson Nascimento, que me apresentou, em diferentes circunstâncias, aos professores Mamede e Tenorio. A ele sou grato pela amizade e oportunidades docentes criadas na ENE/UnB durante o meu doutorado, sem dúvida um dos maiores desafios acadêmicos por que já passei.

Ao amigo Oscar Brandão, por sempre ter-me honrado com sua presença e consideração.

Ao amigo Daniel Oliveira, pelas interessantes discussões sobre visão de mundo e por sempre estar ali ao lado, em uma postura inclusiva e inspiradora de nos fazer ver a importância da colaboração em redes sociais efetivas.

Ao amigo André Siqueira sou devedor de inúmeras eloquentes explicações, apresentações e aulas, com quem muito aprendi.

Agradeço ao Prof. Dr. André Miede pelo belíssima e rica tipografia do modelo latex que utilizei na versão defesa. Devo-lhe um cartão postal de Brasília (a pedido).

Sou muito grato ao Lauro César por ter-me disponibilizado a versão ABNTex2 desenvolvida por ele e em fase de publicação no repositório CTAN.

Agradeço à amizade e companheirismo do Achilles, Alberto, Alfram, Átila, Carlson, Elizabeth, Érica, Flávia, Gilberto, Gustavo, Ismael, Jorge Duarte, Marcelo Ávila, Marcelo Schiessl, Orélio, Paulo Argolo, Ravi, Ricardo e Romani pelos muitos passos que demos juntos.

Sou grato ao Jörg Bliesener por ter-me disponibilizado o banco de dados curriculares (Plataforma Lattes) de sua pesquisa, para a realização de uma das aplicações desta tese.

Ao biólogo Marcus agradeço pela revisão completa das ontologias desenvolvidas e pela construção de uma versão 2.0, já disponível.

A todo o pessoal do CPAI/UnB, com quem convivi nestes anos. O meu muito obrigado à Helena, Maria Ângela, Dora, Ana Lídia, Isadora, Jaqueline e Kamila. Igualmente, às secretárias do PPGCInf/UnB, Jucilene e Martha, pelo apoio e presteza em todos os serviços de sua competência.

Sou muitíssimo grato aos meus colegas de trabalho mais próximos na Agência Espacial Brasileira: Eduardo Alves, Luís Gabriel, Daniel Luchetta, Anderson Mendes, Joesley, Pedro Ferreira; ao Cleodato e Iram.

Dignos de honra são os meus pais, Romualdo e Maria (que hoje descansam no Senhor). Em todo o cuidado que exigem a criação e educação, eles me proporcionaram condições adequadas para uma vida simples e com propósito.

Obrigado Paulo, Vinícius e Elvis pelas frutíferas discussões de fim-de-semana em Corumbá; Augusto e Lélia, Madalena e Paulinha, pela hospedagem, momentos de descontração e todo o apoio que sempre recebemos ali.

Sou grato às orações dos meus irmãos de fé e pelo convívio semanal de casa em casa, no desfrute das insondáveis riquezas de Cristo. A todos vocês, graça e paz, não a que o mundo dá, mas a que excede todo o entendimento.

Por fim, e muito especialmente, sou eternamente grato à Kátia, Renan e Karen, que enfrentaram cada passo desta jornada ao meu lado, e cujos planos familiares muitas vezes tiveram que ser deixados de lado. Obrigado pelo apoio, compreensão sem medida e encorajamento.

*“A ninguém fiquéis devendo coisa alguma, exceto o amor... que é o vínculo da perfeição.” (Rm 13:8; Cl 3:14)*

# A Loucura e Fraqueza de Deus

*“Pois está escrito:*

*’Destruirei a sabedoria dos sábios  
e rejeitarei a inteligência dos inteligentes’.*

*Onde está o sábio? Onde está o erudito?*

*Onde está o questionador desta era?*

*Acaso não tornou Deus louca a sabedoria deste mundo?*

*Visto que, na sabedoria de Deus,  
o mundo não o conheceu por meio da sabedoria humana,  
agradou a Deus salvar aqueles que crêem  
por meio da loucura da pregação. . .*

*Porque **a loucura de Deus** é mais sábia que a sabedoria humana,  
e **a fraqueza de Deus** é mais forte que a força do homem.*

*Irmãos, pensem no que vocês eram quando foram chamados.*

*Poucos eram sábios segundo os padrões humanos;*

*poucos eram poderosos;*

*poucos eram de nobre nascimento.*

*Mas Deus escolheu as coisas loucas do mundo*

*para envergonhar os sábios,*

*e escolheu as coisas fracas do mundo*

*para envergonhar as fortes.*

*Ele escolheu as coisas insignificantes do mundo,*

*as desprezadas e as que nada são,*

*para reduzir a nada as que são,*

*para que ninguém se vanglorie diante dele.*

*. . .*

*Para que, como está escrito:*

*’Quem se gloriar, glorie-se no Senhor’.*

*— Bíblia Sagrada (NVI) - 1 Coríntios 1:19-31*



# Resumo

Esta tese propõe um referencial epistemológico e teórico para a ideia de uma Arquitetura Genética da Informação, fundamentada pela Teoria Geral da Arquitetura da Informação, proposta por Lima-Marques (2011); no âmbito da disciplina de Arquitetura da Informação, proposta por Siqueira (2012); e considerações da Fenomenologia Genética. A *Arquitetura Genética da Informação* compreende dois aspectos fundamentais: a gênese e a genética da informação intencional. A *Gênese da Informação Intencional* ocorre idealmente como fenômeno em uma das fases da redução fenomenológica; e a *Genética da Informação Intencional* é caracterizada em uma analogia qualitativa identificada entre a Ciência da Informação e a Genética, dentro de suas quatro principais abordagens de estudo, a saber: a Clássica, Molecular, Populacional e Quantitativa. Ontologicamente, são considerados principalmente o genoma, genótipo e fenótipo dos objetos (endurantes) e seus processos (perdurantes) em associação aos termos da informação. Como resultados da pesquisa, apresentamos a *Arquitetura Genética da Informação*; desenvolvemos ontologias de alto-nível de analogia entre a Arquitetura da Informação e a Genética, elicitando os termos da informação e processos biológicos que envolvem a informação; e apresentamos algumas aplicações práticas na temática de similaridade fenotípica (textual, visual, acústica e em redes de pesquisa), evolução e inovação tecnológica por memética, hereditariedade da informação e manipulação genética da informação.

**Palavras-chave:** Informação Intencional; Gênese da Informação; Genética da Informação; Arquitetura da Informação; Fenomenologia Genética.



# Abstract

This thesis proposes an epistemological and theoretical framework for the idea of a Genetic Architecture of Information, founded by the General Theory of the Architecture of Information, proposed by Lima-Marques (2011); under the discipline of Architecture of Information, proposed by Siqueira (2012); and considerations of Genetic Phenomenology. The Genetic Architecture of Information comprises two main aspects: the genesis and genetic intentional information. The *Genesis of Intentional Information* ideally occurs as a phenomenon in one of the stages of the phenomenological reduction; and the *Genetic of Intentional Information* is characterized and identified in a qualitative analogy between the Information Science and Genetics, within its four main study approaches, namely: Classical, Molecular, Population and Quantitative. Ontologically, we mainly consider the genome, genotype and phenotype of objects (endurants) and processes (perdurants) in association with terms of information. As results, we present the Genetic Architecture of Information; high-level ontologies developed in analogy between Information Architecture and Genetics aiming to elicit terms of information and biological processes that involve information; and some practical applications, such as phenotypic similarity (textual, visual, acoustic and in research social networks), technological evolution and innovation by means of memetics, information heritability, and genetic manipulation of information.

**Key-words:** Intentional Information; Genesis of Information; Genetics of Information; Architecture of Information; Genetic Phenomenology.





# Sumário

|           |  |           |
|-----------|--|-----------|
| <b>1</b>  | <b>Introdução</b>  | <b>1</b>  |
| 1.1       | Organização da Tese . . . . .  | 1         |
| <b>I</b>  | <b>Requisitos</b>  | <b>3</b>  |
| <b>2</b>  | <b>Requisitos Pré-pesquisa</b>   | <b>5</b>  |
| 2.1       | Objetivos . . . . .  | 5         |
| 2.1.1     | Objetivo Geral . . . . .   | 5         |
| 2.1.2     | Objetivos Específicos . . . . .  | 5         |
| 2.2       | Problemática . . . . .   | 6         |
| 2.3       | Metodologia . . . . .  | 7         |
| 2.3.1     | Classificação da Pesquisa . . . . .  | 7         |
| 2.3.2     | Percurso Metodológico . . . . .  | 7         |
| 2.4       | Pressupostos Básicos . . . . .   | 10        |
| 2.4.1     | Teoria Geral da Arquitetura da Informação - TGAI . . . . .                                 | 10        |
| 2.4.2     | Fenomenologia Genética . . . . .   | 10        |
| 2.4.3     | Intencionalidade . . . . .   | 12        |
| 2.4.4     | Informação Intencional . . . . .   | 13        |
| 2.4.5     | Arquitetura da Informação Intencional . . . . .  | 17        |
| 2.4.6     | Visão de Mundo . . . . .   | 19        |
| <b>II</b> | <b>Fundamentação Teórica</b>   | <b>21</b> |
| <b>3</b>  | <b>Revisão de Literatura e Fundamentos</b>   | <b>23</b> |
| 3.1       | Ontologia e Informação . . . . .   | 23        |
| 3.2       | Fenomenologia . . . . .  | 24        |
| 3.2.1     | O que é Fenomenologia? . . . . .   | 25        |
| 3.2.2     | Desenvolvimento Histórico-Conceitual da Fenomenologia . . . . .                            | 26        |
| 3.2.2.1   | Precursos: Lambert, Kant e Brentano . . . . .  | 27        |
| 3.2.2.2   | Fundador: Husserl . . . . .  | 28        |
| 3.2.2.3   | Seguidores e/ou Críticos: Heidegger, Merleau-Ponty, Sartre,<br>Polanyi e Dennett . . . . . | 31        |
| 3.3       | Epistemologia da Informação . . . . .  | 34        |

|         |   |    |
|---------|---|----|
| 3.3.1   | Paradigma Físico da Informação . . . . .                  | 36 |
| 3.3.2   | Paradigma Cognitivo da Informação . . . . .               | 39 |
| 3.3.3   | Paradigma Histórico-Social da Informação . . . . .        | 40 |
| 3.4     | Arquitetura da Informação . . . . .                       | 41 |
| 3.5     | Organização e Representação do Conhecimento . . . . .     | 44 |
| 3.5.1   | Teoria da Classificação . . . . .                         | 44 |
| 3.5.2   | Representação do Conhecimento com Ontologias . . . . .    | 47 |
| 3.5.2.1 | Ontologias de Domínio e Ontologias Fundamentais . . . . . | 48 |
| 3.5.2.2 | Ontologias Formais e Ontologias Não-Formais . . . . .     | 49 |
| 3.5.2.3 | Basic Formal Ontology - BFO . . . . .                     | 50 |
| 3.5.3   | Metodologia para a Construção de Ontologias . . . . .     | 51 |
| 3.5.4   | Operações com Ontologias . . . . .                        | 52 |
| 3.5.5   | Padrão de Ontologias Biológicas . . . . .                 | 53 |
| 3.5.5.1 | Classes . . . . .   | 54 |
| 3.5.5.2 | Relações . . . . .  | 55 |
| 3.5.5.3 | Web Ontology Language - OWL . . . . .                     | 55 |
| 3.5.5.4 | Princípios da Ontologia BioTop . . . . .                  | 56 |
| 3.6     | Fundamentos da Genética . . . . .                         | 57 |
| 3.6.1   | Genética . . . . .  | 57 |
| 3.6.2   | Genoma . . . . .  | 58 |
| 3.6.3   | Genótipo . . . . .  | 58 |
| 3.6.4   | Fenótipo . . . . .  | 59 |
| 3.7     | Biologia e Informação . . . . .                           | 60 |

### **III Resultados 65**

#### **4 Arquitetura Genética da Informação 67**

|     |  |    |
|-----|--|----|
| 4.1 | Elementos da Arquitetura . . . . .           | 68 |
| 4.2 | O Arquiteto Genético da Informação . . . . . | 73 |

#### **5 Gênese da Informação Intencional 75**

|       |  |    |
|-------|--|----|
| 5.1   | O Fenômeno da Informação Intencional . . . . .                             | 76 |
| 5.1.1 | Informação Causal . . . . .  | 78 |
| 5.1.2 | Informação Intencional, Significativa ou Semântica . . . . .               | 78 |
| 5.2   | O Processo da Gênese da Informação Intencional . . . . .                   | 80 |
| 5.3   | Modelagem da Gênese da Informação Intencional com Redes de Petri . . . . . | 81 |

#### **6 Genética da Informação Intencional 87**

|       |  |    |
|-------|--|----|
| 6.1   | Teorias da Informação e Genética . . . . . | 87 |
| 6.1.1 | Teoria da Informação de Shannon . . . . .  | 87 |

|          |   |            |
|----------|---|------------|
| 6.1.2    | Teoria da Informação de Kolmogorov . . . . .                          | 90         |
| 6.1.3    | Teoria da Informação Algorítmica . . . . .                            | 91         |
| 6.2      | Arquitetura Genética: Mapeamento Genótipo-Fenótipo . . . . .          | 91         |
| 6.3      | Armazenamento da Informação . . . . .                                 | 92         |
| 6.4      | Processamento da Informação . . . . .                                 | 94         |
| 6.4.1    | Informação Genética Intencional . . . . .                             | 95         |
| 6.4.2    | Informação e Seleção Natural . . . . .                                | 95         |
| 6.4.2.1  | Complexidade . . . . .  | 95         |
| 6.4.2.2  | Estrutura Evolucionária . . . . .                                     | 96         |
| 6.4.3    | Informação e Imunização Biológica . . . . .                           | 97         |
| 6.4.4    | Informação e Aprendizado Biológico . . . . .                          | 97         |
| 6.4.5    | Aleatoriedade da Informação na Expressão Genética . . . . .           | 98         |
| 6.4.6    | Redes Biológicas Booleanas . . . . .                                  | 98         |
| 6.4.7    | Informação e Evolução . . . . .                                       | 99         |
| 6.4.7.1  | Memética . . . . .  | 99         |
| 6.4.8    | Programação Genética - PG . . . . .                                   | 100        |
| 6.4.8.1  | Genótipo e Fenótipo . . . . .   | 101        |
| 6.4.8.2  | Blocos de Construção . . . . .  | 101        |
| 6.4.8.3  | Íntrons . . . . .   | 101        |
| 6.4.9    | Programação Evolutiva - PE . . . . .                                  | 101        |
| 6.5      | Leitura e Expressão da Informação . . . . .                           | 102        |
| 6.6      | Busca e Recuperação por Similaridade . . . . .                        | 103        |
| 6.6.1    | Características da Similaridade . . . . .                             | 104        |
| 6.6.2    | Similaridade Molecular . . . . .                                      | 105        |
| 6.6.3    | Auto-Similaridade . . . . .   | 106        |
| 6.6.4    | Similaridade Fenotípica . . . . .                                     | 106        |
| <b>7</b> | <b>Uma Ontologia para a Arquitetura Genética da Informação</b>        | <b>109</b> |
| 7.1      | Genética Clássica da Informação . . . . .                             | 110        |
| 7.2      | Genética Molecular da Informação . . . . .                            | 110        |
| 7.3      | Genética Populacional da Informação . . . . .                         | 111        |
| 7.4      | Genética Quantitativa da Informação . . . . .                         | 112        |
| 7.5      | Construção das Ontologias . . . . .                                   | 112        |
| 7.5.1    | Construção da Ontologia de Termos da Informação na Genética . . . . . | 113        |
| 7.5.1.1  | Fase 1: Identificação do propósito e escopo . . . . .                 | 115        |
| 7.5.1.2  | Fase 2: Aquisição de conhecimento . . . . .                           | 115        |
| 7.5.1.3  | Fase 3: Conceitualização . . . . .                                    | 116        |
| 7.5.1.4  | Fase 4: Integração . . . . .  | 116        |
| 7.5.1.5  | Fase 5: Codificação . . . . .   | 117        |
| 7.5.1.6  | Fase 6: Documentação . . . . .  | 117        |

|           |  |            |
|-----------|--|------------|
| 7.5.1.7   | Fase 7: Avaliação . . . . .  | 117        |
| 7.5.2     | Construção da Ontologia de Processamento da Informação em Sistemas Biológicos . . . . .                              | 118        |
| 7.5.2.1   | Fase 1: Identificação do propósito e escopo . . . . .  | 118        |
| 7.5.2.2   | Fase 2: Aquisição de conhecimento . . . . .  | 118        |
| 7.5.2.3   | Fase 3: Conceitualização . . . . .   | 119        |
| 7.5.2.4   | Fase 4: Integração . . . . .   | 120        |
| 7.5.2.5   | Fase 5: Codificação . . . . .  | 120        |
| 7.5.2.6   | Fase 6: Documentação . . . . .   | 120        |
| 7.5.2.7   | Fase 7: Avaliação . . . . .  | 120        |
| 7.5.3     | Mesclagem das Ontologias . . . . .   | 122        |
| <b>8</b>  | <b>Respostas a Problemas Abertos em Filosofia da Informação</b>  | <b>127</b> |
| 8.1       | P.1 - O Problema Elementar: O que é Informação? . . . . .  | 127        |
| 8.2       | P.2 - O Problema de Entrada/Saída: Qual é a dinâmica da Informação? . . . . .  | 128        |
| 8.3       | P.4 - O Problema Fundamental dos Dados: De que maneira os dados podem adquirir significado? . . . . .                | 128        |
| 8.4       | P.5 - O Problema da Aletização: De que maneira os dados significativos podem adquirir valores verdadeiros? . . . . . | 128        |
| 8.5       | P.6 - Teoria da Verdade Informativa: A informação pode explicar a verdade? . . . . .                                 | 129        |
| <b>IV</b> | <b>Aplicações</b>  | <b>131</b> |
| <b>9</b>  | <b>Aplicações da Genética da Informação</b>  | <b>133</b> |
| 9.1       | Aplicações de Similaridade Fenotípica . . . . .  | 133        |
| 9.1.1     | Descoberta Fenotípica de Literatura Conexa . . . . .   | 133        |
| 9.1.2     | Similaridade Fenotípica entre Imagens . . . . .  | 135        |
| 9.1.3     | Similaridade Fenotípica do Canto dos Pássaros . . . . .  | 140        |
| 9.1.4     | Similaridade Fenotípica em Redes de Pesquisa . . . . .   | 144        |
| 9.2       | Evolução Natural e Artificial de Gêneros Musicais . . . . .  | 147        |
| 9.3       | O Projeto Genoma Musical . . . . .   | 149        |
| 9.4       | Estrutura Fenotípica de Retórica Textual . . . . .   | 150        |
| 9.5       | Algoritmos Genéticos Modelados com Redes de Petri . . . . .  | 154        |
| 9.6       | Índice DNA Brasil . . . . .  | 156        |
| 9.7       | Evolução do Sistema Nacional de Dados Ambientais - SINDA . . . . .   | 159        |
| <b>V</b>  | <b>Considerações Finais</b>  | <b>163</b> |
| <b>10</b> | <b>Conclusão</b>   | <b>165</b> |

|   |            |
|---|------------|
| 10.1 Limitações e Dificuldades Encontradas . . . . .                        | 166        |
| 10.2 Trabalhos Futuros . . . . .  | 166        |
| <b>Referências</b>  | <b>169</b> |
| <br>  |            |
| <b>VI Apêndices</b>   | <b>197</b> |
| <br>  |            |
| <b>A Uma Proposta de Formalização da Arquitetura Genética da Informação</b> | <b>199</b> |
| A.1 Modelo Matemático na Teoria dos Conjuntos . . . . .                     | 199        |
| A.1.1 Definições . . . . .  | 199        |
| A.1.2 Teoremas, Lemas e Demonstrações . . . . .                             | 202        |
| A.1.3 Propriedades . . . . .  | 203        |
| <br>  |            |
| <b>B Diagramas da Ontologia da Arquitetura Genética da Informação</b>       | <b>205</b> |
| <br>  |            |
| <b>C Rede de Petri de Modelagem da Gênese da Informação Intencional</b>     | <b>217</b> |
| <br>  |            |
| <b>D Produção Científica - 2008-2012</b>                                    | <b>225</b> |
| <br>  |            |
| <b>Índice</b>   | <b>237</b> |



# Lista de ilustrações

|           |   |    |
|-----------|---|----|
| Figura 1  | Diagrama de Percurso Metodológico - Fonte: Produzido pelo autor . . . . .   | 9  |
| Figura 2  | Arquitetura da Informação Intencional . . . . .   | 18 |
| Figura 3  | Linha do Tempo de Influência dos Pensadores da Fenomenologia - Fonte: (PEREIRA-JR, 2011) . . . . .                          | 27 |
| Figura 4  | Relógio Excêntrico (Ilustração do Bracketing) . . . . .   | 31 |
| Figura 5  | Esquema de um Sistema Geral de Comunicação - Fonte: (SHANNON, 1948)   | 37 |
| Figura 6  | Informação, Desinformação e Má-informação Derivadas da TMC - Fonte: (FLORIDI, 2007) . . . . .                               | 37 |
| Figura 7  | Cronologia da Arquitetura da Informacao - Fonte: (LEÓN, 2008) . . . . .   | 44 |
| Figura 8  | Ontologia BFO - Classes Hierárquicas (até o terceiro nível) - Fonte: Produzido pelo autor . . . . .                         | 51 |
| Figura 9  | Ontologia BFO - Classe das Entidades Continuantes - Fonte: Produzido pelo autor . . . . .                                   | 52 |
| Figura 10 | Ontologia BFO - Classe das Entidades Ocorrentes- Fonte: Produzido pelo autor . . . . .                                      | 53 |
| Figura 11 | Metodologia de Construção de Ontologias - Modelo-V - Fonte: (OULD, 1990)  | 54 |
| Figura 12 | Ontologia BioTop - Fonte: (HANCOCK; ZVELEBIL, 2004) . . . . .   | 57 |
| Figura 13 | Arquitetura Genética da Informação - Fonte: Produzido pelo autor . . . . .  | 67 |
| Figura 14 | Caracterização de Problemas na Arquitetura Genética da Informação - Fonte: Produzido pelo autor . . . . .                   | 74 |
| Figura 15 | Ontologização: O Caráter Fenomenológico da Informação - Fonte: (SAAB; RISS, 2011) . . . . .                                 | 76 |
| Figura 16 | Elementos de uma Rede de Petri - Fonte: Produzido pelo autor . . . . .  | 82 |
| Figura 17 | Rede de Petri - Abastecimento de Combustível - Fonte: Produzido pelo autor  | 83 |
| Figura 18 | Exemplo de Lugares e Transições em uma Rede de Petri - Fonte: Produzido pelo autor . . . . .                                | 84 |
| Figura 19 | Modelagem da Gênese da Informação Intencional em uma Rede de Petri - Fonte: Produzido pelo autor . . . . .                  | 85 |
| Figura 20 | Diagrama de Comparação da TMC com o DNA - Fonte: (SMITH, 2010) . . .  | 88 |
| Figura 21 | O processo pelo qual as proteínas são feitas pelas instruções codificadas no DNA - Fonte: (RITTNER; MCCABE, 2004) . . . . . | 89 |

|           |   |     |
|-----------|---|-----|
| Figura 22 | Mary Hesse e o Uso de Analogias Científicas - Fonte: (LOSEE, 2001) . . . . .  | 113 |
| Figura 23 | Ciclo de Vida de Construção da Ontologia de Termos da Informação na Genética - Fonte: Produzido pelo autor . . . . .                    | 114 |
| Figura 24 | Extração de termos com mineração textual - Resultado - Fonte: Produzido pelo autor . . . . .  | 116 |
| Figura 25 | Visualização Gráfica Parcial da Ontologia de Termos da Informação na Genética - Fonte: Produzido pelo autor . . . . .                   | 117 |
| Figura 26 | Relações típicas em Ontologias Biomédicas de alto-nível - Fonte: Produzido pelo autor . . . . .   | 118 |
| Figura 27 | Ciclo de Vida de Construção da Ontologia de Processamento da Informação em Sistemas Biológicos - Fonte: Produzido pelo autor . . . . .  | 119 |
| Figura 28 | Processo de Mineração Textual - Fonte: Produzido pelo autor . . . . .   | 120 |
| Figura 29 | Resultado do Processo de Mineração Textual - Fonte: Produzido pelo autor . . . . .  | 121 |
| Figura 30 | Visualização Gráfica Parcial da Ontologia de Processamento da Informação em Sistemas Biológicos - Fonte: Produzido pelo autor . . . . . | 122 |
| Figura 31 | Primeiro Nível das Classes Resultantes da Mesclagem das Ontologias - Fonte: Produzido pelo autor . . . . .                              | 123 |
| Figura 32 | Radical 'inform' Resultante da Mesclagem das Ontologias - Fonte: Produzido pelo autor . . . . .   | 124 |
| Figura 33 | Grafo de Similaridade Fenotípica entre os Livros-Base para a Construção das Ontologias - Fonte: Produzido pelo autor . . . . .          | 134 |
| Figura 34 | Tabela de Distância de Similaridade entre os Livros-Base para a Construção das Ontologias - Fonte: Produzido pelo autor . . . . .       | 135 |
| Figura 35 | Varredura Zig-Zag de uma Imagem - Fonte: Symes (2001) . . . . .   | 136 |
| Figura 36 | Cães que se parecem com os seus donos - Fonte: (TANMONKEY, 2012) . . . . .  | 136 |
| Figura 37 | Resultado da Busca por Similaridade de Imagens - Pergunta 1 - Fonte: Produzido pelo autor . . . . .                                     | 137 |
| Figura 38 | Resultado da Busca por Similaridade de Imagens - Pergunta 2 - Fonte: Produzido pelo autor . . . . .                                     | 137 |
| Figura 39 | Caracterização da Arquitetura Genética da Informação - Parte A - Fonte: Produzido pelo autor . . . . .                                  | 138 |
| Figura 40 | Caracterização da Arquitetura Genética da Informação - Parte B - Fonte: Produzido pelo autor . . . . .                                  | 139 |
| Figura 41 | Caracterização da Arquitetura Genética da Informação - Parte C - Fonte: Produzido pelo autor . . . . .                                  | 139 |
| Figura 42 | Caracterização da Arquitetura Genética da Informação - Parte D - Fonte: Produzido pelo autor . . . . .                                  | 140 |
| Figura 43 | Pássaro Northern Cardinal Grosbeak- Fonte: (WAITE, 2012) . . . . .  | 141 |
| Figura 44 | Pássaros Similares ao Northern Cardinal Grosbeak- Fonte: (WAITE, 2012) . . . . .  | 142 |



|           |  |     |
|-----------|--|-----|
| Figura 45 | Cantos Similares ao do Northern Cardinal Grosbeak - Fonte: (WAITE, 2012)   | 143 |
| Figura 46 | Notação Musical de Messiaen para o Canto do Cardinal Grosbeak- Fonte: (FALLON, 2007)   | 144 |
| Figura 47 | Modelo de Entidades e Relacionamentos (EER) - Lattes - Microeletrônica- Fonte: reduzido de Neves (2010)                                      | 145 |
| Figura 48 | Sistema de Business Intelligence para a Rede de Pesquisadores em Microeletrônica- Fonte: Produzido pelo autor                                | 146 |
| Figura 49 | Evolução do Samba - Fenótipo: Instrumentos Musicais - Fonte: Produzido pelo autor  | 148 |
| Figura 50 | Manipulação Genética da Informação - Fonte: Produzido pelo autor   | 149 |
| Figura 51 | Busca por Similaridade Acústica - Fonte: (SOUNDHOUND, 2012)  | 150 |
| Figura 52 | Esquemas RST - Traços Fenotípicos Textuais- Fonte: Produzido pelo autor  | 152 |
| Figura 53 | Segmentação - Fonte: Produzido pelo autor  | 153 |
| Figura 54 | Estruturação - Fonte: Produzido pelo autor   | 153 |
| Figura 55 | Sumarização - Fonte: Produzido pelo autor  | 154 |
| Figura 56 | Modelagem do Processo de Produção de Proteínas com Rede de Petri- Fonte: (BARJIS; SAMARRAI; AUGUSTIN, 2006)                                  | 155 |
| Figura 57 | Modelagem do Processo de Síntese de Proteínas, Regulação Genética e Biossíntese com Rede de Petri- Fonte: (BARJIS; SAMARRAI; AUGUSTIN, 2006) | 156 |
| Figura 58 | Índice DNA Brasil - Fonte: (BIASOTO-JR. et al., 2006)  | 157 |
| Figura 59 | Árvore Filomemética (Parcial) de Evolução dos Telefones Celulares Nokia - Fonte: (KHANAFIAH; SITUNGKIR, 2004)                                | 159 |
| Figura 60 | Evolução do SINDA - Fonte: (PEREIRA-JR et al., 2012)   | 160 |
| Figura 61 | Árvore Filomemética de Evolução do SINDA - Fonte: Produzido pelo autor   | 161 |
| Figura 62 | Decifrando o Código Genético (Códon) - Fonte: (SOUZA, 2012)  | 190 |
| Figura 63 | Molécula de DNA - Fonte: (SECKO, 2011)   | 190 |
| Figura 64 | Gene - Fonte: (SNUSTAD; SIMMONS, 2008)   | 192 |
| Figura 65 | Mutação - Fonte: (SOUZA, 2012)   | 194 |
| Figura 66 | Penetrância: Um Mapa Genótipo-Fenótipo do Câncer - Fonte: (HERSHBERG; EFRONI, 2001)  | 195 |
| Figura 67 | Transcrição - Fonte: (SNUSTAD; SIMMONS, 2008)  | 196 |
| Figura 68 | OWL - Genômica - Fonte: Produzido pelo autor   | 205 |
| Figura 69 | OWL - Genética Clássica - Fonte: Produzido pelo autor  | 206 |
| Figura 70 | OWL - Genética Molecular (I) - Fonte: Produzido pelo autor   | 206 |
| Figura 71 | OWL - Genética Molecular (II) - Fonte: Produzido pelo autor  | 207 |
| Figura 72 | OWL - Genética Molecular (III) - Fonte: Produzido pelo autor   | 207 |
| Figura 73 | OWL - Genética Molecular (IV) - Fonte: Produzido pelo autor  | 208 |
| Figura 74 | OWL - Genética Populacional (I) - Fonte: Produzido pelo autor  | 208 |

|           |   |     |
|-----------|---|-----|
| Figura 75 | OWL - Genética Populacional (II) - Fonte: Produzido pelo autor . . . . .    | 209 |
| Figura 76 | OWL - Genética Quantitativa - Fonte: Produzido pelo autor . . . . .         | 209 |
| Figura 77 | OWL - Genética da Informação Biológica (I) - Fonte: Produzido pelo autor    | 210 |
| Figura 78 | OWL - Genética da Informação Biológica (II) - Fonte: Produzido pelo autor   | 210 |
| Figura 79 | OWL - Genética da Informação Biológica (III) - Fonte: Produzido pelo autor  | 211 |
| Figura 80 | OWL - Genética da Informação Biológica (IV) - Fonte: Produzido pelo autor   | 212 |
| Figura 81 | OWL - Genética da Informação Biológica (V) - Fonte: Produzido pelo autor    | 212 |
| Figura 82 | OWL - Genética da Informação Biológica (VI) - Fonte: Produzido pelo autor   | 213 |
| Figura 83 | OWL - Engenharia Genética (I) - Fonte: Produzido pelo autor . . . . .       | 213 |
| Figura 84 | OWL - Engenharia Genética (II) - Fonte: Produzido pelo autor . . . . .      | 214 |
| Figura 85 | OWL - Genética Humana (I) - Fonte: Produzido pelo autor . . . . .           | 214 |
| Figura 86 | OWL - Genética Humana (II) - Fonte: Produzido pelo autor . . . . .          | 215 |
| Figura 87 | Imanência ou Transcendência? - Fonte: Produzido pelo autor . . . . .        | 217 |
| Figura 88 | Transcendência habilitada - Fonte: Produzido pelo autor . . . . .           | 218 |
| Figura 89 | Imanência executada - Fonte: Produzido pelo autor . . . . .                 | 219 |
| Figura 90 | Percepção executada . . . . .   | 220 |
| Figura 91 | Início do Bracketing: Noemata executada - Fonte: Produzido pelo autor . .   | 221 |
| Figura 92 | Término do Bracketing: Noese executada - Fonte: Produzido pelo autor . .    | 222 |
| Figura 93 | Emergência do significado executada - Fonte: Produzido pelo autor . . . . . | 223 |

# Lista de Tabelas

|          |   |     |
|----------|---|-----|
| Tabela 1 | Tabela de Analogia Gene-Meme - Fonte: Wilkins (1998) . . . . .  | 100 |
| Tabela 2 | Tabela de Hierarquia da Natureza - Fonte: Sawai (2011) . . . . .  | 115 |
| Tabela 3 | Tabela de Correlação Conceitual de Entes da Biologia e da Arquitetura Genética da Informação - Fonte: Produzido pelo autor . . . . .  | 125 |
| Tabela 4 | Tabela de Correlação Conceitual entre Termos Genômicos e a Arquitetura Genética da Informação - Fonte: Produzido pelo autor . . . . . | 126 |



# Lista de Siglas e Abreviaturas

|      |  |
|------|--|
| AEB  | Agência Espacial Brasileira  |
| BFO  | Basic Formal Ontology  |
| CBIR | Content-Based Image Retrieval  |
| CLD  | Color Layout Descriptor  |
| CMCD | Centro de Missão Coleta de Dados   |
| CNPq | Conselho Nacional de Desenvolvimento Científico e Tecnológico                    |
| CPAI | Centro de Pesquisa em Arquitetura da Informação                                  |
| DL   | Descriptive Logic (Lógica Descritiva)  |
| DNA  | Deoxiribonucleic Acid (Ácido Desoxiribonucléico)                                 |
| EHD  | Edge Histogram Descriptor  |
| FCMS | Forma, Contexto, Manifestação e Significado                                      |
| GO   | Gene Ontology  |
| IDH  | Índice de Desenvolvimento Humano   |
| INPE | Instituto Nacional de Pesquisas Espaciais  |
| KDD  | Knowledge Discovery in Databases (Descoberta de Conhecimento em Bancos de Dados) |
| LRD  | Literature-Related Discovery (Descoberta de Literatura Conexa)                   |
| OWL  | Web Ontology Language  |
| PAM  | Positive Assortative Mating (Acasalamento Associativo Positivo)                  |
| PCD  | Plataforma de Coleta de Dados  |
| PE   | Programação Evolutiva  |
| PG   | Programação Genética   |
| PLN  | Processamento em Linguagem Natural   |

PNAE Programa Nacional de Atividades Espaciais

PPGCInf Programa de Pós-graduação em Ciência da Informação

RNA Ribonucleic Acid (Ácido Ribonucléico)

RST Rhetorical Structure Theory (Teoria da Estrutura Retórica)

SCD Scalable Color Descriptor

SCMCD Software do Centro de Missão Coleta de Dados

SINDA Sistema Nacional de Dados Ambientais

TGAI Teoria Geral da Arquitetura da Informação

TMC Teoria Matemática da Comunicação

UnB Universidade de Brasília







# 1 Introdução

*“Porventura a Deus se ensinaria ciência? Quem primeiro deu a ele para que lhe venha a ser restituído? Porque dele, e por meio dele, e para ele são todas as coisas.”*

— Jó 21:22, Romanos 11:35-36.

“Genética da Informação”, diferentemente de “informação genética”<sup>1</sup>, é a caracterização da informação em termos genéticos, contemplando o fenômeno de sua gênese e os conceitos de genoma, genótipo e fenótipo dos objetos, além dos processos biológicos da informação.

Assumindo-se os pressupostos básicos da Teoria Geral da Arquitetura da Informação - TGAI, proposta por [Lima-Marques \(2011\)](#); a disciplina de Arquitetura da Informação, proposta por [Siqueira \(2012\)](#); o conceito de *informação intencional* e a Fenomenologia Genética, de Husserl, “Uma Proposta de Arquitetura Genética da Informação” é uma tese de caráter epistemológico e teórico que propõe explicar a gênese da *informação intencional* como fenômeno da Fenomenologia e sua consequente caracterização genética.

A Genética é uma área da ciência geralmente estudada sob quatro perspectivas diferentes: Clássica, Molecular, Populacional e Quantitativa. Em linhas gerais, a Genética Clássica (ou Mendeliana) atenta para os indivíduos e famílias, estudando traços físicos e os genes que controlam a aparência, i.e, o fenótipo e sua transmissão para outras gerações; a Genética Molecular se ocupa com as estruturas físicas e químicas do DNA, RNA e proteínas; a Genética Populacional é uma ampliação da genética clássica para grupos maiores; e a Genética Quantitativa é um campo matemático que examina os relacionamentos estatísticos entre os genes e os traços por eles codificados ([ROBINSON, 2006](#)).

## 1.1 Organização da Tese

Esta tese está organizada em seis partes, a saber: I - Requisitos; II - Fundamentação Teórica; III - Resultados; IV - Aplicações; V - Considerações Finais; e VI - Apêndices.

Este Capítulo 1 introduz o trabalho.

A Parte I apresenta os objetivos, problemática, metodologia da pesquisa e pressupostos básicos.

Na Parte II - Fundamentação Teórica, o Capítulo 3 apresenta a revisão de literatura e fundamentos, considerando o estado da arte, referencial teórico, empírico e histórico. A temática está

---

<sup>1</sup>“Informação genética é qualquer informação biológica codificada nas sequências de nucleotídeos do DNA ou RNA” ([STEINBERG; COSLOY, 2009](#))

relacionada à Ontologia; Fenomenologia; Epistemologia; Arquitetura da Informação; Organização e Representação do Conhecimento; Fundamentos da Genética; e Biologia e Informação.

Na Parte III - Resultados, o Capítulo 4 propõe a *Arquitetura Genética da Informação*. O Capítulo 5 discorre sobre o fenômeno da gênese da *informação intencional* no âmbito da *Arquitetura Genética da Informação*. O Capítulo 6 discorre sobre a genética da *informação intencional* no âmbito da *Arquitetura Genética da Informação* considerando: as Teorias da Informação em relação à Genética; a Arquitetura Genética (mapeamento genótipo-fenótipo); os processos biológicos de armazenamento, processamento, leitura, busca e recuperação da informação. O Capítulo 7 propõe a construção de uma ontologia para a Genética da Informação organizada pelas suas principais áreas de estudo. O Capítulo 8 apresenta explicações para cinco dos dez problemas abertos em Filosofia da Informação levantados por Floridi (2007): os problemas 1 (o problema elementar); 2 (o problema de entrada/saída); 4 (o problema fundamental dos dados); 5 (o problema da aletização) e 6 (a teoria da verdade informativa).

Na Parte IV - Aplicações, o Capítulo 9 apresenta algumas aplicações práticas da *Arquitetura Genética da Informação*, na temática de: similaridade fenotípica (descoberta de literatura conexa, similaridade imagética e similaridade acústica); evolução natural e artificial de gêneros musicais; o Projeto Genoma Musical; estrutura fenotípica de retórica textual; algoritmos genéticos modelados com Redes de Petri; o Índice DNA Brasil; e evolução e inovação tecnológica do Sistema Nacional de Dados Ambientais - SINDA.

Na Parte V - Considerações Finais, o Capítulo 10 apresenta a conclusão da pesquisa com seus resultados diretos, limitações e dificuldades encontradas, expectativas de impacto e trabalhos futuros.

Na Parte VI - Apêndices, apresentamos alguns estudos iniciados em relação a uma proposta de formalização matemática da *Arquitetura Genética da Informação*; diagramas da ontologia da *Arquitetura Genética da Informação*; Rede de Petri de modelagem da gênese da *informação intencional* e a produção científica correspondente ao período do curso de doutorado.

**Parte I**

**Requisitos**



## 2 Requisitos Pré-pesquisa

*“No coração da ciência existe um equilíbrio essencial entre duas atitudes aparentemente contraditórias: uma abertura para novas ideias, não importa o quão bizarras ou contra-intuitivas elas sejam; e o exame cético mais implacável de todas as ideias, velhas ou novas. Esta é a maneira como profundas verdades são peneiradas do profundo absurdo.”*

— Carl Sagan

### 2.1 Objetivos

#### 2.1.1 Objetivo Geral

O objetivo geral desta tese é propor um referencial epistemológico e teórico para a ideia de uma *Arquitetura Genética da Informação*, contemplando o fenômeno da *gênese da informação intencional*, sua respectiva caracterização genômica e classificação nas quatro diferentes perspectivas de estudo da Genética, a saber: a Clássica (ou Mendeliana), Molecular, Populacional, e Quantitativa.

#### 2.1.2 Objetivos Específicos

Os objetivos específicos são:

1. No âmbito filosófico, descrever a sua caracterização na Fenomenologia Genética.
2. No âmbito epistemológico, descrever a sua pertinência dentro da teoria do conhecimento sob o ponto de vista fenomenológico.
3. No âmbito científico, relacioná-la à caracterização da Arquitetura da Informação como disciplina, proposta por [Siqueira \(2012\)](#), que, por sua vez, está fundamentada na Teoria Geral da Arquitetura da Informação, proposta por [Lima-Marques \(2011\)](#).
4. No âmbito da práxis, construir ontologias de alto nível dos processos biológicos relacionados à informação e dos conceitos de informação na genética, apresentando aplicações que exemplificam a arquitetura proposta.

## 2.2 Problemática

A definição de uma *Arquitetura Genética da Informação* introduz uma nova abordagem de percepção conceitual fundamentada no âmbito da disciplina de Arquitetura da Informação, sendo passível de ser experienciada, avaliada, validada e falseada. Sob o enfoque de analogia qualitativa entre a Genética e a Arquitetura da Informação, os conceitos de informação são pervasivos e ubíquos nos dois domínios.

A importância científica da *Arquitetura Genética da Informação* se evidencia na contribuição para reflexões fundamentais a respeito da natureza da *informação intencional* sobre os objetos do mundo, desde o fenômeno de sua gênese até a caracterização fenotípica dos objetos, em analogia aos processos dos sistemas de informação biológicos, como armazenamento, processamento, leitura e recuperação da informação por similaridade. Além disso, em conformidade com a disciplina de Arquitetura da Informação proposta por [Siqueira \(2012\)](#), vem fortalecer os trabalhos do grupo do Centro de Pesquisa em Arquitetura da Informação - CPAI da UnB.

Este projeto de pesquisa se filia apropriadamente às linhas de pesquisa do Programa de Pós-Graduação em Ciência da Informação (PPGCIInf) da Faculdade de Ciência da Informação da Universidade de Brasília (UnB), particularmente ao Grupo de Brasília de Arquitetura da Informação, caracterizada por “estudos teóricos e práticos sobre a análise da informação, indexação, estruturas da informação, representação do conhecimento e recuperação da informação”.

Considerando o *Annual Review of Information Science and Technology* - ARIST:

2009 - Volume 43, temos quatro artigos de grande relevância para esta tese: “*Similarity Methods in Chemoinformatics*”, de Willet; “*Information Behavior*”, de Fisher e Julien; “*Information Architecture*”, de Jacob e Loehrlein; e “*Disambiguation*”, de Smalheiser e Torvik.

2010 - Volume 44, um artigo de Jonathan Furne intitulado “*Philosophy and Information Studies*”

2011 - Volume 45, um artigo de Catherine Blake intitulado “*Text Mining*”.

Tais publicações caracterizam o tema desta tese como de interesse atual na Ciência da Informação.

Já, os grupos de pesquisa da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação - ANCIB que se interessam pelo tema são o de “Estudos Históricos e Epistemológicos da Informação”, com relação aos elementos da arquitetura proposta; e o de “Organização e Representação do Conhecimento”, com relação às ontologias desenvolvidas para a verificação de analogia entre domínios distintos.

## 2.3 Metodologia

### 2.3.1 Classificação da Pesquisa

Quanto aos seus objetivos, a presente pesquisa científica caracteriza-se como uma pesquisa descritivo-explicativa, pois, de acordo com [Kahlmeyer-Mertens et al. \(2007\)](#), “tem por objetivo expor as características de determinado fenômeno... e de trazer a explicação de prováveis relações existentes entre uma variável e outra”, no caso, expor as características do fenômeno da gênese da *informação intencional* e trazer explicações das relações analógicas existentes entre a Arquitetura da Informação e a Genética.

### 2.3.2 Percurso Metodológico

A metodologia desenvolvida neste trabalho é uma adaptação de [Green e Rosemann \(2005\)](#) à Metodologia da Meta-Modelagem (M3), cuja proposta é contemplar esferas paralelas e convergentes para as atividades de investigação, visando a solução de problemas, a saber: (1) Filosofia, (2) Epistemologia, (3) Ciência e (4) Práxis.

Apresentamos o diagrama de percurso metodológico na figura 1, página 9, onde podemos observar quatro linhas, que representam as esferas de investigação (Filosófica, Epistemológica, Científica e Pragmática).

O diagrama está dividido em três colunas, que representam as fases do processo (Questões Fundamentais, Referências de Investigação e Resultados Esperados). Assim, temos:

- Filosofia - As principais questões filosóficas nesta pesquisa dizem respeito à natureza ontológica da informação e à gênese da *informação intencional*. Os nossos referenciais de investigação correspondem à análise ontológica da informação e ao fenômeno da *informação intencional*.
- Epistemologia - Recorremos aos conceitos fundamentais de informação, dado e conhecimento e realizamos uma análise epistemológica da informação e do conhecimento.
- Ciência - Propomos um modelo dinâmico da gênese da *informação intencional*, utilizando Redes de Petri; e construímos uma ontologia de similaridade qualitativa entre a Arquitetura da Informação e a Genética; Adicionalmente, iniciamos uma especificação formal matemática da arquitetura proposta (Apêndice A), com a definição de teoremas, propriedades, lemas e corolários, no escopo da Teoria dos Conjuntos;
- Práxis - Desenvolvemos algumas aplicações de similaridade fenotípica (descoberta de literatura conexa, similaridade imagética e similaridade acústica); identificamos processos de evolução natural e artificial de gêneros musicais; apresentamos o Projeto Genoma Musical; desenvolvemos estudo de estrutura fenotípica de retórica textual para a identificação de sumário executivo; apresentamos algoritmos genéticos modelados com Redes de Petri;

o Índice DNA Brasil; e identificamos o processo de evolução e inovação tecnológica do Sistema Nacional de Dados Ambientais - SINDA, com base na memética.



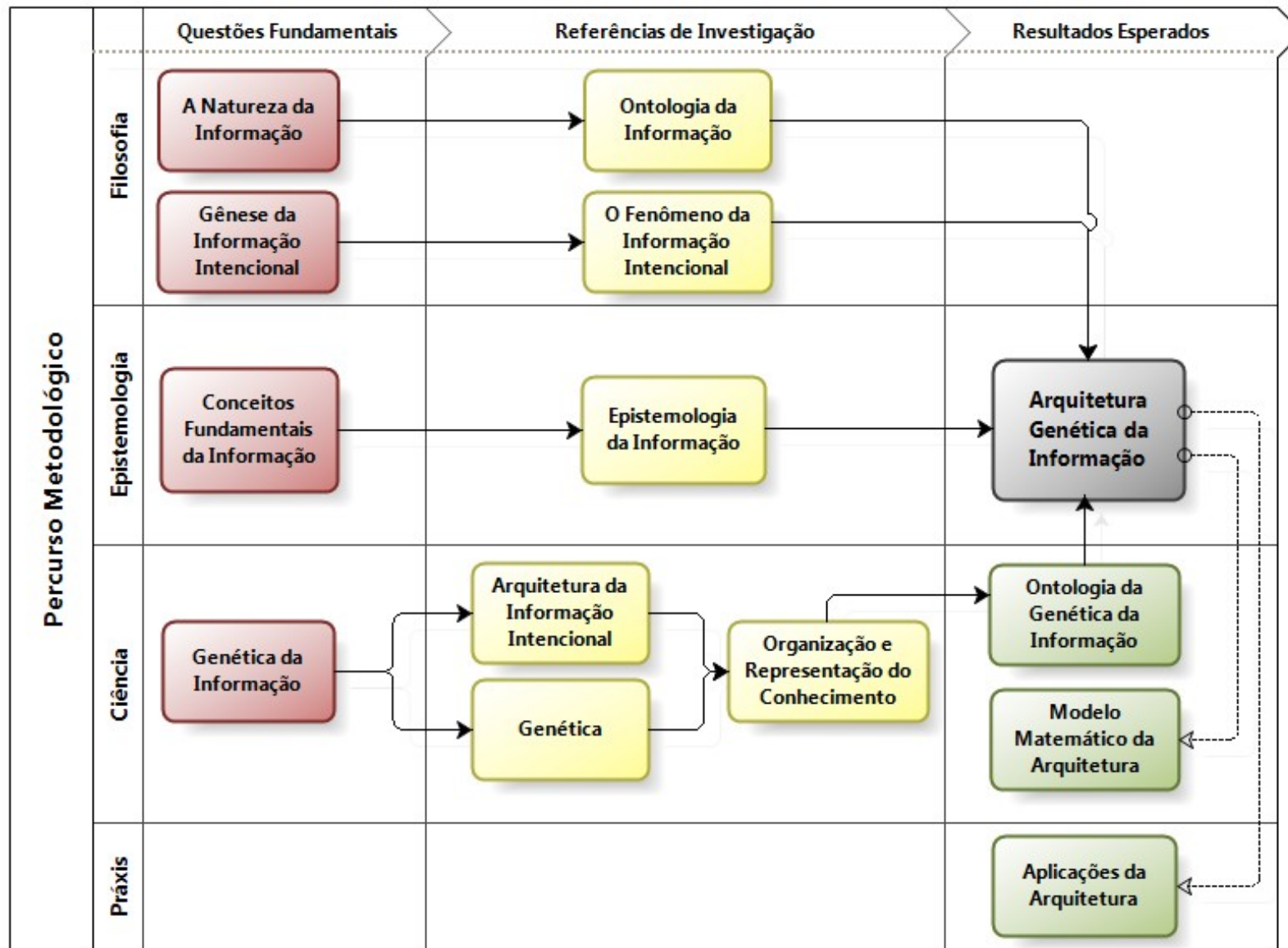


Figura 1: Diagrama de Percorso Metodológico - Fonte: Produzido pelo autor

## 2.4 Pressupostos Básicos

As premissas apresentadas a seguir são fatores externos que estão fora de nossa responsabilidade de intervenção, mas que são muito importantes para a realização dos propósitos desta pesquisa e dos resultados alcançados orientados ao seu objetivo geral. São basicamente as seguintes:

1. Teoria Geral da Arquitetura da Informação - TGAI;
2. Fenomenologia Genética;
3. Intencionalidade Fenomenológica;
4. *Informação Intencional*; e
5. Arquitetura da Informação Intencional.

Estas premissas compõem de forma completa o objeto de estudo desta pesquisa, mantendo-a alinhada à visão de mundo do Centro de Pesquisa em Arquitetura da Informação - CPAI.

Os detalhes são apresentados nos tópicos a seguir.

### 2.4.1 Teoria Geral da Arquitetura da Informação - TGAI

A Teoria Geral da Arquitetura da Informação - TGAI proposta por [Lima-Marques \(2011\)](#) em uma abordagem fenomenológica minimalista<sup>1</sup>, entende a informação em seu caráter objetivo e ontológico, a informação como “coisa”, definida como sendo as propriedades primárias do objeto, independentemente do sujeito. É isso que defende [Dupuy \(1995\)](#), quando afirma que “a informação está na natureza, e a sua existência é, portanto, independente da atividade desses doadores de sentido que são os intérpretes humanos”.

Quanto aos “dados”, estes são definidos como sendo o estado das propriedades primárias do objeto em instante imediatamente anterior à sua apreensão. A demonstração experimental do Princípio de Landauer ([BÉRUT et al., 2012](#)) corrobora com este pressuposto.

No tocante à Fenomenologia, o que interessa à TGAI é que, na relação entre os três elementos – sujeito-objeto-mundo –, há uma limitação na capacidade de apreensão das propriedades ontológicas do objeto por um sujeito no mundo e esta apreensão gera o quarto elemento – conhecimento subjetivo, sendo estes os quatro elementos essenciais de uma epistemologia fenomenológica.

### 2.4.2 Fenomenologia Genética

A fenomenologia pressupõe que a realidade do mundo está além da possibilidade de sua percepção e captação. Assim, os métodos usuais sempre são uma versão empobrecida da realidade observada. Além disso, a fenomenologia defende que é impossível haver uma observação

<sup>1</sup>Em busca de um núcleo conceitual filosófico mínimo e suficiente.

neutra. Suas principais concepções, segundo [Smith \(2011\)](#) são citadas a seguir, sendo de nosso particular interesse, a fenomenologia genética.

1. Fenomenologia Constitutiva Transcendental - Estuda como os objetos são constituídos na consciência pura ou transcendental.
2. Fenomenologia Constitutiva Naturalística - Estuda como a consciência constitui-se das coisas do mundo da natureza, assumindo que a própria consciência é parte da natureza.
3. Fenomenologia Existencial - Estuda a existência concreta do homem, incluindo a experiência de livre-escolha e de ação em situações concretas.
4. Fenomenologia Historicista Generativa - Estuda como o significado encontrado na experiência individual é gerado em processos históricos de experiências coletivas no decorrer do tempo.
5. **Fenomenologia Genética - Estuda a gênese do significado das coisas dentro do fluxo das experiências individuais.**
6. Fenomenologia Hermenêutica - Estuda as estruturas interpretativas da consciência.
7. Fenomenologia Realística - Estuda a estrutura da consciência e a intencionalidade, supondo que esta ocorre em um mundo real que é em grande parte externo à consciência.

[Sandmeyer \(2007\)](#) argumenta que a Fenomenologia Genética caracteriza a proposta final de Husserl, estudando a gênese do significado das coisas dentro do fluxo das experiências individuais.

Dermot [Moran \(2000\)](#) diz que a “fenomenologia genética” foi um projeto que veio à tona na obra de Husserl intitulada “A Crise das Ciências Europeias e a Fenomenologia Transcendental”, mas que já existia ocultamente em manuscritos de muitos anos antes da publicação desse trabalho.

“A constituição estática considera as estruturas noética e noemática que tornam possível aos objetos serem intuídos na consciência, enquanto que a constituição genética examina a maneira pela qual os objetos aparecem dentro do fluxo temporal de nossa experiência. Assim, no projeto de *Idéias III*, Husserl faz uma distinção entre a ‘ontologia’ (que trata os objetos como totalmente formados, com identidades fixas), e a “consideração genética”, que pertence a um mundo totalmente transcendental”. Husserl acreditava que a base da nossa experiência de percepção dos objetos era um processo que ele chamava de ‘gênese passiva’... , que é diferente de ‘síntese passiva’, pois esta última se refere à maneira pela qual nós experimentamos conteúdos sensoriais já estruturados e definidos antes de nós. Assim nos deparamos com objetos formados contra um horizonte de intenções que já estão lá prontas para nós. Nossa experiência é passiva e, no entanto, está estruturada. Gênese passiva, por outro lado, refere-se à estruturação de objetos em camadas sedimentadas umas sobre as outras. Considerando que a constituição estática vê as coisas em seus tipos e organiza-os em uma ordem hierárquica síncrona, a constituição genética examina a estruturação

de uma forma temporal, considerada ativa, quando novos objetos são constituídos na consciência, pelo raciocínio, por exemplo. Mas os objetos são também encontrados como já formados, como objetos culturais (ex.: martelos, tabelas, obras de arte, e assim por diante). Como é que nós percebemos estes objetos de forma imediata e em uma compreensão única? A recepção passiva desses objetos tem sua própria história constitucional, e isso é o que é coberto pelo termo 'gênese passiva'. Gênese passiva é, por assim dizer, a história de uma série de atos de síntese passiva”.

Os aspectos passivo e ativo da gênese são amplamente discutidos por Husserl em “Meditações Cartesianas”, mais especificamente na quarta delas. Sandmeyer (2007) reforça dizendo que, no sentido passivo, ocorre a “constituição” temporal da consciência na dinâmica da intencionalidade, i.e., da sensação dos dados na consciência em um processo de intencionalidade.

A *Arquitetura Genética da Informação* se aplica a esta abordagem de estudo da fenomenologia, ao considerar o aspecto ontológico da informação, quando nos deparamos com os objetos totalmente formados na realidade, independentemente do sujeito; e o aspecto genético de constituição destes objetos na consciência do observador.

### 2.4.3 Intencionalidade

Considerando-se o ser-humano consciente como sujeito da intencionalidade, a *informação intencional* é produzida no processo de intencionalidade fenomenológica, cuja entrada na EBOonline (2012a) diz:

“Intencionalidade (fenomenologia) é a característica de um sujeito estar consciente de alguma coisa, i.e., o seu direcionamento para um objeto, habilitando-o a lidar com o problema da **imanência** e **transcendência**, pela relação entre o que está dentro da consciência e o que se estende além dela. Muitos outros filósofos alegam que um objeto experienciado, representado e lembrado (por exemplo, uma árvore) está dentro da consciência (imane), enquanto que o objeto real em si mesmo está fora da mente (transcendente). Enquanto estes outros filósofos fazem desta distinção o terreno da dúvida (ceticismo) sobre a existência das coisas e da possibilidade de seu conhecimento, **os fenomenologistas distinguem imanência de transcendência no nível do significado**, ou seja, no nível ontológico ou reflexivo. Eles, então, colocam a existência entre parênteses (*bracket*), i.e., excluem da consideração a questão da existência ou não das coisas, em um processo de redução fenomenológica, lidando exclusivamente com o que está indubitavelmente presente na consciência. Assim, nesse nível, o imane é o que é dado efetivamente (por exemplo, o observador vê a frente de uma árvore) e o transcendente é todo o objeto intencionado (a árvore). Daí, o problema de como mover-se do imane para o transcendente é resolvido por uma análise de **como um objeto passa a ter significado na consciência** e de como ela se relaciona com o objeto, procedimento chamado de análise intencional, ou **análise da constituição de significado**.”

Os aspectos de imanência e transcendência são considerados no tópico 4.1, em associação ao genótipo e genoma do objeto, respectivamente.

A *informação intencional* é a informação significativa que tem a sua gênese dentro desse fluxo de experiências individuais de intencionalidade, podendo, com base em suas propriedades

(genéticas), gerar mudanças comportamentais, funcionais ou estruturais no sujeito.

Bayne e Montague (2011) explicam que:

“Há, obviamente, concepções múltiplas de conteúdo intencional. Alguns teóricos concebem conteúdo intencional em termos de objetos e suas propriedades, outros concebem-na em termos de estados de coisas ou conjuntos de mundos possíveis, e outros ainda tomam o conteúdo como envoltura dos modos de apresentação. Para complicar ainda mais as coisas há o fato de que muitos filósofos permitem que o pensamento e a percepção possam ter múltiplas camadas de conteúdo.”

Dennett (1995) desenvolve raciocínio – com certa insinuação de ironia— de que a nossa intencionalidade deriva da intencionalidade dos nossos genes, uma questão de granularidade:

“Com efeito, nós somos artefatos, projetados ao longo da eternidade, como máquinas de sobrevivência de genes que não podem agir prontamente e ‘informadamente’ em busca de seus próprios interesses. Nossos interesses como os concebemos e os interesses dos nossos genes também podem divergir – muito embora, não fosse pelo interesse de nossos genes, nós nem existiríamos: a sua preservação é a nossa ‘razão de ser’ original, mesmo que pudéssemos aprender a ignorá-los e desenvolver o nosso próprio ‘bem maior’... Portanto, a nossa intencionalidade é derivada da intencionalidade dos nossos genes ‘egoístas’! Eles são involuntariamente malvados, não nós!”

#### 2.4.4 Informação Intencional

O nosso pressuposto básico de *informação intencional*, ou *informação significativa*, nas palavras de Reading (2011), “nos oferece uma ponte entre a biologia, a mente e o comportamento”:

“*Informação Significativa* é definida como um padrão de matéria organizada ou energia que é detectada por um receptor animado ou manufaturado e, assim, desencadeia uma mudança no comportamento, funcionamento, ou na estrutura organizacional da entidade receptora – que pode ser uma macromolécula, uma célula, um organismo, uma planta, um animal, ou um dispositivo fabricado.”

Em seguida, o autor explica que “*informação significativa* não é, em si mesma, uma propriedade dos padrões de matéria ou energia, mas algo atribuído pela entidade receptora [o sujeito, o observador] a esta propriedade”.

Ainda, ele reconhece que é difícil compreender como entidades desprovidas de intenções conscientes possam gerar sinais e mensagens intencionalmente. Então, adota uma abordagem de equivalência entre intencionalidade e automação:

“Abordagem esta, em que os sinais nesses organismos são gerados automaticamente... e o fato das máquinas assumirem o papel de agentes intencionais leva-nos a perceber que o comportamento proposital e orientado a metas dos seres humanos e de outros animais também pode ser explicado em termos de causas mecânicas<sup>2</sup>”.

Por fim, ele explica que os conceitos de *informação significativa*, *informação semântica* e *informação pragmática* são similares, “em que a informação não apenas tem um propósito na

<sup>2</sup>O Funcionalismo filosófico reconhece a possibilidade de máquinas terem estados mentais, onde a psicologia do sistema depende não da matéria que ele é feito, mas como o material é arranjado no conjunto” McIntyre (1986)

parte do emissor, mas também tem um significado no receptor de maneira a eliciar a mudança desejada”, como é o caso do programa genético, ou “programa de desenvolvimento, onde os genes orientam as divisões celulares e o amadurecimento do embrião em desenvolvimento” (ENGELN, 2010).

Stonier (1990) explica que os computadores, como os livros e gravações de gramofones podem armazenar informação. Os computadores podem também processar informação: A informação de saída de um processo pode ser totalmente diferente da informação de entrada, dependendo das instruções do programa. Assim, se não houver uma “ânsia antropocêntrica para rejeitar a possibilidade de que os computadores participem de uma forma rudimentar de pensamento”, entenderemos que, de fato, os computadores armazenam e processam informação.

Mais do que isso, os computadores podem ter “inteligência” sofisticada e responder a questões complexas que envolvam o cruzamento de bancos de dados distribuídos.

“Assim como há diferentes formas de energia, também há diferentes formas de informação. . . humanas e não-humanas. . . Os sistemas para armazenamento e processamento de informação contidos dentro do cérebro humano são tão complexos e tão misteriosos que constituem a última grande fronteira das ciências biológicas” (STONIER, 1990)

Sterelny e Griffiths (1999) afirmam que há dois tipos de informação: a *informação causal* e a *informação intencional*, quando entendemos a ideia de genoma como sendo um programa.

“A noção de informação causal deriva da Teoria Matemática da Comunicação. . . e o outro conceito, relativo à *informação intencional*, também chamada de *informação semântica*, contém os objetos do pensamento, sendo a melhor candidata para o sentido de os genes carregarem informação desenvolvimental, e nada mais do que isso”.

Roederer (2005) argumenta que “na biologia, o uso de termos da informação implica **intencionalidade**, em que tanto a forma do sinal como a resposta a ele evoluem pela seleção. Onde um engenheiro vê projeto, um biólogo vê seleção natural”.

Com relação à informação, Siqueira (2012) apresenta argumentos ontológicos e fenomenológicos. Ontologicamente, na constituição dos entes (seres e coisas), tem-se o que ele denomina *informação absoluta*, cuja realidade é inacessível ao observador. Fenomenologicamente, a constituição da *informação intencional* ocorre na aparência do objeto à consciência do observador (na relação sujeito-mundo), tornando-se *dado*. Ou seja, enquanto a *informação absoluta* se caracteriza pela determinação dos estados entre os entes e as relações, compondo a “realidade” essencial, dependente de contexto, mas independente do sujeito; a *informação intencional* é definida pela aparência desta realidade ao sujeito, no fenômeno de observação.

Rakova (2006) argumenta sobre a auto-consciência do sujeito, a possibilidade da falácia fenomenológica, o fenômeno, ‘conceitos perceptivos’ e ‘consciência perceptiva’:

- Auto-consciência: É a consciência de si mesmo. Na Fenomenologia isto tem o significado de estar ciente de si mesmo como sujeito de experiências e como autor de pensamentos e ações, vendo os estados mentais como inseparáveis da genuína consciência. Por outro

lado, a filosofia analítica concentra-se na questão semântica de crer em si mesmo e, também, na questão do auto-conhecimento. Questões mais recentes se preocupam com o que de fato produz um estado mental consciente.

- Fenômeno: É um acontecimento observável, particularmente algo especial, literalmente “algo que pode ser visto” (gr. *phainomenon* = “observável”). Aparece ao indivíduo de tal maneira que produz estados mentais subjetivos (*qualia*).
- Falácia Fenomenológica: É a falácia de inferir partindo-se da aparência para a existência. Trata-se do erro de supor que quando as pessoas descrevem suas experiências elas descrevem as propriedades de alguns objetos mentais internos. Por exemplo, algo que parece verde para um observador correspondendo à existência em sua mente de algo verde.
- Conceitos Perceptivos: São conceitos fenomenais obtidos por nossos estados sensoriais (como sentimento de dor e sensação de cor, por exemplo). A ideia é de que a identidade entre os estados mentais e os estados cerebrais ocorrem necessariamente a posteriori, pela reidentificação das instâncias na imaginação e na introspecção. A existência de diferentes conceitos não implica a existência de diferentes propriedades. Referem-se diretamente às mesmas propriedades que são referenciadas pela correspondência com os conceitos físico-funcionais <sup>3</sup>.
- Consciência Perceptiva: É a experiência do fenômeno como um sentimento subjetivo particular, com características qualitativas individuais (*qualia*). Diz-se que se a experiência faz sentido para alguém, esta pessoa tem consciência perceptiva.

Podemos, então, formular que a *informação intencional*, ou *informação significativa*, é a constituição das propriedades dos conceitos perceptivos no sujeito (seres humanos, animais ou máquinas). Particularmente, no caso dos seres humanos – auto-conscientes –, esta informação tem caráter fenomenológico genético, como ato intencional da consciência perceptiva.

No tocante à informação como fenômeno intencional, diversos outros autores também são protagonistas dessa ideia, como veremos a seguir.

O argumento da Teoria Geral dos Sistemas e do Pensamento Sistêmico é que para dar sentido à complexidade do mundo, precisamos olhar para ele em termos de totalidades e relacionamentos, ao invés de dividi-lo em partes, olhando para cada uma isoladamente. Assim, sob uma perspectiva fenomenológica, as pessoas moldam as ideias com base na interpretação da informação (RAMAGE; SHIPP, 2009) e (FURNER, 2010).

Maturana (1978) explica a informação como fenômeno considerando-a participante de um ciclo fechado e auto suficiente de estruturas que são capazes de transmitir algo compreensível.

---

<sup>3</sup>Questão aberta: Por que dois conceitos distintos podem referenciar a uma mesma propriedade? (CHALMERS, 2004)

A informação é considerada um fenômeno quando possui um contexto específico, sofre mudanças contínuas para manter o seu significado dentro do contexto e mantém a sua identidade por meio de estruturas que foram articuladas para a sua criação.

Bertram **Brookes** (1980) propõe uma extensão e correção da ontologia dos três mundos de Popper para admitir especificamente o conceito de informação e suas relações com o conhecimento subjetivo e objetivo, uma vez que constata que “Popper, infelizmente, ignora o conceito de informação”, talvez porque ele erroneamente identifique informação como sendo os “dados dos sentidos” (**BROOKES, 1980**).

**Losee** (1997) tentou unificar diferentes abordagens de conceitualização da informação e propôs uma visão de informação “independente de domínio” que explicitamente inclui diversas áreas do conhecimento. O princípio de unificação de sua teoria é baseada no pressuposto da *origem* da informação. Ele afirma que o fenômeno fundamental da informação é de que “todos os processos produzem informação”.

**Siewert** (1998) argumenta que a representação mental do conteúdo observado em um fenômeno ocorre pelas experiências sensoriais e são passíveis de uma posterior avaliação de acurácia com base na representação das “propriedades do fenômeno”. A capacidade de julgamento da veracidade de tais propriedades depende de informações adicionais sobre o estado do mundo (**CHALMERS, 2004**).

**Dervin** (1999) entende que a informação se constitui de adições ou mudanças mentais “pela construção de padrões emergentes de organização”.

Ele é o precursor da ideia fenomenológica de “emergência do sentido” (**DERVIN, 1983**), apresentada como elemento de entrelaçamento no centro do diagrama da figura 15, na página 76. Na seção 4.1, discorreremos sobre ‘informação estruturada’, cujas propriedades são o sentido e a semântica e, também, na aplicação da seção 9.4 retomamos esse assunto.

**Herold** (2001) realiza estudos fenomenológicos sobre a natureza da informação, explicando que ela “exerce um papel entre a mente e a matéria, uma interação entre mentes separadas e, também, entre meios físicos... estando ligada à cognição, mente, consciência, evolução e com a identificação da própria vida”.

**Floridi** (2002) expressamente caracteriza a Filosofia da Informação como um ramo da filosofia que combina interesses fenomenológicos de primeira-ordem (a informação) e de segunda-ordem (metodologias e teorias), sendo o fenômeno o componente principal dos estudos sobre informação.

Ele alerta que a informação “é notoriamente um fenômeno polimórfico e polissêmico” (**FLORIDO, 2005**). Mais recentemente, **Floridi** (2007) ainda propõe a sua própria definição de informação como sendo um conjunto constituído de dados “bem-formados” (sintaxe), significativos (semântica) e com uma função de verdade (veridicalidade).

**Marijuán** (2004) explora “uma nova compreensão do fenômeno da informação baseado na organização molecular da vida”. A ideia central é que o inter-relacionamento entre os elementos cruciais da bioinformática podem proporcionar aspectos fundamentais para uma nova síntese



da informação, “representando um microcosmo do fenômeno completo da ’in-formação’”.

Para [Derqui \(2005\)](#), o fenômeno da informação pode ser observado por meio do estabelecimento de diversas estruturas que interagem entre si por meio de articulações.

[Klir \(2006\)](#) propõe uma Teoria Geral da Informação (TGI) fundamentada no fenômeno da informação baseado na incerteza. A TGI está intimamente conectada com duas generalizações na matemática: a da teoria das medidas e a da teoria dos conjuntos. Ele argumenta que nossa atitude intuitiva na vida implica um processo de “tomada de decisões consciente, que talvez seja a mais fundamental capacidade do ser-humano. . . e para que se entenda esta capacidade, é necessário entender a noção de incerteza, primeiramente.”

[Bawden \(2007\)](#), também em busca de uma TGI, entende que a própria vida é um fenômeno da informação. Ele ressalta que esta abordagem é uma tendência devido à crescente ênfase que se dá à compreensão da informação genética.

Sua proposta é uma teoria unificada da informação, contemplando a informação nos domínios da física, biologia e humano e estabelecendo as ligações e interações entre eles, viabilizando as ideias de significado, compreensão e conhecimento.

Impressionado, ele diz:

“O espaço aparentemente vazio à nossa volta está fervendo com informações. Estamos impossibilitados de tomar consciência de grande parte delas porque os nossos sentidos não respondem a elas ou porque temos coisas mais interessantes a atender. Mas não temos o direito de ignorá-las se estamos em busca de uma Teoria Geral da Informação. Não podemos viver apenas lendo e escrevendo livros” ([BAWDEN, 2007](#)).

[Hofkirchner \(2009\)](#) e [Fuchs \(2008\)](#) recentemente vêm despendendo esforços em vista de uma teoria unificada da informação, participando do grupo de pesquisa *Unified Theory of Information Research Group* da Universidade de Salzburgo (Áustria). Fuchs argumenta que, em termos dialéticos (maneira de filosofar que procura a verdade por meio de oposição e conciliação de contradições), “os fenômenos da informação são analisados em termos estruturais, continuidade e descontinuidade, um e muitos, potencialidade, global e local, virtual e local, otimismo e pessimismo, essência e existência, imanência e transcendência etc”. Ele acrescenta que uma teoria da informação que é dialética e realista identifica tendências antagônicas do fenômeno da informação, onde a informação é concebida como um processo contraditório, dinâmico e complexo, que desenvolve e produz resultados, sendo vista como parte do mundo material e pode ser apreendida, descrita e analisada. Com respeito à teoria crítica, esta tem um foco fenomenológico social da informação em contextos de dominação, relações assimétricas de poderes, exploração, opressão e controle, como objetos de estudo. Visando uma teoria unificada da informação, ele conclui dizendo que “há numerosos fenômenos da informação e, portanto, numerosos sub-domínios e sub-teorias”.

[Sagüillo \(2009\)](#) diz que “um dos múltiplos significados do termo ’informação’ é dado implicitamente nos postulados e condições da Lógica Teórica da Informação (I-T-L) baseando-se na tradição de considerar instâncias da informação como um fenômeno lógico”.

[Zaliwski \(2011\)](#) afirma que a definição do termo “informação” deve partir do fenômeno.

O seu argumento fenomenológico é de que a informação é relativa ao observador e é criada subjetivamente, o que não impede que ela exista objetivamente antes ou após a sua percepção. Então, sob essa perspectiva, podemos ter a informação absoluta no nível ontológico e a *informação intencional* no nível fenomenológico.

### 2.4.5 Arquitetura da Informação Intencional

Siqueira (2012) realizou interessante estudo sobre a fundamentação e caracterização da Arquitetura da Informação (proposta de Brasília) como disciplina científica.

Das diversas abordagens realizadas em relação ao problema da demarcação científica, ele encontra um conjunto comum do *corpus* de conhecimento científico, caracterizado por:

“um conjunto de fenômenos que devem ser explicados; um sistema de hipóteses/teorias que delimitam a investigação e que a orientam; ao menos um método de investigação que direciona como as hipóteses/teorias devem ser articuladas para observar, experimentar, explicar e prever resultados”.

Ele apresenta cinco postulados que fundamentam a Arquitetura da Informação como disciplina científica, a saber:

1. Existem Entes;
2. Existem relações entre os Entes. Todas as possíveis relações entre os Entes são denominadas Relações Ontológicas.
3. Dentre os Entes que existem, alguns possuem a especial propriedade de serem capazes de intuir as suas relações com outros Entes, são chamados Seres.
4. A Realidade é composta pelos Entes e suas Relações Ontológicas.
5. O Ser não tem acesso direto ao que o Ente é (em essência), apenas tem acesso ao modo como o Ente lhe aparece na percepção. A esta aparência do Ente para o Ser denomina-se Fenômeno.

A figura 2 apresenta o diagrama da “Arquitetura da Informação Intencional”, proposta por Siqueira (2012), que se constitui o embasamento para a *Arquitetura Genética da Informação*.

A disciplina da Arquitetura da Informação é definida como sendo “a investigação do Mundo considerando as suas manifestações, formas, contextos e significados do ponto de vista de um sujeito”. O problema está relacionado a “perceber, pensar, desenhar e habitar espaços de informação”. Os objetos de investigação são “os fenômenos da configuração do espaço da informação. O método é “fenomenológico, para investigar as distinções do sujeito, do fenômeno e do conhecimento, na construção de configurações do espaço da informação”.

As categorias da disciplina de Arquitetura da Informação, então, são: Forma, Contexto, Manifestação e Significado (FCMS).



Figura 2: Arquitetura da Informação Intencional

1. Forma - É uma estrutura que associa os dados do fenômeno, passando a identificar um objeto na experiência do sujeito;
2. Contexto - Caracteriza uma coleção de relações do fenômeno com o mundo, do ponto de vista do sujeito;
3. Manifestação - É uma coleção de dados no espaço dos objetos que um sujeito é capaz de perceber como um fenômeno;
4. Significado - É a intenção designada por um sujeito para supor um objeto.

Essas categorias, quando combinadas duas a duas, delimitam simultaneamente espaços conceituais de naturezas distintas – subjetividade, materialização, realização, semântica, informação e arquitetura – conforme apresentados por [Albuquerque \(2010\)](#):

- Espaço da Subjetividade: Significado e Forma da natureza do objeto da percepção;
- Espaço da Materialização: Forma e Manifestação da natureza do objeto da materialização;
- Espaço da Realização: Manifestação e Contexto da natureza do objeto da realização;
- Espaço da Semântica: Contexto e Significado da natureza do objeto da semântica
- Espaço da Informação: Manifestação e Significado da natureza do objeto da informação;
- Espaço da Arquitetura: Forma e Contexto da natureza do objeto da arquitetura.

Retomaremos as categorias FCMS e suas relações na seção 4.1, na apresentação dos elementos da *Arquitetura Genética da Informação*.

### 2.4.6 Visão de Mundo

Com base nos postulados da Arquitetura da Informação Intencional (item 2.4.5, a nossa Visão de Mundo, discutida em fórum colaborativo no Portal do CPAI é a seguinte:

“O Mundo é composto por Entes (Seres e Coisas) que se relacionam entre si, cujo Conhecimento é somente acessível intencionalmente por sua aparência (não por sua essência), sendo, portanto, passível de concordância, questionamento e falseamento”.

A distância entre aparência e essência diminui em função de uma condição voluntária de suspensão de julgamento (*bracketing*, ou *epoché*).

Em síntese e orientados a essa visão de mundo, temos que:

1. O tema desta pesquisa é a genética da informação intencional, incluindo a sua gênese e caracterização genética;
2. Informação intencional é a informação significativa gerada no processo de intencionalidade;
3. Intencionalidade (fenomenológica) é um processo de direcionamento para um objeto, em uma experiência na primeira pessoa, tornando o sujeito (observador) consciente de alguma coisa, que particularmente lhe aparece de forma significativa;
4. A Fenomenologia Genética estuda a gênese do significado das coisas dentro do fluxo das experiências individuais de intencionalidade, ou seja, da informação intencional ou significativa, que tem a sua gênese nesse processo.

## **Parte II**

### **Fundamentação Teórica**



## 3 Revisão de Literatura e Fundamentos

*“Na verdade, o efeito poderoso do Novo Testamento sobre minha vida, aos 23 anos de idade, é que geraram em mim um impulso para descobrir o caminho para Deus e para uma vida verdadeira, através de uma investigação filosófica rigorosa.”*

— Edmund Husserl

Em busca do estado da arte, revisão teórica e revisão histórica, como propõe Luna (1998), apresentamos informações sobre o que se sabe sobre o tema, as lacunas existentes e os principais entraves teóricos e metodológicos em seu desenvolvimento histórico.

Para cobrir todos os aspectos envolvidos na *Arquitetura Genética da Informação*, esta revisão teórica está organizada em sete seções:

1. “Arquitetura” - Seção 3.4;
2. “Genética” - Considerando a gênese da *informação intencional* na seção 3.2, o conhecimento no âmbito da fenomenologia, na seção 3.5, a genética propriamente dita nas seções 3.6 e 3.7 e a base para uma ontologia da genética da informação na seção 3.5;
3. “da Informação” - A natureza ontológica da informação na seção 3.1, a fenomenologia da *informação intencional* na 3.2 e os conceitos epistemológicos de informação na seção 3.3;

### 3.1 Ontologia e Informação

Ontologia é o ramo central da Metafísica. Para Aristóteles, a metafísica é a ciência do ser, a filosofia primeira que analisa os métodos e as premissas das ciências particulares, estuda o ser enquanto ser, analisando de modo totalmente abstrato a noção de realidade.

Em nosso objeto de estudo, temos, então, que considerar a metafísica da informação para questionarmos: (1) É a informação um ente ideal? (2) Ela é aquilo que é, em si mesma? (3) O verdadeiro ser da informação coincide com a sua substância, ou seja, com aquilo que não pode deixar de estar presente, sob pena de a informação transformar-se em alguma outra coisa diversa?

Buchler (1955) explica que para Pierce, na Metafísica, informação é a conexão cognitiva entre a forma e a matéria, onde as cognições são logicamente derivadas por indução e hipótese

de prévias cognições que são menos gerais, menos distintas e das quais tem-se menos consciência. Estas, por sua vez, também são derivadas de outras também menos gerais, menos distintas e menos vívidas e assim por diante até que se chegue a uma situação primeira, ideal, singular e completamente fora da consciência. Ele conclui dizendo: “Esse ideal primeiro é a coisa-em-si e não pode existir como tal, no sentido de não ser em relação à mente”. Então, a origem primeira da informação como concepção da realidade não tem limites definidos e o seu estado ideal só poderá ser conhecido em um momento futuro pela combinação cognitiva de vários observadores.

Marroquín e Italo (2007) tratam da questão de uma natureza extra-ontológica e exclusivamente criada e racional das qualidades; a realidade ôntica da essência, questionando se esta é um ente ideal.

Aristóteles também defende que dada a regularidade dos fenômenos naturais nada acontece por acaso ou por acidente. Então, a substância de qualquer ser não é uma questão de casualidade, mas, pelo contrário, coincide necessariamente com sua causalidade, perceptível em quatro dimensões: uma causa material, uma causa eficiente, uma causa formal e uma causa final. Nicola (2005) lembra que as duas últimas causas são as mais importantes. Com relação à informação, traduziremos cada uma destas causas:

1. Causa material: a substância concreta de que a informação é composta;
2. Causa eficiente: o processo que produziu a informação;
3. Causa formal: a estrutura interna que determina a realidade da informação; e
4. Causa final: a finalidade para a qual a informação surgiu ou foi produzida.

Lima-Marques (2011) considera a informação como coisa física, e, portanto, com uma causa material, i.e, substância concreta essencial diretamente associada ao objeto em si, ontologicamente. Esta compreensão da informação absoluta é contemplada na *Arquitetura Genética da Informação* no espaço dos objetos, à parte de qualquer fenômeno de observação, cuja complexidade de propriedades intrínsecas compõe o seu genoma, conforme veremos na seção 4.1.

## 3.2 Fenomenologia

Tradicionalmente, a filosofia inclui quatro áreas ou disciplinas centrais: ontologia, epistemologia, ética e lógica. Smith (2011) propõe incluir a fenomenologia como quinta disciplina:

1. Ontologia é o estudo dos seres ou de seus estados – O que é?
2. Epistemologia é o estudo do conhecimento – Como podemos saber?
3. Lógica é o estudo do raciocínio válido – Como devemos raciocinar?



4. Ética é o estudo do certo e errado – Como devemos agir?
5. Fenomenologia é o estudo de nossa experiência - Como podemos experienciar?

### 3.2.1 O que é Fenomenologia?

No século 18, o filósofo e matemático suíço-germânico Lambert aplica o termo 'Fenomenologia' para fazer a distinção entre a verdade e as ilusões e erros. No século 19, a fenomenologia se tornou associada principalmente à mente, por Hegel, que traçou o desenvolvimento do espírito humano partindo de meras experiências sensoriais para o 'conhecimento absoluto'. No século 20, se deu a sua origem como movimento filosófico, com o objetivo principal de uma direta investigação e descrição de fenômenos experienciados na consciência, livre de possíveis preconceitos e pressuposições. A partir de então, esse movimento filosófico começou a, efetivamente, seguir o seu curso. Mas, há tantas abordagens diferentes que uma caracterização mais abrangente do tema requer que sejam consideradas, não sendo fácil encontrar um denominador comum entre elas, excetuando-se o fato de que, as que se enquadram dentro da filosofia tiveram todas a mesma origem em Husserl.

[Smith \(2011\)](#) explica que a fenomenologia é o estudo de estruturas da consciência em relação a experiências a partir da primeira pessoa. A estrutura central de uma experiência é a sua intencionalidade direcionada em condições adequadas para um objeto, em virtude do seu conteúdo ou significado (que o representa).

[Lacey \(1996\)](#) explica que a fenomenologia é, literalmente, a descrição ou o estudo das aparências. Qualquer descrição de como as coisas aparecem, especialmente se as aparências forem sustentadas pela e penetrantes na consciência, pode ser chamada de fenomenologia. Não se trata apenas de uma descrição da experiência, mas dos objetos da experiência, que Husserl chama de fenômenos.

[Moran \(2000\)](#) argumenta que a fenomenologia é um estilo radical e anti-tradicional de filosofar, que enfatiza a tentativa de se chegar à verdade das questões pela descrição dos fenômenos em um sentido amplo: como eles aparecem, de que forma aparecem, e como se manifestam na consciência do observador.

Como tal, o primeiro passo da fenomenologia é de procurar evitar todos os equívocos e imposições colocados na experiência por antecedência, quer sejam provenientes de tradições religiosas ou culturais, do senso comum do cotidiano, ou mesmo da própria ciência. Quaisquer explicações não devem ser aplicadas antes de os fenômenos terem sido entendidos a partir de dentro deles mesmos.

[Landau, Szudeck e Tomley \(2011\)](#) entendem a fenomenologia como uma abordagem da filosofia que investiga os objetos da experiência (conhecidos como fenômenos) apenas na medida em que eles próprios se manifestam à nossa consciência, sem fazer qualquer suposição sobre sua natureza como algo independente". "Fenômeno é a experiência que é imediatamente presente. Se olho para um objeto, o objeto, experimentado por mim, é um fenômeno.

O objeto da fenomenologia são os fenômenos perceptíveis, tornando sujeito e objeto inseparáveis. Busca alcançar a intuição das essências pela interpretação do mundo através da consciência de um sujeito através de suas próprias experiências de intencionalidade:

- A *consciência* é formada por atos de percepção, imaginação, paixão, emoções e demais atos mentais do homem.
- A *essência* de um determinado fenômeno pode ser obtida por meio do processo de redução fenomenológica.
- *Intencionalidade* é o direcionamento de todas as atividades conscientes para ou sobre um objeto. A intencionalidade possui o componente do ato intencional e do objeto intencionado. As intenções não são meramente mentais, antes são públicas e, em princípio, estão disponíveis à experiência de todos, no mundo. Isto é verdade mesmo que os objetos não existam fisicamente (KHALIFA, 2011).

Marina Rakova (2006) explica que, a fenomenologia, como teoria da consciência é centrada em fenômenos cognitivos em termos de intencionalidade e de representação. O seu princípio fundamental é a impossibilidade de haver consciência desassociada de um conteúdo. Ela explica que a fenomenologia estuda a natureza dos fenômenos mentais (pensamento, crenças, desejos, sensações, percepções, volições, emoções etc), sustentando que a essência da mente é a consciência, com ênfase na perspectiva da primeira pessoa.

Em suma, os componentes amplamente considerados na fenomenologia são: fenômeno; estruturas da consciência; experiência; observador (primeira pessoa); intencionalidade; contexto (condições); objetos da experiência; e representação do objeto (conteúdo e significado).

### 3.2.2 Desenvolvimento Histórico-Conceitual da Fenomenologia

A Fenomenologia, como movimento histórico, é exemplificada por uma grande gama de pensadores extraordinários. Tem em Franz Brentano (1838-1917) o seu precursor direto e foi fundada como ramo da Filosofia por **Edmund Husserl** (1859-1938), com importantes contribuições de **Martin Heidegger** (1889-1976), Hans-Georg Gadamer (1900-), Hannah Arendt (1906-1975), Emmanuel Levinas (1906-1995), Jean-Paul Sartre (1905-1980), **Maurice Merleau-Ponty** (1908-1961) e, mais recentemente, de Jacques Derrida (1930-). Assim, a Fenomenologia não deve ser entendida como um método ou projeto filosófico, mas primariamente como um conjunto de pessoas direta ou indiretamente ligadas às ideias de Husserl, como apresentado na linha do tempo da figura 3, página 27.

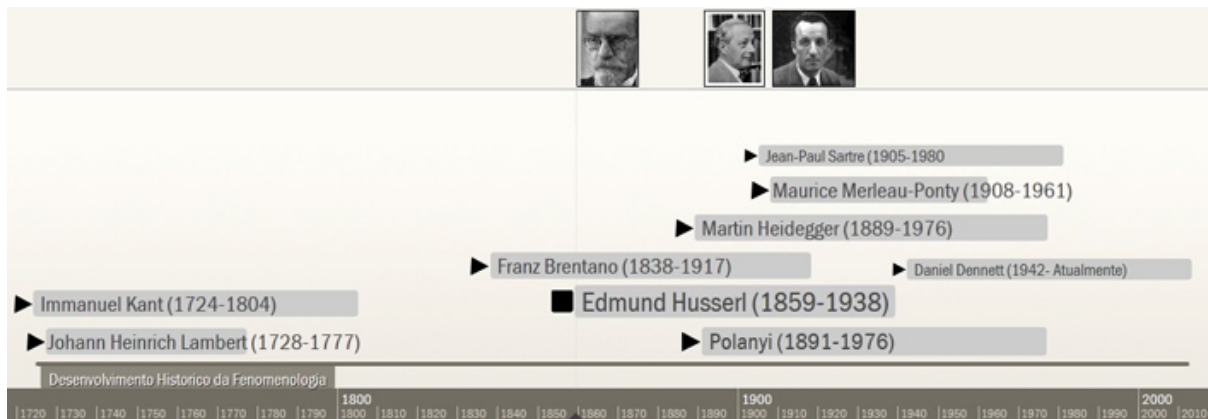


Figura 3: Linha do Tempo de Influência dos Pensadores da Fenomenologia - Fonte: (PEREIRA-JR, 2011)

Com isso em mente, vejamos resumidamente as influências pessoais de cada um de seus mentores, desde os precursores, passando pelo seu fundador e demais contribuintes, até a sua crítica nos dias atuais.

### 3.2.2.1 Precusores: Lambert, Kant e Brentano

Lambert foi o primeiro a empregar o termo “fenomenologia”, colocando-o no título da quarta sessão de seu *Novus Organon*, com o significado de *ciência da aparência*, que nos permite proceder da aparência para a realidade, analogamente aos estudos de perspectiva em ótica, que nos permitem deduzir as características reais de um objeto sob observação.

Kant tinha uma visão fenomenológica baseada no idealismo transcendental, afirmando que tanto a razão como a experiência são necessárias para compreender o mundo (os racionalistas acreditavam que o uso da razão leva à compreensão do mundo enquanto os empiristas acreditavam que o conhecimento provém da experiência dos objetos do mundo).

Esta abordagem fenomenológica lidava com as coisas em suas manifestações, como por exemplo, o movimento relativo, as cores e outras propriedades que dependiam de um observador humano.

“O fato de Kant localizar *a priori* até mesmo em nossas intuições sobre o mundo foi importante para os fenomenologistas do século XX... , que procuraram investigar os objetos da experiência independentemente de quaisquer suposições que possamos ter a respeito deles”. (LANDAU; SZUDECK; TOMLEY, 2011).

Brentano foi um filósofo alemão que lançou as bases contemporânea da filosofia da mente por reintroduzir a noção de intencionalidade ou a mente direcionada para um objeto<sup>1</sup>.

Rakova (2006) ressalta que, hoje, qualquer que seja a teoria da intencionalidade, esta deve abordar o problema levantado por Brentano, qual seja: “Como se estabelecem os estados men-

<sup>1</sup>Intencionalidade é uma doutrina que diz que cada ato mental se relaciona a algum objeto e/ou ideia que se nos apresentam, sejam eles imagináveis, possíveis ou impossíveis

tais sobre as coisas?”. A tese de Brentano é de que os estados intencionais não são redutíveis a estados físicos. Ele tem uma noção psicológica da intencionalidade, em que os atos mentais e conscientes são uma propriedade da consciência voltados para os objetos e, também, para si mesmos, como objetos secundários.

Assim, Brentano desenvolveu uma ciência descritiva da consciência, considerando um tipo peculiar de auto-evidência de estados mentais que poderiam convergir para verdades. Descreveu a psicologia descritiva e a intencionalidade, que se tornaram o fundamento do pensamento de Husserl sobre a Fenomenologia.

Moran (2000) explica que Brentano considerava impossível a atividade da consciência referir-se a algo que não lhe fosse apresentado, sendo, então, a apresentação um ato necessário de visão mental de um objeto ou conceito individual, ou mesmo uma relação abstrata (ideia) complexa. Uma apresentação, então, proporciona o objeto ou conteúdo básicos em torno do qual outros tipos de atividades mentais se cristalizam.

### 3.2.2.2 Fundador: Husserl

Husserl (1859-1938), filósofo alemão, fundou a Fenomenologia. As obras mais relevantes de Husserl foram “*Idéias para uma Fenomenologia Pura*” (1913), *Uma Filosofia Fenomenológica* (1913), e *Investigações Lógicas*. Ele preparou a Fenomenologia como uma nova ciência da percepção e das essências, cujos elementos essenciais são a intencionalidade (a consciência é intencional, direcionada e tem propósito), a redução fenomenológica (quaisquer pressuposições devem ser suspensas no exame da amplitude de diferentes dimensões da experiência) e a impossibilidade de apreensão da totalidade das propriedades do objeto. A complexidade da condição humana envolve a percepção da realidade em determinado ponto no tempo, no espaço, dentro de um contexto social e em determinado estado psicológico. O programa de Husserl é sobre uma ciência que pode resultar em conhecimento, com foco em Epistemologia.

Ele considerou que o objetivo da fenomenologia é descobrir os pré-requisitos da experiência em um estudo descritivo da consciência na primeira pessoa (HUSSERL, 2001). Intencionalidade, ou “consciência de”, é a criação de sentidos nos atos mentais que transcendem o que se oferece na experiência em si, podendo ser estudada em uma abstração da existência real das coisas. Assim, tanto a percepção verídica como a intencionalidade apresentam alucinação.

Em “*Idéias*”, sua principal obra, ele descreve a fenomenologia como uma ciência pura das essências, que se relaciona com a psicologia empírica, assim como a geometria se relaciona com as ciências naturais. A atitude fenomenológica envolve o exercício da “epoché”, i.e., a anulação sistemática de todos os pressupostos sobre a realidade externa, de modo que a redução fenomenológica possibilite revelar a essência pura dos processos mentais.

Segundo Rakova (2006), Husserl afirma que todo processo intencional um ato de auto-concessão que estrutura momentos hiléticos (causa material do objeto) por componentes noéticos (dados do objeto, antes de sua interpretação pelos sentidos, ou seja, sua compreensão

imediate). Paralelo à *noesis* temos o *noema*, que não é inerente aos processos mentais. O *noema* de um processo intencional inclui um conteúdo ou matéria (o “que” noemático) e seu modo de “oferecimento” ou qualidade (obtido na percepção, gosto, julgamento e certeza).

O problema central de Husserl era, então, o mistério da subjetividade. Fink (1939) resalta que a relação entre ser e pensar levou Husserl a uma profunda exploração dos mistérios da subjetividade e da questão de constituição da objetividade, i.e, a maneira como a consciência atinge o conhecimento objetivo. Nesta exploração Husserl teria passado por três estágios: Psicologismo (1887-1901); Fenomenologia Descritiva (1901-1913); e Fenomenologia Transcendental (1913-1938). esse último seria um estágio residual, centrado nos objetos abstratos, em que teríamos uma redução transcendental (*noemata*), como sendo aquelas entidades abstratas ou sentidos que determinam objetos de possíveis experiências subjetivas. Esse idealismo transcendental foi rejeitado por outros fenomenólogos, e em seu trabalho posterior Husserl enfatizou o papel da intersubjetividade na constituição dos objetos que compõem o nosso mundo peculiar na vida.

Pietersma (2000) reforça que “Husserl adota na epistemologia fenomenológica uma atitude cujo enfoque não está nos objetos, mas em sua experiência”.

Em sua fase de maturidade, Husserl resgatou um projeto que mantinha escondido, que reconhecia a necessidade de uma *fenomenologia genética*.

Moran (2000) argumenta que Husserl escreveu muito sobre a natureza e propósito do que intitulou “redução” (latin *reducere*, “levar de volta”).

Nas Husserlianas (coleção de obras do autor compiladas em 1964), vemos que Husserl introduziu o conceito de redução fenomenológica para excluir tudo o que é posto como transcendentalmente existente, falando de uma *redução epistemológica* necessária para focar no fenômeno puro de atos da consciência (“cogitações”) para evitar suposições enganosas sobre a natureza delas. Husserl tem em mente um *bracketing* específico de uma interpretação psicológica do que é dado no ato de conhecer algo, referindo-se a uma redução psicológica (HUSSERL, 1999). Ainda, em “Meditações Cartesianas” (1931), Husserl, Peiffer e Lévinas (1992) falam de uma *redução transcendental fenomenológica*. Em geral, fica difícil distinguir estes diferentes estágios e graus de redução.

Em 1954, Husserl (1970) considerou diferentes formas de redução: redução cartesiana, redução psicológica, redução de intencionalidade, redução de crítica às ciências naturais e redução ontológica, por exemplo. Em 1913, Husserl (1983) já via a necessidade de uma “teoria sistemática de reduções fenomenológicas”.

Husserl caracterizou a prática da *epoché* em diferentes formas: abstenção, deslocamento, desconexão ou exclusão do que está posto no mundo e nossa fé inquestionável normal da realidade que experimentamos. Ele fala em recusar, não dar atenção, abandonar, colocar entre parênteses, colocar fora de ação e deixar de lado todos os julgamentos sobre o mundo observado. Mas a característica essencial é ter sempre uma “mudança de atitude” de tal maneira que deixemos de lado qualquer suposição naturalista sobre o mundo, muito comum em nosso

comportamento diário.

Boudier (1986) explica que a mudança de orientação na prática da redução leva-nos a um “retorno” a um ponto de vista transcendental. A *epoché* é, então, apenas parte de uma estrutura de redução. Para Husserl, experienciar uma redução é experienciar um enriquecimento da vida subjetiva que se abre infinitamente nessa ocasião. Os que mantêm uma atitude natural têm o pensamento firmado em um mundo pré-conhecido, ao invés de moverem-se para a esfera da *epoché*. A experiência transcendental age constituindo o mundo para o observador, na consciência, embora desprovido do ato de reflexão ou juízo.

Na abordagem da redução cartesiana, Husserl frequentemente se refere a um primeiro procedimento de *bracketing* ou *epoché*, significando “cessação”, o que viria a ser um tipo de ceticismo essencial para o conhecimento. Os antigos cétricos recomendavam a suspensão de juízo diante de argumentos conflitantes, aparentemente de mesmo peso, ou seja, suportados pelo mesmo grau de evidência. Isso deixava a pessoa julgando em face de ambas alternativas com uma certa “indecidibilidade”. Daí a recomendação de praticar a abstenção de qualquer julgamento, i.e., a *epoché* cétrica. Sextus Empiricus diz em suas “Outlines of Pyrrhonism”: “A suspensão do intelecto é uma situação ou condição em que não há qualquer movimento ou atividade do intelecto, porque não se rejeita nem se aceita nada”, visando a tolerância e uma mente aberta. Com um diferente propósito, Husserl recomenda a *epoché* fenomenológica objetivando atos de consciência que realizam o *bracketing* de certas estruturas fundamentais.

Husserl usou várias analogias matemáticas para articular o seu senso de *epoché*: é como colocar parênteses em partes de uma equação (por exemplo,  $2+2=(8/4)+2$ ), sustentando que não se deve calcular previamente o que está entre parênteses. Observemos na figura 4, página 31, um relógio excêntrico cujos elementos isolados exigem interpretação e raciocínio. Porém, por um ato voluntário, podemos recusar qualquer interpretação dos dados observados, sem fazer correlações intencionais de associação com conhecimento prévio, focando apenas na estrutura do ato, ou seja, no fenômeno de observação e não em termos de um mundo espaço-temporal existente.

A atitude natural sempre emprega um ato *tético* (gr. *thesis*, proposição), que é um ato de suposição (tomada de posição). Na *epoché*, inibe-se a aceitação desse mundo objetivo, que também inclui o espaço e o tempo (HUSSERL, 1983) (obra de 1913). Esta suspensão de julgamento é um ato livre da mente em escolher alterar o seu ponto de vista.

“Daí, nesse ponto, nós falamos de tais dados absolutos; mesmo que estes dados sejam relacionados à realidade objetiva via suas intenções, caracteres intrínsecos que estão dentro deles; nada se assume com respeito à existência ou não-existência da realidade.”

Husserl explica que a consciência consiste de atos do ego e seus correlatos, as unidades de pensamento. Assim, temos uma tríade fenomenológica formada por ego, cogito e cogitata. O ego fenomenológico é o fluxo de consciência em que se adquire o significado e a realidade do meio em observação. Os atos do ego, denominados *cogito* são atos mentais de dúvida, compreensão, afirmação, recusa etc formados por unidades de pensamento, denominadas *cogitata*, que



Figura 4: Relógio Excêntrico (Ilustração do Bracketing)

se referem à representação dos objetos em observação.

Ele afirma que o seu tema é “mostrar que é possível e como é possível a tarefa da descrição de um método de redução fenomenológica”, processo em que a redução lida diretamente com a subjetividade transcendental. É um processo que parte de um ego individual empírico para o domínio de uma percepção profunda, i.e., parte do objeto dado à consciência para o conteúdo da consciência em si mesma. Ainda, “a redução parte de um mero julgamento empírico para aqueles que, por sua natureza, são corretos em si mesmos e apreendidos de forma adequada”.

Em um rascunho de artigo para a *Encyclopaedia Britannica*, Husserl sugere que “a *epoché* é um primeiro passo na redução, sendo a segunda a identificação, compreensão e descrição das unidades ideais de significado que agora surgem, sendo estas a noemata e a noese”. Noemata (plural de noema) são os objetos do pensamento; e noese é a intuição ou forma de juízo.

### 3.2.2.3 Seguidores e/ou Críticos: Heidegger, Merleau-Ponty, Sartre, Polanyi e Dennett

Glendinning (2007) explica que Martin Heidegger (1889-1976) foi um filósofo alemão, aluno de Husserl, conhecido por suas explorações fenomenológicas e existenciais, preocupando-se com uma ontologia fundamental, ou seja, a questão do ser.

Para ele esta era a única questão importante. Ele explorou os conceitos de mundo e de ambiente (trabalho não publicado de Husserl) e esteve especialmente interessado na discussão do reconhecimento da verdade com o problema do ser. Para ele, a fenomenologia é uma tentativa de entender os objetos do mundo como eles se manifestam a si mesmos e devia ser entendida como uma possibilidade. Brown, Collinson e Wilkinson (1998) enfatizam que o trabalho “Investigações Lógicas” de Husserl exerceu nele uma fascinação até o final de sua vida. Ele nunca aceitou a fenomenologia de Husserl em sua forma transcendental e idealística, mas encontrou nela um meio de prover um método ontológico. Após o seu doutorado, fortemente influenciado por Husserl, passou a desenvolver a sua própria visão de fenomenologia centrada nos fatos da

existência vivida, muito mais do que na consciência transcendental e no puro ego.

A obra mais relevante de Heidegger foi “Ser e Tempo” (1927). Ele propõe uma teoria do conhecimento com foco na Ontologia, conectando a hermenêutica com o “ser-no-mundo”, afirma Moran (2000). Explica o *dasein*, uma qualidade reflexiva sobre a existência e modificação do ser, que detém a possibilidade de enunciar tanto o ser como o modo de ser: o “ser-aí”, em que o ser humano é o único ente capaz de questionar a sua existência (existencialismo); o “ser-no-mundo”, o “ser-a-cada-momento” e o “ser-de-cada-vez”.

Merleau-Ponty (1908-1961), filósofo francês, teve na “Fenomenologia da Percepção” (1945) a sua obra mais relevante, com questões bem particulares: “Como os indivíduos percebem o mundo?” e “Como os indivíduos percebem as próprias percepções?”.

Ele explica a existência de um fluxo contínuo do objetivismo para o subjetivismo e a impossibilidade do dualismo cartesiano de separação entre os dois, em um turvamento da relação sujeito-objeto: Se “eu percebo”, então, “eu percebo algo”. Ele entende que a representação da intenção cognitiva de um indivíduo depende de uma linguagem, cujas palavras lhe sejam previamente conhecidas. Ele explica o *mitsein*: “o ser-com-os outros”, onde ocorre a passagem dos dados visuais simples da sensação para o significado psicológico em nossa experiência interior, levando-se em conta a vida mental dos outros, um todo carregado de significado imanente. O engajamento de uns aos outros ocorre não apenas por meio do ser, mas pelas percepções reveladas nos caminhos de intersecção das várias experiências, alcançando-se o conhecimento.

Para Merleau-Ponty, o lugar central da fenomenologia é a ideia de “personificação” ou “corporificação” (RAKOVA, 2006). Ele se opõe ao objetivismo por ser este incapaz de dar explicações a vários distúrbios da percepção e ações em termos de causas mecanicistas (como o fenômeno do “membro fantasma”). O próprio corpo, com sua intencionalidade motora, não o ego Cartesiano, dão significações ao mundo. Ação e percepção são inseparáveis, e a percepção é realizada pelo corpo, não como uma coisa mecânica, mas como consciência do lugar intimamente conectado com sua capacidade de direcionar os movimentos em direção aos seus objetivos ou metas.

De acordo com o empirismo, “os órgãos dos sentidos são estimulados a receber e transmitir *dados* que são, de alguma maneira, decodificados pelo cérebro de modo a reproduzir uma imagem do estímulo externo original” (SANTAELLA, 2012). Por um lado, temos os dados dos objetos do mundo e, por outro, a sensação e percepção mentais que geram uma representação mental significativa.

Merleau-Ponty (1962) argumenta que “tudo aquilo que é percebido, por sua própria natureza, está *preñe* de ambiguidade e pertence a um contexto, um campo, que lhe dá forma”. O retorno ao fenômeno nos leva a descobrir “um todo já *pregnante* com um significado irreduzível”. Este “todo já *pregnante*” é o exato momento da gênese da informação que só pode ocorrer ao término da suspensão de juízo, ou seja, no final do processo de redução fenomenológica, quando tudo ainda era *dado*. Assim, “antes que a atividade estruturante da consciência entre em cena, a resposta é o caos ou o *noumenon* (“a coisa em si”) kantiano antes do *phenomenon*.



Esta condição é muito relevante, pois nos apresenta a caracterização de *dado* como sendo o caos, em total ausência de estrutura, em contraposição à ocorrência de um fenômeno mental ou intelectual estruturante. Invocando aqui o conceito de informação (i.e., *informação intencional*) como sendo “dado estruturado e significativo”, recorreremos a Merleau-Ponty, que defende que o fenômeno estruturante é a “faculdade de julgamento”, dentre tantas outras possibilidades mentais do intelectualismo, como a “atenção que ilumina a sensação” e a própria “percepção”. Faculdade de julgamento é propriamente o que é suspenso na *epoché*, ou seja, somente ao término desta é que o fenômeno estruturante pode ocorrer, trazendo à existência a *informação*, que estava “preta” ou “pregnante”, aguardando o seu momento oportuno, que denominamos “gênese da informação”, primeiro pilar desta tese.

O fenômeno da observação contempla duas etapas: a **sensação** e a **percepção**. A diferença entre estes dois conceitos varia de acordo com as definições dos termos. Uma distinção comum é de que as sensações são simples experiências sensoriais, enquanto as percepções são construções complexas de elementos simples unidos por alguma associação.

Para Merleau-Ponty (1962), sensação é uma unidade da experiência de um observador, incluindo a maneira como ele é afetado.

As teorias da percepção nasceram com os empiristas ingleses: Locke, Berkeley e Hume. Eles desenvolveram a teoria de que não há outra porta de entrada para o conhecimento humano a não ser a dos sentidos e, segundo Santaella (2012), “é ela que predomina até hoje”, ou seja, “é a mente que constrói o mundo. . . a partir de uma matéria bruta fornecida pelos sentidos”.

Segundo Santaella (2012), sensação “são dados que nos vêm dos sentidos” e estes dados são tomados como elementos primários da percepção, que se constitui de uma soma de sensações, sendo “destituída de significado”. Já, a percepção, “não importa quão elementar possa ser, está sempre carregada de significado”. Ela explica que a percepção é um passo construtivo mental-fisiológico iniciado pela sensação do mundo físico externo, fora do observador. É a mente que tem a tarefa da síntese, i.e., a elaboração daquilo que chamamos de compreensão ou significado, envolvendo algo que se perde e algo que se acrescenta ao que é observado.

Lombardo (2011) diferencia algumas versões do conceito de percepção. No realismo representacional de Descartes e Locke, por exemplo, o conteúdo da percepção copia, mas se distingue do mundo. No realismo crítico, a percepção é o ponto de partida para o conhecimento do mundo, sem copiá-lo. Na linha do construtivismo mental, requer-se um processo de organização fisiológica para que a percepção seja verídica ou significativa. No realismo identitário estrutural, a percepção identifica-se com as qualidades intrínsecas do cérebro. Na emergência psicológica, as qualidades do fenômeno de percepção se mostram como características globais e irreduzíveis de estados cerebrais. E, no noumenalismo kantiano, conteúdo e estrutura do fenômeno da percepção diferem da coisa em si, da qual nada pode ser dito.

Segundo Rakova (2006), Jean-Paul Sartre foi um filósofo francês que desenvolveu a versão existencialista da Fenomenologia.

Com relação aos aspectos de intencionalidade e consciência, Sartre começou estudando a

imaginação onde a consciência do não-existente se mostrava evidente. Ele introduziu na Fenomenologia as categorias de ser-em-si, a transfenomenalidade das coisas além das aparências e o ser-para-si (ou reconstituição constante da consciência e de si mesmo). Ele enfatizava a sua visão filosófica do homem absolutamente livre para escolher a si mesmo como verdadeiro autor de suas ações. Então, pensar que as ações de um observador são determinadas pelo caráter imanente dos objetos do mundo para a sua consciência era um tipo de auto-decepção e má-fé, que levava à angústia.

As obras mais relevantes de Michael Polanyi (1891-1976) foram “Conhecimento Pessoal” (1958) e “Significado” (1976). [Jha \(2002\)](#) explica que Polanyi produziu amplo programa epistemológico e explorou as muitas facetas de interação observador-mundo em que usa o corpo como instrumento para interpretar indícios do ambiente. O ato de interpretação baseia-se no uso habilidoso de indícios provenientes dos órgãos sensoriais: dados sensoriais, contexto e conhecimento subjetivo.

Polanyi argumenta sobre a emergência do significado, corroborando com a Fenomenologia Genética:

“Existem vários níveis de realidade e causalidade. As condições encontradas nas camadas limítrofes fornecem graus de liberdade que, ao invés de serem aleatórias, são determinadas pelas realidades dos níveis mais altos, cujas propriedades são dependentes (mas distintas) dos níveis mais baixos, de onde emergem. O processo pelo qual os significados emergem revela as intenções implícitas nas forças causais” ([POLANYI; PROSCH, 1975](#)).

[Polanyi \(1965\)](#) afirma que “a relação entre o corpo e a mente tem a mesma estrutura lógica das relações entre indícios e a imagem para o qual eles apontam”. Para ele, a consciência alcança coerência pela integração dos indícios às coisas a que se referem em um processo de dois níveis de “conscientização”, envolvendo objetos identificáveis em um ponto focal e um conjunto de raízes subsidiárias que funcionam como indícios para os objetos focais ou parte deles.

Os atos da consciência não são apenas conscientes *de* algumas coisas, mas também *a partir de* certas coisas, incluindo o nosso corpo. “Nós conhecemos uma entidade abrangente por interiorizar suas partes ou por fazer-nos habitar nelas. O processo oposto (externalização) passa a atenção do foco para as partes, tornando as partes em objetos externos, sem significado funcional.

[Rakova \(2006\)](#) diz que Daniel Dennett é um filósofo americano dos dias atuais, crítico dos problemas de intencionalidade e da consciência.

Ele defende que a intencionalidade não é uma propriedade intrínseca aos estados mentais nem aos organismos, mas um construto que emerge do comportamento dos sistemas e de capacidades cognitivas interativas. Em diferentes tempos, os conteúdos adquirem proeminência no sistema cognitivo manifestando-se verbalmente. Os casos de introspecção são entendidos como conteúdos não-privilegiados da consciência sobre evidências neurológicas ou psicológicas. A linguagem nos dá a ilusão de intencionalidade e de que somos conscientes de alguma coisa.

### 3.3 Epistemologia da Informação

Grandes questões da epistemologia são a própria natureza do conhecimento e sua estrutura (ALCOFF, 1998).

A consideração de elementos epistemológicos aqui é importante por nos fazer refletir sobre o que justifica aquilo que definimos ser a “informação” e o que significa quando afirmamos ser isso verdadeiro. Isso torna a sua conceituação semântica um verdadeiro desafio.

Martino (2003) explica que existem várias correntes epistemológicas, destacando-se as que se voltam para questões filosóficas, como a gnosiologia e filosofia das ciências; ou para a produção do conhecimento científico, como a história, psicologia e sociologia.

As condições necessárias para um debate epistemológico produtivo são essenciais para a compreensão madura da analogia entre as áreas de Ciência da Informação e a Genética.

Para Capurro e Hjørland (2003), que realizaram um estudo bem aprofundado quanto à definição do termo informação, esse tem um problema de natureza conceitual que, em uma perspectiva histórica, apresenta aspectos relevantes e complexos. O termo pode ser considerado sob diferentes abordagens, como no uso cotidiano, nas diversas áreas científicas, nos relacionamentos interdisciplinares, epistemologicamente e etimologicamente.

Aspray (1985) realizou ampla pesquisa sobre o desenvolvimento do conceito científico de informação durante a década que seguiu a Segunda Guerra Mundial. Ele examinou, também, as raízes da ciência da informação no período final do século XIX e início do século XX nas áreas da lógica matemática, física, psicologia e engenharia elétrica, enfocando as abordagens de Warren McCulloch, Walter Pitts, Claude Shannon, Alan Turing, John von Neumann, e Norbert Wiener, que combinaram coerentemente diversos estudos sobre a informação em suas disciplinas de domínio.

Zins (2007) realizou uma pesquisa qualitativa entre especialistas da área, usando o Método Delphi, com 57 participantes de 16 países.

Logan (2012) ressalta que “da molécula do DNA a uma pintura cubista, passando por uma peça de Shakespeare e o *New York Times*, muitas coisas diferentes podem ser chamadas de informação”. Ele afirma que a informação é relativista, se relaciona com a materialidade, com o significado e com a propagação de organização, podendo acontecer em quatro esferas distintas: na biosfera, na simbolosfera, na tecnosfera e na econosfera. Ainda, ele pesquisou a origem histórica do termo “informação”, tendo encontrado em 1386 a sua primeira citação por Chaucer, com o significado de uma ação para “dar uma forma para a mente”, um “treinamento ou moldagem da mente”. A próxima noção de “informação” foi, então, encontrada quase um século depois, em 1450, com o significado de “comunicação do conhecimento”. Ele assegura que somente no século XX surgiu a noção de informação como “algo que pode ser armazenado em, transferido ou comunicado a um objeto inanimado... e como uma quantidade definida matematicamente”.

A seguir, começando pela matematização do conceito de informação, apresentamos os paradigmas clássicos da informação, todos encontrando correspondência na *Arquitetura Genética*

da Informação. O paradigma físico, no espaço dos objetos e no fluxo da informação. O paradigma cognitivo, no espaço do sujeito. E o paradigma histórico-social, ele próprio sendo um caso tipicamente cultural, relacionado à memética (abordada no tópico 6.4.7), na genética populacional.

### 3.3.1 Paradigma Físico da Informação

O início do estudo teórico moderno da informação é atribuído a Claude Shannon. Porém, “os trabalhos de Fisher e Hartley (1928) prenunciam o conceito de informação de Shannon, que nada mais é do que uma sequência específica de símbolos, independentemente de seu significado” (LOGAN, 2012).

Para Shannon, informação é a medida da redução da incerteza para um receptor. Krippendorff (1980) reforça, dizendo que quanto maior for a redução de incerteza, mais forte será a correlação entre entrada e saída em um canal de comunicação. O problema fundamental da comunicação é transmitir a mensagem exatamente ou aproximadamente como se quer.

Em sua Teoria Matemática da Comunicação (TMC) Shannon argumenta que a quantidade de informação é inversamente proporcional à probabilidade de ocorrência daquela informação, que é codificada de alguma forma simbólica (como uma sequência de bits) (SHANNON, 1948).

A essa medida Shannon chamou de entropia, que, etimologicamente, tem o significado de “transformação de energia”, do grego *tropo*, “transformação” + o prefixo “en”, de energia, “por causa da estreita associação que acreditava existir entre energia e entropia” (LOGAN, 2012) e, fisicamente, o significado de desordem ou desorganização de um sistema.

Em 1944, Schrödinger (1992) introduziu a noção de entropia negativa:

“Tudo o que está acontecendo na natureza significa um aumento de entropia da parte do mundo onde isso está acontecendo. Assim, um organismo vivo aumenta continuamente a sua entropia – ou, como se pode dizer, produz entropia positiva – e, portanto, tende a aproximar-se do perigoso estado de máxima entropia, que é a morte. Ele só pode manter-se distante dele, ou seja, vivo, continuamente obtendo entropia negativa de seu ambiente. . . O essencial no metabolismo é que o organismo consiga libertar-se de toda a entropia que não pode evitar produzir enquanto estiver vivo”.

Daí, Wiener (1948) introduziu o conceito de informação como sendo “entropia negativa”, que exprime finalidade:

“Assim como a entropia é uma medida de desorganização, a informação transportada por um conjunto de mensagens é uma medida de organização. De fato, . . . o negativo de sua entropia, e o logaritmo negativo da sua probabilidade. Ou seja, quanto mais provável a mensagem, menor a informação que dá.”

As mensagens carregam significados inerentes, os quais são irrelevantes em termos de engenharia. Daí, a ênfase no aspecto da “comunicação” (capacidade de transmissão) e não “informação” (conteúdo ou significado da mensagem). Para Shannon, a solução de engenharia é oferecer no canal de saída um conjunto de possíveis mensagens, permitindo que o destinatário tenha condições de escolher a que julga mais apropriada. Assim, Sveiby (1994) explica que na

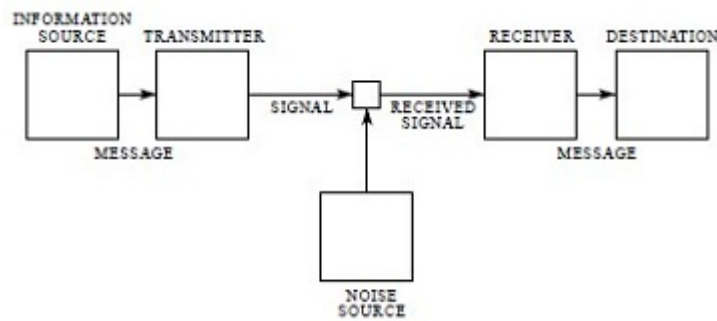


Figura 5: Esquema de um Sistema Geral de Comunicação - Fonte: (SHANNON, 1948)

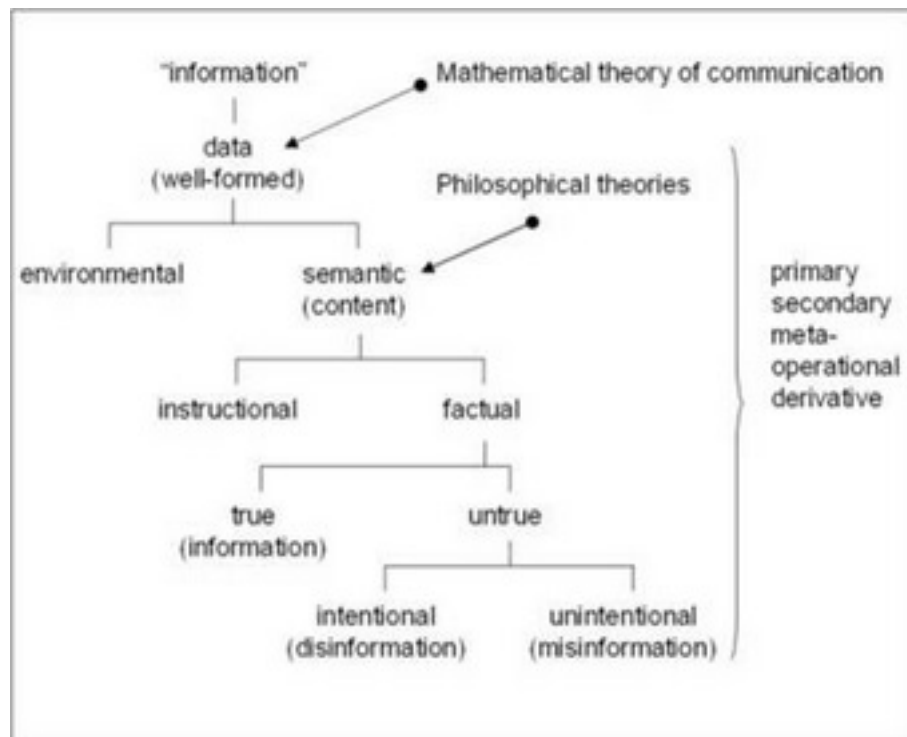


Figura 6: Informação, Desinformação e Má-informação Derivadas da TMC - Fonte: (FLORIDI, 2007)

TMC “informação é a medida da liberdade de escolha para a seleção de uma mensagem”. A figura 5 apresenta o diagrama esquemático de um sistema geral de comunicação.

Para Floridi, que introduziu o termo “Filosofia da Informação”, a informação precisa ser bem-formada (aspecto sintático), ter significado (aspecto semântico) e ser verdadeira (aspecto alético). A figura 6 apresenta os conceitos de informação, desinformação e má-informação derivados da TMC, propostos por Floridi.

O grande mérito de Floridi parece ter sido inovar no uso de um formalismo lógico como metalinguagem, algo que nunca havia sido tentado antes para a definição conceitual de informação. Cançado (2005) ressalta que “a adoção de uma metalinguagem diferente da própria língua elimina prováveis distúrbios na análise linguística”.

Assim, Floridi busca na lógica modal, mais especificamente na lógica das necessidades e das possibilidades, uma demonstração de sua tese de veridicalidade da informação, que pode

ser resumida em cinco passos:

1. Eliminação de informação excedente;
2. Eliminação de tautologias;
3. Eliminação de contradições;
4. Eliminação de inconsistências produzidas por contradições; e
5. Seleção de proposições contingencialmente verdadeiras.

Um fator limitante de sua abordagem é a adoção da definição tarskiana de verdade, que abandona a tentativa de encontrá-la em expressões da linguagem natural, senão somente em linguagens formais. Em seu artigo, Floridi defende-se de outros pesquisadores que o contradizem, como [Dodig-Crnkovic \(2005\)](#), para quem a informação não precisa ser verdadeira, pois encontra na relevância uma suficiência mínima e aceitável.

[Cover e Thomas \(1991\)](#) afirmam que a Teoria da Informação responde a duas questões fundamentais: qual é o limite máximo para (1) a compressão de dados? (resposta: depende da capacidade descritiva) e (2) a transmissão de dados? (resposta: depende da capacidade do canal). Também, enfatizam que os trabalhos mais recentes estão concentrados na teoria da informação em rede, onde acontecem simultâneas taxas de comunicação de vários emissores para vários receptores, na presença de sinais de ruído que interferem na comunicação.

Ainda, explicam o conceito de “informação mútua”, caracterizada pelo prévio conhecimento de determinada variável aleatória quando utilizada para descobrir uma outra variável aleatória por ela condicionada, sendo uma outra maneira de promover a redução de incerteza na saída.

[Vedral \(2010\)](#) discorre sobre a importância da TMC, afirmando categoricamente que independentemente do evento, pode-se aplicar a teoria da informação de Shannon a ele, argumentando que a informação está por trás de cada processo que vemos na Natureza. Ele explica que o comportamento quântico escapa à visualização e ao senso comum, nos forçando a repensar como olhamos o Universo e aceitar uma nova e estranha visão de mundo. Ele entende que:

“Para um físico, tudo no mundo é informação. O Universo e seu funcionamento são o fluxo e refluxo de informação. Nós próprios somos padrões transitórios de informação, passando a receita de nossas formas básicas para futuras gerações através de um código digital de quatro letras chamado DNA”.

Discorreremos mais sobre a aplicação desse conceito na Genética da Informação Intencional, no capítulo 6.

Observando as dimensões sintática e semântica da informação, realizou estudos relacionados à problemática da origem da vida sob uma perspectiva da Teoria da Informação. A primeira dimensão, correspondendo aos problemas estatísticos da origem da vida e a segunda, à complexidade da informação genética. Este estudo é um exemplo que revela a “informação” como sendo o centro de investigações científicas de altíssima complexidade que permeiam as mais

profundas inquietações do homem: “De onde viemos?” “Para onde vamos?” “O que fazemos aqui?”. Ele apresenta alguns exemplos do paradigma físico da informação:

1. O DNA carrega informação genética suficiente para organizar ou controlar o crescimento ordenado de um organismo vivo;
2. Uma mensagem carrega informação suficientemente desconhecida pelos receptores da mesma;
3. A resposta do professor a uma questão do aluno carrega informação somente quando reduz a incerteza do aluno;
4. Uma linha telefônica carrega informação somente quando os sinais enviados têm correlação com aqueles recebidos.

A Associação *Foundations of Information Science* é uma conferência eletrônica sobre os fundamentos da Ciência da Informação, inaugurada em 2002. Desde então, inúmeros artigos relevantes passaram a integrar suas bases. Ele também apresenta um quadro de objetos de estudo da Ciência da Informação (em sentido amplo) que se relacionam mais diretamente com o estudo, aplicações e uso da informação documental, em que a Teoria da Informação de Shannon se destaca como primeiro objeto da lista. Cabe destacarmos alguns trabalhos recentes: de [Darvas \(2005\)](#): “Simetria, ordem, entropia e informação”; [Küppers \(2005\)](#): “Informação e a origem da vida”; [Baeyer \(2005\)](#): “Informação como realidade física”; [Vlasov \(2005\)](#): “Mundo da informação”; e [Burgin \(2005\)](#): “É a informação algum tipo de dado?”.

Ainda, [Cover e Thomas \(1991\)](#) mostram a importância da Teoria da Informação para as outras ciências<sup>2</sup>.

### 3.3.2 Paradigma Cognitivo da Informação

Hornung, citado por [Kopp \(2007\)](#), entende que “informação é o significado da representação de um fato (ou de uma mensagem) para o receptor” (vide seção 2.4.1).

[Bateson \(1979\)](#) complementa que informação é aquilo que provoca mudança em nós. Objetivamente, a definição do termo está mais ligada ao fenômeno de relevância, ou seja, ao que é informativo em determinado contexto. Subjetivamente, o foco central é o fenômeno de interpretação, sendo que a relação entre definição e significado está nas convenções de uso do termo, visando o entendimento, a comunicação e a aplicação. Assim, a questão não é se uma definição é verdadeira ou falsa, mas a sua aplicação e uso, mesmo que convencional.

[Nicola \(2005\)](#) lembra que o pensamento do filósofo Demócrito era de “que as palavras são estranhas às coisas que representam e são sinais puramente convencionais”; e complementa dizendo que “pela primeira vez na história, coloca-se a tese do convencionalismo linguístico: as

---

<sup>2</sup>A Ciência da Informação é desconsiderada no relacionamento da Teoria da Informação a outras áreas do conhecimento humano (Ciência da Computação, Física, Teoria da Comunicação, Teoria da Probabilidade, Estatística, Matemática e Economia)

palavras não possuem, em si, como som, nenhum significado; são puras convenções que adquirem sentido somente pelo uso comum como base no critério de utilidade recíproca”. Assim, colocava-se em cheque se a natureza da linguagem seria naturalista, ou seja, se o som da palavra de alguma maneira realmente exprimiria a natureza do objeto.

Na filosofia moderna, a teoria do convencionalismo linguístico foi herdada e defendida por Locke e Gotlob Frege. Para Locke, as palavras são apenas sinais convencionais, puros símbolos arbitrários eventualmente substituíveis por outros. Portanto, como reforça Nicola (2005), “o escopo das palavras é serem sinais sensíveis das ideias; e as ideias nelas contidas são o seu significado próprio e imediato”.

### 3.3.3 Paradigma Histórico-Social da Informação

A abordagem histórico-social da informação tem forte identidade com o conceito de meme, explorado na seção 6.4.7.

Machlup e Mansfield (1983) explicam os efeitos de diferentes culturas sobre a conceituação de informação: “Informação não é apenas uma coisa. Significa diferentes coisas para aqueles que expõem suas características, propriedades, elementos, técnicas, funções, dimensões e conexões”.

Termos que são definidos de forma peculiar a um indivíduo ou grupo (idiossincrasia) geralmente são negligenciados e não contribuem para o entendimento e comunicação. Logo, é importante conhecer como as pessoas aplicam os termos. Também, definições científicas explícitas podem ter significados diferentes no uso ordinário. Ocorre uma relação de dependência do significado dos conceitos de informação na estrutura da teoria a que está associada, bem como da precisão e grau de coerência desta, devendo-se considerar as limitações de alguns caminhos alternativos em que o conceito possa adquirir significado.

Quanto aos aspectos históricos da definição de Informação, vejamos como Capurro e Hjørland (2003) classificam a informação.

Quanto à origem do termo:

1. Origem Latina (*informatio*; *informo*), ato de dar forma a algo (contexto tangível); ato de educar (contexto intangível: uso moral e pedagógico).
2. Origem Grega (*hypotyposys*; *prolepsis*), ser modelo moral; representação apriori (antes da experiência) de deuses ou coisas em nossas almas; ato a posteriori de descrever algo desconhecido ou lembrança de alguma situação no passado por meio de uma representação pictográfica de uma sentença; ato de instruir, educar.

Quanto ao período histórico-cultural:

1. Idade Média (forte expressão de Tomás de Aquino): ato de prover algo com forma; ato de instruir, educar; ato de “informar”, com alguma qualidade ativa ou essencial, a partir do nada (aspecto teológico); ato de formação de virtudes (aspecto moral).



2. Idade Moderna e Pós-Moderna: ato de moldar a mente ou o caráter (na Europa, desde o sec. XIV); ato de treinar, instruir, ensinar (na Europa, desde o sec. XIV); ato de educar e comunicar conhecimento (na Alemanha, desde o sec. XV)
3. Nos dias atuais: ato subjetivo e seguro de separar o verdadeiro do falso; objetos no mundo são traduzidos em sensações (empiricismo), mas sensação é totalmente diferente de “forma”, sendo a primeira sensual e a última, intelectual; transferência do cosmos para a consciência, da forma para a sensação.

## 3.4 Arquitetura da Informação

Bailey (2003), em sua breve introdução sobre Arquitetura da Informação, discorre sobre a não existência de um consenso para o conceito de Arquitetura da Informação, muitas vezes confundida com desenho de conteúdo visando acessibilidade à informação, sendo, portanto, definida como “a arte e ciência de estruturar e organizar sistemas de informação para ajudar as pessoas a alcançarem suas metas”. Assim, segundo a autora, o arquiteto da informação organiza o conteúdo e desenha sistemas de navegação para ajudar as pessoas a gerenciarem suas informações.

Siqueira et al. (2009) analisam os diferentes conceitos epistemológicos associados ao termo “Arquitetura da Informação”, definindo um argumento para a eliminação de restrições de forma a tornar plausível um conceito abrangente. Isto estabelece um espaço conceitual simultaneamente definido pela disciplina científica, resultado de aplicações tecnológicas desta disciplina, e objeto de estudo. Os elementos comuns identificados nesse espaço e, também, nas definições estabelecidas correspondem à forma, contexto, manifestação e noções de significado. Estas noções são propostas como dimensões fundamentais da Arquitetura da Informação. Com base em suas investigações, eles sugerem que “Arquitetura” seja um conceito constituído por duas dimensões: forma e contexto. O conceito de forma, por sua vez, expressa uma estrutura organizada da matéria no espaço, dando a ela um arranjo de relações entre os seus elementos constituintes. Esta abordagem evoluiu para uma proposta da disciplina científica da Arquitetura da Informação (SIQUEIRA, 2012).

Macedo (2005) explica que no *Asilomar Institute for Information Architecture*, dedicado ao tema, a Arquitetura da Informação é definida como o “desenho estrutural de ambientes de informação compartilhados”. Também, ela ressalta que no glossário publicado na ocasião do *First Annual ASIS&T Information Architecture Summit*, Hagedorn definiu o conceito de Arquitetura da Informação como “a arte e ciência da organização da informação para a satisfação de necessidades de informação, que envolve processos de investigação, análise, desenho e implementação”.

Resmini e Rosati (2011) definem a Arquitetura da Informação como:

“uma prática profissional e campo de estudos com foco na solução de problemas básicos de acesso, uso de vastas quantidades de informação disponíveis hoje...

sendo,principalmente, uma atividade de produção e arte, que conta com um processo indutivo e um conjunto (ou varios conjuntos) de diretrizes, melhores práticas, e especialistas na área... não se tratando de uma ciência, mas de uma arte aplicada, muito parecida com o desenho industrial.”

Apresentaremos uma linha do tempo com relação à evolução do conceito de Arquitetura da Informação:

- 1964 - Em um artigo intitulado “Arquitetura do IBM System/360”, [Amdahl, Blaauw e Brooks \(1964\)](#) definem o termo ‘arquitetura’ como a estrutura conceitual e comportamento funcional, distinguindo a organização de fluxos de dados e controles, projeto lógico, e implementação física (um conceito seminal de arquitetura da computação que veio a se estender para outras áreas, uma vez que não se prendeu apenas a aspectos físicos, mas conceituais e comportamentais);
- 1970 - Na Xerox Palo Alto Research Center (PARC), um grupo de pessoas especializadas em Ciência da Informação se reuniu em assembléia e, então, deu o alvará para o desenvolvimento de tecnologia que poderia suportar a “arquitetura da informação” ([PAKE, 1985](#)). Esse grupo foi responsável por inúmeras contribuições importantes, incluindo o primeiro computador pessoal com interface amigável, impressora a laser, e o primeiro editor de textos *WYSIWIG*. Percebemos aqui a ênfase dada pelo grupo à interação homem-máquina e aos aspectos sociais da computação.
- 1976 - Richard Saul Wurman, em uma conferência no *American Institute of Architecture* usou o termo “informação” juntamente com “arquitetura”, sendo esta a primeira vez na história. Sua definição inicial foi “a organização dos padrões dos dados de forma a tornar claro o complexo”. Mesmo assim, esta expressão que passaria a significar “o desenho de mudanças complexas e dinâmicas” permaneceu em estado dormente por alguns anos, prevalecendo ainda a ideia computacional de sistemas de informação. De fato, a maioria dos artigos do início da década de 1980, refere-se à arquitetura da informação como uma ferramenta para o projeto e criação de infraestruturas de redes de computador ([MORROGH, 2003](#)).
- 1988 - [Smith e Alexer \(1988\)](#) relatam que a Xerox estava entre as primeiras corporações a se engajar na noção de estrutura da informação, usando uma “fraseologia elegante e inspiradora: *’architecture of information’*”, ao invés de *information architecture*. A propósito, esse é um dos conceitos centrais para muitos que escreveram e/ou discutiram sobre o desenvolvimento da arquitetura da informação na década de 1980.
- 1995 - [Weitzmann \(1995\)](#), em sua pesquisa de doutorado no Massachusetts Institute of Technology - MIT sobre “A Arquitetura da Informação: Interpretação e Apresentação da Informação em Ambientes Dinâmicos” fundamenta a noção de arquitetura aplicada à informação em si.

- 1998 - [Morville e Rosenfeld \(2006\)](#), bibliotecários, escreveram o livro “*Information Architecture for the World Wide Web*”, um dos livros mais importantes na área de projeto de sistemas para a web em termos de usabilidade e *design*, visando explorar as novas possibilidades para os profissionais operarem sobre grandes volumes de dados em novas mídias.
- 2005 - O *Information Architecture Institute* - IAI, uma organização profissional internacional dedicada ao avanço do estado da arquitetura da informação através da pesquisa, ensino, apoio e serviços à comunidade, estabelece o conceito de arquitetura da informação em três aspectos:
  1. O desenho estrutural de ambientes de informação compartilhados;
  2. A arte e ciência de organizar e rotular websites, intranets, comunidades online, e software de suporte a usabilidade e recuperação da informação; e
  3. uma comunidade de prática emergente com foco em princípios de desenho e arquitetura em meio digital.
- 2005 - [Barker \(2005\)](#) define arquitetura da informação como “um termo usado para descrever a estrutura de um sistema, i.e., a maneira pela qual a informação é agrupada, os métodos de navegação e a terminologia usada dentro do sistema”. Ele ainda aponta o que entende ser o problema mais comum da arquitetura da informação: “ela é simplesmente uma imitação da estrutura organizacional da empresa onde está inserida”.
- 2008 - [León \(2008\)](#) demonstra que o uso de “informação” juntamente com “arquitetura” iniciou de fato com os especialistas da Ciência da Informação que visavam promover o desenvolvimento de sistemas orientados ao usuário, algo que veio a alcançar grande visibilidade e resultados somente na década de 1990. Ela produziu um gráfico cronológico (figura 7, na página 44) com os marcos da Arquitetura da Informação, principalmente os livros, artigos e conferências em uma hipótese de três partes, cobrindo 30 anos, em que as duas primeiras fases, a do desenho da informação (1960s-1970s) e a do desenho de sistemas (1980s) se projetam para dentro da visão moderna de arquitetura da informação prevalecente nos dias de hoje.
- 2009 - Publicação da Edição 1, Volume 1 do “*Journal of Information Architecture*” ([MADSEN, 2009](#)), uma iniciativa independente do Grupo de Ensino e Pesquisa em Arquitetura da Informação - REG-iA, patrocinado pelo Instituto de Arquitetura da Informação e pela Escola de Negócios de Copenhagen. O Editorial desta edição inaugural enfatiza a necessidade de criação de um fórum permanente onde se possam publicar artigos de interesse central na área de arquitetura da informação. Os artigos diziam respeito à experiência do usuário, análise da arquitetura de documentos, contextualização e incertezas na arquitetura da informação.

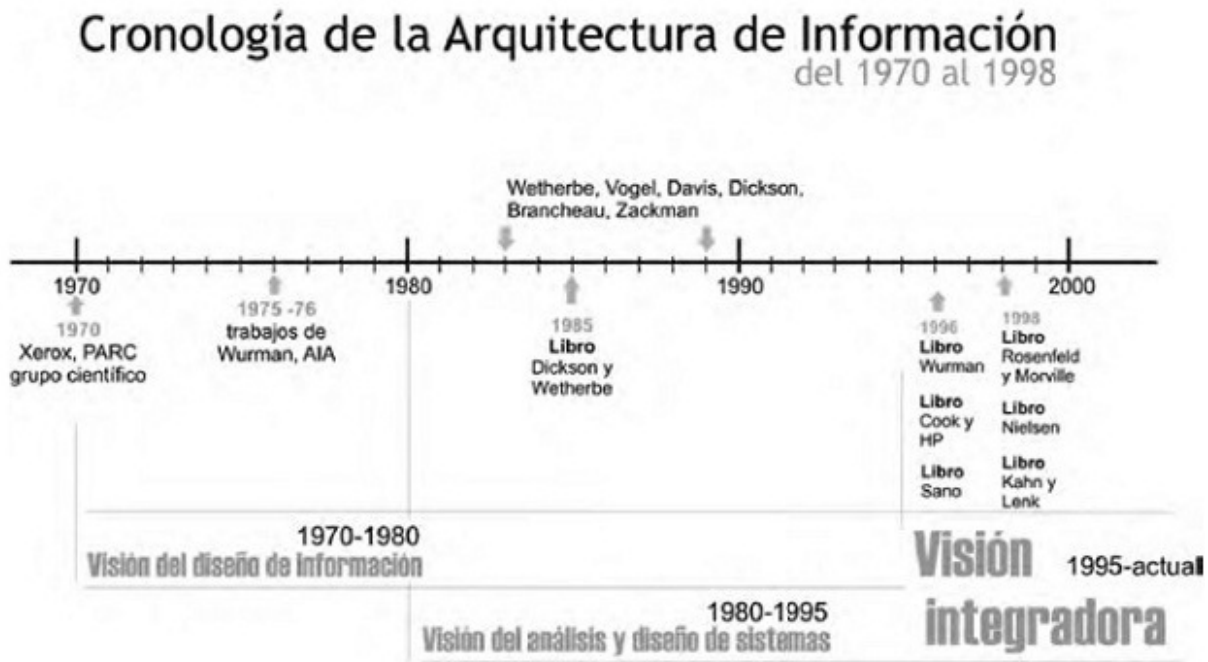


Figura 7: Cronologia da Arquitetura da Informacao - Fonte: (LEÓN, 2008)

As iniciativas para a caracterização de uma Arquitetura da Informação no CPAI tiveram o seu marco em Lima-Marques (2011), que idealiza a “Escola de Brasília em Arquitetura da Informação”, com respaldo epistemológico, filosófico, científico e pragmático para uma Teoria Geral da Arquitetura da Informação - TGAI. Em seguida, Siqueira (2012) defende tese propondo a formalização da Arquitetura da Informação como disciplina científica baseada nas categorias de forma, contexto, manifestação e significado - FCMS.

## 3.5 Organização e Representação do Conhecimento

### 3.5.1 Teoria da Classificação

A Teoria da Classificação teve o seu início na arte de classificar, evoluindo para a ciência de classificar. Como arte, temos os aspectos históricos, uma vez que desde o início da humanidade o homem vem realizando classificações de diversas maneiras, otimizando-as com a evolução do conhecimento (HJØRLAND; PEDERSEN, 2005).

Arranjos sistemáticos podem ser observados na Bíblia; em trabalhos filosóficos; em coleções enciclopédicas, como a de Amenope (1350 a.C.), Caius Plinius Secundus (23-79 d.C.), na Idade Média como Isidoro de Sevilla, Vincent de Beauvais, Bartholomaeus Anglicus e Brunetto Latini. Na Renascença, como Georg Valla, Rafael Maffei, Johann Heinrich Alsted, Wolfgang Ratke, Diderot e D’Alembert.

A Teoria da Classificação como ciência teve sua origem por volta de 1491, quando a classificação passou a ser considerada uma finalidade e não apenas um meio. Inclusive, começaram

a surgir movimentos de elaboração de sistemas de classificação com significativa evidência na busca de relações esquemáticas entre áreas do conhecimento humano.

Alguns arranjos esquemáticos que se destacaram foram os de Angelo Poliziano (1491), em seu trabalho intitulado “Panepistemon”; Francis Bacon (1605), em seu plano de classificação das ciências; Saint-Hilaire (1805-1861), introduzindo a classificação diagramática; Henry Bliss (1910), com os fundamentos filosóficos da classificação e a organização de livros em uma biblioteca.

A moderna teoria da classificação surgiu com Ranganathan, que se inspirou em Bliss e propôs a Classificação dos Dois Pontos, um sistema universal de classificação facetada. Vickery contribuiu com a flexibilidade e generalidade nas citações; Austin, com a indexação por linguagem livre, chamada “PRECIS”.

Dahlberg (2006) aborda a questão da organização do conhecimento, definindo-o como sendo um arranjo de conceitos de acordo com certas características para a obtenção de conhecimento sobre determinados objetos.

Hodge (2000) apresenta sistemas de organização do conhecimento, que são esquemas de organização da informação que promovem a gestão do conhecimento. Fulton (2000) define domínio como sendo o escopo de definição da informação com base em informações apropriadas para um campo de estudo, missão, função ou processo de negócio. Fechando o ciclo, agora com o objeto da pesquisa, temos a definição de análise de domínio, que para De Lorenzo (2005), “é uma busca por unidades de conhecimento culturais (envolvendo pessoas, lugares e coisas), para obter entendimento de relacionamentos semânticos entre termos e categorias”.

Há inúmeras abordagens de análise e estudo de Sistemas de Organização do Conhecimento (KOS), dentre as quais podemos destacar a análise de domínio, análise facetada, análise orientada a solicitações e análise centrada no usuário.

A base para a organização do conhecimento em qualquer domínio de discurso é a Teoria da Classificação, que, historicamente, teve o seu período artístico e, mais recentemente, científico. A Teoria da Classificação como arte teve origem quando o homem classificava objetos de diversas maneiras, evoluindo juntamente com o conhecimento. A Teoria da Classificação como ciência tem a classificação como um fim em si mesma, introduzindo um movimento de elaboração de sistemas de classificação e relações esquemáticas entre áreas do conhecimento humano.

Dentro desta teoria e sob o ponto de vista da análise de domínio, Hjørland e Pedersen (2005) explicam que a classificação é a ordenação dos objetos (ou processos, ideias etc) em classes baseadas em algumas propriedades, que se disponibilizam por meio de descrições. Ressalta-se que a enumeração completa em uma descrição é tarefa impossível, posto que não seria mais um modelo, mas a própria realidade.

Dahlberg (2006) é o pioneiro em Organização do Conhecimento, tendo produzido importantes trabalhos na área, inclusive no âmbito filosófico. Ele define:

“Organização do Conhecimento é a ciência de estruturação e arranjo sistemático de

unidades de conhecimento (conceitos) de acordo com seus elementos de conhecimento inerentes (características) e a aplicação de conceitos e classes de conceitos ordenados para a missão do valoroso conhecimento do conteúdo dos referentes (objetos/assuntos) de todos os tipos.” (Dahlberg, 2006).

Hodge (2000) define os Sistemas de Organização do Conhecimento (KOS) como se constituindo de todos os tipos de esquemas de organização da informação e promoção da gestão do conhecimento, incluindo: esquemas de classificação e categorização (que organizam materiais em um nível genérico); cabeçalhos de assuntos (que proporcionam mais detalhes de acesso); e arquivos de autoridade (que controlam versões de informações-chave, tais como nomes geográficos e nomes pessoais). Também incluem vocabulários altamente estruturados, como os tesouros, e esquemas menos tradicionais, como redes semânticas e ontologias. Por serem mecanismos de organização da informação, estão no coração de cada biblioteca, museu e arquivo.

Algumas definições de domínio encontradas no *Dictionary of Information Science and Technology*, de Khosrow-Pour (2007) inclui a entrada de Fulton, que diz que é o escopo de definição da informação, uma coleção de informações geralmente reconhecidas como apropriadas para um campo de estudo, uma função, missão ou processo de negócio. Na mesma entrada, Tsunoda et al. complementam dizendo que “o domínio combina vários elementos e temas de estruturas secundárias em torno de uma função específica”. E para Coakes e Clarke (2005), é o “escopo de um assunto ou esfera de conhecimento”. Em nosso propósito podemos, então, entender domínio como uma área do conhecimento humano; o assunto principal de determinada área de estudo ou de um setor da atividade humana; ou o corpo de conhecimento de um tema específico.

Explenados os conceitos de organização do conhecimento, de sistemas de organização do conhecimento e de domínio, passemos para a análise de domínio, como definido por De Lorenzo em Khosrow-Pour (2007): “constitui-se de uma busca por unidades de conhecimento culturais chamados domínios, como sinônimo de pessoas, lugares ou coisas, usada para obter entendimento de relacionamentos semânticos de termos e categorias”. Continuam Coakes e Clarke (2005): “o produto da análise de domínio é a linguagem do domínio, que inclui termos técnicos específicos, frases, abreviações e acrônimos que são únicas e específicas para a esfera do conhecimento”, que, Khosrow-Pour (2007) afirma: “pode ser representada por ontologias de domínio, objetos de engenharia que definem modelos de conhecimento em um domínio específico, cujo nível de especificidade pode ser bem profundo”.

Hjørland foi quem promoveu a análise de domínio na Ciência da Informação (HJØRLAND, 1997). A unidade da análise de domínio é uma especialidade, uma disciplina, ou um problema temático. Em contraste com outras abordagens de análise de domínio, Hjørland enfatiza o papel essencial de uma perspectiva social ao invés de uma perspectiva psicológica, mais convencional. Hjørland chamou sua abordagem de atividade-teórica. As abordagens tradicionais focam os indivíduos como usuários únicos de informação em termos de suas estruturas cognitivas e estratégias pessoais. A abordagem de atividade-teórica, por outro lado, enfatiza uma visão holística da recuperação da informação em um contexto mais amplo em que as necessidades de

um usuário devem ser sempre interpretadas no contexto da disciplina. A relevância dos itens recuperados tem ligação direta com a substância do assunto em questão. Esta visão está alinhada com a visualização do domínio, i.e., para prover um contexto significativo em que os cientistas possam explorar o corpo do conhecimento como um todo, em oposição a lidar com fragmentos do conhecimento. A análise de domínio é possível no âmbito de uma comunidade discursiva, identificada por Hjørland como científica, acadêmica ou profissional; com estruturas de comunicação e publicação, tipos de documentos, terminologia e estruturas da informação únicos. Morado Nascimento complementa:

“a análise de domínio reconhece que as comunidades discursivas compõem-se de atores com pontos de vista distintos, estruturas de conhecimento individuais, predisposições, critérios de relevância subjetivos, estilos cognitivos particulares. Mas se fazem presentes no jogo entre as estruturas de domínio e o conhecimento individual e na interação entre o nível individual e social” (MORADO-NASCIMENTO, 2005).

Podemos dizer que, de uma certa forma, toda classificação leva consigo o elemento da arbitrariedade, não no sentido da negligência de racionalidade, mas pelo fato de que é potencialmente subjetiva. Portanto, também é questionável, dada a dificuldade de se alcançar uma verdade consensual, pois está extremamente vinculada à pessoa do organizador do conhecimento, mesmo que ele tenha as melhores intenções e disponha dos melhores métodos. Não poderia ser diferente. A própria ciência está vinculada ao cientista e carrega muito dele (ideologia, valores, cultura etc) em suas teses e teorias. Nem por isso, deixa de ser útil e valiosa, quando bem pensada e clara. Em nossa tentativa de organizar o conhecimento no domínio da Arquitetura da Informação em analogia com a Genética, não teremos como fugir desta constatação. Soergel (1985) afirma que o maior desafio para a recuperação da informação é a identificação do conceito em um domínio específico de interesse.

Para exemplificar, citemos a experiência de Fernando Pessoa, considerado o maior poeta da língua portuguesa, que iniciou carreira como crítico literário na revista “A Águia”, em 1912. O exercício constante da crítica fez dele um grande leitor e teorizador. Comentando sobre a taxonomia aristotélica, que classificava a poesia em lírica, elegíaca, épica e dramática, Fernando Pessoa se deparou com os elementos mais importantes no processo de organização do conhecimento: reflexão, utilidade, clareza e correspondência à verdade. Chegou à seguinte conclusão: “Como todas as classificações bem pensadas, é esta útil e clara; como todas as classificações, é falsa.”

### 3.5.2 Representação do Conhecimento com Ontologias

Studer, Benjamins e Fensel (1998) definem ontologia como sendo “uma especificação formal e explícita de uma conceitualização compartilhada”.

A especificação explícita é dada pelas definições declarativas de conceitos, instâncias, relações, restrições e axiomas; o aspecto formal é alcançado quando tais definições são compreensíveis e manipuláveis por agentes e sistemas de informação; a conceitualização caracteriza o

modelo abstrato de uma área de conhecimento ou de um universo limitado de discurso; e, por fim, o compartilhamento se verifica no conhecimento consensual, seja uma terminologia comum da área modelada, ou acordada entre os desenvolvedores dos agentes que se comunicam.

[Daconta, Obrst e Smith \(2003\)](#) ressaltam que dentre os modelos de representação do conhecimento mais usuais, as ontologias se destacam por limitar as possibilidades de interpretação (semântica) formal dos vocabulários para os significados que realmente são intencionados pelo modelador, evitando-se assim proporcionar liberdade de interpretação.

[Sachs \(2006\)](#) esclarece que uma ontologia em particular é apenas uma maneira de se estruturar os conceitos e as relações de um domínio, não existindo uma maneira correta para o desenvolvimento de ontologias; o que existe são alternativas viáveis. Usualmente, ocorre em um processo iterativo de definições, revisões, detalhamentos e refinamentos. Assim, é possível haver diferentes ontologias para um mesmo domínio. Para ela, os conceitos em uma ontologia deveriam ser restritos aos objetos (físicos ou lógicos) e relacionamentos, indicados respectivamente por nomes e verbos nas sentenças.

As ontologias representam os três níveis de modelos: sintático, semântico e pragmático.

No nível sintático, preocupa-se com a linguagem (alfabeto, vocabulário, regras de combinação e sentenças), formato, ordem e estrutura dos elementos, sempre visando riqueza e precisão de representação, que pode ser por meio de árvores ou grafos. No nível semântico, ocorre a interpretação, que é o mapeamento entre os dados de uma estrutura e os objetos do domínio. É caracterizado pelo fato de que os modelos repousam na mente humana (com representações do mundo e relacionamentos entre os objetos). Tal interpretação transforma símbolos e relações em significados. A formalização das ontologias permite que haja interpretação automatizada por meio de processos lógicos de inferência. No nível pragmático, o foco se desloca da acessibilidade da informação para a utilidade da informação. Daí, podemos associar o conceito de ruído a um dos seguintes aspectos: informação errada, ou momento errado, ou lugar errado ou maneira errada.

As ontologias estão no coração de todas as aplicações semânticas na web. [Horridge \(2011\)](#) reforça dizendo que “ontologias são usadas para capturar o conhecimento sobre um domínio de interesse, descrevendo conceitos nesse domínio e também os relacionamentos que surgem entre estes conceitos”.

### 3.5.2.1 Ontologias de Domínio e Ontologias Fundamentais

[Staab e Studer \(2009\)](#) explicam que:

“Algumas experiências importantes de engenharia de ontologias são capturadas nas estruturas e no uso de ontologias existentes. Os autores distinguem diferentes tipos de ontologias pelos diferentes níveis de generalidade e diferentes propósitos: ‘ontologias de alto-nível’, por vezes, também chamadas de ‘ontologias fundamentais’, capturam conceitos gerais que são de domínio independente. . . As ‘ontologias de domínio’ apresentam os conceitos do modelo e as relações relevantes para um domínio específico. . . As ‘ontologias de tarefas’ descrevem conceitos que são específicos para



uma tarefa específica. . . As 'ontologias de aplicação', no nível mais baixo de abstração, combinam as ontologias de domínio com as ontologias de tarefa."

Esta seção realiza o estudo das ontologias formais fundamentais (ou de alto nível) endereçando a solução para a criação de uma Ontologia Genética da Informação, como um dos resultados desta tese.

Tipicamente, as ontologias são codificadas em linguagens ontológicas, como: Linguagem de Especificação Algébrica Comum; Lógica Comum; Cyc; Gellish; IDEF5; Formato de Intercâmbio de Regras (RIF); *Web Ontology Language* - OWL; Linguagem de Desenho de Aplicações Semânticas (SADL); Semântica de Vocabulários e Regras de Negócio (SBVR); e Ontologias Abertas em Biomedicina - OBO.

### 3.5.2.2 Ontologias Formais e Ontologias Não-Formais

As ontologias também podem ser classificadas como formais ou não-formais, sendo as primeiras definidas por axiomas em uma linguagem formal, tipicamente em uma linguagem de ontologias, visando fornecer uma visão imparcial da realidade e independente de domínio e de aplicação.

As propriedades básicas das ontologias formais são a expansibilidade indefinida, independência de conteúdo e de contexto e acomodação de diferentes níveis de granularidade. Os componentes comuns entre as ontologias formais são os continuantes, perdurantes e qualidades:

- Os continuantes são substâncias, entidades materiais ou abstratas que podem ser observadas ou percebidas como um conceito completo, não importante em que instante de tempo;
- Os perdurantes são os ocorrentes, acidentes ou acontecimentos, entidades as quais apenas uma parte existe em determinado instante de tempo;
- As qualidades são as propriedades que caracterizam os continuantes.

As ontologias formais geralmente servem como modelo para a criação de novas ontologias; são usadas para o mapeamento cruzado entre ontologias existentes; potencializam o processamento em linguagem natural (PLN) e facilitam a normalização de bancos de dados.

As ontologias formais fundamentais (de alto nível) mais conhecidas são a - BFO; *Descriptive Ontology for Linguistic and Cognitive Engineering* - DOLCE; *General Formal Ontology* - GFO; *Object-Centered High-Level REference ontology* - OCHRE; *Suggested Upper Merged Ontology* - SUMO; e a *Upper Mapping and Binding Exchange Layer* - UMBEL.

As linguagens formais de Lógica Descritiva (DLs) são utilizadas para a representação semântica de conhecimento e raciocínio sobre esta representação, primeiro pela definição de conceitos relevantes (terminologia), e depois, usando estes conceitos para especificar propriedades

dos objetos e indivíduos que ocorrem no domínio (a descrição do mundo). O raciocínio automático permite que se realize implicitamente inferência sobre o conhecimento representado explicitamente na base de conhecimento e, também, permite a classificação hierárquica de conceitos e indivíduos (BAADER, 2003). As DLs podem ser consideradas fragmentos decidíveis da lógica de primeira ordem (que por sua vez, estende a lógica proposicional, com operadores de quantificação e poder expressivo suficiente para formalizar praticamente toda a matemática).

### 3.5.2.3 Basic Formal Ontology - BFO

Desenvolvida por Barry Smith e seus associados, a BFO é uma ontologia formal fundamental que consiste de uma série de sub-ontologias em diferentes níveis de granularidade (janelas ou porções da realidade).

As ontologias podem ser de entidades continuantes ou ocorrentes (perdurantes, conforme explicado no item 3.5.2.2. As ontologias continuantes contemplam os objetos tri-dimensionais, que têm características estático-espaciais instantâneas da realidade. Trata-se de um inventário de todas as entidades existentes em determinado momento. As ontologias ocorrentes são compostas por processos que acontecem no tempo (quarta-dimensão), com características dinâmico-temporais da realidade. Trata-se de um inventário de todos os processos existentes em um dado intervalo de tempo.

Algumas aplicações típicas existem em ciências da vida e inteligência (segurança e defesa), além de outras. Smith e Ceusters (2005) apresentam as características principais das relações em ontologias das ciências da vida, muitas das quais podem ser melhoradas com o uso de ferramentas e métodos que proporcionam maior grau de rigor lógico e ontológico. A Biblioteca OBO e a Gene Ontology (GO) são dois exemplos de iniciativas nesse sentido. A primeira é um repositório de vocabulários controlados desenvolvido para o uso compartilhado entre diferentes domínios biológicos e médicos. A segunda consiste de três vocabulários controlados: componentes celulares, de funções moleculares e de processos biológicos, desenhada para ser usada em anotações de genes e seus produtos.

As ontologias BFO são concebidas em estruturas de grafo, onde os nós são os termos e as arestas, as relações entre os termos.

Os **termos** devem ser aplicáveis a todos os organismos em geral e também a grupos taxonômicos bem específicos (ex.: moscas, fungos e levedura).

Há uma variedade de tipos de **relações** utilizadas, dependendo da ontologia: 'é-um', 'parte-de', 'desenvolve-a-partir-de', 'posição-de', 'disjunta-de', 'tem-localização' etc.

Alguns problemas podem acontecer na incorporação de relações entre ontologias similares, pois tipicamente a integração acontece de maneira informal e frequentemente não se fornece qualquer definição.

Ainda, a falta de clareza nas interconexões lógicas entre as várias relações empregadas requerem análise de requisitos para a correção destes defeitos.

Os critérios a serem satisfeitos para a inclusão de uma ontologia na OBO são basicamente as seguintes: definições textuais ou descrições, significado preciso, clareza, emprego de sintaxe, linguagem padrão (OWL ou OBO), ortogonalidade com as outras ontologias já incluídas na biblioteca e consistência no uso das relações entre os termos. Propõe-se uma metodologia de Ontologia de Relações visando maximizar a confiabilidade de curadoria<sup>3</sup>; maximizar a alavancagem na construção de uma sólida base de conhecimento; e permitir a extensibilidade de relações e definições rigorosamente formais, com detalhes técnicos invisíveis aos autores e curadores.

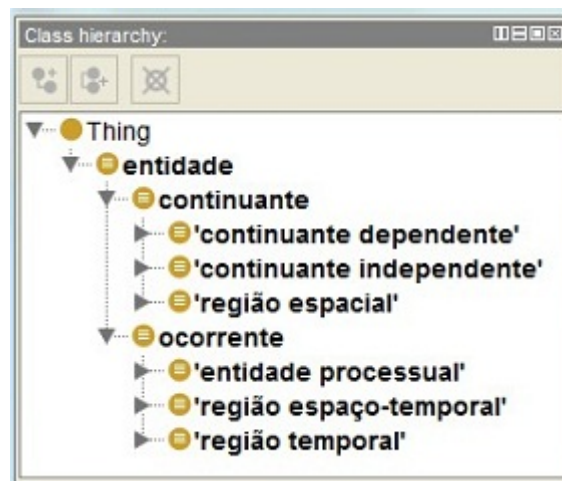


Figura 8: Ontologia BFO - Classes Hierárquicas (até o terceiro nível) - Fonte: Produzido pelo autor

Nesta pesquisa, realizamos a tradução da ontologia BFO para o português e, na figura 8, apresentamos as suas classes hierárquicas até o terceiro nível de detalhe.

A classe das entidades continuantes é apresentada na figura 9.

E, por fim, a classe das entidades ocorrentes é apresentada na figura 10.

### 3.5.3 Metodologia para a Construção de Ontologias

Para a construção das ontologias, adotamos a metodologia do Modelo-V, , apresentada no diagrama da figura 11, na pág. 54. O lado esquerdo do “V” apresenta os processos de construção da ontologia e o lado direito, as diretrizes, princípios e avaliação usados para assegurar a qualidade da ontologia.

Em função do propósito de analogia entre os dois domínios, a ontologia de termos da informação na genética está detalhada até o nível onde aparecem os conceitos de informação. O modelo contempla aspectos de intuitividade, extensibilidade e manutenibilidade.

Na etapa de conceitualização, adotamos a orientação de [Sachs \(2006\)](#): (1) Definir as classes ontológicas; (2) Arranjar hierarquicamente as classes, em superclasses e subclasses; (3) Definir faixas e descrever seus valores permitidos; e (4) Preencher os valores das faixas com suas instâncias.

<sup>3</sup>Curador é o profissional que administra e/ou edita o conteúdo das ontologias.



Figura 9: Ontologia BFO - Classe das Entidades Continuanes - Fonte: Produzido pelo autor

Para a construção da Ontologia de Termos da Informação na Genética e da Ontologia de Processamento da Informação em Sistemas Biológicos, seguimos as orientações do ciclo de vida propostas por Ould (1990), que é composta por: identificação do propósito e escopo; aquisição de conhecimento; conceitualização; integração com outras ontologias; codificação e anotação; documentação; e avaliação. Ressalta-se que esta fase final de avaliação deve ocorrer em aspectos e níveis diferenciados: estrutura, expressividade, granularidade representacional, descritividade x prescritividade, uso de raciocínio automático, e utilidade.

Andersen (2004) diz que as máquinas podem nos ajudar na complexa missão de apresentar modelos de organização do conhecimento, pela construção automática (ou semi-automática) de representações baseadas em técnicas algorítmicas variadas de mineração de dados e descoberta de conhecimento em grandes volumes de bases textuais nos domínios de discurso. Os resultados podem ser surpreendentes, extremamente úteis e funcionais.

As seções 7.5.1 e 7.5.2 apresentam um processo semi-automático de construção das ontologias para viabilizar a analogia entre os domínios da genética e da arquitetura da informação.

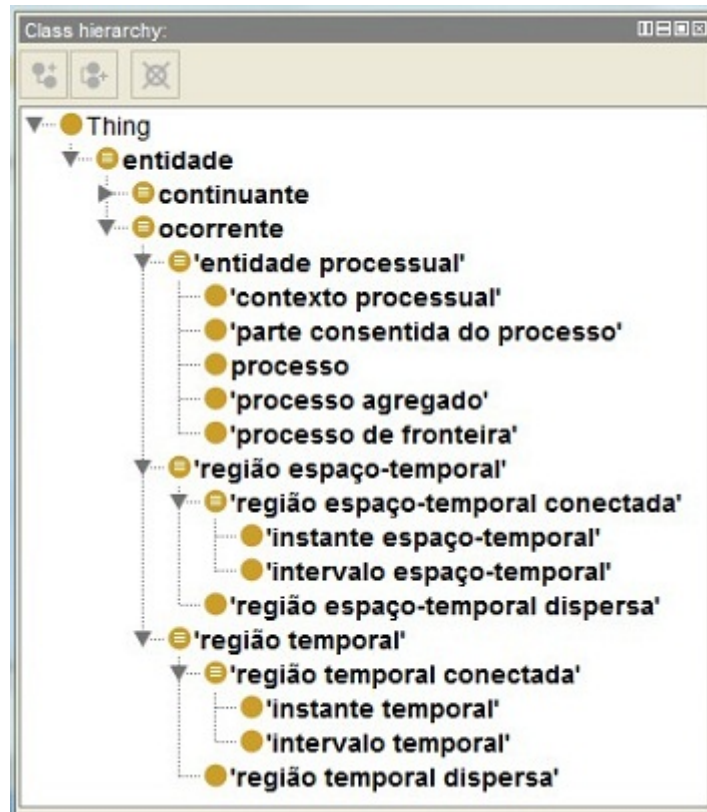


Figura 10: Ontologia BFO - Classe das Entidades Ocorrentes- Fonte: Produzido pelo autor

### 3.5.4 Operações com Ontologias

Studer, Benjamins e Fensel (1998) explicam que as operações básicas entre ontologias são de mesclagem, mapeamento e alinhamento.

- A mesclagem de ontologias é a criação de uma nova ontologia com base em outras pré-existentes, caracterizando uma necessidade convencional, uma vez que a nova ontologia contém todo o conhecimento das ontologias originais. No entanto, esse requisito não precisa ser totalmente preenchido, uma vez que as ontologias originais podem não estar juntas em perfeita sintonia. Nesse caso, as importações para a nova ontologia ocorrem pela seleção dos conceitos das ontologias originais visando consistência. A ontologia resultante da fusão pode ter a introdução de novos conceitos e relações, servindo como uma ponte entre os termos das ontologias originais.
- O mapeamento de uma ontologia para outra expressa como se traduzem declarações de uma ontologia para a outra. Muitas vezes, significa a tradução entre conceitos e relações. No caso mais simples é o mapeamento de um conceito da primeira ontologia para um conceito da segunda ontologia. Nem sempre é possível fazer tal mapeamento e algumas informações podem ser perdidas no mapeamento. O importante é que o mapeamento não introduza inconsistências.

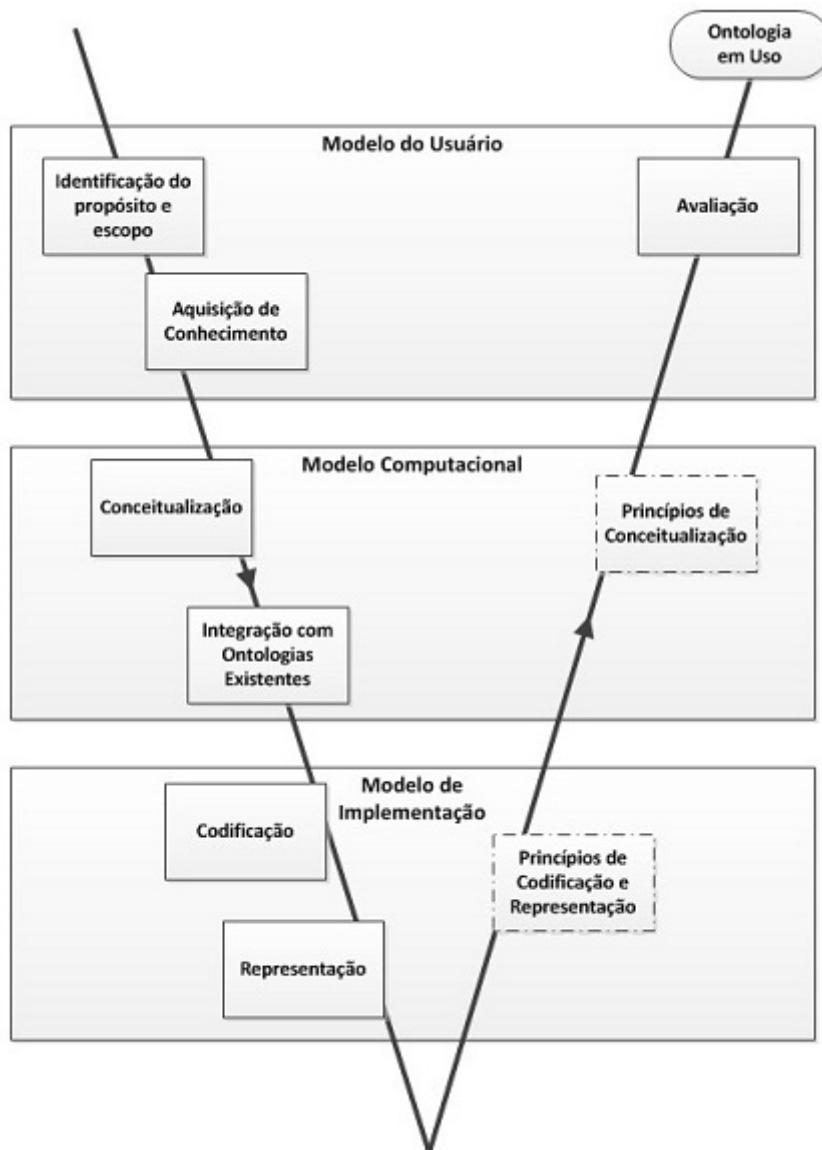


Figura 11: Metodologia de Construção de Ontologias - Modelo-V - Fonte: (OULD, 1990)

- O alinhamento é um processo de mapeamento entre ontologias, modificando ou reorganizando as ontologias originais, de modo que exista uma tradução adequada (isto é, sem perda de informação durante o mapeamento). Pode-se adicionar novos conceitos e relações nas ontologias de maneira a formar equivalências adequadas para o mapeamento.

### 3.5.5 Padrão de Ontologias Biológicas

As ontologias construídas nesta pesquisa seguem o padrão das ontologias biológicas, OBO Flat File. [Smith e Ceusters \(2005\)](#) explicam que as ontologias OBO são projetadas para atender as necessidades da comunidade de biólogos e, então, dispõem de qualidades únicas, proporcionando a habilidade para rastrear grande quantidade de metadados, incluir mecanismos para auditoria de histórico e, principalmente, dispor dos tipos de relacionamentos suficientes para a construção de ontologias de termos biológicos, envolvendo classes e relações.

### 3.5.5.1 Classes

Classes (também chamadas “termos”) são tipos de modelos de objetos ou processos do mundo real. As classes modelam tipos, não instâncias. As classes também podem ter relacionamentos umas com as outras. Por questões de padronização, recomenda-se que todos os termos sejam registrados no singular e em letras minúsculas.

### 3.5.5.2 Relações

As relações, também chamadas de relacionamentos ou, ainda, propriedades, modelam tipos de relacionamentos entre as entidades, podendo ser aplicadas a nível de classes ou a nível de instâncias. No primeiro caso, deve existir pelo menos uma possível aplicação a nível de instância daquela relação.

Uma determinada relação pode ser aplicada de duas maneiras: como relações a nível de classes, ou relações a nível de instâncias. Para cada aplicação do nível de classe de uma relação deve existir pelo menos uma possível aplicação a nível de instância daquela relação.

As relações podem ser:

- **Direcionadas:** as relações são aplicadas em uma única direção;
- **Simétricas:** as relações são aplicadas em ambas as direções;
- **Transitivas:** as relações permanecem verdadeiras através de encadeamentos de ligações;
- **Cíclicas:** as relações formam um ciclo encadeado de ligações entre elas.
- **de Domínio:** o domínio e escopo da relação implicam agrupamentos em tipos.

Assim como as classes, as relações podem ter subtipos.

As relações pré-definidas do formato OBO são as seguintes:

- **is\_a:** Relação de subtipo. Esta relação indica que uma entidade é um subtipo de outra. No caso de classes, denota uma subclasse. No caso de relações, denota uma sub-relação;
- **disjoint\_from:** É uma relação simétrica que indica que duas classes são desunidas. Se duas classes A e B são marcadas como desunidas, nenhuma subclasse de A pode ser uma subclasse de B, e nenhuma subclasse de B pode ser uma subclasse de A.
- **union\_of:** Esta relação indica que uma classe é definida como uma união de outras classes. Não se aplica a relações.
- **inverse\_of:** Esta relação simétrica indica que uma relação é inversa a outra relação (Ex.: “part\_of” é inversa a “has\_part”).

### 3.5.5.3 Web Ontology Language - OWL

As classes OBO são análogas às classes OWL. OWL é uma linguagem baseada em XML, com semântica formal, utilizada para instanciar ontologias na Web, descrevendo indivíduos, propriedades e classes com suas respectivas propriedades e relacionamentos.

- Indivíduos: representam objetos no domínio de discurso.
- Propriedades: são relações binárias que ligam os indivíduos. Podem ser funcionais, transitivas, simétricas ou inversas.
- Classes: são conjuntos que contêm indivíduos com base em algum critério de categorização. Há uma hierarquia de superclasses e subclasses que caracterizam uma taxonomia. As subclasses especializam as superclasses. As superclasses generalizam as subclasses.

OWL é recomendada como padrão W3C e compreende três tipos de sub-linguagens:

- OWL Lite: permite classificação hierárquica e restrições simples, permitindo importação relativamente rápida de tesouros e taxonomias para ontologias.
- OWL DL: baseia-se na Lógica Descritiva e permite máxima expressividade, computabilidade e decidibilidade, com restrições simples.
- OWL Full: permite máxima expressividade, liberdade sintática, porém sem garantia de computabilidade.

[Samuel e Obrst \(2008\)](#) propõem um sistema de tradução de OWL e Web Semântica para regras em Prolog.

### 3.5.5.4 Princípios da Ontologia BioTop

[Hancock e Zvelebil \(2004\)](#) explicam que a enxurrada de dados e o conhecimento factual em biologia e medicina exigem uma abordagem baseada em princípios para a sua análise e gestão. A pedra angular desse esforço consiste da descrição exata e completa das entidades fundamentais dentro do domínio em termos de ontologias formais.

No entanto, o desenvolvimento de ontologias biomédicas, muitas vezes ainda não adere aos princípios básicos de *design*: por exemplo, termos do domínio muitas vezes não têm definições precisas e inequívocas. Esta situação impede o movimento em direção à padronização semântica pretendida. Pelo contrário, ela leva a inconsistências, fragmentações e sobreposições, tanto dentro das próprias ontologias como entre outras diferentes.

À luz disto, a BioTop foi criada como uma ontologia de domínio de alto-nível (top), que fornece definições para as entidades fundamentais da biomedicina, como um vocabulário básico de forma inequívoca, descrevendo fatos nesse domínio. A BioTop pode, além disso, servir como



um modelo de nível superior para a criação de novas ontologias em domínios mais específicos ou como auxílio para alinhar ou melhorar outras ontologias existentes.

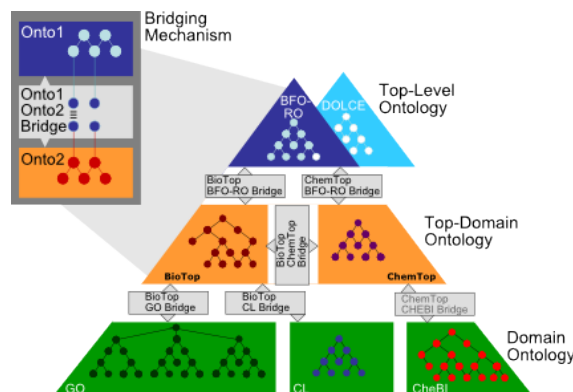


Figura 12: Ontologia BioTop - Fonte: (HANCOCK; ZVELEBIL, 2004)

A BioTop está fundada sobre princípios de *design* formais e implementada em OWL-DL (vide seção 3.5.5.3), a ontologia padrão para a web semântica. Ela faz uso de toda a gama de construtores da OWL-DL. A utilização de OWL-DL torna possível aplicar raciocinadores lógicos, descrição, manter a consistência por classificação contínua durante o desenvolvimento e inferência automática de sua estrutura hierárquica.

## 3.6 Fundamentos da Genética

Como veremos no Capítulo 4, os elementos no domínio da genética da *Arquitetura Genética da Informação* são o genoma, o genótipo e o fenótipo, cujas definições são apresentadas a seguir:

### 3.6.1 Genética

Segundo a “*Encyclopedia of Biology*” (RITTNER; MCCABE, 2004), Genética “é o estudo científico da hereditariedade e variação”. A “*Encyclopedia of Genetics*” (KNIGHT; NESS, 2004) explica que a Genética é um ramo da biologia relativamente novo que explora os mecanismos da hereditariedade, com impacto em praticamente todos os outros ramos da Biologia, bem como na agricultura, farmacologia e medicina. É uma ciência ampla que examina a base molecular da herança em nível celular, a transmissão dos traços de geração em geração e o movimento dos genes dentro de e entre as populações.

Windelspecht (2007) argumenta que mesmo as civilizações mais antigas já aplicavam os conceitos da genética, ao reconhecerem que traços específicos poderiam ser selecionados para o desenvolvimento de melhores linhagens de certos animais. Os mesmos princípios de engenharia genética eram aplicados a plantas, também.

Ao estudar como os seres vivos recebem traços comuns de gerações prévias, a Genética utiliza o termo “informação genética”, que é a informação carregada por uma molécula de DNA. As instruções para a construção e funcionamento de um organismo estão contidas no DNA desse organismo. Toda criatura viva na Terra tem DNA em suas células. Em termos gerais, um gene é uma unidade hereditária que consiste de DNA ocupante de uma posição no cromossomo e que determina uma característica (observável ou não) de um organismo. Os genes são passados de pai para filho sendo uma parte importante na determinação da aparência e do comportamento.

### 3.6.2 Genoma

Genoma é a soma total dos genes de uma espécie (KNIGHT; NESS, 2004). A função dos genes é serem preservados e transmitidos. Os genes controlam várias funções biológicas através da produção de proteínas e RNA. A estrutura dos genes pode ser de dois tipos de ácidos nucleicos: o DNA e o RNA. Os ácidos nucleicos, por sua vez, são compostos por nucleotídeos: açúcar, fosfato e bases nitrogenadas. O açúcar, no caso do DNA é a desoxirribose; e no caso do RNA é a ribose. As bases nitrogenadas do DNA<sup>4</sup>, podem ser a Adenina (A), Guanina (G), Citosina (C) e Timina (T). As possíveis combinações são A-T e C-G, conforme ilustra o diagrama apresentado na figura 63, na pág. 190 do Glossário.

Avery (2003) argumenta que a tradução da informação presente no genoma para um traço fenotípico ocorre de maneira extremamente complexa.

“A nossa compreensão deste processo fabuloso é ainda muito limitado, mas há evidências de que as células mudam os seus estados em resposta aos estados das células vizinhas. O estado de uma célula significa a maneira pela qual ela se diferencia das demais, i.e., quais genes são habilitados e quais não são, ou ainda, que informação do genoma está disponível para leitura e quais segmentos estão bloqueados”.

Essa concepção de genes habilitados e não-habilitados correspondendo a valores binários 1 e 0, respectivamente, é considerada nas redes biológicas booleanas (tópico 6.4.6).

### 3.6.3 Genótipo

Lewontin (1992) definiu o genótipo de um organismo como sendo “a classe da qual é membro, com base no estado de seus fatores hereditários internos, os genes”.

A expressão de um traço é devida à combinação de traços internos ao organismo. As instruções para a produção desses traços é chamada de genótipo e estão contidas dentro do DNA do organismo, em regiões chamadas genes. Cada gene pode ter um número de variantes, chamados de alelos. Um alelo é devido à menor mudança no material genético e o genótipo é a combinação de alelos de um organismo.

Os genes estão localizados nos cromossomos, em posições específicas, chamadas de locus. Cada espécie tem um número específico de cromossomos e os genes são arrançados em uma

<sup>4</sup>As bases do RNA são as mesmas do DNA, exceto que a Uracila (U) substitui a Timina (T)

específica ordem, chamada de *loci*, dentro do cromossomo. Para cada gene pode haver variações na informação que ele contém (alelos). Uma vez que os traços são determinados pelos genes e os genes estão localizados nos cromossomos, podemos dizer que tomamos um conjunto de cromossomos de cada ascendente imediato (pai e mãe). A informação em cada cromossomo é usualmente arranjada da mesma maneira, mas uma vez que os genes podem ter diferentes alelos, os cromossomos em um par são raramente idênticos. A similaridade dos cromossomos chamamos homólogos. Organismos diplóides são aqueles que têm um par de cada cromossomo.

Para cada traço, um diplóide individual pode ter outros dois alelos idênticos para o traço, ou dois diferentes alelos. Tendo-se o mesmo alelo para um dado traço, diz-se que é homozigoto para aquele traço. Tendo-se diferentes alelos para o mesmo traço, diz-se que é heterozigoto para aquele traço. Um indivíduo homozigoto recessivo tem dois alelos recessivos para o traço e um indivíduo dominante homozigoto tem dois alelos dominantes. Assim, Brooks (2009) resume dizendo que, nas interações entre alelos, temos:

- Dominância total - um alelo domina o outro;
- Dominância incompleta - nenhum dos alelos é totalmente manifestado;
- Codominância - os dois alelos manifestam-se por inteiro;
- Alelos múltiplos - mais de dois alelos para um mesmo gene podem ser encontrados na população;
- Epistasia - um gene altera o efeito de outro gene;
- Herança poligênica - muitos genes contribuem para um fenótipo;
- Pleiotropia - um gene influi em vários fenótipos;
- Influência ambientais - o genótipo e o ambiente interagem para formar o fenótipo.

#### 3.6.4 Fenótipo

Lewontin (1992) definiu o fenótipo de um organismo como sendo “a classe da qual é um membro, com base nas qualidades físicas observáveis do organismo, incluindo sua morfologia, fisiologia e comportamento, em todos os níveis de descrição”.

A Enciclopédia Britânica define: “*Fenótipo* são todas as características observáveis de um organismo” (EBONLINE, 2012b). Enfatizando, temos que “fenótipos são quaisquer atributos físicos e químicos observáveis nas células, alguns dos quais, sujeitos à seleção” (SAWAI, 2011). Esses traços ou características podem ser a sua forma, tamanho, cor, e comportamento, que resultam da interação de sua herança genética total (genótipo) com o meio-ambiente. No uso comum, o termo “fenótipo” é muitas vezes empregado como sinônimo para “traço”, mas estritamente falando, não indica um traço, mas o seu estado (por exemplo, o traço ‘cor dos olhos’

tem o fenótipo azul, castanho, verde etc). Um traço pode ser qualquer característica singular ou de medição quantificável de um organismo. Um traço visível é o produto final de inúmeros processos moleculares e bioquímicos.

Na maioria dos casos, a informação começa no DNA, viajando para o RNA e, finalmente, produzindo proteínas (que em última análise, afeta a estrutura e funcionamento dos organismos vivos). Esse fluxo de informação passa do núcleo da célula para o citoplasma, ribossomos, retículo endoplasmático, e finalmente para o Complexo de Golgi, que empacota o produto final para o exterior da célula. Os produtos celulares formam os tecidos e órgãos de um organismo, afetando a sua fisiologia de maneira a produzir um traço fenotípico.

### 3.7 Biologia e Informação

Bawden (2007) relata que Madden (2004) deve ter sido a primeira pessoa a propor que as percepções da Ciência da Informação podem ser valiosas para as ciências mais “básicas”, nesse caso, o das ciências biológicas, por causa da grande semelhança do conceito de “informação”.

Termos relacionados à “informação” são largamente usados na biologia molecular e desenvolvimental.

Smith (2010) explana que os genes são símbolos sem necessária conexão entre sua forma (sequência) e seus efeitos. No processo biológico de seleção natural, o uso de termos da informação implicam intencionalmente tanto na forma como no sinal e sua consequente resposta. Onde os engenheiros vêm *design*, os biólogos veem seleção natural.

“Informação” é uma ideia central na biologia contemporânea. A biologia do desenvolvimento estuda como a informação no genoma pode ser traduzida em uma estrutura adulta.

Tópicos como origem dos genes, das células e da linguagem todos dizem respeito ao armazenamento e transmissão de informação.

Já, Mahner e Bunge (1997) discordam, pois para eles, aparentemente, as noções envolvidas no paradigma físico da informação (codificador, transmissor, receptor e decodificador) não se aplicam a um sistema químico, pois um processo químico não carrega mensagens e no caso da Biologia, estas seriam desprovidas de ruídos, sendo que o conceito de probabilidade, central na teoria da informação, não se aplicaria a esse tipo de alegada transferência da informação.

Smith (2010) (no Cap. 7 do livro de Davies e Gregersen (2010)) é enfaticamente contrário a esta ideia, lembrando que o uso coloquial dos termos de informação é pervasivo na biologia molecular.

Os termos “transcrição”, “translação”, “código”, “leitura”, “cópia”, “redundância”, sinônimo”, “mensagem”, “edição”, “revisão” e “biblioteca”, por exemplo, são todos termos técnicos em Biologia. As similaridades entre os seus significados nas áreas de comunicação humana e genética são surpreendentemente próximas. Por exemplo, na “revisão”, a sequência das quatro bases em uma nova fita de DNA sintetizada é comparada com a sequência correspondente de uma antiga fita que agiu como modelo para esta síntese. Existindo uma correspondência (i.e., se

a base da nova fita não é complementar àquela na antiga fita de acordo com as regras de pares, A-T e C-G), então esta é removida e trocada pela base correta. Parece-nos óbvia a similaridade desse processo com aquele em que as letras em uma cópia são comparadas com as da original, sendo corrigidas quando diferem.

Os biólogos moleculares, então, fazem uso da analogia com a informação em seus trabalhos de rotina. Analogias são usadas na ciência de duas maneiras. Ocasionalmente, existe um isomorfismo formal entre dois diferentes sistemas físicos, apesar de que casos de isomorfismos exatos sejam muito raros.

Muito mais comum é o reconhecimento de uma similaridade qualitativa, útil em dar-nos uma visão para dentro de um sistema não muito familiar pela comparação com outro mais acessível ao entendimento. Mais frequentemente, biólogos se encontram fazendo analogias com a tecnologia atual, ou às vezes, com as ciências sociais. É claro que, se uma analogia é apenas qualitativa, pode tanto nos induzir ao erro como, por outro lado, nos iluminar.

Auletta (2010) afirma que todos os novos desenvolvimentos em Biologia lidam com aspectos de complexidade dos organismos, frequentemente evidenciando a necessidade de atualização das ferramentas que se dispõem para a compreensão, o que seria impossível sem introduzir novos mecanismos explanatórios.

Ele, então passa a apresentar os conceitos de Teleonomia e Teleologia. A Teleonomia diz respeito à habilidade dos organismos de construir a si mesmos por meio de forças e processos internos (na expressão do programa genético), com uma certa liberdade do mundo exterior, embora o organismo seja capaz de integrar restrições internas e externas em um processo de co-adaptação. Por outro lado, a Teleologia<sup>5</sup> é o mecanismo pelo qual um organismo exercita um **controle da informação** sobre um outro sistema de forma a estabelecer uma classe de equivalência e selecionar algumas informações específicas para suas necessidades metabólicas. Finalmente, o autor examina alguns processos interessantes em filogenia, ontogenia e epigenia, em que estes dois mecanismos estão envolvidos.

Um outro conceito associado à informação é o seu contexto (que na *Arquitetura Genética da Informação* é discutido na seção 4.1). Nijhout (2003) argumenta sobre a importância do contexto em Genética, explicando que os efeitos genéticos podem ser alterados pelo seu contexto ambiental, que é determinante na herança de informações ou traços fenotípicos, contribuindo, também, na probabilidade de ocorrência de características de anormalidade. À época dos trabalhos do pai da genética, Gregor Mendel, no sec. XIX, pensava-se que as variações de um traço deviam-se exclusivamente às variações de um único gene. Com o avanço científico, tornou-se claro que a maioria dos traços, ou fenótipos, são herdados de uma maneira mais complexa do que aquela descrita por Mendel. Criou-se, então, o termo “*traço complexo*” para descrever um

---

<sup>5</sup>Este conceito de teleologia tem íntima relação com o conceito marxista de práxis, que em sua obra ‘O Capital’, compara a atividade das abelhas, ao construir a colméia, com o trabalho de um mestre-de-obras ao construir uma casa. Por mais perfeita que seja a construção da colméia, e por mais limitado que seja o trabalho do mestre de obras, esse último possui algo essencialmente diferente: o mestre-de-obras imagina o que vai realizar, criando uma finalidade, um momento ideal, o qual almeja alcançar com seu trabalho.

fenótipo cuja herança de geração em geração não obedece às leis de Mendel. Uma maneira de compreender estes traços complexos é pela análise de seu contexto, i.e., dos mecanismos bioquímicos pelos quais os genes afetam um fenótipo. É comum a realização de estudos, como pela indução artificial de mutações e modificadores, por exemplo, gerando resultados diversos. A importância do contexto pode também ser ilustrada pelos estudos sobre os efeitos de “nocautes” de genes específicos em um animal de laboratório. Estes efeitos eliminam completamente a função de um produto de um gene, podendo causar graves anormalidades em uma linhagem ou, por outro lado, não causar efeito algum em outra. Na hipótese de aditividade, temos que cada fator que afeta uma característica tem um efeito temporal em si mesmo, e a soma destes efeitos pode produzir uma grande influência observável sobre o fenótipo. Na verdade, não se sabe quais são todos os fatores que contribuem para uma característica fenotípica, daí somente podemos trabalhar com probabilidades e previsões. Quando uma saída não corresponde diretamente a uma entrada, tem-se o princípio de não-linearidade e sabe-se que é esse princípio que produz um traço dominante, diante de um outro recessivo. Isso mostra que a dominância não é uma propriedade do gene em si mesmo, ou seja, emerge do contexto.

O efeito da mutação em uma característica, assim como o traço de dominância, depende exatamente onde ele está na paisagem fenotípica. Em outras palavras, o efeito não é uma propriedade da mutação em si, mas uma função de todo o sistema. [Lobo e Shaw \(2008\)](#) reforçam dizendo que é óbvio que os genes não são os únicos fatores que determinam o fenótipo, mesmo gêmeos idênticos não são exatamente iguais. Existe a influência ambiental.

[Reading \(2011\)](#) propõe estabelecer uma ponte que liga a Biologia, o cérebro e o comportamento, introduzindo uma maneira radicalmente nova de pensar sobre a informação e o seu papel central desempenhado nos sistemas vivos. Ele abre novos caminhos para explorar como células e organismos mudam e se adaptam, já que a capacidade de detectar e responder à *informação intencional* é a chave que lhes permite decifrar o seu patrimônio genético, que regula o seu ambiente interno e viabiliza a navegação segura por seus arredores. Ele argumenta que a informação é uma função da maneira pela qual matéria e energia são organizadas e dispostas no espaço e no tempo, como sendo propriedades de sua forma, e não de sua substância, apesar de forma e substância serem inseparáveis (uma vez que é impossível conceber uma forma que não tenha substância, ou uma substância que não tenha forma).

Mas os padrões de matéria e energia apenas transmitem informação para alguém ou algo que possa detectá-la e decifrá-la, tornando-se, então, significativa. Cada tipo de célula e cada espécie viva é dotada de receptores específicos que lhes permitem perceber um subconjunto de tais padrões de informação. A informação significativa é, então, definida como um padrão de matéria ou energia que é detectado por um receptor (entidade viva ou fabricada), que desencadeia mudanças comportamentais, funcionais ou estruturais na própria entidade. Padrões detectados de matéria ou energia que não têm nenhum efeito sobre um receptor são consideradas insignificantes, pelo menos naquele momento particular. Esse conceito fornece uma maneira de entender como os seres vivos interagem com o ambiente e um com o outro.

Os temas abordados por [Reading \(2011\)](#) vão desde a sinalização celular até a tomada consciente de decisões; da formação e organização de categorias até o comportamento orientado a metas; das informações genéticas e emocionais até as formas analógicas e digitais de representação; e da teoria da informação e circuitos neurais de má adaptação comportamental até a interface mente-cérebro. Estes estudos contribuem para esta tese em termos das considerações espaço-temporais; propriedades da forma; transmissão de informação; detecção de padrões de informação; e mudanças comportamentais, funcionais e estruturais no objeto ou organismo.

[Davies e Gregersen \(2010\)](#) argumentam que evidências impressionantes para a visão naturalista da linguagem parecem ser encontradas na disposição semelhante à linguagem de informação genética.

“Assim, como é bem sabido, o alfabeto genético é agrupado em unidades de informação de ordem mais elevadas, o que, na escrita genética assume funções de palavras, frases, e assim por diante. E, como a linguagem humana, a genética tem uma estrutura hierárquica, que se desdobra em um complexo mecanismo de realimentação - que apresenta todas as propriedades de um processo de comunicação entre o genoma e o seu contexto físico.”

[Stonier \(1990\)](#) explica que dentre tantas importantes descobertas, os cientistas biológicos descobriram o fato de que:

“As mensagens transmitidas pelos sistemas de informação biológicos são aparentemente entendidos por todas as formas de vida existente neste planeta. . . A quantidade e natureza da informação contida no DNA pode variar de um para outro organismo, mas o método de codificação dentro de uma molécula de DNA é o mesmo”.

[Meyer \(2007\)](#) explica sobre o DNA e a origem da vida, cujo problema central é a informação, especificação e explanação.

[Vedral \(2010\)](#) afirma que a primeira grande aplicação da informação acontece na Biologia, onde a Genética desenvolveu-se inteiramente usando a linguagem da preservação e transmissão da informação. Na Genética a informação é de fácil entendimento e tem um significado bem definido. Apesar de sua rigidez, tem princípios que podem ser universais para a Arquitetura da Informação. Daí, a analogia que identificamos entre as duas disciplinas.





**Parte III**

**Resultados**



## 4 Arquitetura Genética da Informação

*“As melhores coisas na vida não são coisas”*  
— Art Buchwald

Na Biologia, conforme explicado na seção 6.2, o termo “arquitetura genética” refere-se ao mapeamento genótipo-fenótipo, i.e., à influência implícita do genótipo em um traço fenotípico do organismo.

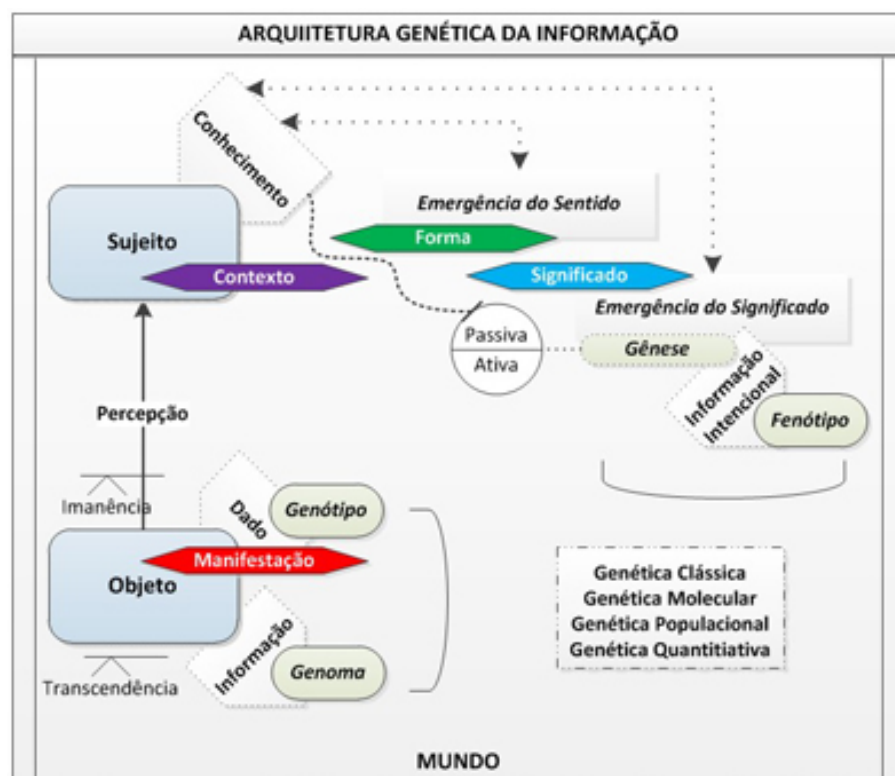


Figura 13: Arquitetura Genética da Informação - Fonte: Produzido pelo autor

Na Arquitetura da Informação, utilizamos o termo “arquitetura genética da informação” para modelar o processo de gênese da informação e, analogamente, realizar o mapeamento do genótipo de quaisquer objetos perceptíveis no mundo para o seu respectivo registro fenotípico, em termos da informação.

A *Arquitetura Genética da Informação*, apresentada no diagrama da figura 13, está fundamentada nos aspectos ontológico, epistemológico, fenomenológico, científico e pragmático da informação, em consonância com o modelo da Teoria Geral da Arquitetura da Informação proposta por Lima-Marques (2011).

No **nível ontológico**, temos o mundo compreendido por entes (seres e coisas) inseridos em um contexto espaço-temporal. Os seres são os sujeitos e as coisas, os objetos, dentro da perspectiva de [Siqueira \(2012\)](#).

A informação corresponde às propriedades ontológicas do objeto, independentemente do sujeito, conforme defende [Lima-Marques \(2011\)](#). A totalidade de possíveis características ontológicas do objeto constitui o *genoma da informação* e aquelas que são perceptíveis (não necessariamente observáveis) correspondem ao seu *genótipo*.

No **nível epistemológico da fenomenologia**, os dados do objeto, quando percebidos pelo sujeito, fazem emergir subjetivamente o sentido e significado ([SAAB; RISS, 2011](#)), em um fenômeno de gênese da *informação intencional*, pela representação do objeto na consciência, em um processo de aquisição de conhecimento. De acordo com a Fenomenologia Genética, a gênese pode ser passiva, quando faz uso de conhecimento prévio, estruturando objetos na consciência, em camadas sedimentadas umas sobre as outras; ou ativa, quando os objetos são totalmente novos à consciência (consultar seção [2.4.2](#)). A *informação intencional*, por sua vez, é caracterizada por seu *fenótipo*, que corresponde a todas as características observáveis do objeto.

No **nível científico**, a analogia dos dois domínios foi realizada com a construção de ontologias de domínio, amparada por ontologia de alto-nível, conforme especificadas na seção [7.5](#).

No **nível da práxis**, estas características genéticas da informação podem ser consideradas em termos clássicos (dominância e herança entre objetos), moleculares (interações internas do objeto), populacionais (o conjunto dos objetos) ou quantitativos (estatísticos). Todos estes aspectos são contemplados nas ontologias construídas. (Sobre o conceito marxista de praxis, que se relaciona à teleologia, consultar seção [3.7](#)).

## 4.1 Elementos da Arquitetura

Retomando a etimologia, o termo “informação” vem do latim *informatio* + *onis*, que significa “dar forma ou moldar uma ideia ou ação na mente”, significando “primeiro” ou “principal” e *tékton*, significando “construção”, ou seja, “construção de um espaço principal ou primitivo”.

Os elementos da *Arquitetura Genética da Informação* apresentados na figura [13](#) são, no aspecto da Teoria Geral da Informação: informação e dado; no aspecto fenomenológico: mundo, sujeito, objeto e conhecimento; no aspecto da disciplina de Arquitetura da Informação: manifestação, forma, contexto e significado; no aspecto genético: genoma, genótipo e fenótipo. Outros componentes da arquitetura são os processos de emergência de sentido e significado no processo de gênese passiva e ativa. Esses processos foram introduzidos no tópico [2.4.1](#) e detalhados na seção [5.1](#).

Então, além do conceito de “informação”, em seu aspecto absoluto e ontológico, existe o conceito da informação intencional, que tem caráter subjetivo, sendo resultante da emergência do sentido e do significado, no processo de cognição pelos dados percebidos do objeto e pelo

conhecimento 'a priori' do sujeito, ocorrendo em sua consciência, após a apreensão e com a possibilidade, mesmo que idealizada, de realização da suspensão de juízo, ou *bracketing*. Esta abordagem enquadra-se no paradigma cognitivo da informação, explicado na seção 3.3.2.

Assim, ao invés de nos posicionarmos pela informação objetiva ou pela informação subjetiva, como ocorre em vários debates importantes na Ciência da Informação, a exemplo do artigo de Hjørland (2007)<sup>1</sup> em oposição a Bates (2005) e, também, do artigo de Zaliwski (2011), consideramos efetivamente o convívio das duas possibilidades, em que no objeto, intrinsecamente existe a informação física objetiva; e no sujeito, após o fenômeno de observação, as propriedades percebidas geram conhecimento em um processo de emergência de sentido e de significado, quando ocorre, então, a gênese da *informação intencional*, subjetiva, situacional.

Logan (2012) afirma que:

“Há três formas de informação que impactam a condição humana: a informação genética de nosso DNA, a informação perceptual que detectamos com os nossos sentidos e as informações conceituais que processamos com a nossa mente”

Na *Arquitetura Genética da Informação*, a informação genética de nosso DNA corresponde ao genoma do objeto, à informação absoluta e ontológica, propriamente dita; a informação perceptual corresponde aos dados sensoriais, genótipo do objeto; e as informações conceituais correspondem à *informação intencional*.

Para Christiansen (1995), a linguagem pode ser considerada como um organismo. Logan (2012) estendeu isso para vários outros aspectos:

“Postulamos que a cultura, como um organismo, evoluiu de tal forma a ser facilmente aprendida. . . Linguagem, cultura, tecnologia, economia e governança e ciência podem ser tratados como organismos que evoluem, propagam sua organização e representam fenômenos emergentes”.

Nesse sentido, a *Arquitetura Genética da Informação* está habilitada a modelar todos esses aspectos, uma vez que os mesmos podem ser considerados organismos vivos.

De fato, se por um lado, temos a explicação de Shannon (1948) afirmando que o problema fundamental da comunicação é reproduzir exata ou aproximadamente em um ponto uma mensagem selecionada em outro ponto, em que os “aspectos semânticos da comunicação são irrelevantes para o problema de engenharia”, onde, no aspecto significativo, a mensagem real é selecionada de um conjunto de mensagens possíveis, para isso, “o sistema deve ser projetado para levar em conta cada seleção possível, não apenas aquela que de fato vai ser escolhida, porque isso é ignorado no momento do *design*”. E, ainda, nas palavras do próprio Shannon (1953):

“A palavra 'informação' recebeu muitos significados diferentes por vários autores no campo geral da teoria da informação. . . Dificilmente se pode esperar que um único

---

<sup>1</sup>Hjørland (2007) contrasta o entendimento de informação como um fenômeno independente do observador, argumentando que informação é situacional, apoiado por Bateson, Yovits, Spang-Hanssen, Brier, Buckland e Goguen. Os conflitos existentes entre os caminhos objetivo e subjetivo de compreensão da informação correspondem ao conflito entre a ideia de informação como coisa e informação como um signo. Trata-se de uma distinção fundamental que envolve toda uma teoria do conhecimento e tem suas raízes nas diferentes metáforas aplicadas à Teoria de Shannon.

conceito de informação vá explicar satisfatoriamente as inúmeras aplicações possíveis desse campo em geral”.

Por outro lado, temos a posição de Mackay, que segundo Hayles (1999):

“Ele propôs que tanto Shannon quanto Bavelas estavam preocupados com o que chamou de ‘informação seletiva’, que é a informação calculada considerando-se a seleção de elementos de mensagem de um conjunto. Mas a informação seletiva por si só não é suficiente; também é necessário um outro tipo de informação que ele chamou de ‘estrutural’. Informação estrutural indica como a informação seletiva deve ser compreendida”.

Disso, Logan (2012) depreende que:

“A informação estrutural deve envolver a semântica e o sentido se ela pretende ser bem-sucedida no seu papel de interpretar a informação seletiva. . . A informação estrutural está preocupada com o efeito e o impacto da informação na mente do receptor e, portanto, é reflexiva. . . tem uma relação com a pragmática, bem como com a semântica, na qual tenta preencher uma lacuna explicativa entre o significado literal de uma frase e o significado pretendido pelo falante ou escritor”.

A gênese diz respeito ao momento em que, subjetivamente, a *informação intencional* surge. O genoma, ao caráter de informação da classe de determinado objeto sob observação. O genótipo, aos traços potencialmente manifestáveis e transmissíveis do objeto. E, o fenótipo, aos traços observáveis do objeto. Com o objetivo de desenhar uma distinção fundamental entre forma, aparência e substrato de um objeto, Smith (2004) buscou “distinguir a intencionalidade da consciência (sua forma), a maneira como experienciamos a consciência (sua aparência, incluindo o qualia), e a base física, biológica e cultural da consciência (seu substrato)”.

Podemos identificar aqui os três elementos da arquitetura no aspecto genético: o genoma do objeto intencionado, o genótipo do objeto percebido e o fenótipo do objeto observado.

Cada perfil particular de um objeto (genótipo) refere-se ao objeto como um todo (genoma), como se todos os seus possíveis perfis pudessem ser percebidos neste. As propriedades do objeto como um todo (genoma), ou seja, o objeto intencionado é o que unifica todos os perfis dados em vários atos de percepção (do genótipo) do objeto. Cada percepção antecipa as outras percepções, em um processo de preenchimento. A totalidade dos fatores não efetivamente ou não imediatamente dados – i.e., o objeto em todos os seus possíveis perfis (genoma) – é chamado de horizonte interno, e o plano de fundo contra o que o objeto aparece (fenótipo) é chamado de horizonte externo (genótipo). Daí, a constituição do objeto é a unidade dos atos da consciência, de todos os perfis percebidos (fenótipo) nos horizontes interno (genoma) e externo (genótipo).

Quanto aos aspectos de transcendência e imanência, assunto apresentado no tópico 2.4.3 e, aqui, aplicado à arquitetura proposta: O aspecto de transcendência (o que diz respeito a todo o objeto intencionado), está relacionado ao genoma do objeto; o aspecto de imanência (o que é dado do objeto no processo de percepção) está relacionado ao genótipo do objeto; e o aspecto da constituição do significado (o que se mostra como característica do objeto na consciência) está relacionado ao fenótipo do objeto.

Todos estes elementos participam de contextos peculiares na Arquitetura Genética da Informação.

Rysiew (2009) explica que, em sua forma dominante, a posição do Contextualismo Epistêmico (EC) é a visão de que a proposição expressa por uma frase de conhecimento 'S conhece que p', 'S não conhece que p' depende do contexto em que ela foi dita. O "contexto" se refere não a algumas características do assunto do conhecimento colocado (evidências, história, crenças etc) ou a sua situação objetiva (o que é falso/verdadeiro, quais alternativas daquilo que se crê estão próximas a serem obtidas etc), mas ao contrário, refere-se às características psicológicas daquele que atribuiu o conhecimento e/ou a situação prático-conversacional. Como resultado, existe uma dependência contextual, que pronunciadas determinadas sentenças em algum contexto diferente, diferem no seu valor de verdade.

Day (1996) ressalta que "a informação não aparece como uma entidade autônoma e é sempre representada em função de um contexto", conforme discorremos, dentre outros assuntos, na seção 3.7.

Quando nos preocupamos com o significado da informação, inevitavelmente nos deparamos com a sua inserção em um contexto (em oposição à dependência exclusiva de definições, fatos e verdade). Derrida (1971), por exemplo, diz que "um texto qualquer pode ser interpretado para significar qualquer coisa, bastando que se mude o seu contexto".

Marroquín e Italo (2007) argumentam que a pretensão de entender a coisa em sua essência constitutiva de maneira desvinculada de sua realidade espaço-temporal inviabiliza a sua idealidade, não tendo qualquer sentido ou apoio para a sua existência em si, sendo na verdade inexistente. Como poderíamos identificar as características essenciais do que não está inserido em uma realidade? Ele insiste que, se algo existe, tem uma razão de ser, comprometendo a sua essência com a sua pretendida fundamentalidade na realidade. Parece mais razoável que a essência, por ser o núcleo fundamental da coisa, longe de ser indiferente à sua existência fática, compromete-se com o seu devir, sem o qual seria inexplicável o fim expresso de sua pretendida fundamentalidade. Se a essência se aplica geralmente a uma pluralidade de casos concretos que a realizam, indistinta, invariável e uniformemente, sem interessar o número de exemplares individuais ou os estados variáveis próprios de cada um; e, ainda, se a essência existe independentemente da existência da realidade por ela essenciada, não significa que a facticidade existencial da coisa essenciada é, por si mesma, incapaz de produzi-la e de explicá-la, tornando-a uma contradição?

Ainda, entendemos que a própria coisa se confunde com o seu contexto, sendo impossível estabelecer os limites entre um e o outro e desassociá-los. Contexto é o ambiente, as circunstâncias, o fundo, ou as configurações que determinam, especificam ou esclarecem o significado de um evento.

Em suma, temos que na relação objeto-mundo, o contexto é extra-sujeito, onde a informação associada ao objeto é contextual e ontológica; e intra-sujeito (na relação sujeito-consciência), o contexto é extra-objeto, onde a informação associada ao sentido e significado é intencional.

Assim, a essência constitutiva da *informação intencional* está totalmente vinculada a uma realidade espaço-temporal, que chamamos de contexto, tendo a sua gênese associada à sua pretendida fundamentalidade (MATURANA, 1978), que ocorre em uma experiência fenomenológica.

Verificamos na definição do Contextualismo Epistêmico que se trata de um contexto discursivo, semântico. Havemos de considerar, também, que existe o contexto observável do mundo real, necessariamente antes da ocorrência da observação propriamente dita, independente, portanto, do observador. É oportuno, portanto, estabelecermos aqui o conceito de contexto em três aspectos distintos: o “contexto observável”, o “contexto perceptível” e o “contexto semântico”.

Em um cenário de natureza ontológica da informação, (na TGAI, estudada na seção 2.4.1, correspondendo ao espaço dos objetos) o aspecto subjetivo não existe, sendo, portanto, anterior à observação. Daí, temos que o “contexto observável” corresponde à ideia de um conjunto infinito de possibilidades de observação dos objetos do mundo, em seu estado latente e dinâmico; anterior ao fenômeno de percepção e, portanto, ontológico e independente do sujeito. Assim como em um organismo biológico, na *Arquitetura Genética da Informação*, cada classe de objetos tem o seu próprio genoma associado à informação. Esse genoma é composto por todas as propriedades ontológicas do objeto, que o tornam diferente dos demais, no nível da informação, sendo anterior ao nível dos dados.

Por outro lado, o “contexto perceptível” vai em direção ao espaço do sujeito e as relações do fenômeno com o mundo ocorrem no nível dos dados, sendo anterior ao nível do conhecimento. Assim, o “contexto perceptível” corresponde à ideia de uma estrutura que se evidencia no processo de percepção, antes que se realize qualquer juízo da parte do sujeito – por isso, também podemos chamá-lo de “contexto epoché” – que existe dentro de um espaço-tempo idealizado fenomenologicamente; finito; contendo todos os seus objetos mensuráveis ou perceptíveis, e somente estes. Na biologia, um indivíduo (instância de um organismo) tem várias possibilidades de manifestação e transferência de informação genética, o que corresponde ao seu genótipo. Analogamente, na *Arquitetura Genética da Informação*, o genótipo ocorre no nível dos dados, em todas as possibilidades de apreensão na relação sujeito-objeto.

A interação do observador com o contexto espaço-temporal revela porções de conhecimento sobre esse contexto visto por ele. Assim, o “contexto semântico” corresponde à ideia de uma estrutura que emerge de um momento específico da observação, tornando-se evidente e significativo, a partir da promoção de uma ordenação mental por coordenadas de natureza semântico-espaço-temporal, intrínsecas ao observador. Nesse contexto, emergem a forma (que dá sentido ao objeto) e o significado, que permitem ao sujeito supor sobre o objeto intencionado. É nesse contexto que ocorre a gênese da *informação intencional*.

Um exemplo prático para a elucidação do conceito de contexto na arquitetura proposta é apresentado no site *Lytro* (CHENG, 2011), onde o autor organizou galerias de fotografias sem intencionalidade, permitindo ao usuário escolher dinamicamente o foco da imagem apresentada, direcionando sua observação para esse foco. Do conjunto de objetos observáveis e mensuráveis uma pequena porção passa a ser o objeto de sua atenção, tornando-se um estado observado em



detalhes, pelas funções *Refocus* ou *Zoom*.

## 4.2 O Arquiteto Genético da Informação

O papel do Arquiteto Genético da Informação está intimamente relacionado com a caracterização dos problemas em conformidade com a *Arquitetura Genética da Informação*, provendo ou apontando soluções adequadas ao que se pretende projetar ou desenvolver, o que pode ser realizado com base no mapa conceitual apresentado na figura 14.

Basicamente, uma vez estudado e compreendido o problema, o arquiteto genético da informação deve identificar e descrever os elementos da arquitetura – sujeito, objeto, contexto, genoma, genótipo, fenótipo etc. – associando-os a uma classe da ontologia genética da informação, que é participante de uma das principais áreas de estudo da genética.

Para tal, propomos a utilização de formulário compreendido por quatro partes: a identificação do problema; os elementos da TGAI; a Arquitetura da Informação Intencional; e a Genética da Informação.

1. Parte I: Problema - descrição do problema;
2. Parte II: TGAI - os elementos são o sujeito, objeto, informação, dado e conhecimento;
3. Parte III: Arquitetura da Informação Intencional - os elementos são as categorias de manifestação, forma, contexto e significado;
4. Parte IV: Genética da Informação - os itens são a classificação no ramo da genética, o genoma (propriedades da classe do objeto), o genótipo (propriedades completas e individuais do objeto, possíveis de participarem de fluxo de informação), fenótipo (propriedades observáveis individuais do objeto), e anotação ontológica (identificação da classe na ontologia proposta).

Um exemplo de caracterização de problemas é apresentado nas figuras 39, 40, 41 e 42, nas páginas 138 a 140.

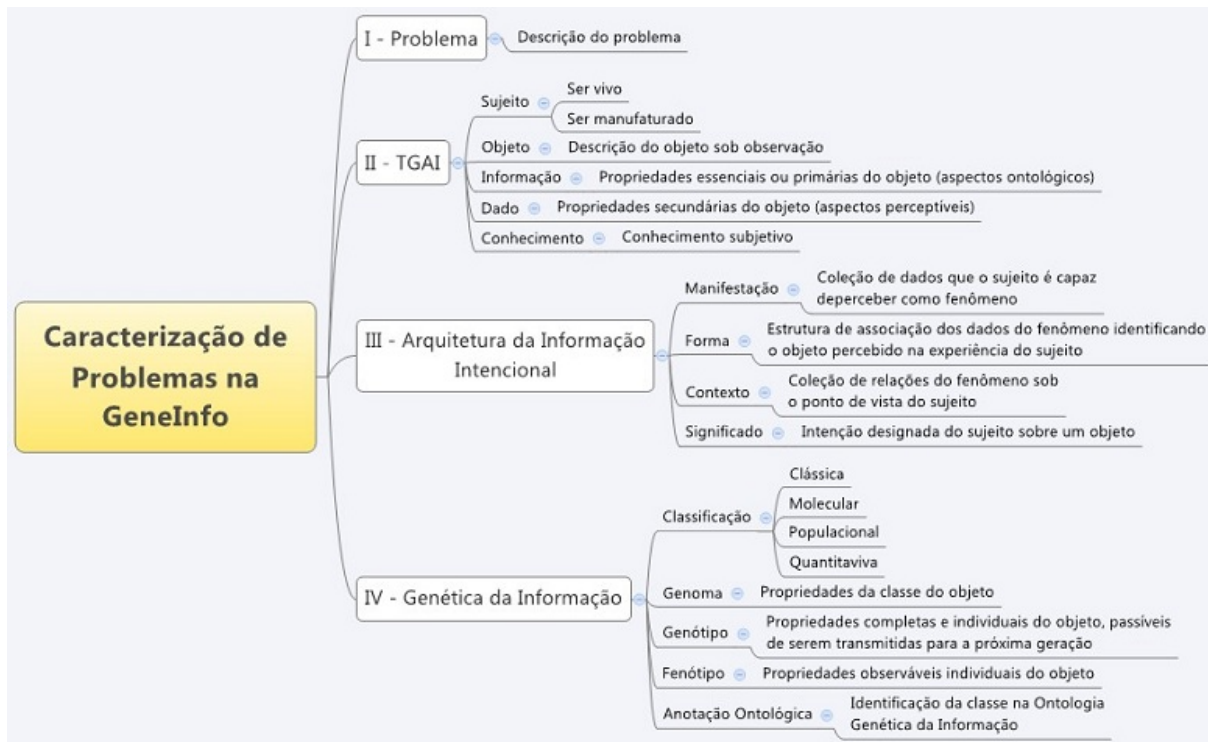


Figura 14: Caracterização de Problemas na Arquitetura Genética da Informação - Fonte: Produzido pelo autor

## 5 Gênese da Informação Intencional

“Onde está todo o conhecimento que perdemos com a informação?”

— T. S. Eliot

O caráter fenomenológico da informação (discutido na seção 3.2) baseia-se na emergência de propriedades de sentido e de significado. Laszlo e Krippner (1998) explicam que “uma propriedade emergente é marcada pelo aparecimento de novas características exibidas no nível da totalidade de seus itens, mas não em seus componentes isolados”. Há dois aspectos importantes das propriedades emergentes: primeiro, elas se perdem quando o sistema é dividido em partes; segundo, quando um componente é removido do todo, aquele componente isolado perde as propriedades emergentes do todo. Esta noção nos conduz ao conceito de sinergia, sugerindo que, o todo é mais do que a soma das partes <sup>1</sup>.

Desta forma, a emergência do sentido (ou padrões) e a emergência do significado dependem da completude e coesão do sistema fenomenológico como um todo, incluindo os componentes da intencionalidade, que são o ato intencional (consciência) e o objeto intencionado, onde os dados provenientes do objeto intencionado inserido em um contexto e o conhecimento prévio da consciência atuam como insumos para a gênese da *informação intencional*, que idealizamos em Pereira-Jr, Lima-Marques e Tenorio (2012).

Quanto aos blocos básicos da Ciência da Informação – dado, informação e conhecimento – Bates (2005) explica que na literatura científica, os conceitos de *conhecimento* e *dado* são definidos e discutidos em termos do conceito de *informação*.

A abordagem DIKW entende que dado, informação e conhecimento formam uma hierarquia. Isso tem promovido o entendimento inadequado expresso pelas equações definidas por Rowley (2007):

$$\left\{ \begin{array}{l} \text{informação} = \text{dado} + \text{significado} \\ \text{conhecimento} = \text{informação} + \text{contexto} \end{array} \right\}$$

Enquanto isso, outros filósofos e cientistas da informação têm questionado se esta hierarquização realmente faz sentido, pois entendem que “dado, informação e conhecimento podem ser vistos como tipos específicos um do outro, ou como insumo de um para produzir os outros, ou ainda, como uma saída do processamento dos outros” (MACHLUP, 1984).

---

<sup>1</sup>Zaliwski (2011) relata que Andrzej Chmielecki “desenvolveu uma ontologia com a convicção de que quando a totalidade da realidade é dividida em partes, a informação sofre perdas consideráveis”

Como explanado na seção 2.4, informação, dado e conhecimento na *Arquitetura Genética da Informação* estão em conformidade com a TGAI (2.4.1), que os explica de forma totalmente independente um do outro. Siqueira (2012) argumenta que isso tem gerado um ciclo vicioso referencial, resolvido em sua proposta de uma “Arquitetura da Informação Intencional”, ao considerar os conceitos isoladamente um do outro, mas relacionados aos seus respectivos espaços arquiteturais. O conhecimento, no espaço do sujeito; a informação, no espaço do objeto; e o dado no espaço do fenômeno da relação sujeito-objeto.

## 5.1 O Fenômeno da Informação Intencional

Robredo (2003) argumenta que a “informação tem natureza fenomenológica e econômica”. Saab e Riss (2011), por sua vez, explicam em detalhes esse caráter, realizando o que chamaram de “ontologização” do caráter fenomenológico da informação, conforme ilustrado na fig. 15, em que a informação envolve um entrelaçamento de dois fenômenos: a emergência do sentido e a emergência do significado. A emergência do sentido é um processo individual contínuo de discernimento dos dados em um contexto. Os dados proporcionam padrões que poderão ser indícios de possíveis ações, i.e., um viés de orientação para determinado assunto. Os esquemas, que são necessariamente ligados a experiências prévias de conhecimento, geram estes indícios.

Bates (2005), em seu discurso sobre uma estrutura evolucionária da informação (consultar o item 6.4.2), explica o conceito de “padrão”, este relacionado à primeira e segunda ordens. Um padrão de primeira ordem é “uma marcação natural ou ao acaso, configuração ou *design*”. Um padrão de segunda ordem é “uma combinação de qualidades, atos, tendências etc, formando um arranjo consistente e característico”, dando, portanto, sentido ao objeto intencionado em um fenômeno intencional.

Esse entrelaçamento da emergência do sentido com a emergência do significado nos remete a Kant (2001), respectivamente aos conceitos de “sensibilidade” e de “entendimento”.

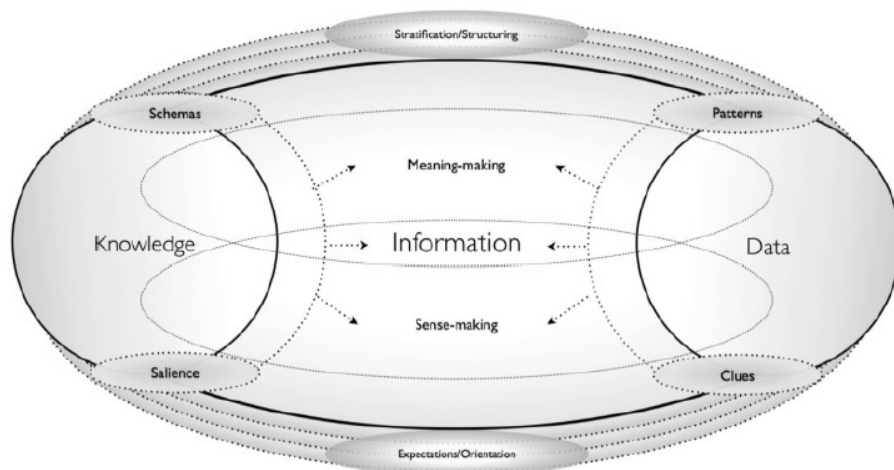


Figura 15: Ontologização: O Caráter Fenomenológico da Informação - Fonte: (SAAB; RISS, 2011)

Ele afirma que “pensamentos sem conteúdo são vazios; sensibilidade sem conceitos são cegos... somente a partir de sua união pode surgir a cognição”. Ainda, argumenta que nossa experiência de mundo envolve, então, o elemento da sensibilidade (capacidade de experimentar diretamente coisas particulares no espaço e no tempo); e entendimento (nossa capacidade de ter e usar conceitos) (LANDAU; SZUDECK; TOMLEY, 2011). Assim, a teoria do idealismo transcendental kantiana afirma que tanto a razão como a experiência são necessárias para a compreensão do mundo.

Considerando a evidência das propriedades emergentes (sentido e significado) nesta ontologia do caráter fenomenológico da informação, também havemos de considerar como os ontologistas veem as características emergentes de um objeto. Nesse sentido, O'Connor e Wong (2012), corroborando com a Fenomenologia Genética, em especial com a gênese e síntese passiva (consultar seção 2.4.2), afirmam que:

“Os ontologistas emergentistas veem o mundo físico constituído inteiramente por estruturas físicas, simples ou compostas. Mas, as composições não são apenas agregações das simples. Há camadas de estratos, ou níveis de objetos, baseados em crescente complexidade. Cada nova camada é uma consequência do aparecimento de uma faixa interativa de 'novas qualidades'. Estas novidades não são meramente temporais, nem a primeira instância de um determinado particular de um familiar determinável. Ao invés disso, são juntamente um tipo de propriedade nova e fundamental... Além disso, novidade de propriedade implica novos poderes primitivos causais, refletidos nas leis que ligam complexas estruturas físicas às características emergentes”.

Aplicando esse conceito, o aparecimento de novidades (propriedades emergentes de sentido e significado) resultam em informação, como explicam Saab e Riss (2011):

“O ponto de partida de nossa investigação é o discernimento de que informação não pode ser descrita simplesmente como dado que possui um significado objetivo. Antes, vemos que informação resulta em grande parte a partir de uma interpretação baseada no esquema em que o sujeito produz o sentido associado a um padrão observado nos dados.”

Nessa linha de raciocínio, Searle (1997) afirma que o sentido que se atribui a tais padrões depende do conhecimento prévio do sujeito (por síntese passiva). descreve isso da seguinte maneira: “a informação que alguém recebe é uma função do que previamente sabe”.

Ao questionarmos David Saab sobre a aderência de seu artigo “Informação como Ontologização” à Fenomenologia, responde-nos:

“A linhagem do pensamento fenomenológico que resultou na figura 15, [na página 76], flui de volta de Polanyi para Heidegger, que é fundamental para a minha compreensão de Fenomenologia. A crítica heideggeriana da distinção sujeito-objeto foi muito influente no desenvolvimento da caracterização do referido diagrama. Apesar de não fazer referência a Heidegger, especificamente no artigo 'Informação como Ontologização', pode-se ver a sua influência na forma como fica evidente o obscurecimento entre os padrões do *conhecimento* subjetivo e os padrões objetivos dos *dados*, como base do processo de entrelaçamento de emergência do sentido e emergência do significado”.

A Fenomenologia Polanyiana da emergência do significado é apresentada em detalhes em Polanyi (1965) e Polanyi e Prosch (1975).

Nos tópicos a seguir, consideraremos os conceitos de informação causal e *informação intencional* (também chamada de significativa ou semântica).

### 5.1.1 Informação Causal

Sterelny e Griffiths (1999) explicam que ao tomarmos o genoma como um programa, há essencialmente dois conceitos de informação, que podemos rotular como informação causal e *informação intencional*.

As noções causais da informação derivam da Teoria da Informação de Shannon (consultar seção 6.1.1), que em suma, estuda apenas a quantidade de informação em um sistema físico, não dizendo nada sobre do que se trata a informação. A quantidade de informação pode ser entendida grosso modo como a quantidade de ordem no sistema, ou o inverso da entropia (desordem) que todos os sistemas físicos fechados acumulam com o passar do tempo.

A informação flui em um canal conectando dois sistemas: o receptor, que contém a informação; e o emissor, a respeito de quem se trata aquela informação. Havendo a possibilidade de inferência do estado do emissor pelo estado do receptor, pode-se concluir que existe um canal de comunicação entre os dois sistemas. Toda instrumentação científica é projetada para garantir a confiabilidade do fluxo de informação nesse senso causal do emissor para o receptor. Daí, podemos dizer que existe um canal conectando um dispositivo de sensoriamento a um estado sensorial do mundo, se o estado apresentado nesse dispositivo é causado pelo estado do mundo. Quanto aos canais, estes também têm suas próprias condições e carregam ruído.

Sterelny e Griffiths (1999) explicam que:

“A ideia de informação como dependência causal pode ser usada para explicar como os genes transportam informação desenvolvimental. . . Os biólogos exploram esse fato quando eles usam uma população clonal para medir a variação em algum fator ambiental. A seleção natural explora esse fato quando diferentes castas são produzidas em diferentes condições”.

Eles prosseguem afirmando que uma outra forma de ver a paridade entre os genes e outras causas desenvolvimentais é retornando-se às ideias de ruído e sinal. Como, tipicamente, muitos fatores mudam ao mesmo tempo, diferenciar o que é ruído e o que é sinal não é tarefa trivial, ou seja, por vezes não o sinal, mas o ruído é que será considerado informação. Em termos causais, informação é simplesmente covariação, e todos os fatores que causam covariação são fontes de informação desenvolvimental.

### 5.1.2 Informação Intencional, Significativa ou Semântica

Dentre os pressupostos básicos desta tese, temos o conceito de *informação intencional* (tópico 2.4.4), apresentado por Reading (2011). Em complementação, ele afirma que:

“Boa parte do que é atualmente chamado de ‘informação’ não se encaixa nessa definição, uma vez que não muda ou afeta os destinatários. A capacidade para detectar e responder à informação significativa é essencialmente um fenômeno biológico, já que

não há detectores de informação inanimados na natureza. Informação e energia são as duas propriedades fundamentais da matéria organizada que refletem a complexidade da sua organização, mas enquanto a energia é uma função da massa de uma entidade ou substância, a informação é uma função da sua forma (isto é, da forma como a sua estrutura é organizada e disposta no espaço ou no tempo).

Ele também argumenta que os biólogos cunharam o termo 'teleonomia' referindo-se à "informação significativa", que é direcionada a uma meta e aos processos auto-regulatórios dos organismos, sem relações de causalidade.

Ele ressalta que "da mesma forma que a informação acumulada em um milênio no genoma molda as metas específicas de uma espécie, a informação acumulada no cérebro dos indivíduos durante suas vidas molda suas personalidades".

[Stegmann \(2005\)](#) argumenta que "os ácidos nucléicos contêm informação em um sentido semântico, mesmo que com efeitos limitados. . . O conteúdo da informação genética se identifica com as propriedades de um modelo, o qual determina o produto pela constituição de instruções", i.e., informação genética é conteúdo instrucional. Ele explica que há três características semânticas da informação genética buscadas na Filosofia da Biologia:

1. Entidades da informação devem ser intencionais, em um sentido amplo: dizendo respeito a outras entidades;
2. A ideia de erro ou má-representação deve fazer sentido, sem ser considerada como efeito incomum ou mal-funcionamento;
3. Deve-se levar em conta o significado da informação genética ser armazenada no DNA, garantindo a sua persistência no decorrer do tempo ou, ao contrário, a possibilidade de a informação não ser usada ou expressa.

Ele ainda explica que os três principais tipos de transferência de informação ocorrem nos processos moleculares de replicação ou duplicação (DNA  $\rightarrow$  DNA); transcrição (DNA  $\rightarrow$  RNA); e tradução (RNA  $\rightarrow$  Proteína).

[Sterelny e Griffiths \(1999\)](#) exemplificam, dizendo que quando um organismo desenvolve diferentes fenótipos sob diferentes condições ambientais, a mensagem interna dos genes é 'Faça isto nas circunstâncias A; Faça aquilo nas circunstâncias B' (um programa genético disjuntivo). "Nesta abordagem de informação semântica, faz sentido falar de erro ou de má-representação. . . o que não é verdade no contexto da noção causal de informação, pois o erro ou a má-representação são entendidos como mudanças nas condições do canal".

[Niiranen e Ribeiro \(2011\)](#) exploram o conceito de *informação intencional*, argumentando que o processamento e propagação da informação pervadem os sistemas biológicos em vários níveis: dentro do gene (redes regulatórias), dentro das células (rotas químicas), entre as células de um tecido e entre organismos de uma população. Os sistemas biológicos podem ser percebidos em termos do processamento da informação, desde a percepção e cognição, até a representação da informação e o uso da linguagem. Ele explica que os processos básicos de

sistemas biológicos podem ser entendidos em termos de expressão da informação, execução de programas e interpretação de códigos, sempre dentro de um contexto <sup>2</sup>, onde o gene é visto como “portador de informação”.

As aplicações do conceito de informação ocorrem:

- Na descrição dos traços fenotípicos dos organismos, como “informação especificada ou codificada”.
- Na descrição dos processos que ocorrem no interior das células, como “programa” armazenado nos genes.
- No tratamento da transmissão dos genes e de outras estruturas de hereditariedade como “propagação da informação” entre gerações.
- No tratamento dos genes em si mesmos, no contexto da teoria evolucionária, no sentido de “informação constituída”.

## 5.2 O Processo da Gênese da Informação Intencional

No tópico 3.2.2.2, vimos que a redução fenomenológica é estruturada em duas etapas:

1. *Epoché*, ou *bracketing*: suspensão de julgamento sobre os objetos do pensamento (noemata), que são os dados sensíveis desprovidos de intencionalidade (ou dados hiléticos). Os elementos exteriores dos objetos são os “caracteres noemáticos”, que exprimem pela percepção presencial ou pela recordação e imaginação os modos como se apresentam a si mesmos.
2. Emergência do significado dos objetos por uma forma de juízo (noese), ou seja, pela intencionalidade que os “informa”. São os “caracteres de ser”, ou “caracteres dóxicos” (atitude de certeza ou dúvida relativamente à existência do objeto por meio de uma espécie de crença ou opinião, i.e, juízo) indicando como são concebidos na mente. Morujão (1980) explica que “neles e por eles é que o objeto encontra significado” e que a origem do significado se dá na consciência, “como primeira expressão de ideias”.

Na primeira etapa, com foco no *noema*, ocorre a suspensão de julgamento, caracterizada pela “destruição” de qualquer possível relação prévia entre observador e objeto; na segunda etapa, com foco na *noese*, o significado aparece em toda riqueza, pois o objeto se manifesta como total novidade.

---

<sup>2</sup>Na definição de Shannon não há contexto. Já, em sistemas biológicos, informação é dependente de contexto. Os conceitos da Teoria da Informação adquiriram um papel proeminente em vários campos da biologia, particularmente na genética, biologia do desenvolvimento e teoria evolucionária.



Esta preocupação com a emergência do significado, segundo [Oliveira \(2000\)](#), é privilegiada por Husserl (em “Investigações Lógicas”) em sua teoria do significado na construção monumental de sua fenomenologia. A análise que Husserl faz sobre “significado” é de que este não pode ser compreendido apenas em termos de funções simbólicas e linguísticas, mas deve realizar um retorno às “coisas nelas mesmas” na determinação de sua “irrevogável identificação”.

Assim, se estabelece a conexão essencial entre o significado intencional do objeto (objetivo) e o significado de realização no sujeito (subjetivo). A apreensão sempre ocorre pela sensação do objeto que se apresenta na realidade ou pela imaginação do objeto em uma imagem mental.

O fenômeno concreto da expressão compreende, então, o aspecto físico, que informa o objeto aos sentidos; e os atos mentais que lhe conferem significado. Daí, considerando-se o princípio da ausência de pressupostos e suspensão de julgamento, é plausível o entendimento de que o surgimento do significado ocorre em um processo de redução fenomenológica, após a etapa da *epoché*, quando os dados noemáticos (ou hiléticos) adquirem significado, gerando a *informação intencional*.

[Merleau-Ponty \(1962\)](#), argumentando sobre a questão do espaço observável, explica que “uma perturbação não afeta a “informação” derivada da percepção, mas revela por trás da percepção, um aspecto mais profundo da consciência”. Esta derivação refere-se à *informação intencional*.

[Siegel \(2011\)](#) propõe uma reflexão sobre a fenomenologia da particularidade, enfatizando que nas experiências de observação de objetos, a própria experiência também é representada.

Ela explica que “a noção de conteúdo da percepção pode acomodar diretamente a ideia de que existe tal coisa como sendo o testemunho dos sentidos”. A experiência, certamente fazendo parte do contexto perceptível, conforme explicamos na seção 4.1.

Em 1913, [Husserl \(1983\)](#) explica que a *epoché* é vista como questionadora da base de validade do mundo objetivo, procurando suspender nossa tendência natural de validar o que é apresentado na experiência. Um aspecto crucial da redução, como aplicado por Husserl, é que todas as características da experiência consciente devem ser tomadas como aparecem, sem nossa tentativa de categorizá-las como falsa, ilusória e assim por diante, não avaliando sua validade como tal.

A redução remove a referência ao mundo real de entidades existentes e todas as aparências são tomadas devidamente como genuínas. As experiências factuais não são estabelecidas por prioridades: Memória, fantasia, e outras formas de atenção igualmente revelam-se como possíveis atos de percepção, apesar de que, percepções têm a presença real dos objetos e as fantasias, apenas uma presença fictícia. Na redução, o mundo inteiro torna-se para a consciência um campo de possíveis experiências. A fenomenologia requer um procedimento de atenção cuidadosa em relação às dimensões da experiência.



Figura 16: Elementos de uma Rede de Petri - Fonte: Produzido pelo autor

### 5.3 Modelagem da Gênese da Informação Intencional com Redes de Petri

David e Alla (2010) explicam que os principais usuários de Redes de Petri são cientistas da computação e de automação de processos.

Apesar disso, constitui-se uma ferramenta suficientemente genérica para modelar fenômenos de tipos extremamente variados, oferecendo duas características interessantes: Em primeiro lugar, tornam possível modelar e visualizar comportamentos com paralelismo, concorrência, sincronização e compartilhamento de recursos; e, em segundo lugar, os resultados teóricos que o concernem são abundantes.

Balasubramanian, Yeh e Chang (2005) em artigo intitulado “Redes de Petri Hierárquicas para a Modelagem de Fenótipos Metabólicos em Procariontes” apresentam os conceitos básicos das Redes de Petri, descrevem as funções celulares hierárquicas de processos bioquímicos, discute vários mecanismos de reação metabólica e suas correspondentes Redes de Petri. Também, considera os processos de transporte através das membranas celulares e realiza simulações dos efeitos de modificação genética com Redes de Petri, comparando os resultados obtidos com dados experimentais disponíveis na literatura.

As redes de Petri são compostas pelos elementos identificados no diagrama da figura 16.

Nas redes de petri, os modelos expressam o fluxo da estrutura dos processos. O tipo mais simples de modelos usam somente três construtos: o lugar, a transição e o arco, caracterizando uma rede de petri básica.

A unidade de modelagem mais importante representa um passo no processo, aparecendo no diagrama como um quadrado verde.

A estrutura do fluxo do processo significa que nem todos os passos podem ocorrer a qualquer momento: eles estão sujeitos a certas condições, que precisam ser explicitadas.

As **transições** são representadas por quadrados, que significam ações. De um modo geral, representam as transições entre estados de coisas no mundo.

Os círculos representam as condições e, todos juntos, compõem os estados de coisas possíveis, na medida em que são relevantes para o fluxo do processo a ser descrito. Eles são conhecidos como **lugares** ou locais.

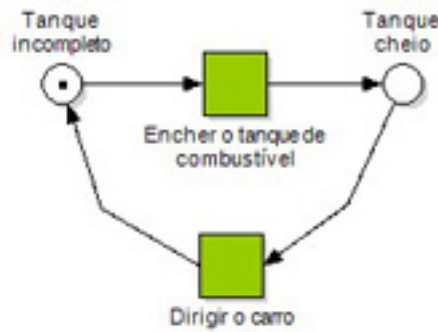


Figura 17: Rede de Petri - Abastecimento de Combustível - Fonte: Produzido pelo autor

Transições e lugares se conectam para formar uma rede, como por exemplo, a que é apresentada no diagrama da figura 18, onde temos quatro lugares e três transições ligadas por **arcos direcionados** por setas.

Um estado específico das coisas do mundo é representado pela marcação das condições que possuem: Os pontos na rede são conhecidos como **fichas**; uma distribuição de fichas sobre lugares caracteriza uma marcação.

A execução do processo em redes básicas é definida por uma única regra: uma transição pode ocorrer sempre que todos os lugares predecessores contenham pelo menos uma ficha; sua ocorrência remove uma ficha de cada um destes lugares, e coloca uma ficha em cada um dos locais sucessores. Por exemplo, no modelo representado pelo diagrama da figura 17, abastecer o carro só pode ocorrer quando o carro estiver com tanque incompleto, ou seja, esse lugar estiver marcado com uma ficha, e quando ele dispara a transição, ou seja a ação de abastecer o carro, passa para a condição ou estado de tanque cheio de combustível. Logo que a ficha chega lá, o carro pode ser dirigido. Quando isso ocorre, a ficha se move para o lugar de tanque incompleto, podendo ser abastecido novamente, a qualquer momento. Esse modelo é um exemplo muito reduzido, mas expressa bem o conceito de que o fluxo do processo é completamente definido pelas regras de execução das transições.

Embora as redes de Petri e suas operações sejam relativamente fáceis de compreender através de ilustrações gráficas, é necessário mencionar que o campo repousa sobre uma matemática rigorosa subjacente. Em termos formais, uma rede de Petri é definida por seus lugares, transições, a função de entrada e a função de saída:

$C = (P, T, I, O)$ , onde:

- $P$  é um conjunto de lugares;
- $T$  é um conjunto de transições;
- $I$  é uma função de entrada que faz o mapeamento de uma transição  $t_j$  para um conjunto de lugares de entrada desta transição:  $I(t_j)$ .
- $O$  é uma função de saída que faz o mapeamento de uma transição  $t_j$  para um conjunto de lugares de saída desta transição:  $O(t_j)$ ;

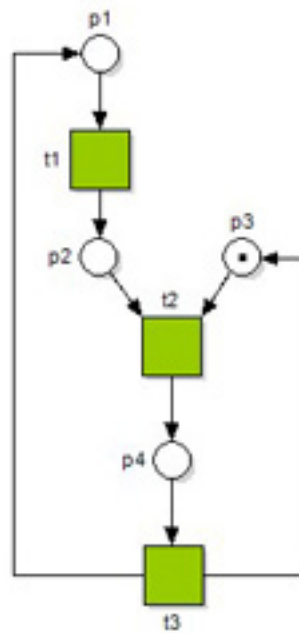


Figura 18: Exemplo de Lugares e Transições em uma Rede de Petri - Fonte: Produzido pelo autor

A rede de Petri na figura 18, por exemplo, é construída por essas regras. A rede tem quatro lugares:  $p1$ ,  $p2$ ,  $p3$  e  $p4$ ; e três transições:  $t1$ ,  $t2$  e  $t3$ . Arcos direcionados conectam lugares e transições, e  $p3$  é um lugar ativo, pois tem uma ficha nele. Assim, na notação matemática, temos:

$P = \{p1; p2; p3; p4\}$  - Os quatro lugares;  
 $T = \{t1; t2; t3\}$  - As três transições;  
 $I(t1) = \{p1\}$  - As entradas de  $t1$ ;  
 $I(t2) = \{p2; p3\}$  - As entradas de  $t2$ ;  
 $I(t3) = \{p4\}$  - As entradas de  $t3$ ;  
 $O(t1) = \{p2\}$  - As saídas de  $t1$ ;  
 $O(t2) = \{p4\}$  - As saídas de  $t2$ ;  
 $O(t3) = \{p1; p3\}$  - As saídas de  $t3$ .

Apresentamos na figura 19, página 85, o modelo da gênese da *informação intencional* com o formalismo das Redes de Petri.

Considerando o pressuposto básico (Seção 2.4) do conceito de “Informação Significativa”, temos os seguintes requisitos básicos para esta modelagem:

1. O padrão de matéria organizada ou energia de um objeto é intrínseca ao seu contexto espaço-temporal;
2. O fenômeno de observação acontece pela experiência de sensoriamento desse padrão, seguida pela experiência de percepção;
3. O padrão de matéria organizada ou energia detectada adquire significado na mente do observador, tornando-se *informação intencional*.

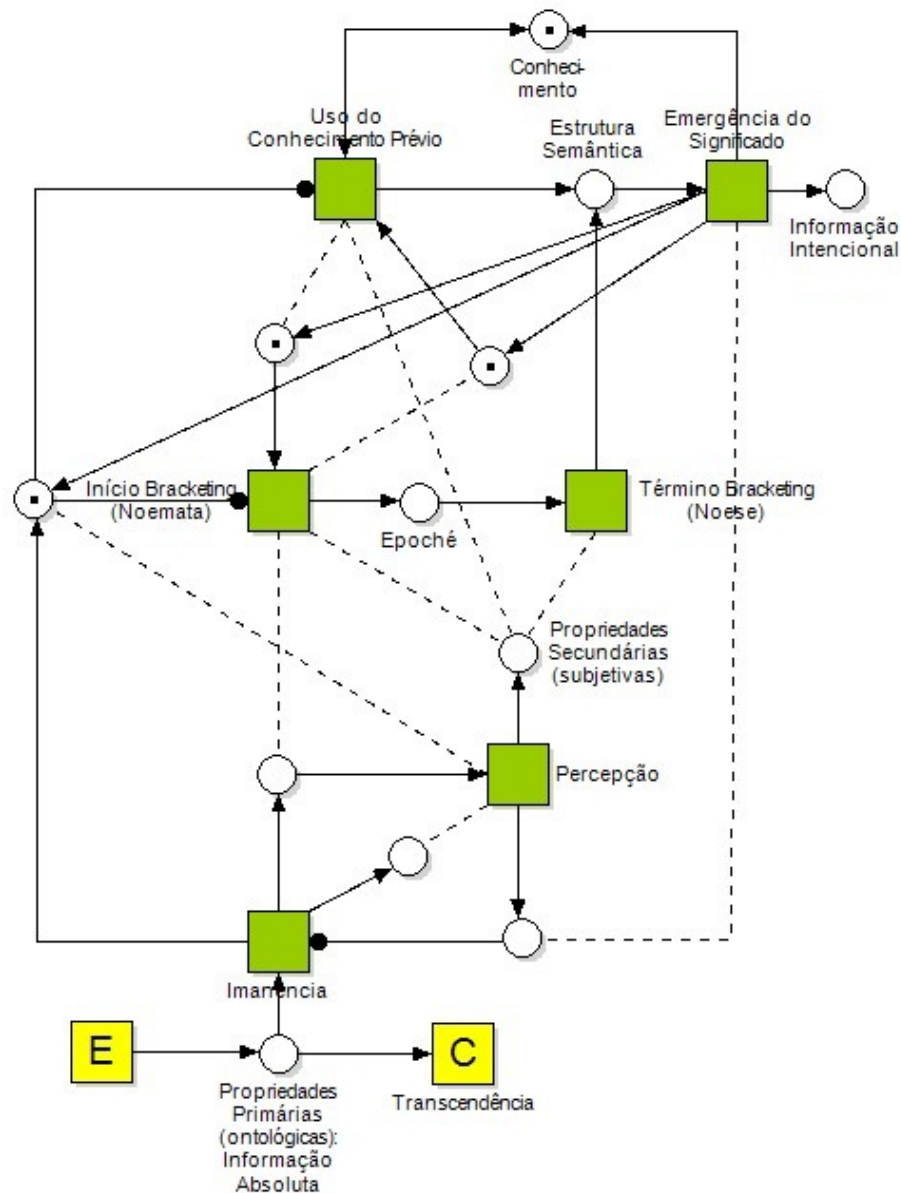


Figura 19: Modelagem da Gênese da Informação Intencional em uma Rede de Petri - Fonte: Produzido pelo autor

As possibilidades de imanência ou transcendência são mutuamente exclusivas, conforme apresentado no diagrama da figura 87, página 217.

Os diagramas do processo completo são apresentados no Apêndice C, conforme indicado a seguir.

No caso de transcendência, o estado da rede fica como apresentado no diagrama da figura 88, página 218.

Na ocorrência de imanência, o estado da rede de petri passa para a condição apresentada no diagrama da figura 89, página 219.

Ao se habilitar a percepção, chega-se ao estado apresentado no diagrama da figura 90, página 220.

Optando-se pela Epoché Fenomenológica, temos o início do *bracketing*, ou seja, a suspen-

são de juízo (figura 91, página 221).

Daí, a única transição habilitada no modelo é a finalização da Epoché Fenomenológica, pelo término do *bracketing* (figura 92, página 222).

Finalmente, temos a gênese da *informação intencional* (dado) (figura 93, página 223).

Logo após esse fenômeno de gênese, a *informação* pode ser caracterizada por seu genoma e características fenotípicas, segundo pilar da tese, a saber “a genética da informação”.

# 6 Genética da Informação Intencional

*“A forma como uma palavra é usada é o seu fenótipo, mas os seus significados profundamente imutáveis, muitas vezes ocultos, constituem o seu genótipo.”*

— Lewis Thomas

Neste capítulo discorreremos sobre algumas teorias da informação em relação à Genética; sobre a Arquitetura Genética, definida pelo mapeamento do genótipo para o fenótipo; e sobre processos biológicos da informação, como o armazenamento, processamento, leitura e expressão, busca e recuperação por similaridade.

O processo da gênese à genética da informação tem início na segunda etapa da redução fenomenológica (explicada na seção 5.2 e, também, no tópico 3.2.2.2).

## 6.1 Teorias da Informação e Genética

Na seção 3.7, introduzimos o assunto de como a informação é vista na Biologia. E, em busca de encontrar o papel da informação na Biologia, [Niiranen e Ribeiro \(2011\)](#) procuram aplicar em sistemas biológicos as duas principais teorias da informação, a de Shannon e a de Kolmogorov. Os conceitos dessas teorias adquiriram um papel proeminente em vários campos da biologia, particularmente na genética, biologia do desenvolvimento e na teoria evolucionária. Um outra abordagem para o tratamento da informação na Biologia é a Teoria da Informação Algorítmica de [Chaitin \(2012\)](#).

### 6.1.1 Teoria da Informação de Shannon

[Niiranen e Ribeiro \(2011\)](#) argumentam que na Teoria da Informação de Shannon há três aspectos importantes:

1. A semântica é irrelevante (não temos a menor ideia do que 0 ou 1 possam significar na mensagem);
2. O decodificador precisa ter poder computacional indefinido;
3. O decodificador deve ter a capacidade de decodificar qualquer código arbitrário, o que é irrealístico.

Na definição de Shannon, não há contexto (em sistemas biológicos, informação é dependente de contexto). Shannon oferece uma medida de quanta informação há na fonte (entropia da fonte), mas nunca afirma o que é a informação propriamente dita. Ainda, a expressão matemática de Shannon é essencialmente a mesma da expressão de Boltzmann para a entropia de sistemas físicos.

A interpretação da teoria da informação dentro da biologia ocorre em conexões da informação entre eventos ou variáveis que envolvem nada mais do que correlações ordinárias, possivelmente relacionadas a causas físicas. Um sinal carrega informação sobre uma fonte se pudermos prever o estado da fonte a partir do sinal. Isso equivale à contingência e correlação na tecnologia de comunicação digital. Ainda, ocorre na descrição da ação do gene na tentativa de adotar uma estrutura quantitativa para descrever correlações ordinárias ou conexões causais. Nesse caso, os processos biológicos não estão sendo explicados em termos do uso ou manipulação da informação. É necessário um conceito de informação mais rico, não apenas causal, mas “semântico” ou “intencional” (consultar 2.4.4).

Smith (2010) elaborou um diagrama que apresentamos (com pequenas adaptações na tradução) na figura 20, p. 88, onde se propõe uma analogia entre a TMC e a transmissão biológica de informação genética.



Figura 20: Diagrama de Comparação da TMC com o DNA - Fonte: (SMITH, 2010)

No exemplo da TMC, a mensagem é primeiramente codificada e, depois, decodificada.

O primeiro codificador é a pessoa que converte uma mensagem semântica em uma sequência de fonemas, e depois, converte para o código Morse. No caso genético, o codificador é o RNA, que produz a sequência de bases de inúmeras possibilidades<sup>1</sup> de sequências que, via um canal de informação, especifica uma proteína que tem um “significado” intencional e funcional, de forma a favorecer a sobrevivência do organismo.

O processo de produção de proteínas codificadas pelo DNA é apresentado na figura 21.

Rittner e McCabe (2004) explicam que o RNA pode ser informacional ou não-informacional. O primeiro é composto por moléculas codificadoras (4% do conteúdo do RNA) e o segundo, de moléculas não-codificadoras (96% do conteúdo do RNA). O mensageiro tem um papel imprescindível à vida da célula, visto que é o responsável por carregar a informação contida na molécula de DNA, que está em seu filamento.

Na maior parte dos casos, a molécula gerada no processo de transcrição ainda não é a molécula com atividade funcional, consistindo de um precursor que requer modificações para adqui-

<sup>1</sup>As inúmeras possibilidades configuram os cenários de seleção natural



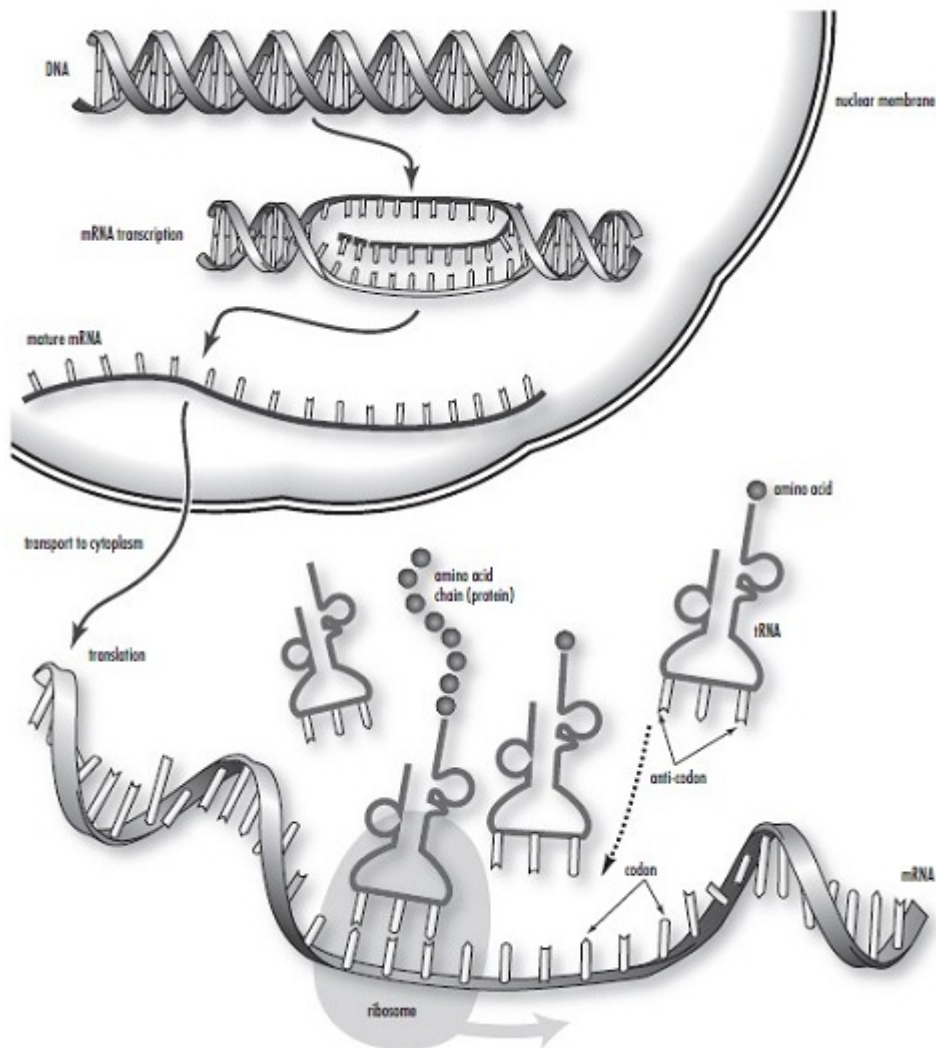


Figura 21: O processo pelo qual as proteínas são feitas pelas instruções codificadas no DNA - Fonte: (RITTNER; MCCABE, 2004)

rir seu estado biologicamente ativo. Ao conjunto de modificações sofrida pelos RNAs chamamos processamento. Os tipos mais comuns de RNA são os RNAs mensageiros (mRNA), que são da informação; e o RNA ribossomal (rRNA) e transportador (tRNA), não-da informação. Estas moléculas ocorrem em todos os organismos celulares, sendo indispensáveis à vida. O mRNA carrega a informação genética para os ribossomos (expressão gênica)

Enquanto o engenheiro da TMC vê um esquema projetado para a realização da comunicação, o biólogo vê um esquema da seleção natural.

Ambos com processamento de informação. Assim, nesta analogia, uma mensagem pode ser carregada por meios físicos inorgânicos (onda de som, onda eletromagnética, ou corrente alternada em um fio), como por meios químicos orgânicos (através de um conjunto de moléculas). A grande contribuição da TMC é permitir que a informação possa ser transmitida por diferentes tipos de canais.

Marijuán (2004) fala sobre a 'Metáfora da Informação', percebendo indiscutível hegemonia de visões reducionistas dentro de um mosaico de abordagens da informação nas ciências

naturais:

“A Metáfora da Informação é uma nova conceituação de fenômenos de informação, principalmente com base no panorama da dinâmica celular, muito além de ingênuas metáforas de informação da biologia molecular clássica.”

Na teoria da informação, Shannon mede a quantidade de informação pela capacidade de transmissão do canal, dada pelo número de diferentes mensagens que pode ser enviada. Os aspectos probabilísticos da teoria de Shannon são utilizados em neurobiologia, mas raramente em genética.

“Dada uma cadeia de símbolos  $n$ , cada um dos quais pode ser qualquer uma das quatro alternativas igualmente possíveis, a medida de Shannon dá  $2 * n$  bits de informação. Na mensagem genética, por sua vez, existem quatro possibilidades de bases. Se fossem igualmente prováveis, e se cada símbolo fosse independente de seus vizinhos, a quantidade de informação seria de dois bits por base. Com efeito, as bases não são igualmente prováveis, e existem correlações entre os vizinhos, havendo, assim, alguma redução na quantidade de informação, mas não é muito grande e é geralmente ignorada: uma maior redução resulta da redundância do código. Em suma, não precisamos nos preocupar com a medida de Shannon, porque dois bits por base é suficientemente próximo. Existe um isomorfismo formal, não apenas uma analogia qualitativa. Há dificuldades em aplicar a teoria da informação em genética. Tais dificuldades surgem na questão do significado, pois a teoria ignora o significado e se preocupa somente com a quantidade de informação” (MARIJUÁN, 2004).

### 6.1.2 Teoria da Informação de Kolmogorov

Na Teoria da Informação de Kolmogorov, o conteúdo da informação é o comprimento do menor programa em um computador universal que pode produzir a sequência de símbolos.

Fica evidente que se tira vantagem da redundância. Por exemplo, o menor programa para apresentar a sequência '1111111111' é: “Apresente 1 dez vezes”. Da mesma forma, o menor programa para apresentar a sequência '1010101010' é: Apresente 10 cinco vezes. Kolmogorov também não fala o que é informação. A teoria assume que a informação pode estar presente tanto na linguagem (consciência) como nas células de um organismo.

Wilkins (1998) afirma que o teorema de Kolmogorov mostra que entre quaisquer dois conjuntos arbitrariamente grandes, pode haver pelo menos um que realiza o mapeamento em um conjunto intermediário de ligações. Esse é um resultado importante para sistemas de pesquisa conexionistas. Esses mapeamentos de Kolmogorov são aplicáveis ao mapeamento entre genes e memes, significando que, não sendo direta a identificação de relacionamento entre genes e memes, pode haver um mapeamento intermediário que o faça, mesmo que a relação seja extremamente complexa. A memética é considerada no tópico 6.4.7.1.

Riolo, O'Reilly e McConaghy (2010) apresentam várias aplicações da teoria da informação de Kolmogorov à programação genética evolucionária.

Niiranen e Ribeiro (2011) argumentando sobre a complexidade do fluxo da informação em sistemas biológicos explicam que:

“A teoria da informação fornece uma lente comum, através da qual se pode estudar tanto a estrutura como a dinâmica de sistemas complexos dentro de uma estrutura unificada. Na verdade, desde suas estruturas em rede, até as trajetórias dos estados dinâmicos. . . Ao contrário da informação de Shannon, que se preocupa apenas com a distribuição dos objetos no sistema, a complexidade de Kolmogorov é uma estrutura adequada para capturar não somente a informação incorporada em objetos individuais de comprimento finito, mas também a informação compartilhada entre eles.”

### 6.1.3 Teoria da Informação Algorítmica

Na mesma linha de Kolmogorov (tópico 6.1.2), [Chaitin \(2007\)](#) propõe uma Teoria da Informação Algorítmica<sup>2</sup>, que é uma outra abordagem de quantificação da informação. Nesta visão, o conteúdo da informação de um objeto qualquer é definido como o programa binário mais curto que o descreve adequadamente. Em um dos trabalhos pioneiros nesse campo, Gregory Chaitin afirmou que seu método era teoricamente capaz de descrever sistemas biológicos, mas ele reconheceu as suas limitações práticas: “Nós não podemos realizar estas tarefas [isto é, descrições biológicas] por computador, porque elas ainda são muito complexas para nós, e os programas seriam muito longos” ([CHAITIN, 2012](#)).

[Spetner \(1998\)](#) usou esta abordagem para determinar se mutações genéticas poderiam produzir novas informações, concluindo que as mutações levam a uma perda de informação genética:

“Portanto, vemos que a mutação reduz a especificidade da proteína do ribossomo, o que significa uma perda de informação genética. . . Em vez de dizer que a bactéria ganhou resistência ao antibiótico, é mais correto afirmar que se perde a sensibilidade ao mesmo. . . Todas as mutações pontuais que têm sido estudadas no nível molecular mostraram uma redução da informação genética e não o seu aumento.”

[Williams \(2005\)](#) enfatiza que outros pesquisadores afirmam que uma aplicação mais rigorosa do método de Chaitin mostra o resultado oposto, ou seja, que as mutações podem, sim, produzir novas informações. “Isto ilustra a potencial complexidade dos argumentos da informação”.

## 6.2 Arquitetura Genética: Mapeamento Genótipo-Fenótipo

Arquitetura genética refere-se à influência implícita do genótipo em um traço fenotípico.

Um termo sinônimo é o ‘mapa genótipo-fenótipo’, i.e., a maneira como os genótipos são mapeados para os fenótipos. Esse tipo de mapeamento tem sido analisado em termos de vários eixos principais, como : poligenia, epistasia, pleiotropia, quasi-continuidade, modularidade, plasticidade fenotípica, robustez, evolubilidade e transgressão fenotípica:

---

<sup>2</sup>A Teoria da Informação Algorítmica foi fundada por Ray Solomonoff. Mais tarde, foi desenvolvida de forma independente por Andrey Kolmogorov, em 1965 e Gregory Chaitin, por volta de 1966.

- Epistasia: quando os alelos interagem em um locus e alteram os efeitos fenotípicos em outro lugar.
- Poligenia: quando múltiplos genes contribuem para um determinado carácter fenotípico. A arquitetura genética pode ser ainda caracterizada pelo espectro de contribuições dos genes no efeito final.
- Pleiotropia: quando vários caracteres fenotípicos são afetados por uma única variação genética.
- Quasi-continuidade: quando pequenas mudanças genéticas mapeiam pequenas alterações fenotípicas.
- Modularidade: quando duas diferentes características fenotípicas são mapeadas para a maioria dos conjuntos não sobrepostos de genes.
- Plasticidade: quando um único genótipo dá origem a um espectro de fenótipos. A plasticidade pode se manifestar como diferentes fenótipos entre os indivíduos de mesmo genótipo, ou diferentes fenótipos dentro da vida de um único indivíduo, ou, ainda, diferentes fenótipos em resposta a condições ambientais específicas.
- Robustez mutacional: quando o mesmo fenótipo ocorre num organismo, apesar de uma variedade de perturbações ambientais ou apesar de mutações ocorridas nos genes envolvidos.
- Evolubilidade: quando há uma chance significativa de que a variação genética possa ser produzida, juntamente com o aumento da possibilidade de adaptação do organismo.
- Transgressão: quando um fenótipo é mais extremo do que o fenótipo exibido por qualquer uma das linhas parentais.

Segundo [Kell \(2002\)](#):

“Os efeitos dos genes no fenótipo são mediados por processos que são tipicamente desconhecidos, mas cuja determinação é desejável. A conversão do gene para o fenótipo não é uma função simples de genes individuais, mas envolve complexas interações de muitos genes, problema conhecido como mapeamento não-linear. Um método computacional chamado programação genética permite a representação de mapeamentos de candidatos não-lineares em várias árvores possíveis. Para encontrar o melhor modelo, as árvores ‘evoluem’ por processos semelhantes de mutação e recombinação, e árvores que melhor representam os dados reais são preferencialmente selecionadas. O resultado é uma árvore melhorada de regras que representam diretamente o mapeamento não-linear. Desta forma, a codificação das atividades celulares de ordem superior pelos genes é vista como diretamente análoga a programas de computador.”

## 6.3 Armazenamento da Informação

A [National Coalition for Health Professional Education in Genetics \(2012\)](#) explica que:

“Cada pessoa tem enorme quantidade de informação genética que, se pudesse ser esticada, teria uma extensão equivalente a mais de setenta vezes a distância da Terra ao Sol. Tal informação precisa ser organizada e empacotada hermeticamente e permitir acesso somente aos genes apropriados.”

Há, portanto, diferentes níveis de armazenamento e de organização da informação genética:

- DNA: A informação genética é armazenada dentro de estruturas químicas da molécula de DNA. A estrutura do DNA pode ser comparada com uma escada em espiral, onde os degraus são os pares de bases e o corrimão é formado por açúcar e fosfato. Uma propriedade importante do DNA é que ele pode se replicar, ou seja, fazer uma cópia de si mesmo.
  - Dois *backbones* que formam um espiral em torno um do outro em uma formação bem conhecida de dupla hélice;
  - Sequências de quatro produtos químicos, chamadas de “bases”<sup>3</sup> – A (Adenina), T (Timina), C (Citosina) e G (Guanina) –, correndo ao longo dos backbones, cuja ordem, literalmente, “dita” o “código genético”. A variada sequência de bases de indivíduo para indivíduo é um fonte de variação biológica;
  - Pontes entre as bases dos backbones opostos, chamadas pares de bases (A com T e C com G). A ordem ou sequência destas bases determina a informação disponível para a construção e manutenção de um organismo, similar à maneira em que as letras do alfabeto aparecem em certa ordem, formando palavras e sentenças.
- Genes: Sequências de bases são agrupadas em segmentos específicos, chamados de “genes”. Cada gene fornece um modelo para uma ou mais proteínas específicas. Cada pessoa geralmente tem duas cópias de cada gene, um herdado de cada progenitor. As cópias, no entanto, podem ser ligeiramente diferentes.
  - Alelos: Diferentes versões de um mesmo gene são chamadas alelos. Um gene pode ter muitos alelos possíveis, mas cada pessoa terá apenas dois alelos de cada gene (um de cada progenitor). Todos os seres humanos têm aproximadamente os mesmos genes, mas não há duas pessoas com as mesmas versões (alelos) de todos os genes (exceto no caso de gêmeos idênticos). Ou seja, não existem duas pessoas que tenham exatamente a mesma sequência de pares de bases.
  - Homozigotos: Quando em um indivíduo dois alelos para um gene são o mesmo, diz-se que ele é homozigótico para aquele gene.

---

<sup>3</sup>O DNA humano consiste de 3 bilhões de bases e quase a sua totalidade absoluta é a mesma em todas as pessoas.

- Heterozigotos: Quando um indivíduo tem dois alelos diferentes para um gene, diz-se que ele é heterozigótico para aquele gene.
- Cromossomos: Resultam da compactação da molécula de DNA, de forma a caberem dentro do núcleo da célula. Isso é feito pelo enrolamento em torno de si mesmo e devido a algumas proteínas especializadas.
- Mitocôndrias: São uma pequena quantidade de material genético adicional, encontrada fora do núcleo da célula, produtoras de energia da célula. O DNA mitocondrial (mtDNA) tem a forma de anel e não é empacotado em cromossomos.

## 6.4 Processamento da Informação

Niiranen e Ribeiro (2011) explicam que o processamento e propagação da informação pervadem os sistemas biológicos em vários níveis: dentro do gene (redes regulatórias); dentro das células (rotas químicas); entre as células de um tecido; e entre os organismos de uma população. Em todos estes casos, os sistemas biológicos podem ser identificados em termos de processamento da informação (ex.: percepção e cognição); representação da informação; e uso da linguagem. Os processos básicos dos sistemas biológicos podem ser entendidos em termos de expressão da informação, execução de programas e interpretação de códigos.

Bajic e Wee (2005) dizem que:

“Todos os seres vivos buscam perpetuar sua espécie e linhagem. Os descendentes herdam características dos progenitores e a base para esta herança reside no material genético. A partir do material genético herdado dos pais, a prole deve decodificar a informação no interior do material genético que então se manifesta como vários traços que observamos entre a grande diversidade de organismos no mundo dos seres vivos. Isto envolve o processamento e fluxo de informações no nível mais fundamental, durante todo o curso de desenvolvimento dos organismos vivos. Organismos não vivem isoladamente, mas interagem uns com os outros constantemente dentro de um ecossistema complexo. As relações entre os organismos... são complexas e multi-dimensionais. Em todos os casos, há uma comunicação constante e fluxo de informação em vários níveis.”

A National Coalition for Health Professional Education in Genetics (2012) explica que:

“Diferentemente da maioria das moléculas biológicas, o DNA pode replicar a si mesmo, com a ajuda de proteínas especializadas. Cada filamento de bases pode servir como um modelo para a duplicação durante a divisão celular. A unidade básica da hereditariedade celular é o cromossomo<sup>4</sup>. O DNA copiado não é transmitido apenas de uma célula para outra, mas também de pai para filho. Muitas seqüências específicas de DNA são conservadas de espécie para espécie. Além de viabilizar a variação, o DNA também é responsável pela continuidade.”

Sawai (2011) apresenta inúmeras analogias de funções biológicas com as Tecnologias da Informação e Comunicação.

---

<sup>4</sup>Os cientistas podem observar a replicação e transmissão de cromossomos entre células utilizando um microscópio

Além disso, com base nos estudos de [Niiranen e Ribeiro \(2011\)](#), [Sterelny e Griffiths \(1999\)](#), [Munz \(1993\)](#) e [Bajic e Wee \(2005\)](#), identificamos uma analogia qualitativa do processamento da informação com sistemas biológicos nos seguintes aspectos:

1. Informação Intencional
2. Seleção Natural
3. Imunização
4. Aprendizado
5. Expressão Genética
6. Redes Booleanas
7. Evolução
8. Programação Genética
9. Programação Evolutiva

#### **6.4.1 Informação Genética Intencional**

O conceito de gene nas teorias modernas explica o funcionamento dos organismos e a herança de características entre gerações. O gene é visto como “portador de informação”, para a descrição dos traços fenotípicos dos organismos (“informação especificada” ou “informação codificada”); descrição dos processos causais dentro das células, em termos de um “programa” armazenado nos genes; tratamento da transmissão dos genes e de outras estruturas de hereditariedade (propagação da informação entre gerações); tratamento dos genes em si mesmos, no contexto da teoria evolucionária, no sentido de “informação constituída”.

As correlações ordinárias da Teoria da Informação de Shannon (tópico 6.1.1) são insuficientes para descrever a semântica implícita na informação genética. Daí, há a necessidade de uma nova abordagem, um conceito de informação mais rico, não apenas causal, mas “semântico” ou “intencional”. Por exemplo, de maneira a quantificar o total de informação contida em uma mensagem, é necessário saber o espaço inteiro de estados possíveis, para computar a probabilidade de ocorrência do conteúdo recebido na mensagem. Na maioria dos cenários realísticos, esse espaço de estados é desconhecido ou mesmo, não-computável. Daí, o valor da probabilidade é desconhecido. Uma possibilidade é o conhecimento prévio de ocorrências similares fornecer um valor da probabilidade da presente mensagem, mas sem qualquer garantia ou precisão.

## 6.4.2 Informação e Seleção Natural

Seleção natural envolve questões interessantes relacionadas à informação, como complexidade e estrutura evolucionária, descritos a seguir.

### 6.4.2.1 Complexidade

Com relação à complexidade, a seleção natural tende a produzir sistemas de transmissão da informação biológica, por convenções simplificadas.

Trata-se de um mecanismo que envolve a complexidade de um organismo em todo o seu ciclo de vida, contemplando todo o seu genoma e os aspectos contextuais do mundo em que está inserido, ou seja, o ambiente. As partes da codificação do genoma são consideradas ruído até que o ambiente se modifique de tal maneira que ele se torne útil, quando então se torna um sinal.

Mutações podem ser consideradas como ruídos ocasionalmente informativos. [Brooks \(2009\)](#) argumenta que as mutações são mudanças aleatórias e permanentes no DNA. A existência de todas as adaptações elaboradas de reparo do genoma mostram que as mutações em células individuais realmente são apenas ruídos a serem suprimidos e filtrados como qualquer outro ruído aleatório. A chance de uma mutação ser uma nova informação biológica ou apenas ruído não pode ser descoberta apenas inspecionando o organismo em que ela ocorre. O que é adaptativo depende do ambiente em que o organismo se encontra. A mesma alteração no mesmo genoma pode ser fatal ou providencial, dependendo das circunstâncias externas. A complexidade total do organismo e do ambiente age como um filtro para tornar um ruído mutacional em informação biológica, e nenhuma dessas categorias é particularmente significativa na ausência da complexidade total.

Quando existe um relacionamento entre seleção natural, complexidade e informação então diz-se que temos uma “paisagem adaptativa”.

[Brooks \(2009\)](#) também afirma que quando consideramos a célula como um sistema de informação, ela pode ser comparada a uma rede neural ampla e obscura.

As moléculas agem como nós da rede. As interações entre as moléculas podem ser vistas como as conexões entre os nós. A propensão de um par de moléculas interagir pode ser identificada pelo peso da conexão. A participação de um gene em um fator de transcrição é análogo à função de resposta de um nó. Enfim, praticamente tudo pode, em princípio, ser modelado, mesmo que imperfeitamente, em um sistema estocástico de equações diferenciais.

Ele ainda explica que uma rede regulatória de genes basicamente processa informação biológica, em que moléculas de DNA codificam a informação em uma sequência específica (sintaxe), têm uma existência física repleta de significado (semântica) em um ambiente (contexto). Cada sequência pode ser associada a estados físicos diferentes. Moléculas tRNA funcionam como convenções semânticas, facilitando o uso disseminado de códons particulares como autoridades para aminoácidos particulares, que são os blocos de construção básicos dos tecidos, órgãos e



organismos. A informação decodificada é o que caracteriza o genoma de um organismo.

#### 6.4.2.2 Estrutura Evolucionária

Bates (2005) argumenta que a informação pode ser definida como o padrão de organização de matéria e energia que recebe significado de um ser vivo e o conhecimento como sendo esta informação integrada com outros conteúdos de entendimento.

Nas palavras dela:

“Esta abordagem tem suas raízes em uma estrutura evolucionária; i.e., modos de percepção, processamento, transmissão e armazenamento da informação são vistos como tendo sido desenvolvido como uma parte da evolução geral dos integrantes do reino animal. Suportar um cérebro tem um alto custo para os animais. Por conseguinte, o armazenamento eficiente... é poderoso e eficaz para os animais.”

#### 6.4.3 Informação e Imunização Biológica

O sistema imunológico do corpo humano pode ser visto como um modelo computacional de processamento da informação de um sistema biológico altamente distribuído, orquestrado sobre múltiplas escalas de tempo e espaço. A informação é processada em diferentes escalas de tempo – de microssegundos para a interação de uma molécula de proteína a dias para o desdobramento de uma resposta de imunização. O corpo humano e seu ambiente é um sistema bem-orquestrado de “enxames” de células, moléculas mensageiras, bactérias e vírus<sup>5</sup>.

Hershberg e Efroni (2001) explicam que a questão de como um sistema pode aprender características fundamentais que são posteriormente recombinadas em recursos de ordem elevada não é confinada à genética. Esse problema é comum a qualquer sistema cognitivo. Em quaisquer dos tais sistemas, podemos assumir a perspectiva de que existem dois processos de alto-nível simultâneos e não-discretos: Um que identifica características como elementos; e outro que forma conceitos ou aprende por meio da integração e diversificação dos recursos. O sistema imunológico pode ser visto como tal sistema cognitivo.

#### 6.4.4 Informação e Aprendizado Biológico

Munz (1993) argumenta que a transferência de informação entre células pode ser chamada de um processo de aprendizado.

“Tal aprendizado não resulta da observação do mundo nem de um armazenamento resultante de informações que bateram repetidamente a membrana da célula. Não há nada na célula que a habilite a aprender pela instrução ou observação. O mundo fora da membrana não consiste de bits discretos de informação que possam ser absorvidos por cuidadosa observação. O mundo exterior “não é rotulado”. Qualquer rótulo que encontramos hoje é resultante da resposta da célula ao mundo e da maneira como o conhecimento é incorporado nela, fazendo-a reagir” (MUNZ, 1993).

---

<sup>5</sup>Simulações virtuais de sistemas biológicos de alta complexidade estão mudando os rumos e possibilidades da pesquisa biológica.

Ele ainda afirma que Dretske estava enganado ao pensar que “a matéria-prima é a informação”. Uma vez que o mundo não consiste de pedaços discretos de informação, o estabelecimento de um relacionamento cognitivo entre o conhecedor e o conhecido não pode ser ligado pela transferência técnica de informação do conhecido para o conhecedor.

E mais, ele argumenta que o conhecedor não armazena simplesmente pedaços de informação. Em um relacionamento cognitivo e não-causal ele detém a habilidade de categorizar e abstrair, daí criando uma imagem de seu mundo através dessas abstrações e a responder a ele em termos dos resultados de tais abstrações.

Daí, conclui, dizendo:

“Esse relacionamento cognitivo é muito diferente do processamento de informação. Informação pode ser entendida simplesmente como pedaços de conhecimento. Transferência de informação deve ser tomada como aquisição de conhecimento. A consciência é intencional. É sobre coisas e, necessariamente, tem um conteúdo. Consiste em geral da existência interdependente entre sujeito e conteúdo. A consciência é sinônima de mapeamentos e estratégias de processamento da informação. O curso inteiro da evolução depende do aparecimento incessante de erros, i.e., de chances de mutações e da retenção seletiva daquelas que se mostram mais adaptativas dos que as outras” (MUNZ, 1993).

### 6.4.5 Aleatoriedade da Informação na Expressão Genética

Munz (1993) explica que os mecanismos estocásticos desempenham um papel importante na dinâmica de sistemas biológicos<sup>6</sup>. A aleatoriedade na expressão de produtos dos genes pode resultar em células geneticamente idênticas, apresentando variações significantes nos atributos físicos e bioquímicos.

Estas variações podem influenciar a aptidão no nível individual e populacional. As células também recebem sinais de ruído de seus ambientes, realizando detecção e transdução com bioquímica estocástica.

As células sentem e processam informação usando redes bioquímicas de interação entre genes e proteínas, diferenciando o que é ruído. Vários mecanismos, incluindo rotinas em cascata e ciclos de realimentação, permitem à célula manipular esses sinais de ruído identificados e manter a fidelidade do sinal.

Redes genéticas são definidas como conjuntos de moléculas e interações que controlam a expressão genética, produzem e regulam a dinâmica da célula.

Em um nível fundamental, um gene é informação codificada em uma sequência de nucleotídeos. Essa informação é processada pelo maquinário da célula para executar as instruções que ela contém. Assim, o comportamento da célula pode ser entendido a partir do processo como a informação é produzida, processada e propagada.

---

<sup>6</sup>Qualquer tipo de evolução temporal (determinística ou essencialmente probabilística) que seja analisável em termos de probabilidade pode ser chamada de processo estocástico.

### 6.4.6 Redes Biológicas Booleanas

Redes biológicas booleanas são os principais tipos de sistemas dinâmicos finitos que têm sido usados com sucesso na modelagem de redes biológicas. Os modelos dinâmicos discretos de redes biológicas foram introduzidos primeiramente por [Kaufman \(1993\)](#), que usou redes booleanas para estudar a dinâmica de redes regulatórias.

Um gene é assumido como estando em um dos dois estados: “expresso” ou “não expresso” e é modelado por um valor binário 1 ou 0, respectivamente. O próximo estado de um gene é determinada por uma função booleana em termos dos estados atuais do gene e os seus vizinhos imediatos na rede. O estado de uma rede de  $n$  variáveis é então um vetor binário de comprimento  $n$ , que representa o estado de cada nó da rede. Assim, existem  $2^n$  estados possíveis. A dinâmica da rede é representada por um grafo direcionado para os estados  $2^n$ , onde cada estado é mapeado para um outro estado.

### 6.4.7 Informação e Evolução

[Sterelny e Griffiths \(1999\)](#) afirmam que o conceito de gene evolucionário é de “unidade de informação pura”, onde o DNA é o meio e o gene é a informação transcrita e codificada pela molécula de DNA, i.e., um pacote de informação.

As entidades que exercem um papel importante na evolução são os replicadores e os interagentes: Os replicadores são genes copiados para a próxima geração (cadeias de replicadores copiados compõem as linhagens). Carregam a receita ou o programa de construção dos organismos. Seus padrões sobrevivem intactos, dando origem a uma sequência potencialmente ilimitada de cópias descendentes. Eles interagem com o meio-ambiente com sucesso diferenciado, construindo entidades com propósito específico, por meio de veículos que transportam informação.

Os interagentes são usualmente os organismos, colônias e populações. Reproduzem-se de maneira diferenciada, podendo ser simples ou agressivos.

Os sistemas de herança compõem-se de elementos da célula que causam variações que serão herdadas nas próximas gerações (epigenética) ou a informação manifesta no ciclo de vida de um organismo é reconstruída por si mesma no desenvolvimento da próxima geração (ontogenia).

[Gibbs \(2003\)](#) afirma que os biólogos supunham que apenas as proteínas regulassem os genes dos seres humanos e de outros organismos complexos. Mas um sistema controlador negligenciado, baseado no RNA, pode conter as chaves do desenvolvimento e da evolução, emergente de um programa adicional oculto no DNA-lixo. As proteínas realmente ajudam a regular a expressão dos genes, mas esse sistema paralelo e oculto compõe uma rede sinalizadora que permite aos seres humanos, por exemplo, atingir uma complexidade estrutural muito além de qualquer outra vista no mundo unicelular. A quantidade de DNA de um organismo não corresponde exatamente à sua complexidade. Menos de 1,5% do genoma humano codifica proteínas. Muitos genes em organismos complexos não codificam proteínas, mas dão origem a RNAs

com funções reguladoras. Ou o genoma destes organismos está repleto de transcrição inútil, ou esses RNAs desempenham alguma função desconhecida ou inesperada, podendo transmitir informações cruciais, em especial para o desenvolvimento e a evolução.

#### 6.4.7.1 Memética

Avery (2003) explica que a evolução genética é muito lenta, enquanto a evolução cultural é extremamente rápida e acelerada.

O genoma humano mudou muito pouco durante milhares de anos, mas em décadas ou séculos a evolução cultural alterou sobremaneira o nosso modo de vida. Como todas as outras formas de vida na terra, os seres humanos passam informação de uma geração para a seguinte, codificada pelas sequências de bases de DNA. No entanto, os seres humanos desenvolveram um outro modo altamente eficaz de transmissão de informação: linguagem e cultura.

Em 1976, Dawkins (1990) cunhou o termo “meme”, como equivalente cultural ao “gene”, aplicando conceitos da teoria da evolução, especialmente da genética populacional, à cultura humana, identificando-se com o paradigma histórico-social da informação, explicado na seção 3.3.3. Sua intenção é mostrar que os seres humanos são produto não apenas de uma evolução biológica, cuja unidade são os genes, mas, também de uma evolução cultural, cuja unidade são os memes. Para tal, ele argumenta que os genes usam os corpos dos seres vivos para promover a sua própria sobrevivência, em uma perspectiva das moléculas competindo por espaço e recursos limitados para produzir e disseminar mais de seu próprio tipo.

Socialmente, a memética tenta explicar vários assuntos controversos, como religião e sistemas políticos, usando modelos matemáticos. Nesse sentido, deve ser distinguida da sociobiologia, onde as entidades envolvidas são os genes, enquanto na memética são os memes.

A tabela 1, pág. 100, apresenta uma analogia elaborada por Wilkins (1998) (traduzida e reduzida).

Sobre a Teoria da Informação de Kolmogorov em relação à memética, consultar o tópico 6.1.2.

#### 6.4.8 Programação Genética - PG

A Programação Genética (PG) é um dos ramos descendentes dos algoritmos genéticos.

Bittencourt (2006) explica que “os cromossomos dos indivíduos da população não são sequências de *bits*, mas sim programas de computador, geralmente armazenados na forma de árvores sintáticas... candidatos à solução do problema proposto”.

A PG utiliza o operador de “recombinação genética”, pela troca de subárvores entre dois indivíduos candidatos à solução.

Daida et al. (1999) apresentam uma metáfora para a PG em pelo menos três aspectos: do genótipo e fenótipo; dos blocos de construção; e dos íntrons.

| Termo       | Genética               | Memética               |
|-------------|------------------------|------------------------|
| Interagente | Fenótipo               | Femótipo               |
| Replicador  | Gene                   | Meme                   |
| Caractere   | Traço Mendeliano       | Traço Femotípico       |
| Linhagem    | Espécie                | Tradição, instituição  |
| Economia    | Sistema ecológico      | Cultura                |
| Reprodução  | Reprodução orgânica    | Constituição de perfil |
| Indivíduo   | Organismo, colônia     | Indivíduo memético     |
| Código      | Alfabeto DNA           | Semântica              |
| Mídia       | Aminoácidos, proteínas | Estrutura neural       |

Tabela 1: Tabela de Analogia Gene-Meme - Fonte: [Wilkins \(1998\)](#)

#### 6.4.8.1 Genótipo e Fenótipo

As metáforas do genótipo e fenótipo têm sido amplamente usadas na literatura de PG. A definição usual para genótipo é aquela relacionada a um traço específico ou a um conjunto de traços, i.e. uma árvore sintática em PG.

Para o fenótipo, a definição habitual é de comportamento observado. Alguns autores definiram o fenótipo como equivalente aos valores de um vetor e outros, ainda, optaram por uma definição mais abstrata, igualando fenótipo com semântica.

#### 6.4.8.2 Blocos de Construção

A célula é a unidade básica de vida. Todos os organismos são compostos de uma ou mais células. Os seres humanos são compostos de muitos milhões de células. Todas as células, independentemente da sua função ou localização no corpo, partilham características comuns e processos. Surpreendentemente, as células são constituídas quase inteiramente por apenas quatro tipos básicos de moléculas, que são conhecidos como blocos de construção: carboidratos, proteínas, lipídios e ácidos nucleicos. Analogamente, na PG, blocos de construção são entendidos como componentes simples a partir dos quais as coisas mais complexas podem ser produzidas (como em um brinquedo para crianças).

#### 6.4.8.3 Íntrons

Os íntrons referem-se a porções codificadas dentro de um indivíduo que não contribuem diretamente para uma aptidão nele. Na programação genética, são seções de código funcionalmente inertes.

### 6.4.9 Programação Evolutiva - PE

[Bittencourt \(2006\)](#) explica que “a Programação Evolutiva (PE) visa prever o comportamento de máquinas de estado finitas”, seguindo o modelo conceitual de evolução natural para a solução de problemas complexos.

Há apenas três operadores evolutivos: seleção; mutação; e recombinação.

A seleção permite escolher um indivíduo ou um par deles para gerar descendência. A mutação realiza mudanças aleatórias no material genético. A recombinação simula a troca de material genético entre os ancestrais, determinando a carga genética dos descendentes.

As ideias básicas são:

1. A criação de uma população de soluções (geralmente, a primeira geração é aleatória);
2. A criação de uma entidade (função de avaliação) capaz de julgar a aptidão de cada um dos indivíduos, atribuindo-lhes uma nota de desempenho; e
3. A criação de uma série de operadores evolutivos (seleção, mutação e recombinação).

Assim, uma população inicial de soluções evolui, ao longo das gerações que são simuladas no processo, em direção a soluções mais adaptadas (melhor avaliadas) por meios dos operadores evolutivos.

## 6.5 Leitura e Expressão da Informação

A expressão máxima de um gene é a produção de uma proteína, e o seu efeito sobre a fisiologia contribui para o traço ou 'fenótipo' que observamos. A informação codificada em um gene passa por um processo de ser copiada, transportada, lida, e, finalmente, utilizada como um modelo para a construção de proteínas.

A [National Coalition for Health Professional Education in Genetics \(2012\)](#) resume este processo em “transcrição” e “tradução”, conforme explicado a seguir:

1. Transcrição - proteínas especializadas anexam-se à fita de DNA e fazem uma cópia da sequência de DNA, na forma de uma molécula chamada de RNA mensageiro (mRNA). A mensagem genética é transferida para fora do núcleo através do mRNA.
2. Tradução - A maquinaria celular lê as instruções genéticas do mRNA e sintetiza as proteínas.

O RNA é semelhante ao DNA, exceto que consiste apenas de uma cadeia simples, não tendo a forma de hélice torcida.

Como o DNA, o RNA contém sequências de bases, que lhe permitem codificar mensagens genéticas. Há muitas formas de RNA, promovendo a diversidade de aspectos da expressão do

gene. O RNA mensageiro é a molécula intermediária que transfere a informação a partir do DNA através da membrana nuclear (no núcleo da célula) para o local de tradução exterior à célula.

A sequência de bases no mRNA reflete a sequência de bases no gene. No entanto, o mRNA não corresponde exatamente ao DNA a partir do qual foi transcrito. Após a transcrição no núcleo, a sequência é processada para remover e embaralhar segmentos. Isto aumenta o número de proteínas que podem ser produzidas a partir de um único gene. Há apenas de 20.000 a 25.000 genes, mas mais de 200.000 diferentes tipos de proteínas no corpo humano. O embaralhamento de segmentos dos genes no mRNA ajuda a explicar esse fenômeno.

As proteínas podem ser estruturais, o que significa que contribuem para as formas de materialização do corpo. Por exemplo, o colagênio é uma proteína que fornece uma estrutura e suporte para células e tecidos. Outras proteínas são enzimas, que facilitam as reações bioquímicas necessárias. Por exemplo, a lactase é uma enzima que decompõe hidratos de carbono em produtos lácteos para serem usados pelo corpo. Diferenças na expressão de todos os tipos de proteínas podem ter impacto no fenótipo final.

Todas as células têm os mesmos genes, mas nem todos os genes são expressos em todas as células em todos os momentos.

Por exemplo, as células musculares contêm o gene para o hormônio da insulina pancreática, mas os músculos não produzem insulina. De modo semelhante, os genes que produzem certos hormônios ou sinais nervosos só podem ser expressos na presença de estímulos específicos, a partir do ambiente. Alguns dos fatores que regulam a temporização, taxa, e localização da expressão do gene, chamados de “interruptores genéticos”, são codificados no DNA e podem contribuir para a variação biológica.

Em geral, a regulação dos genes envolve tipicamente uma proteína ou alguma outra molécula que se associa a um pedaço de material genético, mudando a sua função. Isto pode ocorrer em qualquer nível da rota entre o gene e a proteína, por exemplo:

- **Regulação transcricional:** Proteínas especializadas podem se anexar ao DNA, bloqueando ou melhorando a máquina que copia a mensagem genética.
- **Regulação pós-transcricional:** A mensagem genética pode ser alterada ou desativada na rota entre o gene e a proteína.
- **Regulação pós-traducional:** As proteínas podem ser alteradas ou desativadas mesmo depois de sua síntese.

Esses fatores de regulação de genes, juntamente com os seus disparadores ambientais, significam que a mesma variante genética pode ser expressa de forma diferente em diferentes indivíduos, mesmo dentro da mesma família.

Na verdade, algumas condições genéticas e características são bem conhecidas pela sua expressividade variável e penetrância incompleta.

Alguns genes têm maior impacto no fenótipo do que outros. A clarificação do relacionamento genótipo-fenótipo pode ser complicada pelo fato de vários genes, que sozinhos podem não ter tantas consequências, mas juntos contribuem significativamente para o fenótipo. Ou seja, a interação de múltiplas proteínas alteradas podem ter um efeito fisiológico cumulativo.

Há amplas evidências de que os genes não operam fora de um ambiente, começando pelo contexto interno da célula, e estendendo-se para as influências do mundo externo. O fato é que o ambiente não apenas contribui para a constituição do fenótipo, mas pode também modificar a regulação do gene que participará da contribuição genética.

## 6.6 Busca e Recuperação por Similaridade

A recuperação genética da informação baseia-se em buscas por similaridade.

McGee (2003) afirma que uma das operações mais básicas em bioinformática envolve a busca por similaridades, ou homologias, entre um recente pedaço sequenciado de DNA e segmentos de DNA previamente identificados em vários organismos. Encontrar associações aproximadas entre os segmentos de DNA comparados permite que os pesquisadores façam previsões sobre a proteína que aquela nova sequência de DNA poderá produzir<sup>7</sup>.

É igualmente importante o esforço para encontrar em uma população as diferenças de DNA, e não apenas as similaridades.

Por exemplo, empresas farmacêuticas se interessam na identificação de “Polimorfismos de Nucleotídeo Único” (SNPs), que são lugares em que o DNA de uma pessoa difere, mesmo que ligeiramente, da maioria das outras. Assim, pode-se identificar a propensão genética de um indivíduo ao desenvolvimento de alguma doença e tratá-lo com uma droga mais específica. Estudos de variação genética em populações humanas começaram a ser realizados na Primeira Guerra Mundial I, com a identificação da frequência do grupo sanguíneo. Como resultado, a identificação de algumas das diferenças entre os povos permitiu a realização de tratamentos mais especializados, notavelmente no que diz respeito a transplantes e transfusão de sangue.

A diversidade entre os povos é extremamente importante para as novas formas de medicina e pesquisa, e isso acaba por constituir uma enorme base de dados genéticos. O Grupo de Trabalho Internacional de Mapeamento de SNPs (*The International SNP Map Working Group*) é um consórcio massivo de empresas e grupos governamentais empenhados em encontrar SNPs. No início do projeto, eles identificaram e mapearam 1.42 milhões de SNPs em questão de meses, uma forte indicação de que devem existir muitos outros milhões de diferentes variações herdadas. Segundo o *SNP-Consortium* (2001):

“SNPs são indícios importantes de que as pessoas têm ascendência bastante diversificada, como nunca se suspeitou antes. Compreende-se hoje que as diferenças humanas são muito mais complexas do que as classificações das raças e tribos realizadas ao longo dos últimos mil anos. A ausência de diferenças genéticas entre a maioria dos

<sup>7</sup>O software BLAST compara sequências de DNA através de informações disponíveis *online* e de amostras previamente armazenadas para determinar onde as homologias ocorrem. (NCBI, 2012)



humanos e para a maioria dos genes sugere, também, que há muitas características que são compartilhadas de forma comum, mesmo quando se pensava anteriormente que os genes em questão distinguia um grupo de outro.”

A atividade de identificar as diferenças de traços entre os seres humanos não pode ser conseguida utilizando-se apenas os genes. McGee (2003) enfatiza, que “é necessário procurar em todo o genoma, incluindo as regiões chamadas de ‘lixo’ (ou *junk*), que correspondem de 90 a 95% dos três bilhões de bits de DNA que compõem o genoma”.

### 6.6.1 Características da Similaridade

“O senso de similaridade é o elemento mais estrutural de nosso pensamento. A compreensão do processo de solução de problemas, categorização, recuperação da memória, raciocínio indutivo e outros processos cognitivos requerem que entendamos como o ser humano lida com a similaridade” (WILSON; KEIL, 2006).

A similaridade desempenha um papel fundamental nas teorias do conhecimento e comportamento. Serve como um princípio organizador pelo qual os indivíduos classificam objetos, conceitos, formas e fazem generalizações.

Tversky (2004) explica que:

“Na verdade, o conceito de similaridade é onipresente na teoria psicológica. Ela é a base da verificação de estímulo e resposta na aprendizagem, da explicação de falhas de memória e reconhecimento de padrões, e da análise do significado conotativo. Dados de similaridade ou dissimilaridade aparecem em diferentes formas: avaliações de pares, classificação de objetos, comunalidade entre associações, erros de substituição e de correlação entre as ocorrências. A análise desta informação tenta explicar a semelhança observada nas relações, visando capturar a estrutura subjacente dos objetos em estudo.”

Quanto à modelagem de relações de similaridade, Tversky (2004) continua:

“A análise teórica de relações de similaridade tem sido dominada por modelos geométricos. Esses modelos representam objetos como pontos em um espaço de coordenadas tais que as dissimilaridades observadas entre os objetos correspondem às distâncias métricas entre os respectivos pontos. Praticamente todas as análises de dados de proximidade têm alguma métrica associada, embora algumas (agrupamento hierárquico, por exemplo) utilizam estruturas de árvore ao invés de espaços dimensionalmente organizados. No entanto, a maior parte das análises teóricas e empíricas de similaridade assumem que os objetos podem ser adequadamente representados como em alguns pontos de coordenadas no espaço e a dissimilaridade se comporta como uma métrica pela função de distância. . . Tem sido argumentado que representações tridimensionais são apropriadas para certos estímulos (por exemplo, cores, tons), mas não para outros. Parece mais adequado para representar as faces, países ou personalidades em termos de muitas características qualitativas em detrimento das dimensões quantitativas. . . A função de distância métrica é uma escala que atribui a distância positiva entre cada par de pontos, de acordo com três axiomas: minimalidade, simetria e desigualdade triangular”.

ZeZula et al. (2006) argumentam que as técnicas de recuperação da informação tradicionais, tipicamente baseadas em rotinas de classificação e tabelas de índices não são apropriadas para

o crescente número de domínios de dados emergentes que surgem a todo o tempo<sup>8</sup> A solução se encontra no uso de índices de estruturas para a busca por similaridades métricas no espaço do domínio de discurso, também conhecida como “busca por similaridade”.

Goldstein (2012) apresenta vários exemplos de similaridade no processo psicológico de percepção: acústica; estrutural, entre figuras; entre tarefas; atratividade etc.

Algumas aplicações de similaridade fenotípica são apresentadas na seção 9.1.

## 6.6.2 Similaridade Molecular

Carbó-Dorca e Mezey (1996) organizaram uma publicação contemplando os avanços na área de similaridade molecular, como medidas de similaridade na química molecular, similaridade dos átomos nas moléculas, análise de conformidade sob o ponto de vista da similaridade molecular, mecanismos de busca automática para sub-estruturas de similaridade, combinação canônica para medidas de similaridade entre moléculas, e outros. Eles explicam que:

“Similaridade molecular é um conceito fundamental da química. . . Produtos químicos que exibem propriedades semelhantes dentro de um contexto muitas vezes apresentam propriedades similares dentro de um contexto diferente. Usando a similaridade, previsões podem ser feitas mesmo sem entender muito do comportamento molecular, ou mesmo sem saber nada da existência de moléculas. Apoiando-se apenas na similaridade, boa parte do conhecimento químico acumulado poderia ser organizado de maneira sistemática, assumindo um papel importante no eventual reconhecimento de tendências, relações, regras e muitas das leis fundamentais da química.”

## 6.6.3 Auto-Similaridade

Schroeder (2006) afirma que:

“A natureza está repleta de fenômenos periódicos: do movimento de um balanço às oscilações de um átomo, do chilrear de um gafanhoto às órbitas dos corpos celestes. E os nossos próprios corpos, também, participam deste minueto universal – desde o batimento cardíaco até ciclos de vida bem mais longos. É claro que nada na natureza é exatamente periódico. Todo movimento tem um começo e um fim, de modo que, no sentido matemático, a periodicidade estrita não existe. No entanto, a periodicidade provou ser um conceito extremamente útil para elucidar as leis e os mecanismos subjacentes. Mas existem numerosos outros fenômenos em que estes princípios fundamentais são quebrados e, em vez de periodicidade, temos aperiodicidade ou mesmo um movimento caótico. . . O paradigma da previsibilidade, é ofuscado por fenômenos que são um refúgio para o caos – ainda que um caos determinista. Mas não importa o quanto a vida possa ficar caótica. . . , um outro baluarte fundamental muitas

---

<sup>8</sup>Um fator crítico para a utilidade da informação é a sua qualidade e a velocidade com que ela é disponibilizada. Em todas as áreas do conhecimento a quantidade de informação disponível está crescendo rapidamente, gerando um estado psicológico que Turban, McLean e Wetherbe (2004) denomina de “ansiedade informacional”, caracterizado pela frustração com a incapacidade de se manter atualizado; a incapacidade de qualificar a informação disponível na web; o excesso de fontes online; a culpa associada ao fato de não estar mais bem informado ou informar-se tarde demais. O perfil infométrico do usuário é dinâmico, modificando-se continuamente no tempo, o que resulta em atualização e refinamento permanente de seus interesses conceituais.

vezes permanece inabalável: a auto-similaridade, que é uma invariância em relação à escala. . . Muitas leis da natureza são independentes, ou quase isso, de um fator de escala. O fato de qualquer ampliação ter sempre um limite (a constante de Planck, quando as coisas ficam muito pequenas, ou a velocidade da luz, quando os objetos se movimentam muito rapidamente), não desestimulam pensarmos na utilidade da auto-similaridade, assim como a falta de periodicidade rigorosa fora da matemática não traz grandes consequências para o mundo real. Na verdade, auto-similaridade nada mais é do que a periodicidade em uma escala logarítmica<sup>9</sup>.”

#### 6.6.4 Similaridade Fenotípica

Estudos da UNL (2012) apontam que “casais muitas vezes se assemelham em uma variedade de características fenotípicas a um nível significativamente maior do que o esperado por acaso, indicando que existe uma busca consciente ou não por similaridade fenotípica na escolha dos parceiros conjugais”. Por exemplo, os homens mais altos tendem a se casar com mulheres mais altas e os casais tendem a ter excesso de peso ou ambos são magros. Os casais também tendem a ser semelhantes em seus escores de QI, traços de personalidade, e muitas outras características. Tal observação é chamada de Acasalamento Associativo Positivo (PAM).

Os casais não são mais semelhantes do que o esperado, simplesmente porque eles convivem e compartilham um ambiente físico, social e intelectual comum. Tanto os recém-casados como os com mais vivência conjugal se assemelham um ao outro. Quando os casais são estudados, identifica-se uma consistência com a estrutura de preferência abstrata entre eles: As pessoas são mais atraídas para o par “ideal”, mas depois descobrem a dificuldade da conquista do ideal e se realizam com companheiros que tenham mais similaridades com elas. Quanto mais abrangente é o conjunto de características fenotípicas comparadas, menor se verifica a distância de similaridade métrica entre eles, ou seja, mais próximos são um do outro.

---

<sup>9</sup>Uma escala logarítmica é uma escala que usa o logaritmo de uma grandeza em vez da grandeza propriamente dita, facilitando a visualização e realização de cálculos, sendo útil quando os dados cobrem uma grande gama de valores. Alguns de nossos sentidos operam de maneira logarítmica. Outros exemplos incluem o cálculo de intensidade de terremotos, movimento da terra, intervalos de notas musicais, entropia (Termodinâmica) e informação (Teoria da Informação) (ENCYCLOPEDIA.COM, 2004).



# 7 Uma Ontologia para a Arquitetura Genética da Informação

*“O essencial é saber ver – Mas isso (triste de nós que trazemos a alma vestida!) Isso exige um estudo profundo. Uma aprendizagem de desaprender. . . procuro despir-me do que aprendi, procuro esquecer-me do modo de lembrar que me ensinaram, E raspar a tinta com que me pintaram os sentidos, desencaixotar as minhas emoções verdadeiras, desembrulhar-me e ser eu.”*

— Fernando Pessoa.

Como vimos nas seções 2.2 e 3.7, o termo “informação” é pervasivo e ubíquo tanto na Ciência da Informação como na Biologia (particularmente na Genética). Em busca de uma analogia semântica de conceitos diretamente relacionados à “informação” entre os dois domínios verificamos a necessidade de construção de duas ontologias formais, de alto-nível que possam ser alinhadas, evidenciando o papel fundamental que a informação exerce nestas duas ciências, fundamento da Arquitetura Genética da Informação (cap. 4, i.e, uma arquitetura da gênese da *informação intencional* e de sua caracterização sob uma perspectiva genética. Esta caracterização compreende o genoma, genótipo e fenótipo da informação, além de processos biológicos relacionados à informação.

O desenvolvimento de ontologias visa compartilhar o entendimento da estrutura da informação entre pessoas ou agentes de software; permitir o reuso do domínio do conhecimento; explicitar suposições; e analisar o domínio do conhecimento. No processo de construção de ontologias, deve-se delimitar bem o escopo da ontologia; considerar o reuso de ontologias existentes; enumerar importantes termos da ontologia; definir as classes de conceitos, suas propriedades, restrições e instâncias.

Como produto de uma ontologia, temos o metadados semântico, cuja função é descrever o conteúdo do documento, considerando os aspectos descritivos textuais ou relacionados especificamente às entidades a que estão se referindo. O metadados semântico torna viável as buscas por significado; a apresentação de dados em agrupamentos de conceitos; a combinação de informação, eliminação de redundância e sumarização; o relacionamento entre entidades-chaves; as inferências automáticas para a produção de novo conhecimento; a integração de informações de fontes heterogêneas; e a identificação de processos que utilizam a mesma informação, dentre tantas outras funções.

## 7.1 Genética Clássica da Informação

Em 1866, Mendel descreveu os genes através de seus efeitos finais, i.e, sua expressão gênica, ou fenotípica. As mensagens escondidas no gene constituem as instruções de construção de um organismo, controlando sua aparência, comportamento e reprodução. Na *Arquitetura Genética da Informação* isso se relaciona à expressão (representação), características (atributos), hereditariedade (derivação) e comportamento do objeto.

Um gene pode controlar um ou múltiplos fenótipos, nesse último caso, é chamado de gene pleiotrópico. A maioria dos defeitos genéticos são associados a efeitos de pleiotropia. Isto é particularmente importante, ao considerarmos a nossa arquitetura em termos de desqualificação da informação em contextos de seleção por relevância conceitual.

## 7.2 Genética Molecular da Informação

Na introdução de seu livro “Uma História da Biologia Molecular”, (MORANGE, 2000) explica:

“Biologia Molecular não é meramente uma descrição da biologia em termos de moléculas... Mais do que isso, consiste de todas aquelas técnicas e descobertas que tornam possível a análise molecular dos processos biológicos mais fundamentais – aqueles envolvidos na estabilidade, sobrevivência, e reprodução dos organismos... A Biologia Molecular nasceu quando os geneticistas, não mais satisfeitos com a visão quasi-abstrata do papel dos genes, focaram no problema da natureza dos genes e seu mecanismo de ação.”

Morange acredita que a justificação para uma dicotomia, para um status distinto, é a “nova maneira de olhar os organismos como reservatórios e transmissores de informação”. Não se trata de uma nova maneira, mas mais do que isso, uma extensão molecular das noções fenomenológicas deterministas dos cientistas genéticos.

A transmissão genética formulada em termos da Teoria da Informação e cibernética é uma metáfora enganosa. A noção cibernético-matemática da Teoria da Informação lida com a confiança probabilística da transmissão de sinais, ao passo que a metáfora da informação da análise genética se refere à comparação de transmissão de informação semântica, como tendo sido obtida pela metodologia de hibridização, quer seja de organismos ou de ácidos nucleicos (FALK, 2009). Sterelny e Griffiths (1999) referem-se à última como *informação intencional*, que, sendo immanentemente independente de contexto, promove o determinismo genético.

Genética molecular, segundo Burian (2005), é uma disciplina construída em torno de questões centrais como: Qual é o material genético? Como ele está organizado? Como a estrutura do gene se relaciona com sua expressão e função? Por que mecanismos os genes (ou a informação genética) são transmitidos de uma geração para a próxima? Como o material genético eventualmente afeta traços particulares dos organismos? Na Ciência da Informação, podemos fazer as mesmas perguntas em relação à informação, i.e, Qual é o objeto da informação? Como é a sua organização? Qual é a relação entre a sua estrutura e a sua expressão e função? Como acontece

a transmissão de informação? Como a informação caracteriza o objeto? Então, da genética molecular, aproveitamos as estruturas moleculares (físicas e químicas) para explicar a estrutura da informação; suas interações, ou seja, os relacionamentos entre objetos da informação; e o seu comportamento.

O foco da genética molecular são as estruturas físicas e químicas do DNA. explica que o DNA é a base da vida e que todos os seres vivos (e alguns seres não-vivos) utilizam-no para armazenar informação genética e transmiti-la de uma geração para a próxima. Uma cópia de parte ou de todo o DNA é passada para a sua descendência.

Os genes são expressos através de um complexo sistema de interações que começam por copiar as mensagens do DNA para um lugar temporário chamado RNA. O RNA carrega a mensagem do DNA pelo processo de translação, que, em essência, é como um anteprojeto ou guia de fabricação, fornecendo as proteínas para a concepção do organismo [Robinson \(2006\)](#) por meio desta arquitetura genética.

Na *Arquitetura Genética da Informação*, o processo em que a informação é percebida em seus atributos peculiares mensuráveis (dados sensoriais) que o caracterizam como indivíduo no mundo (indivíduo aqui no sentido da unicidade e singularidade existencial de um objeto) ocorre em um contexto espaço-temporal. A falta do sensoriamento de determinado atributo leva à incapacidade de se criar uma classificação em torno desse atributo e/ou de outros diretamente a ele associados. Essa falta pode ocorrer, também, havendo a presença de grande quantidade de ruído no sensoriamento. Em ambos os casos, temos o que [Floridi \(2007\)](#) denomina “desinformação” e Shannon, de mensagem com alto grau de entropia ou desordem.

Além da transcendência do objeto, há as limitações sensoriais no processo de percepção. Assim, a apreensão das propriedades do objeto em sua totalidade (genoma) será sempre impossível. Apenas um subconjunto do seu genótipo (possivelmente satisfatório) serão suficientes para a sua concepção como *informação intencional* (nesse contexto). Daí, um objeto real pode ter diferentes percepções correspondentes a contextos diversos. Exemplificando, podemos perceber uma música e classificá-la pelo seu estilo, ritmo, tipo de instrumentos, arranjo etc. Temos, então, vários planos dimensionais passíveis de serem considerados em um objeto.

A tecnologia de sequenciamento de DNA tem sido usada para descobrir a natureza oculta do código genético. A informação genética é, então, usada para rastrear, diagnosticar e tratar distúrbios genéticos. As mutações são responsáveis por variações fenotípicas, que ocorrem espontaneamente o tempo todo, e, muitas vezes, aleatoriamente ([ROBINSON, 2006](#)). Pode acontecer de estas mutações modificarem a atividade normal do gene, causando doenças e defeitos. Aqui, fazemos a correspondência com o que [Floridi \(2007\)](#) entende ser “má-informação” (conforme apresentado anteriormente na figura 6, pág. 37).

### 7.3 Genética Populacional da Informação

Da genética populacional, podemos verificar a similaridade de expressão fenotípica em um corpus de documentos, visando obter agrupamentos orientados por características comuns parametrizadas entre os seus diversos objetos da informação. A realidade existente é percebida através de um conjunto de sensores, representado por um vetor de características que inclui tanto as medições diretas dos sensores como possíveis pós-processamentos destas medidas.

A similaridade de objetos é obtida pela comparação da representação do fenótipo dos objetos, onde as características fenotípicas de um objeto são comparadas às características fenotípicas de outro objeto. Assim, a Teoria da Informação pode ser útil para a busca de paridade entre documentos, visando agrupamentos e recuperação da informação. [Aslam e Frost \(2003\)](#) explicam que, segundo o paradigma físico da informação, a similaridade é encontrada pela medição de quanto de informação dois objetos têm em comum e quanto têm de diferença.

A memética tem inúmeras aplicações na Genética Populacional, como, por exemplo, o caso de inovação e evolução tecnológica, como na aplicação apresentada na seção [9.7](#).

### 7.4 Genética Quantitativa da Informação

Da genética quantitativa, podemos aplicar técnicas de descoberta de conhecimento em bancos de dados (KDD), particularmente de mineração de dados, para a descoberta de padrões da informação não evidentes.

A aplicação apresentada na seção [9.6](#) diz respeito a indicadores de desenvolvimento combinados para a composição do Índice DNA Brasil baseado em vários componentes estatísticos.

### 7.5 Construção das Ontologias

Segundo [Losee \(2001\)](#), Mary Hesse discorre sobre o uso científico de analogias.

Ela explica que há dois tipos de relações entre sistemas análogos: relações de similaridade entre propriedades e relações causais ou funcionais que acontecem tanto em um sistema como em seu análogo.

Apresentamos na figura [22](#) a proposta de Hesse para o uso de analogias científicas entre domínios do conhecimento humano.

Com base neste diagrama, podemos afirmar que propriedades correspondentes em cada domínio são similares e que há relações causais/funcionais do mesmo tipo que ligam os termos dentro de cada domínio. Hesse enfatiza que estas duas afirmações podem ser objeto de indagação de sua aplicabilidade. Hesse não pretende estabelecer uma analogia formal, mas uma analogia material entre os dois domínios, onde se estabelecem critérios para ligações análogas apropriadas.





Figura 22: Mary Hesse e o Uso de Analogias Científicas - Fonte: (LOSEE, 2001)

As ontologias foram construídas sob os princípios da *Basic Formal Ontology - BFO*, desenvolvida pelo grupo de pesquisa de Smith (1998), que consiste de uma série de sub-ontologias em diferentes níveis de granularidade.

As variedades dos termos ontológicos se limitam aos continuantes (ou instantâneos) e ocorrentes (ou perdurantes). Os continuantes têm características estático-espaciais da realidade, tratando-se de um inventário de todas as entidades existentes em determinado momento. Já os ocorrentes são processos que acontecem no tempo, com aspectos dinâmico-temporais da realidade, tratando-se de um inventário de todos os processos existentes em um dado intervalo de tempo.

Este trabalho foi conduzido usando-se os recursos da ferramenta Protégé, que é suportada pela concessão LM007885 da Biblioteca Nacional de Medicina dos Estados Unidos (STANFORD UNIVERSITY SCHOOL OF MEDICINE, 2012).

Construímos duas ontologias:

1. Ontologia de Termos da Informação na Genética;
2. Ontologia de Processamento da Informação em Sistemas Biológicos.

### 7.5.1 Construção da Ontologia de Termos da Informação na Genética

Identificamos sete grandes categorias da Genética nas fontes de referência para aquisição do conhecimento: Engenharia Genética, Genética Clássica, Genética Molecular, Genética Humana, Genética Populacional, Genética Quantitativa e Genômica, que descrevemos a seguir:

- A Engenharia Genética refere-se a um conjunto de tecnologias usadas para alterar a composição genética das células e mover genes através das fronteiras de espécies para a produção de novos organismos. As técnicas envolvem manipulações altamente sofisticadas de material genético e de outros produtos químicos biologicamente importantes.

- A Genética Clássica (ou Mendeliana) atenta para os indivíduos e famílias, estudando traços físicos e os genes que controlam a aparência, i.e, o fenótipo e sua transmissão para outras gerações.
- A Genética Molecular se ocupa com as estruturas físicas e químicas do DNA, RNA e proteínas.
- A Genética Humana estuda a transmissão genética entre os seres humanos.
- A Genética Populacional é uma ampliação da genética clássica e humana, considerando grupos maiores.
- A Genética Quantitativa é um campo matemático que examina os relacionamentos estatísticos entre os genes e os traços por ele codificados.
- A Genômica está envolvida com o estudo dos genomas dos organismos. O campo inclui esforços para determinar o sequenciamento de DNA completo dos organismos e o seu mapeamento genético.

As propriedades fundamentais de cada uma destas categorias nortearam a construção da ontologia de termos da informação na genética.

O diagrama da figura 23 apresenta os passos do processo de construção da ontologia pela metodologia do Modelo-V (fig. 11).

#### 7.5.1.1 Fase 1: Identificação do propósito e escopo

O propósito da ontologia é contemplar os termos no domínio da Biologia com alguma relação ou associação aos conceitos de informação, como na comunicação e linguagem, por exemplo.

O escopo cobre os dez primeiros níveis da hierarquia da natureza (desde as partículas e moléculas até o nível de populações), conforme mostrado na tabela 2. Estes níveis formam uma taxonomia proposta por Sawai (2011), onde foram identificados termos relacionados à informação.

#### 7.5.1.2 Fase 2: Aquisição de conhecimento

Realizamos um processo de mineração textual em duas enciclopédias, a “*Encyclopedia of Genetics*” (KNIGHT; NESS, 2004) e a “*Encyclopedia of Biology*” (RITTNER; MCCABE, 2004); e, complementarmente, em Steinberg e Cosloy (2009), de maneira a contemplar os termos relevantes a serem elicitados na construção da ontologia.

Além dos termos ali selecionados, outros surgiram na verificação de sinonímia (sinônimos) e hiperonímia (significados).

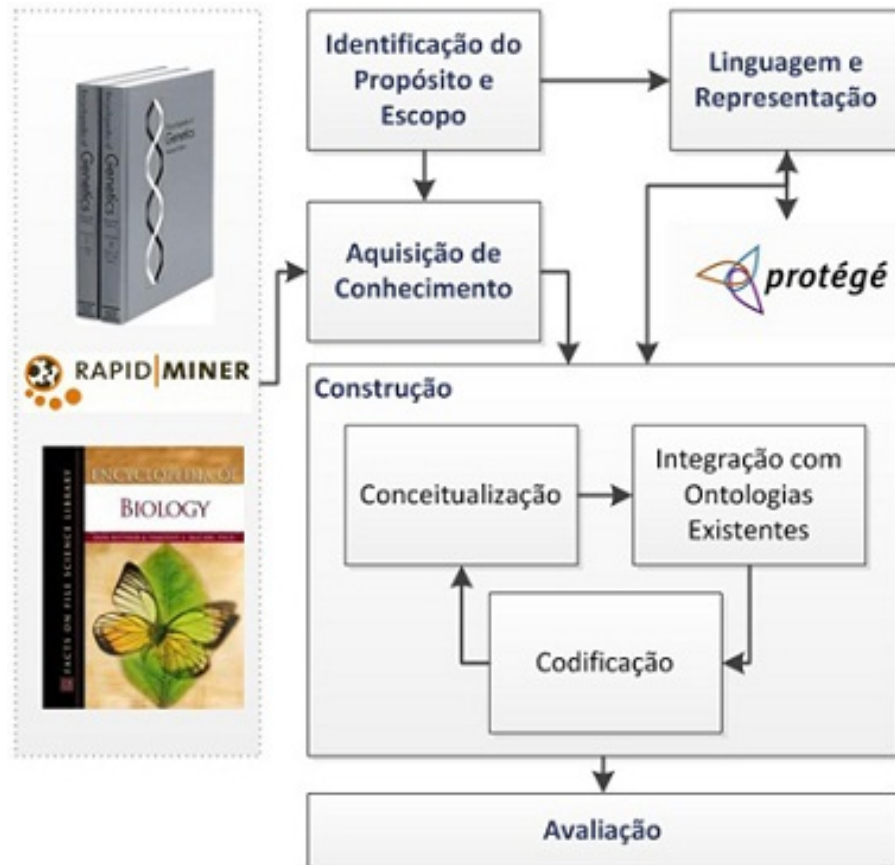


Figura 23: Ciclo de Vida de Construção da Ontologia de Termos da Informação na Genética - Fonte: Produzido pelo autor

Foram eliminados, por exemplo, termos relacionados a especificidades de doenças ou síndromes, técnicas e métodos da engenharia genética, ou aqueles cuja relação com a “informação” não se verificou evidente em primeira análise.

Para realizar esse processo semi-automático desenvolvemos o algoritmo de mineração textual no software RapidMiner (RAPID-I, 2012), extraindo os termos de entrada nas enciclopédias, com filtragem de expressões regulares (“regex”), que é uma forma concisa e flexível de identificar cadeias de caracteres particulares, palavras ou padrões de caracteres.

Na primeira etapa do processo é especificado o padrão da expressão regular que separa as entradas da enciclopédia das outras palavras no texto.

A figura 24 apresenta o resultado final obtido neste processo.

Os termos da “*Encyclopedia of Genetics*” estão explicitamente associados às áreas da genética em que se aplicam, garantindo assim a sua validação na classificação dentro das classes principais da ontologia, que são estas áreas, conforme apresentado na seção 7.5.1.5.

### 7.5.1.3 Fase 3: Conceitualização

Uma vez identificadas e extraídas as palavras, elas foram traduzidas para o português, juntamente com um resumo de sua definição. Tanto os termos como as definições fizeram parte

| <i>Nível</i>        | <i>Aplicações</i>  |
|---------------------|--|
| 1. Partículas       | Nano-bio ciência; química artificial; comunicação molecular.   |
| 2. Genes            | Biologia molecular; teoria neutra da evolução; evolução de vírus.  |
| 3. Aminoácidos      | Algoritmo genético químico (CGA/CGP).  |
| 4. Proteínas        | Engenharia de proteínas; comunicação molecular (proteínas motoras).  |
| 5. Células          | Comunicação molecular e de neurons.  |
| 6. Tecidos e Órgãos | Engenharia de tecidos, cérebro e mente; consciência.   |
| 7. Organismo        | Engenharia de tecidos, cérebro e mente; consciência.   |
| 8. População        | Sistemas de imunização; Estratégia evolucionária estável; evolução de comportamento altruístico; sistemas multi-agentes. |
| 9. Espécies         | Teoria darwiniana da evolução; neo-darwinianismo; co-evolução.   |
| 10. Ecossistema     | Teoria de segregação de habitats; mimetismo; estratégia de migração de população.  |
| 11. Terra           | A hipótese de Gaia; problemas ambientais.  |
| 12. Universo        | Origem da vida.  |

Tabela 2: Tabela de Hierarquia da Natureza - Fonte: [Sawai \(2011\)](#)

| Word         | Attribute Name | Total Occurrences | Document Occu... | Biologia |
|--------------|----------------|-------------------|------------------|----------|
| Aggression   | Aggression     | 1                 | 1                | 1        |
| Aging        | Aging          | 1                 | 1                | 1        |
| Agricultural | Agricultural   | 1                 | 1                | 1        |
| Albinism     | Albinism       | 1                 | 1                | 1        |
| Alcoholism   | Alcoholism     | 1                 | 1                | 1        |
| Alleles      | Alleles        | 1                 | 1                | 1        |
| Allergies    | Allergies      | 1                 | 1                | 1        |
| Alphabetical | Alphabetical   | 1                 | 1                | 1        |
| Altruism     | Altruism       | 1                 | 1                | 1        |

Figura 24: Extração de termos com mineração textual - Resultado - Fonte: Produzido pelo autor

das anotações da ontologia.

#### 7.5.1.4 Fase 4: Integração

Nesta fase de integração com ontologias existentes, foram consideradas as classes, relações e anotações da *Gene Ontology* - GO e a BioTop, uma ontologia de alto-nível da Biologia, ambas no idioma Inglês.

A versão mais recente da BioTop é totalmente modularizada, obedecendo aos princípios de limites de domínio; tamanho adequado dos módulos, visando manutenibilidade; estabelecimento de pontes entre domínios; e compatibilidade com a BFO 2.0.

#### 7.5.1.5 Fase 5: Codificação

Na fase de codificação, as classes de primeiro nível da ontologia correspondem às principais áreas de estudo da Genética: a Engenharia Genética, Genética Molecular, Genética Humana, Genética Populacional, Genética Quantitativa e Genômica.

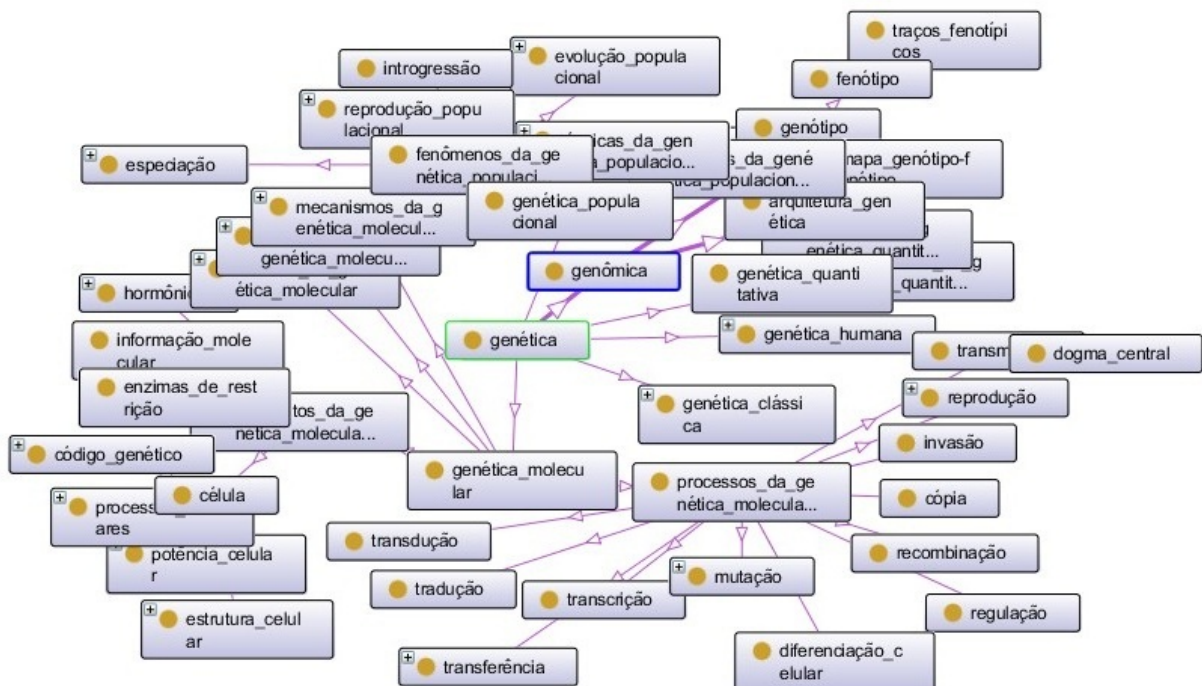


Figura 25: Visualização Gráfica Parcial da Ontologia de Termos da Informação na Genética - Fonte: Produzido pelo autor

Parte do gráfico da ontologia, gerado pelo módulo *OntoGraf* do *Protégé* é apresentado na figura 25.

As relações típicas em ontologias biomédicas de alto-nível são apresentadas na figura 26

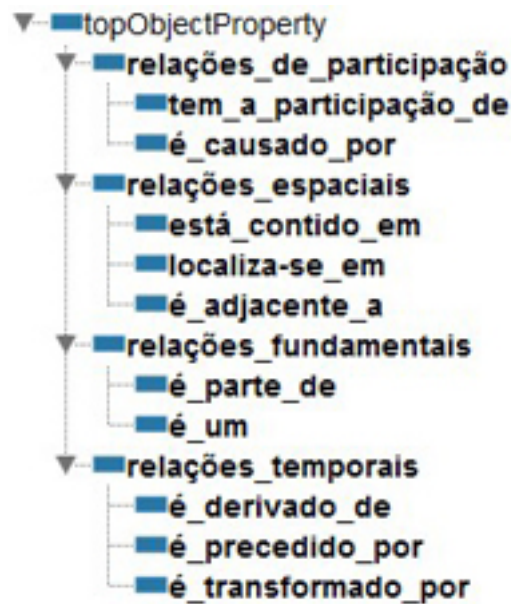


Figura 26: Relações típicas em Ontologias Biomédicas de alto-nível - Fonte: Produzido pelo autor

#### 7.5.1.6 Fase 6: Documentação

As ontologias desenvolvidas com a ferramenta Protégé ([STANFORD UNIVERSITY SCHOOL OF MEDICINE, 2012](#)) são auto-documentáveis, contemplando estrutura, termos, relações, regras, anotações diversificadas, possibilidade de visualização gráfica, mecanismos de raciocínio, verificação de integridade etc.

#### 7.5.1.7 Fase 7: Avaliação

Considerando as orientações no item [3.5.3](#), a estrutura é orientada às especialidades da Genética, dentro da área do conhecimento de Biologia.

Sua expressividade se evidencia na cobertura das principais abordagens de estudo da Genética.

Sua granularidade é de até cinco níveis, dentro de cada classe principal.

Trata-se de uma ontologia descritiva.

Permite raciocínio automático e sua utilidade é em termos de conceitos da informação dentro da Genética, podendo ser integrada a outras ontologias de domínio.

## 7.5.2 Construção da Ontologia de Processamento da Informação em Sistemas Biológicos

A ontologia de processamento da informação em sistemas biológicos centrou-se nos sistemas biológicos (componentes genéticos, execução de programas, expressão da informação, interpretação de códigos e genoma da informação) e teoria da informação (de [Shannon \(1948\)](#), de [Kolmogorov \(1950\)](#) e molecular, de [Nalewski \(2006\)](#)).

O diagrama da figura 27 apresenta os passos do processo de construção da ontologia pela metodologia do Modelo-V (fig. 11).

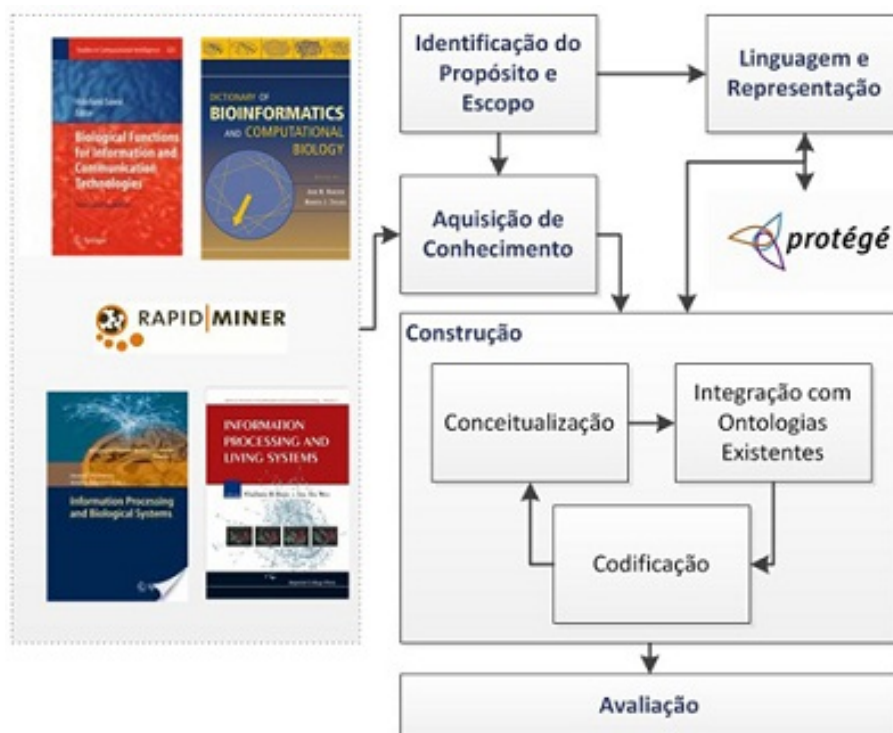


Figura 27: Ciclo de Vida de Construção da Ontologia de Processamento da Informação em Sistemas Biológicos - Fonte: Produzido pelo autor

### 7.5.2.1 Fase 1: Identificação do propósito e escopo

O propósito da ontologia é contemplar os termos relacionados à informação em sistemas biológicos e nas teorias da informação de Shannon (1948), Kolmogorov (1950) e molecular, de Nalewski (2006).

### 7.5.2.2 Fase 2: Aquisição de conhecimento

Por convenção, suficiência e praticidade, realizamos um processo de mineração textual nas seguintes fontes: “*Dictionary of Bioinformatics and Computational Biology*” (HANCOCK; ZVELEBIL, 2004) (289 p.), “*Biological Functions for Information and Communication Technologies: Theory and Inspiration*” (SAWAI, 2011) (226 p.), “*Information Processing and Biological Systems*” (NIIRANEN; RIBEIRO, 2011) (229 p.) e “*Information Processing and Living Systems*” (BAJIC; WEE, 2005) (778 p.).

A maioria dos conceitos emergiu desta mineração, cujo processo é apresentado no diagrama da figura 28.

Esse processo inclui um módulo de criação de vetores, compreendendo: transformação para letras minúsculas, *tokens*, filtro de *stop-words* (palavras irrelevantes, como artigos, preposições

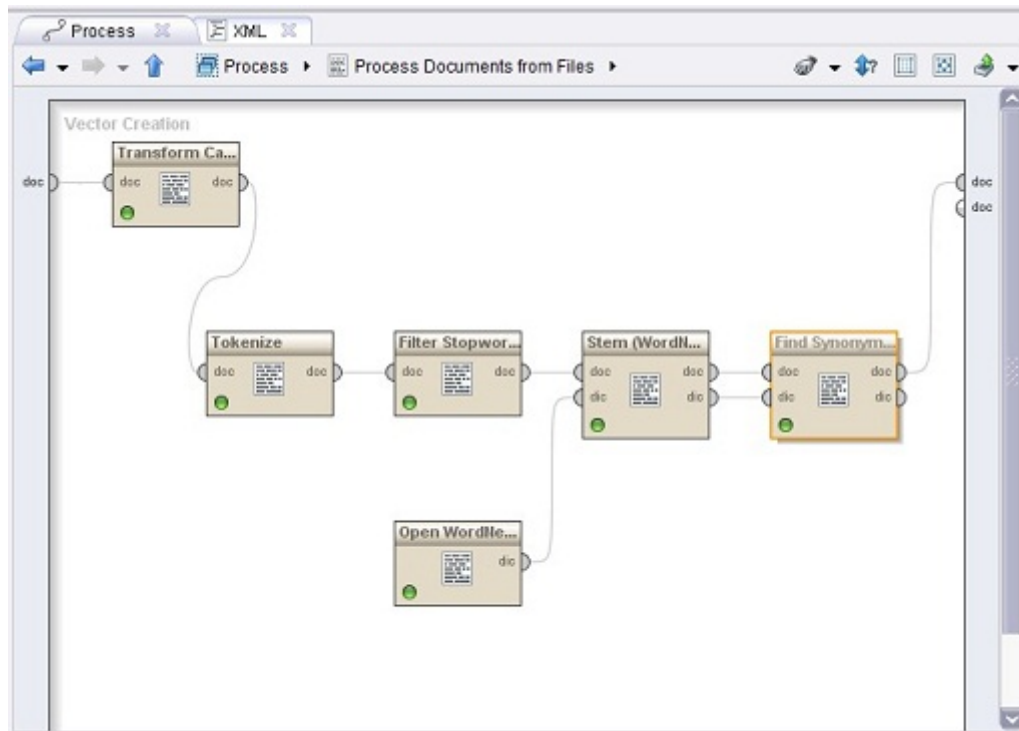


Figura 28: Processo de Mineração Textual - Fonte: Produzido pelo autor

e advérbios) e identificação dos prefixos das palavras (de 3 a 25 caracteres) por consulta aos substantivos que constam do *Open WordNet Dictionary*. Como resultado do processo, foram identificados <sup>1</sup> automaticamente, 8.884 atributos regulares, dos quais, os mais relevantes fazem parte do gráfico da figura 29.

### 7.5.2.3 Fase 3: Conceitualização

Da mesma maneira como foi feito com a primeira ontologia, uma vez identificadas e extraídas as palavras, elas foram traduzidas para o português, juntamente com um resumo de sua definição. Tanto os termos como as definições fizeram parte das anotações da ontologia.

### 7.5.2.4 Fase 4: Integração

Esta ontologia foi desenvolvida considerando-se os aspectos da primeira, explanada na seção 7.5.1.

### 7.5.2.5 Fase 5: Codificação

Na fase de codificação, a ontologia foi construída com seis níveis de profundidade. As classes de primeiro nível da ontologia correspondem aos sistemas biológicos, com os seus componentes, execução de programas, expressão da informação, genoma da informação, e interpreta-

<sup>1</sup>Tempo de execução desse processo no RapidMiner: 80 segundos.



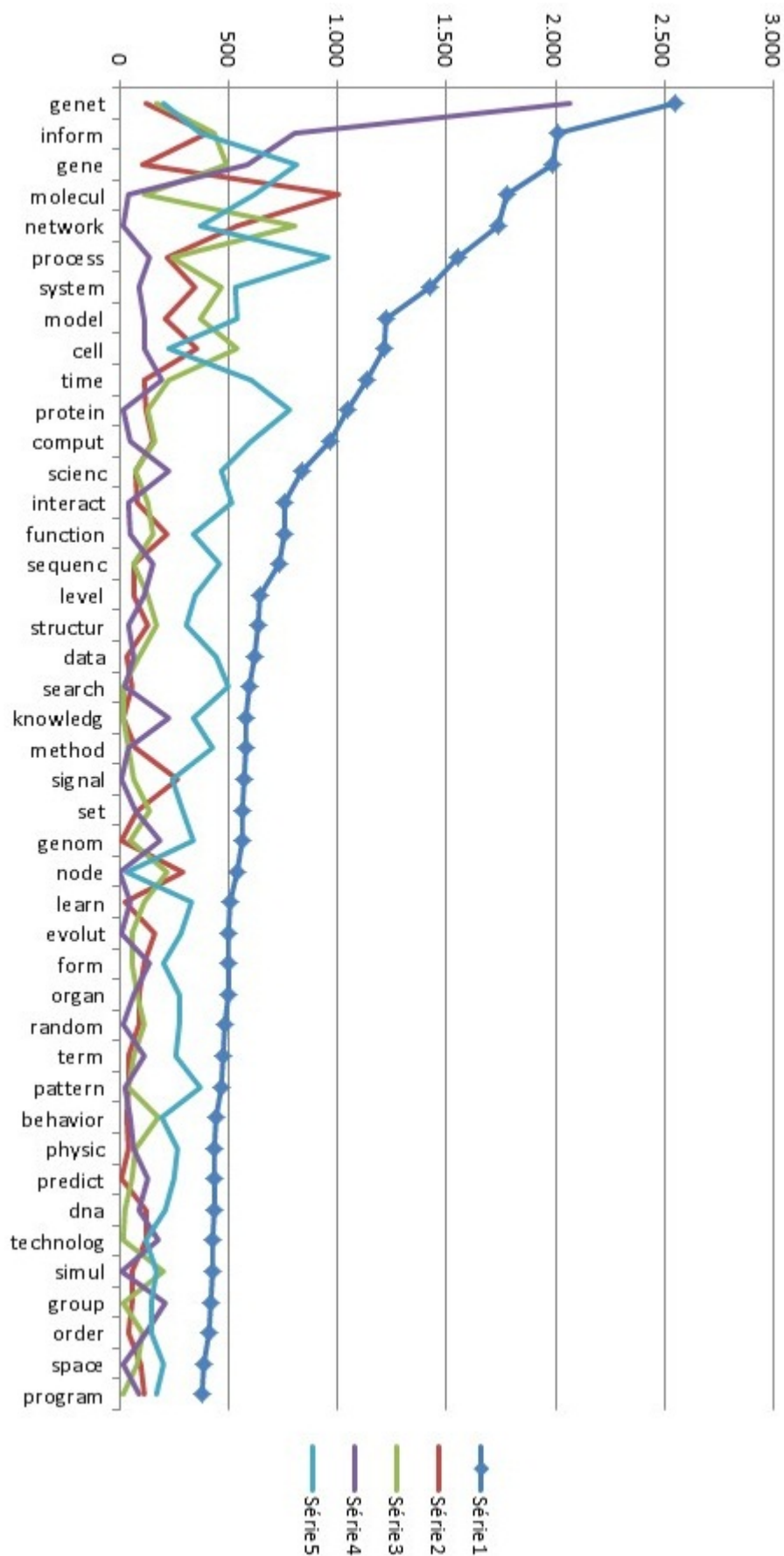


Figura 29: Resultado do Processo de Mineração Textual - Fonte: Produzido pelo autor

ção de códigos; e às teorias da informação (Shannon (1948), Kolmogorov (1950) e Molecular, de Nalewski (2006)).

Parte do gráfico da ontologia, gerado pelo módulo OntoGraf é apresentado na figura 30.

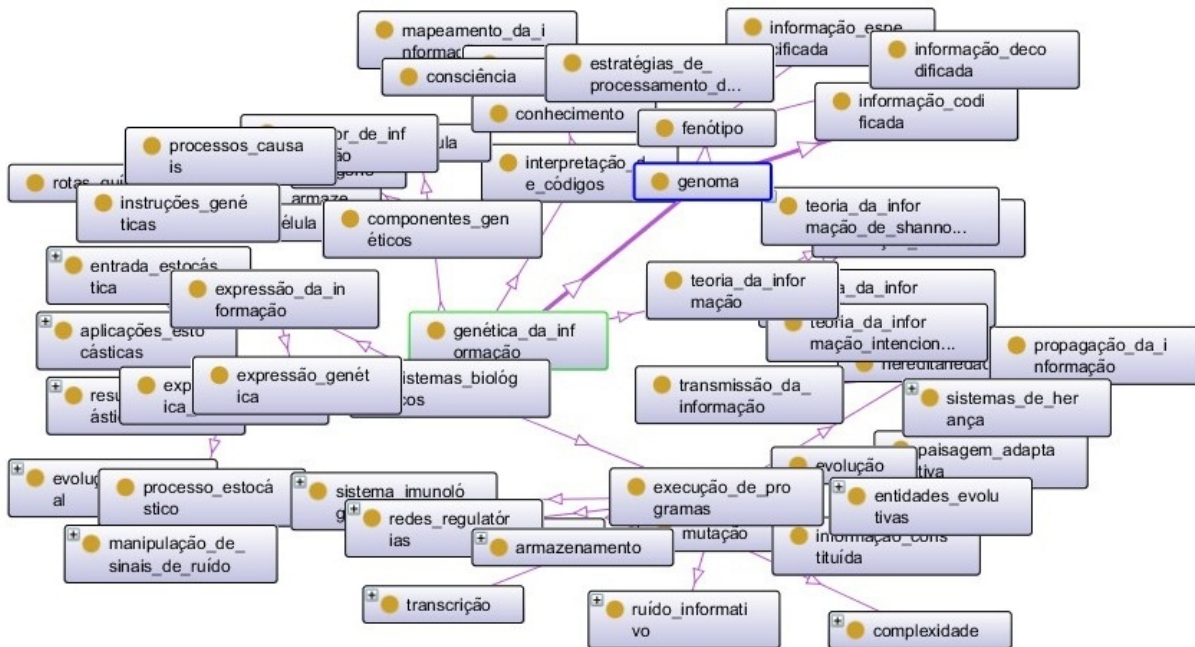


Figura 30: Visualização Gráfica Parcial da Ontologia de Processamento da Informação em Sistemas Biológicos - Fonte: Produzido pelo autor

#### 7.5.2.6 Fase 6: Documentação

Conforme explicado anteriormente no item 7.5.1.6, as ontologias desenvolvidas com a ferramenta Protégé (STANFORD UNIVERSITY SCHOOL OF MEDICINE, 2012) são auto-documentáveis, contemplando estrutura, termos, relações, regras, anotações diversificadas, possibilidade de visualização gráfica, mecanismos de raciocínio, verificação de integridade etc.

#### 7.5.2.7 Fase 7: Avaliação

Considerando as orientações do item 3.5.3, a estrutura é orientada aos processos biológicos da informação e à teoria da informação molecular.

Sua expressividade se verifica na identificação dos principais termos da informação na Biologia.

Sua granularidade é de até cinco níveis, dentro de cada classe principal. Trata-se de uma ontologia descritiva.

A ontologia permite raciocínio automático e é útil para a web semântica e modelagem de processos e sistemas de informação, podendo ser integrada a outras ontologias de domínio.

### 7.5.3 Mesclagem das Ontologias

Temos até aqui, então, duas ontologias construídas com base em fontes de informação distintas, uma na categoria de referências e outra baseada em livros-texto de sistemas biológicos da informação.

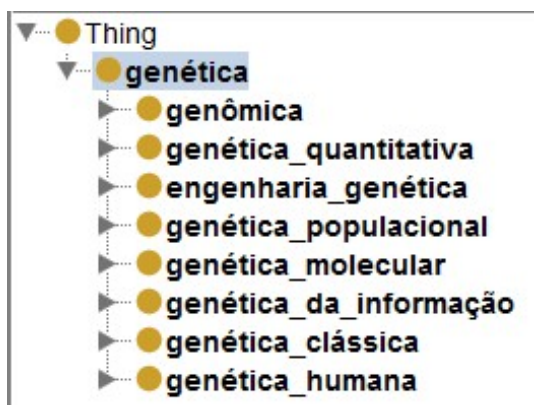


Figura 31: Primeiro Nível das Classes Resultantes da Mesclagem das Ontologias - Fonte: Produzido pelo autor

Conforme vimos no tópico 3.5.4, por questões convencionais, realizamos a mesclagem das ontologias construídas, em que os conceitos das ontologias originais foram selecionadas visando consistência.

Euzenat (2007) propõe solução de alinhamento de ontologias diante da heterogeneidade semântica em um mesmo domínio. A mesclagem das duas ontologias ora construídas é baseada nas correspondências semânticas e de suas entidades relacionadas. Tais correspondências amparam-se nas relações de equivalência, consequência e assunções sob o ponto de vista da mineração textual realizada nas diferentes categorias de fontes de informação (enciclopédias e dicionários x livros-texto no domínio de discurso).

O primeiro nível da ontologia resultante é apresentado na figura 31.

Os demais diagramas de classes da ontologia resultante são apresentados no Apêndice B

Os termos com o radical 'inform', por exemplo, resultante desta mesclagem são apresentados no diagrama da figura 32, pág. 124.

A tabela 3 da pág. 125 apresenta uma analogia conceitual dos entes da Biologia com a Arquitetura Genética da Informação e a tabela 4 apresenta uma correlação conceitual de alguns termos genômicos da Biologia com a Arquitetura Genética da Informação.

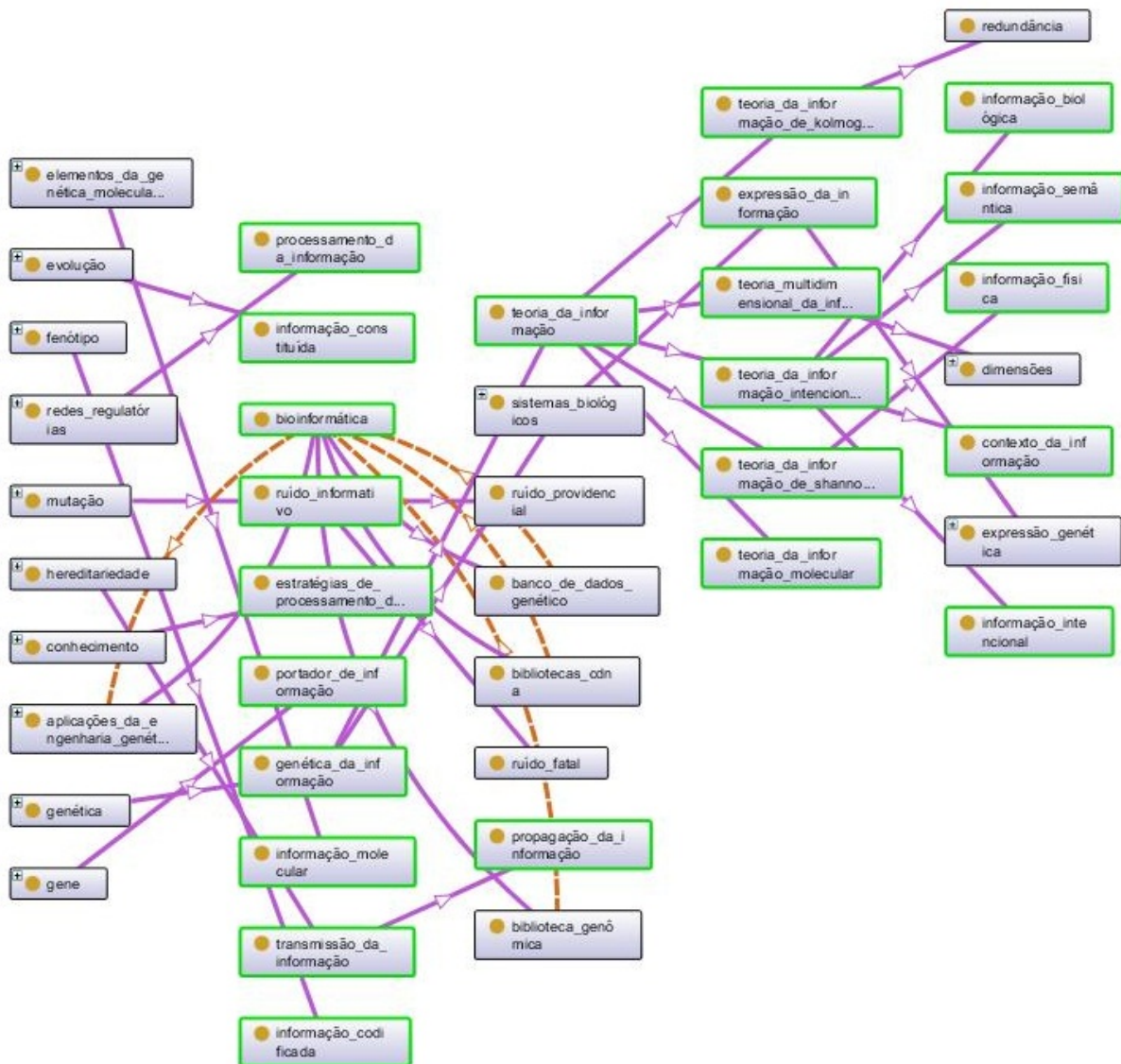


Figura 32: Radical 'inform' Resultante da Mesclagem das Ontologias - Fonte: Produzido pelo autor

|           | <i>Biologia</i>   | Arquitetura Genética da Informação   |
|-----------|---|--|
| DNA       | Programa que contém as instruções genéticas usadas no desenvolvimento e funcionamento de todos os organismos vivos.                               | Programa que contém a informação intencional, cujo padrão de matéria organizada ou energia é detectada por um receptor animado ou manufaturado, desencadeando nele uma mudança comportamental, funcional, ou estrutural. |
| RNA       | Ácido nucleico que desempenha um papel importante na transferência de informação a partir do DNA para o sistema de síntese de proteína da célula. | Mecanismo de transferência de informação para a caracterização de aspectos específicos de um objeto.   |
| Gene      | Unidade física fundamental e funcional da hereditariedade.  | Unidade física fundamental e funcional de informação.  |
| Indivíduo | Instância única e separada de um organismo vivo distinto dos outros de uma mesma espécie.   | Objeto único e separado dentre outros da mesma classe.   |
| Organismo | Coisa viva que pode reagir a estímulos, reproduzir-se, crescer e manter o equilíbrio interno diante de variações ambientais.                      | Classe homogênea de objetos passível de interrelacionamentos e descrição baseada em suas comunalidades, cujas características essenciais são preservadas diante de mudanças externas.                                    |
| População | Grupo de organismos de uma espécie que se cruzam e vivem em um mesmo lugar, ao mesmo tempo.   | Grupo de objetos de uma classe comum e que participam de um mesmo contexto espaço-temporal.  |

Tabela 3: Tabela de Correlação Conceitual de Entes da Biologia e da Arquitetura Genética da Informação

- Fonte: Produzido pelo autor

|          | <i>Biologia</i>  | Arquitetura Genética da Informação   |
|----------|--|--|
| Genética | É o estudo dos genes e de seus papéis na hereditariedade, i.e., o meio pelo qual certos traços são passados entre gerações.  | É o estudo das características da informação e de suas possíveis combinações na produção de novos traços de informação.  |
| Genômica | É o estudo de todos os genes de um organismo, incluindo as interações destes genes com cada um dos outros e com o ambiente onde vive o organismo.  | É o estudo de todas as características da informação de uma classe de objetos, incluindo suas interações, em determinado contexto.                                       |
| Genoma   | É a totalidade de informação genética que possui um organismo.   | É a totalidade de informação que caracteriza uma classe de objetos, diferenciando-a das demais.  |
| Genótipo | São as informações hereditárias de um indivíduo contidas em seu genoma, i.e., o conjunto de genes de um indivíduo.   | São as informações perceptíveis e transmissíveis de um objeto contidas na totalidade das características da informação em sua classe.                                    |
| Fenótipo | É o conjunto de características observáveis ou caracteres de um indivíduo, como a sua morfologia, desenvolvimento, propriedades bioquímicas ou fisiológicas e comportamento, sob a influência de fatores ambientais. | É o conjunto de características observáveis ou caracteres de um objeto, como suas propriedades primárias e secundárias, além de seu comportamento dentro de um contexto. |

Tabela 4: Tabela de Correlação Conceitual entre Termos Genômicos e a Arquitetura Genética da Informação - Fonte: Produzido pelo autor

# 8 Respostas a Problemas Abertos em Filosofia da Informação

*“Nós não podemos resolver problemas usando o mesmo tipo de pensamento que tivemos quando os criamos.”*

— Albert Einstein

Os problemas abertos em Filosofia da Informação foram levantados por [Floridi \(2007\)](#).

Esta tese propõe explicações aos Problemas nº 1, 2, 4, 5 e 6, a saber: (1) O que é informação? (2) Qual é a dinâmica da Informação? (4) De que maneira os dados podem adquirir significado? (5) De que maneira os dados significativos podem adquirir valores verdadeiros? e (6) A informação pode explicar a verdade?

## 8.1 P.1 - O Problema Elementar: O que é Informação?

[Floridi \(2007\)](#) afirma ser este “o problema central e mais difícil da Filosofia da Informação. . . Sabemos que a informação deve ser quantificável, aditiva, armazenável e transmissível.

Mas, além disso, parece que nós ainda não temos uma ideia muito clara sobre a sua natureza específica”. Na visão de Floridi, a informação pode ter três abordagens: informação como a própria realidade, ecológica e quantificável (como os padrões de sinais físicos, por exemplo); informação sobre a realidade, semântica e qualificável; ou informação para a realidade, instrucional (como a informação genética, por exemplo).

A Arquitetura Genética da Informação considera a “informação” em dois aspectos: objetivo (informação absoluta), correspondendo à informação como realidade na visão de Floridi e subjetivo (*informação intencional*), correspondendo à informação sobre e para a realidade na visão de Floridi.

No aspecto objetivo, a “informação” é absoluta e essencial, identificada pelas propriedades ontológicas e invariantes do objeto, o qual, por sua vez, participa de um contexto do mundo.

A informação é, então, anterior a qualquer fenômeno de observação ou percepção e caracteriza o genoma do objeto, conforme apresentamos no diagrama da [fig. 13](#) e explicamos na [seção 4.1](#).

No aspecto subjetivo a *informação intencional* se manifesta dinamicamente, o que responde o Problema 2 ([seção 8.2](#)).

## 8.2 P.2 - O Problema de Entrada/Saída: Qual é a dinâmica da Informação?

Floridi (2007) ressalta que este é um problema “diz respeito aos processos da informação em si mesmos”, ao fluxo da informação em termos semióticos, hermenêuticos e situacionais.

Em resposta, na Arquitetura Genética da Informação a dinâmica da *informação intencional* tem início no fenômeno de percepção. É dependente de um sujeito consciente ou de uma entidade capaz de realizar direta ou indiretamente o sensoriamento, armazenamento e processamento de dados, no caso, os dados sensoriais imanentes do objeto que participa do fenômeno. Esses dados caracterizam o genótipo do objeto, conforme explicado na seção 4.1.

Além do aspecto de imanência (acessível), existe sempre o aspecto transcendental (inacessível), explicados nos tópicos 2.4.3 e 2.4.4, seção 4.1 e representados dinamicamente na Rede de Petri apresentada na figura 87, seção 5.1, impossibilitando que as propriedades do objeto sejam percebidas em sua totalidade. Uma vez percebidos os dados sensoriais, estes passam a constituir o conhecimento no sujeito, ocorrendo a gênese (ativa ou passiva) da *informação intencional*, ao emergir o sentido e o significado na consciência do sujeito ou em um processo na entidade de observação.

## 8.3 P.4 - O Problema Fundamental dos Dados: De que maneira os dados podem adquirir significado?

Este é um problema de “significado intrínseco ou intencionalidade”, com referência em [Searle \(1984\)](#). Conforme explicado na seção 8.2, a aquisição de significado dos dados ocorre no fenômeno de intencionalidade de percepção do objeto, na gênese (ativa ou passiva) da *informação intencional*, caracterizando o fenótipo do objeto em um processo que pode ocorrer voluntariamente (e idealmente) em uma estrutura de redução fenomenológica, em sua segunda etapa, na noese, conforme explicado na seção 5.2 e modelado dinamicamente com Redes de Petri na seção 5.3.

Assim, o significado apresentável do objeto – o seu fenótipo – depende da essência completa do objeto – o seu genoma – e do que foi transferido no processo de percepção dos dados sensoriais – parte ‘herdada’ disponível no seu genótipo.

## 8.4 P.5 - O Problema da Aletização: De que maneira os dados significativos podem adquirir valores verdadeiros?

Floridi (2007) questiona se “a verdade e o significado seriam compreensíveis com base em uma abordagem da informação, ou, pelo contrário, deveria ser a informação analisada em termos de outras teorias do significado e da verdade?”



Uma vez que “aletização” refere-se às várias modalidades da verdade, como: possibilidade, necessidade, ou impossibilidade, na Arquitetura Genética da Informação, a correspondência dos dados significativos (*informação intencional*) com a verdade depende do conceito de verdade (tema explorado na resposta ao problema P.6 (seção 8.5)). De qualquer forma, enfatizamos aqui que a apreensão dos dados do objeto é sempre limitada e parcial, conforme resposta apresentada ao problema P.1 (seção 8.1).

## 8.5 P.6 - Teoria da Verdade Informativa: A informação pode explicar a verdade?

A nossa resposta a esta pergunta limita-se às teorias da verdade como apresentadas por Glanzberg (2009) na Enciclopédia de Filosofia de Stanford, que resumidamente são as seguintes:

- Teoria da Correspondência: “Uma crença é verdadeira se existe uma entidade apropriada, fato para o qual ela corresponde”;
- Teoria da Coerência: “Uma crença é verdadeira se, e somente se, é parte de um sistema coerente de crenças”;
- Teoria Pragmática: “A verdade é o que satisfaz uma crença”;
- Teoria Semântica (de Tarski): Uma asserção de uma sentença é verdadeira se, e somente se, uma interpretação formal da linguagem utilizada é verdadeira”;
- Teoria do Desvelamento: “A verdade está sempre em construção”.

A resposta da Arquitetura Genética da Informação para esta pergunta é que a informação pode explicar a verdade na Teoria da Correspondência, se a informação diz respeito a uma entidade apropriada (objeto); na Teoria da Coerência, se a informação é coerente com o sistema de crenças a respeito do objeto; na Teoria Pragmática, se a crença a respeito do objeto é satisfeita pela informação; na Teoria Semântica se a sentença que pronuncia a informação a respeito do objeto tem sua interpretação no próprio objeto; e na Teoria do Desvelamento, a verdade não pode ser 100% alcançada pela informação, pois está sempre em um processo dinâmico de construção e em conformidade com as razões apresentadas na resposta ao P.5 (seção 8.4), em que as propriedades do objeto só podem ser conhecidas parcialmente.



**Parte IV**

**Aplicações**



# 9 Aplicações da Genética da Informação

*“Música é como soam os sentimentos.”*

— Autor desconhecido

Apresentamos nas seções a seguir exemplos de aplicações da Arquitetura Genética da Informação.

São três aplicações de similaridade fenotípica: descoberta de literatura conexas; similaridade entre imagens de cães e seus donos; e similaridade do canto dos pássaros. Duas aplicações na área de Música: evolução de gêneros musicais; e o Projeto Genoma Musical. Uma aplicação textual: análise de estrutura retórica de documentos. Uma aplicação de algoritmos genéticos modelados por Redes de Petri. Uma aplicação relacionada a indicadores de desenvolvimento do Brasil, o Índice DNA Brasil. E uma aplicação de evolução tecnológica.

## 9.1 Aplicações de Similaridade Fenotípica

Nos itens a seguir, utilizaremos algoritmos de similaridade métrica para a comparação de similaridade fenotípica (explicada na seção 6.6) entre objetos com conteúdo textual e entre objetos com conteúdo imagético. No primeiro caso, a distância de similaridade é calculada baseando-se no fenótipo de relevância de termos nos documentos e no corpus. No segundo caso, a distância de similaridade é calculada com base no fenótipo da cor e textura das imagens.

### 9.1.1 Descoberta Fenotípica de Literatura Conexas

[Kostoff et al. \(2009\)](#) apresentam o conceito de Descoberta de Literatura Conexas (*Literature-Related Discovery - LRD*) afirmando ser a ligação de dois ou mais conceitos na literatura que até então não haviam sido conectados, de modo a produzir conhecimento novo, interessante, plausível e inteligível.

Ele afirma que a “LRD é uma abordagem sistemática para conectar literatura desconexas com base em procedimentos de mineração textual, permitindo a hipótese de descoberta potencialmente radical pelo uso de literatura técnica em si, ou da literatura e seus autores.”

[Aslam e Frost \(2003\)](#) argumentam que a Teoria da Informação pode ser útil para a busca de paridade entre documentos, visando agrupamentos e recuperação da informação, verificando quanta informação dois objetos têm em comum e quanto têm de diferença.

Implementações em bancos de dados têm sido realizadas para suportar de forma nativa a busca por similaridade.

Segundo Böhm e Krebs (2004), a primitiva “*similarity join*” (união por similaridade) combina dois conjuntos de objetos complexos cujo resultado contém todos os pares de objetos similares, com suas faixas de distância métrica (definidas previamente pelo usuário). Eles utilizaram vários algoritmos de Descoberta do Conhecimento em Bancos de Dados (KDD), como agrupamentos por k-means, k-medoid, classificação pelo vizinho mais próximo etc, implementando estruturas de indexação visando melhorar o desempenho destes algoritmos.

Aplicando o conceito de LRD, formulemos a seguinte questão: “Qual é o índice de similaridade entre os livros de determinada estante de uma biblioteca?”.

Para responder a esta questão, podemos adotar critérios variados, como: autoria, contemporaneidade, assunto, editora, ou até mesmo a cor da capa, etc. Adotando-se o critério fenotípico de termos relevantes, por exemplo, exigirá uma indexação de todos os termos presentes nos livros selecionados e cálculo do número de ocorrência destes nos livros, individualmente, e no corpus.

Com o intuito de responder a questões desse tipo, desenvolvemos algoritmo de similaridade métrica no RapidMiner e utilizamos como fonte de informação alguns livros (*e-book*) para realizar o cálculo das distâncias de similaridade. A propósito, alguns destes livros serviram de base para a construção das ontologias. O resultado é apresentado no diagrama da figura 34. O critério adotado foi com relação aos termos relevantes identificados em cada livro e a comparação foi realizada dois-a-dois, entre todos eles.

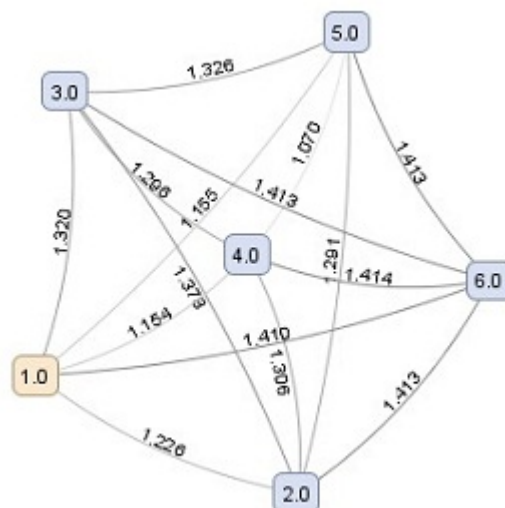
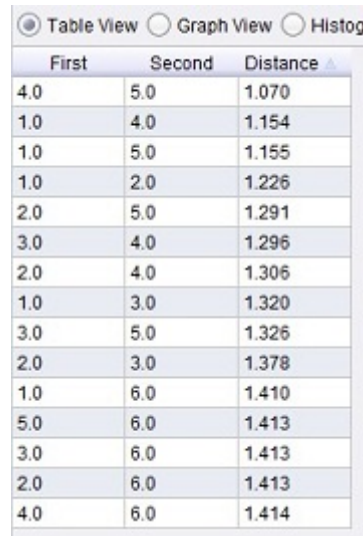


Figura 33: Grafo de Similaridade Fenotípica entre os Livros-Base para a Construção das Ontologias -  
Fonte: Produzido pelo autor

Na figura 34 podemos verificar que o *ranking* de similaridade é dado pelas menores distâncias encontradas, i.e., os três livros mais similares são: 4.0 com o 5.0; 1.0 com 4.0; e 1.0 com 5.0.



| First | Second | Distance |
|-------|--------|----------|
| 4.0   | 5.0    | 1.070    |
| 1.0   | 4.0    | 1.154    |
| 1.0   | 5.0    | 1.155    |
| 1.0   | 2.0    | 1.226    |
| 2.0   | 5.0    | 1.291    |
| 3.0   | 4.0    | 1.296    |
| 2.0   | 4.0    | 1.306    |
| 1.0   | 3.0    | 1.320    |
| 3.0   | 5.0    | 1.326    |
| 2.0   | 3.0    | 1.378    |
| 1.0   | 6.0    | 1.410    |
| 5.0   | 6.0    | 1.413    |
| 3.0   | 6.0    | 1.413    |
| 2.0   | 6.0    | 1.413    |
| 4.0   | 6.0    | 1.414    |

Figura 34: Tabela de Distância de Similaridade entre os Livros-Base para a Construção das Ontologias -  
Fonte: Produzido pelo autor

Outra observação diz respeito aos livros que têm maior similaridade com os demais, que são respectivamente o 5.0 (com 4 ocorrências) e o 4.0 (com 3 ocorrências), entre as dez menores distâncias verificadas. Ainda, fica claro que o livro com a maior distância de similaridade (o mais distinto) entre os demais deste grupo é o 6.0, ocupando as últimas posições na lista.

Os nós de distância de similaridade seguem a seguinte legenda, que correspondem aos livros-fonte: (1.0) [Sawai \(2011\)](#); (2.0) [Niiranen e Ribeiro \(2011\)](#); (3.0) [Nill \(2002\)](#); (4.0) [Bajic e Wee \(2005\)](#); (5.0) [Thompson e Chadwick \(1997\)](#); e (6.0) [Nalewalski \(2006\)](#).

### 9.1.2 Similaridade Fenotípica entre Imagens

[Feng, Siu e Zhang \(2003\)](#) explicam que a técnica de Recuperação de Imagens Baseada em Conteúdo (*Content-Based Image Retrieval - CBIR*) utiliza as propriedades visuais para a recuperação de imagens em grandes bases de dados.

Para isso, o sistema transforma as características fenotípicas de uma imagem, como cor, forma, textura etc em dados numéricos e/ou textuais, em uma representação semântica da imagem.

A extração de um atributo dessa representação semântica é realizada por um descritor. Podemos ter, por exemplo, um descritor fenotípico de extração do histograma das cores HSV da imagem: o *Scalable Color Descriptor - SCD*; extração da distribuição espacial da cor na imagem: *Color Layout Descriptor - CLD*; e a extração da distribuição espacial das bordas da imagem: *Edge Histogram Descriptor - EHD*. A cor é a qualidade fenotípica mais básica dos conteúdos visuais. Por isso, é possível a utilização de cores como um descritor representativo da semântica de uma imagem.

O algoritmo que utilizamos nesta pesquisa para a similaridade fenotípica de imagens baseia-se no CLD. Segundo [Burget et al. \(2010\)](#), o processo de extração de características consiste de

quatro etapas: (1) particionamento da imagem em uma matriz de 64 quadros (8x8); (2) seleção representativa de cores; (3) transformação com quantização; e (4) varredura zig-zag dos quadros (conforme ilustra a fig. 35).

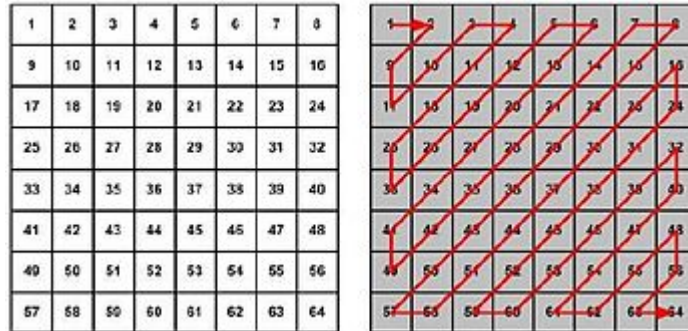


Figura 35: Varredura Zig-Zag de uma Imagem - Fonte: [Symes \(2001\)](#)

Consideremos questões do tipo “Qual é a pessoa que mais se parece com este cachorro?”. Vejamos as imagens da figura 36, obtidas na internet, em [TanMonkey \(2012\)](#).



Figura 36: Cães que se parecem com os seus donos - Fonte: ([TANMONKEY, 2012](#))

Cada par dono-cachorro foi separado em duas imagens, com a seguinte nomenclatura de

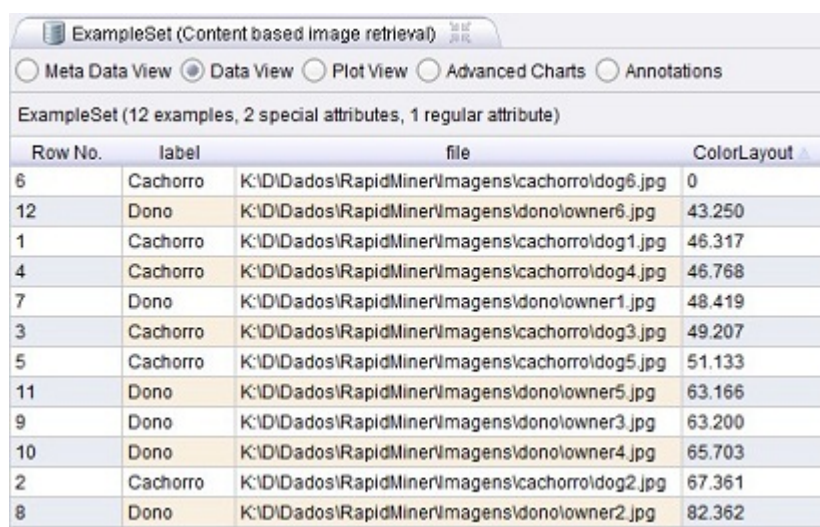


arquivo: owner1.jpg é o dono do cão dog1.jpg; owner2.jpg é o dono do cão dog2.jpg, e assim por diante. As imagens foram armazenadas em pastas diferentes para a aplicação do algoritmo de similaridade métrica.

Com o intuito de responder a questões desse tipo, desenvolvemos rotina de similaridade métrica no RapidMiner e utilizamos como fonte de informação as duas pastas criadas com as imagens dos cães e de seus donos.

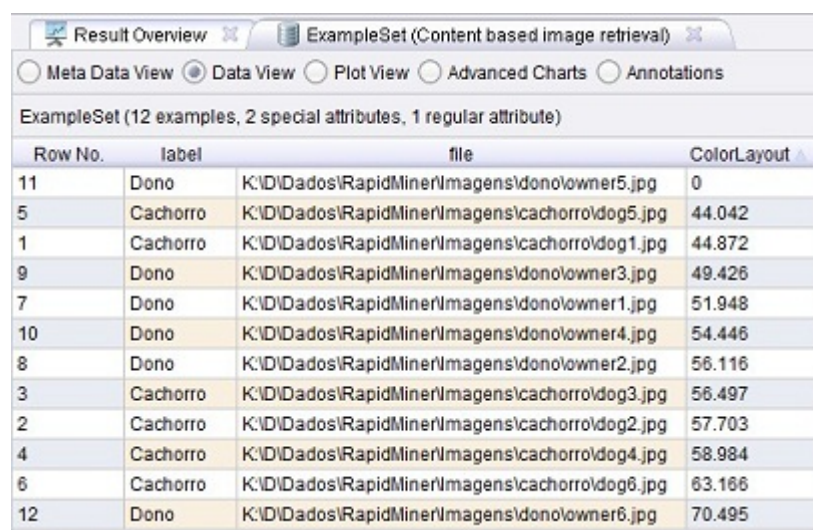
*Pergunta 1: “Nas duas pastas (de cães e donos) quem mais se parece com o cão nº 6?”  
Apresente um ranking de similaridade métrica.*

O resultado, baseado no critério de similaridade CLD é apresentado no diagrama da figura 37.



| Row No. | label    | file   | ColorLayout |
|---------|----------|--|-------------|
| 6       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog6.jpg | 0           |
| 12      | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner6.jpg   | 43.250      |
| 1       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog1.jpg | 46.317      |
| 4       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog4.jpg | 46.768      |
| 7       | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner1.jpg   | 48.419      |
| 3       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog3.jpg | 49.207      |
| 5       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog5.jpg | 51.133      |
| 11      | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner5.jpg   | 63.166      |
| 9       | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner3.jpg   | 63.200      |
| 10      | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner4.jpg   | 65.703      |
| 2       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog2.jpg | 67.361      |
| 8       | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner2.jpg   | 82.362      |

Figura 37: Resultado da Busca por Similaridade de Imagens - Pergunta 1 - Fonte: Produzido pelo autor



| Row No. | label    | file   | ColorLayout |
|---------|----------|--|-------------|
| 11      | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner5.jpg   | 0           |
| 5       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog5.jpg | 44.042      |
| 1       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog1.jpg | 44.872      |
| 9       | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner3.jpg   | 49.426      |
| 7       | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner1.jpg   | 51.948      |
| 10      | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner4.jpg   | 54.446      |
| 8       | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner2.jpg   | 56.116      |
| 3       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog3.jpg | 56.497      |
| 2       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog2.jpg | 57.703      |
| 4       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog4.jpg | 58.984      |
| 6       | Cachorro | K:\ID\Dados\RapidMiner\Imagens\cachorro\dog6.jpg | 63.166      |
| 12      | Dono     | K:\ID\Dados\RapidMiner\Imagens\dono\owner6.jpg   | 70.495      |

Figura 38: Resultado da Busca por Similaridade de Imagens - Pergunta 2 - Fonte: Produzido pelo autor

Vejam que a imagem que mais se parece com dog6.jpg é o próprio dog6.jpg (valor 0 na tabela, ou seja, nenhuma diferença).

Depois, na ordem de similaridade, aparece owner6.jpg (com uma distância métrica de aproximadamente 43 pontos) que de fato corresponde ao caráter fenotípico do dono deste cão, sendo o owner2.jpg o mais distante (com uma distância superior a 82 pontos).

*Pergunta 2: “Nas duas pastas (de cães e donos) quem mais se parece com a jovem (nº 5)?”  
Apresente um ranking de similaridade métrica.*

O resultado, baseado no mesmo critério de similaridade é apresentado no diagrama da figura 38, pág. 137.

Vejamos que a imagem que mais se parece com owner5.jpg é o próprio owner5.jpg (valor 0 na tabela, ou seja, nenhuma diferença). Depois, na ordem de similaridade, aparecem dog5.jpg e dog1.jpg (ambos, com uma distância métrica de aproximadamente 44 pontos) que de fato correspondem ao caráter fenotípico da jovem (cabelos que se parecem com as orelhas daqueles cães, por exemplo), sendo o owner6.jpg o mais distante (com uma distância superior a 70 pontos).

A caracterização do problema é apresentada nas figuras 39, 40, 41 e 42.

**Arquitetura GeneInfo**

Arquitetura Genética da Informação  
\*Obrigatório

**Problema**  
Descrição do problema

Encontrar as similaridades entre as imagens de cães e donos.

Figura 39: Caracterização da Arquitetura Genética da Informação - Parte A - Fonte: Produzido pelo autor

### Elementos da TGA

Teoria Geral da Arquitetura da Informação

**Sujeito**  
Identificação do sujeito

Ser animado  
 Ser manufaturado

**Objeto**  
Descrição do objeto sob observação

imagens de cães e donos

**Informação**  
Propriedades essenciais ou primárias do objeto (aspectos ontológicos)

forma, tamanho, posição, quantidade e movimento

**Dado**  
Propriedades secundárias do objeto (aspectos perceptíveis)

textura e cor

**Conhecimento**  
Conhecimento subjetivo

Reconhecimento do relacionamento de similaridade entre cães e donos, com base nos dados observados

Figura 40: Caracterização da Arquitetura Genética da Informação - Parte B - Fonte: Produzido pelo autor

### Elementos da Arquitetura da Informação Intencional

Disciplina de Arquitetura da Informação

**Manifestação**  
Coleção de dados que o sujeito é capaz de perceber como fenômeno.

textura e cor

**Forma**  
Estrutura de associação dos dados do fenômeno identificando o objeto percebido na experiência do sujeito

emergência de sentido com base nos padrões percebidos

**Contexto**  
Coleção de relações do fenômeno sob o ponto de vista do sujeito

diretório de imagens de cães e diretório de imagens de donos

**Significado**  
Intenção designada do sujeito sobre um objeto.

Cálculo de distância de similaridade métrica entre os objetos pelo algoritmo ColorLayoutDescriptor - CLD

Figura 41: Caracterização da Arquitetura Genética da Informação - Parte C - Fonte: Produzido pelo autor



Figura 42: Caracterização da Arquitetura Genética da Informação - Parte D - Fonte: Produzido pelo autor

### 9.1.3 Similaridade Fenotípica do Canto dos Pássaros

A similaridade fenotípica de áudio também tem sua aplicação na identificação de cantos similares de pássaros.

O genoma dos pássaros inclui corpo (peso, faixa de comprimento, tamanho, cor, partes inferiores, partes superiores, padrão da barriga, padrão das costas, padrão peitoral, cor peitoral); cabeça (cor dos olhos, padrão, cor da coroa, cor da testa, cor da nuca, cor do pescoço e cere); voo (padrão de voo, envergadura, forma da asa, forma da cauda, cauda superior, cauda inferior, cor dos pés); habitat, reprodução e aninhamento; forrageamento e alimentação; alimentos imediatos; **vocalização**; locais de reprodução; tipo de reprodução; cor dos ovos, incubadora dos ovos; material de ninho; migração; e aspectos gerais.

Em nossa aplicação, consideraremos o fenótipo de vocalização, que inclui os cantos e chamadas dos pássaros. “Algumas aves da mesma espécie são tão semelhantes que muitas vezes se torna impossível identificá-las, até mesmo nas mãos de um ornitólogo, com um conjunto preciso de pinças de medição. Mas toda a incerteza desaparece assim que o pássaro começa a cantar” (THE CORNELL LAB OF ORNITHOLOGY, 2012).



Figura 43: Pássaro Northern Cardinal Grosbeak- Fonte: (WAITE, 2012)

O fenótipo da vocalização diz respeito ao ritmo, escala, frequência e repetição. A maioria dos pássaros canta em uma escala característica: os pássaros menores (como o *Waxwing Cedar*) normalmente cantam com vozes mais altas e os pássaros maiores (como o corvo comum), geralmente com vozes mais profundas. Muitos pássaros mudam de escala ao cantarem, produzem sons difusos, doces, assobios ascendentes ou descendentes e se diferenciam na firmeza da vocalização.

Para ilustrar, tomemos o exemplo do pássaro *Northern Cardinal Grosbeak*<sup>1</sup> (fig. 43), que canta uma variedade de claras melodias e borbulhantes assobios, com mais de vinte e cinco diferente sons.

<sup>1</sup>Este pássaro foi escolhido pelo fato de que Messiaen também compôs músicas com o seu canto, em um processo de tradução e transcrição para uma notação musical peculiar ao canto dos pássaros (FALLON, 2007).

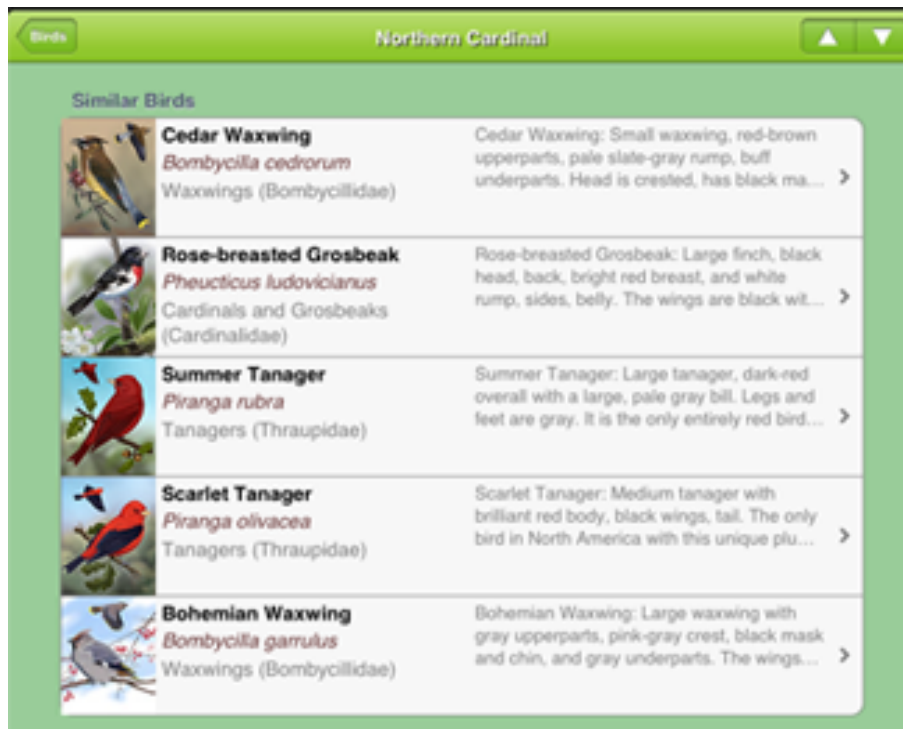


Figura 44: Pássaros Similares ao Northern Cardinal Grosbeak- Fonte: (WAITE, 2012)

As fêmeas cantam duetos com os machos durante o namoro. O chamado é um metálico e abrasivo “chip” ou “pik”. O texto fonético inclui frases do tipo: “whoit cheer, whoit cheer, cheer-cheer-cheer”, “cheer, whoit-whoit-whoit-whoit”, “wheat-wheat-wheat-wheat”, e “bir-dy,bir-dy,bir-dy,bir-dy” (WAITE, 2012).

Em termos gerais, os pássaros fenotipicamente similares ao *Northern Cardinal Grosbeak*, incluindo os traços de corpo, cabeça, voo, habitat, reprodução etc, são o *Cedar Waxwing*, *Rose-breasted Grosbeak*, *Summer Tanager*, *Scarlet Tanager* e o *Bohemian Waxwing*, conforme ilustra a figura 44.

Já no traço fenotípico específico de vocalização, as similaridades do *American Goldfinch* se evidenciam com o *Carolina Wren*, e o *Tufted Titmouse*, conforme apresentado na figura 45.

O canto das aves muitas vezes foi utilizado como material temático na composição musical. Olivier Messiaen, não foi o primeiro a fazer isso, mas fê-lo com dedicação e rigor incomparáveis, integrando similaridades e dissimilaridades em composições para orquestras sinfônicas com resultados belíssimos e intrigantes. Messiaen catalogou a notação musical específica para o canto de milhares de pássaros. O mapeamento espectômetro-notação musical do canto do pássaro Cardinal é apresentado na figura 46 e, representa um processo de transcrição e tradução genética.

Fallon (2007) esclarece que o espectrograma foi produzido com a frequência em Hz/s. Os números abaixo do espectrograma indicam a frequência indicada pela seta. As setas alinham o espectrograma com a transcrição.



Figura 45: Cantos Similares ao do Northern Cardinal Grosbeak - Fonte: (WAITE, 2012)

As frequências foram determinadas prestando-se atenção à audibilidade do sinal e à combinação da duração com a intensidade. As frequências são acompanhadas pelas letras correspondentes das notas musicais. O sinal “+” depois da letra indica que a frequência tem uma microtonalidade maior do que a indicada. O sinal “-” depois da letra indica que a frequência tem uma microtonalidade menor do que a indicada. A preponderância de tons combinados sugere similaridade entre a velocidade das gravações de Messiaen com o equipamento de transferência digital utilizado.

Para Olivier Messiaen as aves representavam a ideia de liberdade e ocupavam o topo de uma hierarquia musical em escala planetária. Embora fosse sensível a muitos outros sons da Natureza (como os do mar e do vento, por exemplo) e admirasse profundamente os compositores que, como Debussy, Wagner ou Berlioz, os tinham invocado na música, Messiaen considerava que o canto das aves era o mais musical, o mais próximo de nós e o mais fácil de reproduzir (FALLON, 2007).

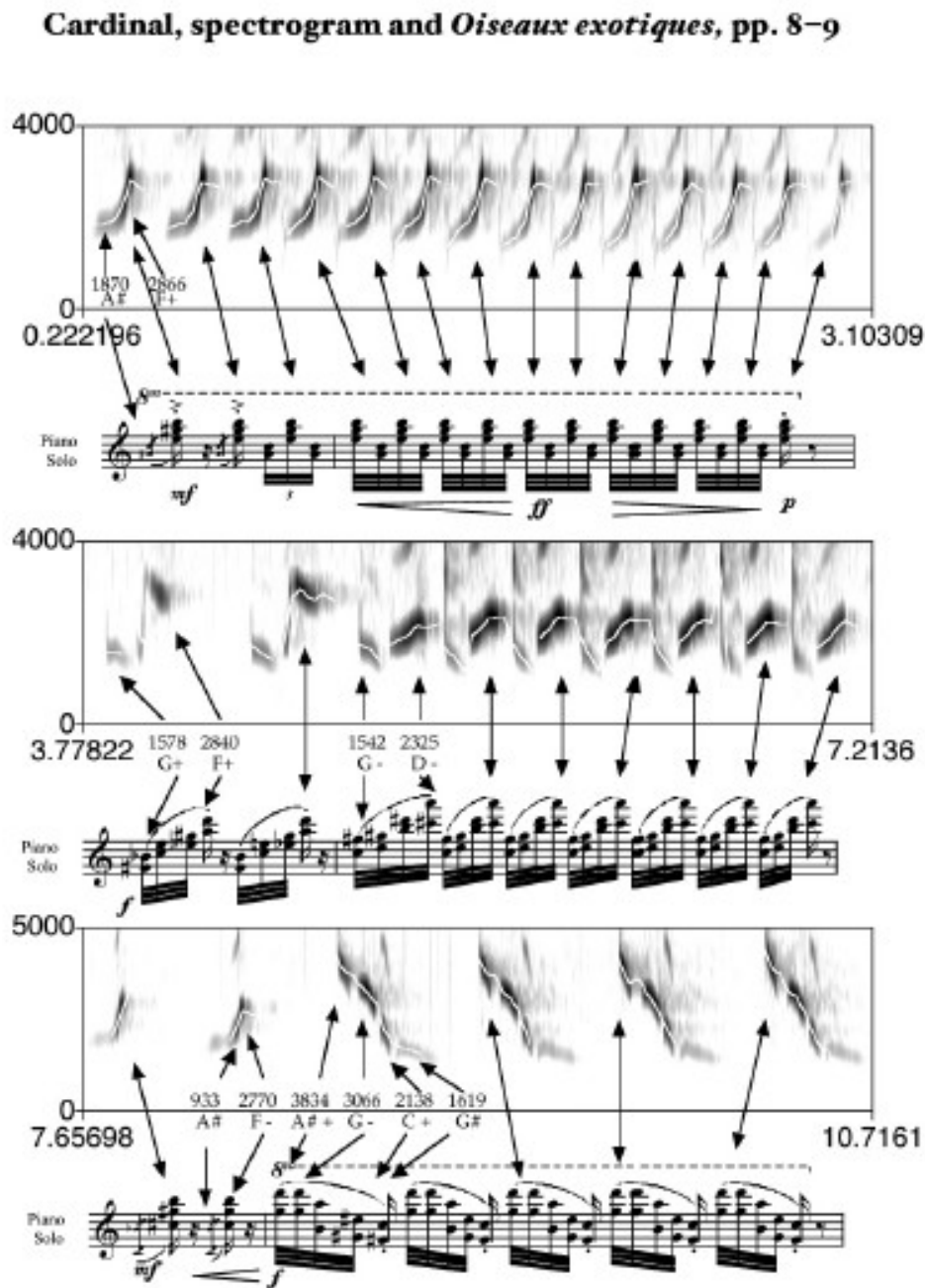


Figura 46: Notação Musical de Messiaen para o Canto do Cardinal Grosbeak- Fonte: (FALLON, 2007)

#### 9.1.4 Similaridade Fenotípica em Redes de Pesquisa

Neves (2010) realizou trabalho de identificação de grupos de pesquisa com base em análise infométrica de autores e co-autores com cadastro na Plataforma Lattes do CNPq, considerando pesquisadores e produção científica (artigos, eventos e orientações).

Ela relata que a metodologia utilizada para mapear os doutores que formam a rede de microeletrônica no Brasil consistiu em selecionar em 2009 uma amostra de doutores cadastrados na Plataforma Lattes por meio das palavras-chave: semi-condutor, chip, cmos, fpga, circuito integrado e microeletrônica; complementar esta amostragem com os registros obtidos em pesquisa



anterior, em 2007, à época, com validação por meio de questionários; limpeza de dados e cruzamento de dados. A abordagem de similaridade métrica adotada que obteve melhor resultado foi a de Levenshtein.

Obtivemos a base de dados completa desse trabalho, que contém 43.459 artigos, 1.540 pesquisadores, 100.769 autores e relação pesquisador-artigo de 42.764 registros.

Ressaltamos que eventos e orientações não foram considerados em nossa pesquisa e somente os artigos publicados no período de 2004 a 2009, para viabilizar uma comparação com o trabalho de [Neves \(2010\)](#). Neste período o número de pesquisadores foi de 1.454.

O Modelo reduzido de Entidades e Relacionamentos (EER) do projeto é apresentado na figura 47, que contempla os pesquisadores e as relações de autoria e co-autoria em artigos científicos.

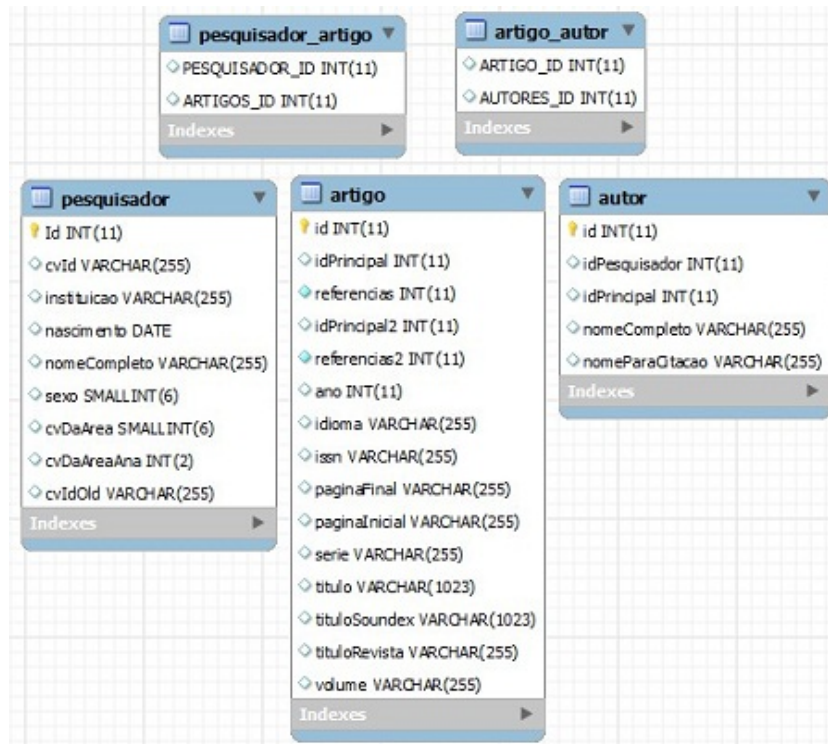


Figura 47: Modelo de Entidades e Relacionamentos (EER) - Lattes - Microeletrônica- Fonte: reduzido de [Neves \(2010\)](#)

Cada pesquisador foi associado à um índice de relevância TF/IDF de todos os *stems* de seus artigos. A similaridade fenotípica foi calculada por análise combinatória de todos os pesquisadores, dois a dois, ou seja:

$$C_n^p = \frac{n!}{p!(n-p)!} \quad (9.1)$$

Onde  $n$  é o número total de itens, i.e., 1.454 pesquisadores e  $p$  é o agrupamento entre eles, i.e., 2.

$$C_{1.454}^2 = \frac{1.454!}{2!(1454 - 2)!} \quad (9.2)$$

$$\frac{(1.454)(1.453)}{2} = 1.056.331 \quad (9.3)$$

Portanto, temos que há mais de um milhão de relações entre todos os 1.454 pesquisadores, comparados dois a dois.

Considerando-se as similaridades maiores ou iguais a 20%, conseguimos produzir uma rede de pesquisadores ampla, totalmente baseada em *stems* relevantes, coerente com a rede de autores e co-autores produzida por Neves (2010). Os resultados foram apresentados em sistema de Business Intelligence que desenvolvemos especificamente para esta aplicação, conforme ilustra a figura 48.

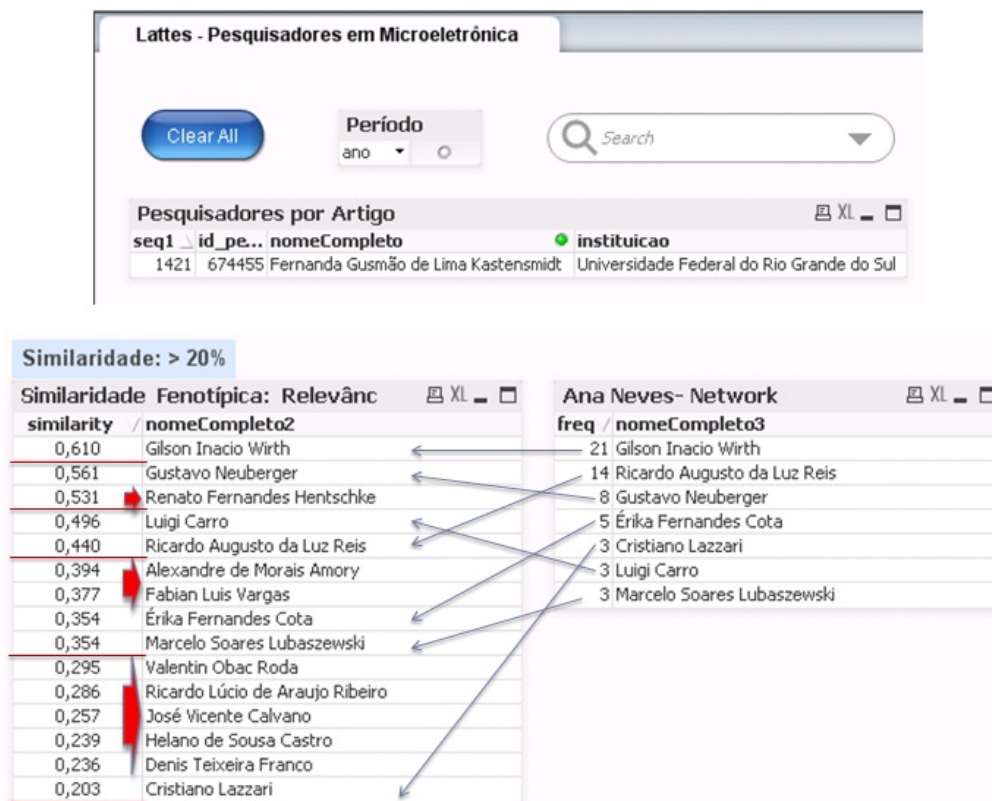


Figura 48: Sistema de Business Intelligence para a Rede de Pesquisadores em Microeletrônica- Fonte: Produzido pelo autor

A rede produzida por Neves, por exemplo, identificou sete pesquisadores com algum vínculo de autoria e co-autoria com a pesquisadora Fernanda Gusmão de Lima Kastensmidt. A rede produzida por similaridade fenotípica de relevância de *stems* encontrou relação de similaridade desta pesquisadora com 15 outros, cobrindo todos os já identificados na rede de Neves, o que de certa forma valida a rede de autoria e co-autoria, ampliando-a qualitativa e substancialmente.

Encontrar uma nova<sup>2</sup> lei de potência característica dessa rede equivale, por exemplo, a aplicar a lei genética Hardy-Weinberg para populações naturais. Dentre algumas mudanças evolutivas populacionais, o equilíbrio de Hardy-Weinberg pode nos ajudar a proceder análises para identificar: ocorrência de mutações; indivíduos que saíram ou entraram na população; ocorrência de acasalamentos não aleatórios e seleção natural; existência de populações pequenas.

## 9.2 Evolução Natural e Artificial de Gêneros Musicais

O termo “gênero” significa uma “classe ou categoria de esforço artístico que possui forma, conteúdo e técnicas particulares” (DICTIONARY.COM, 2012b), com origem no termo biológico “genus”, significando “a maior subdivisão de uma família ou subfamília de organismos, que consiste geralmente em mais de uma espécie” (DICTIONARY.COM, 2012a).

Dentre os diversos elementos que concorrem para a definição dos gêneros musicais podemos destacar:

1. instrumentação (que instrumentos são mais frequentemente usados);
2. texto (conteúdo sacro, profano, romântico, idílico, etc.);
3. função (prelúdio, encerramento, dança, ritual, etc.);
4. estrutura (linear, segmentada, repetitiva, etc.);
5. contextualização (local de interpretação, contextualização geográfica, contextualização cronológica, contextualização etnográfica, etc.)

Na Arquitetura Genética da Informação, cada um destes itens pode ser considerado um estado fenotípico de um genótipo com diversas possibilidades de combinação. A porcentagem dos descendentes (novos estilos dentro do gênero) com um determinado genótipo que manifesta a característica fenotípica pelo menos em algum grau é chamada de penetrância, conforme apresentado na figura 66, no Glossário.

Consideremos, então, o processo de evolução natural do Samba com base na instrumentação.

O Samba é tocado basicamente por instrumentos de percussão e acompanhado por instrumentos de corda. Em vertentes como o samba-exaltação e o samba-de-gafieira, foram acrescentados instrumentos de sopro. Os instrumentos básicos do samba são: cavaquinho, violão, pandeiro, surdo, tamborim, tantã e bandolim. As vertentes do samba incluem os instrumentos: cuíca e trompete.

---

<sup>2</sup>Nas relações de autoria, Neves (2010) encontrou uma lei de potência de quantidade de co-autores por autor.

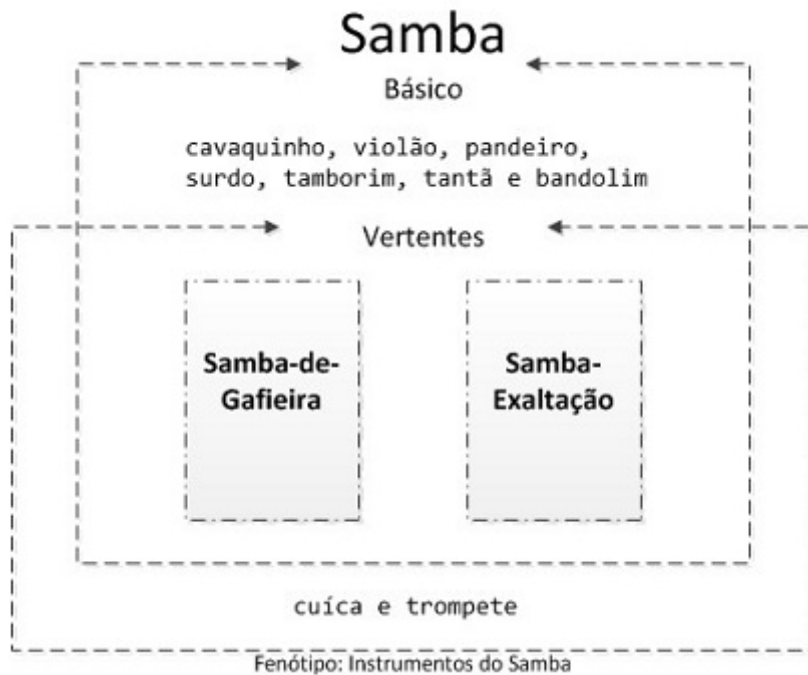


Figura 49: Evolução do Samba - Fenótipo: Instrumentos Musicais - Fonte: Produzido pelo autor

Em uma evolução artificial, podemos considerar a mistura genética de dois gêneros diferentes. Consideremos a regra de complementaridade de Watson-Crick (apresentada na seção 9.5), em que a síntese de proteínas seja uma analogia para a síntese dos elementos individuais de cada gênero musical. A sequência de ligações A-T e C-G representa as inúmeras possibilidades de desenvolvimento de um novo gênero.

Por exemplo, na união do Jazz com o Samba, a Bossa Nova herdará as características fenotípicas dominantes de cada um deles presentes nesta relação. A combinação de características recessivas também poderá acontecer, prevalecendo quaisquer delas. Ainda, a Bossa Nova terá suas próprias características, que a diferenciam dos demais gêneros ascendentes em questão.

Exemplificando, no contexto “*estilo musical*” existe o genoma ótimo que perfeitamente o caracteriza pela estrutura da composição, ritmo, obstinação, raízes, tonalidade, instrumentação, nuança e outros. O novo gênero, da “*Bossa Nova*” é delimitado por fenótipos como ritmo = 2/4; instrumentos = violão de nylon; raízes = “Samba”, “Jazz”; tonalidades = “escalas maiores”; harmonia = “acordes com extensões de sétimos maiores”; dentre outros.

O que podemos apreender aqui é que cada gênero musical tem suas próprias características fenotípicas, algumas das quais herdadas geneticamente de outros gêneros. Ainda, a manipulação genética da informação entre gêneros diferentes pode gerar um novo gênero.

Características fenotípicas incompatíveis inseridas (por mutação) em um gênero musical podem gerar desarmonia, em um sistema de correção de erro ineficiente (seção 9.5).

Apresentamos na figura 50 o Diagrama de Venn com a representação da manipulação genética das informações do Jazz e do Samba, gerando a Bossa Nova.



Figura 50: Manipulação Genética da Informação - Fonte: Produzido pelo autor

Outros estudos musicais também podem ser caracterizados na Arquitetura Genética da Informação, caso da invenção da música Folk.

[Gelbart \(2007\)](#) explica a grande similaridade existente entre as músicas chinesa, escocesa e o antigo grego, supondo que seja uma espécie de música natural para um povo de costumes simples.

### 9.3 O Projeto Genoma Musical

Com o advento de serviços de descoberta de música baseados na web, a arte de encontrar nova música mudou.

Sites da internet de rádio como TagWorld, Last.fm, SoundHound e Pandora permitem ao usuário digitar o nome de uma música ou artista e instantaneamente encontrar outra música de características similares. O Pandora permite também a realização de buscas por áudio, baseado em similaridade de trechos musicais.

[Layton \(2012\)](#) explica que: “O sistema Pandora se baseia em um genoma musical, que consiste de 400 atributos musicais cobrindo as qualidades de melodia, harmonia, ritmo, forma, composição e letras”.

É um projeto que começou em janeiro de 2000 que envolveu 30 especialistas em teoria musical, levando cinco anos para ser concluído. O genoma baseia-se numa análise complexa por seres humanos (em média, cerca de 20 a 30 minutos dedicados para quatro minutos da

música) em uma base de músicas de 10.000 artistas dos últimos 100 anos. A análise de novas músicas continua a cada dia desde o lançamento on-line do Pandora, em agosto de 2005. Em maio de 2006, a biblioteca do genoma musical continha 400.000 músicas<sup>3</sup> analisadas de 20.000 artistas contemporâneos.

O SoundHound, por exemplo, realiza a busca por texto, voz, ou música na entrada de um SmartPhone, processando uma busca por similaridade no banco de dados da empresa, fazendo o 'casamento' das características selecionadas e apresentando o resultado no SmartPhone, conforme apresentado na figura 51



Figura 51: Busca por Similaridade Acústica - Fonte: (SOUNDHOUND, 2012)

## 9.4 Estrutura Fenotípica de Retórica Textual

Segundo Logan (2012), “a informação de Shannon tem sido aceita como a definição canônica de informação” e é desprovida de qualquer subjetividade e semântica. Ele continua dizendo que “parte da resistência à informação de MacKay era porque a sua definição de [informação estruturada] envolvia subjetividade que os cientista ortodoxos não podiam incluir em suas teorias”, bom para os seus objetivos de engenharia.

Com essa finalidade de comunicação subjetiva, analisaremos aqui a aplicabilidade da *Arquitetura Genética da Informação* para a identificação do genoma de um texto produzido com o uso da retórica.

Ele argumenta perguntando: “Qual é o fator causal para a existência da informação? O que nos leva a escrever uma carta, um cartão, uma nota de felicitação, um diário, ou um comentário em um documento? O pré-requisito mais importante é a nossa própria vontade, ou a vontade de um supervisor”.

Asher e Lascarides (1994) explicam que a retórica pode ser definida como a técnica ou arte de convencer o interlocutor através da oratória, i.e., de discurso que pode ser verbal, escrito ou visual.

<sup>3</sup>Os estilos clássico e latino não foram contemplados na primeira etapa do projeto.

Assim, no processo de produção do texto, a organização e apresentação do discurso têm como meta última atingir o público-alvo de forma convincente com o conteúdo apresentado, seja qual for a sua natureza. A Teoria da Estrutura Retórica (RST)<sup>4</sup> é, então, retórica no sentido de que “as relações da estruturação do texto refletem as opções do produtor de organização e apresentação”. O analista tem como pressupostos as possíveis finalidades do produtor textual, as plausíveis suposições do produtor a respeito do leitor (público-alvo) e o conhecimento de padrões proposicionais em relação ao conteúdo do texto.

Aplicando-se o princípio matemático de Pareto, ou Princípio 80-20, podemos deduzir que em um texto, 20% dele representa a riqueza da informação necessária para a sua compreensão geral e os 80% restantes, são apenas complementação.

Há duas abordagens de automatização de sumarização: superficial ou profunda.

A **abordagem superficial** utiliza métodos “cegos”, exclusivamente de relevância, em função da posição e do contexto. Normalmente, esta abordagem apresenta problemas de coesão e coerência. Por sua vez, a **abordagem profunda**, adotada nesta aplicação, utiliza algoritmos de Processamento em Linguagem Natural - PLN para produzir extratos ou abstrações com base em técnicas de interpretação de teorias linguísticas formais.

A RST descreve a estrutura hierárquica do texto a partir de um modelo de relações que se estabelecem recursivamente entre unidades de discurso (espaços mínimos dotados de significado). Na Arquitetura Genética da Informação, tais espaços significativos caracterizam o fenótipo do documento por suas unidades de discurso. Azar (1999) explica que as unidades de discurso em um texto são denominadas núcleos e satélites. Os núcleos têm conteúdo centrado no propósito do texto; os satélites têm conteúdo secundário ou complementar. As unidades de discurso são unidas por relações de discurso, de forma hierárquica e recursiva. São funcionalmente significantes, têm papéis definidos no texto, e contribuem para uma finalidade específica do produtor textual. É importante fazer analogia entre os fenômenos de combinação textual com as regras gramaticais: Sempre ocorrem aos pares; ocorrem entre partes textuais de tamanhos diversificados; são recorrentes; são assimétricos (a parte central é o núcleo e a parte periférica é o satélite); e a relação é formada pelo núcleo + satélite.

---

<sup>4</sup>A Teoria da Estrutura Retórica foi criada por Bill Mann no *Information Sciences Institute* e é mantida por Maite Taboada.

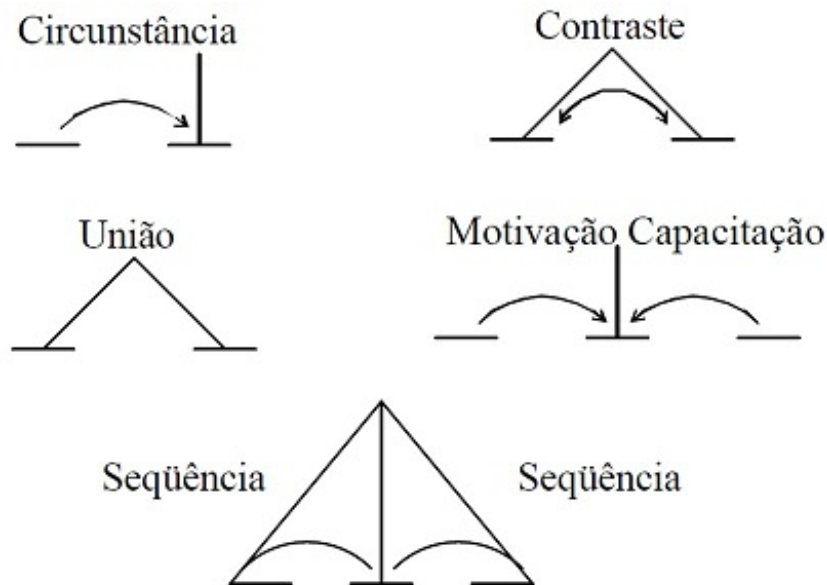


Figura 52: Esquemas RST - Traços Fenotípicos Textuais- Fonte: Produzido pelo autor

A RST é uma teoria organizacional do discurso baseada nos seguintes princípios:

- Princípio da Unidade: As partes do texto contribuem para uma finalidade específica e o texto não sofre descontinuidade (nonsequitur);
- Princípio da Coerência: O texto tem uma estrutura com fenômenos de combinação entre as partes;
- Princípio da Homogeneidade: Há um grupo de padrões estruturais chamados esquemas de relações;
- Princípio da Hierarquia: Partes elementares se organizam recursivamente em partes maiores até chegarem à escala do texto inteiro.

Os esquemas possíveis (fig. 52 para as relações RST, entendidos na Arquitetura Genética da Informação como os traços fenotípicos retóricos de um texto, são: circunstância, contraste, união, motivação/capacitação e sequência. Circunstância é uma relação de causa e efeito, ocorrendo quando o satélite é efeito do núcleo. “Contraste” é uma relação de oposição de ideias, quando núcleo e satélite são contrários entre si. “União” é uma relação de afirmação mútua entre núcleo e satélite. “Motivação/Capacitação” é uma relação onde diversos fatores encontrados nos satélites reforçam a ideia do núcleo. E “Sequência” é uma relação temporal de eventos cronológicos.

A título de exemplo, procuramos identificar no texto do Programa Nacional de Atividades Espaciais - PNAE, da Agência Espacial Brasileira - AEB, os traços fenotípicos retóricos de um sumário executivo (Onde?, Por que?, Quando?, Quanto? e Como?) com base nos possíveis esquemas das relações RST, realizando a segmentação, estruturação e sumarização do texto, etapas apresentadas nas figuras 53, 54 e 55, respectivamente.



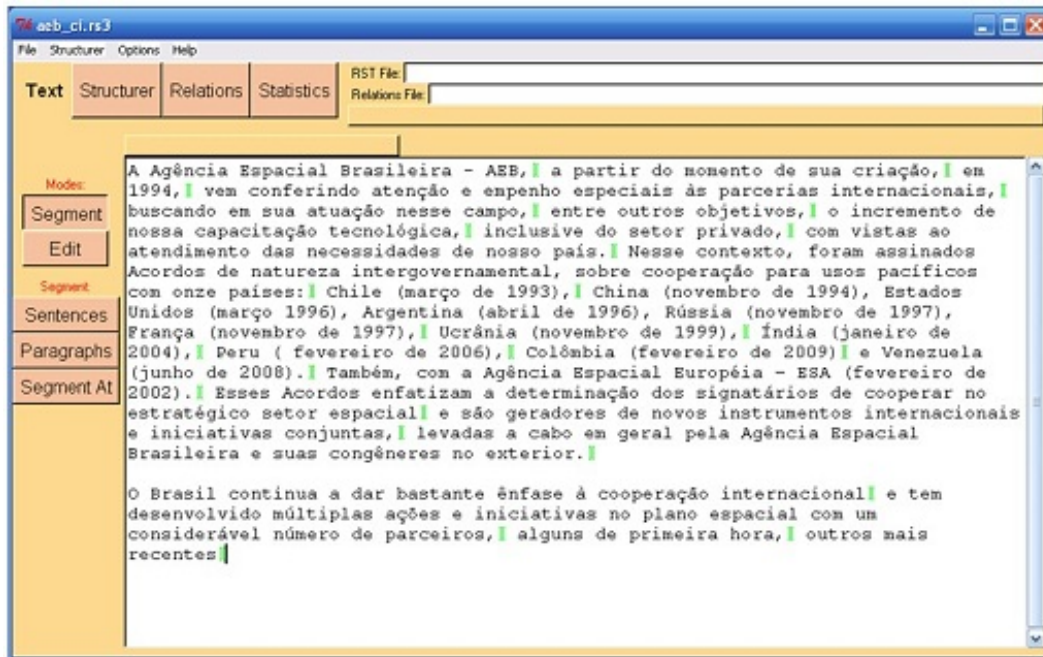


Figura 53: Segmentação - Fonte: Produzido pelo autor

Todos os fenômenos de combinação textual presentes e observados no texto passam a caracterizar o seu fenótipo retórico, podendo ser útil, por exemplo, para identificar entidades, opinião, autoria, relevância etc.

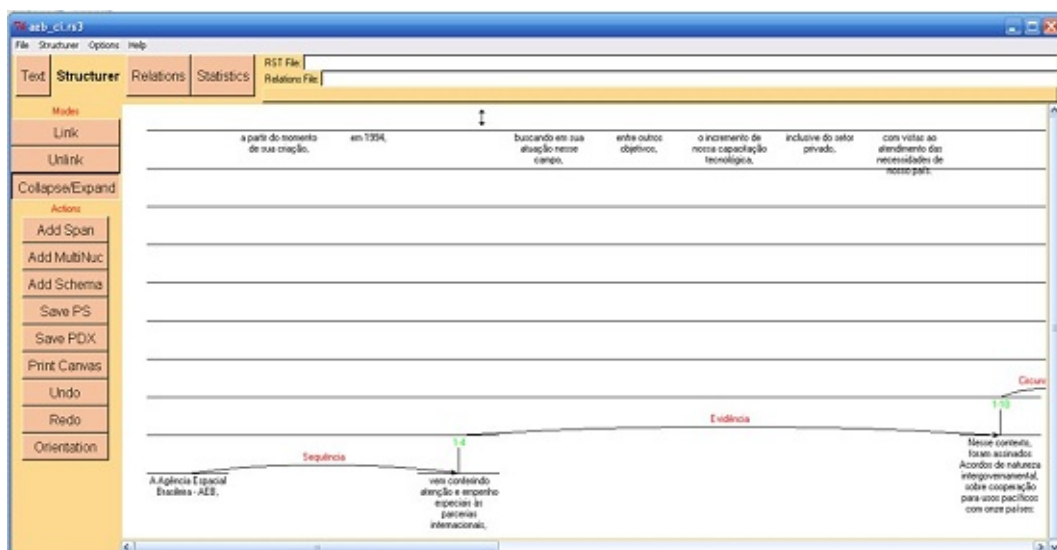


Figura 54: Estruturação - Fonte: Produzido pelo autor

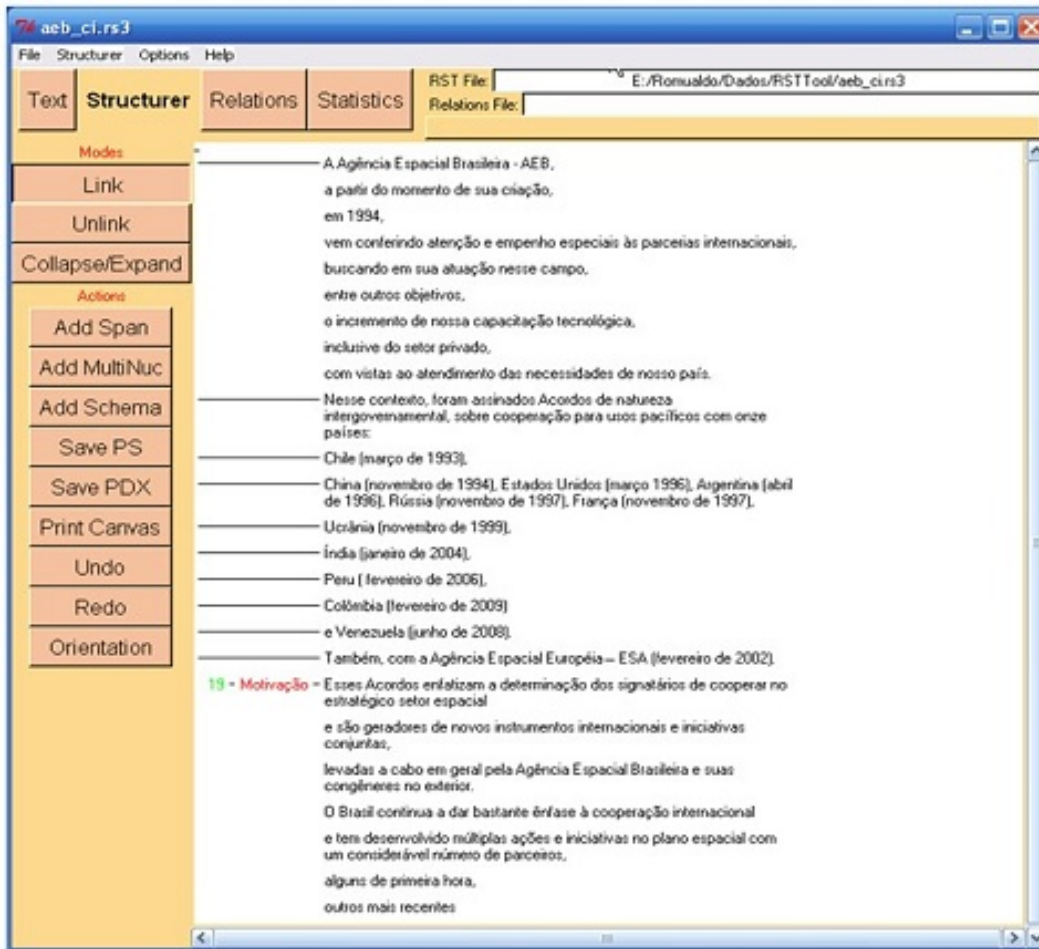


Figura 55: Sumarização - Fonte: Produzido pelo autor

## 9.5 Algoritmos Genéticos Modelados com Redes de Petri

Schuster (2008) aplica, em uma investigação teórica, o paradigma da computação genética à modelagem de redes de Petri.

Ele explica que o potencial desse tipo de computação está em sua capacidade inerente para um vasto paralelismo, no escopo de armazenamento de alta densidade e sua habilidade intrínseca para resolver muitos problemas combinatórios. Simplificando, a computação genética é baseada no *design*, manipulação e processamento de nucleotídeos que podem ser combinados em uma fita simples de DNA ou em uma fita dupla de DNA. A fita simples é gerada através de subsequentes ligações de qualquer um dos quatro tipos de nucleotídeos, e pode ser representada por uma sequência de letras (por exemplo, TATCGGATCGGTATATCCGA). A regra de complementaridade de Watson-Crick diz que uma base A se liga apenas com uma base T, uma base G com uma base C, e vice-versa. Por exemplo, a cadeia ATAGCCTAGCCATATAGGCA é o complemento Watson-Crick da cadeia de DNA mencionada anteriormente.

“Tipicamente, os quatro nucleotídeos A-T e C-G proporcionam a base para um al-

fabeto<sup>5</sup> A partir desse alfabeto, uma linguagem particular pode ser construída. Essa linguagem é utilizada para definir algoritmos e programas de computador. Em termos práticos, um computador de DNA tem semelhança com uma máquina bioquímica que executa algoritmos e programas através da manipulação de cadeias de DNA de uma série cuidadosamente orquestrada de processos bioquímicos. . . Uma vez que esta seqüência de eventos seja executada corretamente, o resultado estará disponível na forma de seqüências de DNA. A fim de extrair o resultado do algoritmo, é necessário ler essas cadeias e decodificar a informação em um processo similar à leitura de uma seqüência do genoma humano (por exemplo, a identificação de um gene que codifica uma proteína)” (SCHUSTER, 2008).

Barjis, Samarrai e Augustin (2006) elaboraram uma modelagem simples de manipulação de cadeias de DNA com Rede de Petri, que apresentamos na figura 56.

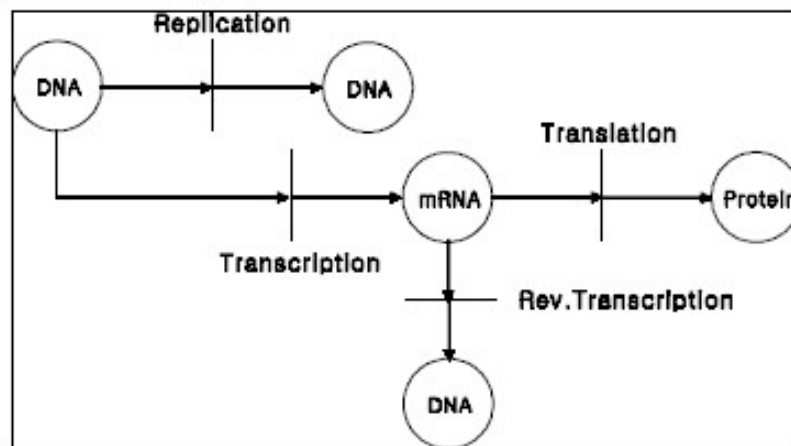


Figura 56: Modelagem do Processo de Produção de Proteínas com Rede de Petri- Fonte: (BARJIS; SAMARRAI; AUGUSTIN, 2006)

No artigo original de Watson e Crick (1953), eles afirmam que “a seqüência de bases em uma cadeia simples não parece ser restrita de forma alguma. Apesar disso, se apenas pares específicos de bases podem ser formados, disso segue que se uma seqüência de bases em uma cadeia é dada então a seqüência na outra cadeia é automaticamente determinada”.

A complementaridade dos dois filamentos torna o DNA unicamente adequado a armazenar e transmitir a informação genética.

<sup>5</sup>O alfabeto do DNA tem 4 letras. As proteínas são formadas por 20 tipos de aminoácidos, cada um codificado por um códon (seqüência de 3 das 4 letras do DNA).

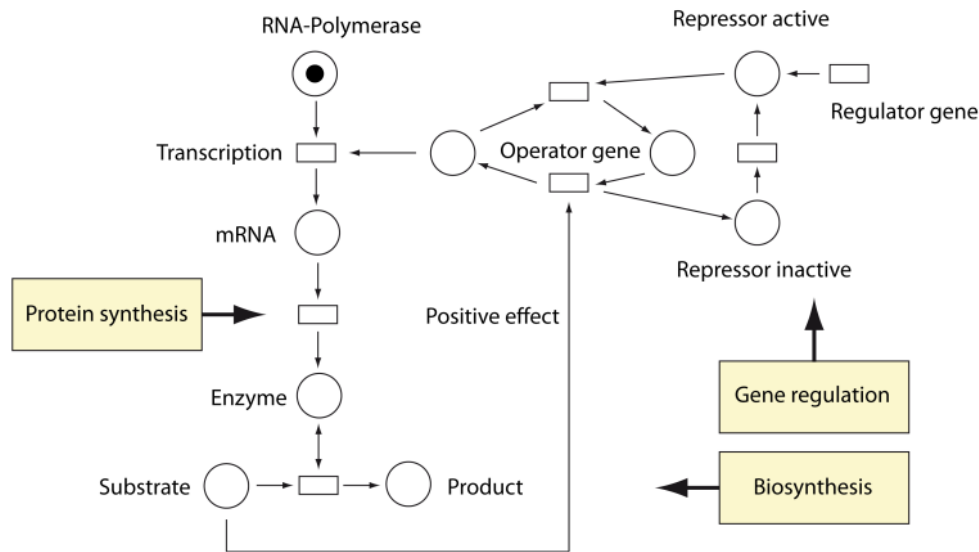


Figura 57: Modelagem do Processo de Síntese de Proteínas, Regulação Genética e Biossíntese com Rede de Petri- Fonte: (BARJIS; SAMARRAI; AUGUSTIN, 2006)

Cada fita de DNA contém uma cópia desta informação (codificada pela sequência de nucleotídeos complementares). Os genes carregam a informação biológica que deve ser precisamente copiada e transmitida durante a divisão celular de maneira a viabilizar a replicação do DNA. A cada divisão, a célula deve copiar o seu genoma para transmiti-lo a ambas as células filhas.

A acurácia da replicação do DNA é crítica para a reprodução celular. Se a polimerase não pudesse corrigir erros haveria a inserção de uma base errada a cada  $10^4$  nucleotídeos incorporados.

Uma atividade extra de revisão de prova (*proof-reading*) reduz a taxa de erro para uma base errada a cada  $10^9$ . E o sistema completo de reparo reduz a taxa de erro para uma base errada a cada  $10^{10}$ .

## 9.6 Índice DNA Brasil

O Núcleo de Estudos de Políticas Públicas (Nepp) da Unicamp propôs em 2002 a inclusão de indicadores de saúde e de educação ao Índice de Desenvolvimento Humano - IDH. Estas duas novas dimensões, além da dimensão econômica do IDH compõem o Índice DNA Brasil, um sistema mais analítico de medidas baseado em um Sistema Múltiplo de Indicadores (SMI).

Biasoto-Jr. et al. (2006) explicam que esse índice é útil não apenas para a comparação com outros países, mas “para balizar o estágio em que nos encontramos em relação ao que a sociedade brasileira possa estabelecer como resultado de um projeto ético-político nacional para um horizonte temporal”. O índice principal constitui-se de uma representação gráfica conjunta da situação dos indicadores que o compõem, em que os valores em cada eixo representam o estágio em que o indicador se encontra em relação ao desejado ou projetado como possível num horizonte de tempo, ou ainda, como comparação com o valor do indicador em outro país.

O índice é composto por sete dimensões: bem-estar; competitividade econômica; condições sócio-ambientais; educação; saúde; proteção social; e coesão social, conforme apresenta a figura 58.

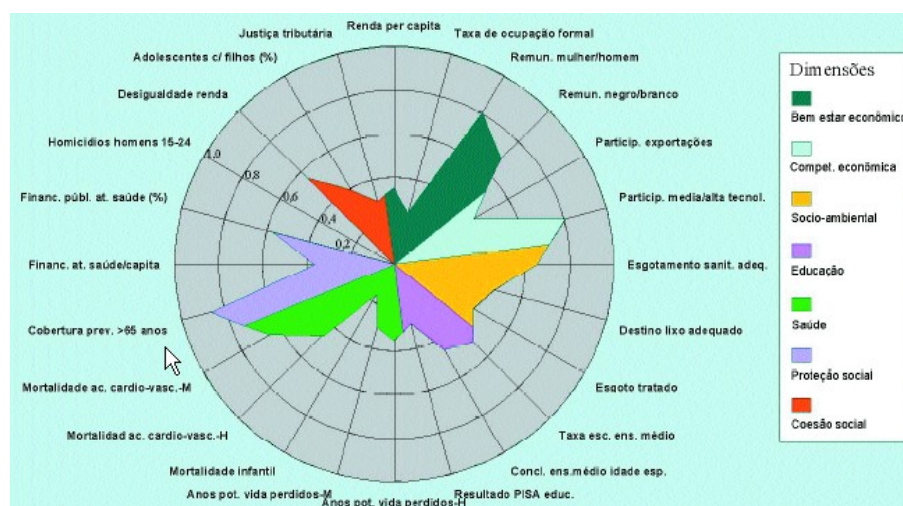


Figura 58: Índice DNA Brasil - Fonte: (BIASOTO-JR. et al., 2006)

Genética Quantitativa é a parte da genética que estuda os caracteres quantitativos dos organismos (HOLSINGER, 2012).

Na Arquitetura Genética da Informação, os indicadores representam os aspectos quantitativos de herança, nível de estudo e variações/efeitos do meio. No aspecto de herança, esta é poligênica, em que, em geral, os caracteres quantitativos são regulados por vários genes, no caso, as sete dimensões do índice. As características quantitativas são estudadas no nível de população, no caso, a população brasileira, e são descritas através de parâmetros tais como média, variância e covariância, exibindo variações contínuas (às vezes descontínuas) e parcialmente de origem não-genética, ou seja, são fortemente afetadas pelo ambiente, no caso, os fatores econômicos, políticos, sociais e éticos nacionais e internacionais.

A seguir apresentamos os indicadores do Índice DNA Brasil:

Indicadores de bem-estar:

1. Renda per capita
2. Relação entre as remunerações médias das mulheres e dos homens;
3. Relação entre as remunerações médias de negros e brancos;
4. Taxa de ocupação formal.

Indicadores de competitividade econômica:

1. Participação do Brasil no mercado externo;
2. Produtos com médio e alto valor agregado na pauta de exportações.

Indicadores de condições sócio-ambientais:

1. Instalações adequadas de esgotamento sanitário;
2. Destinação adequada do lixo urbano;
3. Tratamento do esgoto sanitário;

Indicadores de educação:

1. Taxa de escolarização líquida no ensino médio;
2. Faixa etária matriculada no nível de ensino adequado e obrigatório à sua categoria;
3. Concluintes do ensino médio na idade esperada;
4. Desempenho no Programa Internacional para Avaliação do Estudante.

Indicadores de saúde:

1. Anos potenciais de vida;
2. Mortalidade infantil;
3. Coeficientes de mortalidade por acidentes cardiovasculares (ACVs);
4. Acidentes Vasculares Cerebrais (AVCs).

Indicadores de proteção social:

1. Cobertura previdenciária para maiores de 65 anos;
2. Financiamento da atenção à saúde;

Indicadores de coesão social:

1. Distribuição de renda interpessoal;
2. Morte por homicídio em homens, na faixa de 15 a 24 anos de idade;
3. Taxa de mães adolescentes;
4. Justiça tributária.

## 9.7 Evolução do Sistema Nacional de Dados Ambientais - SINDA

O Sistema Nacional de Dados Ambientais - SINDA passou por três gerações de arquitetura da informação (PEREIRA-JR et al., 2012), cujo desenvolvimento ao longo dos anos foi registrado em trabalhos que publicamos em (CARVALHO; PEREIRA-JR; MATTIELLO-FRANCISCO, 2000) e (MATTIELLO-FRANCISCO; NERI; PEREIRA-JR, 1997). Veremos a seguir como a Arquitetura Genética da Informação pode ser utilizada para explicar a evolução tecnológica do SINDA.

Khanafiah e Situngkir (2004) explicam que “em sistemas vivos, o processo de evolução envolve a mutação genética e um processo de recombinação de uma espécie, de modo que resulte em uma espécie diferente”.

Brooks (2009) explica que mutações são mudanças aleatórias e permanentes do DNA, podendo ser gênicas ou cromossômicas, conforme ilustrado na figura 65, na pág. 194 do Glossário.

Muitas são malélicas, algumas não têm efeito algum e poucas realmente beneficiam o organismo. A natureza seleciona as mutações benéficas ou adaptativas nos organismos para ajudar no curso da evolução.

A história evolutiva de um organismo pode ser identificada a partir da mudança de suas características, sendo a característica de similaridade a base de análise da relação entre as espécies. Neste caso, o diagrama de árvore é um meio lógico para mostrar a relação evolutiva entre os organismos. “Filogenética<sup>6</sup> é o estudo ou estimativa da evolução histórica que está por trás da diversidade biológica. . . Diagramas em árvore representam a relação ancestral aproximada dos organismos” (HALL, 2008).

Uma representação de evolução tecnológica usando “árvore filomemética, baseada em memes tecnológicos de telefones celulares para inferir ou estimar a história evolutiva e relação entre os aparelhos da Nokia” pode ser observado no diagrama da figura 59, resultado do estudo de Khanafiah e Situngkir (2004), sobre a inovação como evolução memética.

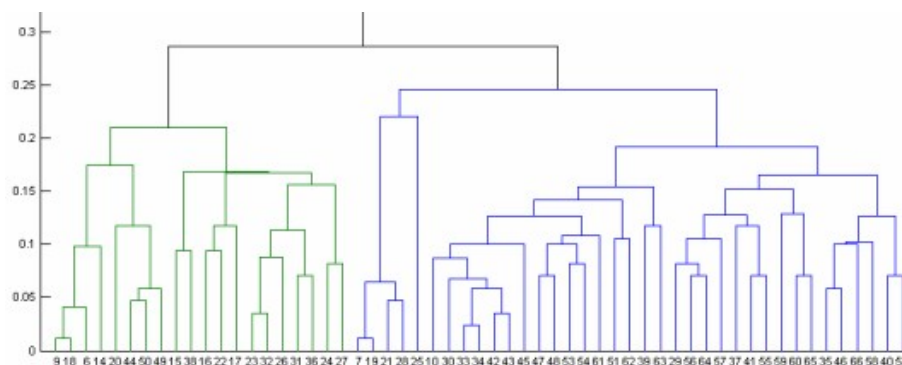


Figura 59: Árvore Filomemética (Parcial) de Evolução dos Telefones Celulares Nokia - Fonte: (KHANAFIAH; SITUNGKIR, 2004)

Na abordagem da Arquitetura Genética da Informação, podemos identificar e caracterizar o genoma de cada uma das gerações, conforme ilustra o diagrama da figura 60.

<sup>6</sup>Filomemética é o equivalente à Filogenética, porém baseada em memes

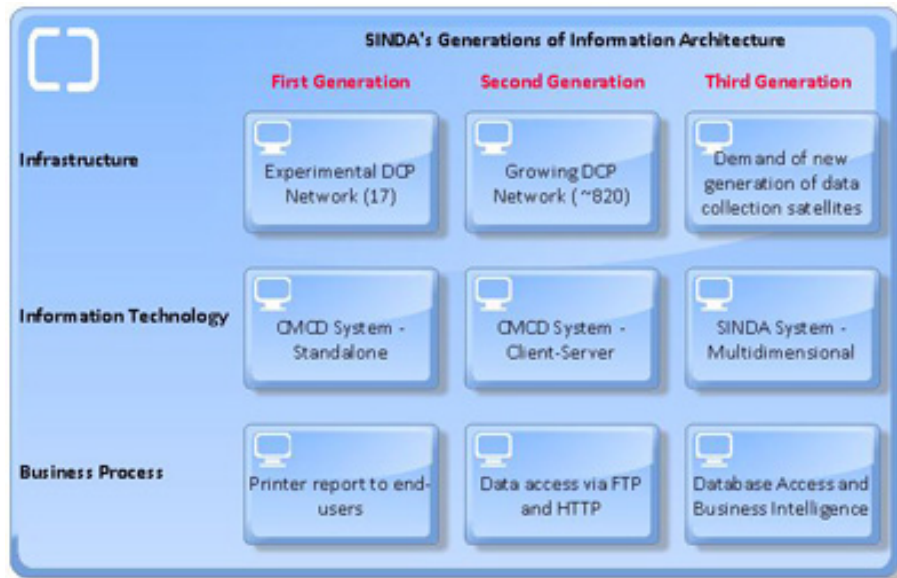


Figura 60: Evolução do SINDA - Fonte: (PEREIRA-JR et al., 2012)

O genoma de cada geração é caracterizado pelas possibilidades de infraestrutura tecnológica conjunturais de cada geração.

A evolução vem ocorrendo de forma adaptativa às necessidades dos usuários em função da disponibilidade de recursos tecnológicos, que também passaram por processos de evolução específicos, compreendendo satélites, plataformas de coleta de dados, sensores, infraestrutura de rede, bancos de dados e sistemas de informação.

A primeira geração do SINDA começou com o lançamento, em 1997, do satélite SCD1, e um sistema de segmento solo que contava com uma rede experimental de 17 Plataformas de Coleta de Dados (PCDs).

O objetivo primeiro era estudar a camada de ozônio, efeito estufa, regeneração da floresta tropical e monitoramento de queimadas. Estava habilitado para uma coleta sistemática de dados do Atlântico Sul (costa oceânica e área terrestre), sensoriamento do nível do mar, pressão atmosférica e temperatura da água.

A segunda geração do SINDA começou com as melhorias em matéria de divulgação de dados utilizando a Rede Nacional de Pesquisa - RNP, Bitnet e outras formas de acesso fornecidos pelo Centro de Missão Coleta de Dados - CMCD.

A fim de apoiar a rede de PCDs em crescimento (mais de sete vezes em dez anos), foram implementadas melhorias no software original SCMCD, proporcionando aos usuários interfaces mais amigáveis e uma maneira eficiente de acessar os dados ambientais coletados pelos SCDs.

A terceira geração do SINDA está relacionada com a reformulação geral de todo o sistema, inclusive com a transferência do Sistema Brasileiro de Coleta de Dados Ambientais para o Centro Regional do INPE na região Nordeste do Brasil, em 2008.

O aumento do número de usuários e da demanda por investimentos na manutenção das PCDs levou à necessidade de uma nova geração de satélites de coleta de dados. A crescente de-



manda de usuários interessados nos dados ambientais para fins de Pesquisa & Desenvolvimento motivou a concepção e desenvolvimento da nova versão do sistema de centro de missão, pela primeira vez, efetivamente chamado de SINDA.

A comparação de similaridade entre as gerações pode ser construída em uma árvore filomemética, como apresentada na figura 61.

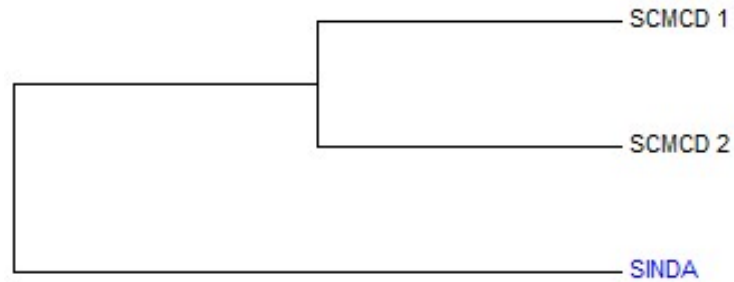


Figura 61: Árvore Filomemética de Evolução do SINDA - Fonte: Produzido pelo autor



## **Parte V**

### **Considerações Finais**



# 10 Conclusão

*“E o fim de todas as nossas explorações  
Será chegar ao lugar de onde saímos  
E conhecê-lo, então, pela primeira vez.”*  
— Thomas Stearns Eliot

Em nossa Proposta de Arquitetura Genética da Informação, visando atender ao objetivo geral, descrevemos a arquitetura com seus elementos e processos associados, enfatizando o papel do arquiteto (genético) da informação. Em seguida, explicamos que o fenômeno da gênese da *informação intencional* ocorre pela emergência do sentido e significado do objeto apreendido pela consciência do sujeito (sendo factível de acontecer, pelo pressuposto da *informação intencional*, em máquinas capazes de realizar o sensoriamento, armazenamento e processamento da informação, o que amplia sobremaneira o escopo de aplicabilidade da arquitetura proposta).

A analogia qualitativa entre os domínios da Arquitetura da Informação e a Genética (para a caracterização da *Arquitetura Genética da Informação*) foi realizada por meio da construção e mesclagem de ontologias de domínio de alto-nível<sup>1</sup> relacionadas a sistemas de informação biológicos e de termos da informação na Genética. O processo foi realizado de maneira semi-automática com ferramentas e algoritmos de mineração textual na fase de aquisição do conhecimento do Modelo-V. Além dos ramos de estudo tradicionais da Genética (Clássica, Molecular, Populacional e Quantitativa) propostos inicialmente, avançamos em outros ramos (Genômica, Humana e Engenharia Genética). Uma vez que as ontologias construídas são de alto-nível, poderão comportar sub-ontologias embutidas e interligadas por pontes relacionais ou por operações de mesclagem, mapeamento e alinhamento, em um processo natural de ampliação sustentável de escopo.

Algumas aplicações foram apresentadas (similaridade fenotípica textual, imagética e acústica; inovação e evolução tecnológica por memética, análise de estrutura retórica textual; hereditariedade da informação; dentre outros) em diversas áreas do conhecimento. Isso evidencia coerência e intuitividade naturais na modelagem de problemas nos mais diversos níveis de dificuldade em termos da *Arquitetura Genética da Informação*. Particularmente relevantes são as considerações a respeito do genoma, genótipo e fenótipo dos objetos em analogia a processos biológicos associados à informação nos organismos vivos, amplamente discutidos na pesquisa.

A *Arquitetura Genética da Informação* foi desenvolvida em total conformidade com a Teoria Geral da Arquitetura da Informação e a disciplina científica da Arquitetura da Informação

---

<sup>1</sup>Uma tendência nas Ciências Biomédicas. Diversas iniciativas nessa área têm impulsionado a construção de ontologias como a Basic Formal Ontology - BFO, GeneOntology - GO e BioTop, por exemplo.

(ambas propostas pelo grupo de Brasília), podendo vir a contribuir com a Tecnologia e Ciência da Informação em todas as fases do ciclo da informação, em termos de apreensão, armazenamento, processamento e recuperação da informação.

## 10.1 Limitações e Dificuldades Encontradas

Como limitações desta tese, entendemos que o esforço para mapear a ontologia com as quatro principais áreas de estudo da genética, apesar de ser válido, acabou prejudicando o aprofundamento naquelas áreas de analogia mais imediata, i.e., com a Genética Clássica e Molecular.

A resposta ao problema P.6 (seção 8.5), limitou-se apenas às teorias da verdade elicitadas por Glanzberg (2009).

A aplicação de similaridade fenotípica entre imagens (item 9.1.2), apesar de ter apresentado resultados muito satisfatórios, foi testada apenas com imagens de mesma resolução e tamanho, no caso, os cães e seus donos e, também, entre imagens urbanas e da natureza. As possibilidades de aplicações na área de reconhecimento facial, por exemplo, não foram consideradas nesta tese.

## 10.2 Trabalhos Futuros

Em trabalhos futuros, pretendemos formalizar matematicamente a *Arquitetura Genética da Informação*, dando continuidade ao trabalho iniciado nesta pesquisa) e, também, verificar a possibilidade de se estabelecer um isomorfismo em um escopo mais reduzido entre as disciplinas de Arquitetura da Informação e a Genética, pela Teoria das Categorias.

Pretendemos continuar o processo de construção, revisão e validação da ontologia da *Arquitetura Genética da Informação* com a formação de grupo de pesquisa específico para esta atividade de organização e representação do conhecimento e traduzir a ontologia contruída para regras em Prolog, conforme explanado no subtópico 3.5.5.3.

Conforme comentado no tópico 9.1.4, esperamos encontrar uma Lei de Potência baseada em *stems* para a configuração de rede de pesquisadores em determinada área do conhecimento humano<sup>2</sup>.

Outras interessantes possibilidades de aplicações incluem:

- Biomatemática<sup>3</sup>, na Genética Quantitativa e dinâmica de populações.
- Compactação de dados, em analogia ao processo cromossômico de compactação do DNA;
- Sistemas de anti-vírus de computador, em relação aos sistemas de imunização biológica;

<sup>2</sup>Nem toda rede de pesquisadores, particularmente as não consolidadas, apresentam leis de potência

<sup>3</sup>A Biomatemática é compreendida como a aplicação de modelos matemáticos para a resolução e compreensão de fenômenos biológicos (BASSANEZI, 2012)

- Computabilidade da aprendizagem automática, em relação ao aprendizado celular;
- Criptografia, com base no DNA e em sistemas biológicos.

Assim, os resultados apresentados nesta pesquisa são panorâmicos e apenas a primeira versão de uma série de iterações possíveis.





## Referências

- ALBUQUERQUE, A. R. R. *Discurso sobre fundamentos de Arquitetura da Informação*. Dissertação (Mestrado) — Faculdade de Ciência da Informação. Universidade de Brasília, Brasília, DF, Setembro 2010. Citado na página 19.
- ALCOFF, L. M. *Epistemology: the big questions*. Oxford: Blackwell, 1998. Citado na página 34.
- AMDAHL, G. M.; BLAAUW, G. A.; BROOKS, F. P. Architecture of the ibm system/360. *IBM Journal for Research Development*, v. 8(2), 1964. (Reprinted in *IBM Journal for Research Development*, 44(1/2), 2000). Citado na página 41.
- ANDERSEN, J. *Analyzing the role of knowledge organization in scholarly communication: An inquiry into the intellectual foundation of knowledge organization*. Dissertação (PhD dissertation) — Royal School of Library and Information Science. Department of Information Studies, Copenhagen, 2004. Citado na página 52.
- ASHER, N.; LASCARIDES, A. Intentions information in discourse. In: *Proceedings of 32nd Meeting of the Association for Computational Linguistics (ACL'94)*. Las Cruces, New Mexico: [s.n.], 1994. p. 34–41. Citado na página 150.
- ASLAM, J. A.; FROST, M. An information-theoretic measure for document similarity. In: ACM (Ed.). *SIGIR'03*. [S.l.]: Publishing Press, 2003. Citado 2 vezes nas páginas 112 e 133.
- ASPRAY, W. The scientific conceptualization of information: a survey. *Annals of the History of Computing*, v. 7, n. 2, p. 1, April 1985. Citado na página 35.
- AULETTA, G. A paradigm shift in biology? *Information MDPI*, v. 1, p. 28–59, 2010. Citado na página 61.
- EVERETT, J. *Information Theory Evolution*. World Scientific, 2003. ISBN 9789812383990. Disponível em: <<http://books.google.com.br/books?id=1agMKiYn2CkC>>. Citado 2 vezes nas páginas 58 e 99.
- AZAR, M. Argumentative text as rhetorical structure: An application of rhetorical structure theory. *Argumentation*, v. 13, n. 1, p. 97–144, 1999. Citado na página 151.
- BAADER, F. *The Description Logic Handbook: Theory, Implementation, Applications*. Cambridge: Cambridge University Press, 2003. 555 p. Citado na página 49.

BAEYER, H. C. von. Information as physical reality. In: *FIS2005*. [s.n.], 2005. Disponível em: <<http://www.mdpi.org/fis2005>>. Citado na página 39.

BAILEY, S. *Information architecture: a brief introduction*. [S.l.], march 2003. Disponível em: <<http://iainstitute.org/tools/download/Bailey-IAIntro.pdf>>. Acesso em: abril de 2007. Citado na página 41.

BAJIC, V. B.; WEE, T. T. *Information Processing Living Systems*. [S.l.]: Imperial College Press - ICP, 2005. (Series on Advances in Bioinformatics Computational Biology, v. 2). Citado 3 vezes nas páginas 94, 119 e 135.

BALASUBRAMANIAN, N.; YEH, M.-L.; CHANG, C.-T. Hierarchical petri nets for modeling metabolic phenotype in prokaryotes. *American Chemical Society*, v. 44, p. 2218–2240, 2005. Disponível em: <[http://www.che.ncku.edu.tw/FacultyWeb/ChangCT-/publications/Hierarchical Petri Nets for Modelling Metabolic Phenotype in Prokaryotes.pdf](http://www.che.ncku.edu.tw/FacultyWeb/ChangCT-/publications/Hierarchical%20Petri%20Nets%20for%20Modelling%20Metabolic%20Phenotype%20in%20Prokaryotes.pdf)>. Citado na página 82.

BARJIS, I.; SAMARRAI, W.; AUGUSTIN, I. Modeling of dna transcription gene regulation using petri nets. In: \_\_\_\_\_. [S.l.]: Idea Group, 2006. v. 1,2, cap. Emerging Trends Challenges in Information Technology Management. Citado 3 vezes nas páginas xxiii, 155 e 156.

BARKER, I. What is information architecture? *Step Two*, May 2005. Disponível em: <[www.steptwo.com.au/papers/kmc\\_whatinfoarch/index.html](http://www.steptwo.com.au/papers/kmc_whatinfoarch/index.html)>. Acesso em: abril de 2007. Citado na página 43.

BASSANEZI, R. C. *Uma Introdução à Biomatemática*. 2012. Disponível em: <<http://posmat.ufabc.edu.br/inverno/images/stories/notas/minicursoufabc.pdf>>. Citado na página 166.

BATES, M. J. Information knowledge: an evolutionary framework for information science. In: *Information Research*. [s.n.], 2005. v. 10, n. n. 4. Disponível em: <<http://InformationR.net/ir-/10-4/paper239.html>>. Citado 4 vezes nas páginas 69, 75, 76 e 96.

BATESON, G. *Mind Nature - A Necessary Unity*. [S.l.]: Advances in Systems Theory, Complexity, the Human Sciences, 1979. Citado na página 39.

BAWDEN, D. Organised complexity, meaning understanding: An approach to a unified view of information for information science. *Emerald Pub.*, v. 59,4/5, p. 307–327, June 2007. Citado 2 vezes nas páginas 17 e 60.

BAYNE, T.; MONTAGUE, M. *Cognitive phenomenology: An introduction*. OUP, 2011. Citado na página 12.

- BÉRUT, A. et al. Experimental verification of lauer's principle linking information thermodynamics. *Nature*, v. 483, n. 10872, p. 187–189, 08 March 2012 2012. Citado na página 10.
- BIASOTO-JR., G. et al. O Índice dna brasil. *Revista da Sociedade Brasileira para o Progresso da Ciência*, n. Ano 58 - 2, p. 4–5, Abril-Maio-Junho 2006. Editor: Carlos Vogt. Citado 3 vezes nas páginas xxiii, 156 e 157.
- BITTENCOURT, G. *Inteligência Artificial - Ferramentas e Teorias*. 3a edição revisada. ed. Florianópolis: Editora da UFSC, 2006. Citado 2 vezes nas páginas 100 e 101.
- BÖHM, C.; KREBS, F. The k-nearest neighbour join: Turbo charging the kdd process. In: *Knowledge Information Systems*. University of Munich, Germany: Springer-Verlag, 2004. v. 6, p. 728–749. Citado na página 134.
- BOUDIER, C. E. M. S. Phenomenology anthropology. *Man World*, Springer Netherls, v. 19, p. 95–101, 1986. ISSN 0025-1534. 10.1007/BF01248380. Disponível em: <<http://dx.doi.org/10.1007/BF01248380>>. Citado na página 29.
- BROOKES, B. C. The foundations of information science. *Journal of Information Science. Part I: Philosophical Aspects*, v. 2 (3/4), p. 125–133, 1980. Citado na página 15.
- BROOKS, R. *Resumão Biologia*. 6a. ed. São Paulo: BFA, 2009. (Ciências Biológicas n. 1). Trad.: Mônica Tambelli; Revisão: Paulo R. Pompeo. Citado 3 vezes nas páginas 59, 96 e 159.
- BROWN, S.; COLLINSON, D.; WILKINSON, R. *One Hundred Twentieth-Century Philosophers*. New York: Routledge, 1998. Citado na página 31.
- BUCHLER, J. *Philosophical Writings of Pierce*. New York: Dover Publications, 1955. Citado na página 23.
- BURGET, R. et al. Rapidminer image processing extension:a platform for collaborative research,. In: *International Conference on Telecommunications Signal Processing*. Baden, Austria: [s.n.], 2010. Citado na página 135.
- BURGIN, M. Is information some kind of data? In: *FIS2005*. [S.l.: s.n.], 2005. Citado na página 39.
- BURIAN, R. M. *The Epistemology of Development, Evolution, Genetics*. New York: Cambridge University Press, 2005. Citado na página 110.
- CANÇADO, M. *Manual de semântica: noções básicas e exercícios*. Belo Horizonte: [s.n.], 2005. Citado na página 37.

- CAPURRO, R.; HJØRLAND, B. The concept of information. *Annual Review of Information Science Technology*, v. 37, p. 343–411, 2003. Disponível em: <<http://www.capurro.de/infoconcept.html>>. Citado 2 vezes nas páginas 35 e 40.
- CARBÓ-DORCA, R.; MEZEY, P. G. *Advances in Molecular Similarity*. [S.l.: s.n.], 1996. Citado na página 105.
- CARVALHO, M. J. M. de; PEREIRA-JR, R. A.; MATTIELLO-FRANCISCO, F. Open source to small satellites ground station. *5th International Symposium on Small Satellites Systems Services*, June 2000. La Baule - France. Citado na página 159.
- CHAITIN, G. *Proving Darwin: Making Biology Mathematical*. New York: Pantheon, 2012. 144 p. ISBN10: 0375423141. ISBN13: 9780375423147. Disponível em: <<http://www.openisbn.com/isbn/9780375423147/>>. Citado 2 vezes nas páginas 87 e 91.
- CHAITIN, G. J. Speculations on biology, information complexity. *Bulletin of the EATCS*, p. 231–237, 2007. Citado na página 91.
- CHALMERS, D. J. The representational character of experience. In: LEITER, B. (Ed.). *The Future for Philosophy*. Oxford: Oxford University Press, 2004. p. 153–181. Citado 2 vezes nas páginas 15 e 16.
- CHENG, E. *Lytro Picture Revolution*. 2011. Citado na página 72.
- CHRISTIANSEN, M. *Language as an Organism: Implications for the Evolution Acquisition of Language*. Washington: Washington University, 1995. Citado na página 69.
- COAKES, E.; CLARKE, S. *Encyclopedia of Communities of Practice in Information Knowledge Management*. Hershey: IRM Press, 2005. Citado na página 46.
- MADSEN, D. (Ed.). *Journal of Information Architecture*. Spring 2009. Disponível em: <<http://journalofia.org/>>. Citado na página 43.
- COVER, T.; THOMAS, J. *Elements of Information Theory*. [S.l.]: Wiley-Interscience, 1991. Citado 2 vezes nas páginas 38 e 39.
- DACONTA, M. C.; OBRST, L. J.; SMITH, K. T. *The Semantic Web: A guide to the future of XML, Web Services Knowledge Management*. [S.l.]: Wiley, 2003. Citado na página 47.
- DAHLBERG, I. Current trends in knowledge organization. *Indeks*, 2006. Disponível em: <[http://dialnet.unirioja.es/servlet/fichero\\_articulo?codigo=2341306](http://dialnet.unirioja.es/servlet/fichero_articulo?codigo=2341306)>. Citado na página 45.
- DAIDA, J. M. et al. Of metaphors darwinism: Deconstructing genetic programming chimera. *CEC*, 1999. Citado na página 100.

DARVAS, G. Symmetry, order, entropy information. *FIS2005*, 2005. Disponível em: <<http://www.mdpi.org/fis2005>>. Citado na página 39.

DAVID, R.; ALLA, H. *Discrete, Continuous, Hybrid Petri Nets*. 2nd. ed. Berlin: [s.n.], 2010. Citado na página 81.

DAVIES, P.; GREGERSEN, N. H. *Information the Nature of Reality: From Physics to Metaphysics*. 1. ed. [S.l.]: Cambridge University Press, 2010. Hardcover. ISBN 0521762251. Citado 2 vezes nas páginas 60 e 63.

DAWKINS, R. *The Selfish Gene*. [S.l.]: Oxford University Press, 1990. Paperback. ISBN 0192860925. Citado na página 99.

DAY, R. Lis, method postmodern science. *Journal of Education for Library Information Science*, v. 37, n. 4, Fall 1996. Citado na página 71.

DENNETT, D. C. *The Intentional Stance*. Cambridge, Massachussets: MIT Press, 1995. Fifth printing. Citado na página 13.

DERQUI, P. M. O paradigma biológico do conhecer e a questão da informação. *Revista de Ciência da Informação*, v. 6, n. 6, Dez 2005. Disponível em: <<http://www.dgz.org.br/dez05-/Arte05.htm>>. Citado na página 16.

DERRIDA, J. Signature, event, context. In: <<HTTP://HYDRA.HUMANITIES.UCI.EDU/DERRIDA/SEC.HTML>>, T. H. . U. (Ed.). *Margins of Philosophy*. [S.l.]: Congr s international des Soci t s de Philosophie de Langue Fran aise, 1971. p. 307–330. Citado na p gina 71.

DERVIN, B. An overview of sense-making research: concepts, methods, results to date. *International Communication Association Annual Meeting*, 1983. Disponível em: <<http://communication.sbs.ohio-state.edu/sense-making/art/artdervin83.html>>. Citado na p gina 16.

DERVIN, B. On studying information seeking methodologically: the implications of connecting metatheory to method. *Information Processing Management*, v. 35(6), p. 727–750, 1999. Citado na p gina 16.

DICTIONARY.COM. The american heritage new dictionary of cultural literacy, third edition. Oct 2012. 'genus'. Disponível em: <<http://dictionary.reference.com/browse/genre>>. Citado na p gina 147.

DICTIONARY.COM. Collins english dictionary - complete & unabridged 10th edition. Oct 2012. 'genre'. Disponível em: <<http://dictionary.reference.com/browse/genre>>. Citado na p gina 147.

- DODIG-CRNKOVIC, G. System modelling information semantics. In: AL, I. J. B. et (Ed.). *Proceedings of the Fifth Conference for the Promotion of Research in IT*. Studentlitteratur, Lund, 2005. Citado na página 38.
- DUPUY, J.-P. *Nas Origens das Ciências Cognitivas*. [S.l.]: Unesp, 1995. 157 p. Citado na página 10.
- EBONLINE. Intentionality (phenomenology). In: *Encyclopaedia Britannica*. Phoenix College Library, Phoenix.: [s.n.], 2012. Citado na página 12.
- EBONLINE. Phenotype. In: *Encyclopaedia Britannica*. Phoenix College Library, Phoenix.: [s.n.], 2012. Citado na página 59.
- ENCYCLOPEDIA.COM. *Logarithmic Scale*. July 25, 2012 2004. A Dictionary of Biology. (Retrieved July 25 2012). Disponível em: <<http://www.encyclopedia.com/doc/1O6-logarithmicscale.html>>. Citado na página 106.
- ENGELN, H. O que são genes? *Geo*, n. 14, p. 32, Junho 2010. Citado na página 13.
- EUZENAT, J. *Ontology Matching*. Physica-Verlag, 2007. ISBN 9783540496120. Disponível em: <<http://books.google.com.br/books?id=qYVpA2t2EtQC>>. Citado na página 123.
- FALK, R. Book. *Genetic Analysis: A History of Genetic Thinking*. [S.l.]: Cambridge University Press, 2009. 330 p. Citado na página 110.
- FALLON, R. The record of realism in messiaen's bird style. In: *Olivier Messiaen: Music, Art Literature*. Christopher Dingle & Nigel Simeone, 2007. Disponível em: <<http://www.oliviermessiaen.org/birdsongs.html>>. Citado 5 vezes nas páginas xxiii, 141, 142, 143 e 144.
- FENG, D.; SIU, W.; ZHANG, H. J. *Multimedia Information Retrieval Management: Technological Fundamentals Applications*. [S.l.]: Springer, 2003. Citado na página 135.
- FINK, E. The problem of the phenomenology of edmund husserl. *Apriori World.*, I, p. 21–55, 1939. Citado na página 29.
- FLORIDI, L. What is the philosophy of information? *Metaphilosophy*, v. 33(1/2), p. 123–145, 2002. Citado na página 16.
- FLORIDI, L. Is semantic information meaningful data? *Philosophy Phenomenological Research*, v. 70(2), p. 351–370, 2005. Citado na página 16.
- FLORIDI, L. In defence of the veridical nature of the semantic information. In: . [s.n.], 2007. Disponível em: <<http://www.philosophyofinformation.net/pdf/idotvnosi.pdf>>. Citado 7 vezes nas páginas xxi, 2, 16, 37, 111, 127 e 128.

- FUCHS, C. Towards a critical theory of information. *TripleC*, v. 7(2), p. 243–292, 2008. Citado na página 17.
- FULTON, K. *News isn't always journalism*. [S.l.: s.n.], 2000. p. 30-35. p. Citado na página 45.
- FURNER, J. Philosophy information studies. the concept of information. In: CRONIN, B. (Ed.). *Annual Review of Information Science Technology*. Medford, NJ: ARIST, 2010. (Information Today), p. 161–200. Citado na página 15.
- GELBART, M. *The Invention of Folk Music Art Music: Emerging Categories from Ossian to Wagner*. New York: Cambridge University Press, 2007. Citado na página 149.
- GIBBS, W. W. The unseen genome: gems among the junk. *Scientific American*, v. 289, n. 5, p. 46–53, November 2003. Citado na página 99.
- GLANZBERG, M. Truth. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Spring 2009. [s.n.], 2009. Disponível em: <<http://plato.stanford.edu/archives/spr2009/entries/truth/>>. Citado 2 vezes nas páginas 129 e 166.
- GLENDINNING, S. *In the Name of Phenomenology*. New York: Routledge, 2007. Citado na página 31.
- GOLDSTEIN, E. B. *Encyclopedia of Perception*. Thous Oaks, California: SAGE Publications, 2012. Citado na página 105.
- GREEN, P.; ROSEMAN, M. *Business systems analysis with ontologies*. New York: Idea Group, 2005. 379p. Citado na página 7.
- HALL, B. *Phylogenetic Trees Made Easy: A How-To Manual*. Sunderland, Massachusetts: Sinauer Associates, 2008. 196 p. 137 illus. ISBN 9780878933105. Disponível em: <<http://books.google.com.br/books?id=sLCpQgAACAAJ>>. Citado na página 159.
- HANCOCK, J. M.; ZVELEBIL, M. J. *Dictionary of Bioinformatics Computational Biology*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2004. Thomas D. Schneider contributed 50 entries to the dictionary. The web links to Tom Schneiders web site are incorrect but this has been hled on the server computer. See <http://www.ccrnp.ncifcrf.gov/toms/toms/>. ISBN 0-471-43622-4. Citado 4 vezes nas páginas xxi, 56, 57 e 118.
- HAYLES, N. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, Informatics*. University of Chicago Press, 1999. ISBN 9780226321462. Disponível em: <<http://books.google.com.br/books?id=JqB6Qy9z3TcC>>. Citado na página 70.
- HEROLD, K. R. Librarianship the philosophy of information. *Library Philosophy Practice*, v. 3(2), 2001. Disponível em: <<http://www.webpages.uidaho.edu/~mbolin%20-/herold.pdf>>. Citado na página 16.

HERSHBERG, U.; EFRONI, S. The immune system other cognitive systems. *Complexity*, v. 6, p. 14–21., 2001. Citado 3 vezes nas páginas [xxiii](#), [97](#) e [195](#).

HJØRLAND, B. *Information Seeking Subject Representation: An Activiti-Theoretical Approach to Information Science*. Westport: Greenwood Press, 1997. Citado na página [46](#).

HJØRLAND, B. Information: Objective or subjective/situational? *Journal of the American Society for Information Science Technology*, v. 58, p. 1448–1456, August 2007. Citado na página [69](#).

HJØRLAND, B.; PEDERSEN, K. N. A substantive theory of classification for information retrieval. In: *Journal of Documentation*. [S.l.: s.n.], 2005. v. 65, p. p. 582–597. Citado 2 vezes nas páginas [44](#) e [45](#).

HODGE, G. Systems of knowledge organization for digital libraries. beyond traditional authority files. In: . Washington, DC: The Council on Library Information Resources, 2000. Disponível em: <http://www.clir.org/pubs/reports/pub91/contents.html>. Citado na página [45](#).

HOFKIRCHNER, W. How to achieve a unified theory of information. *TripleC*, v. 7(2), p. 357–368, 2009. Citado na página [17](#).

HOLSINGER, K. E. Introduction to quantitative genetics. *UCONN*, 2012. Creative Commons. Disponível em: <http://darwin.eeb.uconn.edu/eeb348/lecturenotes/quant-intro.pdf>. Citado na página [157](#).

HORRIDGE, M. *A Practical Guide to Building OWL Ontologies using Protégé4 CO-ODE Tools*. 1.3. ed. Manchester, March 2011. Contributor: Sebastian Brt. Citado na página [48](#).

HUSSERL, E. *Crisis of European Sciences Transcendental Phenomenology*. Northwestern University Press, 1970. (Northwestern University Studies in Phenomenology & Existential Philosophy). ISBN 9780810104587. Disponível em: <http://books.google.com.br/books?id=Ca7GDwz5IF4C>. Citado na página [29](#).

HUSSERL, E. *Ideas Pertaining to a Pure Phenomenology to a Phenomenological Philosophy*. Boston: Martinus Nijhoff, 1983. Translation of: *Ideen zu einer reinen Phénomenologie und phénomenologischen Philosophie* (1913), by F. Kersten. Citado 3 vezes nas páginas [29](#), [30](#) e [81](#).

HUSSERL, E. *The Idea of Phenomenology*. Dordrecht, The Netherls: Kluwer Academic Pub, 1999. Translated by Lee Hardy. Citado na página [29](#).

HUSSERL, E. *Logical Investigations*. Taylor & Francis, 2001. (International Library of Philosophy, v. 1). ISBN 9780415241892. Disponível em: <http://books.google.com.br/books?id=9KNIII0\9JYC>. Citado na página [28](#).



- HUSSERL, E.; PEIFFER, G.; LÉVINAS, E. *Méditations cartésiennes.: Introduction à la phénoménologie*. Isd, 1992. (Bibliothèque des textes philosophiques). ISBN 9782711611331. Disponível em: <[http://books.google.com.br/books?id=CT0extbackslash\\_PgAACAAJ](http://books.google.com.br/books?id=CT0extbackslash_PgAACAAJ)>. Citado na página 29.
- JHA, S. R. *Reconsidering Michael Polanyi's Philosophy*. [S.l.]: University of Pittsburgh Press, 2002. Citado na página 34.
- KAHLMAYER-MERTENS, R. S. et al. *Como elaborar projetos de pesquisa: linguagem e método*. Rio de Janeiro: FGV Editora, 2007. Citado na página 7.
- KANT, I. *Crítica da Razão Pura*. 5a. ed. Lisboa: Fundação Calouste Gulbenkian, 2001. Trad.: Manuela Pinto dos Santos. Introdução e notas: Alexre Fradique Morujão. Tradução do original alemão intitulado Kritik der Reinen Vernunft de Immanuel Kant, baseada na edição crítica de Raymund Schmidt, confrontada com a edição da Academia de Berlim e com a edição de Ernst Cassirer. ISBN 972-31-0623-X. Citado na página 76.
- KAUFMAN, S. A. *The Origins of Order: Self Organization Selection in Evolution*. New York; Oxford.: Oxford University Press, 1993. Citado na página 98.
- KELL, D. B. Genotype-phenotype mapping: genes as computer programs. *PubMed Elsevier*, 2002. Dept of Chemistry, UMIST. PMID: 12414184. MEDLINE. Citado na página 92.
- KHALIFA, K. *What is Phenomenology?* Middlebury, 2011. Department of Philosophy. Disponível em: <<http://s08.middlebury.edu/PHIL0260A/What%20is%20phenomenology-.ppt>>. Citado na página 26.
- KHANAFIAH, D.; SITUNGKIR, H. *Innovation as Evolution*. [S.l.], 2004. Disponível em: <<http://cogprints.org/3991/>>. Citado 2 vezes nas páginas xxiii e 159.
- KHOSROW-POUR, M. *Dictionary of Information Science Technology*. Hershey: Idea Group, 2007. Citado na página 46.
- KLIR, G. J. *Uncertainty Information*. [S.l.]: Wiley-Interscience, 2006. Citado na página 16.
- KNIGHT, J. A.; NESS, B. D. *Encyclopedia of Genetics*. [S.l.]: Salem Press, 2004. Citado 3 vezes nas páginas 57, 58 e 115.
- KOLMOGOROV, A. N. *Foundations of the theory of probability*. Chelsea: Chelsea Pub. Co., 1950. 71 p. Citado 2 vezes nas páginas 118 e 120.
- KOPP, C. *FIT2043 - Technical Documentation for Software Engineers*. [S.l.], 2007. Disponível em: <<http://www.infotech.monash.edu>>. Citado na página 39.

- KOSTOFF, R. N. et al. Literature-related discovery. *Annual Review of Information Science Technology*, v. 43, p. 1–71, 2009. ISSN: 0066-4200 ; E-ISSN: 1550-8382 ; DOI: 10.1002/aris.2009.1440430112. Citado na página 133.
- KRIPPENDORFF, K. *Content analysis: an introduction to its methodology*. Beverly Hills, CA: Sage, 1980. Citado na página 36.
- KÜPPERS, B.-O. Information the origin of life. In: *FIS2005*. [s.n.], 2005. Disponível em: <<http://www.mdpi.org/fis2005>>. Citado na página 39.
- LACEY, A. R. *A Dictionary of Philosophy*. Third edition. [S.l.]: Routledge, 1996. First published in 1976. Citado na página 25.
- LANDAU, C.; SZUDECK, A.; TOMLEY, S. *O Livro da Filosofia*. 2ª reimpressão, 2012.. ed. São Paulo: Globo, 2011. Trad. Rosemarie Ziegelmaier. Citado 3 vezes nas páginas 25, 27 e 77.
- LASZLO, A.; KRIPPNER, S. Systems theories: Their origins, foundations, development. In: \_\_\_\_\_. Amsterdam: Elsevier Science, 1998. (Systems Theories A Priori Aspects of Perception), cap. 3, p. 47–74. Citado na página 75.
- LAYTON, J. *How Pandora Radio Works*. July 2012. Disponível em: <<http://computer-howstuffworks.com/internet/basics/pandora.htm>>. Citado na página 149.
- LEÓN, R. Arquitectura de información: análisis histórico-conceptual. *No Solo Usabilidad Journal*, April 28 2008. Disponível em: <[http://www.nosolousabilidad.com/articulos-/historia\\_arquitectura\\_informacion.htm](http://www.nosolousabilidad.com/articulos-/historia_arquitectura_informacion.htm)>. Citado 3 vezes nas páginas xxi, 43 e 44.
- LEWONTIN, R. C. Genotype phenotype. In: KELLER, E. F.; LLOYD, E. A. (Ed.). *Keywords in Evolutionary Biology*. Cambridge: Harvard University Press, 1992. p. 137–44. Citado 2 vezes nas páginas 58 e 59.
- LIMA-MARQUES, M. Logic without frontiers: Festschrift for walter alexre carnielli on the occasion of his 60th birthday. In: \_\_\_\_\_. London: College Publications, 2011. (Tribute Series, v. 17), cap. Outline of a Theoretical Framework of Architecture of Information: A School of Brasilia Proposal., p. 311–320. Citado 9 vezes nas páginas xi, xiii, 1, 5, 10, 24, 43, 67 e 68.
- LOBO, I.; SHAW, K. Phenotypic range of gene expression: Environmental influence. *Nature Education*, I, p. 1, 2008. Citado na página 62.
- LOGAN, R. K. *Que é Informação? A Propagação da Organização na Biosfera, na Simbiosfera, na Tecnosfera e na Econosfera*. Rio de Janeiro: Editora PUC Rio, 2012. Trad. Adriana Braga. ISBN: 978-85-7866-046-8. Citado 5 vezes nas páginas 35, 36, 69, 70 e 150.

- LOMBARDO, T. *The Ecological Cosmology of Consciousness*. [S.l.], 2011. Disponível em: <[http://www.centerforfutureconsciousness.com/pdf\\_files/Articles-/Consciousness\\_Exped\\_Final\\_May2011.pdf](http://www.centerforfutureconsciousness.com/pdf_files/Articles-/Consciousness_Exped_Final_May2011.pdf)>. Citado na página 33.
- LOSEE, J. *A Historical Introduction to the Philosophy of Science*. Fourth. Oxford: Oxford University Press, 2001. Citado 3 vezes nas páginas [xxi](#), [112](#) e [113](#).
- LOSEE, R. M. A discipline independent definition of information. *Journal American Society of Information Science*, v. 48, p. 254–269, 1997. Citado na página [16](#).
- LUNA, S. V. A revisão de literatura como parte integrante do processo de formulação do problema. In: *Planejamento de pesquisa: uma introdução*. São Paulo: Educ, 1998. Citado na página [23](#).
- MACEDO, F. L. O. *Arquitetura da informação: aspectos epistemológicos, científicos e práticos*. Dissertação (Mestrado em Ciência da Informação) — Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2005. Citado na página [41](#).
- MACHLUP, F. *The economics of information human capital*. Princeton, NJ: Princeton University Press, 1984. Citado na página [75](#).
- MACHLUP, F.; MANSFIELD, U. *The study of information: interdisciplinary messages*. New York: John Wiley, 1983. Citado na página [40](#).
- MADDEN, A. D. Evolution information. *Journal of Documentation*, v. 60, n. 1, p. 9–23, 2004. Citado na página [60](#).
- MAHNER, M.; BUNGE, M. *Foundations of Biophilosophy*. [S.l.]: Springer-Verlag, 1997. Citado na página [60](#).
- MARIJUÁN, P. C. Information life: Towards a biological understanding of informational phenomena. *TripleC*, v. 2(1), p. 6–19, 2004. Citado 3 vezes nas páginas [16](#), [89](#) e [90](#).
- MARROQUÍN, B.; ITALO, O. *El Status Ontológico de la Esencia o del "Que" de la Cosa*. [S.l.]: IIPCIAL, 2007. 236 p. p. Citado 2 vezes nas páginas [24](#) e [71](#).
- MARTINO, L. C. As epistemologias contemporâneas e o lugar da comunicação. In: LOPES, M. I. V. (Ed.). *Epistemologia da Comunicação*. São Paulo: Loyola., 2003. p. pp. 69–101. Citado na página [35](#).
- MATTIELLO-FRANCISCO, F.; NERI, J. A. d. C. F.; PEREIRA-JR, R. A. Relational database for the space plasma data collected by the saci-1 mission. In: LURIE J. B. DELANEY, T. (Ed.). *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. [S.l.], 1997. (Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, v. 3119), p. 162–171. SAO/NASA Astrophysics Data System. Citado na página [159](#).

MATURANA, H. R. Biology of language: The epistemology of reality. *Psychology Biology of Language Thought*, p. 27–63, 1978. Citado 2 vezes nas páginas 15 e 72.

MCGEE, G. *Beyond Genetics: The User's Guide to DNA*. Pymble, Australia: Harper-Collins e-books, 2003. Citado 2 vezes nas páginas 103 e 104.

MCINTYRE, R. Husserl the representational theory of mind. *Topoi*, Springer Netherlands, v. 5, n. 2, p. 101–113–113, set. 1986. ISSN 0167-7411. Disponível em: <<http://dx.doi.org/10.1007/BF00139224>>. Citado na página 13.

MERLEAU-PONTY, M. *Phenomenology of Perception*. [S.l.]: Routledge, 1962. Citado 3 vezes nas páginas 32, 33 e 81.

MEYER, S. C. Dna the origin of life - information, specification explanation. *Discovery Institute, Darwinism, Design, Public Education*, p. 223–285, 2007. Citado na página 63.

MORADO-NASCIMENTO, D. A abordagem sócio-cultural da informação. In: . Marília: UNESP, 2005. Disponível em: <<http://www.portalppgci.marilia.unesp.br/enancib/viewpaper.php?id=71>>. Citado na página 47.

MORAN, D. *Introduction to Phenomenology*. London: Routledge, 2000. Citado 5 vezes nas páginas 11, 25, 28, 29 e 32.

MORANGE, M. *A History of Molecular Biology*. Harvard University Press, 2000. Paperback. ISBN 0674001699. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\path=ASIN/0674001699>>. Citado na página 110.

MORROGH, E. *Information Architecture, an Emerging 21st Century Profession*. [S.l.]: Prentice Hall, 2003. Citado na página 42.

MORUJÃO, A. F. *Significado e Estrutura da Redução Fenomenológica*. [S.l.]: Biblos, 1980. Citado na página 80.

MORVILLE, P.; ROSENFELD, L. *Information Architecture for the World Wide Web*. [S.l.]: O'Reilly, 2006. Citado na página 42.

MUNZ, P. Book. *Philosophical Darwinism : on the origin of knowledge by means of natural selection*. London: Routledge, 1993. ix, 252 p. ; p. ISBN 0415086027. Citado 3 vezes nas páginas 94, 97 e 98.

NALEWALSKI, R. F. *Information Theory of Molecular Systems*. Amsterdam: Elsevier, 2006. Citado 3 vezes nas páginas 118, 120 e 135.

NATIONAL COALITION FOR HEALTH PROFESSIONAL EDUCATION IN GENETICS. *How is genetic information stored?* Lutherville, MD, July 29 2012. Disponível em:

<[http://www.nchpeg.org/bssr/index.php?option=com\\_k2\view=item\id=86:how-is-genetic-information-stored?\Itemid=126](http://www.nchpeg.org/bssr/index.php?option=com_k2\view=item\id=86:how-is-genetic-information-stored?\Itemid=126)>. Citado 3 vezes nas páginas 92, 94 e 102.

NCBI. *Basic Local Aligement Search Tool - BLAST*. 2012. Disponível em: <[http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web\PAGE\\_TYPE=BlastHome](http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web\PAGE_TYPE=BlastHome)>. Citado na página 104.

NEVES, A. M. S. *A Rede Social Acadêmica da Microeletrônica: Rumo ao Aprendizado Ativo*. Dissertação (Mestrado em Desenvolvimento Sustentável) — Universidade de Brasília, Brasília - DF, 2010. Citado 5 vezes nas páginas xxiii, 144, 145, 146 e 147.

NICOLA, U. *Antologia Ilustrada de Filosofia - das Origens à Idade Moderna*. 1a. ed. Sao Paulo: Globo, 2005. (3a. reimpressão). Citado 3 vezes nas páginas 24, 39 e 40.

NIIRANEN, S.; RIBEIRO, A. *Information Processing Biological Systems*. 1st. ed. Berlin: Springer, 2011. 230 p. (Intelligent Systems Reference Library, X). 83 illus., 44 in color. Citado 6 vezes nas páginas 79, 87, 90, 94, 119 e 135.

NIJHOUT, H. F. The importance of context in genetics. *American Scientist*, v. 91, p. 416–430, 2003. Citado na página 61.

NILL, K. *Glossary of Biological Terms*. 3rd. ed. Boca Raton: CRC Press, 2002. Citado na página 135.

O'CONNOR, T.; WONG, H. Y. Emergent properties. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Spring 2012. [S.l.: s.n.], 2012. Citado na página 77.

OLIVEIRA, N. F. Husserl's phenomenology of meaning in the logical investigations. *Veritas*, v. 44/1, p. 117–135, 2000. Citado na página 80.

OULD, M. A. *Strategies for Software Engineering : The Management of Risk Quality*. Chichester: Wiley (series in software engineering practice), 1990. Citado 3 vezes nas páginas xxi, 51 e 54.

PAKE, G. E. Research at xerox parc: a founder's assessment. In: *IEEE Spectrum*. [S.l.: s.n.], 1985. v. 22(10), p. 54–61. Citado na página 42.

PEREIRA-JR, R. A. *Linha do Tempo do Desenvolvimento Histórico da Fenomenologia*. Brasília, novembro 2011. TimeGlider. Disponível em: <[https://timeglider.com/timeline-/line\\_481461e2e7961a13ab7270f9852a03b1](https://timeglider.com/timeline-/line_481461e2e7961a13ab7270f9852a03b1)>. Citado 2 vezes nas páginas xxi e 27.

PEREIRA-JR, R. A.; LIMA-MARQUES, M.; TENORIO, M. F. d. M. The genesis of information in the phenomenological epoché. *Society for Phenomenology Media - SPM*, February 2012. The Fourteenth International Conference. National University, San Diego, California. Host: Paul Majkut . Keynoter Marc van der Bossche. Abstract. Citado na página 75.

- PEREIRA-JR, R. A. et al. Improving satellite data archiving facility for environmental r&d purposes based on architecture of information approach. In: *SpaceOps*. Stockholm: [s.n.], 2012. Disponível em: <<http://www.spaceops2012.org/proceedings/documents/id1295180-Paper-003.pdf>>. Citado 3 vezes nas páginas xxiii, 159 e 160.
- PIETERSMA, H. *Phenomenological Epistemology*. New York: Oxford University Press, 2000. Citado na página 29.
- POLANYI, M. The structure of consciousness. *Brain*, LXXXVIII, p. 799–810, 1965. Citado 2 vezes nas páginas 34 e 77.
- POLANYI, M.; PROSCH, H. *Meaning*. [S.l.]: University of Chicago Press, 1975. ISBN 0-226-67294-8. Citado 2 vezes nas páginas 34 e 77.
- RAKOVA, M. *Philosophy of Mind A–Z*. [S.l.]: Edinburgh University Press, 2006. Citado 7 vezes nas páginas 14, 26, 27, 28, 32, 33 e 34.
- RAMAGE, M.; SHIPP, K. *Systems Thinkers*. London: Springer, 2009. 316 p. Citado na página 15.
- RAPID-I. *Intelligent Business Intelligence - Rapid Miner*. July 2012. Disponível em: <[http://rapid-i.com/component/option,com\\_rline\\_docman/task,doc\\_download/gid,46/Itemid,165](http://rapid-i.com/component/option,com_rline_docman/task,doc_download/gid,46/Itemid,165)>. Citado na página 116.
- READING, A. *Meaningful Information - The Bridge Between Biology, Brain, Behavior*. [S.l.]: Springer, 2011. Citado 3 vezes nas páginas 13, 62 e 78.
- RESMINI, A.; ROSATI, L. *Pervasive Information Architecture*. Burlington, MA: Morgan Kaufman, 2011. Citado na página 41.
- RIOLO, R.; O'REILLY, U.-M.; MCCONAGHY, T. *Genetic Programming Theory Practice VII*. [S.l.]: Springer, 2010. Citado na página 90.
- RITTNER, D.; MCCABE, T. L. *Eyclopedia of Biology*. New York: Facts on File, 2004. Citado 5 vezes nas páginas xxi, 57, 88, 89 e 115.
- ROBINSON, T. *Genetics for Dummies*. 1st. ed. [S.l.]: Wiley, 2006. Citado 2 vezes nas páginas 1 e 111.
- ROBREDO, J. *Da Ciência da Informação Revisitada aos Sistemas Humanos de Informação*. Brasília: [s.n.], 2003. Citado na página 76.
- ROEDERER, J. G. *Information Its Role in Nature*. 1. ed. Springer, 2005. Hardcover. ISBN 3540230750. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/3540230750>>. Citado na página 14.

ROWLEY, J. The wisdom hierarchy: Representations of the dikw hierarchy. *Journal of Information Science*, v. 33(2), p. 163–180, 2007. Citado na página 75.

RYSIEW, P. Epistemic contextualism. In: Z., E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. [s.n.], 2009. Disponível em: <[http://plato.stanford.edu/archives/spr2009/entries-contextualism-epistemology/](http://plato.stanford.edu/archives/spr2009/entries/contextualism-epistemology/)>. Citado na página 71.

SAAB, D. J.; RISS, U. V. Information as ontologization. *Journal of the American Society for Information Science Technology*, 2011. Citado 4 vezes nas páginas xxi, 68, 76 e 77.

SACHS, E. Getting started with protégé-frames. *Stanford*, 2006. Disponível em: <[http://protege.stanford.edu/doc/tutorial/get\\_started/table\\_of\\_content.html](http://protege.stanford.edu/doc/tutorial/get_started/table_of_content.html)>. Citado 2 vezes nas páginas 47 e 51.

SAGÜILLO, J. M. One sense of 'information': A quick tutorial to information-theoretic logic. *TripleC*, v. 7(2), p. 179–184, 2009. Citado na página 17.

SAMUEL, K.; OBRST, L. *Translating OWL Semantic Web Rules into Prolog: Moving Toward Description Logic Programs*. [S.l.], 2008. Citado na página 56.

SANDMEYER, B. *Husserl's Constitutive Phenomenology: its Problem and Promise*. New York: Routledge, 2007. Citado 2 vezes nas páginas 11 e 12.

SANTAELLA, L. *Percepção - Fenomenologia, Ecologia, Semiótica*. São Paulo: Cengage Learning, 2012. Citado 2 vezes nas páginas 32 e 33.

SAWAI, H. *Biological Functions for Information Communication Technologies: Theory Inspiration*. Berlin: Springer, 2011. Citado 6 vezes nas páginas xxv, 59, 94, 115, 119 e 135.

SCHRÖDINGER, E. *What is Life?* Cambridge: Cambridge University, 1992. Citado na página 36.

SCHROEDER, M. R. *Number Theory in Science Communication*. Fourth. [S.l.]: Springer, 2006. Citado na página 106.

SCHUSTER, A. Dna algorithms for petri net modeling. *Informatica*, v. 32, p. 421–427, 2008. Citado 2 vezes nas páginas 154 e 155.

SEARLE, J. *The Construction of Social Reality*. New York: The Free Press, 1997. Citado na página 77.

SEARLE, J. R. *Minds, Brains, Science*. Harvard University Press, 1984. (Reith Lectures). ISBN 9780674576339. Disponível em: <[http://books.google.com.br/books?id=yNjN-\\\_jznw4C](http://books.google.com.br/books?id=yNjN-\_jznw4C)>. Citado na página 128.

- SECKO, D. A monk's flourishing garden: The basics of molecular biology explained. *The Science Creative Quarterly*, 2011. Disponível em: <<http://www.scq.ubc.ca/a-monks-flourishing-garden-the-basics-of-molecular-biology-explained/>>. Citado 2 vezes nas páginas xxiii e 190.
- SHANNON, C. *A Mathematical Theory of Information*. [S.l.], 1948. P. 379-423, 623-656. Citado 6 vezes nas páginas xxi, 36, 37, 69, 118 e 120.
- SHANNON, C. The lattice theory of information. *IEEE Transactions on Information Theory*, v. 1, p. 105–107, 1953. Citado na página 69.
- SIEGEL, S. The content of perception. In: *Stanford Encyclopedia of Philosophy*. [S.l.]: SEP, 2011. Citado na página 81.
- SIEWERT, C. *The Significance of Consciousness*. Princeton: Princeton University Press, 1998. Citado na página 16.
- SIQUEIRA, A. H. *Arquitetura da Informação: Proposta de Brasília para a fundamentação e caracterização de uma disciplina científica*. Tese (Doutorado) — Faculdade de Ciência da Informação. Universidade de Brasília, Brasília, 2012. Citado 12 vezes nas páginas xi, xiii, 1, 5, 6, 14, 17, 18, 41, 43, 68 e 76.
- SIQUEIRA, A. H. et al. Concerning architecture of information its applicability: Proposal of a comprehensive definition. October 2009. Citado na página 41.
- SMITH, B. The basic tools of formal ontology. In: GUARINO, N. (Ed.). *Frontiers in Artificial Intelligence Applications*. Amsterdam, Oxford, Tokyo, Washington, DC.: IOS Press, 1998. p. 19–28. Citado na página 112.
- SMITH, B.; CEUSTERS, W. Relations in biomedical ontologies. *Genome Biology*, v. 6, n. 5, p. R46:1–15, 2005. Citado 2 vezes nas páginas 50 e 53.
- SMITH, D. K.; ALEXER, R. C. *Fumbling the Future: How Xerox Invented, Then Ignored, the First Personal Computer*. [S.l.]: William Morrow & Co., 1988. Citado na página 42.
- SMITH, D. W. *Mind World: Essays in Phenomenology Ontology*. [S.l.]: Cambridge University Press, 2004. Citado na página 70.
- SMITH, D. W. Phenomenology. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Fall 2011. [S.l.: s.n.], 2011. Citado 3 vezes nas páginas 11, 24 e 25.
- SMITH, J. M. The concept of information in biology. *Information the Nature of Reality: From Physics to Metaphysics*, I, p. 123–145, 2010. Citado 3 vezes nas páginas xxi, 60 e 88.



- SNP-CONSORTIUM. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, v. 409, p. 928–933, February 2001. Disponível em: <[http://genepath.med.harvard.edu/~reich%20-SNP\\_Map.pdf](http://genepath.med.harvard.edu/~reich%20-SNP_Map.pdf)>. Citado na página 104.
- SNUSTAD, D. P.; SIMMONS, M. J. *Fundamentos da Genética*. Rio de Janeiro: [s.n.], 2008. Citado 3 vezes nas páginas xxiii, 192 e 196.
- SOERGEL, D. *Organizing information: principles of data base retrieval systems*. [S.l.]: Elsevier, 1985. (Library Information Science). ISBN: 0126542619. Citado na página 47.
- SOUNDHOUND. *SoundHound - Instant Music Search Discovery*. 2012. Disponível em: <<http://www.soundhound.com/index.php?action=s.sound2sound>>. Citado 2 vezes nas páginas xxiii e 150.
- SOUZA, R. F. de. *Mutações gênicas e cromossômicas e suas consequências*. Londrina, 2012. Disponível em: <[http://www.uel.br/pessoal/rogerio/genetica/respostas/pratica\\_05.html](http://www.uel.br/pessoal/rogerio/genetica/respostas/pratica_05.html)>. Citado 3 vezes nas páginas xxiii, 190 e 194.
- SPETNER, L. M. *Not by Chance!: Shattering the Modern Theory of Evolution*. Judaica Press, 1998. ISBN 9781880582244. Disponível em: <<http://books.google.com.br/books?id=29BGPwAACAAJ>>. Citado na página 91.
- STAAB, S.; STUDER, R. *Hbook on Ontologies*. Springer, 2009. (International Hbooks on Information Systems). ISBN 9783540709992. Disponível em: <<http://books.google.com.br/books?id=W6ZNCaolVbwC>>. Citado na página 48.
- STANFORD UNIVERSITY SCHOOL OF MEDICINE. *Prótegé*. Stanford, July 2012. Grant LM007885. Disponível em: <<http://protege.stanford.edu>>. Citado 3 vezes nas páginas 113, 117 e 120.
- STEGMANN, U. E. Genetic information as instructional content. *Philosophy of Science*, v. 72, p. 425–443, July 2005. Citado na página 79.
- STEINBERG, M.; COSLOY, S. D. *The Facts on File Dictionary of Biotechnology Genetic Engineering*. Facts On File, Incorporated, 2009. (The Facts on File Science Dictionary Series). ISBN 9781438109367. Disponível em: <<http://books.google.com.br/books?id=UXZGroS1hBQC>>. Citado 2 vezes nas páginas 1 e 115.
- STERELNY, K.; GRIFFITHS, P. E. Book. *Sex death : an introduction to philosophy of biology / Kim Sterelny Paul E. Griffiths*. [S.l.]: University of Chicago Press, Chicago, Ill. :, 1999. xv, 440 p. : p. ISBN 0226773035 0226773043 0226773035 0226773043 0226773035 0226773043. Citado 6 vezes nas páginas 14, 78, 79, 94, 99 e 110.
- STONIER, T. *Information the Internal Structure of the Universe: An Exploration Into Information Physics*. London: Springer-Verlag, 1990. 155 p. ISBN 9783540198789.

Disponível em: <<http://books.google.com.br/books?id=MaVKPgAACAAJ>>. Citado 2 vezes nas páginas 14 e 63.

STUDER, R.; BENJAMINS, V.; FENSEL, D. Data knowledge engineering. *Knowledge Engineering: Principles Methods*, v. 25(1-2), p. 161–197, 1998. Citado 2 vezes nas páginas 47 e 52.

SVEIBY, K.-E. *Towards a Knowledge Perspective on Organisation*. Dissertação (PhD dissertation) — University of Stockholm, 1994. Citado na página 36.

SYMES, P. *Video Compression Demystified*. New York: McGraw-Hill, 2001. Citado 2 vezes nas páginas xxii e 136.

TANMONKEY. *Owners who look like their dogs*. 2012. Disponível em: <<http://www.tanmonkey.com/fun/dog-looks-like-owner.php>>. Citado 2 vezes nas páginas xxii e 136.

THE CORNELL LAB OF ORNITHOLOGY. *All About Birds*. Ithaca, NY, 2012. Disponível em: <<http://www.allaboutbirds.org/page.aspx?pid=1059>>. Citado na página 140.

THOMPSON, A. K.; CHADWICK, R. F. *Genetic Information - Acquisition, Access Control*. Preston, UK: [s.n.], 1997. Citado na página 135.

TURBAN, E.; MCLEAN, E.; WETHERBE, J. *Tecnologia da Informação para Gestão*. 3a. ed. Porto Alegre: Bookman, 2004. Trad. Renate Schinke. Citado na página 105.

TVERSKY, A. *Preference, Belief, Similarity*. Cambridge, Massachusetts: The MIT Press, 2004. Citado na página 105.

UNL. *Phenotypic Similarity*. 2012. Disponível em: <<http://www1.unl.edu/search/?q=phenotypic+similarity&submit=Go>>. Citado na página 106.

VEDRAL, V. *Decoding Reality: The Universe as Quantum Information*. [S.l.]: Oxford University Press, 2010. Citado 2 vezes nas páginas 38 e 63.

VLASOV, S. V. Information world. In: *FIS2005*. MDPI, 2005. Disponível em: <<http://www.mdpi.org/fis2005/>>. Citado na página 39.

WAITE, M. *iBird the Evolution of the Digital Field Guide*. 2012. Magazine Birdwatcher's Digest. Disponível em: <<http://ibird.com/episode-20-ibird-the-evolution-of-the-digital-field-guide/>>. Citado 4 vezes nas páginas xxii, 141, 142 e 143.

WATSON, J. D.; CRICK, F. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, v. 171, p. 737–738, 25 April 1953. Citado na página 155.

- WEITZMANN, L. M. *The Architecture of Information: Interpretation Presentation of Information in Dynamic Environments*. Dissertação (PhD Dissertation) — Massachusetts Institute of Technology, 1995. Disponível em: <<http://dspace.mit.edu/hle/1721.1/29085>>. Citado na página 42.
- WIENER, N. *Cybernetics; or Control Communication in the Animal the Machine*. 2 (1961).. ed. Cambridge, MA: MIT Press, 1948. Citado na página 36.
- WILKINS, J. S. What's in a meme? reflections from the perspective of the history philosophy of evolutionary biology. *Journal of Memetics - Evolutionary Models of Information Transmission*, v. 2, 1998. Disponível em: <<http://cfpm.org/jom-emit/1998/vol2/wilkins.js.html>>. Citado 3 vezes nas páginas xxv, 90 e 100.
- WILLIAMS, A. Inheritance of biological information - part i: The nature of inheritance of information. *Technical Journal*, v. 19(2), p. 29–35, 2005. Citado na página 91.
- WILSON, R. A.; KEIL, F. C. *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press, 2006. 763-765 p. Citado na página 104.
- WINDELSPECHT, M. *Genetics 101*. Westport, Connecticut: Greenwood Press, 2007. Citado na página 57.
- ZALIWSKI, A. S. Information - is it subjective or objective? *TripleC*, v. 9(1), p. 77–92, 2011. Citado 3 vezes nas páginas 17, 69 e 75.
- ZEZULA, P. et al. Advances in database systems, vol. 32. In: *Similarity Search: The Metric Space Approach*. 15th. ed. [S.l.]: Springer, 2006. Citado na página 105.
- ZINS, C. Conceptual approaches for defining data, information knowledge. *Journal of The American Society for Information Science Technology - InterScience*, v. 58(4), p. 479–493, 2007. Citado na página 35.



# Glossário Ilustrado

## **Arquitetura da Informação Intencional**

Disciplina que investiga o mundo considerando as sua Formas, Contextos, Manifestações e Significados (FCMS) do ponto de vista de um sujeito e por um método fenomenológico, para investigar as distinções do sujeito, do fenômeno e do conhecimento, na construção de configurações do espaço da informação.

## **Arquitetura Genética**

Refere-se ao mapeamento genótipo-fenótipo, i.e., a influência implícita do genótipo em um traço fenotípico do organismo.

## **Arquitetura Genética da Informação**

Refere-se à arquitetura do processo de gênese da *informação intencional* e realiza o mapeamento do genótipo de quaisquer objetos perceptíveis no mundo para o seu respectivo estado fenotípico, em termos da informação.

## **Contexto**

Caracteriza uma coleção de relações do fenômeno com o mundo, do ponto de vista do sujeito.

## **Código Genético**

É um código universal que estabelece a equivalência entre a sequência de bases no DNA e a sequência correspondente de aminoácidos, na proteína. É estabelecido através da complementaridade entre as bases nitrogenadas: Adenina (A), Guanina (G), Timina (T) e Citosina (C).

CÓDONS - m RNA

|                   |   | Segunda Letra                             |                                      |  |  | U<br>C<br>A<br>G |                   |
|-------------------|---|---|--------------------------------------|--|--|------------------|-------------------|
|                   |   | U   | C                                    | A  | G  |                  |                   |
| Primeira<br>Letra | U | UUU } Phe<br>UUC }<br>UUA } Leu<br>UUG }  | UCU }<br>UCC } Ser<br>UCA }<br>UCG } | UAU } Tyr<br>UAC }<br>UAA } Sem<br>UAG } sentido | UGU } Cys<br>UGC }<br>UGA } Sem sentido<br>UGG } Try | U<br>C<br>A<br>G | Terceria<br>Letra |
|                   | C | CUU }<br>CUC } Leu<br>CUA }<br>CUG }      | CCU }<br>CCC } Pro<br>CCA }<br>CCG } | CAU } His<br>CAC }<br>CAA } Gln<br>CAG }         | CGU }<br>CGC } Arg<br>CGA }<br>CGG }                 |                  |                   |
|                   | A | AUU } Ileu<br>AUC }<br>AUA }<br>AUG } Met | ACU }<br>ACC } Thr<br>ACA }<br>ACG } | AAU } Asn<br>AAC }<br>AAA } Lys<br>AAG }         | AGU } Ser<br>AGC }<br>AGA } Arg<br>AGG }             |                  |                   |
|                   | G | GUU }<br>GUC } Vai<br>GUA }<br>GUG }      | GCU }<br>GCC } Ala<br>GCA }<br>GCG } | GAU } Asp<br>GAC }<br>GAA } Glu<br>GAG }         | GGU }<br>GGC } Gly<br>GGA }<br>GGG }                 |                  |                   |

Figura 62: Decifrando o Código Genético (Códon) - Fonte: (SOUZA, 2012)

O código possui poucos elementos, mas é necessário que haja conexão química entre as bases que constituem ambos os filamentos, sendo A-T e C-G as únicas possíveis ligações no DNA. No RNA, a Uracila (U) substitui a Timina (T).

## DNA

É um programa que contém as instruções genéticas usadas no desenvolvimento e funcionamento de todos os organismos vivos.

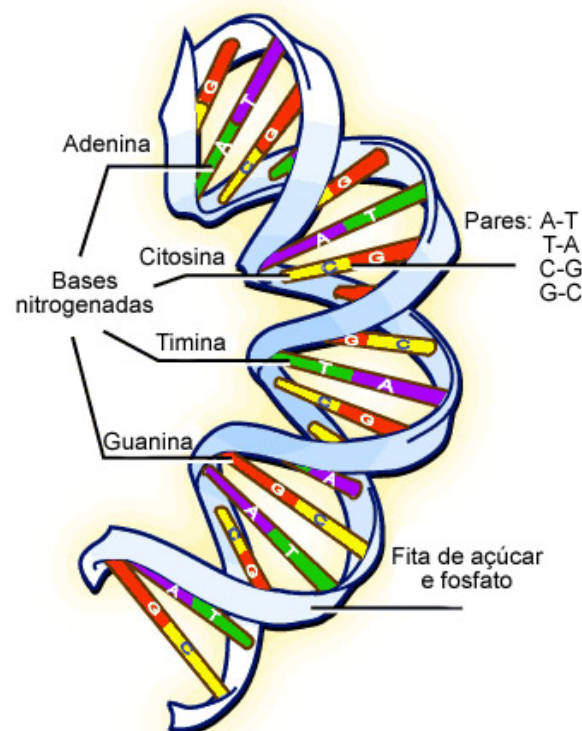


Figura 63: Molécula de DNA - Fonte: (SECKO, 2011)

Na *Arquitetura Genética da Informação*, é um programa que contém a *informação intencional*, cujo padrão de matéria organizada ou energia é detectada por um receptor animado ou manufaturado, desencadeando nele uma mudança comportamental, funcional, ou estrutural.

**Femótipo**

É o conjunto de características culturais observáveis em um indivíduo ou população.

**Fenomenologia Genética**

Estuda a gênese do significado das coisas dentro do fluxo das experiências individuais.

**Fenótipo**

É o conjunto de características observáveis ou caracteres de um indivíduo, como a sua morfologia, desenvolvimento, propriedades bioquímicas ou fisiológicas e comportamento, sob a influência de fatores ambientais.

Na *Arquitetura Genética da Informação*, é o conjunto de características observáveis ou caracteres de um objeto, como suas propriedades primárias e secundárias, além de seu comportamento dentro de um contexto.

**Fenômeno**

É um acontecimento observável, particularmente algo especial, literalmente “algo que pode ser visto” (gr. *phainomenon* = “observável”). Aparece ao indivíduo de tal maneira que produz estados mentais subjetivos (*qualia*).

**Forma**

É uma estrutura que associa os dados do fenômeno, passando a identificar um objeto na experiência do sujeito.

**Gene**

É a unidade física fundamental e funcional da hereditariedade.

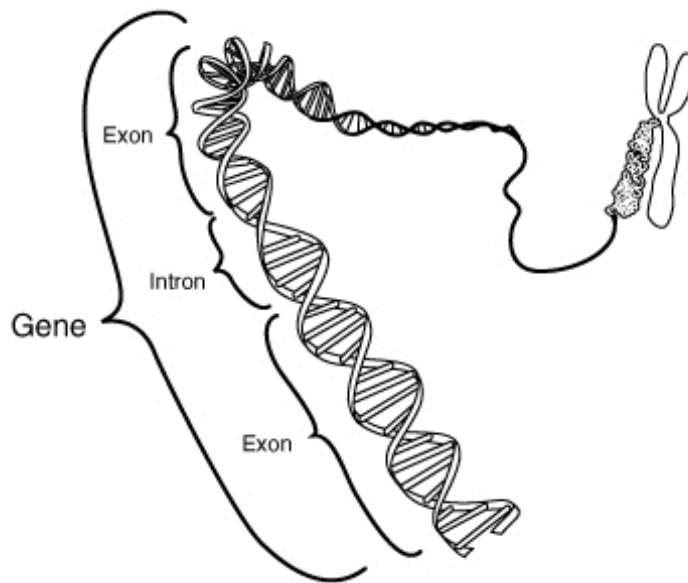


Figura 64: Gene - Fonte: (SNUSTAD; SIMMONS, 2008)

Na *Arquitetura Genética da Informação*, é a unidade física fundamental e funcional de informação.

### Genoma

É a totalidade de informação genética que possui um organismo.

Na *Arquitetura Genética da Informação*, é a totalidade de informação que caracteriza uma classe de objetos, diferenciando-a das demais.

### Genética

É o estudo dos genes e de seus papéis na hereditariedade, i.e., o meio pelo qual certos traços são passados entre gerações.

Na *Arquitetura Genética da Informação*, é o estudo das características da informação e de suas possíveis combinações na produção de novos traços de informação.

### Genótipo

São as informações hereditárias de um indivíduo contidas em seu genoma, i.e., o conjunto de genes de um indivíduo.

Na *Arquitetura Genética da Informação*, são as informações perceptíveis e transmissíveis de um objeto contidas na totalidade das características da informação em sua classe.

### Genômica

É o estudo de todos os genes de um organismo, incluindo as interações destes genes com cada um dos outros e com o ambiente onde vive o organismo.

Na *Arquitetura Genética da Informação*, é o estudo de todas as características da informação de uma classe de objetos, incluindo suas interações, em determinado contexto, contemplando, também, o genótipo e fenótipo dos objetos.



**Imanência**

É a manutenção dos limites da experiência do possível; a existência da causa na própria causa; permanência.

**Indivíduo**

É uma instância única e separada de um organismo (animal ou vegetal) distinto dos outros de uma mesma espécie.

Na *Arquitetura Genética da Informação*, é um objeto único e separado dentre outros da mesma classe.

**Informação Genética**

É qualquer informação biológica codificada nas sequências de nucleotídeos do DNA ou RNA.

**Informação Intencional**

É um padrão de matéria organizada ou energia que é detectada por um receptor animado ou manufaturado e, assim, desencadeia uma mudança no comportamento, funcionamento, ou na estrutura organizacional da entidade receptora – que pode ser uma macromolécula, uma célula, um organismo, uma planta, um animal, ou um dispositivo fabricado; Informação Significativa.

**Intencionalidade**

É a característica de um sujeito estar consciente de alguma coisa, i.e., o seu direcionamento para um objeto, habilitando-o a lidar com o problema da imanência e transcendência, pela relação entre o que está dentro da consciência e o que se estende além dela.

**Manifestação**

É uma coleção de dados no espaço dos objetos que um sujeito é capaz de perceber como um fenômeno.

**Meme**

É a unidade fundamental e funcional da hereditariedade cultural.

Na *Arquitetura Genética da Informação*, é a unidade fundamental e funcional da evolução cultural.

**Memética**

É o estudo formal dos memes e de seus papéis na evolução cultural, por meio da transmissão de informação ou de ideias entre as gerações.

**Mutação**

São as modificações súbitas que ocorrem no material genético e que são passíveis de serem herdadas pelos descendentes.

|                             |                                  |                  |      |     |     |     |     |     |     |
|-----------------------------|----------------------------------|------------------|------|-----|-----|-----|-----|-----|-----|
|                             |                                  | Seqüência normal | ACA  | ATG | CAA | CGA | CAA |     |     |
|                             |                                  |                  | Thr  | Met | Gln | Arg | Gln |     |     |
| Troca de um par de bases    | Sem a troca do aminoácido        | ACG              | ATG  | CAA | CGA | CAA |     |     |     |
|                             |                                  | Thr              | Met  | Gln | Arg | Gln |     |     |     |
|                             | Com a troca do aminoácido        | ATA              | ATG  | CAA | CGA | CAA |     |     |     |
|                             | Ileu                             | Met              | Gln  | Arg | Gln |     |     |     |     |
|                             | Mutação para um códon de término | ACA              | ATG  | TAA | CGA | CAA |     |     |     |
|                             |                                  | Thr              | Met  |     |     |     |     |     |     |
| Deleção/inservação de bases | Inserção de 1 base               | ACA              | GAT  | GTA | ACG | ACA | A   |     |     |
|                             |                                  | Thr              | Asp  | Val | Thr | Thr |     |     |     |
|                             | Perda de 1 base                  | ACA              | ATC  | AAC | GAC | AA  |     |     |     |
|                             |                                  | Thr              | Ileu | Asn | Asp |     |     |     |     |
|                             | Inserção de uma seqüência        | ACA              | ATG  | AGG | GGG | CTA | CAA | CGA | CAA |
|                             |                                  | Thr              | Met  | Arg | Gly | Leu | Gln | Arg | Gln |

Figura 65: Mutação - Fonte: (SOUZA, 2012)

O termo mutação pode se referir tanto à modificação do material genético quanto ao processo que lhe deu origem. Um organismo que exiba um novo fenótipo como resultado de uma mutação é chamado de mutante. A mutação pode ser gênica ou cromossômica. Na mutação gênica, as alterações ocorrem em um único par de bases numa determinada região do DNA. Na mutação cromossômica, segmentos de cromossomos, cromossomos inteiros, ou conjuntos de cromossomos podem estar envolvidos em alterações genéticas.

## Organismo

É uma coisa viva que pode reagir a estímulos, reproduzir-se, crescer e manter o equilíbrio interno diante de variações ambientais (homeostase). Pode ser um vírus, bactéria, protista, fungo, planta ou animal

Na *Arquitetura Genética da Informação*, é uma classe homogênea de objetos passível de interrelacionamentos e descrição baseada em suas comunalidades, cujas características essenciais são preservadas diante de mudanças externas.

## Penetrância

Penetrância é um termo utilizado em genética para descrever a proporção de indivíduos portadores de uma variação particular de um gene que igualmente expressa no seu fenótipo.

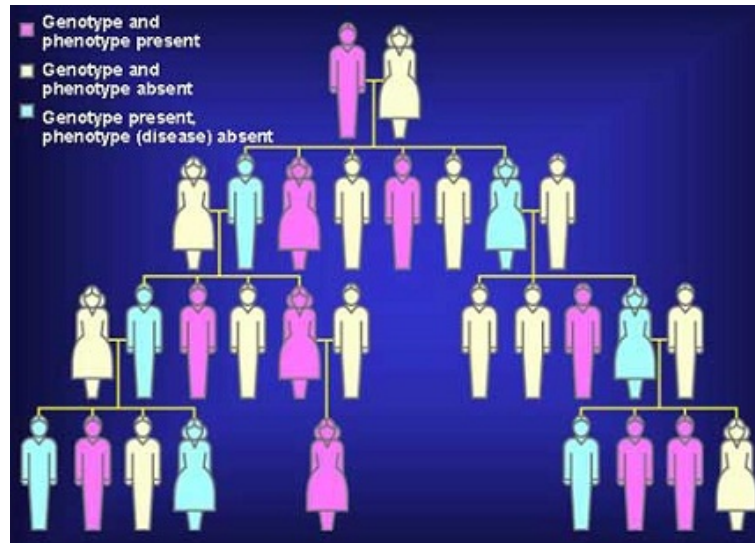


Figura 66: Penetrância: Um Mapa Genótipo-Fenótipo do Câncer - Fonte: (HERSHBERG; EFRONI, 2001)

É a porcentagem de descendentes com um determinado genótipo que manifesta a característica fenotípica pelo menos em algum grau. .

### População

É um grupo de organismos de uma espécie que se cruzam e vivem em um mesmo lugar, ao mesmo tempo.

Na *Arquitetura Genética da Informação*, é um grupo de objetos de uma classe comum e que participam de um contexto comum..

### RNA

É um ácido nucleico que desempenha um papel importante na transferência de informação a partir do DNA para o sistema de síntese de proteína da célula.

Na *Arquitetura Genética da Informação*, é um mecanismo de transferência de informação para a caracterização de aspectos específicos de um objeto.

### Significado

É a intenção designada por um sujeito para supor um objeto.

### Tradução

É o processo de síntese de proteínas, ou seja, a formação de proteínas específicas e correspondentes à porção da fita de DNA copiada na etapa anterior, de transcrição.

### Transcendência

É a caracterização dos objetos que não podem derivar das impressões que temos deles; é o caráter do que está fora do alcance de nossa ação ou até de nosso pensamento.

## Transcrição

É o processo de formação de uma molécula de RNA com o código genético originário de uma molécula molde de DNA.

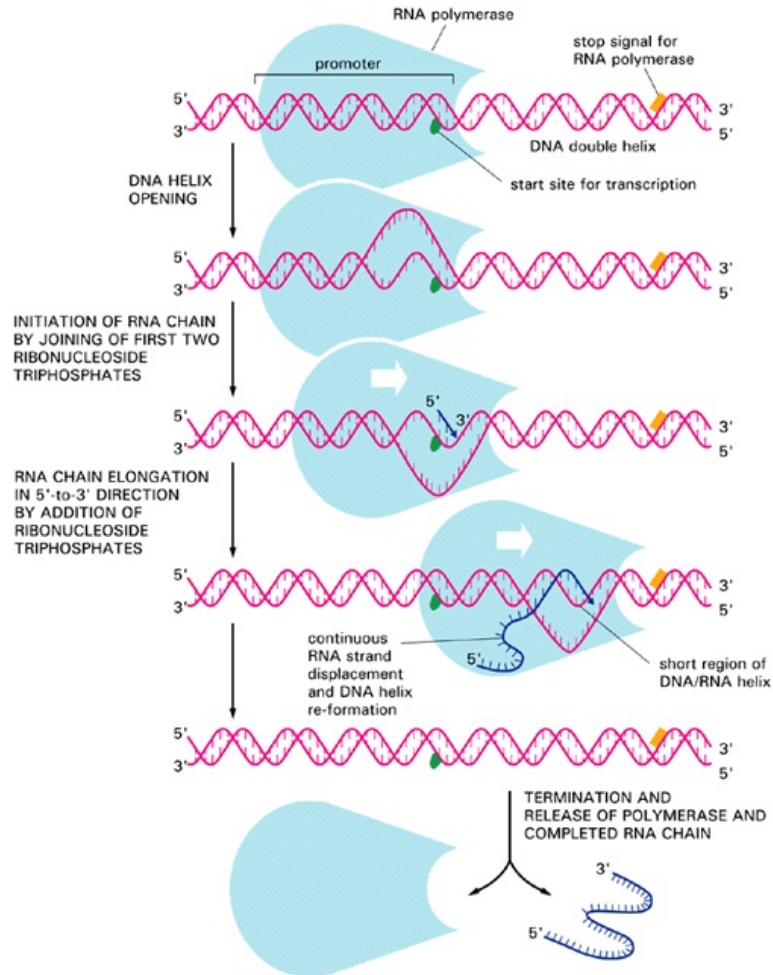


Figura 67: Transcrição - Fonte: (SNUSTAD; SIMMONS, 2008)

Neste processo, as fitas do DNA se separam e uma serve de molde para o RNA, enquanto a outra fica inativa. Ao fim da transcrição, as fitas que foram separadas voltam a se unir. A transcrição é um processo altamente seletivo, pois apenas pequenas porções da fita de DNA são copiadas.

**Parte VI**

**Apêndices**



# A Uma Proposta de Formalização da Arquitetura Genética da Informação

Apresentamos neste capítulo uma proposta de formalização matemática da *Arquitetura Genética da Informação*, na Teoria dos Conjuntos.

O modelo matemático da arquitetura proposta estabelece um vocabulário notacional com respeito à epistemologia, fenomenologia e às quatro abordagens da Genética da Informação (Clássica, Molecular, Populacional e Quantitativa).

Definem-se, também, matematicamente: contexto, similaridade, fenótipo e genoma. Com base no vocabulário e nas definições, apresentamos os teoremas relativos à similaridade métrica, com suas demonstrações.

Ainda, caracterizamos as propriedades das condições de simetria, dissimilaridade, identidade dos indiscerníveis, positivo definida, discernibilidade e classificabilidade.

## A.1 Modelo Matemático na Teoria dos Conjuntos

Para a fundamentação matemática de nossa teoria dentro da Teoria dos Conjuntos, utilizaremos algumas definições, teoremas, lemas, notas e propriedades, descritos a seguir.

### A.1.1 Definições

**Definição 1.** *Especificação do vocabulário notacional:*

*A1. Com respeito à Epistemologia da Informação, temos:*

(a)  $c = \bar{\varepsilon}(\phi(r), \aleph)$ ; onde a função  $\phi$  ordena  $r$  (objetos definidos em  $A2(a)$ , no momento  $\aleph$  (espaço-tempo estático), por coordenadas intrínsecas ao observador e a função  $\bar{\varepsilon}$  extrai a estrutura desta ordenação, criando o contexto semântico  $c$ .

*A2. Com respeito à Fenomenologia da Informação, temos:*

(a)  $\mathcal{D}_c = \{r_m\}, 0 \leq m \leq n$ , onde  $m \in \mathbb{N}_{\geq 0}$ ,  $r_0 = \emptyset$ ; onde  $\mathcal{D}_c$  é um domínio de discurso que contém  $n$  objetos  $r$  do mundo real em determinado contexto espaço-temporal  $c$ .

(b)  $\mathcal{V}_c = \{v_m\}, 0 \leq m \leq n$ , onde  $m \in \mathbb{N}_{\geq 0}$ ,  $v_0 = \emptyset$ ; onde  $\mathcal{V}_c$  é a representação virtual de  $n$  objetos  $v$  apreendidos por redução fenomenológica em determinado contexto espaço-temporal  $c$ , em  $\mathcal{D}_c$ .

**A3.** Com respeito à Genética Clássica da Informação, temos:

- (a)  $\mathfrak{p}_c = \{p_m\}, 0 \leq m \leq n$ , onde  $m \in \mathbb{N}_{\geq 0}$ ,  $p_0 = \emptyset$ ; onde  $\mathfrak{p}_c$  é o conjunto de  $n$  características fenotípicas  $p$  de um objeto  $r$  de  $\mathfrak{D}_c$ , representado como  $v$  em  $V_c$ .
- (b)  $\mathfrak{p}_c = \mathfrak{p}_c(r)$ , onde  $\mathfrak{p}_c$  representa uma instância das características fenotípicas de  $r$ .

**A4.** Com respeito à Genética Molecular da Informação, temos:

- (a)  $\beta_c = \{b_m\}, 0 \leq m \leq n$ , onde  $m \in \mathbb{N}_{\geq 0}$ ,  $b_0 = \emptyset$ ; onde  $\beta_c$  é o conjunto de  $n$  características comportamentais  $b$ , em um contexto  $c$ , de um objeto  $r$  em  $\mathfrak{D}_c$ , representado como  $v$  em  $V_c$ .

**A5.** Com respeito à Genética Populacional da Informação, temos:

- (a)  $\Delta_c$  é uma função de cálculo da distância de similaridade métrica contextual entre dois objetos  $a$  e  $b$  tal que  $d(a, b) \in \mathbb{R}_{\geq 0}$ ;  $a, b \in V, \mathfrak{D}$  (condição de não-negatividade)
- (b)  $(a \doteq b)$ ; significa que há uma relação de similaridade contextual entre dois objetos  $a$  e  $b$  tal que  $a, b \in V_c, \mathfrak{D}_c$ .
- (c)  $(a \vDash b)$ ; significa que há uma relação de dissimilaridade contextual entre dois objetos  $a$  e  $b$  tal que  $a, b \in V_c, \mathfrak{D}_c$ .

**A6.** Com respeito à Genética Quantitativa da Informação, temos:

- (a) Algoritmos de mineração de dados para a descoberta de conhecimento.

**A7.** Com respeito à Epistemologia da Informação, temos:

- (a) O genoma  $\Gamma$  do contexto  $c$  é caracterizado por um subconjunto de propriedades mensuráveis e dedutíveis, conhecidas ou não,  $\mathfrak{p}_c$  em um conjunto de objetos  $\{r\}$  daquele contexto.
- (b)  $\Lambda_c = (V_c, \Gamma_c, \Delta_c)$  é a Arquitetura da Informação do contexto  $c$ .

*Nota 1.1.* A representação virtual  $v$  de um objeto do mundo real  $r$  é feita pela associação a um vetor  $\mathfrak{p}$  de suas características fenotípicas  $p$  e a um vetor  $\beta$  de comportamentos  $b$ , em determinado contexto espaço-temporal  $c$ .

*Nota 1.2.* Um conjunto de fenótipos  $\mathfrak{p}$  representativo de  $\mathfrak{D}$  só pode ser conhecido dentro do seu contexto  $c$ .

*Nota 1.3.* Para que as características fenotípicas  $\mathfrak{p}$  de um objeto  $r$  em  $\mathfrak{D}$  sejam conhecidas e representadas em  $V$  elas devem ser necessariamente mensuráveis.

*Nota 1.4.* O genoma  $\Gamma$  caracteriza completamente determinado contexto  $c$ , onde subentende-se que todas as características fenotípicas são definidas e as suas relações de distância válidas.

*Nota 1.5.* Na prática, o genoma ideal, livre de contexto, por ser infinito, não existe.



*Nota 1.6.* É extremamente difícil definir o genoma ótimo no contexto  $c$ , pois este depende do conhecimento de características mensuráveis em todos os objetos. Então usamos uma versão  $\gamma$  como uma aproximação do verdadeiro genoma  $\Gamma$ , que é suficiente em termos pragmáticos. Para um genoma ser dito suficiente, sua aplicação deve atender à Definição 1, item A4, sub-itens (b) e (c).

**Definição 2.** *Um contexto específico  $c$  estabelece um conjunto finito de características fenotípicas  $\mathfrak{p}$  para um objeto  $r$ .*

*Nota 2.1.* Livre de contexto, o número de possíveis características mensuráveis  $p$  em  $\mathfrak{p}$  (e derivativos pós-processados destas) pode ser infinito, assim como o número de seus possíveis subconjuntos  $\mathfrak{p}_c$ .

**Definição 3.** *Em todo o espaço  $\mathcal{D}_c$  e  $V_c$  podemos definir um grafo orientado  $G_c(V_c, \Delta_c)$  tal que os nós  $a$  e  $b$  são ligados por arcos com peso atribuído por  $d(a,b)$ , onde  $V$  é o conjunto de vértices ou nós (objetos de  $\mathcal{D}_c$  representados em  $V_c$ ) e  $\Delta_c$  (Definição 1, item A4-a) é a função de distância de similaridade métrica contextual entre pares de objetos.*

**Definição 4.** *O fenótipo  $\mathfrak{p}_c$  de um objeto  $r$  é a instância única de propriedades mensuráveis  $\mathfrak{p}_c$  que caracterizam este objeto no contexto  $c$ .*

*Nota 4.1.* O fenótipo  $\mathfrak{p}$  derivado do genoma  $\Gamma_c$  caracteriza as instâncias  $v$  dos objetos  $r$  no contexto  $c$ .

*Nota 4.2.* A incapacidade de medições precisas de  $\mathfrak{p}$  em  $r$  deve-se ao *princípio de incerteza de Heisenberg*, que diz que a noção de incerteza pode se referir aos seguintes fatores: falta de conhecimento de uma quantidade por um observador; erros experimentais na medida da quantidade; ambiguidades na definição da quantidade; e definição de intervalos em um conjunto de medições em sistemas similares. (<http://plato.stanford.edu/entries/qt-uncertainty/>).

*Nota 4.3.* Devido às considerações das notas anteriores, o espaço virtual é uma representação ruidosa do espaço real, ou seja, sempre vai haver incerteza entre os dois.

*Nota 4.4.* No tocante à veridicalidade da informação (mencionada no item 2.2) e, apesar da nota anterior, o nível de informação no espaço vetorial pode ser considerado suficiente para a compreensão da realidade em termos práticos, ou seja, somos capazes de estabelecer relações de similaridade ou dissimilaridade entre dois objetos quaisquer em um contexto  $c$  utilizando um genoma  $\gamma_c$ , representativo de  $\Gamma_c$ .

### A.1.2 Teoremas, Lemas e Demonstrações

Por construção das definições anteriores, temos os seguintes teoremas:

**Teorema 1.** *Por A4, temos que  $(a \doteq b) \longleftrightarrow d(a, b) = 0$  (similaridade);*

*Demonstração.* Por construção de A5, temos que  $d(a, b)$  descreve uma distância de dissimilaridade entre  $a$  e  $b$ . Ora, em A3 temos que a medida desta distância é feita no conjunto  $\mathfrak{p}_c$  de características fenotípicas dos objetos  $a$  e  $b$ . Assim, fica claro que dois objetos são similares se seus atributos de comparação (características fenotípicas selecionadas) são iguais no mesmo contexto, ou seja, distância igual a zero, o que é expresso por  $(a \doteq b) \longrightarrow d(a, b) = 0$ . O inverso também é verdadeiro, ou seja, sendo a distância igual a zero, podemos concluir que os dois objetos comparados são similares nos atributos (características fenotípicas) selecionados, o que é expresso por  $[d(a, b) = 0 \longrightarrow (a \doteq b)] \implies [(a \doteq b) \longleftrightarrow d(a, b) = 0]$ . □

**Lema 1.1.** *Toda comparação entre objetos  $a$  e  $b$  do mundo real e/ou virtual só pode acontecer em  $V_c$ ;*

*Demonstração.* Pela Definição 2, o contexto  $c$  determina a existência de um conjunto único de características fenotípicas para um objeto. Para a identificação destas características é necessário registrar estas medidas dentro de  $V_c$ . No objeto em  $\mathfrak{D}_c$  estas características fenotípicas são inerentes ao objeto, como essência de sua realidade  $e$ , em  $V_c$ , de forma explícita, ou seja, registrada com base na observação fenomenológica, possibilitando a comparação entre dois objetos quaisquer em  $V_c$  pela distância de similaridade métrica entre eles, dada por  $d(a, b)$ . Assim, como só há registros em  $V_c$ , então, toda comparação feita em registros só pode acontecer em  $V_c$ . □

**Lema 1.2.** *Pelo lema anterior, a distância de similaridade no mundo real  $\mathfrak{D}_c$  é medida no mundo virtual  $V_c$ , em uma relação substitutiva:  $d(r_1, r_2) = d(v_1, v_2)$  e, da mesma maneira, a comparação de um objeto do mundo real  $r_1$  com um objeto do mundo virtual  $v_2$  é dada por  $d(r_1, v_2) = d(v_1, v_2)$ .*

*Demonstração.* Na verdade, a representação de um objeto em  $V_c$  é o conjunto de características fenotípicas que em sua totalidade capturam a essência dele, tal que o objeto em  $\mathfrak{D}_c$  e sua representação em  $V_c$  são indistinguíveis no contexto  $c$ . Logo, por construção temos que  $d(r_1, r_2) = d(v_1, v_2)$ . Os atributos dos objetos do mundo real existem em sua essência independentemente dos registros. Porém, os objetos do mundo real  $\mathfrak{D}_c$  só podem ser comparados em  $V_c$ , porque somente ali é que estão registrados. Isto evidencia que podemos comparar objetos do mundo real entre si em  $V_c$  e objetos do mundo virtual entre si, também em  $V_c$ . Sendo intercambiáveis, também podemos comparar objetos do mundo real com objetos do mundo virtual

e vice-versa, o que é expresso por  $d(r1, v2)$  e  $d(v1, r2)$ , respectivamente.

□

*Até aqui, formalizamos as relações de similaridade. Formalizaremos, posteriormente, as relações de hereditariedade, transmissibilidade da informação.*

### A.1.3 Propriedades

Esta formalização é válida para qualquer domínio do mundo real e seu registro, ou seja, em  $\mathfrak{D}_c$  e em  $V_c$ , respectivamente; porém, a partir deste ponto nos restringiremos ao estudo de distâncias de similaridade métrica entre fenótipos que atendam as seguintes propriedades de similaridade métrica entre objetos.

**Propriedade 1.** A condição de simetria é dada por  $d(v_1, v_2) = d(v_2, v_1)$ .

**Propriedade 2.** A condição de dissimilaridade é dada por  $(a \doteq b) \iff d(a, b) > 0$ ; Isso quer dizer que dois objetos

**Propriedade 3.** A condição de identidade dos indiscerníveis é dada por  $(a \doteq b) \iff \mathfrak{p}_c(a) = \mathfrak{p}_c(b) \iff (a = b)$ ;  $a, b \in V_c, \mathfrak{D}_c$ . Isso quer dizer que dois objetos são considerados iguais se tiverem os mesmos registros fenotípicos em um dado contexto, ou seja, em  $V_c$ , apesar de serem essencialmente diferentes no mundo real, em  $\mathfrak{D}_c$ , o que dá base para a formação de conjuntos. Todo conjunto existente é formado de elementos indiscerníveis em algum contexto. Isso também caracteriza a *condição de fungibilidade*, permitindo a permutação entre os objetos no espaço e, também, no tempo (Ex.: mercado de futuros, câmbio de moedas etc).

**Propriedade 4.** A condição de positivo definida ocorre quando  $\Delta$  satisfaz as condições de não-negatividade e de identidade dos indiscerníveis. Isso quer dizer que a distância mínima entre dois objetos está na sua igualdade e, no máximo, representa a sua dissimilaridade e é sem limite.

**Propriedade 5.** A condição de discernibilidade é dada por  $(a \doteq b) \iff \mathfrak{p}_c(a) \neq \mathfrak{p}_c(b) \iff (a \neq b)$ ;  $a, b \in V_c, \mathfrak{D}_c$ . Isso quer dizer que existe sempre um contexto  $c$  em que dois objetos sejam infinitamente discerníveis em suas características fenotípicas.

**Propriedade 6.** A condição de classificabilidade é suficiente em  $\mathfrak{B}_c$  e  $\mathfrak{p}_c$ .



## B Diagramas da Ontologia da Arquitetura Genética da Informação

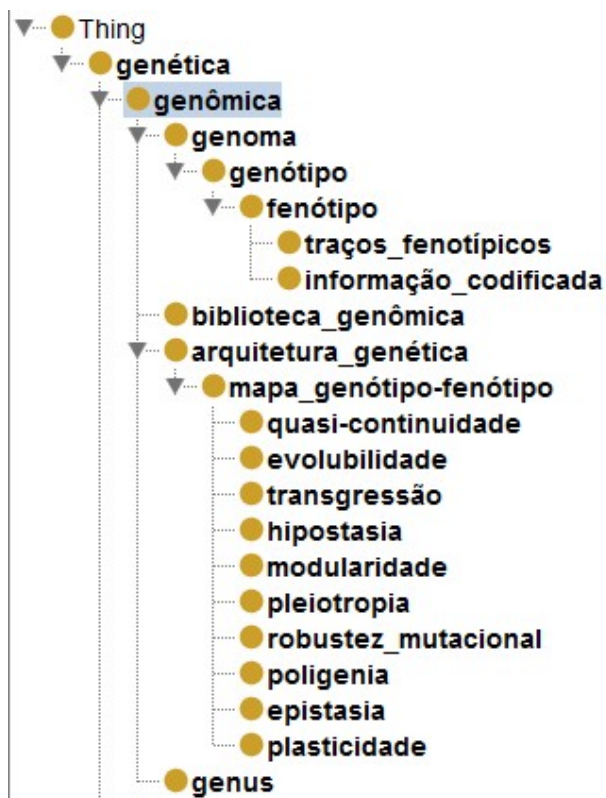


Figura 68: OWL - Genômica - Fonte: Produzido pelo autor

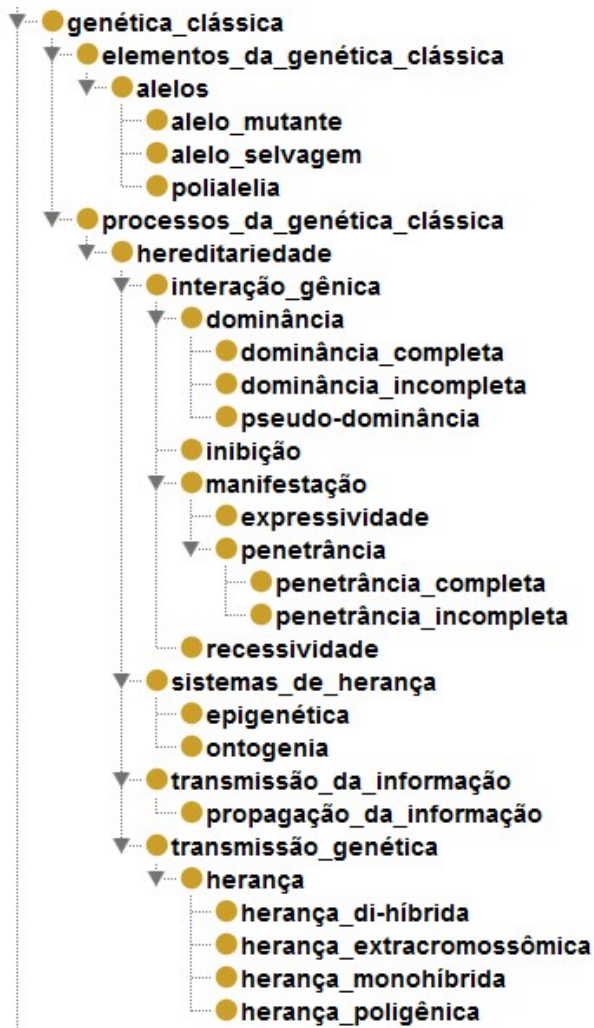


Figura 69: OWL - Genética Clássica - Fonte: Produzido pelo autor

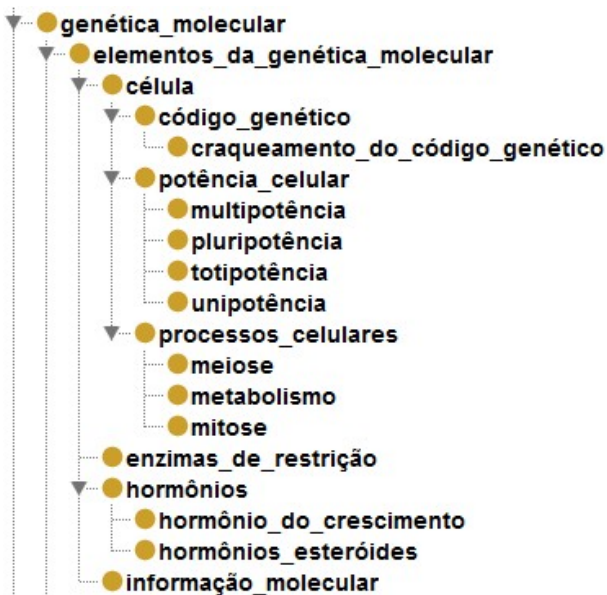


Figura 70: OWL - Genética Molecular (I) - Fonte: Produzido pelo autor

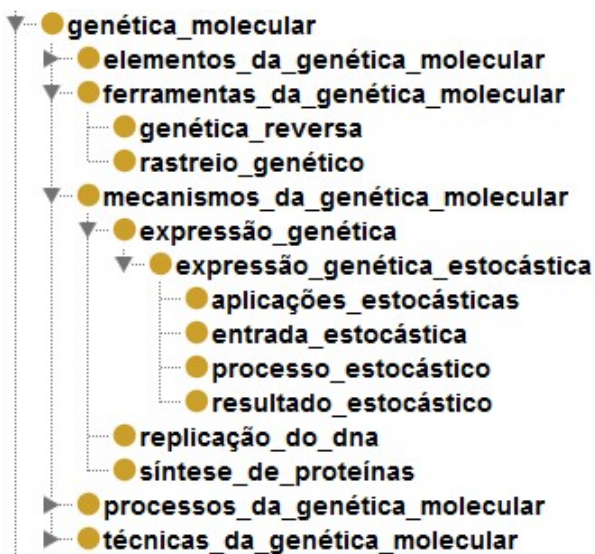


Figura 71: OWL - Genética Molecular (II) - Fonte: Produzido pelo autor



Figura 72: OWL - Genética Molecular (III) - Fonte: Produzido pelo autor

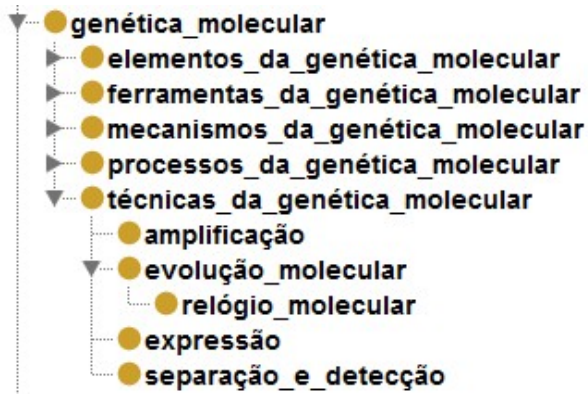


Figura 73: OWL - Genética Molecular (IV) - Fonte: Produzido pelo autor

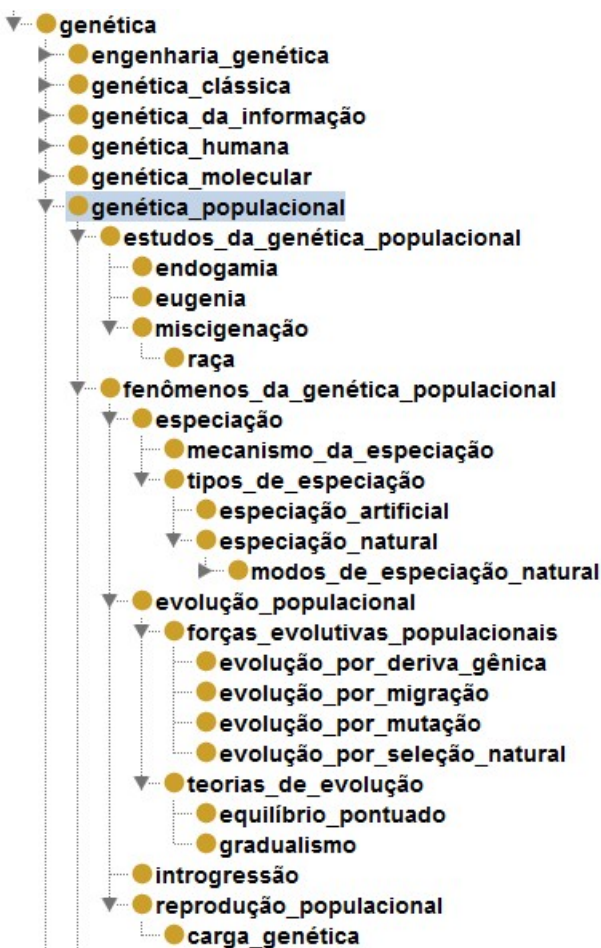


Figura 74: OWL - Genética Populacional (I) - Fonte: Produzido pelo autor



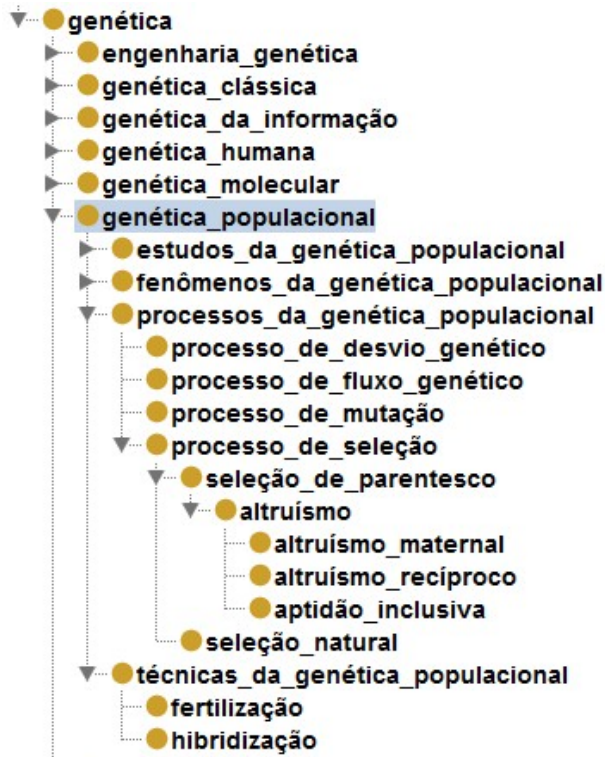


Figura 75: OWL - Genética Populacional (II) - Fonte: Produzido pelo autor

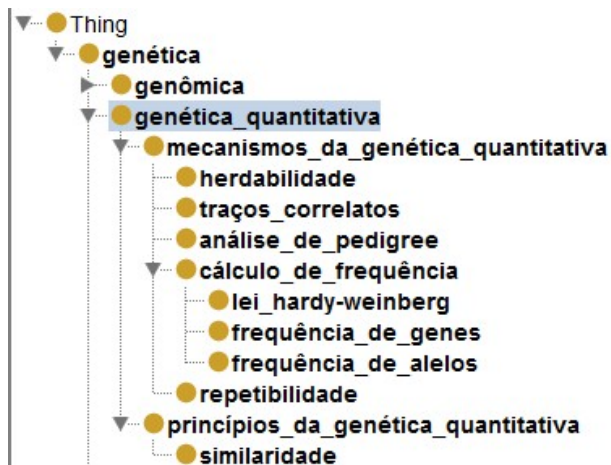


Figura 76: OWL - Genética Quantitativa - Fonte: Produzido pelo autor



Figura 77: OWL - Genética da Informação Biológica (I) - Fonte: Produzido pelo autor



Figura 78: OWL - Genética da Informação Biológica (II) - Fonte: Produzido pelo autor



Figura 79: OWL - Genética da Informação Biológica (III) - Fonte: Produzido pelo autor



Figura 80: OWL - Genética da Informação Biológica (IV) - Fonte: Produzido pelo autor



Figura 81: OWL - Genética da Informação Biológica (V) - Fonte: Produzido pelo autor

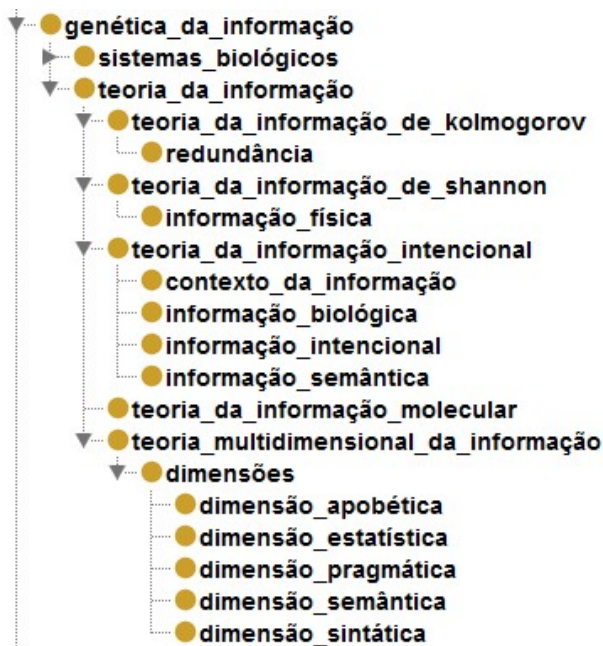


Figura 82: OWL - Genética da Informação Biológica (VI) - Fonte: Produzido pelo autor

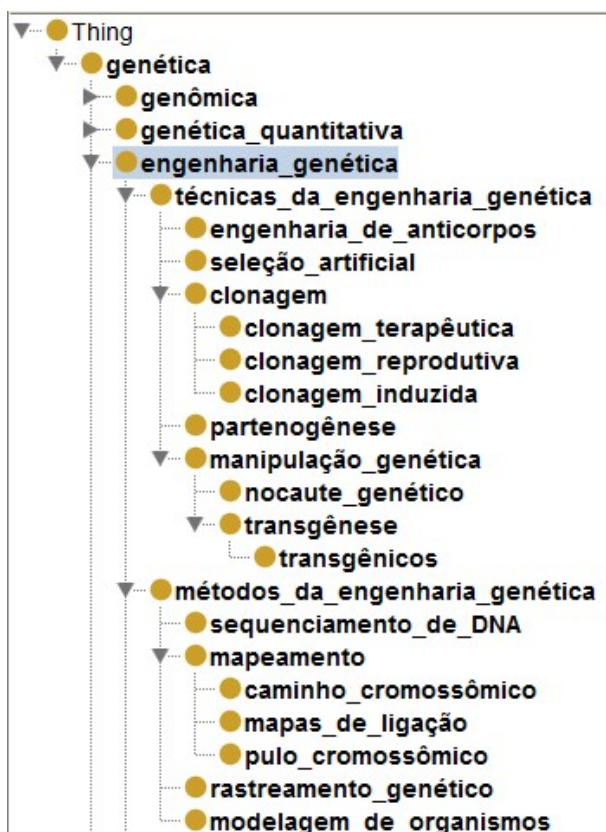


Figura 83: OWL - Engenharia Genética (I) - Fonte: Produzido pelo autor

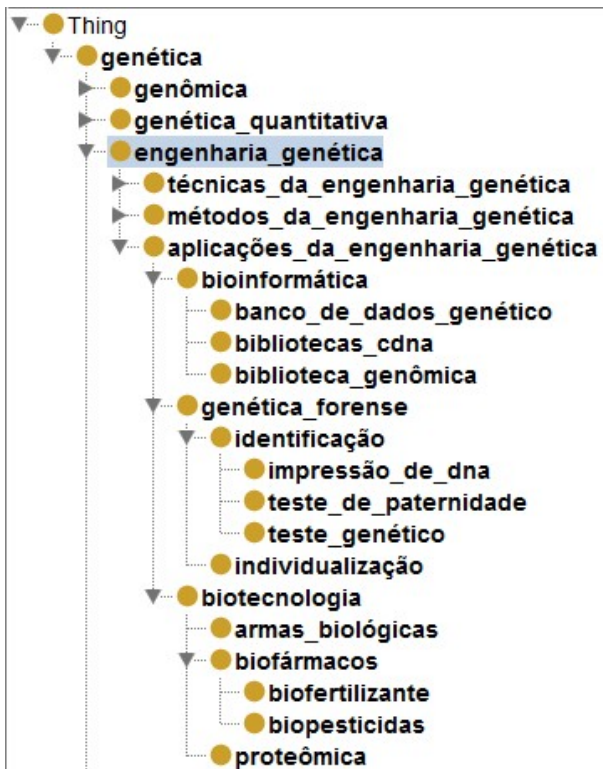


Figura 84: OWL - Engenharia Genética (II) - Fonte: Produzido pelo autor

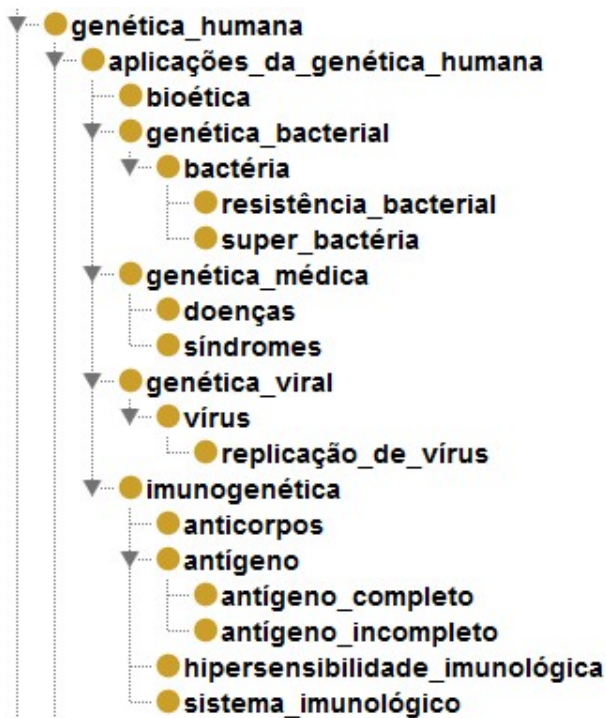


Figura 85: OWL - Genética Humana (I) - Fonte: Produzido pelo autor

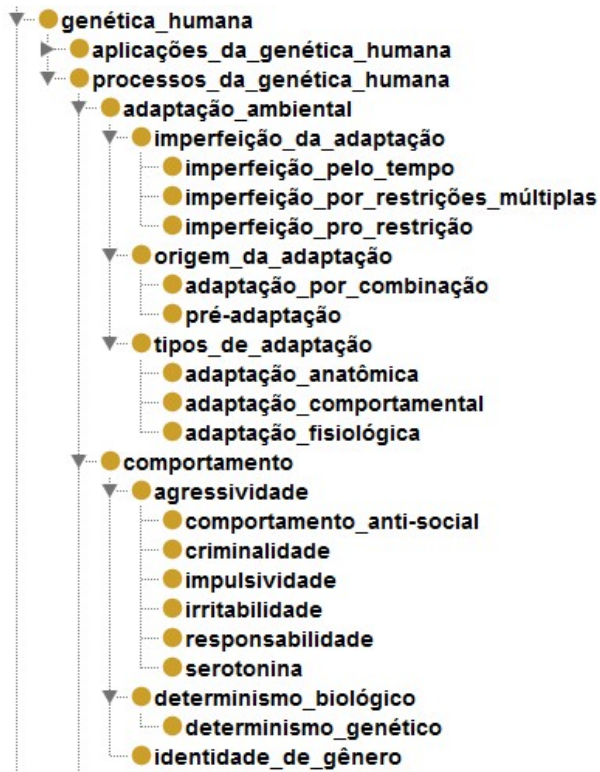


Figura 86: OWL - Genética Humana (II) - Fonte: Produzido pelo autor





## C Rede de Petri de Modelagem da Gênese da Informação Intencional

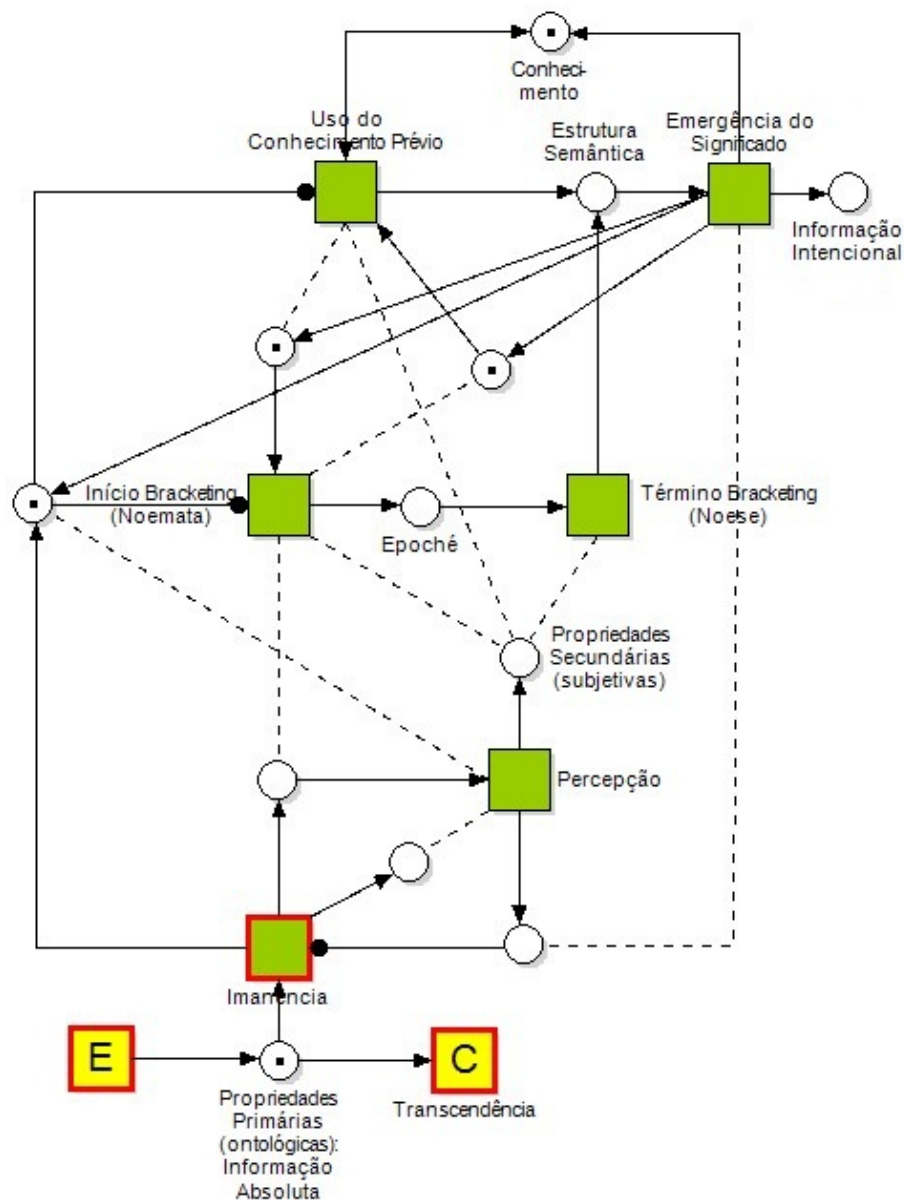


Figura 87: Imanência ou Transcendência? - Fonte: Produzido pelo autor



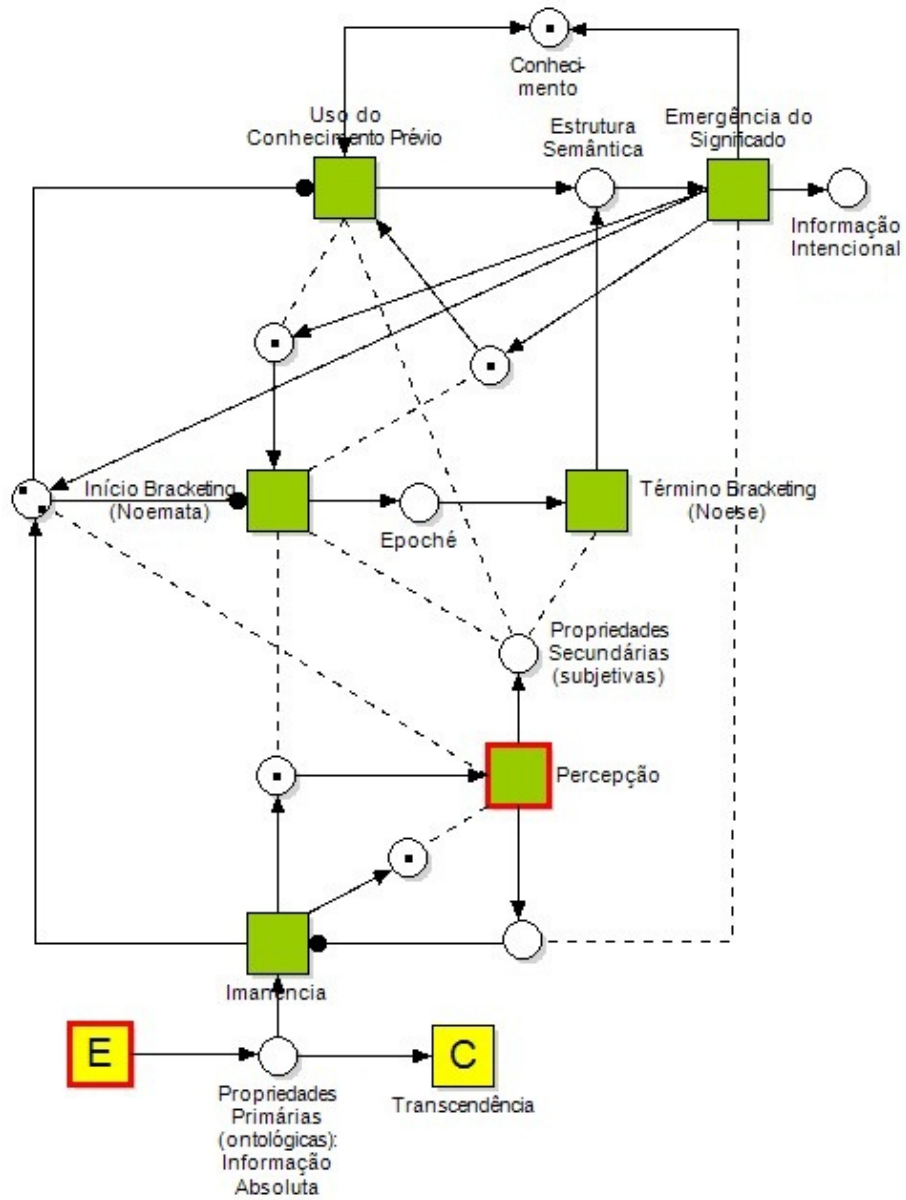


Figura 89: Imanência executada - Fonte: Produzido pelo autor

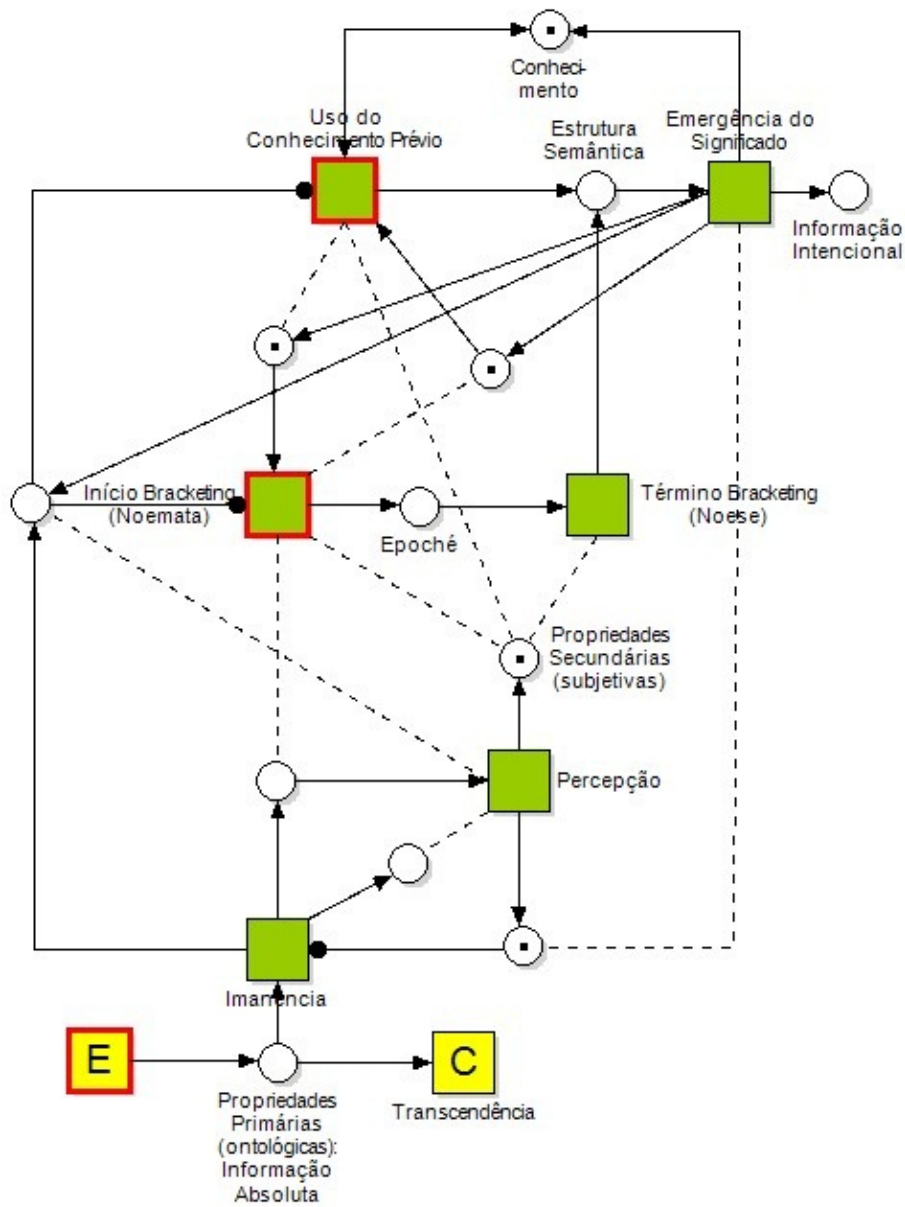


Figura 90: Percepção executada



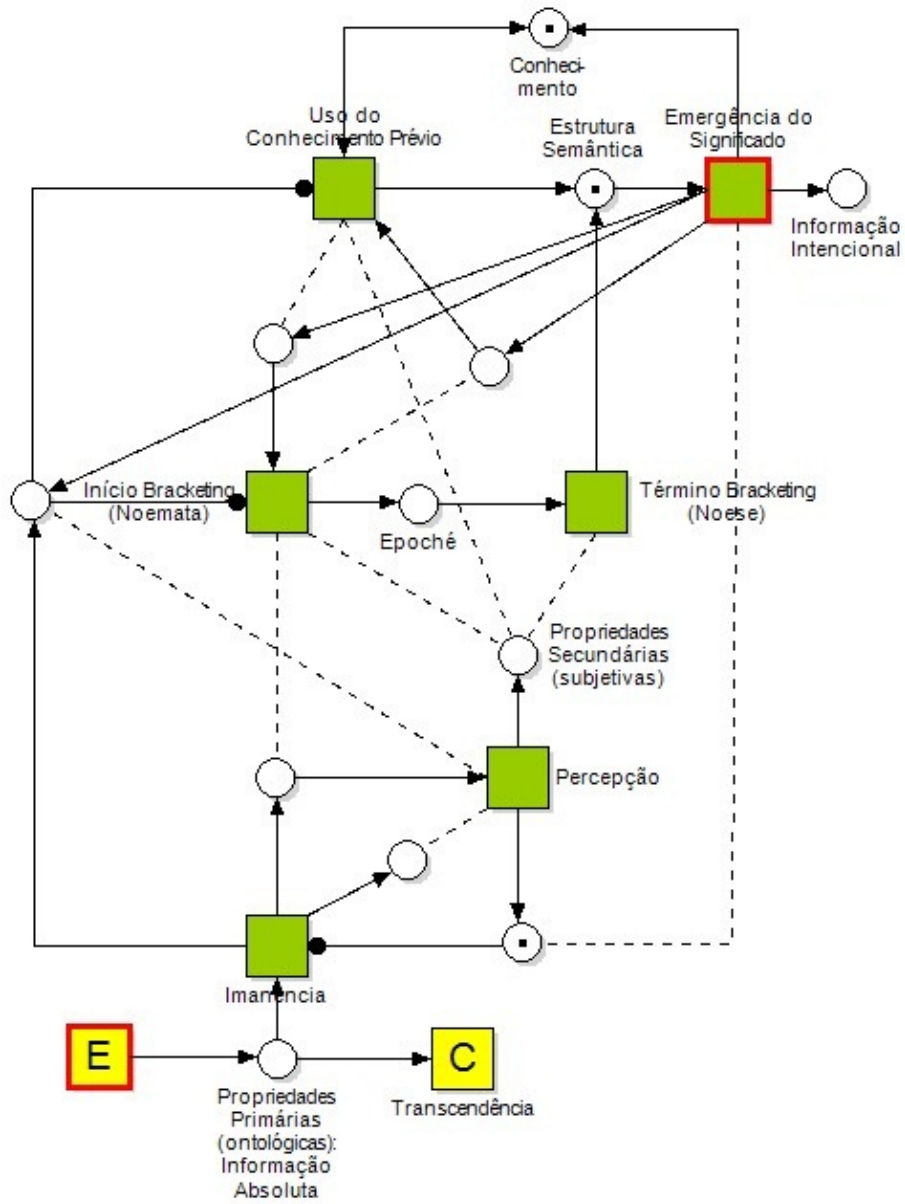


Figura 92: Término do Bracketing: Noese executada - Fonte: Produzido pelo autor

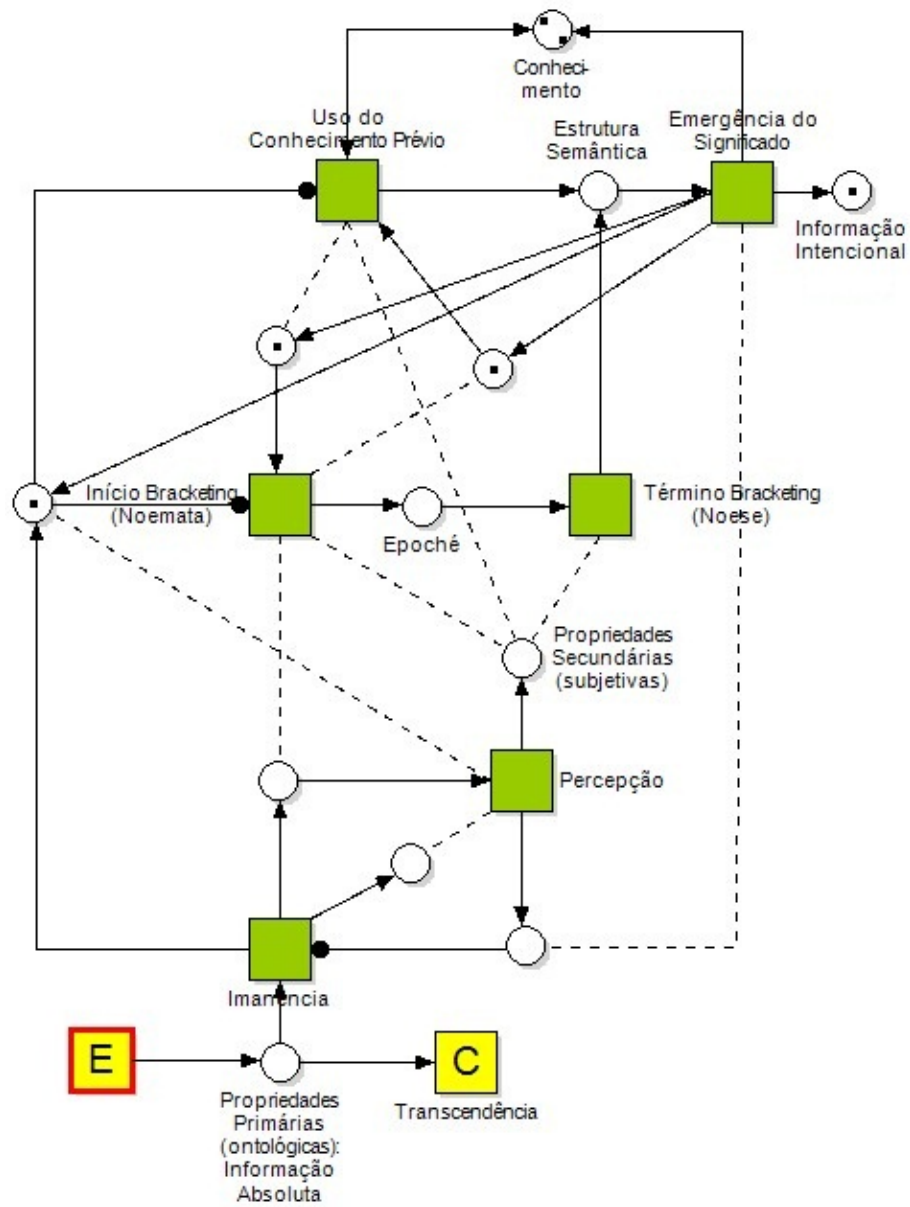


Figura 93: Emergência do significado executada - Fonte: Produzido pelo autor





## **D Produção Científica - 2008-2012**

# PRODUÇÃO CIENTÍFICA – 2008-2012

Romualdo Alves Pereira Júnior

---

---

## Society for Phenomenology and Media 14th Annual International Conference

Paper: **The Genesis of Information in the Phenomenological Epoche**

Romualdo Alves Pereira Júnior (AEB)

Mamede-Lima Marques (CPAI/UnB)

Manoel Fernando da Mota Tenorio (CPAI/UnB)

*February, 16, 17 and 18, 2012 (La Jolla – California)*



**SpaceOps 2012**

**11-15 June in Stockholm, Sweden**

**Paper: “Improving Satellite Data Archiving Facility for Environmental R&D Purposes Based on Architecture of Information Approach.”**

- Romualdo Alves Pereira Júnior (AEB)
- Maria de Fátima Mattiello-Francisco (INPE)
- Marcos Aurelio Ferreira dos Santos (INPE)
- Mamede Lima-Marques (CPAI/UnB)

<http://www.spaceops2012.org/proceedings/documents/id1295180-Paper-003.pdf>

---

**INTED 2012**

5th - 7th March  
Valencia - Spain 

INTED2012 - INTERNATIONAL TECHNOLOGY, EDUCATION AND DEVELOPMENT CONFERENCE

This is to certify that:

**Romualdo Pereira Jr.**

has been a member of the International Advisory Board at  
the *International Technology, Education and Development Conference*  
held in Valencia, Spain, 5th-7th of March 2012.



INTED2012, Local Organizing Committee  
5/3/2012



Valencia, 4th of September, 2012

### **ABSTRACT ACCEPTANCE LETTER**

This is a confirmation that the abstract entitled:

"A PHENOMENOLOGICAL ARCHITECTURE OF INFORMATION APPROACH TO  
ONLINE EDUCATION"

Author(s): Romualdo Alves Pereira Jr., Daniel Oliveira, Elvis D. Moreno Seimetz,  
Mamede Lima-Marques

has been accepted as ORAL presentation at ICERI2012.

**Name of event:** ICERI2012 (5th International Conference of Education, Research and  
Innovation)

**Dates:** 19th - 21st of November, 2012

**Place:** Madrid (Spain)

**Organising entity:** IATED



**IATED**  
*International Association of  
Technology, Education and  
Development*

ICERI2012 Local Organising Committee

---



### Oral presentations:

THE VIRTUAL SCHOOL PROJECT FOR THE BRAZILIAN SPACE PROGRAM  
*R. Alves Pereira Jr., E. Nascimento, L. Fernandes, M. Lima-Marques*

### Poster presentations:

HOME-TEACHERS - A DEFINITE PORTAL FOR PERSONAL AND GROUP LEARNING OPENED TO ALL KNOWLEDGE DOMAINS  
*R. Pereira Jr., E. Alves Nascimento, F. Hideo Mikami*



Ministério da  
Agricultura, Pecuária  
e Abastecimento



## APRESENTAÇÃO

### **SISTEMA EXPERTISE DE INTELIGÊNCIA COMPETITIVA PARA GRUPOS COLABORATIVOS**

- Romualdo Alves Pereira Júnior – Agência Espacial Brasileira (AEB)
- Luis Gabriel de Melo Fernandes – Agência Espacial Brasileira (AEB)
- Manoel Fernando da Mota Tenorio – Universidade de Brasília (CPAI/UnB)
- Mamede Lima-Marques – Universidade de Brasília (CPAI/UnB)
- Joshua Emanuel Tenorio – Agência Espacial Brasileira (AEB)



Data: 02/12/2011

Palestra: "Uma Teoria da Gênese e Genética da Informação – Aspectos Epistemológicos, Fenomenológicos e Ontológicos".

Autores: Romualdo Alves Pereira Júnior e Mamede Lima-Marques

---



Data: 16/11/2011

Título: "Paragidma Genético da Arquitetura da Informação"

Autores: Romualdo Alves Pereira Júnior e Mamede Lima-Marques

---

**Orientador: Romualdo Alves Pereira Júnior**

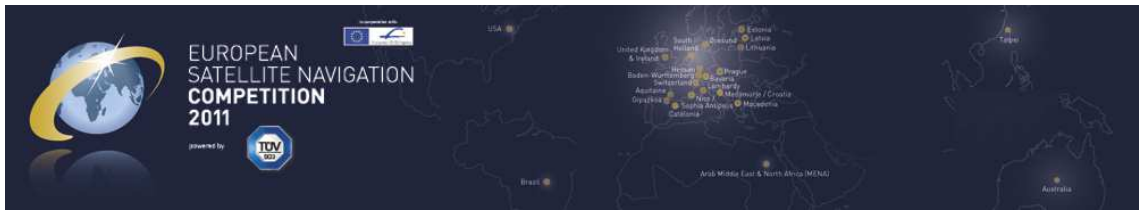
1. **Sistema de Gestão de Bagagens em Aeroportos do Grupo I do Brasil**  
(Marcelo Augusto da Cruz Motta – 11/2012)
  2. **Brasília Smart City**  
(Leonardo Ramiro de Oliveira – 11/2012)
  3. **Sistema de Conversão de Dados Brutos das Plataformas de Coleta de Dados do Sistema Nacional de Dados Ambientais – SINDA**  
(Thiago Serra – 11/2012)
  4. **Gestão de Relacionamento de Clientes (CRM) da Caixa Econômica Federal - CEF**  
(Weber Stevan Roder Kai – 11/2012)
  5. **Gestão Ágil de Culturas do Conhecimento na Empresa Brasileira de Comunicação - EBC**  
(Joilson Marques – 11/2012)
  6. **CrowdScience Scientia 2.0**  
(Evandro Maraschin – 11/2012)
  7. **Business Intelligence – UnB em Números**  
(Paulo Sérgio Pires – 06/2011)
  8. **Implementação de Business Intelligence para a Análise de Planos de Vôo Repetitivos do Aeroporto Internacional de Brasília**  
(Gustavo Henrique Pimentel Freitas – 08/2010)
  9. **Análise de Desempenho de Grid Computing com a Solução BOINC**  
(Rodrigo Ferreira Ruiz – 08/2010)
  10. **Utilização do Market Basket Analysis no Processo de Tomada de Decisões**  
(Guilherme Coelho Rodrigues – 04/2010)
  11. **Aplicação de Algoritmo Genético para Solução de Sistemas de Equação Não-Lineares**  
(Luís Gabriel de Melo Fernandes – 01/2010)
  12. **Uso de Redes de Petri para Modelagem do Sistema de Transporte Urbano do DF**  
(Daiana Nascimento Teixeira de Oliveira – 01/2009)
  13. **Mineração de Dados: Uma Aplicação na Base de Dados de Artigos de Periódicos Científicos das Áreas de Informação – ABCDM**  
(Held Barbosa de Souza e Daniela Leite Naglis – 11/2008)
  14. **Web Semântica para Sistemas de Redução da Demanda, Oferta e dos Danos Associados ao Uso de Drogas**  
(Paulo Cesar Saturno Corrêa e Rafael Valentim Fonseca -11/2008)
  15. **Rede Neural para Assimilação de Dados Meteorológicos**  
(Thiago Tavares Braga – 10/2008)
-

- 
16. **Projeto do Sistema Integrado de Controle Cambial do Banco Nacional de Angola**  
(Zorobabel Teixeira Cavalcante – 10/2008)
  17. **Aplicando BPM para Mapear Processos da Divisão de Atendimento ao Usuário do Gabinete do Ministro do Ministério do Planejamento**  
(Roberto Ferreira Irmão e Bruno Cardoso – 10/2008)
  18. **Etiquetas Inteligentes: Estado da Arte e Questões Éticas**  
(Jany Dantas de Almeida – 01/2008)

#### **Co-Orientador: Romualdo Alves Pereira Júnior**

1. Dissertação de Mestrado em Gestão de Segurança da Informação (UnB): **“Utilização da Computação Distribuída para o Armazenamento e Indexação de Dados Forenses”**  
(Marcelo Antônio da Silva – defesa prevista para 02/2012)
  2. Monografia de Especialização em Gestão de TI (UnB): **Método de Pré-Qualificação de Sistemas de Informação Geográfica para a Agência Nacional de Telecomunicações – ANATEL** (João Bosco Vasconcelos de Souza – defesa realizada em 04/2011)
-





- **ACCEPTED PAPERS**
- **Manoel Fernando da Mota Tenorio- University of Brasilia – CPAI/UnB**  
**Romualdo Alves Pereira Júnior – Brazilian Space Agency - AEB**  
**Mamede Lima-Marques – University of Brasilia – CPAI/UnB**

| ID     | Title   | Region                   | Special Prize                                | Status    |
|--------|---|--------------------------|--|-----------|
| 111412 | GNSS as a complementary instrument to automated flight systems            | United Kingdom & Ireland | ESA: ESA Innovation Prize                    | Published |
| 111410 | Use of LBS for monitoring, maintenance and fuel control in car fleets     | Brazil                   | NAVTEQ: Mobile Enterprise Solutions          | Published |
| 111409 | Complete Management and Monitoring of Forest Resources for Sustainability | USA                      | University Challenge/Location Based Services | Published |



## ACCEPTED PAPERS

>> **Utilização da Computação Distribuída para o Armazenamento e Indexação de Dados Forenses**

**Marcelo Antônio da Silva – Brazilian Federal Police, and  
 Romualdo Alves Pereira Júnior – Brazilian Space Agency**



## Presentations

### **ForeIndex: A Framework for Analysis and Triage of Data Forensics**

- **Marcelo Silva**, *Forensic Expert of Brazilian Federal Police*, and
  - **Romualdo Pereira**, *Researcher of the Brazilian Space Agency*
- 

## ICT<sub>in</sub>Forensics

### *Accepted Papers (2011)*

A Distributed System for Storage and Indexing Forensics Data

**Authors** : Marcelo Silva, Romualdo Pereira .



#### **DIA 12/08/2010 (QUINTA-FEIRA)**

- **17:00h – 18:00h**
  - **Romualdo Alves Pereira Júnior** (Projeto do Sistema Nacional de Dados Ambientais do INPE)

#### **DIA 13/08/2010 (SEXTA-FEIRA)**

- **08:00h – 09:55h**
    - **Daniel Luchetta** (Configuração do Servidor de Rede do CPAI)
    - **Ricardo Kehrlé Miranda** (Portais Colaborativos)
    - **Eduardo A. Nascimento** (Businesss Intelligence)
    - **Luís Gabriel M. Fernandes** (Sala de Situação e Fatos Comentados)
    - **Igor Pessoa Rocha** (PLN para o Programa Nacional de Atividades Espaciais - PNAE)
    - **Cristiane Alves Wanderley** (Metodologias Ágeis de Projetos e Documentação)
  - **10:30h – 11:55h**
    - **Dino Macedo Amaral** (Grid Computing CPAI-AEB – Solução Hadoop)
    - **Marcelo Holtz** (Grid Computing CPAI-AEB – Análise de Desempenho)
    - **Vinícius M Alves** (Grid Computing CPAI-AEB – Configuração da Rede)
    - **Pedro Ferreira Andrade** (Podcasting em áudio e vídeo do Programa Espacial Brasileiro, com tecnologia RIA)
    - **Laiana Emanuelle Corrêa** (Gestão de Projetos com ProjectOpen)
-

## 3º Colóquio de Arquitetura da Informação

---

- **Manoel Fernando da Mota Tenorio** (Aplicações de Otimização com Restrições Utilizando Programação Lógica)
  - **Daniel Oliveira Júnior** (Arquitetura da Informação de Redes Sociais que atendam às necessidades da Olimpíada Brasileira de Astronomia - OBA)
  - **Romualdo Alves Pereira Júnior** (Orientador)
    - **Evandro Maraschin** (Parcerias Estratégicas – Crowd Sourcing and Crowd Funding)
    - **Gustavo Henrique Freitas** (Business Intelligence para Aeronavegação)
    - **Thiago de Matos e Eduardo Alves do Nascimento** (Business Intelligence – UnB em Números)
    - **Daniel Luchetta** (A Infra-estrutura do Portal do CPAI)
    - **Marcelo Antônio da Silva** (Um Sistema Distribuído para o Armazenamento e Indexação de Dados Forenses)
    - **Kaio Rocha** (SICSTUS Prolog e Busca com Restrições)
- 
-



# Índice

- Aleatoriedade, 96
- Ambiente
  - Influência do, 102
- Analogia, 123
- Arquiteto da Informação, 73
- Arquitetura da Informação, 41
  - Disciplina da, 18, 68
  - Linha do Tempo, 41
  - Teoria Geral da, 10, 41
- Arquitetura da Informação Intencional, 18, 73
  - Contexto, 18
  - Forma, 18
  - Manifestação, 18
  - Significado, 18
- Arquitetura Genética, 91
- Arquitetura Genética da Informação, 67
  - Elementos da, 68
  - Ontologia da, 109
- Automação, 13
- Blocos de construção, 96
- Bracketing, 12, 82
- Brentano, 27
- Complexidade, 13, 27, 28, 36, 39, 41, 58, 60, 76, 90, 96, 104, 110
- Comportamento, 12, 106
- Comunicação, 36, 87
- Conceitos, 49
  - Perceptivos, 13
- Conhecimento, 10, 75
  - Organização do, 44
  - Representação do, 44, 47
- Consciência, 12, 68
  - Perceptiva, 13
- Contexto, 25, 36, 39–41, 44, 49, 60, 67, 68, 73, 76, 78, 80, 82, 95, 96, 102, 106, 110, 123, 150
  - Independência de, 110
  - Irrelevância do, 87
  - Social, 28
- Dados, 10, 75, 128
- Descoberta de Literatura Conexa, 133
- DIKW, 75
- DNA, 102
- Epoché, 82
- Estrutura Evolucionária, 96
- Evolução
  - Tecnológica, 159
- Falácia Fenomenológica, 13
- FCMS, 18, 68
- Femótipo, 100
- Fenótipo, 59, 68, 73, 102
- Fenômeno, 13
  - Intencional, 78
- Fenomenologia, 10, 25
  - Constitutiva
    - Naturalística, 10
    - Transcendental, 10
- Desenvolvimento Histórico-Conceitual da, 26
- Existencial, 10
- Fundador da, 28
- Genética, 10, 67
- Hermenêutica, 10
- Historicista
  - Generativa, 10

- Precusores da, 27  
 Realística, 10  
 Filomemética, 159  
 Filosofia da Informação, 127  
   Problemas Abertos em, 127  
 Forma, 68, 73  
 Gênese  
   Ativa, 10  
   Passiva, 10  
 Genética, 57  
   Clássica, 1, 87, 110, 113  
   Fundamentos da, 57  
   Molecular, 1, 87, 110, 113  
   Populacional, 1, 35, 87, 99, 112–114  
   Quantitativa, 156  
 Genética da Informação, 87, 95  
   Aplicações, 133  
     Estrutura Retórica de Documentos, 150  
     Evolução Artificial, 147  
     Evolução Natural, 147  
     Genoma Musical, 149  
     Indicador DNA Brasil, 156  
     Redes de Petri, 154  
     Similaridade Fenotípica, 133  
     SINDA, 159  
   Clássica, 110  
   Molecular, 110  
   Populacional, 112  
   Quantitativa, 112  
 Genótipo, 58, 68, 73, 102  
 Genes, 100, 102  
 Genoma, 58, 68, 73  
 Husserl, 28  
 Imanência, 12  
 Informação, 10, 75, 127  
   Busca da, 104  
   Causal, 13, 78  
   Constituída, 78, 95  
   Dinâmica da, 128  
   e Verdade, 129  
   em Sistemas Biológicos, 118  
   Entropia da, 36  
   Estrutural, 68  
   Expressão da, 102  
   Leitura da, 102  
   Metáforas da, 87  
   Natureza Ontológica da, 23  
   Paradigmas Epistemológicos da, 35  
   Pragmática, 13  
   Propagação da, 78  
   Recuperação da, 104  
   Seletiva, 68  
   Semântica, 78  
   Significativa, 13, 78  
   Teorias da, 87  
 Informação Biológica, 60, 113  
   Armazenamento da, 93  
   e Aleatoriedade, 98  
   e Aprendizado, 97  
   e Complexidade, 99  
   e Evolução, 99  
   e Imunização, 97  
   e Redes Booleanas, 99  
   e Seleção Natural, 96  
   e Semântica, 96  
   e Sintaxe, 96  
   Processamento da, 94  
 Informação Intencional, 1, 7, 10, 12, 13, 18, 20, 23, 24, 36, 60, 67, 68, 73, 75, 76, 78, 80, 82, 87, 91, 94, 95, 109, 110, 123, 185  
   Gênese da, 75, 80, 82  
   O Fenômeno da, 76  
 Inovação  
   Tecnológica, 159  
 Intencionalidade, 10, 12, 68

- Conteúdo da, 12
- Kant, 27
- Kolmogorov, 90
- Lógica
  - Descritiva, 49
- Lambert, 27
- LRD, 133
- Manifestação, 68, 73
- Mapeamento Genótipo-Fenótipo, 91
- Memética, 100, 159
- Memes, 100
- Mineração
  - de Dados, 51, 112
  - Textual, 114, 119, 133, 144
- Mutação, 96
- Noema, 80
- Noese, 80
- Ontologias, 47
  - Basic Formal Ontology, 50
  - BFO, 50
  - Biológicas, 54, 55
  - BioTop Ontology, 56
  - da Arquitetura Genética da Informação, 109
  - de Domínio, 48
  - Formais, 49
  - Fundamentais, 48
  - Método de Construção de, 51, 112, 113, 118
  - Mesclagem das, 123
  - Não-Formais, 49
  - Operações em, 53
  - OWL, 56
  - Web Ontology Language, 56
- Paisagem
  - Adaptativa, 96
  - Fenotípica, 60
- Paradigma
  - Cognitivo da Informação, 39
  - Físico da Informação, 36, 87
  - Histórico-Social da Informação, 40
- PLN, 150
- Programação Evolutiva, 102
  - e Complexidade, 102
- Programação Genética, 100
  - e Blocos de Construção, 101
  - e Complexidade, 101
  - e Fenótipo, 101
  - e Genótipo, 101
  - e Introns, 101
- Qualia, 68
- Redes
  - de Petri, 82, 154
- Redundância, 90
- Regulação Genética, 102
- Retórica, 150
- RNA, 102
- RST, 150
- Ruído
  - Informativo, 96
- Síntese
  - Passiva, 10
- Seleção
  - Natural, 13, 76, 87
- Shannon, 87
- Significado, 12, 68, 73, 80
  - Irrelevância do, 36, 87
- Similaridade, 104
  - Auto-, 106
  - Características da, 105
  - Fenotípica, 107
  - Molecular, 106
- Subjetividade, 13

Teleonomia, 78

Teoria

da Evolução, 95

da Classificação, 44

da Evolução, 78

Geral da Arquitetura da Informação, 10,  
41

Geral da Informação, 13

Geral dos Sistemas, 13

Matemática da Comunicação, 87

Teoria da Informação

Algorítmica, 91

de Kolmogorov, 90

de Shannon, 87

Tese

Fundamentos, 23

Introdução, 1

Metodologia, 7

Classificação da Pesquisa, 7

Percurso Metodológico, 7

Objetivos, 5

Específicos, 5

Gerais, 5

Organização da, 1

Pressupostos Básicos, 10

Problemática, 6

Requisitos Pré-Pesquisa, 5

Revisão de Literatura, 23

TMC, 36

Tradução, 102

Transcendência, 12

Transcrição, 102

Visão de Mundo, 20