Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Autenticação Semântica Gráfica

Leonardo dos Santos Dourado

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientador
Prof. Dr. Edison Ishikawa

Brasília
2021

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

DD739a

Dourado, Leonardo dos Santos
Autenticação Semântica Gráfica / Leonardo dos Santos Dourado; orientador Edison Ishikawa. -- Brasília, 2021.
99 p.

Dissertação (Mestrado - Mestrado Profissional em Computação Aplicada) -- Universidade de Brasília, 2021.

1. Autenticação Semântica Gráfica. 2. Autenticação Semântica Humana. 3. Autenticação Gráfica. 4. Lógica Descritiva. I. Ishikawa, Edison, orient. II. Título.

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Autenticação Semântica Gráfica

Leonardo dos Santos Dourado

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Prof. Dr. Edison Ishikawa (Orientador)
CIC/UnB

Prof. Dr. Anderson Fernandes Pereira dos Santos
Instituto Militar de Engenharia

Prof. Dr. João José Costa Gondim
Universidade de Brasília

Prof. Dr. Marcelo Ladeira
Coordenador do Programa de Pós-graduação em Computação Aplicada

Brasília, 28 de janeiro de 2021

Dedicatória

Dedico este trabalho aos meus pais, a minha esposa e a toda minha família, que no decorrer da minha vida não mediram esforços para que eu chegasse até esta etapa.

Agradecimentos

Agradeço a minha família pelo incentivo e pela compreensão nos diversos momentos que deveríamos estar juntos, porém tive que abdicar destes momentos em prol do desenvolvimento deste trabalho.

Agradeço ao programa de pós-graduação em computação aplicada da Universidade de Brasília e aos professores que em algum momento fizeram parte desta minha jornada.

Gostaria de registrar um agradecimento especial ao meu orientador Prof. Dr. Edison Ishikawa, pelo apoio, paciência e por compartilhar parte do seu vasto conhecimento em prol do desenvolvimento deste trabalho, pois sem esse auxílio e orientação, esse trabalho não teria sido possível.

Por último, mas não menos importante, gostaria de agradecer ao meu Coordenador-Geral na época da inscrição no programa de mestrado Moisés Henrique Castro da Silva e ao Diretor de Tecnologia na época Edvaldo Noleto Perna Filho, pelo apoio neste desafio e aos colegas da Diretoria de Tecnologia que de alguma forma contribuíram para que eu conseguisse desenvolver esse trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

A autenticação utilizando somente usuário e senha atualmente não é suficiente para garantir um nível de segurança aceitável em um sistema de informação. Para reforçar a segurança no processo de autenticação estão sendo utilizados múltiplos fatores, contudo o fator do tipo que o usuário possui gera inconveniente para o usuário, pois necessitará de algum dispositivo auxiliar sempre que for autenticar no sistema. Os fatores biométricos podem mudar com o passar do tempo, necessitam de dispositivos auxiliares que podem aumentar o custo e também podem depender de aspectos ambientais para funcionar corretamente. Visando a criação de um sistema de autenticação complementar para a autenticação com usuário e senha e eliminar alguns inconvenientes da autenticação por múltiplos fatores, este trabalho propõe a autenticação por meio de relações semânticas, por meio de lógica descritiva, de conceitos identificáveis em imagens como forma de aumentar a segurança no processo de autenticação. Uma prova de conceito desse método de autenticação foi implementada e os resultados obtidos demonstram que o sistema de autenticação complementar proposto é robusto e fácil de usar.

Palavras-chave: Autenticação Semântica Gráfica, Autenticação Semântica Humana, Autenticação Gráfica e Lógica Descritiva.

Abstract

Authentication on systems using only an authentication method based on username and password is not enough to ensure an acceptable level of information security to a information system. To fortify the safety during the authentication process multi factors have been used, however factor as *what you have* creates inconveniences for the users, because the users during the authentication process always will need to have a device that complements the authentication process in their possession. Biometric factors might change during the time, they need an auxiliary device that might increase the costs and also it might depend from environmental conditions to work appropriately. Intending to create an complementary authentication system to fortify user and password authentication and to solve some inconveniences that there are in the multi factors authentication systems, this work purposes the authentication using semantic representations of concepts identifiable in images using descriptive logic as a form to increase the safety during the authentication process. A proof of concept was implemented and the resulting complementary authentication system proved to be robust and easy to use.

Keywords: Graphical Semantic Authentication, Human Semantic Authentication, Graphic Authentication and Description Logic.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Autenticação de um usuário é o ato de confirmação de que a pessoa que interage com o serviço é quem diz ser [1]. O processo de autenticação é geralmente o primeiro controle de acesso lógico que o usuário precisa superar para ter acesso ao sistema almejado. Atualmente existem três fatores que podem ser utilizados no processo de autenticação, os fatores são: o que você sabe, o que você é e o que você tem.

Para acessar qualquer serviço ou recurso na infraestrutura de tecnologia da informação, o usuário deveria estar autenticado. Os usuários deveriam utilizar credenciais de acesso e o sistema deveria comparar essas credenciais fornecidas com um banco de dados para identificar o usuário [2]. Geralmente neste processo de autenticação o usuário digita o seu usuário e a sua senha, e o sistema valida a credencial inserida pelo usuário no sistema.

Na época dos primeiros computadores de grande porte *(Mainframes)* problemas de segurança da informação envolvendo autenticação e autorização já existiam. Segundo Bonneau [3], a senha *(Password)* foi originalmente implantada na década de 60 no acesso aos computadores de grande porte *(Mainframes)* de tempo compartilhado *(Time Sharing)*. A senha era utilizada como forma de proteger o computador de grande porte contra o acesso não autorizado e limitar o acesso aos recursos do sistema.

Na década de 70 foi desenvolvido o controle de acesso no MULTICS e no Unix, e a senha foi projetada para proteger dados sensíveis e recursos computacionais. O MULTICS protegia as senhas armazenando em formato *hash*. Roger Needham e Mike Guy da Universidade de Cambridge criaram a prática de armazenar senhas em formato *hash* [3]. O MULTICS *(Multiplexed Information and Computing Service)* foi criado em 1964 e foi o primeiro sistema operacional de tempo compartilhado.

Na década de 60 ocorreram incidentes de segurança da informação relacionados a usuários de computadores de grande porte *(Mainframes)* que tiveram suas senhas comprometidas porque essas senhas eram fracas ou por terem sido descobertas e até mesmo

senhas que eram armazenadas sem estarem criptografadas [3]. Esses problemas relatados envolvendo senha há quase 60 anos atrás ocorrem até os dias atuais.

A autenticação utilizando usuário e senha é a forma mais comum de autenticação [4], mas a autenticação em sistemas de informação apenas com usuário e senha não é aconselhada, e é insuficiente para garantir um nível de segurança aceitável. A senha pode ser comprometida por meio do compartilhamento com outros usuários, por meio de ataques de dicionário [5], *rainbow table* [5], engenharia social [6], *keyloggers* [4] e *shoulder surfing* [1]. A complexidade da senha é uma necessidade a ser considerada quando é utilizado esse tipo de autenticação [6].

Segundo Almulhem [7], impor uma política de senhas fortes pode levar o usuário a escrever a sua senha em algum lugar e expor a senha diretamente. Senhas muito grandes e complexas dificilmente serão memorizadas pelos usuários. Para mitigar os riscos relacionados a utilização de apenas usuário e senha, é aconselhável que seja utilizada uma segunda autenticação.

De acordo com Ometov [6], usuário e senha não são adequados para prover um nível de segurança adequado em face a quantidade de ameaças a que esse método de autenticação é suscetível. A autenticação de dois fatores foi proposta para resolver o problema da fragilidade da autenticação somente com usuário e senha, e para agregar à autenticação tradicional um fator que o usuário possui ou um que usuário é, visando aumentar a robustez no processo de autenticação.

A autenticação por dois fatores e por múltiplos fatores aumentaram a segurança no processo de autenticação significativamente, contudo a autenticação de mais de um fator possui alguns problemas conhecidos, como os seguintes que foram citados por Ometov [6]: nem todo usuário pode utilizar autenticação biométrica, alguns tipos de autenticação precisam de sensores como: leitor de smartcard, leitor de impressão digital, microfone para o reconhecimento de voz, câmera para o reconhecimento facial etc, e isso gera custo adicional, outros tipos de autenticação exigem a disponibilização de dispositivos para os usuários como: *smartcards*, *tokens*, telefones celulares etc. A tecnologia de reconhecimento facial pode não funcionar adequadamente em ambientes com iluminação ruim ou com câmeras de baixa qualidade e a tecnologia de reconhecimento de voz pode não funcionar adequadamente em ambientes com muito barulho.

O uso de autenticação que dependa de celular pode não funcionar com a falta de Internet dependendo do tipo de autenticação ou com a falta de carga na bateria do celular [8].

[1]No ataque *shoulder surfing* o atacante observa o usuário durante a autenticação e a segurança é quebrada [4]. Neste ataque o atacante tem como objetivo descobrir a senha que o usuário está utilizando por meio da observação durante o processo de autenticação.

Diante dos possíveis problemas expostos e relacionados com a utilização de autenticação com apenas usuário e senha ou da autenticação com mais de um fator, outros tipos de autenticação pouco comuns podem ser utilizados como alternativa para mitigar os problemas relatados, como as autenticações gráficas e semânticas.

Almulhem [7] afirma que senhas gráficas proveem uma alternativa promissora às senhas tradicionais alfanuméricas, são atrativas porque as pessoas geralmente lembram com mais facilidade de imagens do que de palavras. A senha gráfica consiste em utilizar fotos como senha ou padrões que são definidos pelo usuário.

Em seu trabalho Le Bouder [4] afirma que a senha conceitual (semântica) é menos suscetível ao ataque de *shoulder surfing*, porque o usuário utiliza um conceito como sua senha e um conceito é mais difícil de identificar do que uma senha textual.

Este trabalho tem como objetivo propor um sistema de autenticação semântica utilizando lógica descritiva em conjunto com imagens para melhorar a segurança da autenticação com usuário e senha, mas sem a necessidade de utilização de dispositivos auxiliares e sem gerar custos adicionais.

1.1 Pergunta de Pesquisa

É possível melhorar a segurança do sistema de autenticação usuário e senha com um sistema de autenticação complementar que utiliza relações semânticas existentes em imagens?

1.2 Objetivos

O objetivo geral desse trabalho é a proposição de um sistema de autenticação complementar para qualquer sistema, aumentando a segurança no processo de autenticação, mas sem gerar impacto significativo no processo de autenticação da solução para os usuários finais e sem gerar custo adicional.

A execução do trabalho foi dividida entre os seguintes objetivos específicos:

- criar um modelo para o sistema de autenticação complementar;
- realizar prova de conceito do sistema de autenticação semântica gráfico;
- avaliar o sistema de autenticação semântica gráfica quanto a sua robustez e usabilidade.

1.3 Contribuições

Este trabalho tem como objetivo introduzir um novo sistema de autenticação contribuindo para a ampliação do conhecimento e possivelmente no aumento da segurança no processo de autenticação.

O sistema de autenticação proposto utiliza lógica descritiva na descrição das relações existentes em imagens que serão utilizadas posteriormente no processo de autenticação, reduzindo os riscos relacionados à segurança no processo de autenticação.

Neste trabalho foi possível responder à pergunta de pesquisa e cumprir todos os objetivos com êxito. Os resultados mostram que o sistema de autenticação proposto é viável, robusto e fácil de usar.

1.4 Metodologia

Foi utilizada neste trabalho a metodologia *Design Science Research* [9] seguindo as etapas descritas na Figura 1.1.

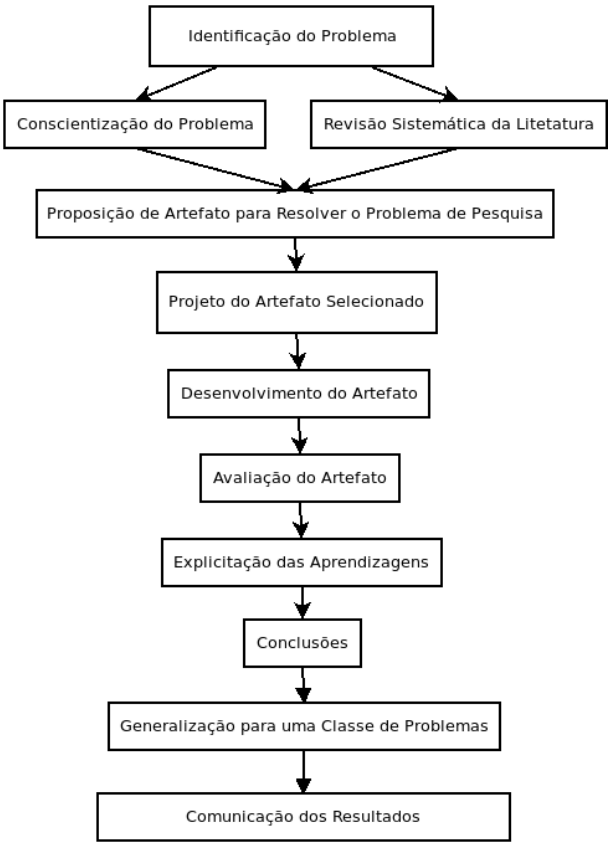

Figura 1.1: Etapas do *Design Science Research* adaptado de Dresch.

A Tabela 1.1 descreve as ações executadas neste trabalho em cada uma das fases do *Design Science Research* [9] mostradas na Figura 1.1.

Etapa	**Ação**
Identificação do problema	Foram identificadas as vulnerabilidades atuais relacionadas a autenticação com usuário e senha.
Conscientização do problema	Foram identificadas as vulnerabilidades de forma geral dos sistemas de autenticação baseados em usuário e senha, e as limitações dos mecanismos de autenticação de mais de um fator.
Revisão da literatura	Foram realizadas pesquisas de artigos científicos relevantes e no estado da arte sobre mecanismos de autenticação.
Proposição do artefato	Foi proposto o artefato para resolver o problema de pesquisa, foi proposto a criação do sistema de autenticação semântica gráfica.
Projeto do artefato	Foi definido o algoritmo de funcionamento do sistema.
Desenvolvimento do artefato	Foi desenvolvido o sistema de autenticação semântica gráfica.
Avaliação do artefato	O sistema desenvolvido foi avaliado quanto a robustez e usabilidade em prova de conceito.
Explicitação das aprendizagens	Foi registrado na dissertação o aprendizado obtido na execução deste trabalho.
Conclusão	Foram realizadas as considerações a respeito dos resultados obtidos neste trabalho.
Generalização para uma classe de problemas	Foi desenvolvida uma API para possibilitar a integração do sistema de autenticação proposto com outros sistemas. A API desenvolvida foi testada com o WordPress versão 5.6.
Comunicação dos resultados	Comunicação dos resultados obtidos neste trabalho para a comunidade acadêmica por meio de pedido de publicação de patente, publicação de artigo e a publicação da dissertação.

Tabela 1.1: Ações executadas neste trabalho em cada etapa do *Design Science Research* (Dresch).

1.5 Estrutura do Trabalho

O capítulo 1 apresenta a introdução e aborda as contribuições, pergunta de pesquisa, objetivos, metodologia utilizada e a estrutura deste trabalho.

O capítulo 2 aborda o tema segurança cibernética, a relevância do tema e algumas ameaças existentes no contexto da segurança cibernética, também serão apresentados conceitos importantes sobre segurança da informação.

O capítulo 3 apresenta uma revisão sobre sistemas de autenticação. O objetivo deste capítulo é mostrar os métodos de autenticação mais utilizados e alguns métodos de autenticação que não são tão comuns. Esse capítulo também mostra alguns sistemas de autenticação que utilizam imagens durante o processo de autenticação. São mostrados nesta revisão o método de autenticação usuário e senha, a autenticação biométrica, conceitos sobre autenticação de múltiplo fator, autenticação gráfica, autenticação semântica humana e CAPTCHA.

O capítulo 4 apresenta conceitos importantes sobre computação semântica, aborda temas importantes para esse trabalho como lógica descritiva, RDF e ontologias. A lógica descritiva, o RDF e as ontologias foram utilizadas como base no sistema de autenticação proposto.

O capítulo 5 apresenta o sistema de autenticação proposto, a autenticação semântica gráfica. Este capítulo também abordará as melhorias que serão obtidas no sistema de autenticação proposto, será apresentado o estudo de caso envolvendo a utilização da API no sistema proposto.

O capítulo 6 mostra os resultados obtidos na avaliação do sistema de autenticação proposto com o grupo de usuários.

O capítulo 7 apresenta a conclusão, sugestões de melhorias em trabalhos futuros e considerações finais.

Por fim um apêndice com uma descrição detalhada do experimento realizado para avaliar o sistema proposto.

Capítulo 2

Segurança da Informação

Este capítulo revisa conceitos relevantes para este trabalho. Serão abordados no decorrer desse capítulo conceitos relacionados à segurança da informação, como: segurança cibernética, ameaças, disponibilidade, integridade, confidencialidade, autenticidade, autorização e auditoria.

2.1 Segurança Cibernética

Em seu trabalho Kremer et al. [10] definem segurança cibernética como garantir a confidencialidade, integridade e disponibilidade da informação, serviços e infraestrutura de TIC. A confidencialidade, integridade e disponibilidade são aspectos importantes da segurança da informação citados pelo *NIST (National Institute of Standards and Technology)*, pelo Departamento de Segurança da Informação do Gabinete de Segurança Institucional da Presidência da República e por diversos autores em suas publicações na área de segurança da informação. A seção 2.3 irá abordar esse assunto.

Em seu trabalho Mandarino e Canongia [11] afirmam que a segurança cibernética vem se tornando cada vez mais uma função estratégica de estado e possui papel primordial na manutenção e preservação das infraestruturas críticas de um país. Os autores citam como parte da infraestrutura crítica aquela relacionada à: saúde, energia, defesa, transporte, telecomunicações, informação, etc.

O espaço cibernético é descrito por Mandarino e Canongia [11] como uma metáfora que descreve o espaço não físico criado por redes de computadores, sendo a Internet o maior espaço cibernético existente atualmente, pois trata-se da maior rede de computadores. O escopo de atuação da segurança cibernética abrange os aspectos e ações de prevenção e repressão aos ataques cibernéticos.

Utilizando os conceitos citados nos parágrafos anteriores, é possível concluir que a segurança cibernética deve garantir a confidencialidade, integridade e disponibilidade de ativos e informações existentes no espaço cibernético.

Na atualidade a dependência de tecnologia da informação por parte do estado e de empresas é cada vez maior. Existe uma tendência irreversível de automação/informatização do que for possível, visando a redução de custo e ganho em eficiência, porém essa informatização pode deixar vulneráveis aos ataques cibernéticos serviços que anteriormente não eram. Entre os serviços que podem estar ou se tornar vulneráveis aos ataques cibernéticos estão aqueles da infraestrutura crítica citados por Mandarino e Canongia [11]. A segurança cibernética deve garantir a segurança no espaço cibernético por meio da utilização das melhores práticas de segurança da informação. Além disso, é fundamental estabelecer no país as bases necessárias para o fomento de um ecossistema de negócios que permita o florescimento de uma cadeia de produção nacional mais autônoma em relação ao exterior, uma vez que é quase impossível garantir a segurança cibernética nacional com produtos e serviços estrangeiros, segundo Ishikawa et al. [12].

2.2 Ameaças

Segundo Kremer et al. [10], a primeira tarefa da segurança cibernética é identificar as ameaças e definir os tipos de ataques correspondentes. Podem ser consideradas ameaças: *software* malicioso, dano físico ou engenharia social. Essas ameaças podem ter como alvo: *hardware*, rede de comunicação, sistema operacional, aplicações ou os próprios usuários e o atacante pode ser alguém dentro da própria organização atacada *(insider)* ou externo *(outsider)*.

Kizza [13] afirma em seu trabalho que existem vários tipos de ameaças, incluindo vírus de computador, *worms* (vírus de computador que se auto propaga), ataque de negação de serviço distribuído *DDoS (Distributed Denial of Service)* e bombas eletrônicas. As motivações para os atacantes utilizarem essas ameaças em ataques cibernéticos pode estar relacionadas à: vingança, ganho financeiro, ódio, diversão, ganho de notoriedade e outras.

As ameaças aos sistemas computacionais surgem segundo Kizza [13] de:

- vulnerabilidade na infraestrutura de rede;
- vulnerabilidade em protocolos de comunicação;
- crescimento da comunidade *hacker*;
- vulnerabilidade em sistema operacional ou em serviços providos por meio do sistema operacional;

- *insider*;
- engenharia social;
- roubo ou perda de dispositivos.

De acordo com Kizza [13], as ameaças cibernéticas são categorizadas pelo FBI *Federal Bureau of Investigation* como: terrorismo, espionagem militar, espionagem econômica, direcionadas à infraestrutura nacional de informação, vingança e ódio.

Harris e Maymí [14] em seu livro citaram as seguintes ameaças para a senha:

- monitoramento eletrônico: durante a autenticação caso o atacante consiga interceptar a comunicação de rede entre o usuário e o sistema alvo, o atacante pode capturar a comunicação e autenticar no sistema utilizando ataque conhecido como ataque de repetição *(Replay Attack)*, caso o sistema alvo esteja vulnerável. Neste cenário caso a comunicação entre o usuário e o servidor não seja criptografada, o atacante poderia obter o usuário e senha utilizados pelo usuário durante o processo de autenticação;
- ataque de força bruta: neste ataque o atacante tenta autenticar utilizando todas as combinações de caracteres possíveis até descobrir a senha;
- ataque de dicionário: neste ataque o atacante utiliza um arquivo com muitas palavras e tenta descobrir a senha do usuário alvo tentando com cada uma das palavras;
- engenharia social: neste ataque o atacante tenta convencer o usuário a revelar a sua senha;
- *rainbow table*: neste ataque o atacante tenta descobrir a senha do usuário utilizando arquivo com *hashes* de várias senhas.

O sistema de autenticação proposto tem como objetivo minimizar os riscos relacionados às ameaças para o método de autenticação usuário e senha, implementando uma segunda autenticação, a autenticação semântica gráfica.

2.3 Princípios Básicos de Segurança da Informação

O *NIST (National Institute of Standards and Technology)* [15][16] define segurança da informação como proteger a informação e sistemas da informação de: acesso, utilização, divulgação, interrupção, modificação e destruição não autorizadas. Segurança da informação significa proteger ativos de atacantes, desastres naturais, condições ambientais adversas, falhas de energia, roubo, vandalismo e qualquer outra ameaça indesejável [17].

Segundo Andress [17], quanto mais elevado for o nível de segurança, geralmente menor será o nível de produtividade, então se o nível de segurança for muito alto, o nível de produtividade tende a zero. Equilibrar os requisitos mínimos de segurança com a usabilidade e a produtividade é um desafio.

A correta implantação dos controles de segurança da informação é primordial para proteger os ativos tangíveis e intangíves de uma organização, dados pessoais, a reputação da organização e cumprir leis. Estar em conformidade com marcos legais é uma obrigação das empresas, pois o descumprimento da legislação pode acarretar em prejuízo para a organização devido a penalização com multas [16].

Quando são aplicados controles de segurança em ativos, sistemas ou ambientes, deve ser considerado o nível de segurança em relação ao valor do ativo a ser protegido [17]. Aplicar um nível de segurança maior que o necessário para um ativo é um desperdício de recursos, inclusive financeiro para a organização. A categorização e a classificação dos ativos definirão os níveis de segurança necessários e ajudarão na escolha dos controles de segurança que deverão ser aplicados em cada ativo [18].

A política de segurança da informação é um componente essencial da governança de segurança da informação e sem a política a governança não há os elementos necessários para impor regras. A política de segurança é um conjunto de diretivas, regras e práticas que ditam como a organização gerencia, protege e distribui informações [19]. As diretrizes definidas na política de segurança da informação devem ser detalhadas nas normas e procedimentos.

O decreto nº 9.637, de 26 de dezembro de 2018 [20] institui a Política Nacional de Segurança da Informação no âmbito da administração pública federal e tem como objetivo assegurar a disponibilidade, a integridade, a confidencialidade e a autenticidade da informação a nível nacional. Segundo o Gabinete de Segurança Institucional da Presidência da República [21], Segurança da Informação e Comunicações são ações que objetivam viabilizar e assegurar a disponibilidade, a integridade, a confidencialidade e a autenticidade das informações.

O equilíbrio entre confidencialidade, integridade e disponibilidade é algo muito difícil de alcançar. Quando a disponibilidade é priorizada, provavelmente isso irá afetar a integridade e a confidencialidade, e quando a confidencialidade e a integridade são priorizadas, inevitavelmente a disponibilidade será afetada [22]. É importante priorizar a confidencialidade, integridade e a disponibilidade de forma que o nível de segurança da informação requerido para um determinado ativo seja alcançado.

2.3.1 Disponibilidade

A disponibilidade garante o acesso oportuno e confiável às informações [16]. A disponibilidade é a propriedade de que a informação está acessível e utilizável sob demanda para uma pessoa física ou determinado sistema, órgão ou entidade. Para haver disponibilidade no contexto da segurança da informação, a informação deve estar acessível no momento que ela é necessária [21].

2.3.2 Integridade

A integridade é a proteção contra modificações ou destruições indevidas de informação e garante o não repúdio da informação e a sua autenticidade [16]. Segundo o Gabinete de Segurança Institucional da Presidência da República[21], a integridade é a propriedade de que a informação não foi modificada ou destruída de maneira não autorizada ou acidental.

O NIST [16] define dois tipos de integridade, a integridade do dado e a integridade do sistema. Integridade do dado é a propriedade de que o dado não foi alterado de maneira não autorizada. A integridade do dado é aplicável para o dado armazenado, durante o processamento e enquanto está em trânsito. A integridade do sistema é a qualidade de que quando o sistema está em funcionamento ele funciona de forma integra, sem manipulações não autorizadas, intencionais ou acidentais.

2.3.3 Confidencialidade

A confidencialidade preserva restrições autorizadas de acesso e divulgação de informação, incluindo os meios para proteger a privacidade pessoal e informações proprietárias [16]. Segundo o Gabinete de Segurança Institucional da Presidência da República [21], confidencialidade é a propriedade de que a informação não esteja disponível ou não seja revelada a pessoa física, sistema, órgão ou entidade não autorizada e credenciada. Para existir a confidencialidade no contexto da segurança da informação, a informação só pode estar acessível para os usuários que devem ter o acesso.

2.3.4 Autenticidade

A autenticidade é a propriedade de que a informação foi produzida, expedida, modificada ou destruída por uma determinada pessoa física, ou por um determinado sistema, órgão ou entidade [21]. A autenticidade no contexto de segurança da informação apresentado pelo Gabinete de Segurança Institucional da Presidência da República garante o não repúdio.

2.4 Autorização

O processo de autorização verifica como um sistema pode ser utilizado. Durante a fase de autorização, o nível de acesso do usuário será verificado e será concedido o acesso de acordo com o que foi definido durante o processo de concessão da autorização. A autorização define quais são as permissões do usuário dentro de sistema [2]. A autorização é realizada após o processo de autenticação.

A autorização possibilita que uma vez que o usuário do sistema esteja autenticado, seja determinado o que exatamente este usuário pode fazer. A autorização é geralmente implementada através da utilização de controles de acesso [17].

2.5 Auditoria

Quando um usuário autentica em um sistema todas as suas atividades devem ser registradas e a auditoria registra essas atividades dos usuários no sistema [2]. Os registros relacionados a auditoria podem ser armazenados em banco de dados e em arquivo. São gravadas as ações executadas no sistema, quem executou, quando foi executada e onde foi executada.

A auditoria provê mecanismos para rastrear atividades no ambiente até a sua origem. A auditoria depende que a identificação, autenticação e controle de acesso estejam presentes para que o responsável por uma dada transação seja descoberto e associado a ela [17].

2.6 Considerações Finais

Este capítulo abordou temas relevantes para este trabalho como segurança cibernética, ameaças, conceitos de segurança da informação, autorização e auditoria, tendo em vista que o objetivo principal deste trabalho é contribuir na melhoria da segurança da informação no processo de autenticação utilizado em sistemas, mitigando o risco de comprometimento do sistema por meio de ameaças existentes para o método de autenticação usuário e senha. Foi necessário abordar temas como autorização e auditoria, pois esses dois temas estão diretamente ligados ao tema deste trabalho (autenticação). Para o processo de autenticação é importante que só sejam autenticados no sistema usuários com a devida autorização e todas às ações executadas pelo usuário no sistema devem ser rastreáveis (auditoria), iniciando no momento da autenticação.

O próximo capítulo mostra trabalhos na área de autenticação e realiza comparações entre esses trabalhos e o sistema de autenticação proposto.

Capítulo 3

Métodos de Autenticação

O capítulo 1 contemplou a fase de identificação do problema do *Design Science Research*, este capítulo contempla as fases de conscientização do problema e revisão da literatura do *Design Science Research* e mostra trabalhos na área de autenticação. Serão realizadas comparações entre os trabalhos citados e o trabalho proposto.

Autenticação de um usuário é o ato de confirmação de que a pessoa que interage com o serviço é quem diz ser [1]. O processo de autenticação é geralmente o primeiro controle de acesso lógico que o usuário precisa superar para ter acesso ao sistema almejado. Atualmente existem três fatores que podem ser utilizados no processo de autenticação, os fatores são: o que você sabe, o que você é e o que você tem.

Para acessar qualquer serviço ou recurso na infraestrutura de tecnologia da informação, o usuário deveria estar autenticado. Os usuários deveriam utilizar credenciais de acesso e o sistema deveria comparar essas credenciais fornecidas com um banco de dados para identificar o usuário [2]. Geralmente neste processo de autenticação o usuário digita seu usuário e sua senha, e o sistema valida a credencial inserida pelo usuário no sistema.

3.1 Autenticação com usuário e senha

Na época dos primeiros computadores de grande porte *(Mainframes)* problemas de segurança da informação envolvendo autenticação e autorização já existiam. Segundo Bonneau [3], a senha *(Password)* foi originalmente implantada na década de 60 no acesso aos computadores de grande porte *(Mainframes)* de tempo compartilhado *(Time Sharing)*. A senha era utilizada como forma de proteger o computador de grande porte contra o acesso não autorizado e limitar o acesso aos recursos do sistema.

Na década de 70 foi desenvolvido o controle de acesso no MULTICS e no Unix, e a senha foi projetada para proteger dados sensíveis e recursos computacionais. O MULTICS protegia as senhas armazenando em formato *hash*. Roger Needham e Mike Guy na

Universidade de Cambridge criaram a prática de armazenar senhas em formato *hash* [3]. O MULTICS *(Multiplexed Information and Computing Service)* foi criado em 1964 e foi o primeiro sistema operacional de tempo compartilhado.

Entropia da informação é o padrão da indústria para determinar se uma senha é forte ou fraca. A entropia é aferida baseando-se na quantidade de bits de informação utilizados na senha. Esta métrica avalia o quão imprevisível uma senha é. A avaliação é realizada de acordo com as seguintes características [23]:

- Conjuntos de símbolos utilizados;
- Distribuição dos conjuntos de símbolos entre os caracteres maiúsculos e minúsculos;
- Tamanho da senha.

Foi divulgado em publicação realizada no site da empresa de segurança da informação ESET [24] um ranking com as senhas mais comuns no ano de 2018. A Tabela 3.1 mostra as 25 senhas mais utilizadas em 2018 [24]:

Após a análise do ranking exibido na Tabela 3.1 fica claro que parte significativa dos usuários não se preocupam com segurança ou não tem consciência sobre os riscos de segurança relacionados a utilização de senha fraca. As senhas utilizadas no ranking foram obtidas por meio de divulgação de senhas vazadas após ataques segundo os autores. Parte dessas senhas são palavras de dicionário, sequências lógicas ou sequências de teclas que podem ser facilmente quebradas por um atacante. As senhas listadas não estão de acordo com o conceito de senha forte segundo a entropia da informação.

Há duas formas principais para gerar senhas, essas formas são: a senha gerada aleatoriamente por meio de computador e a senha gerada por humano [23].

A fórmula utilizada para o cálculo da entropia de uma senha é:

$$H = \log_2(b^l) \tag{3.1}$$

Na Fórmula 3.1 H é a entropia calculada em bits da senha, b é o número possível de símbolos no conjunto e l é o número de símbolos em uma senha ou o tamanho [25]. A Tabela 3.2 mostra a entropia para cada símbolo em um conjunto de símbolos específico:

A Tabela 3.2 mostra que quanto maior for o número de símbolos disponíveis na criação de uma senha, maior será a entropia por símbolo da senha e mais difícil será quebrar a senha.

Função *Hash*

Funções de *hash* recebem como entrada uma grande quantidade de dados e produzem uma saída curta chamada de *hash value* ou *digest* [26]. No método de autenticação usuário

Posição	**Senha**	**Diferente de 2017**
1	123456	Não.
2	password	Não.
3	123456789	Subiu 3.
4	12345678	Caiu 1.
5	12345	Não.
6	111111	Nova.
7	1234567	Subiu 1.
8	sunshine	Nova.
9	qwerty	Caiu 5.
10	iloveyou	Não.
11	princess	Nova.
12	admin	Caiu 1.
13	welcome	Caiu 1.
14	666666	Nova.
15	abc123	Não.
16	football	Caiu 7.
17	123123	Não.
18	money	Caiu 5.
19	654321	Nova.
20	!@#$%^&*	Nova.
21	charlie	Nova.
22	aa123456	Nova.
23	donald	Nova.
24	password1	Nova.
25	qwerty123	Nova.

Tabela 3.1: As 100 piores senhas de 2018. Fonte: SplashData.

e senha a função de *hash* é muito importante para mitigar riscos relacionados a quebra da senha dos usuários, pois utilizando hash a senha do usuário nunca é armazenada diretamente pela aplicação, é armazenado somente o *hash* da senha.

A Figura 3.1 mostra o fluxo de entrada e saída na função de *Hash*.

A Figura 3.1 mostra que o *Hash* é uma função de um único caminho, teoricamente não é possível descobrir a senha utilizando o *Hash*. Além de não permitir a descoberta da senha a partir do *Hash* da senha, a função de *Hash* deve ser resistente a colisão. Resistência a colisão significa que duas entradas diferentes não irão gerar o mesmo *Hash* na saída da função, contudo colisões irão inevitavelmente ocorrer de acordo com o princípio do buraco de pombo, na qual afirma que se existem m buracos e n pombos para entrar nos buracos, se n for maior que m, no mínimo um buraco abrigará mais de um pombo [26]. No cenário das funções de *Hash*, os buracos são as saídas da função de *Hash* e os pombos são as mensagens de entrada na função de *Hash*.

Nome do conjunto de símbolos	**Quantidade de símbolos no conjunto**	**Entropia por símbolo em bits**
Algarismos arábicos (0-9)	10	3.322
Numerais hexadecimais (0-9,A-F)	16	4.000
Case-insensitive alfabeto latino (a-z or A-Z)	26	4.700
Case-insensitive alfanumérico (a-z or A-Z,0-9)	36	5.170
Case-sensitive alfabeto latino (a-z,A-Z)	52	5.700
Case-sensitive alfanumérico (a-z,A-Z,0-9)	62	5.954
Todos os caracteres ASCII	95	6.570
Todos os caracteres estendidos ASCII	218	7.768
Binários (0-255 or 8 bits or 1 byte)	256	8.000
Lista de palavras Diceware	7776	12.925

Tabela 3.2: Entropia para cada símbolo em um conjunto de símbolos.

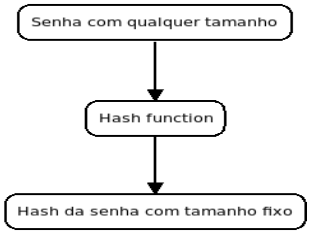

Figura 3.1: Entrada e saída da função de *Hash*, adaptado de Aumasson

Como existem mais possibilidades de mensagens de entrada para a função de *Hash* do que a quantidade de *Hash* de saída que podem ser gerados, porque o *Hash* gera sempre uma saída com tamanho fixo, por isso a colisão de *Hash* irá existir, contudo as colisões de *Hash* são tão difíceis de encontrar que as funções de *Hash* são consideradas resistentes a colisão.

No momento em que o usuário digita sua senha no sistema, esta senha passa pela função de *Hash* e só posteriormente será armazenada no banco de dados do sistema. No momento da autenticação o usuário irá digitar a senha e esta senha passará pela função de *Hash* e será comparada com o *Hash* da senha que foi armazenado no banco de dados anteriormente, caso os *Hashes* sejam iguais, a validação da senha será realizada com sucesso.

A autenticação com usuário e senha pode ser comprometida de diversas formas, para

minimizar o risco outros fatores de autenticação estão sendo utilizados em conjunto com o usuário e senha, porém esses fatores também têm os seus problemas. Este trabalho propõe que o sistema de autenticação semântica gráfica seja utilizado em conjunto com a autenticação por meio de usuário e senha. A autenticação semântica gráfica não possui os problemas existentes nos sistemas de autenticação utilizados como o segundo fator. Esses problemas são abordados no decorrer desse trabalho.

3.2 Autenticação biométrica

A autenticação biométrica é um sistema de controle de acesso que autentica indivíduos baseando-se em seus traços biométricos. Diferentemente da senha ou de qualquer sistema de autenticação convencional, traços biométricos como impressões digitais, íris e características comportamentais ligadas a um indivíduo são de difícil manipulação [27].

A autenticação biométrica possui as seguintes fases [28]:

- aquisição do dado biométrico com sensor;
- conversão do dado biométrico para um formato digital;
- comparação do dado biométrico coletado e convertido com o dado biométrico armazenado.

A autenticação biométrica está dividida em dois grupos principais: fisiológico e comportamental. A impressão digital, características do rosto, DNA, formato do corpo e íris são características fisiológicas. O padrão de digitação, voz, letra, maneira de andar são características comportamentais [28].

Tecnologias de verificação da íris e do rosto são frequentemente utilizadas em sistemas de vigilância de vídeo, controle de fronteira e no gerenciamento do controle de acesso em dispositivos individuais [29].

Segundo Abe e Shinzaki [29], dados biométricos como: formato das sombrancelhas, movimentos dos olhos, formato do nariz, formato das orelhas e movimentos labiais podem ser utilizados para aumentar a acurácia de outros sistemas de autenticação biométricos, como por exemplo a autenticação facial e íris.

No trabalho de Abe e Shinzaki [29] também são abordadas formas de autenticação como: impressão da junta do dedo, impressão da palma da mão, impressão digital e unha.

Na autenticação com a junta do dedo, o dedo é digitalizado, informações como a distância entre a junta maior e a menor, junta menor e a ponta do dedo, formato do dedo e a curvatura do dedo são registradas e utilizadas no processo de validação [29].

Na autenticação com a palma da mão informações sobre o formato da mão e das artérias são armazenadas e utilizadas no processo de validação [29].

Na autenticação com impressão digital informações sobre a impressão digital são armazenadas e utilizadas no processo de autenticação [29].

A autenticação com a unha é sugerida como uma forma de autenticação temporária, pois as informações sobre o formato das unhas são coletadas e utilizadas no processo de validação, contudo as informações coletadas podem mudar rapidamente, caso as unhas cresçam ou o usuário corte as unhas e neste caso o usuário não irá conseguir autenticar [29].

Dados biométricos como os batimentos cardíacos podem ser utilizados na autenticação biométrica, por meio da validação do padrão do eletrocardiograma [29].

De acordo com Ometov et al. [6], na implementação de um sistema de autenticação biométrica é importante observar que alguns usuários podem não ser capazes de utilizar autenticação com validação das impressões digitais por terem perdido as impressões digitais, usuários com problemas visuais podem não ser capazes de utilizar autenticação com validação da íris e usuários mudos podem não ser capazes de utilizar autenticação com validação da voz.

Autenticação biométrica requer integração com novos serviços e dispositivos que resultam na necessidade de treinamento durante a adoção, o que é mais complicado para pessoas mais velhas por causa do entendimento em relação aos problemas relacionados a segurança [6].

Para utilizar autenticação biométrica é importante ter um ambiente operacional robusto. Um exemplo de falta de robustez do ambiente é no caso do reconhecimento de voz que foi testado em um ambiente silencioso e foi implantado em um ambiente barulhento, podendo impossibilitar a utilização desse tipo de biometria [6]. Algumas soluções de autenticação biométrica dependem de condições ambientais ideais para o seu funcionamento.

As duas principais métricas utilizadas para aferir a precisão de um sistema de autenticação biométrico são: *FAR (False Acceptance Rate)* e *FRR (False Recognition Rate)* [6]. O *FAR* é o percentual de usuários não autorizados que são identificados como usuários autorizados pelo sistema de autenticação biométrica e o *FRR* é o percentual de usuários autorizados que são identificados como usuários não autorizados pelo sistema de autenticação biométrica [6].

A autenticação semântica gráfica não necessita de dispositivos auxiliares durante o processo de autenticação e nem de condições ambientais específicas para funcionar, diferente da autenticação biométrica.

3.3 Autenticação com múltiplo fator

De acordo com Colnago et al. [8], uma forma de reduzir os riscos relacionados a insegurança das senhas é utilizar a senha em conjunto com um outro fator de autenticação. Os fatores de autenticação são identificados em três categorias: alguma coisa que você sabe, alguma coisa que você tem e alguma coisa que você é.

O fator do tipo alguma coisa que você sabe é o conhecimento que o usuário tem e que será utilizado no sistema de autenticação, por exemplo: senha ou respostas para questões de segurança [8].

O fator do tipo alguma coisa que você tem são dispositivos como: *token*, celular, *smartcard* ou objetos como um papel com códigos de acesso [8] que são utilizados no sistema de autenticação.

O fator do tipo alguma coisa que você é são dados biométricos que serão utilizados no sistema de autenticação como: impressão digital, retina, íris, reconhecimento de voz, reconhecimento facial [8].

A autenticação de dois fatores é definida como a utilização de duas categorias de métodos de autenticação diferentes em conjunto [8], por exemplo: autenticação biométrica e autenticação com senha, autenticação com senha e autenticação com *smartcard* ou autenticação biométrica e autenticação com *token*. A autenticação utilizando múltiplos fatores é definida como a utilização de duas ou mais categorias de métodos de autenticação diferentes em conjunto.

Segundo Ometov et al. [6], os fatores de autenticação podem ser avaliados de acordo com os seguintes critérios: universalidade, unicidade, coletividade, performance, aceitabilidade e falsificação. A Tabela 3.3 descreve cada um dos critérios.

Critério	**Descrição**
Universalidade	É a presença do fator em cada pessoa.
Unicidade	Indica como o fator se diferencia entre as pessoas.
Coletividade	Mede qual é a facilidade de obter o dado para o processamento do fator.
Desempenho	Indica a acurácia, velocidade e robustez.
Aceitabilidade	É o nível de aceitação da tecnologia pelas pessoas em sua vida.
Falsificação	Indica o nível de dificuldade para capturar e falsificar os dados.

Tabela 3.3: Critérios de avaliação dos fatores de autenticação

Segundo Colnago et al. [8], os trabalhos sobre *2FA (Two Factor Authentication)* focam mais em áreas onde os usuários esperam um nível de segurança da informação elevado, como em sistemas financeiros.

No trabalho de Colnago et al. [8], a pesquisa mostrou que a maioria dos usuários que adotaram o *2FA* acharam que a solução causa inconvenientes, mas também acharam que é fácil de utilizar e acreditam que suas contas ficaram mais seguras após a adoção.

Em seu trabalho Colnago et al. [8] realizaram pesquisa com estudantes e funcionários da Universidade Carnegie Mellon para descobrir quais são os motivos para não utilizarem a solução do tipo *2FA (Two Factor Authentication)* ofertada pela Universidade e o resultado foi o seguinte:

- falta de utilidade, preocupações relacionadas com a usabilidade e inconvenientes;
- não acreditam que sua conta precisa de segurança, pois não tem informações importantes nestas contas ou sistemas;
- tempo extra para realizar o processo de autenticação;
- não querem depender de um outro dispositivo para realizar a autenticação;
- boatos negativos sobre a solução espalhados por usuários que já adotaram.

As duas principais desvantagens apontadas pela solução de autenticação de *2FA* identificadas na pesquisa de Colnago et al. [8] são: o tempo extra para autenticar e o inconveniente causado pela solução.

Os problemas identificados durante a implantação da solução de *2FA* na Universidade Carnegie Mellon foram os seguintes [8]:

- o usuário esqueceu o telefone e não consegue autenticar;
- acabou a bateria do celular do usuário e não consegue autenticar;
- não tem conexão de dados no celular e não consegue autenticar;
- o aplicativo do celular não está sincronizando com a conta no servidor.

Durante a implantação da solução de *2FA* na Universidade Carnegie Mellon alguns usuários reclamaram que tinham que utilizar seus próprios dispositivos para realizar a autenticação em sistemas da Universidade e acreditavam que a Universidade deveria fornecer esses dispositivos [8].

A autenticação semântica gráfica tem como proposta ser um sistema de autenticação complementar que utiliza relações que podem ser identificadas facilmente pelo usuário por meio da inferência durante o processo de autenticação.

3.4 Autenticação Gráfica

Almulhem [7] em seu trabalho propôs um sistema de autenticação intitulado sistema de autenticação de senha gráfica. O sistema de Almulhem é um sistema de autenticação gráfica simples que combina senhas gráficas com senhas textuais.

No sistema proposto por Almulhem [7], o usuário utiliza livremente a foto escolhida, os pontos de interesse e as palavras que quiser vincular. A ordem e o número de pontos de interesse podem aumentar a segurança do mecanismo de autenticação. Estes parâmetros juntos permitem senhas muito complexas segundo o autor. O autor acredita que a abordagem proposta é promissora e é única por duas razões, combina senha gráfica com senha textual para conseguir o melhor dos dois mundos e provê uma autenticação de múltiplo fator, com imagem, texto e os pontos de interesse.

No processo de autenticação utilizando o sistema proposto por Almulhem [7], o usuário primeiro digita o seu usuário e então o sistema exibe a imagem associada ao seu usuário. O usuário deve selecionar corretamente os pontos de interesse e digitar as palavras corretas em cada ponto.

A Figura 3.2 mostra como seria a autenticação utilizando o sistema de autenticação gráfica proposto por Almulhem [7].

Na Figura 3.2, o usuário selecionou três áreas na foto, inseriu textos identificando cada cômodo na imagem em sua respectiva posição e em uma ordem específica, criando a sua senha gráfica. O usuário selecionou pontos de sua preferência na foto e inseriu textos nestes pontos, contudo não existe nenhuma semântica na imagem que possa ser utilizada pelo computador, apenas pelo usuário que criou a senha gráfica. As setas estão sendo utilizadas apenas para mostrar em qual ordem os pontos devem ser selecionados durante o processo de autenticação.

No trabalho de Almulhem [7] o conhecimento sobre a imagem está com o usuário e o computador não possui esse conhecimento. Na autenticação semântica humana, o computador possui informações que o permite ter algum conhecimento sobre a imagem e o conhecimento do usuário será validado durante o processo de autenticação pelo computador.

3.5 Autenticação Semântica Humana

Em seu trabalho Lorant et al. [30] introduziram um sistema de autenticação gráfica intitulado *(HSA - Human Semantic Authentication)*. Durante o processo de autenticação

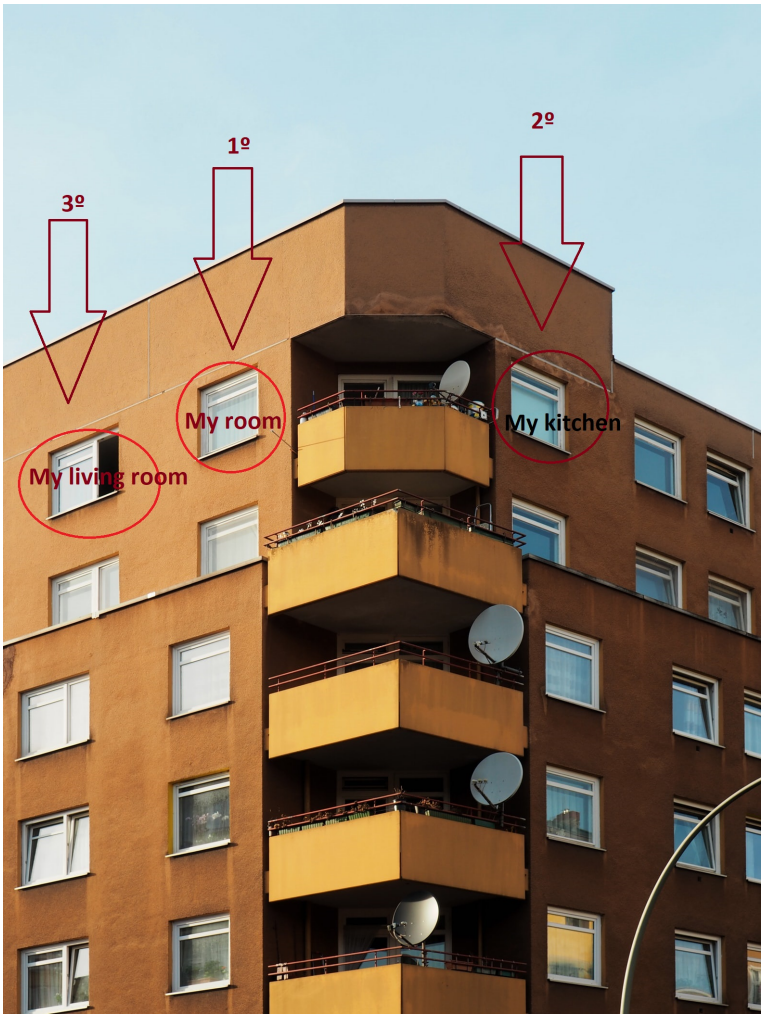

Figura 3.2: Autenticação utilizando o sistema proposto por Almulhem em 2011

no sistema de Lorant et al. [30] o usuário deve identificar quatro conceitos que constituem a sua senha conceitual clicando nas áreas da imagem onde os conceitos estão presentes.

O *HSA* tem como objetivo aumentar a resistência do sistema de autenticação contra ataques humanos e automatizados [30]. O *HSA* é um sistema do tipo *recall-based*[1] e *cued recall-based*[2]. São definidas zonas nas imagens e os conceitos são codificados nestas zonas [30].

No *HSA* a codificação não fica visível para o usuário, no processo de autenticação uma imagem é apresentada para o usuário e o mesmo deve identificar as zonas na imagem que possuem a sua senha conceitual na ordem correta. Em cada autenticação, uma imagem diferente é apresentada para o usuário por motivos de segurança [30].

[1]Na autenticação do tipo *recall-based* os usuários desenham padrões ou repetem uma sequência de ações [4]

[2]Na autenticação do tipo *cued-recall-based* os usuários devem selecionar pontos em uma imagem ou selecionar uma sequência de imagens [4]

No *HSA* se um atacante possuir as imagens utilizadas na autenticação do usuário e as respostas, a possibilidade do atacante conseguir inferir a senha do usuário é alta, sendo um ponto fraco dos sistemas de autenticação gráfica. Uma forma de dificultar a identificação da senha conceitual por meio da observação é não utilizar conceitos que são fáceis de identificar e utilizar conceitos que são mais abstratos [30].

O sistema proposto por Salembier et al. [31] tem como finalidade prover alternativa a autenticação com senha alfanumérica, melhorar a usabilidade no processo de autenticação e facilitar para o usuário a memorização da senha conceitual. Os autores projetaram o sistema de autenticação para tornar mais difícil o comprometimento por ataques que utilizam *key logger* e *spyware*.

No sistema de autenticação proposto por Salembier et al. [31], o usuário memoriza quatro conceitos que poderiam ser: amarelo, ferramenta, animal e comida, e deve selecionar durante o processo de autenticação áreas na imagens que possuem esses conceitos para ser autenticado.

Os conceitos são associados às áreas em cada imagem e a descrição dos conceitos associados não ficam visíveis para o usuário, o mesmo deve identificar apenas o conceito relacionado a sua senha conceitual. Além da utilização de símbolos, cores, é possível utilizar palavras em imagens na senha conceitual [31].

As seguintes regras foram utilizadas no trabalho de Salembier et al. [31]:

- a utilização de cores como senha conceitual foi abandonada, pois cores podem ser reconhecidas automaticamente por computadores, enfraquecendo o sistema;
- conceitos que envolvem um trabalho de inferência complexo ou incerto não foram utilizados;
- as imagens devem ser suficientemente visíveis, ter boa nitidez e tamanho.

Em seu trabalho Salembier et al. [31] criaram três aplicações fictícias e os usuários que participaram do experimento tiveram que memorizar 3 conjuntos com quatro senhas conceituais que foram geradas pelos pesquisadores e enviadas para os participantes, uma para cada aplicação.

O tempo médio de autenticação obtido no sistema de Salembier et al. [31] foi de 24,7 segundos, estando um pouco acima do que é considerado razoável pelos autores, que seria abaixo de 20 segundos. O sistema atual dificulta para o usuário a memorização das senhas conceituais, pois os usuários não tem condições de gerar sua própria senha, as senhas conceituais foram geradas pelo sistema e posteriormente enviadas para os usuários, caso os usuários pudessem gerar suas próprias senhas conceituais, isso teria um efeito positivo na memorização.

O estudo mostrou que os melhores resultados foram obtidos quando as senhas conceituais não exigiam muito trabalho inferencial dos usuários ou não geravam incertezas. O sistema proposto apresentou resultados significativos no que diz respeito a memorização das senhas conceituais pelos usuários durante o experimento, contudo não foi avaliada a usabilidade em dispositivos móveis neste experimento [31].

Os resultados obtidos no trabalho de Salembier et al. [31] mostram taxa de associação correta entre imagem e conceito pelos usuários de 98% e a percepção de dificuldade de 1,77 de uma escala de 1 até 5, onde 5 representa o nível mais difícil.

Bouder et al. [4] em seu trabalho analisaram a segurança do protocolo semântico humano *(HSM - Human Semantic Authentication)* introduzido por Salembier et al. [31]. Este estudo tenta descobrir o quão resistente o sistema de autenticação semântico humano é ao ataque *shoulder surffing* ou qualquer outro que capture a tela do usuário durante a autenticação.

O estudo de Bouder et al. [4] comparou o processo de autenticação utilizando senha convencional com o processo utilizando senha conceitual. Diferentemente da autenticação com senha alfanumérica, onde os usuários escolhem caracteres para serem cada dígito da sua senha, na autenticação semântica humana, cada conceito de senha seria o equivalente a um caractere da senha alfanumérica. Quando um usuário utilizando senha alfanumérica é observado, é possível identificar os caracteres utilizados como senha, diferente da senha conceitual, pois uma imagem pode ter n conceitos associados.

Segundo Bouder et al. [4], o mesmo número de conceitos deve ser apresentado na tela de autenticação, para não ter um conceito mais seguro do que outro. Para ter o mesmo nível de segurança do *PIN (Personal Identification Number)*, pelo menos 10 conceitos deveriam ser apresentados em cada imagem.

De acordo com Bouder et al. [4], um problema com o *HSA* é a capacidade dos usuários reconhecerem os conceitos nas fotos, pois se o conceito é muito complexo ou abstrato o usuário terá dificuldade para identificar e se o conceito for muito simples, o atacante facilmente conseguirá descobri-lo.

A melhoria na segurança do *HSA* em relação ao ataque *shoulder surffing* é mínima, o sistema deveria possuir um número suficiente de fotos para que as fotos dificilmente se repetissem ou as fotos utilizadas no sistema de autenticação deveriam ser geradas ou obtidas em tempo de execução [4].

Este trabalho propõe a criação de um sistema de autenticação com abordagem diferente do *HSA (Human Semantic Authentication)*, introduz no sistema múltiplas imagens, onde o usuário selecionará aquela que possui a relação da sua senha semântica em uma ordem determinada, sem a necessidade de clicar exatamente na área que indica a relação de sua senha semântica na foto durante o processo de autenticação. O sistema desde tra-

balho utiliza uma imagem diferente para cada relação da senha semântica que o usuário precisará selecionar durante o processo de autenticação. O sistema utiliza na descrição das imagens lógica descritiva e ontologias para descrever as relações existentes nas imagens.

3.6 CAPTCHA

O CAPTCHA é um acrônimo de *Completely Automated Public Turing test to tell Computers and Humans Apart*, é conhecido também como teste de Turing inverso. O CAPTCHA é um teste comum de desafio e resposta com propósito de garantir que quem está interagindo com o sistema é um humano e não um *software* que pode utilizar inteligência artificial ou não [32].

De acordo com Xu [33], uma pessoa gasta em média 10 segundos para identificar um CAPTCHA que utiliza caracteres. O projeto do CAPTCHA deve considerar a segurança e sua usabilidade simultaneamente.

Na década de 50 o pesquisador Alan Turing defendia a hipótese de que os computadores poderiam se passar por humanos, para provar elaborou um teste onde o avaliador deveria dizer se é humano ou computador baseando-se em perguntas simples respondidas por humanos e computadores. Para ser considerado um CAPTCHA, o código e os dados do programa precisam estar disponíveis publicamente [34].

Em setembro de 2000 pesquisadores da Universidade Carnegie Mellon desenvolveram o primeiro CAPTCHA comercial baseado em texto para resistir as propagandas geradas por programas maliciosos em salas de bate-papo do Yahoo de acordo com Chen [35].

Os CAPTCHAs podem ser divididos nos seguintes grupos de acordo com Brodić [36]:

- baseado em texto;
- baseado em imagem;
- baseado em áudio;
- baseado em vídeo.

De acordo com Ogiela [32], a maior vantagem dos CAPTCHAs baseados em textos é o fato de humanos naturalmente terem a habilidade para reconhecer padrões visuais, esses padrões são difíceis de serem reconhecidos por computadores e exige do computador para isso no mínimo implementações envolvendo inteligência artificial.

Em seu trabalho Chen [35] afirma que CAPTCHAs baseados em texto para resistir a ataques de reconhecimento de máquina geralmente possuem algumas das características a seguir:

- utiliza grande quantidade de caracteres, para resistir ataques de força bruta;

- distorce e/ou sobrepõe caracteres;
- utiliza caracteres com: tamanhos, comprimentos, ângulos, localização e fontes diferentes;
- utiliza caracteres sem preenchimento ou contorno.

Segundo Hasan [37], o CAPTCHA baseado em texto é o tipo de CAPTCHA mais simples e foi o primeiro a ser inventado.

Segundo Xu [33], CAPTCHAs baseados em imagens proveem desafios com intuito de diferenciar humanos de computadores, exigindo o entendimento dos conteúdos existentes em imagens. São utilizados comumente em desafios CAPTCHAs com imagens a detecção de objetos, o reconhecimento de alvos e o entendimento de alguma cena [33].

Gao [38] afirma em seu trabalho que CAPTCHAs baseados em imagens tem as seguintes desvantagens:

- não são gerados automaticamente;
- as imagens precisam ser identificadas de alguma forma;
- muitos sistemas dependem de idioma e precisam dar muitas informações para o usuário saber o que fazer;
- os sistemas tradicionais consomem muita banda na rede para carregar as imagens.

De acordo com Hasan [37], CAPTCHAs baseados em imagens podem gerar os seguintes problemas para os seus usuários:

- alguns usuários possuem pouca visão ou algum déficit que dificulta que o desafio do CAPTCHA seja resolvido com sucesso;
- a probabilidade de um programa malicioso quebrar um CAPTCHA irá aumentar se o número de escolhas for diminuindo, porém quanto mais imagens disponíveis, mais recurso de armazenamento será consumido;
- se o CAPTCHA está disponível somente em um idioma, somente os conhecedores daquele idioma saberão resolver o desafio.

A principal diferença entre o CAPTCHA baseado em imagem e o trabalho proposto é que no CAPTCHA o sistema dá instruções para os usuários durante a validação para que os mesmos sejam capazes de selecionar as imagens corretas, no nosso sistema de autenticação o sistema apenas exibe as imagens para os usuários e os mesmos são os únicos que sabem como selecionar as imagens corretadas por meio de suas senhas semânticas.

3.7 Considerações Finais

A Tabela 3.4 mostra um quadro comparativo entre os sistemas de autenticação exibidos neste capítulo e o sistema de autenticação proposto neste trabalho.

	Fácil memorização	**Resistente ao ataque *shoulder surfing***	**Resistência ao ataque de força bruta**	**Gera problema de mobilidade**	**Exige investimento**	**Usuário utiliza inferência**
Usuário e senha	Não	Não	Não	Não	Não	Não se aplica
Autenticações biométricas	Não se aplica	Não se aplica	Não se aplica	Sim	Sim	Não se aplica
Smartcard / token	Não se aplica	Não se aplica	Não se aplica	Sim	Sim	Não se aplica
Autenticação semântica humana	Sim	Não	Não se aplica	Não	Não	Sim
Autenticação gráfica	Sim	Não	Não se aplica	Não	Não	Não
Autenticação semântica gráfica	Sim	Sim	Sim	Não	Não	Sim
CAPTCHA	Não se aplica	Não se aplica	Sim	Não	Não	Sim

Tabela 3.4: Quadro comparativo entre sistemas de autenticação

A Tabela 3.4 mostra que o sistema de autenticação proposto possui características importantes que não foram contempladas integralmente pelos outros sistemas de autenticação utilizados na comparação, demonstrando a robustez do sistema. O ponto forte deste sistema de autenticação é a necessidade do usuário utilizar a sua capacidade de inferência para autenticar no sistema, pois realizar um ataque automatizado explorando essa caraterística é muito difícil sem a utilização de inteligência artificial. Outras caraterísticas muito importantes são: a resistência do sistema ao ataque de força bruta, a resistência do sistema ao ataque *shoulder surfing*, a facilidade que o usuário tem para memorizar a senha semântica e a possibilidade do usuário autenticar em qualquer dispositivo, não tendo a sua mobilidade afetada pelo sistema de autenticação, além de não exigir investimento em dispositivos.

Este capítulo mostrou o método de autenticação usuário e senha, a autenticação biométrica, a autenticação com múltiplo fator, a autenticação gráfica, a autenticação semântica humana e o CAPTCHA, tais métodos/sistemas de autenticação possuem suas fragilidades e pontos fortes, por este motivo foram abordados neste capítulo, tendo em vista que o

trabalho proposto visa justamente a proposição de um sistema de autenticação complementar para o método de autenticação usuário e senha. Foi importante mostrar quais são as alternativas de segundo fator de autenticação ou segunda autenticação disponíveis, também foi importante mostrar em que situação está atualmente a autenticação gráfica e a autenticação semântica. A autenticação é um dos processos mais importantes em qualquer sistema de informação, pois é por meio da autenticação que os usuários legítimos obtêm o acesso ao sistema pretendido e os atacantes também, caso consigam burlar o sistema de autenticação. A autenticação contribui na proteção do sistema contra acessos indevidos, consequentemente a autenticação contribui na proteção das informações existentes no sistema. O sistema de autenticação é um dos controles de segurança da informação mais importantes em qualquer sistema de informação que exija autenticação.

No próximo capítulo serão abordados os seguintes assuntos relacionados à computação semântica: lógica descritiva, RDF *(Resource Description Framework)* e ontologias.

Capítulo 4

Computação Semântica

Este capítulo revisa conceitos importantes para esse trabalho sobre computação semântica. No decorrer desse capítulo serão abordados os seguintes assuntos relacionados à computação semântica: lógica descritiva, RDF *(Resource Description Framework)* e ontologias.

4.1 Lógica Descritiva

Segundo Kumar [39] a Lógica Descritiva *(Description Logic - DL)* fornece uma construção lógica de bases de conhecimento. A Lógica Descritiva possui conceitos, papéis e indivíduos em seu bloco básico de construção.

De acordo com Baader [40], o nome Lógica Descritiva foi dado porque importantes noções do domínio são descritas e também está equipado com uma semântica formal baseada em lógica.

Kumar [39] afirma que a descrição lógica tem sido comprovada como a mais promissora para o processamento, compartilhamento e a interpretação de conhecimento especialmente na Internet. As ontologias tem um papel chave na construção de bases de conhecimento de forma hierárquica de conceitos e papéis de domínios em particular.

Baader [40] afirma que Lógica Descritiva é o mais recente nome para a família de formalismos de representação do conhecimento que representa o conhecimento do domínio de uma aplicação, definindo os conceitos relevantes do domínio e utilizando esses conceitos para especificar propriedades dos objetos e indivíduos no domínio.

A Lógica Descritiva possui os construtores exibidos na Tabela 4.1[40][39]:

As descrições dos conceitos são utilizadas para construir afirmações na base de conhecimento da lógica descritiva, onde tipicamente possui duas partes, uma terminológica e a outra afirmativa. A parte terminológica é chamada de *TBox* e a parte afirmativa é chamada de *ABox* [40].

Construtor	**Símbolo**	**Descrição**
Conjunção	($\sqcap$)	Intersecção.
Disjunção	($\sqcup$)	União.
Negação	($\neg$)	Conjunto complemento.
Restrição existencial	($\exists$r.C)	Pelo menos um valor da restrição deve existir.
Restrição de valor	($\forall$r.C)	O valor da restrição deve existir em todos os casos.
Conceito universal	($\top$)	Superconceito ou conceito pai de todos os conceitos.
Conceito de insatisfação	($\bot$)	Conceito de insatisfação.

Tabela 4.1: Construtores da lógica descritiva

Na Lógica Descritiva ontologias tomam a forma de *TBox* e os dados são armazenados no *ABox* [41].

O *TBox* possui o conhecimento intencional em forma de uma terminologia, o vocabulário do domínio de uma aplicação, é construído através de declarações que descrevem as propriedades gerais dos conceitos. O conhecimento intencional não muda. O vocabulário possui conceitos que definem o conjunto de indivíduos, papéis que definem os relacionamentos entre indivíduos [42].

Um axioma *TBox* descreve relações entre conceitos [43]. O exemplo a seguir representa que todos os carros são veículos utilizando o conceito de inclusão:

$$\text{Carro} \sqsubseteq \text{Veículo}$$

O axioma anterior afirma que o conceito Carro está contido no conceito Veículo.

O axioma *TBox* a seguir utiliza o conceito de equivalência para afirmar que dois conceitos têm a mesma instância [43]:

$$\text{Pai} \equiv \text{Homem} \sqcap \exists\ \text{temFilho.Pessoa}$$

O axioma anterior afirma que o conceito Pai é equivalente ao conceito Homem que tem algum relacionamento do tipo temFilho com Pessoa.

O *ABox* possui o conhecimento estendido, também chamado de conhecimento afirmativo, contém afirmações sobre indivíduos utilizando o vocabulário, esse conhecimento é específico de indivíduos de um domínio. O conhecimento estendido é geralmente dependente de um conjunto de circunstâncias e pode mudar [42].

Um axioma *ABox* captura o conhecimento sobre individuos, especialmente conceitos que eles possuem e relacionamentos mútuos. O mais comum axioma é a afirmação de conceito como [43]:

Professor(Paulo)

O axioma anterior afirma que Paulo é professor ou mais precisamente que um indivíduo chamado Paulo é uma instância do conceito Professor.

As afirmações de papéis descrevem no *ABox* relações entre indivíduos [43]. A afirmação a seguir afirma que Otávio é primo de Gustavo ou mais precisamente que um indivíduo chamado Otávio está em uma relação que é representada por primo com um indivíduo chamado Gustavo.

primo(Otávio,Gustavo)

O *TBox* possui axiomas relacionais e referem-se as propriedades dos papéis, axiomas de inclusão de papéis e equivalência entre papéis [43]. Um exemplo de inclusão é:

$$\text{Gato} \sqsubseteq \text{Felino}$$

O axioma anterior afirma que Gato é um sub-papel de Felino.

O *RBox* possui o papel de inclusão de axiomas do tipo composição de papel que pode ser utilizado para descrever axiomas complexos de inclusão de papel. O exemplo a seguir mostra como a composição de papéis pode ser descrita.

$$\text{casadoCom} \circ \text{irmaoDe} \sqsubseteq \text{CunhadoDe}$$

O axioma anterior afirma que a composição dos papéis casadoCom e irmaoDe estão contidos ou é um sub-papel do papel CunhadoDe.

De acordo com Kumar [39], a criação de uma base de conhecimento com lógica descritiva inicia com a definição de conceitos atômicos e papéis atômicos. Conceitos e papéis atômicos existem em um domínio específico, são entidades autoexplicáveis que não são definidas utilizando outros conceitos ou papéis.

Outros conceitos gerais e papéis são definidos utilizando conceitos atômicos e conceitos gerais, papéis atômicos e papéis gerais e construtores [39].

Os conceitos (C) são construídos de conceitos atômicos utilizando o conceito universal ($\top$), o conceito de insatisfação ($\bot$), a negação ($\neg$A), a união ($C_1 \sqcup C_2$), a intersecção ($C_1 \sqcap C_2$), o quantificador existencial ($\exists$R.C), o quantificador universal ($\forall$R.C), restrições de cardinalidade ($\geq_n$ R.C, $\leq_n$ R.C), etc [39].

Os papéis (R) são construídos de papéis atômicos (P), negação ($\neg$R), transição (R^+), papéis inversos (R^-), etc [39].

Conceitos e papéis possuem relações unárias ou binárias na lógica descritiva como: C(x) e R(y,z), onde x satisfaz o conceito C e y e z estão em uma relação R [39].

Desde o início a Lógica Descritiva tem sido considerada uma linguagem de propósito geral para a representação do conhecimento e do raciocínio, sendo adequada para muitas aplicações [40].

O primeiro sistema de representação do conhecimento baseado em lógica descritiva foi o KL-ONE criado por Ronald J. Brachman e Schmolze em 1985 [40]. O KL-ONE sinalizou a transição das redes semânticas para uma terminologia lógica mais bem fundamentada [40].

As redes semânticas surgiram por volta de 1966 como uma representação de conceitos básicos em inglês e se tornaram o tipo de modelo mais popular na época para a representação de uma grande variedade de conceitos em aplicações de inteligência artificial. Os sistemas de redes semânticas tinham problemas relacionados à imprecisão e inconsistência de significado [40].

A Figura 4.1 mostra a arquitetura do sistema de representação de conhecimento baseado em lógica descritiva adaptado de Baader [40].

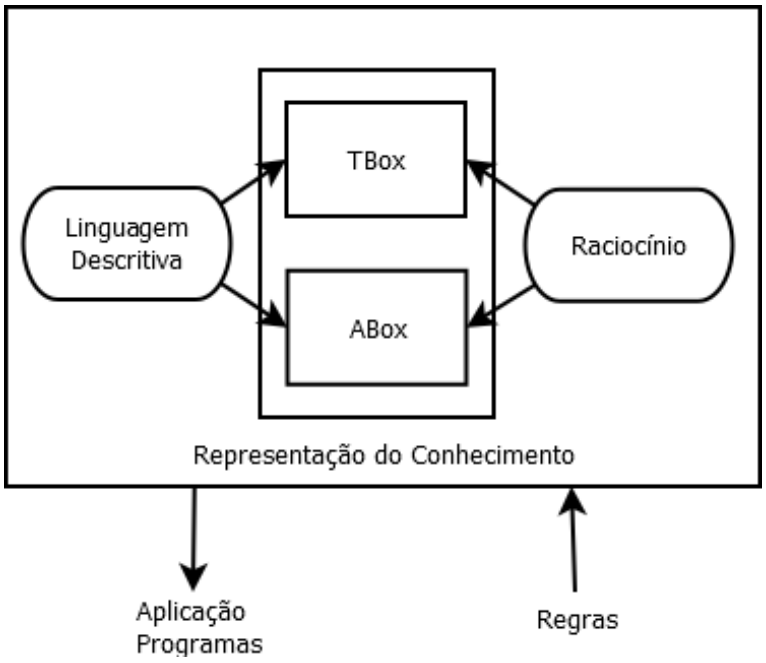

Figura 4.1: Arquitetura do Sistema de Representação de Conhecimento baseado em Lógica Descritiva adaptado de Baader.

O diagrama mostrado na Figura 4.1 mostra que a lógica descritiva tem como entrada conjuntos de regras e como saída aplicações e programas. A arquitetura internamente possui duas divisões, o *TBox* que possui o conhecimento intencional, a descrição dos conceitos e o *ABox* que possui o conhecimento estendido, específico de indivíduos em um domínio.

A linguagem utilizada na descrição é uma característica de cada sistema de lógica descritiva e vários sistemas são distinguidos pela sua linguagem de descrição. A linguagem descritiva é um modelo teórico semântico, assim as afirmações no *TBox* e *ABox* podem ser identificadas como uma forma de lógica de primeira ordem ou em alguns casos, uma extensão dela [40].

A lógica descritiva não armazena somente terminologias e afirmações, também oferece serviços que dá significado para eles. Uma típica tarefa da terminologia é determinar se uma descrição está sendo satisfeita, se uma descrição é mais generalista que outra, se a primeira está contida na segunda. Um importante problema para o *ABox* é descobrir se o conjunto de afirmações são consistentes, se ela está no modelo, se uma afirmação de um indivíduo particular é uma instância de um conceito definido [40].

A linguagem de atribuição AL *(Attributive Language)* foi introduzida por Schmidt-Schauß e Smolka em 1991 como uma linguagem minimalista de interesse prático, outras linguagens são uma extensão da linguagem AL [40].

Há uma convenção de nomes bem estabelecida para a família da lógica descritiva, o esquema de nomes pode ser resumido da seguinte forma [44]:

$$((\text{ALC}|\text{S})[\text{H}]|\text{SR})[\text{O}][\text{I}][\text{F}|\text{N}|\text{Q}]$$

ALC: é uma abreviação da linguagem de atribuição com complemento *(Attributive Languagem with Complements)*. A ALC não permite axiomas *RBox*, papel universal, papel inverso, cardinalidade fixa, conceitos nominais e autoconceitos [44].

S: adiciona à linguagem de atribuição com complemento (ALC) afirmações de transitividade, especifica axiomas de cadeia de papéis da forma $r \circ r \sqsubseteq r$ para $r \in N_R$ [44].

H: a ALC e o S podem ser estendidos pelo papel hierárquico obtendo o ALCH ou SH, na qual é permitida a inclusão de papéis simples, o encadeamento de axiomas de papéis na forma de $r \sqsubseteq s$ [44].

SR: é o ALC estendido com todos os tipos de axiomas RBox e autoconceitos [44].

O: permite que conceitos nominais sejam suportados [44].

I: permite papel inverso na linguagem descritiva [44].

F: habilita o suporte a declarações de funcionalidade do papel que podem ser expressadas como $\top \sqsubseteq \leq 1.\top$ [44].

N: permite restrições de número não qualificadas, conceitos na forma de $\geq nr.\top$ e $\leq nr.\top$ [44].

Q: permite restrições arbitrárias de números qualificados [44].

Segundo Hellmann [45], a lógica descritiva ALC *(Attributive Language with Complement)* é um fragmento da OWL *(Ontology Web Language)*, padrão da Web Semântica,

descrito na próxima seção. A ALC permite a construção de conceitos complexos de uma forma simples utilizando várias construções na linguagem.

4.1.1 Construtores de Conceitos Booleanos

Os construtores de conceitos booleanos são formados pelas operações booleanas de interseção, união, conjunto complemento, conjunção, disjunção e negação [44].

Uma inclusão de conceito nos permite afirmar que todas as mães são do sexo feminino e que todas as mães são pais. Essa afirmação pode ser construída utilizando a intersecção, também conhecida como conjunção [44]. Segue o exemplo com a referida afirmação:

$$\text{Mother} \equiv \text{Female} \sqcap \text{Parent}$$

O exemplo anterior também poderia ser escrito da seguinte forma:

$$\text{Mother} \equiv \neg\ \text{Male} \sqcap \text{Parent}$$

O exemplo anterior afirma que todas as mães não são do sexo masculino e são pais.
O exemplo a seguir utiliza a união para afirmar que pai e mãe definem pais.

$$\text{Parent} \equiv \text{Mother} \sqcup \text{Father}$$

Às vezes é útil fazer afirmações sobre cada indivíduo e afirmar que é uma coisa ou outra. O exemplo a seguir utiliza o conceito universal para afirmar que cada indivíduo é do sexo masculino ou feminino [44].

$$\top \sqsubseteq \text{Male} \sqcup \text{Female}$$

Para expressar que nenhum indivíduo pode ser simultaneamente dos sexos masculino e feminino, podemos utilizar a intersecção e descrever utilizando o seguinte axioma [44]:

$$\text{Male} \sqcap \text{Female} \sqsubseteq \bot$$

O axioma anterior compara a intersecção dos sexos masculinos e femininos com o conceito de insatisfação, afirmando que nenhum indivíduo pode ser ao mesmo tempo do sexo masculino e feminino.

4.1.2 Restrições de Papéis

Utilizando a Lógica Descritiva é possível criar afirmações que ligam conceitos e papéis. Um exemplo é a relação entre o conceito Parent e o papel *parentOf*. O papel *parentOf* representa alguém que é pai ou mãe de pelo menos um indivíduo da instância $\top$. Essa relação pode ser obtida utilizando o conceito de equivalência a seguir [44]:

$$\text{Parent} \equiv \exists \text{parentOf}.\top$$

Para representar o conjunto de indivíduos que todos seus filhos são do sexo feminino, pode ser utilizada a restrição universal de acordo com o exemplo a seguir [44]:

$$\forall \text{parentOf.Female}$$

Para representar o conjunto de indivíduos que não têm nenhum filho, pode ser utilizada a negação no axioma de acordo com o exemplo a seguir:

$$\neg\ \exists \text{parentOf.T}$$

O axioma anterior afirma que não são pais de nenhum indivíduo da instância $\top$.

A restrição numérica permite restringir o número de indivíduos que são alcançados por algum papel [44]. O exemplo a seguir descreve o conjunto de indivíduos que são filhos e têm mais pelo menos dois pais:

$$\leq 2\ \text{childOf.Parent}$$

A reflexibilidade local permite que um conjunto de indivíduos se relacionem com eles mesmos por meio de um papel [44]. O exemplo a seguir descreve um conjunto de indivíduos que conversam com eles mesmos:

$$\exists \text{talksTo.Self}$$

O sistema proposto utilizada lógica descritiva na busca das relações existentes em cada imagem descrita.

4.2 RDF

O *Resource Description Framework* (RDF) é uma tecnologia flexível capaz de resolver uma série de problemas [46]. O RDF torna possível que o conhecimento seja representado em um formato que computadores conseguem entender.

Segundo Marzano [46], o *Resource Description Framework* (RDF) é um recurso de descrição de um modelo de dados e consegue armazenar dados em um formato inteligível para computadores, dessa forma possibilita a troca de dados e processos automatizados.

De acordo com Ma [47], o *Resource Description Framework* é um modelo flexível para representar informações sobre recursos na Internet. Atualmente o RDF é bastante utilizado na representação do conhecimento e na web semântica.

Em seu trabalho Liyang [48] afirma que o *Resource Description Framework* (RDF) foi proposto em 1999 como um modelo que seria utilizado como base para a criação e processamento de metadados. Seu objetivo é ser um mecanismo utilizado na descrição de recursos sem a associação com um domínio de conhecimento em particular, dessa forma pode ser utilizado para promover a interoperabilidade entre aplicações na troca de informações inteligíveis para computadores na Internet [48].

O RDF decompõe a informação ou conhecimento em partes menores com regras semânticas para cada uma dessas partes, utiliza os seguintes componentes chave [48]:

- afirmações;
- sujeito e objeto;
- predicado.

A Figura 4.2 mostra como o conhecimento é representado no formato RDF, utilizando sujeito, predicado e objeto.

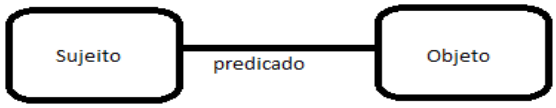

Figura 4.2: Representação do formato RDF

O sistema proposto utiliza o formato RDF para descrever as relações semânticas existentes nas imagens.

4.3 Ontologias

O termo ontologia tem origem na Filosofia e significa na Filosofia o estudo da existência [49]. A ontologia cria conhecimento comum e uma linguagem semântica não ambígua

para representar o conhecimento [50]. A definição de ontologia de Wu [50] foi aplicada no contexto da computação, mais precisamente na área de segurança da informação. A ontologia descreve conceitos e relacionamentos de alguns fenômenos do mundo [51].

Ontologias são muito utilizadas para diversas finalidades e em diferentes áreas de estudo. Ontologias são utilizadas para o processamento de linguagem natural, gerenciamento do conhecimento, comércio eletrônico, integração inteligente de informações e Web Semântica em áreas como engenharia de conhecimento, banco de dados, engenharia de *software*, etc [51].

Uma ontologia define o vocabulário que pode ser utilizado para criar consultas e afirmações em um domínio de conhecimento de forma independente dos recursos tecnológicos que serão utilizados [52] no sistema.

Através da definição do vocabulário, as ontologias permitem o compartilhamento de um entendimento comum na conversação entre os agentes de coleta, processamento, fusão e troca de informações [52].

De acordo com Kendall [52], ontologias criadas para o compartilhamento de informações podem ser utilizadas com os seguintes propósitos, mas não estão restritas somente a esses:

- nas comunicações dentro e entre grupos de pessoas utilizando seu próprio vocabulário;
- na codificação, extensão e melhoria na flexibilização dos esquemas XML e/ou RDF;
- para a organização de informação;
- para descrever recursos em um sistema de gerenciamento de conteúdo, para o seu arquivamento, no gerenciamento de portais corporativos ou em experimentos científicos e reuso.

Ontologias que descrevem informações de recursos, processos ou aplicações são frequentemente construídas para suportar a resposta de questões, através de linguagens de consulta tradicionais como SQL ou SPARQL, ou através de regras de negócio, podendo utilizar linguagens de regras como RuleML, Jess, Flora-2 e entre outras linguagens comerciais [52].

De acordo com Man [53], ontologias são utilizadas na inteligência artificial, web semântica, engenharia de sistemas, engenharia de *software*, arquitetura da informação, como uma forma de representar o conhecimento do mundo ou de parte dele.

Pandey [54] afirma que as ontologias são componentes chave na representação do conhecimento na web semântica.

Euzenat [55] em seu trabalho afirma que ontologias são expressadas em uma linguagem ontológica, a semântica provê regras para interpretar a sintaxe que não fornece significado diretamente e limita as possibilidades de interpretação do que foi declarado.

A maioria das linguagens ontológicas utilizam os mesmos tipos de entidades, com diferentes nomes, mas com interpretação compatível [55].

Linguagens ontológicas possuem geralmente os seguintes tipos de entidades [55]:

- classes ou conceitos: são as entidades principais em uma ontologia;
- indivíduos, objetos ou instâncias: são interpretadas como um indivíduo particular em um domínio;
- relações: são relacionamentos, são interpretadas como o subconjunto de um produto no domínio;
- tipos de dados: são a parte do domínio onde os tipos de valores são especificados;
- valores de dados: são simplesmente valores no domínio.

Segundo Euzenat [55], entidades em uma ontologia podem ser conectadas por meio de vários tipos de relacionamentos, incluindo os seguintes:

- especialização ou absorção entre duas classes ou propriedades;
- exclusão ou disjunção entre duas classes ou duas propriedades;
- instanciação entre indivíduos e classes.

Keet [56] afirma que a linguagem OWL *(Web Ontology Language)* é a linguagem ontológica mais utilizada com propósito computacional e é o padrão recomendado pelo W3C *(World Wide Web Consortium)* . A linguagem OWL foi padronizado primeiro em 2004, a nova versão foi padronizada em 2009 e possibilitou o desenvolvimento de ferramentas e o desenvolvimento de ontologias para sistemas de informação orientados a ontologias [56].

A Tabela 4.2 mostra a sintaxe ALC e a semântica com a construção da OWL correspondente [45].

A Tabela 4.2 mostra a equivalência entre a *OWL*, a lógica descritiva e a sintaxe semântica. A *OWL* em sua essência utiliza lógica descritiva.

Segundo Grigoris [57], as mais importantes linguagens ontológicas da atualidade são o RDF *(Resource Description Framework)* e OWL.

A linguagem RDF é uma linguagem de descrição de vocabulário para descrever propriedades e classes de recursos RDF [57].

A linguagem OWL é uma rica linguagem de descrição de vocabulário que descreve propriedades e classes, relações entre classes, cardinalidade, igualdade, características de propriedades e classes enumeradas [57].

OWL	DL	Sintaxe DL	Semântica
named class	Conceito atômico	A	$A^I \sqsubseteq \triangle^I$
object property	Papel abstrato	r	$r^I \sqsubseteq \triangle^I x \triangle^I$
Thing	Conceito universal	$\top$	$\triangle^I$
Nothing	Conceito de insatisfação	$\bot$	$\oslash$
intersectionOf	Conjunção	$C \sqcap D$	$(C \sqcap D)^I = C^I \sqcap D^I$
unionOf	Disjunção	$C \sqcup D$	$(C \sqcup D)^I = C^I \sqcup D^I$
complementOf	Negação	$\neg C$	$(\neg C^I) = \triangle^I \backslash C^I$
someValuesFrom	Restrição existencial	$\exists r.C$	$(\exists r.C)^I = \{a \mid \exists b.(a,b) \in r^I and\ b \in C^I\}$
allValuesFrom	Restrição de valor	$\forall r.C$	$(\forall r.C)^I = \{a \mid \forall b.(a,b) \in r^I implica\ b \in C^I\}$

Tabela 4.2: Sintaxe ALC e semântica com construção OWL correspondente.

De acordo com Keet [56], a linguagem OWL considera os mesmos princípios básicos da lógica descritiva.

Ontologias são fundamentais para a construção de uma linguagem com entendimento comum em um domínio específico e para a web semântica na representação do conhecimento.

A atividade de descrição das relações existentes nas imagens foi realizada utilizando ontologia criada neste trabalho. Tal ontologia possui o vocabulário necessário para descrever as relações existentes nas imagens que pretendemos descrever.

4.4 Considerações Finais

Este capítulo abordou temas relevantes para este trabalho como lógica descritiva, RDF e Ontologia. A lógica descritiva foi utilizada no sistema proposto na busca das relações semânticas existentes em cada imagem durante o processo de autenticação. O RDF foi utilizado no sistema proposto como o formato de descrição das relações semânticas existentes nas imagens. A ontologia foi utilizada no trabalho proposto como o vocabulário disponível para descrever as relações semânticas existentes nas imagens.

O próximo capítulo mostra as melhorias implementadas no sistema de autenticação proposto em relação aos outros sistemas de autenticação exibidos neste trabalho.

Capítulo 5

Autenticação Semântica Gráfica

Este capítulo contempla as seguintes fases do *Design Science Research*: proposição de artefato, projeto do artefato, desenvolvimento do artefato e generalização para uma classe de problemas. Serão exibidas as melhorias implementadas no sistema de autenticação proposto em relação aos outros sistemas de autenticação exibidos neste trabalho.

Este capítulo tem como objetivo mostrar o funcionamento do sistema de autenticação desenvolvido e o estudo de caso realizado envolvendo a utilização da API desenvolvida.

5.1 Melhorias Propostas

Esta seção apresenta as melhorias que este trabalho propõe em relação aos trabalhos no estado da arte pesquisados e considerados mais relevantes nesta pesquisa.

Almulhem [7] propôs um sistema de autenticação gráfica, onde o usuário pode:

- escolher pontos específicos em uma foto;
- inserir texto em pontos específicos;
- definir a ordem de seleção dos pontos.

No sistema de autenticação de Almulhem [7], a imagem definida pelo usuário é apresentada, o usuário deve marcar os pontos na posição correta e na ordem definida, além de ter que inserir os textos correspondentes nos pontos, caso tenha cadastrado algum texto no processo de criação da senha.

Diferente do sistema proposto por Almulhem [7], este trabalho propõe um sistema onde o usuário não terá a relação de sua semântica associada somente a uma imagem específica ou terá que clicar em pontos específicos predefinidos em imagem que foi selecionada anteriormente, pois utilizará uma relação que estará presente em imagens diferentes e quanto maior for a quantidade de imagens na base de dados, menor será a probabilidade de repetição para o usuário durante o processo de autenticação.

Salembier et al. [31] e Lorant et al. [30] em seus trabalhos propuseram o sistema de autenticação semântica humana onde o usuário possui uma senha conceitual com quatro conceitos em uma ordem definida.

No sistema de Salembier et al. [31] e Lorant et al. [30] durante o processo de autenticação, o usuário deve identificar os quatro conceitos de sua senha conceitual clicando nas áreas das imagens onde os conceitos estão presentes. É exibida para o usuário apenas uma imagem por vez e o mesmo deve identificar o respectivo conceito clicando em sua área correspondente na imagem.

O sistema proposto por Salembier et al. [31] e Lorant et al. [30] necessita que os conceitos sejam descritos com a sua codificação em cada imagem e em sua respectiva área na imagem. O sistema proposto neste trabalho não utilizará codificação em áreas da imagem e nem apresentará apenas uma imagem para o usuário por vez, conforme no sistema de autenticação semântica humana de Salembier et al. [31] e Lorant et al. [30].

Este trabalho propõe a exibição de imagens simultaneamente, para que o usuário selecione aquela que tenha a relação correspondente em sua senha semântica e não será necessário que o usuário clique na área onde está presente a relação, será necessário apenas a seleção da imagem correta, aumentando a dificuldade na identificação da relação escolhida por um observador externo, aumentando a resistência do sistema de autenticação contra o ataque *shoulder surfing* em relação ao sistema autenticação semântica humana.

Diferentemente dos trabalhos de Almulhem [7], Salembier et al. [31] e Lorant et al. [30], este trabalho não vincula as relações da senha semântica a áreas específicas nas imagens onde os usuários devem clicar durante o processo de autenticação para identificar as relações, as relações são associadas as imagens e descritas utilizando lógica descritiva, mais especificamente, utilizando o formato RDF *(Resource Description Framework)*.

5.2 Genoma Visual

Em seu trabalho Krishna et al. [58] apresentam coleção de dados intitulada Genoma Visual *(Visual Genome dataset)* com objetivo de possibilitar que os objetos e seus relacionamentos nas imagens possam ser descritos de forma a gerar conhecimento de máquina. O conjunto de dados do Genoma Visual é utilizado no reconhecimento de objetos em imagens, seus possíveis relacionamentos com outros objetos na imagem e na reposta de perguntas relacionadas aos objetos, e seus relacionamentos.

O conjunto de dados do Genoma Visual [58] possui aproximadamente 108 mil imagens, onde cada imagem possui uma média de 35 objetos, 26 atributos e 21 relacionamentos entre objetos. O *dataset* do Genoma Visual já foi utilizado em diversos trabalhos acadêmicos [59], [60], [61], [62], [63], [64], [65], [66], [67].

A Figura 5.1 será descrita utilizando o modelo utilizado no trabalho de Krishna et al. [58] na qual os objetos são identificados na imagem e o relacionamento entre esses objetos são descritos. Os objetos neste modelo podem possuir atributos.

Figura 5.1: Imagem do *dataset* do Genome Visual

A Figura 5.2 mostra como a Figura 5.1 pode ser descrita utilizando o modelo utilizado no trabalho de Krishna et al. [58]. Na descrição realizada da Figura 5.1, o objeto *Man* possui os relacionamentos: *Wearing*, *Playing*, *Holding*, *Using* com os objetos *Tennis*, *Shorts*, *T-shirt*, *Racket*, *Watch* e *Sunny Glasses*, e os objetos *Tennis*, *Shorts*, *T-shirt* possuem o atributo *White*.

A Figura 5.2 mostra como seria a descrição do homem na Figura 5.1 utilizando o modelo proposto por Krishna et al..

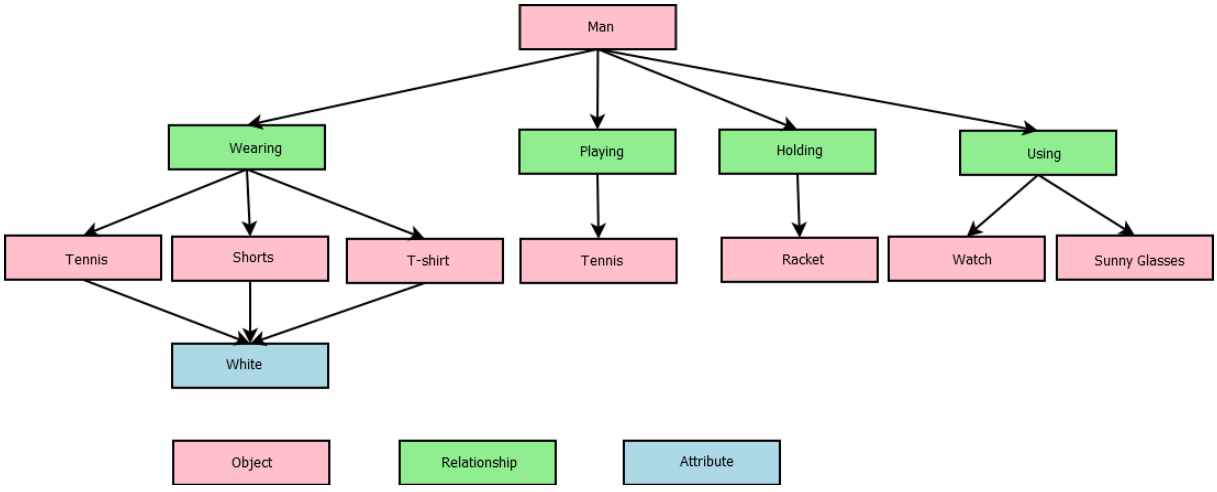

Figura 5.2: Descrição da Figura 5.1 utilizando o modelo proposto por Krishna et al.

Neste trabalho foram utilizadas imagens do *dataset* do Genoma Visual, não foi possível realizar o aproveitamento das descrições das imagens porque foram descritas utilizando

linguagem natural e não foi utilizado um vocabulário comum nas descrições. Krishna et al. [58] em seu trabalho utilizaram objeto, relacionamento, objeto e atributo, este trabalho propõe na descrição das imagens a utilização de uma ontologia e construções utilizando o formato sujeito, predicado e objeto.

5.3 Trabalho Proposto

A solução proposta tem como objetivo simplificar a utilização de uma segunda autenticação em conjunto com a autenticação utilizando usuário e senha, eliminando a necessidade de implementação de dispositivos auxiliares de autenticação ou a exigência de que o usuário tenha algum dispositivo em sua posse para utilizar de forma complementar no processo de autenticação.

O objetivo deste trabalho é desenvolver uma solução de autenticação semântica gráfica que possa ser utilizada como um sistema de autenticação complementar para qualquer sistema, aumentando a segurança no processo de autenticação, mas sem gerar impacto significativo no processo de autenticação da solução para os usuários finais e sem gerar custo adicional com aquisição de dispositivos em sua implantação. A validação é realizada com a comparação das relações vinculadas a imagem com as relações da senha semântica definidas e armazenadas em banco de dados para o referido usuário.

A Figura 5.3 mostra fluxo proposto na utilização da autenticação semântica gráfica.

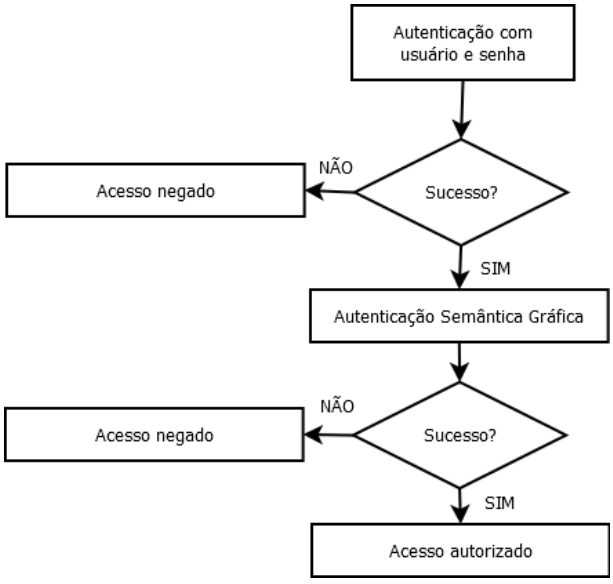

Figura 5.3: Fluxo proposto na utilização da autenticação semântica gráfica.

A Figura 5.3 mostra o fluxo proposto para a autenticação semântica gráfica quando utilizada em conjunto com a autenticação por meio de usuário e senha.

O fluxo mostrado na Figura 5.3 mostra as etapas do processo de autenticação no sistema proposto:

- o usuário deve autenticar utilizando o seu usuário e senha;
- se a autenticação executada no passo anterior ocorrer com sucesso, a autenticação semântica gráfica é iniciada e se a autenticação executada no passo anterior não ocorrer com sucesso, o usuário não inicia a autenticação semântica gráfica e o acesso é negado;
- se a autenticação semântica gráfica ocorrer com sucesso, o acesso ao sistema será autorizado, se a autenticação semântica gráfica não ocorrer com sucesso, o acesso é negado.

Na solução proposta existe a necessidade de realização de descrição das imagens adotando um padrão. O padrão adotado é definido por meio de ontologia.

O conhecimento será descrito utilizando lógica descritiva, mas de forma que o dado possa ser armazenado em banco de dados relacional, possa ser vinculado a imagem que será armazenada no sistema de arquivos. É importante que possam ser pesquisadas neste banco de imagens, imagens que possuam as relações da senha semântica do usuário que serão exibidas durante o processo de autenticação.

Após o usuário selecionar a imagem que possui a primeira relação de sua senha semântica, uma nova pesquisa será realizada pelo sistema no banco de dados para escolher a imagem que possui a segunda relação da senha semântica e esse processo se repetirá até que todas as relações da senha semântica sejam validadas.

Durante o processo de autenticação, o usuário deve autenticar utilizando usuário e senha normalmente, após a validação destas credenciais, o usuário terá que realizar a autenticação complementar no sistema de autenticação semântica gráfica.

A Figura 5.4 mostra o processo de descrição semântica de uma imagem no sistema de autenticação semântica gráfica.

A Figura 5.5 mostra como são apresentadas as imagens para o usuário escolher aquela que possui a relação de sua senha semântica durante o processo de autenticação.

Se cada imagem exibida no processo de autenticação no sistema proposto possuir a mesma quantidade de relações que de conceitos no sistema de autenticação semântica humana *HSA*, o sistema proposto terá como quantidade de relações disponíveis a soma das relações existentes em cada uma das imagens exibidas simultaneamente, excluindo as relações repetidas, dificultando a inferência humana da senha semântica por meio da observação. É importante ressaltar que no processo de autenticação podem ser exibidas várias imagens simultaneamente.

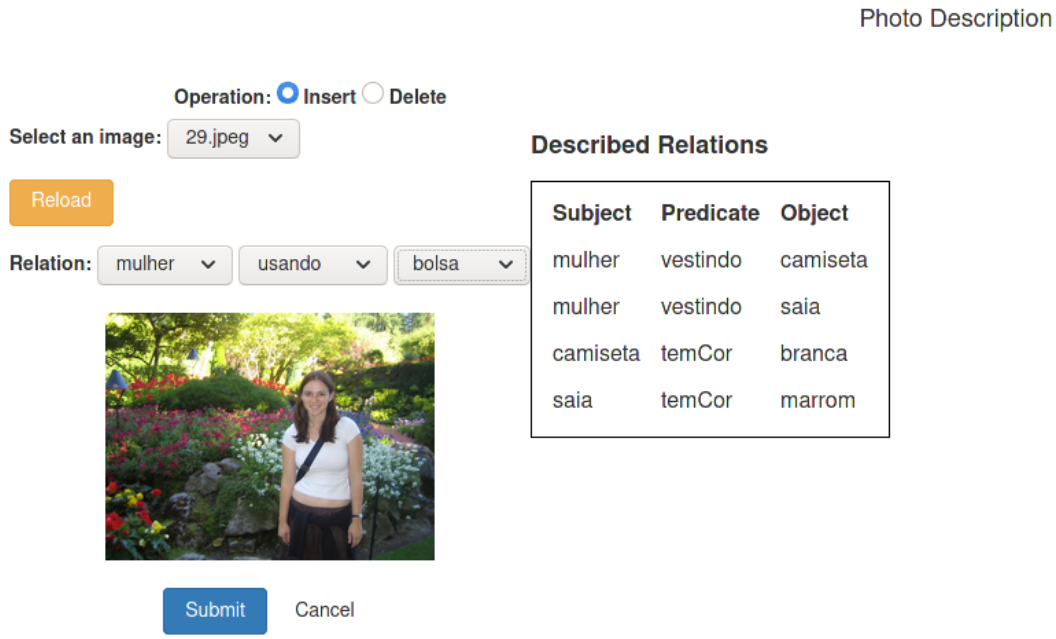

Figura 5.4: Processo de descrição semântica de imagem

A operação 5.1 mostra axioma que descreve a operação realizada para selecionar no banco de dados as imagens que possuem a relação da senha semântica do usuário. Após o retorno da operação, uma imagem entre as retornadas pela consulta será selecionada de forma aleatória e a sua posição na tela de apresentação para o usuário também será selecionada de forma aleatória.

$$Predicado(Sujeito, Objeto) \sqsubseteq Imagem \tag{5.1}$$

O axioma anterior retornará apenas imagens que possuem o sujeito, predicado e objeto informados.

A operação 5.2 mostra axioma que descreve a operação realizada para selecionar no banco de dados as imagens que não possuem a relação da senha semântica do usuário. Após o retorno da operação, três imagens entre as retornadas pela consulta serão selecionadas de forma aleatória e as suas posições na tela de apresentação para o usuário serão selecionadas de forma aleatória.

$$\neg\exists Predicado(Sujeito, Objeto) \sqsubseteq Imagem \tag{5.2}$$

O axioma anterior retornará apenas imagens que não possuem o sujeito, predicado e objeto informados.

Figura 5.5: Tela de autenticação

Após as operações exibidas em 5.1 e 5.2, o sistema de autenticação proposto terá as imagens necessárias para realizar a validação da primeira relação da senha semântica do usuário e este processo será repetido até que o sistema valide todas as relações da senha semântica do usuário ou até que o usuário erre a quantidade de vezes necessária para concluir o processo de autenticação sem sucesso.

A Figura 5.6 mostra o fluxo seguido durante o processo de autenticação na autenticação semântica gráfica.

No fluxo descrito na Figura 5.6, a autenticação semântica gráfica segue os seguintes passos:

- o sistema pesquisa a senha semântica do usuário no banco de dados;
- o sistema pesquisa no banco de dados imagens que possuem a relação n da senha semântica do usuário;
- o sistema seleciona aleatoriamente uma imagem entre as imagens obtidas no passo anterior;
- o sistema pesquisa no banco de dados imagens que não tenham a relação n da senha semântica do usuário;

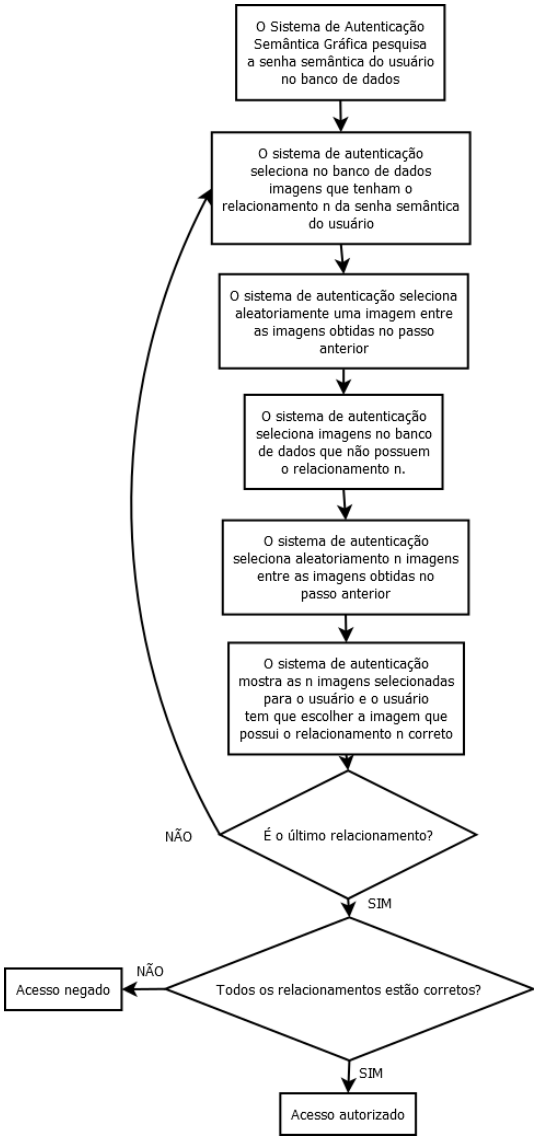

Figura 5.6: Fluxo da autenticação semântica gráfica.

- o sistema seleciona aleatoriamente três imagens entre as imagens obtidas no passo anterior;
- o sistema exibe as quatro imagens selecionadas para o usuário;
- o usuário deve que escolher a imagem correta;
- este processo é repetido até a última relação da senha semântica;
- se o usuário selecionar corretamente todas as imagens, ele será autenticado com sucesso no sistema;
- se o usuário não selecionar corretamente todas as imagens, a autenticação irá reiniciar ou falhar.

A Figura 5.7 mostra o mapa conceitual da ontologia que será utilizada para realizar a descrição das imagens no estudo de caso e na avaliação.

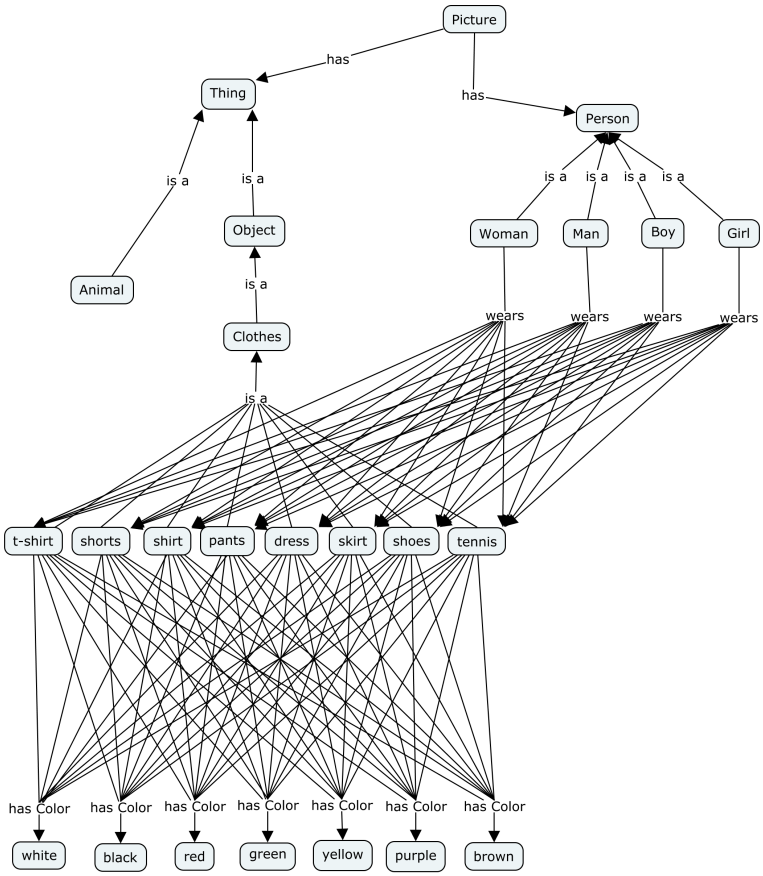

Figura 5.7: Mapa conceitual da ontologia utilizada no estudo de caso e na avaliação.

É possível observar na Figura 5.7 que utilizando esta ontologia é possível utilizar mais de 200 relações distintas na descrição de elementos nas imagens, mostrando a robustez da ontologia criada.

O esquema de descrição da Figura 5.1 utilizando o modelo utilizado no trabalho de Krishna et al. [58] foi exibido na Figura 5.2 no capítulo 3. Os axiomas a seguir mostram como a Figura 5.1 poderia ser descrita utilizando lógica descritiva.

$$wearing(Man, Tennis) \circ hasColor(Tennis, White) \tag{5.3}$$

$$wearing(Man, Shorts) \circ hasColor(Shorts, White) \tag{5.4}$$

$$wearing(Man, T-shirt) \circ hasColor(T-shirt, White) \tag{5.5}$$

$$wearing(Man, T-shirt) \tag{5.6}$$

$$wearing(Man, Shorts) \tag{5.7}$$

$$wearing(Man, Tennis) \tag{5.8}$$

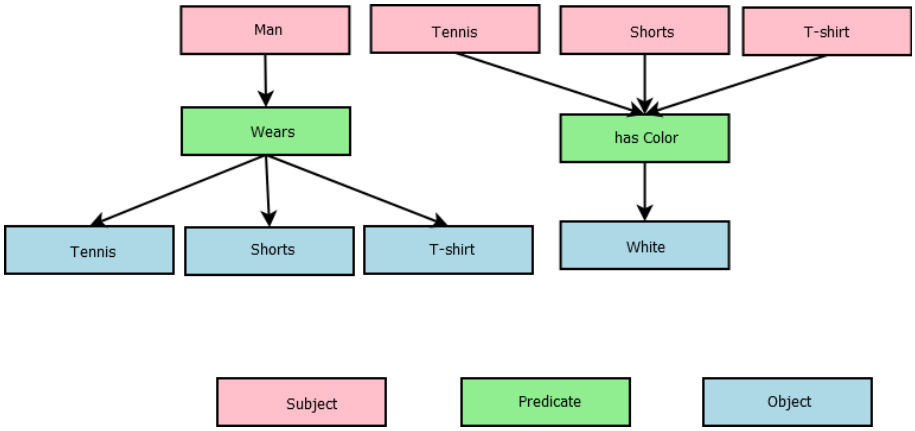

Figura 5.8: Relações da Figura 5.1 no formato RDF.

Utilizando o mapa conceitual da ontologia exibida na Figura 5.7, é possível identificar as relações associadas ao homem que poderiam ser utilizadas como uma relação da senha semântica. Poderiam ser utilizadas as seguintes relações:

- Homem calçando tênis branco;
- Homem vestindo shorts branco;
- Homem vestindo camiseta branca;
- Homem vestindo camiseta;
- Homem vestindo shorts;
- Homem calçando tênis.

Apenas com o único sujeito que foi descrito na Figura 5.8 foi possível descrever várias relações existentes, mostrando que uma imagem pode possuir várias relações, possibilitando que imagens sejam utilizadas no processo de autenticação.

O sistema proposto utiliza relações binárias para conectar sujeito ao objeto utilizando o predicado. Não foram utilizadas relações unárias, porque este é um tipo de relacionamento fácil de descobrir e deixaria o sistema de autenticação mais vulnerável.

Não foram utilizadas relações ternárias porque elas aumentam a complexidade do sistema e a simplicidade do sistema proposto é uma vantagem, mas não descartamos sua utilização no futuro. Um problema que identificamos com a relação ternária é que ela iria deixar mais difícil a tarefa de criar a senha semântica para o usuário e a descrição das imagens.

5.4 Segurança do Sistema

A equação 5.9 descreve as possibilidades de descoberta da senha semântica na autenticação semântica gráfica quando o atacante conhece a ontologia utilizada e os conceitos mostrados para o usuário não se repetem entre as imagens na mesma tela.

$$L = (C \times I)^P \tag{5.9}$$

Na equação 5.9, L é a quantidade de possibilidades durante a autenticação, C é o número de relações por imagem, I é o número de imagens mostradas simultaneamente para o usuário e P é o tamanho da senha semântica.

A equação 5.10 descreve as possibilidades se o usuário tentar autenticar clicando nas imagens aleatoriamente.

$$L = I^P \tag{5.10}$$

Na equação 5.10, L é a quantidade de possibilidades, I é o número de imagens mostradas simultaneamente para o usuário e P é o tamanho da senha semântica.

Mesmo que o atacante escolhendo imagens de forma aleatória durante o processo de autenticação consiga autenticar uma vez, o mesmo não tem garantia de que conseguirá autenticar com sucesso novamente, pois não conseguiu identificar as relações da senha semântica do usuário alvo do ataque.

É importante ressaltar que em um sistema de autenticação utilizando usuário e senha, cada vez que o atacante tenta autenticar, aumenta a possibilidade de sucesso, pois na medida que vai tentando, as possibilidades diminuem, contudo no sistema de autenticação semântica gráfica, cada seleção de imagens é um evento independente. Assim, um atacante não acumula informação sobre o espaço de busca. Logo, se o usuário não for capaz de identificar cada conceito existente nas imagens, o sistema de autenticação não irá se comportar estatisticamente da mesma forma que o sistema de autenticação com usuário e senha, impossibilitando o ataque de força bruta.

No sistema de autenticação semântica gráfica, cada vez que o atacante tenta autenticar escolhendo de forma aleatória as imagens, a probabilidade de sucesso é a mesma, pois o mesmo não é capaz de eliminar combinações, porque em cada autenticação imagens diferentes são exibidas e as imagens corretas sempre são exibidas em posições selecionadas aleatoriamente, mantendo as possibilidades exibidas na equação 5.10 inalteradas.

Como forma de mitigar a possibilidade de quebra do mecanismo de autenticação semântico gráfico, o sistema pode bloquear a autenticação caso n erros na autenticação sejam identificados. Tal bloqueio pode gerar alerta para o administrador do sistema e/ou

usuário, mostrando que o usuário é alvo de ataque e que a sua senha textual foi comprometida, pois foi utilizada com sucesso no primeiro mecanismo de autenticação.

5.5 Estudo de Caso

O estudo de caso teve como objetivo integrar o sistema de autenticação proposto com a autenticação do WordPress. A integração do sistema proposto com o WordPress é por meio de API *(Application Programming Interface)* utilizada pelo sistema WordPress após a validação do usuário e senha inseridos na primeira fase da autenticação. Durante o processo de autenticação no sistema de Autenticação Semântica Gráfica, são exibidas para o usuário quatro imagens de forma aleatória e uma dessas imagens possui a relação da senha semântica do usuário. O usuário precisa validar os três conceitos de sua senha semântica. Apesar do WordPress ter sido escolhido para o teste de integração com a API do sistema proposto, essa autenticação por meio de API funciona para qualquer sistema *Web*.

5.5.1 WordPress

O WordPress é um sistema de gestão de conteúdo *(CMS - Content Management System)* de código aberto *(Open Source)* desenvolvido em PHP e nativamente utiliza banco de dados MySQL. O WordPress é bastante utilizado como plataforma para blogs e portais na Internet. A Figura 5.9 mostra a tela da gestão de conteúdo do WordPress.

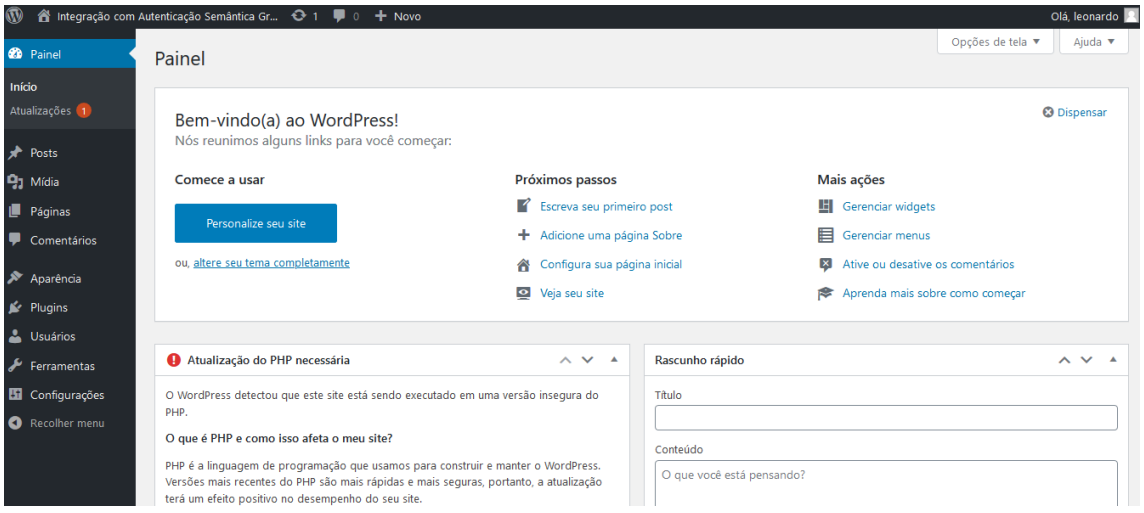

Figura 5.9: Tela Inicial da Gestão de Conteúdo do WordPress

O WordPress é utilizado por vários órgãos da União, dos Estados e Municípios em sites institucionais e até sistemas. Devido a popularidade do WordPress na Internet, este

(CMS) tem sido alvo constante de ataques que exploram novas vulnerabilidades em seu código principal e/ou plugins.

Existem ferramentas disponíveis na Internet que auxiliam os atacantes em ataques contra o WordPress. Visando mitigar os riscos relacionados à quebra do sistema de autenticação padrão do WordPress, usuário e senha, o WordPress foi integrado ao sistema de Autenticação Semântica Gráfica desenvolvido, por meio de API.

A Figura a seguir 5.10 mostra uma das arquiteturas mais simples do WordPress.

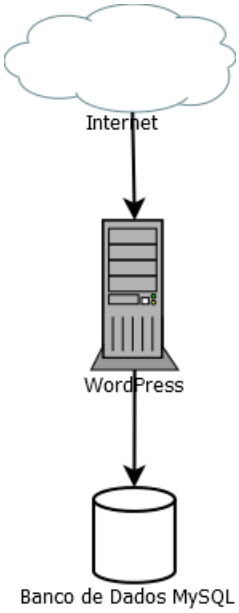

Figura 5.10: Arquitetura do WordPress

A Figura 5.10 mostra a simplicidade na implantação do WordPress. Nesta arquitetura o usuário conecta por meio da Internet ou rede local na única instância de WordPress que conecta no banco de dados MySQL que pode estar no mesmo servidor do WordPress ou em um servidor de banco de dados dedicado/remoto.

Integração

A integração com a solução de autenticação proposta entre o WordPress e o servidor do Sistema de Autenticação Semântica Gráfica será por meio de API REST.

A Figura 5.11 descreve a integração do WordPress com o Sistema de Autenticação Semântica Gráfica proposto.

O Sistema de Autenticação Semântica Gráfica utiliza banco de dados relacional PostgreSQL e neste banco são armazenados:

- metadados com informações sobre as imagens;
- ontologias;

- sujeitos possíveis nas imagens;
- predicados das ontologias;
- objetos possíveis nas imagens;
- descrições semânticas das imagens;
- senhas semânticas dos usuários.

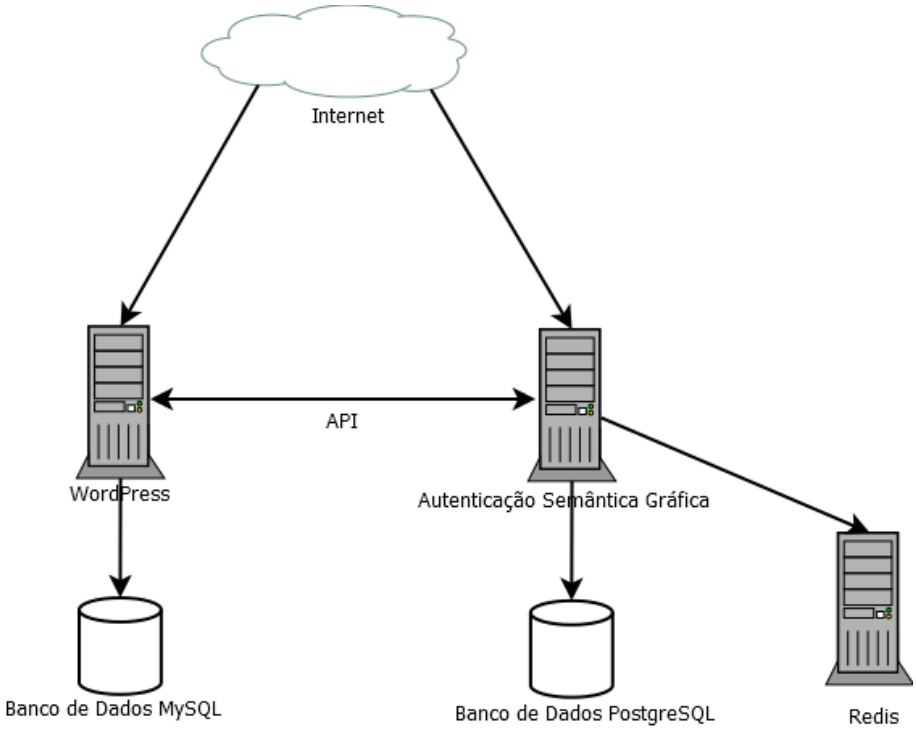

Figura 5.11: Arquitetura de Autentição Proposta

5.5.2 API

Esta subseção descreve de forma detalhada a integração realizada entre o WordPress e o sistema de autenticação proposto por meio de API.

A autenticação por meio de API ocorre da forma descrita a seguir.

Após a autenticação do usuário com sucesso no WordPress por meio de seu usuário e senha, as seguintes operações são realizadas pelo WordPress:

- gera código alfanumérico de 16 caracteres que será utilizado como vetor de inicialização pelo algoritmo de criptografia AES 128;
- gera código alfanumérico de 64 caracteres que será utilizado como código pelo usuário na URL de destino (Sistema de autenticação semântica gráfica);
- cifra com chave de criptografia e o vetor de inicialização gerado informações como: data e hora em UTC, e-mail do usuário autenticado e endereço IP do cliente;
- envia via método HTTP POST para o sistema de autenticação semântica gráfica a identificação do sistema, código alfanumérico gerado de 64 caracteres, o vetor de inicialização e os dados criptografados;
- redireciona o usuário para o contexto /auth_to_3parties/código_de_64_caracteres no sistema de autenticação semântica gráfica.

O sistema proposto recebe a requisição do usuário por meio do contexto:
/auth_to_3parties/código_de_64_caracteres.
As seguintes operações são realizadas pelo sistema de autenticação proposto:

- obtém da requisição o código de 64 caracteres informado por meio do método HTTP GET;
- utilizando o código de 64 caracteres, localiza informações como: identificação do sistema, vetor de inicialização utilizado na criptografia dos dados e os dados criptografados que foram recebidos anteriormente via método HTTP POST do WordPress;
- localiza no banco de dados a chave de criptografia associada ao identificador do sistema informado;
- utiliza a chave de criptografia associada ao identificador do sistema e o vetor de inicialização informado para decifrar os seguintes dados criptografados: data e hora, e-mail do usuário autenticado e endereço IP;
- se a diferença de tempo entre o horário informado pelo WordPress e o horário do servidor de autenticação for maior que 5 minutos para mais ou para menos, o acesso é negado;

- verifica se o redirecionamento foi realizado da origem esperada, caso não tenha sido, o acesso é negado;
- verifica se o endereço IP do cliente é o mesmo informado pelo WordPress, caso não seja, o acesso é negado;
- verifica se o código passado na requisição pelo usuário ainda existe, porque um código após a sua utilização é removido e não pode mais ser utilizado, caso não exista, o acesso é negado;
- verifica se o código utilizado no link está correto, caso não esteja, o acesso é negado;
- verifica se o e-mail informado possui conta válida no sistema de autenticação, caso não possua, o acesso é negado;
- se passou em todas as verificações com sucesso, o usuário é redirecionado para o contexto de autenticação /check_semantic_password_to_3parties, caso não tenha passado nas verificações, o acesso é negado.

Após a finalização da autenticação do usuário no sistema de autenticação semântica gráfica, as seguintes operações são realizadas:

- o sistema envia via método HTTP POST para o WordPress o código de 64 caracteres recebido e o resultado da autenticação cifrado com a chave de criptografia associada ao identificador do sistema e com o vetor de inicialização informado pelo WordPress;
- o usuário é redirecionado para o WordPress. Caso a autenticação semântica gráfica tenha ocorrido com sucesso, o usuário já acessa o WordPress autenticado, caso a autenticação semântica gráfica não tenha ocorrido com sucesso, o usuário visualizará a tela de autenticação inicial do WordPress.

5.6 Considerações Finais

Este capítulo expos as melhorias realizadas na autenticação semântica gráfica em relação aos outros trabalhos abordados no capítulo anterior, foi apresentado o projeto Genoma Visual e como esse projeto contribuiu com essa pesquisa, foi apresentada a autenticação semântica gráfica de forma detalhada, foi apresentado o estudo de caso com o WordPress e o desenvolvimento da API que possibilitou essa integração com o WordPress.

O próximo capítulo mostra a avaliação do sistema de autenticação proposto e os resultados obtidos nessa avaliação.

Capítulo 6

Avaliação do Sistema de Autenticação Semântica

Este capítulo contempla as seguintes fases do *Design Science Research*: avaliação do artefato e explicitação das aprendizagens. O objetivo deste capítulo é mostrar os resultados obtidos nessa avaliação.

6.1 Resultado da Avaliação do Sistema de Autenticação

Durante a realização da avaliação várias informações foram coletadas para que os seguintes aspectos do sistema proposto fossem avaliados:

- usabilidade do sistema de acordo com a escala de usabilidade de sistemas *(System Usability Scale)* [1];
- grau de dificuldade na criação da senha semântica;
- grau de dificuldade nas autenticações utilizando de 2 até 9 imagens;
- tempo gasto no processo de autenticação e sua relação com a quantidade de imagens exibidas simultaneamente;
- percentual de erros e acertos de acordo com a quantidade de imagens exibidas;
- resistência do sistema ao ataque *shoulder surfing*.

[1] Gao et al. [68] em seu trabalho afirmam que a escala de usabilidade de sistema *(System Usability Scale - SUS)* é um instrumento utilizado na avaliação da usabilidade de produtos e sistemas. O SUS é um instrumento de medida muito utilizado atualmente.

A Figura 6.1 mostra gráfico em pizza com o percentual da distribuição dos participantes da avaliação de acordo com a faixa etária.

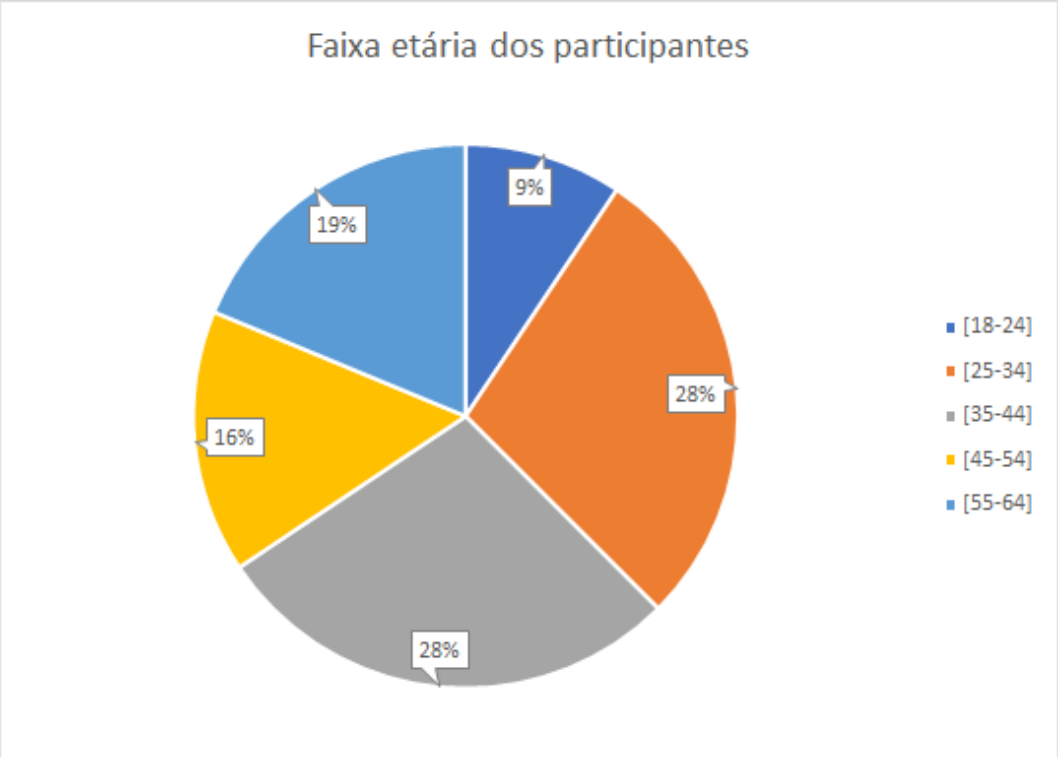

Figura 6.1: Gráfico com a faixa etária dos participantes

Na Figura 6.1 é possível visualizar os percentuais relacionados à distribuição dos participantes de acordo com a sua faixa etária. Participaram da avaliação, participantes de diversas idades e após a análise dos resultados, concluímos que a idade não influencia na usabilidade do sistema.

Não participaram da avaliação menores de 18 anos, porque para que menores de idade participem de qualquer tipo de pesquisa, é necessário obter a autorização de um dos responsáveis legais. Seria complexo assegurar que realmente foi um dos responsáveis legais que autorizou a participação do menor na pesquisa e teríamos que implementar alguns controles extras, para validar essas liberações, aumentando a complexidade do sistema de avaliação e sem trazer algum ganho significativo para pesquisa que justifique esse aumento na complexidade.

A avaliação de usabilidade do sistema foi realizada por meio de questionário respondido no final da avaliação pelos participantes. Foi utilizada como resposta a escala de Likert[2] variando de 1 até 5.

[2]Segundo Wu e Leung [69], a escala de Likert é muito utilizada em pesquisas e é comumente construída utilizando de 4 até 7 pontos. A escala de Likert é uma escala de medida ordinal e é um padrão de mercado utilizado por muito tempo.

A Tabela 6.1 mostra os resultados obtidos por meio do questionário de avaliação da usabilidade com a margem de erro utilizando nível de confiança de 95%.

Pergunta	**Média**	**Resultado**
Poderia usar esse sistema com frequência.	4,1875 ± 0,26	Concordo parcialmente
Acho o sistema muito complexo.	2,5 ± 0,34	Discordo parcialmente-Indiferente
Achei o sistema fácil de usar.	4.09375 ± 0,31	Concordo parcialmente
Acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema.	1.71875 ± 0,34	Discordo totalmente-Discordo parcialmente
Acho que as várias funções do sistema estão muito bem integradas.	3.59375 ± 0,36	Indiferente-Concordo parcialmente
Acho que o sistema apresenta muita inconsistência.	2.28125 ± 0,38	Discordo parcialmente
Imagino que as pessoas aprenderão como usar esse sistema rapidamente.	4.09375 ± 0,31	Concordo parcialmente
Achei o sistema atrapalhado de usar.	1.96875 ± 0,33	Discordo parcialmente
Me senti confiante ao usar o sistema.	4.15625 ± 0,26	Concordo parcialmente
Precisei aprender várias coisas novas antes de conseguir usar o sistema.	1.5625 ± 0,29	Discordo totalmente-Discordo parcialmente

Tabela 6.1: Perguntas sobre a avaliação da usabilidade.

Os resultados obtidos na pesquisa e mostrados na Tabela 6.1 demonstram que segundo a avaliação dos participantes:

- o sistema poderia ser utilizado com frequência. Esse resultado mostra que esse sistema de autenticação poderia ser utilizado diariamente;
- o grau de complexidade do sistema não é fácil e nem difícil. Isso mostra que é um sistema viável sob o ponto de vista de sua complexidade;
- o sistema é fácil de utilizar. A facilidade na utilização é um critério muito importante para o sucesso de um produto ou sistema, pois sistemas muito complexos podem dificultar a sua utilização e gerar resistência quanto a sua utilização;

- a maior parte dos usuários não precisariam de auxílio técnico para utilizar o sistema. É importante que seja realizada uma demonstração do funcionamento do sistema para que os poucos usuários que informaram que precisariam de auxílio técnico consigam assimilar as informações necessárias para utilizar o sistema sem dúvidas;
- não existe concordância sobre a integração das funções do sistema, os participantes não concordaram e nem discordaram. Esse resultado mostra que não existe crítica da parte dos participantes sobre a integração das funções do sistema;
- o sistema não apresenta inconsistências de acordo com a maior parte dos usuários. Sobre os usuários que informaram que o sistema apresenta algum tipo de inconsistência, alguns desses usuários informaram algum tipo de falha lógica no *software* descoberta durante a realização da avaliação ou/e algum tipo de inconsistência na descrição semântica de algumas imagens;
- os usuários do sistema aprenderão a utilizá-lo rapidamente de acordo com a maior parte dos usuários;
- o sistema não é atrapalhado de utilizar de acordo com a maior parte dos usuários. Sobre os usuários que tiveram alguma crítica sobre o sistema ser atrapalhado de usar, alguns informaram que o sistema apresentou algum tipo de inconsistência;
- os usuários se sentiram confiantes na utilização do sistema. A confiança é importante, pois diminui a resistência na utilização do sistema e mostra que os usuários compreenderam como devem utilizar o sistema;
- os usuários não precisariam de nenhum conhecimento extra para utilizar o sistema, apenas as informações fornecidas de acordo com a maior parte dos usuários. Alguns usuários tiveram alguns problemas durante a utilização do sistema, aparentemente por não terem lido ou assimilado as informações passadas nas instruções, por causa desse tipo de problema, seria importante passar também as instruções por meio de vídeo.

Na avaliação de usabilidade o sistema ficou com pontuação média de 80,06 e margem de erro de 4,41 para mais ou para menos utilizando nível de confiança de 95%. A escala utilizada é de 0 até 100, onde 68 já é considerado um nível de usabilidade aceitável. A avaliação de usabilidade mostra que o sistema proposto tem boa usabilidade de acordo com a avaliação dos usuários. A pontuação média foi obtida por meio das avaliações individuais calculadas utilizando a fórmula da escala de usabilidade de sistema.

A Tabela 6.2 mostra os resultados obtidos por meio do questionário com a margem de erro utilizando nível de confiança de 95%, onde foi avaliada a dificuldade de cria-

ção da senha semântica e a dificuldade nas autenticações utilizando de 2 até 9 imagens simultaneamente.

Pergunta	**Média**	**Resultado**
Qual foi o grau de dificuldade na criação de sua senha semântica.	3.84375 ± 0,31	Fácil
Qual foi o grau de dificuldade na memorização de sua senha semântica.	3.5625 ± 0,24	Moderado-Fácil
Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando duas imagens foram exibidas simultaneamente.	4.09375 ± 0,23	Fácil
Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando três imagens foram exibidas simultaneamente.	4 ± 0,21	Fácil
Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando quatro imagens foram exibidas simultaneamente.	3,59375 ± 0,23	Moderado-Fácil
Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando seis imagens foram exibidas simultaneamente.	3,25 ± 0,27	Moderado
Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando nove imagens foram exibidas simultaneamente.	2,875 ± 0,28	Moderado
Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica de forma geral.	3.6875 ± 0,22	Moderado-Fácil

Tabela 6.2: Perguntas sobre a dificuldade de utilização do sistema.

Os resultados obtidos na pesquisa e mostrados na Tabela 6.2 demonstram que segundo a avaliação dos participantes:

- é fácil criar a senha semântica;
- não é difícil memorizar a senha semântica;
- é fácil identificar as relações existentes nas imagens durante a autenticação com 2 ou 3 imagens sendo exibidas simultaneamente, porém na medida que a quantidade de imagens aumenta, o grau de dificuldade na identificação das relações também aumenta, chegando ao nível moderado quando 9 imagens são exibidas simultaneamente;
- o grau de dificuldade na utilização do sistema não é difícil.

A Figura 6.2 mostra o tempo médio gasto na autenticação semântica gráfica para 2, 3, 4, 6 e 9 imagens exibidas simultaneamente quando o usuário utiliza uma senha semântica pré-definida e quando utiliza sua própria senha semântica.

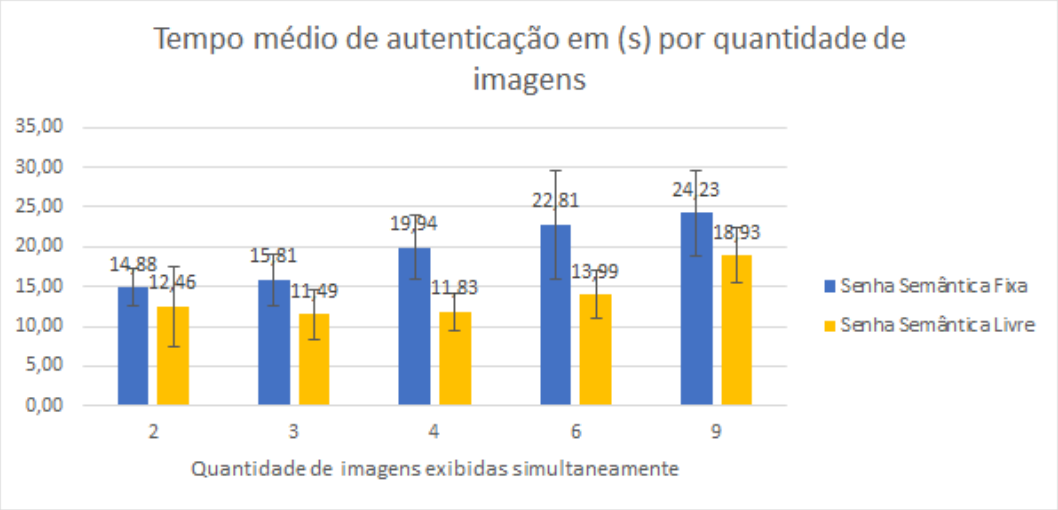

Figura 6.2: Tempo médio gasto na autenticação em segundos por quantidade de imagens exibidas simultaneamente e margem de erro utilizando intervalo de confiança de 95%.

No gráfico mostrado na Figura 6.2 é possível observar que à medida que a quantidade de imagens exibidas simultaneamente aumenta, o tempo médio gasto pelo usuário no processo de autenticação tende a aumentar. O gráfico também mostra que o usuário gasta em média menos tempo durante o processo de autenticação quando utiliza uma senha semântica própria

A Figura 6.3 mostra o percentual médio de acertos para 2, 3, 4, 6 e 9 imagens exibidas simultaneamente quando o usuário utiliza uma senha semântica pré-definida e quando utiliza sua própria senha semântica.

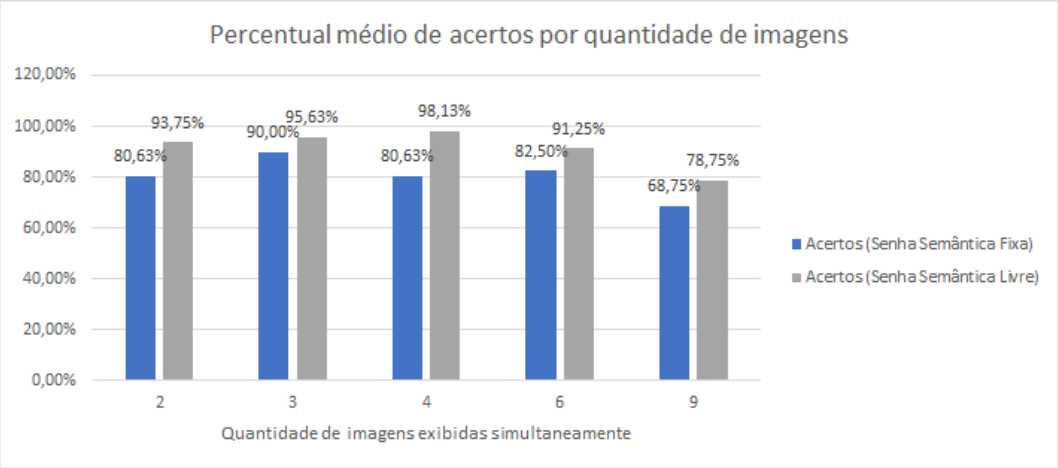

Figura 6.3: Percentual de acertos por quantidade de imagens exibidas simultaneamente.

No gráfico mostrado na Figura 6.3 é possível observar que o percentual de autenticações com sucesso foi maior quando os usuários utilizaram uma senha semântica própria. O gráfico também mostra que os piores resultados foram obtidos quando 9 imagens foram exibidas simultaneamente. No questionário alguns usuários informaram que na autenticação com 9 imagens é difícil identificar as relações da senha semântica devido as imagens serem exibidas em um tamanho bem menor do que nas autenticações com menos imagens.

No cenário onde teoricamente seria mais fácil para o atacante descobrir a senha semântica do usuário, nenhum dos participantes da avaliação conseguiu descobrir a senha semântica após observar 5 autenticações distintas com a mesma senha semântica. Essa avaliação foi realizada com intuito de verificar a resistência do sistema de autenticação proposto ao ataque *shoulder surfing*. Nesta avaliação de resistência o sistema foi avaliado em sua condição mais desfavorável para o sistema proposto e mais favorável para o atacante, na autenticação onde apenas duas imagens são exibidas simultaneamente. O resultado dessa avaliação mostra que a autenticação semântica gráfica possui boa resistência ao ataque *shoulder surfing*.

6.2 Considerações Finais

Os resultados mostrados neste capítulo demonstram que o sistema de autenticação proposto é viável, pois possui boa usabilidade, resistência ao ataque *shoulder surfing*, a maioria dos usuários conseguiram aprender rapidamente como utilizar o sistema, sem terem passado por qualquer tipo de capacitação prévia, o sistema pode disponibilizar uma quantidade grande de alternativas durante o processo de autenticação, dificultando a autenticação bem sucedida de forma aleatória e também pode prover uma quantidade enorme de combinações possíveis de senhas semânticas.

O próximo capítulo apresenta a conclusão e quais trabalhos podem ser desenvolvidos como forma de aperfeiçoar este trabalho.

Capítulo 7

Conclusão e Trabalhos Futuros

Este capítulo contempla a fase de conclusão do *Design Science Research* e lista possíveis trabalhos futuros.

7.1 Conclusões

Este trabalho teve como objetivo principal propor um sistema de autenticação complementar para o método de autenticação usuário e senha por meio de representações semânticas, com o uso da lógica descritiva, de conceitos identificáveis em imagens como forma de aumentar a segurança no processo de autenticação.

Na fase de levantamento bibliográfico de artigos na área de autenticação identificou-se uma lacuna em relação a usabilidade das soluções atuais que podem ser utilizadas como um segundo fator no processo de autenticação para o método de autenticação usuário e senha (fator **saber** - conhecer a senha ou respostas a perguntas). Ao invés de usar como segundo fator algo que se **tenha** ou seja, usar uma variação do saber, no caso a habilidade humana de **saber** identificar a existência ou não de relações semânticas binárias entre elementos de uma imagem.

Este trabalho demonstrou que a quantidade de relações semânticas binárias possíveis em imagens é suficientemente grande para tornar o sistema robusto a ataques que escolham imagens de forma aleatória. O modelo analítico mostra que quando a quantidade de imagens exibidas simultaneamente aumenta, a quantidade de combinações exibidas durante o processo de autenticação também aumenta e o sistema de autenticação se torna mais robusto.

A avaliação do sistema de autenticação proposto com o grupo de usuários mostrou que o sistema possui boa usabilidade, obtendo 80 pontos, em uma escala de 0 até 100. Para a usabilidade do sistema ser considerada boa, bastaria atingir 68 pontos na escala de

usabilidade de sistema. Portanto, baseando-se no resultado da pesquisa, é possível afirmar que o sistema de autenticação proposto não possui problema relacionado a usabilidade.

O resultado da pesquisa relacionado à senha semântica mostrou que a maioria dos participantes consideraram fácil criá-la, não consideraram difícil memorizá-la e nem identificá-la durante o processo de autenticação. Com este resultado é possível concluir que esse sistema de autenticação poderia ser utilizado sem criar muitos problemas para a maioria dos usuários. De forma geral, a pesquisa mostrou que não é difícil utilizar o sistema de autenticação proposto.

No início do trabalho, havia a preocupação do tempo gasto pelos usuários durante o processo de autenticação, pois se os usuários demorassem muito tempo no processo de autenticação, a utilização do sistema proposto seria inviabilizada. Após a realização da avaliação do sistema proposto com o grupo de usuários, foi identificado que o tempo médio gasto no processo de autenticação no cenário onde os usuários receberam uma senha semântica definida foi de 14,88 segundos na autenticação com duas imagens sendo exibidas simultaneamente e 24,23 segundos na autenticação com nove imagens sendo exibidas simultaneamente. No cenário onde os usuários criaram a sua própria senha semântica, o tempo médio consumido no processo de autenticação foi de 12,46 segundos na autenticação com duas imagens sendo exibidas simultaneamente e 18,93 segundos na autenticação com nove imagens sendo exibidas simultaneamente. Esse resultado mostra que os usuários precisaram de menos tempo para autenticar no sistema proposto quando utilizaram sua própria senha semântica. O resultado também mostra que os usuários levaram um tempo médio aceitável durante o processo de autenticação, mostrando a viabilidade na utilização do sistema no quesito tempo.

A avaliação mostrou que no cenário onde os usuários receberam a senha semântica definida, o percentual médio de sucesso nas autenticações foi de 80% até 90% quando foram exibidas para os usuários durante o processo de autenticação de 2 até 6 imagens simultaneamente e o percentual médio de sucesso foi de 68,75% quando foram exibidas 9 imagens simultaneamente. No cenário onde os usuários criaram a sua própria senha semântica, o percentual médio de sucesso nas autenticações foi de 91% até 98% quando foram exibidas para os usuários durante o processo de autenticação de 2 até 6 imagens simultaneamente e o percentual médio de sucesso foi de 78% quando foram exibidas 9 imagens simultaneamente. Esse resultado mostrou que os usuários tiveram dificuldade na identificação das relações semânticas nas imagens na autenticação com 9 imagens sendo exibidas simultaneamente. Na questão discursiva da avaliação, vários usuários informaram que na autenticação com 9 imagens tiveram dificuldade, pois as imagens foram exibidas em um tamanho menor para caber na tela. Esse resultado de forma geral mostra que os usuários obtiveram um percentual de sucesso maior durante o processo de autenticação

quando autenticaram utilizando uma senha semântica definida por eles mesmos.

A resistência do sistema proposto ao ataque *shoulder surfing* foi testada na avaliação realizada com o grupo de usuários em um cenário onde os participantes observaram 5 autenticações distintas com a mesma senha semântica, duas imagens foram exibidas simultaneamente durante o processo de autenticação e nenhum dos participantes conseguiu descobrir a senha semântica utilizada durante as suas 5 tentativas. Com esse resultado é possível concluir que o sistema de autenticação proposto possui boa resistência ao ataque *shoulder surfing*. É importante ressaltar que foi avaliada a resistência do sistema proposto ao ataque *shoulder surfing* no pior cenário para o sistema, onde a autenticação foi realizada com a exibição de apenas duas imagens simultaneamente.

Foi desenvolvida uma API para que o sistema de autenticação proposto possa ser integrado com outros sistemas. Foi realizada prova de conceito para testar a integração do WordPress com o sistema proposto. Por meio da API desenvolvida foi possível autenticar no WordPress utilizando a autenticação semântica gráfica em complemento a autenticação por meio do método usuário e senha. É importante ressaltar que devido o pouco tempo disponível para a realização dos testes com a API, seria importante que fosse realizada uma avaliação mais profunda e possivelmente a realização de melhorias.

Baseando-se nos resultados obtidos neste trabalho, é possível afirmar que a autenticação semântica gráfica pode aumentar a segurança no processo de autenticação de sistemas que utilizam o método de autenticação usuário e senha, e possui vantagem em relação a usabilidade quando comparado com outros sistemas de autenticação utilizados como segundo fator.

O sistema tem como desvantagem a necessidade de uma quantidade grande de imagens em sua base para diminuir a possibilidade de repetições das imagens durante uma mesma autenticação e em autenticações diferentes para um mesmo usuário, pois se o sistema exibir as mesmas imagens para um dado usuário em autenticações distintas, facilmente um atacante conseguirá autenticar no sistema, apenas memorizando a sequência de imagens, sem a necessidade de ter que saber qual é a senha semântica.

O processo de seleção de imagens foi demorado e o processo de descrição está suscetível ao erro se o vocabulário não for bem definido na ontologia criada. O processo de descrição das imagens também é demorado e deve ser realizado com cautela, para evitar erros, pois um erro na descrição pode induzir os usuários ao erro durante o processo de autenticação. É importante que as imagens selecionadas tenham uma quantidade considerável de relações semânticas, para dificultar que a senha semântica seja descoberta por meio de observações durante o processo de autenticação.

Na ontologia utilizada no sistema de autenticação proposto, é possível utilizar algumas centenas de relações distintas na descrição das imagens e essas mesmas relações estão

disponíveis durante a criação da senha semântica. A quantidade de combinações possíveis pode ser muito grande tornando o sistema robusto a ataques de força bruta.

O sistema de autenticação proposto foi apresentado na Conferência Ibérica de Sistemas y Tecnologias de Información (CISTI) no dia 26/06/2020.

Foi iniciado processo de registro de patente deste sistema de autenticação junto ao INPI (Instituto Nacional da Propriedade Industrial).

O *software* desenvolvido será registrado e será disponibilizado após a sua refatoração.

Com os resultados obtidos neste trabalho é possível responder à pergunta de pesquisa afirmando que é possível melhorar a segurança do método de autenticação usuário e senha com um sistema de autenticação complementar que utiliza relações semânticas existentes em imagens.

7.2 Trabalhos Futuros

É possível identificar várias direções em que este trabalho pode ser relacionado.

Integração ao SEI - este trabalho inicialmente tinha como um dos objetivos específicos a integração do Sistema Eletrônico de Informações (SEI) com o sistema proposto via API, porém devido o SEI utilizar um *framework* desenvolvido no TRF-4 e haver pouca documentação disponível sobre esse *framework* na Internet, avaliamos que o tempo necessário para aprender sobre o funcionamento do *framework* e posteriormente integrá-lo com o sistema de autenticação proposto seria maior que o disponível para tal atividade, por este motivo o sistema foi integrado com o WordPress. Como trabalho futuro, o SEI poderia ser integrado com o sistema de autenticação proposto, visando aumentar a sua segurança, tendo em vista que atualmente o SEI autentica seus usuários somente por meio do usuário e senha

Semântica em textos - em sistemas com pouca largura de banda, ao invés de imagens poder-se-ia utilizar diversos textos curtos. Pode ser a descrição da própria imagem ou coisas mais abstratas para a identificação da existência ou não das relações semânticas.

Adequação para dispositivos móveis – este trabalho avaliou a utilização do sistema de autenticação proposto apenas em estações de trabalho, não era o objetivo deste trabalho avaliar a sua utilização em dispositivos móveis devido o tempo disponível. Um trabalho futuro poderia trabalhar na responsividade da solução para dispositivos móveis ou até mesmo desenvolver uma interface própria para dispositivos móveis.

Avaliação de acessibilidade e a implementação de melhorias – este trabalho não avaliou a acessibilidade da solução para deficientes visuais devido a limitação de tempo e também não estava previsto no escopo. Um trabalho futuro poderia avaliar a

acessibilidade da solução para deficientes visuais e propor como o sistema de autenticação proposto poderia ser aperfeiçoado para atender esse público.

Referências

[1] Belk, M., C. Fidas, P. Germanakos e G. Samaras: *The interplay between humans, technology and user authentication: A cognitive processing perspective*. Computers in Human Behavior, 76:184 – 200, 2017, ISSN 0747-5632. http://www.sciencedirect.com/science/article/pii/S0747563217304120. 1, 13

[2] Mercl, L., V. Sobeslav, P. Mikulecky e M. Macinka: *Infrastructure authentication, authorization and accounting solutions for an openstack platform*. Em *Mobile Web and Intelligent Information Systems*, páginas 123–135, Cham, 2019. Springer International Publishing, ISBN 978-3-030-27191-6. 1, 12, 13

[3] Bonneau, J., C. Herley, P. C. van Oorschot e F. Stajano: *Passwords and the evolution of imperfect authentication*. Commun. ACM, 58(7):78–87, junho 2015, ISSN 0001-0782. http://doi.acm.org/10.1145/2699390. 1, 2, 13, 14

[4] Bouder, H. Le, G. Thomas, E. Bourget, M. Graa, N. Cuppens e J. Lanet: *Theoretical Security Evaluation of the Human Semantic Authentication Protocol*. páginas 332–339, 2018. 2, 3, 22, 24

[5] Luo, W., Y. Hu, H. Jiang e J. Wang: *Authentication by encrypted negative password*. IEEE Transactions on Information Forensics and Security, 14(1):114–128, Jan 2019, ISSN 1556-6013. 2

[6] Ometov, A., S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen e Y. Koucheryavy: *Multi-Factor Authentication: A Survey*. Cryptography, 2(1):1, 2018, ISSN 2410-387X. http://www.mdpi.com/2410-387X/2/1/1. 2, 18, 19

[7] Almulhem, A.: *A graphical password authentication system*. 2011 World Congress on Internet Security (WorldCIS-2011), (January):223–225, 2011. 2, 3, 21, 40, 41

[8] Colnago, J., S. Devlin, M. Oates, C. Swoopes, L. Bauer, L. Cranor e N. Christin: *" It's not actually that horrible " : Exploring Adoption of Two-Factor Authentication at a University*. Proc. of CHI, páginas 1–12, 2018. 2, 19, 20

[9] Dresch, A., D. Lacerda e J. Júnior: *Design Science Research: A Method for Science and Technology Advancement*. setembro 2014, ISBN 978-3-319-07373-6. 4, 5

[10] Kremer, S., L. Mé, D. Rémy e V. Roca: *Cybersecurity : Current challenges and Inria's research directions*. janeiro 2019. 7, 8

[11] Junior, R. M. e C. Canongia: *Segurança cibernética: o desafio da nova sociedade da informação*. Parcerias Estratégicas, 14(29):21–46, 2009, ISSN 2176-9729. http://seer.cgee.org.br/index.php/parcerias_estrategicas/article/view/349/0. 7, 8

[12] Ishikawa, E., E. W. Vianna, J. M. da Silva, J. H. C. Fernandes, P. R. L. Gondim e R. Zelenovsky: *Modeling a Cyber Defense Business Ecosystem of Ecosystems: Nurturing Brazilian Cyber Defense Resources*. IGI Global, 2021. http://doi:10.4018/978-1-7998-5728-0.ch021. 8

[13] Kizza, J. M.: *Guide to Computer Network Security*. Springer Publishing Company, Incorporated, 2nd edição, 2013, ISBN 1447145429. 8, 9

[14] Harris, S. e F. Maymi: *CISSP All-in-One Exam Guide, Eighth Edition*. McGraw-Hill Education, New York, 8ª edição, 2018, ISBN 1260142655. https://mhebooklibrary.com/doi/book/10.1036/9781260142648. 9

[15] Barker, W.: *Guideline for identifying an information system as a national security system*. NIST Special Publication 800-59, 2003. https://doi.org/10.6028/NIST.SP.800-59. 9

[16] Nieles, M., K. Dempsey e V. Pillitteri: *An Introduction to Information Security*. NIST Special Publication 800-12, 2017. https://doi.org/10.6028/NIST.SP.800-12r1. 9, 10, 11

[17] Andress, J.: *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*. Syngress Publishing, 1st edição, 2011, ISBN 9781597496537, 9781597496544. 9, 10, 12

[18] University, C. M.: *Guidelines for data classification*. https://www.cmu.edu/iso/governance/guidelines/data-classification.html, 2018. 10

[19] Bowen, P., J. Hash e M. Wilson: *Information security handbook: A guide for managers*. NIST Special Publication 800-100, 2006. https://doi.org/10.6028/NIST.SP.800-100. 10

[20] SG-PR: *Política nacional de segurança da informação*. http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2018/Decreto/D9637.htm, 2018. 10

[21] GSI-PR: *Gestão de segurança da informação e comunicações na administração pública federal*. http://dsic.planalto.gov.br/legislacao/in_01_gsidsic.pdf, 2018. 10, 11

[22] Aminzade, M.: *Confidentiality, integrity and availability – finding a balanced it framework*. Network Security, 2018(5):9 – 11, 2018, ISSN 1353-4858. http://www.sciencedirect.com/science/article/pii/S1353485818300436. 10

[23] LeBlanc, J. e T. Messerschmidt: *Identity and Data Security for Web Development: Best Practices*. O'Reilly Media, Inc., 1st edição, 2016, ISBN 1491937017, 9781491937013. 14

[24] Foltýn, T.: *The most popular passwords of 2018 revealed: Are yours on the list?* https://www.welivesecurity.com/2018/12/17/most-popular-passwords-2018-revealed/, 2018. 14

[25] Lim, H. W. e G. Yang: *Authenticated key exchange protocols for parallel network file systems.* IEEE Transactions on Parallel and Distributed Systems, 27(1):92–105, Jan 2016, ISSN 1045-9219. 14

[26] Aumasson, J.: *Serious Cryptography: A Practical Introduction to Modern Encryption.* No Starch Press, San Francisco, CA, USA, 2017, ISBN 1593278268, 9781593278267. 14, 15

[27] Zhou, K. e J. Ren: *Passbio: Privacy-preserving user-centric biometric authentication.* IEEE Transactions on Information Forensics and Security, 13(12):3050–3063, Dec 2018. 17

[28] Wójtowicz, A. e K. Joachimiak: *Model for adaptable context-based biometric authentication for mobile devices.* Personal Ubiquitous Computing, 20(2):195 – 207, 2016, ISSN 16174909. https://doi.org/10.1007/s00779-016-0905-0. 17

[29] Abe, N. e T. Shinzaki: *A survey on newer prospective biometric authentication modalities.* 2014. 17, 18

[30] Lorant, G., J. Wary, P. Salembier, Z. Moustafa e C. Mathias: *Evaluation ergonomique d'un système d'authentification graphique*, fevereiro 2016. 21, 22, 23, 41

[31] Salembier, P., M. Zouinar, R. Héron, C. Mathias, G. Lorant e J. Wary: *Experimental studies of a graphical authentication system based on semantic categorisation.* Em *Actes De La 28IèMe Conference Francophone Sur L'Interaction Homme-Machine*, IHM '16, páginas 134–143, New York, NY, USA, 2016. ACM, ISBN 978-1-4503-4243-8. http://doi.acm.org/10.1145/3004107.3004121. 23, 24, 41

[32] Ogiela, M. R., N. Krzyworzeka e L. Ogiela: *Application of knowledge-based cognitive captcha in cloud of things security.* Concurrency and Computation: Practice and Experience, 30(21):e4769, 2018. https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.4769, e4769 CPE-18-0166.R1. 25

[33] Xu, X., L. Liu e B. Li: *A survey of captcha technologies to distinguish between human and computer.* Neurocomputing, 408:292 – 307, 2020, ISSN 0925-2312. http://www.sciencedirect.com/science/article/pii/S0925231220304896. 25, 26

[34] Aldwairi, M., S. Mohammed e M. L. Padmanabhan: *Efficient and secure flash-based gaming captcha.* Journal of Parallel and Distributed Computing, 142:27 – 35, 2020, ISSN 0743-7315. http://www.sciencedirect.com/science/article/pii/S0743731519308858. 25

[35] Chen, J., X. Luo, Y. Guo, Y. Zhang e D. Gong: *A survey on breaking technique of text-based captcha.* Security and Communication Networks, 2017:6898617, Dec 2017, ISSN 1939-0114. https://doi.org/10.1155/2017/6898617. 25

[36] Brodić, D. e A. Amelio: *Analysis of the human-computer interaction on the example of image-based captcha by association rule mining*, 2016. 25

[37] Hasan, W. K. A.: *A survey of current research on captcha.* International Journal of Computer Science Engineering Survey, 7(3):1–21, 2016. https://app.dimensions.ai/details/publication/pub.1072616257andhttps://doi.org/10.5121/ijcses.2016.7301. 26

[38] Gao, H., F. Cao e P. Zhang: *Annulus: A novel image-based captcha scheme.* Em *2016 IEEE Region 10 Conference (TENCON)*, páginas 464–467, 2016. doi.org/10.1109/TENCON.2016.7848042. 26

[39] Kumar, S. K. e J. A. Harding: *Description logic–based knowledge merging for concrete- and fuzzy-domain ontologies.* Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 230(5):954–971, 2016. https://doi.org/10.1177/0954405414564404. 29, 31

[40] Baader, F., I. Horrocks e U. Sattler: *Description Logics.* Em van Harmelen, Frank, Vladimir Lifschitz e Bruce Porter (editores): *Handbook of Knowledge Representation*, capítulo 3, páginas 135–180. Elsevier, 2008. download/2007/BaHS07a.pdf. 29, 32, 33

[41] Lutz, C. e F. Wolter: *The Data Complexity of Description Logic Ontologies.* Logical Methods in Computer Science, Volume 13, Issue 4, novembro 2017. https://lmcs.episciences.org/4060. 30

[42] Baader, F., D. Calvanese, D. Mcguinness, D. Nardi e P. Patel-Schneider: *The Description Logic Handbook: Theory, Implementation, and Applications.* janeiro 2007. 30

[43] Krötzsch, M., F. Simančík e I. Horrocks: *Description Logics.* IEEE Intelligent Systems, 29(1):12–19, 2014. download/2014/KrSH14.pdf. 30, 31

[44] Rudolph, S.: *Foundations of description logics.* Volume 6848, páginas 76–136, janeiro 2011. 33, 34, 35

[45] Hellmann, S., J. Lehmann e S. Auer: *Learning of owl class descriptions on very large knowledge bases.* Int. J. Semantic Web Inf. Syst., 5:25–48, abril 2009. 33, 38

[46] Marzano, G.: *Using resource description framework (rdf) for description and modeling place identity.* Procedia Computer Science, 77:135 – 140, 2015, ISSN 1877-0509. http://www.sciencedirect.com/science/article/pii/S1877050915038806, ICTE in regional Development 2015 Valmiera, Latvia. 35, 36

[47] Ma, Z., M. A. M. Capretz e L. Yan: *Storing massive resource description framework (rdf) data: a survey.* The Knowledge Engineering Review, 31(4):391–413, 2016. 36

[48] Yu, L.: *A Developer's Guide to the Semantic Web.* janeiro 2011, ISBN 978-3-642-15969-5. 36

[49] Alqahtani, S. S., E. E. Eghan e J. Rilling: *Tracing known security vulnerabilities in software repositories – a semantic web enabled modeling approach.* Science of Computer Programming, 121:153 – 175, 2016, ISSN 0167-6423. http://www.sciencedirect.com/science/article/pii/S0167642316000253, Special Issue on Knowledge-based Software Engineering. 36

[50] Wu, S., Y. Zhang e W. Cao: *Network security assessment using a semantic reasoning and graph based approach.* Computers Electrical Engineering, 64:96 – 109, 2017, ISSN 0045-7906. http://www.sciencedirect.com/science/article/pii/S0045790617302409. 37

[51] Kurilovas, E. e A. Juskeviciene: *Creation of web 2.0 tools ontology to improve learning.* Computers in Human Behavior, 51:1380 – 1386, 2015, ISSN 0747-5632. http://www.sciencedirect.com/science/article/pii/S0747563214005494, Computing for Human Learning, Behaviour and Collaboration in the Social and Mobile Networks Era. 37

[52] Kendall, E. F. e D. L. Mcguinness: *Ontology engineering.* Synthesis Lectures on the Semantic Web: Theory and Technology, 9:i–102, abril 2019. 37

[53] Man, D.: *Ontologies in computer science.* DIDACTICA MATHEMATICA, 31:43–46, 2013. 37

[54] Pandey, M. e R. Pandey: *Provenance linking using bundles in owl ontology.* International Journal of Computer Applications, 164(11):5–9, Apr 2017, ISSN 0975-8887. http://www.ijcaonline.org/archives/volume164/number11/27525-2017913721. 37

[55] Euzenat, J. e P. Shvaiko: *Ontology Matching.* Springer Publishing Company, Incorporated, 2nd edição, 2013, ISBN 3642387209. 38

[56] Keet, C. M.: *An Introduction to Ontology Engineering.* College Publications, 2018, ISBN 9781848902954. https://books.google.com.br/books?id=RUqzvgEACAAJ. 38, 39

[57] Antoniou, G., P. Groth, F. Harmelen van van e F. V. Hoekstra: *A Semantic Web Primer.* The MIT Press, 2012, ISBN 0262018284. 38

[58] Krishna, R., Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein e L. Fei-Fei: *Visual genome: Connecting language and vision using crowdsourced dense image annotations.* International Journal of Computer Vision, 123(1):32–73, May 2017, ISSN 1573-1405. https://doi.org/10.1007/s11263-016-0981-7. 41, 42, 43, 48

[59] Krause, J., J. Johnson, R. Krishna e F. Li: *A hierarchical approach for generating descriptive image paragraphs.* páginas 3337–3345, julho 2017. 41

[60] Hata, K., R. Krishna, L. Fei-Fei e M. S. Bernstein: *A glimpse far into the future: Understanding long-term crowd worker quality.* Em *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW

'17, páginas 889–901, New York, NY, USA, 2017. ACM, ISBN 978-1-4503-4335-0. http://doi.acm.org/10.1145/2998181.2998248. 41

[61] Sun, X., Y. Zi, T. Ren, J. Tang e G. Wu: *Hierarchical visual relationship detection.* Em *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, páginas 94–102, New York, NY, USA, 2019. ACM, ISBN 978-1-4503-6889-6. http://doi.acm.org/10.1145/3343031.3350921. 41

[62] Anderson, P., B. Fernando, M. Johnson e S. Gould: *Spice: Semantic propositional image caption evaluation.* Em Leibe, B., J. Matas, N. Sebe e M. Welling (editores): *Computer Vision – ECCV 2016*, páginas 382–398, Cham, 2016. Springer International Publishing, ISBN 978-3-319-46454-1. 41

[63] Fukui, A., D. H. Park, D. Yang, A. Rohrbach, T. Darrell e M. Rohrbach: *Multimodal compact bilinear pooling for visual question answering and visual grounding.* Em *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, páginas 457–468, Austin, Texas, novembro 2016. Association for Computational Linguistics. https://www.aclweb.org/anthology/D16-1044. 41

[64] Johnson, J., A. Karpathy, L. Fei-Fei e D. Zekrifa: *Densecap: Fully convolutional localization networks for dense captioning.* páginas 4565–4574, junho 2016. 41

[65] Zhu, Y., O. Groth, M. S. Bernstein e L. Fei-Fei: *Visual7w: Grounded question answering in images.* Em *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, páginas 4995–5004, 2016. https://doi.org/10.1109/CVPR.2016.540. 41

[66] Bowman, S. R., G. Angeli, C. Potts e C. D. Manning: *A large annotated corpus for learning natural language inference.* Em *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP).* Association for Computational Linguistics, 2015. 41

[67] Schuster, S., R. Krishna, A. X. Chang, L. Fei-Fei e C. D. Manning: *Generating semantically precise scene graphs from textual descriptions for improved image retrieval.* Em *Proceedings of the Fourth Workshop on Vision and Language, VL@EMNLP 2015, Lisbon, Portugal, September 18, 2015*, páginas 70–80, 2015. https://doi.org/10.18653/v1/W15-2812. 41

[68] Gao, M., P. Kortum e F. L. Oswald: *Multi-language toolkit for the system usability scale.* International Journal of Human–Computer Interaction, 36(20):1883–1901, 2020. https://doi.org/10.1080/10447318.2020.1801173. 56

[69] Wu, H. e S. Leung: *Can likert scales be treated as interval scales?—a simulation study.* Journal of Social Service Research, 43(4):527–532, 2017. https://doi.org/10.1080/01488376.2017.1329775. 57

Apêndice A

Apêndice

A.1 Concepção do Experimento de Avaliação

Esta subseção descreve de forma detalhada a avaliação do sistema de autenticação proposto realizada com um grupo de 32 usuários.

O objetivo desta avaliação foi avaliar a usabilidade de forma geral, a capacidade do usuário memorizar relações binárias semânticas como senha, a capacidade do usuário identificar estas mesmas relações em imagens e a resistência contra o ataque *shoulder surfing* do sistema de autenticação proposto.

Considerando que a avaliação deste sistema de autenticação envolve a participação de seres humanos, sendo um experimento que:

- avalia a capacidade dos participantes em memorizar relações semânticas binárias;
- avalia a capacidade cognitiva dos participantes em identificar relações binárias semânticas em imagens;
- mede a usabilidade do sistema; e
- verifica a capacidade dos participantes em tentar adivinhar relações semânticas binárias por meio de associações de imagens (forma de avaliar a resistência do sistema proposto ao *shoulder surfing*)

chegou-se a conclusão de que este experimento deveria ser analisado pelo Comitê de Ética em Pesquisa em Ciências Humanas e Sociais (CEP/CHS) da UnB.

A Resolução Nº 510, de 7 de abril de 2016 do Conselho Nacional de Saúde, subordinado ao Ministério da Saúde, parte do princípio que:

- "a ética em pesquisa implica o respeito pela dignidade humana e a proteção devida aos participantes das pesquisas científicas envolvendo seres humanos";

- "que o agir ético do pesquisador demanda ação consciente e livre do participante";
- "que a pesquisa em ciências humanas e sociais exige respeito e garantia do pleno exercício dos direitos dos participantes, devendo ser concebida, avaliada e realizada de modo a prever e evitar possíveis danos aos participantes";
- "que as Ciências Humanas e Sociais têm especificidades nas suas concepções e práticas de pesquisa, na medida em que nelas prevalece uma acepção pluralista de ciência da qual decorre a adoção de múltiplas perspectivas teórico-metodológicas, bem como lidam com atribuições de significado, práticas e representações, sem intervenção direta no corpo humano, com natureza e grau de risco específico"; e
- "que a produção científica deve implicar benefícios atuais ou potenciais para o ser humano, para a comunidade na qual está inserido e para a sociedade, possibilitando a promoção de qualidade digna de vida a partir do respeito aos direitos civis, sociais, culturais e a um meio ambiente ecologicamente equilibrado".

Levando em conta estes princípios, a citada resolução normatiza as pesquisas em Ciências Humanas e Sociais que em sua metodologia utilize dados diretamente obtidos com os participantes ou de informações identificáveis ou que possam acarretar riscos maiores do que os existentes na vida cotidiana.

Do exposto, este experimento foi elaborado seguindo toda a legislação vigente e de forma que o mesmo não acarretasse riscos maiores do que os existentes na vida cotidiana. Desta forma, tomou-se o cuidado de utilizar relações semânticas binárias existentes no cotidiano das pessoas, evitando-se relações semânticas binárias que envolvessem situações anormais ou que evocassem sentimentos indesejados, como de violência, repulsa etc. Da mesma forma, o mesmo cuidado foi tomado na seleção das imagens utilizados nos experimentos. Este cuidado, também deve ser tomado na utilização do sistema de autenticação, caso este seja adotado em algum sistema de informação.

Cabe ressaltar que a Computação seja, talvez, a mais humana das ciências exatas. Que os sistemas computacionais normalmente interagem com seres humanos, e que nesta interação, muitas vezes, emergem efeitos colaterais imaginados e inimaginados. No primeiro caso poder-se-ia ter, por exemplo, o uso de mensagens subliminares pelo uso de uma ontologia não neutra em termos de conciência para as relações binárias. No segundo caso, mais comuns em sistemas complexos ou sistemas de sistemas, temos o que chamamos de comportamentos emergentes, que podem ser positivos ou negativos. Este fenômeno pode ser constatado, por exemplo, nas redes sociais, que de forma não prevista incrementou a proliferação das fakenews, da segmentação da sociedade em guetos radicais, da invasão da privacidade e, até mesmo, no resultados das eleições. Obviamente, ocorreram comportamentos emergentes positivos nas redes sociais, como o surgmento de novos negócios

e formas de trabalho. Por isso, talvez, a necessidade das pesquisas e produtos de computação se preocuparem com estes aspectos que influem no comportamento social, não só com os aspectos de segurança, *shoulder surfing* no caso do presente trabalho, mas na modificação do cotidiano dos seres humanos pela disponibilização e uso de programas de computadores aparentemente inofensivos.

O processo de solicitação de autorização para avaliar o sistema proposto com usuários foi realizado da seguinte forma:

- foi realizado o cadastro como pesquisador na Plataforma Brasil do Ministério da Saúde;
- o projeto de pesquisa foi cadastrado na Plataforma Brasil e foram fornecidas todas as informações solicitadas pela plataforma;
- foram anexados na plataforma os seguintes documentos exigidos pelo Comitê de Ética em Pesquisa em Ciências Humanas e Sociais da Universidade de Brasília:
  - carta de encaminhamento: nesta carta os pesquisadores fornecem algumas informações sobre os pesquisadores, a pesquisa, o público alvo, a data de início da avaliação. Neste documento os pesquisadores se comprometem a iniciar a pesquisa somente após a aprovação da comissão;
  - folha de rosto: é um documento obtido na Plataforma Brasil onde o pesquisador deve preencher e assinar se comprometendo a seguir os requisitos da resolução CNS 466/12 e o responsável pelo programa de graduação ou pós-graduação também deve assinar autorizando a pesquisa em nome da instituição;
  - instrumento de coleta de dados: descreve de forma detalhada a pesquisa/avaliação está sendo solicitada a autorização;
  - aceite institucional: documento onde o responsável pela instituição que será alvo da pesquisa autoriza a pesquisa. Este formulário é necessário somente se a pesquisa for realizada em alguma empresa ou instituição. Nesta pesquisa não foi necessária a apresentação deste formulário, porque a pesquisa não foi direcionada para uma empresa ou instituição específica, essa pesquisa foi aberta para qualquer participante na Internet;
  - currículo lattes: currículo dos pesquisadores na plataforma lattes do CNPQ em formato PDF;
  - carta de revisão de ética: nesta carta é necessário informa o período de início e término da pesquisa, o tipo de pesquisa (quantitativa, qualitativa, etc), o público alvo da pesquisa, o objetivo primário da pesquisa, o objetivo secundário da pesquisa, a descrição da avaliação, como será realizada a coleta de dados,

os benefícios pretendidos com a pesquisa, os riscos que os participantes podem estar expostos durante a avaliação e como fazer para mitigá-los;

- termo de consentimento livre e esclarecido (TCLE): esse documento informa os participantes da pesquisa sobre como será realizada, informa os contatos dos pesquisadores e do comitê de ética para tirar dúvidas. O participante deve assinar o termo concordando com o que está escrito e após essa assinatura, o pesquisador tem a autorização para iniciar a avaliação com o participante;
- cronograma: no cronograma os pesquisadores devem informar de forma detalhada o cronograma de execução da avaliação.

No portal do Comitê de Ética em Pesquisa em Ciências Humanas e Sociais da Universidade de Brasília é possível obter modelos da maioria dos documentos citados no parágrafo anterior.

Foi criado termo de consentimento que está de acordo com modelo utilizado e aprovado pelo conselho de ética do CEP/CHS da UnB por meio do processo 39802820.0.0000.5540 da Plataforma Brasil.

No pré-teste da avaliação, identificamos algumas inconsistências na descrição de algumas imagens que foram corrigidas antes da avaliação final e identificamos que realizar 10 autenticações em cada um dos cenários onde 2, 3, 4, 6 e 9 imagens são exibidas simultaneamente levava de 15 até 20 minutos, porém nessa pesquisa tínhamos o intuito de realizar a mesma avaliação com os usuários utilizando uma senha semântica já definida e uma senha semântica criada por ele mesmo, visando avaliar o desempenho nos dois cenários, por este motivo tivemos que diminuir a quantidade de repetições de 10 para 5 nas autenticações de 2 até 9 imagens, objetivando viabilizar a avaliação nos dois cenários de senha semântica e para manter o tempo de realização da avaliação não muito longo, em torno de 15 até 20 minutos conforme medido no pré-teste. Era importante que o tempo da avaliação não fosse superior a 20 minutos, para evitar que os participantes desistissem da avaliação antes do término, prejudicando a pesquisa.

Inicialmente o usuário deve cadastrar-se na solução de autenticação e fornecer algumas informações de acordo com as figuras a seguir.

O participante acessa a página inicial de acordo com a Figura A.1 e clicar no link para realizar o cadastro.

Antes de carregar a tela de cadastro, o participante deve aceitar o Termo de Consentimento caso queira avançar para a tela de cadastro. A Figura A.2 mostra como é exibido o Termo de Consentimento para o participante.

Figura A.1: Tela inicial da avaliação

Após o participante concordar com o termo de consentimento, a tela de cadastro será exibida. O participante informar obrigatoriamente e-mail e cadastra uma senha, informações como nome e idade são opcionais. A Figura A.3 mostra a tela de cadastro.

Após o participante realizar o cadastro, o sistema direciona para a página de autenticação de acordo com a Figura A.4. Nesta página o usuário deve informar o e-mail e a senha cadastrada para autenticar no sistema.

Após o participante inserir o e-mail e senha cadastrada, o sistema direciona para página com instruções sobre a avaliação de acordo com a Figura A.5.

Após o participante ler as instruções, o usuário deve iniciar a primeira fase da avaliação que é a autenticação no sistema com uma senha pré-definida com três relações semânticas constituídas de <sujeito, predicado e objeto>, por exemplo, <homem, usando, meia>, <mulher, vestindo, calça> e <criança, segurando, bola>. Em seguida o participante verá na tela do seu computador duas imagens, uma com a primeira relação semântica e a outra sem a primeira relação semântica. Ele deve escolher a imagem que julga ter a primeira relação semântica. Em seguida são apresentadas na tela do computador mais duas imagens, uma com a segunda relação semântica e outra sem a segunda relação semântica. O participante da pesquisa deve escolher a imagem que julga ter a segunda relação semântica. Em seguida são apresentadas na tela do computador mais duas imagens, uma com a terceira relação semântica e outra sem a terceira relação semântica. O participante da pesquisa deve escolher a imagem que julga ter a terceira relação semântica. São armazenados o tempo que o participante do experimento levou para realizar a autenticação semântica, as imagens que escolheu e se ele acertou a senha semântica correlacionando corretamente as três relações semânticas com as respectivas imagens. Este processo é repetido por 5 vezes.

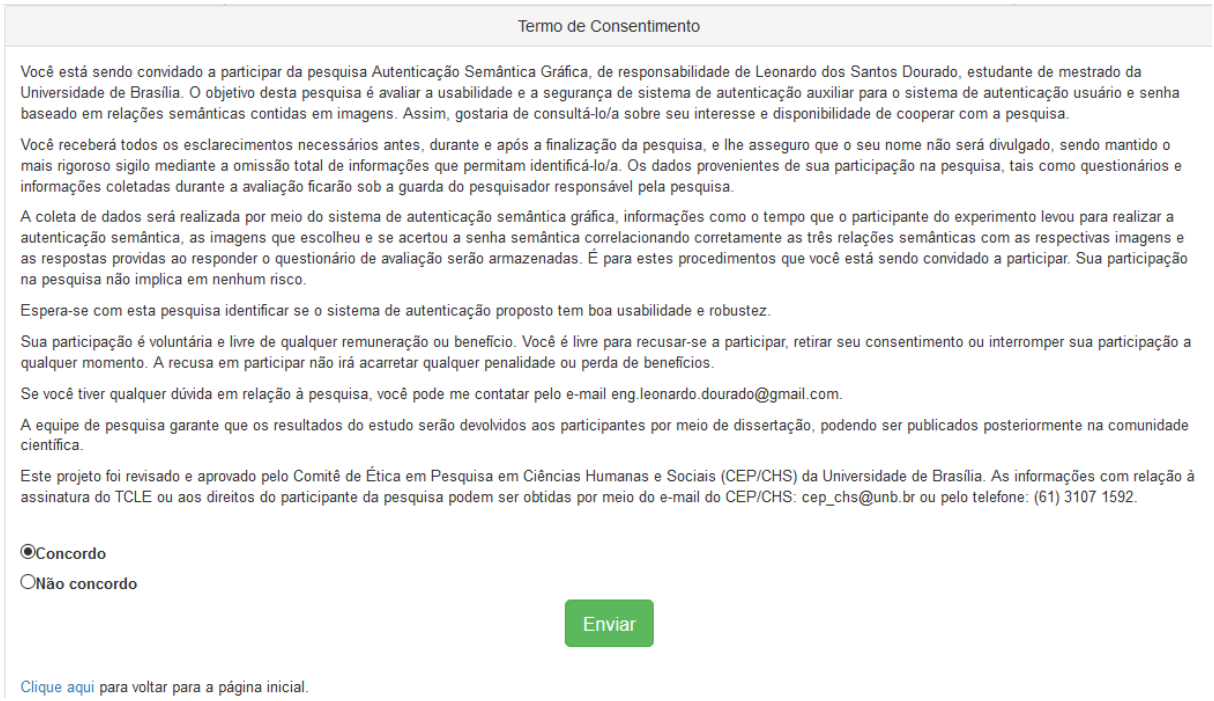

Figura A.2: Tela do Termo de Consentimento

A Figura A.6 mostra o processo de validação de um dos conceitos da senha semântica do participante na autenticação com duas imagens.

Em seguida o mesmo processo é repetido exibindo ao invés de duas imagens, três, quatro, seis e nove imagens, para avaliar a usabilidade do sistema com três, quatro, seis e nove imagens, respectivamente uma vez que com mais imagens o sistema é mais seguro.

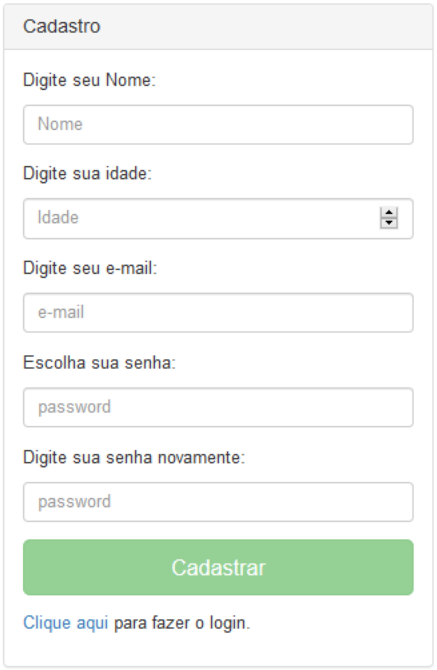

Figura A.3: Tela de cadastro

As Figuras A.7, A.8, A.9 e A.10 mostram o processo de validação de um dos conceitos da senha semântica do participante na autenticação com três, quatro, seis e nove imagens respectivamente.

Após o participante finalizar as 5 autenticações com nove imagens terá finalizado a primeira etapa da pesquisa e será direcionado para tela onde deverá criar sua própria senha semântica, essa senha é utilizada na segunda etapa da pesquisa. A Figura A.11 mostra a tela de criação da senha semântica.

A segunda etapa da pesquisa é realizada da mesma forma que a primeira etapa, exceto que nesta etapa o participante deve autenticar com a sua própria senha semântica. São armazenados o tempo que o participante do experimento levou para realizar a autenticação semântica, as imagens que escolheu e se ele acertou a senha semântica correlacionando corretamente as três relações semânticas com as respectivas imagens. O participante autentica 5 vezes com duas, três, quatro, seis e nove imagens.

Após finalizar a segunda etapa da pesquisa, o participante é direcionado para a terceira etapa, o questionário de avaliação do sistema.

São realizadas perguntas com intuito de avaliar a dificuldade na utilização da autenticação semântica gráfica conforme aumenta a quantidade de imagens, a dificuldade na criação da senha semântica, a dificuldade na utilização do sistema e perguntas sobre a

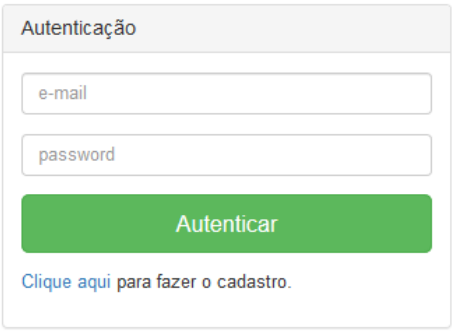

Figura A.4: Tela de autenticação

usabilidade do sistema.

Seguem abaixo as perguntas realizadas para os participantes na avaliação do sistema:

- 1. Qual foi o grau de dificuldade na criação de sua senha semântica?;
- 2. Qual foi o grau de dificuldade na memorização de sua senha semântica?;
- 3. Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando duas imagens foram exibidas simultaneamente?
- 4. Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando três imagens foram exibidas simultaneamente?
- 5. Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando quatro imagens foram exibidas simultaneamente?
- 6. Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando seis imagens foram exibidas simultaneamente?
- 7. Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica quando nove imagens foram exibidas simultaneamente?
- 8. Qual foi o grau de dificuldade na identificação dos relacionamentos de sua senha semântica de forma geral?
- 9. Neste campo podem ser realizados comentários, sugestões e críticas caso tenha algum(a)?

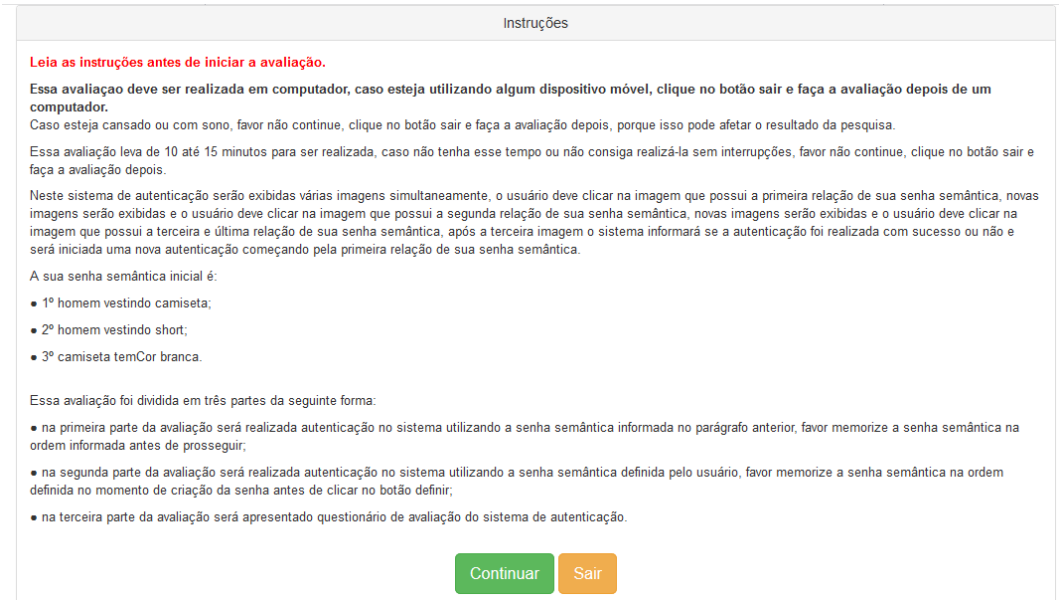

Figura A.5: Tela com instruções

Figura A.6: Autenticação semântica com duas imagens

- 10. Poderia usar esse sistema com frequência.
- 11. Acho o sistema muito complexo.
- 12. Achei o sistema fácil de usar.
- 13. Acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema.
- 14. Acho que as várias funções do sistema estão muito bem integradas.
- 15. Acho que o sistema apresenta muita inconsistência.
- 16. Imagino que as pessoas aprenderão como usar esse sistema rapidamente.
- 17. Achei o sistema atrapalhado de usar.

Figura A.7: Autenticação semântica com três imagens

Figura A.8: Autenticação semântica com quatro imagens

- 18. Me senti confiante ao usar o sistema.
- 19. Precisei aprender várias coisas novas antes de conseguir usar o sistema.

Os usuários têm como opção de resposta cinco alternativas com pesos variando de 1 até 5 de acordo com a escala de Likert.

As perguntas de 10 até 19 são perguntas da escala de usabilidade de sistema *(System Usability Scale - SUS)*. Esse sistema de avaliação de usabilidade é bastante utilizado e caso o peso final seja igual o superior a 68, a usabilidade do sistema avaliado é considerada boa.

Após responder as perguntas do questionário, o usuário envia o questionário clicando no botão responder e finaliza a terceira etapa.

Figura A.9: Autenticação semântica com seis imagens

Após responder o questionário, o participante é direcionado para tela que pergunta se o usuário deseja participar da última etapa de acordo com a Figura A.12, esta etapa é opcional. Caso o usuário não queria participar, basta clicar na opção não e sai do sistema, caso o usuário queira prosseguir para a última etapa, basta clicar no botão "sim" de acordo com tela a seguir.

Na quarta e última etapa da pesquisa, o participante tenta descobrir as três relações semântica observando duas imagens por 4 segundo, uma imagem é realçada com uma moldura amarela por 1 segundo indicando que é a imagem que possui a primeira relação semântica, em seguida são apresentadas mais duas imagens por 4 segundo, uma imagem é realçada com uma moldura amarela por 1 segundo indicando que é a imagem que possui a segunda relação semântica, por fim são apresentadas mais duas imagens por 4 segundo, uma imagem é realçada com uma moldura amarela por 1 segundo indicando que é a imagem que possui a terceira relação semântica. A Figura A.13 mostra como a imagem correta será exibida para o participante.

Após o participante visualizar as 3 imagens que contém as relações da senha semântica, é direcionado para tela onde visualizará menu com as opções possíveis para a primeira, segunda e terceira relações semânticas que ele acha que podem ser as corretas. A Figura A.14 mostra tela exibida para o participante.

Após o participante enviar a senha semântica que acha que é a correta, o sistema informa se acertou ou não.

O participante da pesquisa repetirá este processo de tentativa de descoberta da senha semântica por cinco vezes. O sistema gravará a senha semântica, as imagens exibidas em

Figura A.10: Autenticação semântica com nove imagens

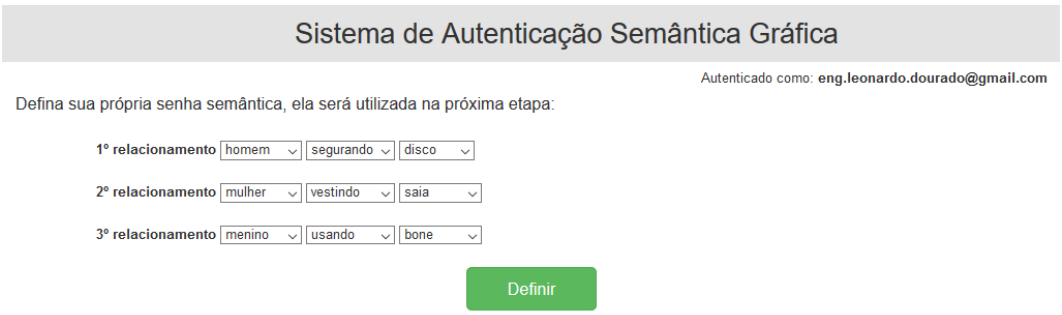

Figura A.11: Tela de definição da senha semântica

cada tentativa e a senha semântica que ele acha que estava sendo usada.

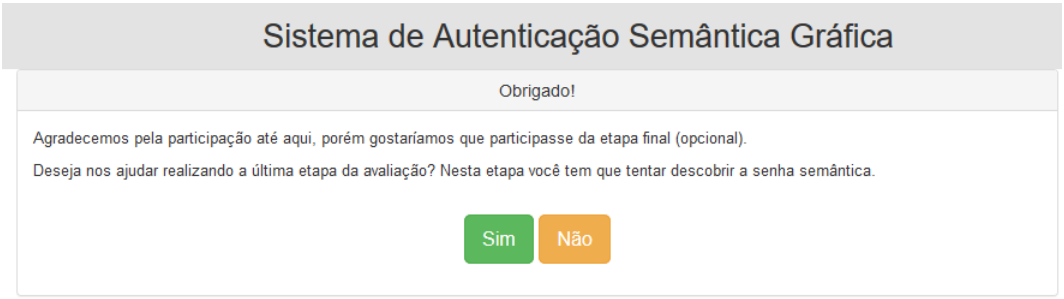

Figura A.12: Tela de encerramento ou de direcionamento para última etapa

Figura A.13: Tela da quarta etapa da pesquisa

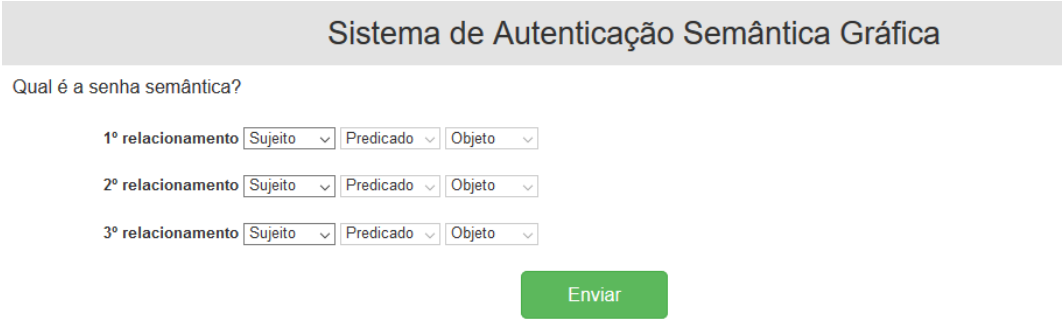

Figura A.14: Tela onde o usuário seleciona a possível senha semântica