



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DO TOCANTINS
CAMPUS PORTO NACIONAL
CURSO SUPERIOR EM LICENCIATURA EM COMPUTAÇÃO**

WESLEY SOUSA GALVÃO

**SEGURANÇA DA INFORMAÇÃO: análise e recolhimento de dados públicos
através da internet**

**PORTO NACIONAL
2022**

WESLEY SOUSA GALVÃO

**SEGURANÇA DA INFORMAÇÃO: análise e recolhimento de dados públicos
através da internet**

Trabalho de Conclusão de Curso apresentado à
Coordenação do Curso de Licenciatura em
Computação, da Unidade de Porto Nacional, do
Instituto Federal do Tocantins, como exigência à
obtenção do título de Licenciado em Computação

Orientador: Prof. Dr. Heleno Ayres Manduca

**PORTO NACIONAL
2022**

Dados Internacionais de Catalogação na Publicação (CIP)
Bibliotecas do Instituto Federal do Tocantins

G182s Galvão, Wesley Sousa
Segurança da Informação: Análise e Recolhimento de dados
públicos através da internet / Wesley Sousa Galvão. – Porto
Nacional, TO, 2022.
39 p. : il. color.

Trabalho de Conclusão de Curso (Licenciatura em Computação)
– Instituto Federal de Educação, Ciência e Tecnologia do Tocantins,
Campus Porto Nacional, Porto Nacional, TO, 2022.

Orientador: Me. Heleno Ayres Manduca

1. Osint. 2. Dados Públicos. 3. Segurança da Informação. I.
Manduca, Heleno Ayres. II. Título.

CDD 004

A reprodução total ou parcial, de qualquer forma ou por qualquer meio, deste documento é autorizada para fins
de estudo e pesquisa, desde que citada a fonte.

**Elaborado pelo sistema de geração automática de ficha catalográfica do IFTO com os dados fornecidos
pelo(a) autor(a).**

WESLEY SOUSA GALVÃO

**SEGURANÇA DA INFORMAÇÃO: análise e recolhimento de dados públicos
através da internet**

Trabalho de Conclusão de Curso apresentado à
Coordenação do Curso de Licenciatura em
Computação, da Unidade de Porto Nacional, do
Instituto Federal do Tocantins, como exigência à
obtenção do título de Licenciado em Computação

Aprovado em: 24/11/2022.

BANCA AVALIADORA

Prof. Dr. Heleno Ayres Manduca
Orientador
IFTO – Campus Porto Nacional

<<Titulação>> << Nome>>
<<Nome da instituição onde atua>>

<<Titulação>> << Nome>>
<<Nome da instituição onde atua>>

RESUMO

O Brasil possui cerca de 134 milhões de usuários conectados na internet. De acordo com a pesquisa da TIC Domicílios 2019, de cada 4 brasileiros, 3 acessam a internet. Sendo assim, o terceiro país que mais faz uso das redes sociais, o que por si só, gera uma grande quantidade de dados, muitos disponibilizados publicamente devido à falta de preocupação dos usuários com configurações de privacidade, vazamento de dados ou falhas das aplicações. Esses dados públicos podem ser acessados por ferramentas, as quais na prática evita que tenhamos que fazer tarefas repetitivas, automatizando parte do trabalho, assim como plataformas específicas que ajudam e facilitam todo o processo, frisando partes importantes em vez de optar por buscas com conteúdo menos relevantes e com algum valor específico. Logo, tem-se como objetivo geral do trabalho a análise e recolhimento de dados públicos através da internet, com foco nas ferramentas de Osint que além de ferramenta, é uma maneira de recolher informações pessoais de fontes abertas, tais como: google, revistas, redes sociais. Abordando vários assuntos descritos de forma explicativa. Este estudo, também, apresenta como metodologia a pesquisa exploratória, bibliográfica e com investigação das ferramentas, tendo como intuito explicar as diversas formas que os dados pessoais e outros tipos de dados são obtidos e apresentá-los, assim como podem ser extraídos da internet de forma livre e lícita, tanto com a utilização de ferramenta para a busca de dados quanto com técnicas e o uso da internet para fazer buscas avançadas.

Palavras-chaves: Osint. Dados públicos. Segurança da informação.

ABSTRACT

O Brasil has about 134 million users connected to the Internet. According to the TIC Domicílios 2019 survey, out of every 4 Brazilians, 3 access the Internet. Being likewise, the third country that makes the most use of social networks, or that by itself, generates a large amount of data, many made publicly available due to the lack of concern of two users with privacy settings, data leaks or application failures. These public data can be accessed by tools, as a practical matter it prevents us from having to do repetitive tasks, automating part of the work, as well as specific platforms that help us facilitate the whole process, avoiding important parts instead of opting for searches with less content relevant and with some specific value. Logo, the general objective of the work is the analysis and collection of public data through the Internet, with a focus on Osint tools that, in addition to tools, is a way to collect information from public sources, such as: Google, magazines, social networks Addressing several issues described in an explanatory way. This study also presents as a methodology exploratory, bibliographical research and with the investigation of tools, intended to explain the various ways that personal data and other types of data are obtained and presented, as well as how they can be extracted from the internet. free and lawful way, both with the use of tools to search for data and with techniques and the use of the Internet to do advanced searches.

Keywords: Osint. Public data. Information security.

LISTA DE ILUSTRAÇÕES

Figura 1 – Meios em que a Osint pode ser aplicada e canais de extração de informações.....	22
Figura 2 – Processo Osint	23
Figura 3 – Osint framework	25
Figura 4 – Fases da Osint	30
Figura 5 – Dados vazado de determinado e-mail	32
Figura 6 – Sistemas que possuem em sua base de dados o nome de usuário “jhondoe”	33
Figura 7 - Endereços de e-mail encontrados vinculados ao domínio uft.edu.br	35
Figura 8 - Resultado de conferência de e-mail	35
Figura 9 – Alguns subdomínios encontrados para o site uft.edu.br	36
Figura 10 – Localização de usuário do Twitter com a ferramenta Creepy.....	37
Figura 11 – Resultado da ferramenta Recon Tool.....	38
Figura 12 – Geração de “usuário” pela ferramenta justdelete.me	39
Figura 13 – Verificação das contas em que um usuário está logado	39
Figura 14 – Informações abertas em portal brasileiro	40
Figura 15 – Grafo de transações em bitcoin.....	41

LISTA DE TABELAS

Tabela 1 – Possíveis falhas que podem ser pesquisadas através do Google	24
Tabela 2 – Dados vazados de determinado e-mail	32

SUMÁRIO

1. INTRODUÇÃO	10
2. METODOLOGIA	11
2.1 PROCESSO PARA COLETA DE DADOS	12
3. REFERENCIAL TEÓRICO.....	14
4. LEI DE PROTEÇÃO DE DADOS.....	15
4.1 DADOS PESSOAIS	17
4.2 DADOS SENSÍVEIS.....	17
4.3 DADOS PÚBLICOS	18
4.4 DADOS ANONIMIZADOS	18
4.5 BIG DATA.....	18
4.6 FONTES DE INTELIGÊNCIA.....	19
4.7 FONTES ABERTAS DE INTELIGÊNCIA	20
4.8 OSINT	22
5. COLETA DE DADOS ATRAVÉS DO OSINT	24
5.1 GOOGLE HACKING	24
5.2 OSINT FRAMEWORK.....	25
5.3 MALTEGO.....	29
6. UTILIZAÇÃO DOS DADOS COLETADOS.....	30
6.1 BUSCA POR DADOS VAZADOS	31
6.2 BUSCA POR NOMES DE USUÁRIO.....	33
6.3 BUSCA POR EMAILS	34
6.4 BUSCA POR NOMES DE DOMÍNIO	36
6.5 BUSCA EM REDES SOCIAIS.....	36
6.6 CRIAÇÃO DE IDENTIDADES FALSAS	38
6.7 PRIVACIDADE	39
6.8 REGISTROS PÚBLICOS	40
6.9 MOEDAS DIGITAIS	40
7. CONCLUSÃO E TRABALHOS FUTUROS	42
REFERÊNCIAS	43
GLOSSÁRIO.....	46

1. INTRODUÇÃO

De acordo com Valente (2020), o Brasil possui cerca de 134 milhões de usuários na Internet, ou seja, 3 em cada 4 pessoas utilizam serviços online. Desses, 90% afirmam estar online todos os dias, e 7% ao menos uma vez no mês. Como recursos mais utilizados a fonte menciona os comunicadores como WhatsApp, Skype ou Facebook Messenger (92%), redes sociais como Facebook ou Snapchat (76%) e chamadas de vídeo por Skype ou WhatsApp (73%).

Se tratando das redes sociais, a Revista Seguro Total (2021) sintetizou um estudo realizado pela plataforma CupomValido, que reuniu dados da *Hootsuite* e *WeAreSocial*, na qual o Brasil ocupa a 3ª posição a nível mundial em quantidade de acesso às redes, com uma média de 6 horas e 20 minutos conectados diariamente.

Tais dados são importantes, pois demonstram uma mudança no comportamento das pessoas, que passaram a dar preferência a vários serviços online como o pagamento de contas, reserva de hotéis, conversa com os amigos, compras, entre outros.

Por outro lado, essa presença online acaba tornando público muitos dados dos usuários, principalmente dos que não utilizam as configurações das redes sociais para esconder algumas informações pessoais, como: nome da mãe, data de nascimento, gostos pessoais, endereço, local de trabalho, entre outros.

Muitos desses dados públicos podem ser utilizados para realização de fraudes, sendo que somente em 2020 os crimes reportados envolvendo fraudes financeiras através de *phishing* somaram 85,75%, conforme o Cert.br (2020). Esse tipo de crime é conhecido pela tentativa de enganar um usuário, ao se passar por uma página/portal conhecido, como no caso de bancos, ou então familiares pedindo dinheiro. Sendo que muitas das informações referentes ao parentesco podem ser obtidos através de dados públicos disponíveis na Internet.

Uma forma de obter essas informações, é através do *framework osint* (*Open source intelligence* - Inteligência de Código Aberto), que segundo Dokman e Ivanjko (2019) é uma forma de inteligência de se coletar dados públicos de diversas fontes, e após processados permitem a tomada de decisão. Osint é um conjunto de atividades para coletar, armazenar e analisar as informações de fontes públicas — basicamente, qualquer dado sobre uma empresa ou pessoa que possa ser

encontrado por meio de ferramentas da Internet ou *framework osint*, como os buscadores.

Wallner (1993) afirma que cerca de 80% do conhecimento que provêm da inteligência vêm de dados públicos, e podem ser utilizados: por empresas, para monitoramento de funcionários; por pessoas, para evitarem que existam informações pessoais demais online e que possam dar brechas para criminosos; e, também, por criminosos, entre outros.

Portanto, o objetivo geral deste estudo é a análise e o recolhimento de dados públicos através da internet e de ferramentas e plataformas. Para isso, tem-se os seguintes objetivos específicos:

- Entender o conceito de fontes abertas de inteligência;
- Analisar a ferramenta Osint; e
- Explanar como os dados podem ser coletados através da ferramenta Osint.

2. METODOLOGIA

Para a análise qualitativa, traz-se explicações para a exposição de dados e configurações de privacidade, de forma a entender os motivos pelo qual pessoas, empresas e hackers buscam por esses dados, que são expostos de forma pública em toda a internet. Já na pesquisa quantitativa, busca-se por resultados objetivos e palpáveis, baseados em experimentos e testes realizados antes e após a coleta de informações.

Os resultados das pesquisas partem das percepções de profissionais em cyber segurança e empresas de vendas online, o famoso *e-commerce*, observando as indiferenças em campo, os conflitos e os aspectos subjetivos identificados.

Este estudo configura-se como estudo de caso, uma vez que se trata de uma pesquisa que busca analisar as diversas formas que existem para recolher dados pessoais, de maneira aprofundada e completa.

Desse modo, a presente pesquisa tem por base entender o objeto de maneira completa, interpretando o contexto em que se insere os métodos. Assim, as fontes utilizadas foram diversas com foco principal em falhas de segurança de dados, além de pesquisa documental.

A metodologia que também foi desenvolvida no trabalho é a bibliográfica e exploratória, conforme Silva e Menezes (2005). Assim, a pesquisa exploratória envolve várias atividades, entre elas a pesquisa bibliográfica em livros e artigos da área, bem como a investigação da ferramenta mencionada e utilização.

Este estudo, realizado através de pesquisa bibliográfica tem como finalidade: apresentar conceitos sobre fontes abertas e fontes privadas; distinguir as diferenças delas; analisar possíveis meios em que as informações podem ser descobertas e coletadas através de pesquisas na internet, sejam elas pesquisas simples ou pesquisas avançadas.

Quanto aos levantamentos de dados serão utilizadas pesquisas investigativas que visam trazer dados reais publicados de fontes confiáveis, além de apresentar ferramentas que auxiliam nas buscas por informações, que, por muitas vezes, podem estar exposta em páginas indexadas na internet não necessitando de autenticação.

A pesquisa exploratória tem como principal objetivo a interação com tema proposto no trabalho aprofundando de forma específica nos assuntos principais citados.

Quanto ao objetivo de apresentar técnicas de recolhimento de dados tanto na internet como com ferramentas de uso simples, as quais podem ser instaladas nos computadores. Assim, apresentou-se dados pessoais de usuários da Universidade Federal do Tocantins (UFT), através da busca nos dados públicos e a utilização da ferramenta framework osint, maltego e google hacking.

2.1 PROCESSO PARA COLETA DE DADOS

A coleta de dados é um processo que visa reunir o máximo de informações possíveis para uso posterior através de determinadas técnicas. Os dados serão utilizados na pesquisa, estudo e nas investigações.

Primeiramente, analisou-se as ferramentas citadas de forma objetiva com o intuito de demonstrar como é feito o processo. Simplificando o máximo todo o conteúdo, cada ferramenta, aplicação ou “site” terá uma forma de buscar dados pessoais públicos e privados, dependendo da segurança em que se encontra no

momento, conforme a busca vai se tornando mais extensa e mais dados serão obtidos, com isso iremos ter uma vasta quantidade de informações.

3. REFERENCIAL TEÓRICO

O referencial teórico contempla os tópicos em que essa pesquisa está baseada, como a big data e as fontes de inteligência, google hacking e demais ferramentas, para então tratarmos das fontes abertas de inteligência, com foco no Osint.

Assim como o Google Dork que é também conhecido como google hacking, com ele através de técnicas mais avançadas conseguimos filtrar informações importantes e bem mais relevantes para o tipo de pesquisa ou informação que estamos procurando na internet. É válido ressaltar que a web pode ser rastreada por qualquer pessoa. O google como um *Search Engine* ou motor de busca como é conhecido por alguns, indexa sites através de robôs conhecidos como googlebots, que conseguem buscar milhões de páginas na internet e ficam atualizando constantemente as buscas realizadas pelos usuários.

As demais ferramentas servem como auxílio principal para buscas mais específicas, filtrando o máximo de informações relevante bem antes da pré-análise ser feita, a fim de reduzir todo o trabalho, minimizando, assim, o tempo que poderia ser perdido nessa etapa.

4. LEI DE PROTEÇÃO DE DADOS

A Lei Geral de Proteção de Dados (LGPD), nº 13.709/2018, tem como principal objetivo proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

O objetivo de suspender o uso e a comercialização indevida de dados pessoais, inclusive em meios digitais. O intuito imediato deve buscar proteger os direitos e a privacidade de todas as pessoas.

Existem as regras sobre o uso de dados pessoais em todas as transações online. A lei proíbe o uso indiscriminado de dados pessoais. Para estar de acordo com a legislação todas as empresas terão que definir a base legal para uso dos dados e deverão informar qual a finalidade específica que serão utilizados.

Além disso a Lei Geral de Proteção de Dados prevê sanções de 2% sobre o faturamento ou até R\$ 50 milhões. Vale frisar que esse valor é calculado por infração. Portanto, se for cometido mais de uma infração a multa é dobrada. No plano da LGPD, a legislação não diferencia entre dados e informações. Qualquer tipo de arquivo, ao conter dados pessoais, deve seguir as regras da lei, independentemente de serem dados brutos ou já terem passado por algum tipo de análise também chamado de pró-análise ou pós-análise.

No Art. 6º, a LGPD determina 10 princípios que devem nortear o tratamento de dados pessoais. Estes princípios é que irão ajudar a garantir que a empresa esteja em conformidade e adequação à lei.

Esses princípios são: Finalidade, Adequação, Necessidade, Livre acesso, Qualidade dos dados, Transparência, Segurança, Prevenção, Não discriminação, Responsabilização e Prestação de Contas.

O primeiro é a finalidade, a LGPD faz com que seja uma obrigação as empresas terem propósitos bem elaborados de como irão tratar esses dados. Além disso, têm que deixar escrito com clareza quais são as intenções para o proprietário dos dados, justificando efetivamente e mostrando a finalidade do uso de tais dados. O segundo é a origem da adequação, de acordo com a LGPD, refere-se à “uniformidade do tratamento com os objetivos informados ao titular, de acordo com o cenário do tratamento”.

A necessidade também é importante no ponto de vista da LGPD, pois leva em consideração a responsabilidade das empresas acerca dos dados tratados.

Em aplicação, quanto maior for a quantidade de dados pessoais a ser tratada maior será a responsabilidade. Dessa forma, como consequência maior será a cobrança e mais elevado o valor das multas, em caso de erros, falhas e vazamentos.

O princípio do livre acesso é um dos pontos fundamentais da LGPD. De acordo com a lei, o livre acesso é a “garantia, aos titulares, de consulta rápida e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais”.

A qualidade dos dados deve ser de acordo com a garantia de exatidão. Devem ser claros ter relevância e atualização dos dados de acordo com a necessidade, o comprimento e finalidade do tratamento.

Simplificando, deve se respeitar as normas da LGPD e garantir que a base de dados seja verdadeira e esteja sempre atualizada além de estar alinhada com o propósito principal.

O princípio da transparência, de acordo com a lei, é a “garantia, aos titulares, de informações claras, precisas e de fácil acesso sobre a execução do tratamento”. Transparência é o que se deve informar aos proprietários dos dados sobre os respectivos agentes de processamento, que são basicamente outras empresas envolvidas no processo de processamento de dados.

A segurança envolve a adoção de procedimentos, tecnologias e soluções que garantem maior proteção dos dados pessoais. A LGPD regulamenta que as empresas façam o “uso de medidas técnicas e administrativas capazes de proteger os dados pessoais de acesso não autorizado e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou divulgação”

Sempre é melhor prevenir do que remediar. O princípio da prevenção determina a “adoção de medidas para prevenir a ocorrência de danos decorrentes do tratamento de dados pessoais”. A LGPD se compromete a estar preparada para lidar com possíveis problemas envolvendo o Processamento de Dados Pessoais, antes mesmo que eles surjam.

O tratamento de dados pessoais nunca poderá ser realizado com o objetivo de discriminar ou promover abusos contra os seus titulares. O princípio da não discriminação, segundo a LGPD, refere-se à “impossibilidade de realizar o tratamento para fins discriminatórios ilícitos ou abusivos”

O princípio da responsabilização e prestação de contas, dispõe sobre o cumprimento da lei, tendo em vista provas e evidências de quais medidas e procedimentos foram tomados pela empresa a fim de garantir a proteção dos dados.

Segundo o ministério da cidadania os dados pessoais que são abordados pela LGPD são os seguintes: CPF, RG, número de telefone. Devido a Lei n.º 13.709/2018, a proteção de dados passou a ser um dever dos cidadãos, da administração pública e das empresas que utilizam esses dados.

4.1 DADOS PESSOAIS

Dado pessoal é aquele que identifica a pessoa direta ou indiretamente, conforme a quantidade que é extraído para análise. São exemplos de dados pessoais: nome e sobrenome, data e local de nascimento, RG, CPF, retrato em fotografia, endereço residencial, endereço de e-mail, número de cartão bancário, renda, histórico de pagamentos, hábitos de consumo, dados de localização, como dados de localização no celular, endereço de IP (protocolo de internet), testemunhos de conexão (cookies), número de telefone.

4.2 DADOS SENSÍVEIS

Além dos dados pessoais, existem aqueles que devem ter um maior nível de segurança no tratamento, são os tipos de dados relacionados a crianças e adolescentes. Classificam-se, também, como dados sensíveis, os que revelam origem racial ou étnica, convicções religiosas ou filosóficas, opiniões políticas, filiação sindical, questões genéticas, biométricas e sobre a saúde ou a vida sexual de uma pessoa.

Se o dado for de indivíduos menores de idade, deve haver consentimento específico obtido por alguns dos pais ou responsável legal, limitando ao conteúdo estritamente necessário, sem repassar tais dados a terceiros. Os dados pessoais coletados de menores sem consentimento só podem ser obtidos apenas para contatar aos pais ou responsável, podendo ser utilizados uma única vez sem armazenar nos bancos de dados.

4.3 DADOS PÚBLICOS

O tratamento de dados pessoais públicos deve ser explícita e demonstrar qual a finalidade específica, que justifica a disponibilização. A LGPD define que uma organização pode, tratar dados tornados públicos pelo(a) titular em momento anterior e de forma evidente. Porém, se a organização porventura quiser compartilhar esses dados com outras empresas, deverá novamente pedir o consentimento para tal fim.

É importante destacar que a LGPD também se relaciona com a Lei de Acesso à Informação (LAI), Lei nº 12.527/11, e com princípios constitucionais, a exemplo do inciso XXXIII, do artigo 5º da Constituição Federal, de 1988, em que: “todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado”.

4.4 DADOS ANONIMIZADOS

A anonimização é uma técnica de processamento de dados que remove ou modifica informações que possam identificar a pessoa. Nesses casos, a LGPD não se aplicará aos dados. É importante salientar que os dados só são considerados anonimizados se não permitirem, por meios técnicos ou outros, reconstruir o caminho para revelar quem é o titular dos dados.

4.5 BIG DATA

Em uma sociedade cercada pelo uso da tecnologia para os mais diversos fins, como - interação com os amigos, estudo, pesquisas em geral, compras entre outros -, é gerado uma vasta quantidade de dados conhecida como *big data*, conforme Nascimento (2016).

Casalinho (2015) então define *big data* como um grande repositório de dados, caracterizado por:

- Valor: os dados adquiridos devem ter valor e fazer sentido dentro da organização. Assim, uma empresa pode capturar, por exemplo, opiniões de consumidores de marcas concorrentes de forma a

entender como os clientes estão interagindo com os concorrentes, ou então dados internos de produção, vendas, entre outros;

- Veracidade: os dados devem ser reais e confiáveis, de forma que caso não sejam, podem ser feita uma análise com base em informações falsas, comprometendo o resultado e uma posterior tomada de decisão da empresa;
- Variedade: deve existir uma grande variedade de dados, como: áudio, vídeo, imagens, mensagens de texto, logs de sistemas, entre outros. Um exemplo de variedade são os dados contidos em um aparelho celular ou então em redes sociais.
- Velocidade: apesar da quantidade de dados dentro do repositório, a extração e filtragem devem ser rápidas; e
- Volume: deve haver um grande volume de dados.

Assim, esses grandes repositórios de dados vão possibilitar que sejam feitas as mais diversas análises envolvendo mineração de dados e inteligência artificial. É com base nesses dados, os quais são públicos, que as ferramentas de inteligência de fontes abertas atuam.

4.6 FONTES DE INTELIGÊNCIA

A Doutrina Nacional de Segurança Pública (Brasil, 2009), estabelece que existem dois principais tipos de obtenção de dados: as abertas e as protegidas. As abertas são, como o nome já sugere, as fontes de dados de livre acesso, enquanto as protegidas são aquelas que exigem credenciais específicas para acesso, por exemplo, não podendo ser acessadas por qualquer pessoa.

Ainda acerca da obtenção dos dados, esses podem ser obtidos por meio humano ou eletrônico, conforme Brasil (2009). O humano tem como foco a obtenção da inteligência e conhecimento do homem, enquanto o meio eletrônico ainda se divide em: inteligência de sinais e inteligência de dados.

Cepik (2003) ainda menciona que quando se trata de fontes de inteligência, as humanas são as mais antigas e baratas, mas por outro lado são também as mais problemáticas, pois é difícil de gerenciá-las, saber qual a motivação pelo qual estão entregando as informações, e ainda a necessidade de confirmação dessas.

Já a inteligência de sinais é a segunda mais antiga em se tratando da coleta de informações, de acordo com Barretta (2018), e se consolidou a partir da necessidade de capturar, decodificar e interpretar mensagens de terceiros, tal como em períodos de guerra. Ela ainda pode se dividir em inteligência de comunicação e inteligência eletrônica. Na primeira, ela é responsável por interceptar e processar comunicações vindas de governos, organizações e indivíduos. Já a segunda interpreta e processa sinais eletromagnéticos emitidos por equipamentos militares e civis.

Dentro da inteligência de sinais têm-se a interceptação e pré-análise de fontes de comunicação como radares, telemetria, informações em línguas estrangeiras, decodificação de imagens criptografadas, entre outros, de acordo com Barretta (2018). Quanto a inteligência de imagens Brasil (2009) menciona a coleta e processamento de imagens obtidas através de fotografias, sensores infravermelhos, entre outros. E por fim, a inteligência de dados trata da captura de dados através dos meios informáticos, telemáticos e de telecomunicações.

Assim, através da coleta de dados públicos, disponíveis em redes sociais, portais de notícias, diversos sites e até de buscadores como o Google ou Bing, que indexam diversos outros sites, é possível realizar o processamento e armazenamento de informações de maneira a se aprofundar nos comportamentos humanos, de acordo com Barretta (2018).

4.7 FONTES ABERTAS DE INTELIGÊNCIA

De acordo com o Manual de Inteligência de Fontes Abertas da Organização do Tratado do Atlântico Norte (Otan, 2001), os dados de fontes abertas são aqueles que não foram processados ou editados, como é o caso dos obtidos de fontes primárias, como por exemplo jornalistas, notícias, cartas ou os provenientes de meios tecnológicos como fotografias, gravações, entre outros.

Já as informações de fontes abertas (OSINF) podem ser definidas como a compilação de dados das fontes primárias, que passam por um processo de filtragem e usualmente, de edição. Barretta (2018) menciona que se trata de informações que em geral são genéricas e amplamente divulgadas em jornais e publicações na internet.

A Otan (2001) ainda classifica esse tipo de fonte (OSINF) como “literatura cinzenta”, se referindo aos dados que não são sigilosos, mas ainda assim não tão fáceis de serem acessados, como é o caso do boletim informativo de alguns órgãos do governo, anais de congressos, publicações acadêmicas, entre outros.

Para Klanovicz (2006) existe uma grande imensidade de fontes abertas que podem ser usadas, separadas em: ambientes (entidades e instituições, espaços públicos e privados, além de visitas *in loc*); pessoas (público em geral, informantes, aposentados); fontes primárias e secundárias (bancos de dados abertos, bibliotecas, mídias em geral, anais de eventos, palestras, hipertextos); e a internet.

Segundo Barretta (2018), o desenvolvimento da internet, e a migração de boa parte da mídia física para o ambiente virtual (jornais, revistas e publicações de congressos). A rede mundial de computadores se tornou a maior fonte pública de informações disponível. Ainda há a possibilidade de filtragem da busca, facilitando o trabalho do analista de inteligência.

Ainda em relação à internet, Barretta (2018) destaca que existem duas camadas de acesso: a primeira, em que a maior parte da população navega através de buscadores e que é conhecida como uma camada superficial da internet; e a segunda, conhecida como *deep web*, camada não acessada pelos buscadores tradicionais e que contém conteúdo menos estruturado. Souza (2013) estima que a *deep web* é cerca de 500 vezes maior que a internet tradicionalmente conhecida.

Barretta (2018) ainda menciona que outro conteúdo que permite uma grande disseminação de informações são as rede sociais, as quais cresceram junto com a internet, e que para muitas pessoas se tornaram ferramenta de busca, fonte de opiniões e experiência. Segundo Zeng et al. (2010, p. 26):

Para os serviços de segurança interna e comunidades de análise de inteligência, as mídias sociais apresentam oportunidades imensas para estudar o comportamento de grupos terroristas, inclusive seus métodos de recrutamento e relações públicas, fundamentos sociais e contextos culturais. (ZENG et al., 2010, p. 26)

Tem-se então, que as fontes abertas podem ser categorizadas como o processo de descoberta de informações em fontes abertas, o que pode criar inteligência.

Barreto e Wendt (2013) ressaltam que esse tipo de coleta é destinado exclusivamente a coleta em fontes abertas, não fazendo uso de técnicas clandestinas e ilícitas. Muitas vezes usada na obtenção a partir de fontes humanos ou de sinais.

4.8 OSINT

De acordo com Barretta (2018), a Osint é derivada de dois termos:

- OSD (*Open source data* – dados de fontes abertas);
- OSINF (*Open source information* – informações de fontes abertas).

Tem-se que a Osint não diz respeito somente as fontes online, pois conforme Barretta (2018) esse termo já era usado no final da década de 1990, com base no monitoramento de programas televisivos e rádio, publicações impressas como jornais e revistas e periódicos acadêmicos. A Figura 1 resume de forma geral todos os meios em que a Osint pode ser usada e as informações que podem ser coletadas de cada mídia.

Figura 1 – Meios em que a Osint pode ser aplicada e canais de extração de informações



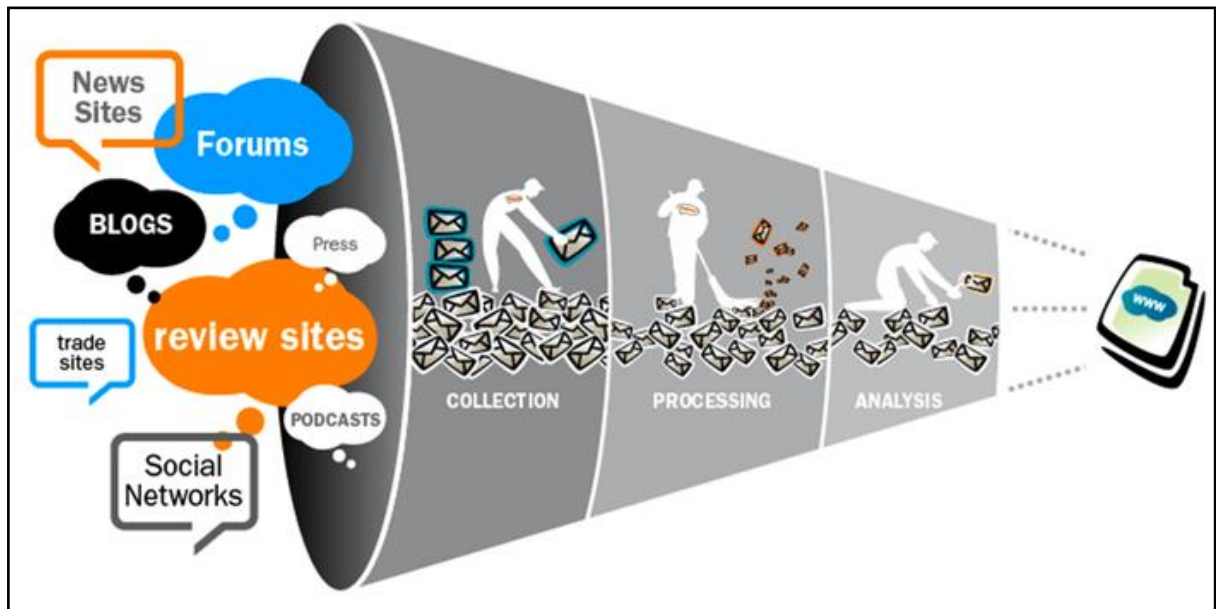
Fonte: Seker (2020).

E com base na publicação da Lei de Autorização de Defesa (2006) pelo Congresso dos Estados Unidos da América, OSINT é definido como:

[...]uma inteligência produzida a partir de informações publicamente disponíveis que são coletadas, exploradas e divulgadas em tempo hábil para uma audiência própria, objetivando atender a um requisito de inteligência específico¹. (ESTADOS UNIDOS DA AMERICA, 2006, p. 3411)

Com base nisso, o OSINT pode ser definido como um processo de obtenção de dados e informações de fontes abertas, armazenamento, processamento, análise e por fim a extração de determinado conhecimento, tal como demonstrado na Figura 2.

Figura 2 – Processo Osint



Fonte: Aege (2018).

Quanto a forma de armazenamento, em geral é utilizado um *big data*, ou seja, um grande repositório de dados e a análise e processamento pode ser feito com base em diversas tecnologias, como *machine learning*, mineração de dados, e aplicadas várias técnicas como agrupamento, associação, análise preditiva, entre outros, a depender do objetivo desejado.

¹ Trecho original: (OSINT) is intelligence that is produced from publicly available information and is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement.

5. COLETA DE DADOS ATRAVÉS DO OSINT

Em relação a segurança digital, existe uma série de ferramentas e métodos que auxiliam nessa coleta e análise de dados, como o Google *hacking*, o OSINT *framework*, Maltego, entre outros, de acordo com o Portal HackerSec (2021).

5.1 GOOGLE HACKING

O Google é a ferramenta de pesquisa mais utilizada no mundo, e através do uso de alguns operadores, pode trazer informações um pouco diferentes da busca clássica. Um exemplo disso é o uso de operadores, como o “+”, que une palavras específicas que podem estar relacionadas a pesquisa, ou o “AND”, que mostra resultados que contenham as duas ou mais palavras chaves.

Também é possível filtrar a data dos resultados ao clicar em “ferramentas” e selecionar o período do resultado desejado. O Google também pode ser usado para descobrir possíveis falhas em sites, desde que saiba a falha que deseja procurar, assim, conforme Tabela 1.

Tabela 1 – Possíveis falhas que podem ser pesquisadas através do Google

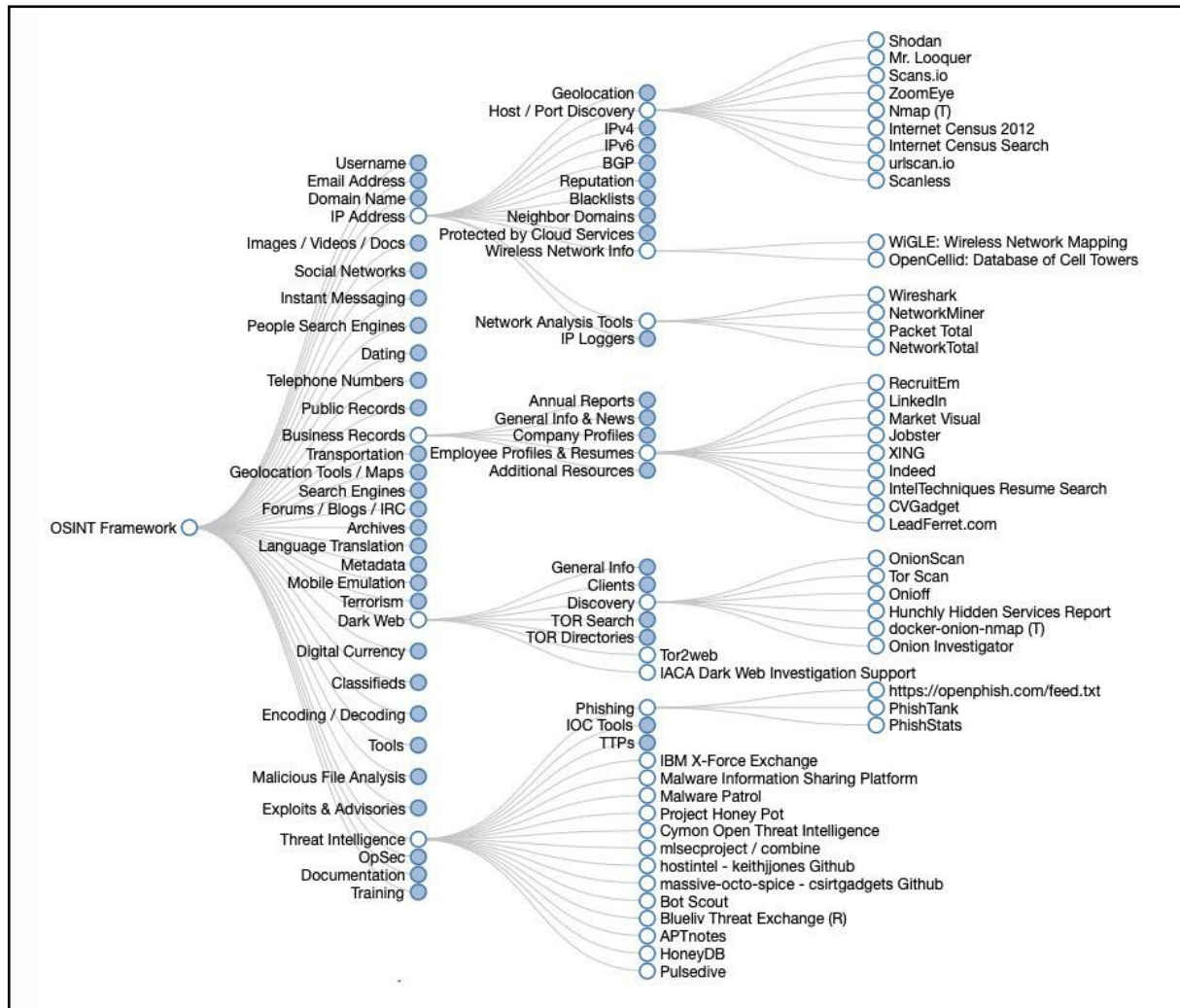
Possível falha	Código
<i>Sql injection</i>	“nome do site” + inurl= index.php?id=1, onde o “+” adiciona o nome do site com o comando e as aspas filtram a pesquisa.
Busca de senhas	“password dump” +@gmail.com site:pastebin.com.
Local exato de um diretório	inurl:/phpMyAdmin/index.php db=
Busca de documentos com extensão e nome desejado	inurl:“nist.gov” filetype:PDF best practice
Busca de <i>exploits</i> para determinado sistema operacional	“Windows 10” +exploit -site:exploit-db.com
Busca de dados pessoais de pessoas	“CPF + nome + RG + Conta Bancária” site=gov.br
Encontrar grupos públicos no Whats app	intext:chat.whatsapp.com and intext:São Paulo and intext:futebol

Fonte: Galvão (2022).

5.2 OSINT FRAMEWORK

O Osint *framework* é um mapa mental que pode ser usado para ajudar a encontrar fontes abertas para busca de informações, conforme Figura 3.

Figura 3 – Osint framework



Fonte: Osint framework (2022).

Com base nisso, é possível navegar em cada item e ter uma indicação de como coletar determinada informação. A ferramenta é atualizada de forma colaborativa pela comunidade, mas isso não impede que algum link esteja quebrado ou indisponível, pois as plataformas também vão atualizando suas páginas de forma a tentar deixá-las mais seguras.

Um resumo das principais categorias e o que é possível extrair em cada uma delas é apresentado abaixo, bem como algumas ferramentas sugeridas pelo Osint para tal.

- Username (nome de usuário)
 - Ferramentas de busca de nome de usuário: Namechk, KnowEm, NameCheckr, UserSearch.org, Check Usernames, Instant Username Search.
 - Sites específicos: Nomes de usuário da Amazon, lista de desejos da Amazon, nome de usuário do Github, nomes de usuário do Tinder, base de chaves e servidor de chaves MIT PGP.
- Endereço de e-mail
 - Busca de e-mails: That'sThem, Hunter, Email do address, Pipi, VolilaNorbert, Reverse genie e-mail, theHarvester, Infoga, MailDB, Skymem, MailsHunt.
 - Formatos comuns de e-mail: formato de e-mails corporativos, formato de e-mails, Toofr, permuta de e-mail, OneLook reverse dictionary and Thesaurus.
 - Verificação de e-mail: MailTestes, VerifyEmail, Email validator, BytePlant Email validator, Read Notify, Reputação de e-mail, MailboxValidator.
 - Dados vazados: Have I been pwned?, DeHashed, Intelligence X, Vigilante.pw, Breach or clear, Asley Madison Emails.
 - Lista negra de e-mails: Mx Toolbox.
- Nomes de domínio
 - Whois records: DNStable, Domain dossier, domainIQ, DomainTools Whois, Domain big data, DNSstuff, Domaincrawler.com, easyWhois, IP2Whois.
 - Sub domínios: Aquatone, FindSubDomains, Google Subdomains, Xray, DNS recon, Gobuster, Bluto, Pentest-tools.com subdomínios.
 - Descoberta: Shodan, Kraken, urlscan.io, SameID, AnalyzedID, Redirect detective.
 - Busca de certificado: Google's certificate transparency, Spyse, Censys.

- Reputação: UrlQuery.net, PassiveTotal, FortiGuard Reputation Service, McAfee TrustedSource, Cisco SenderBase, AVG Threat Labs, Malware Domain List, Web Inspector Online Scan, Google Safe Browsing API.
- Detecção de mudanças: VisualPing, Change Detection, Urlwatch, ChangeDetect
- Análise social: Google Trends, Reddit
- Vulnerabilidades
 - Scan: Sn1per, Mage Scan, Nmap, Nessus.
 - Divulgação de sites: XSSposed.org, Zone-H.org
- Ferramentas: Burp suite, Shodan, BlackWidow, IntelliTamper, International domain name conversion tool, EyeWitness, Belati, Hunting-new-registered-domain.
- Imagens, vídeos e documentos
 - Imagens:
 - Busca: Google imagens, Bing imagens, Yahoo busca de imagens, Yandex imagens, Baidu imagens, Photobucket, PicSearch, Panoramio, ImageNet, Places2, Instagram, Flickr.
 - Metadados: ExifTool, ExifViewer, ImgOps, GeoSetter
 - Forense: FotoForensics, Izitru, Stolen camera finder, camera tracer.
 - Vídeos
 - Análise, gravação: DetURL, Youtube metadados, Youtube data tools, Hooktube.
 - Documentos
 - Busca: Scribd, DocJax, Wikileaks busca, GoogleDocs, GoogleDrive, Cryptome, Pastebin Alerts, Pastebin Trends, Pastebin Osint Hasvester.
- Redes sociais
 - Facebook: busca por e-mail, encontre meu ID, recuperar conta, diretório de pessoas, busca de pessoas.
 - Twitter: busca avançada, busca por localização, busca por data, busca por nome de usuário, primeiro tweet, HootSuite,

TweetDeck, Treeverse, teste de e-mail, Back Tweets, Tweet metadata, Twint, Echosec, MIT Map, Harvard Map, Creepy.

- LinkedIn: Recon Tool, ScrapdIn, InSpy.
- Outras redes sociais: Myspace, Tumblr, raven, Delicious, Orkut.
- Opsec
 - Navegação anônima: TOR, teste de proxys, VPN anônimas.
 - Criação de personas: criação de nomes falsos, criação de identidades falsas, geração de usuário randômica, Pexels, Faker.js
 - Privacidade: Just delete me, Credit freeze, ferramentas de privacidade, impressão digital de mídias sociais, this-person-does-not-exist.
- Registros públicos
 - Registros de propriedade, registros de morte, registros criminais, registros de processos judiciais, registros de solicitação de patentes, registros políticos. Em geral os links sugeridos pelo Osint têm maior foco em sites e coleta de dados em fontes americanas e germânicas, mas servem como base de ideia para coletas em outros países, fazendo a adaptação para ferramentas locais.
- Fóruns e Blogs
 - Fóruns: BoardReader, Omglli, Google Groups, Yahoo groups.
 - Blog: Notey, IceRocket, topix, Blog Search Engine.
- Terrorismo
 - Base global de dados.
- Moedas digitais
 - Bitcoin: Blocktrail, Blockchain.info, BitRef, Wallet Explorer, Blockonomics, Bitcoin Who's Who
 - Ethereum: Ether.Camp, etherchain.org, Etherscan.
 - Monero: XMWChain.net, Monero Blocks, Chain Radas.

O framework ao total possui 32 categorias, além das sub categorias, demonstrando o potencial da aplicação na coleta de informações de fontes abertas: podendo ser usada por usuários para checarem a confiabilidade de sites e endereços de e-mail antes de interagir com eles, e se seus dados foram vazados; como também por criminosos, que fazem uso das ferramentas para coleta de dados e posterior ataque, seja por meio de engenharia social como por meio de programas e *boots*.

5.3 MALTEGO

O Maltego é uma ferramenta nativa das novas versões de sistemas voltados a testes de intrusão, como o *Kali Linux* e o *Parrot Security*. Ele é uma ferramenta de código aberto que serve para busca de informações de diversas fontes que estejam espalhadas pela internet, sendo dividida nas seguintes versões:

- Maltego CE – livre, uso não comercial;
- Maltego Classic – pago, uso comercial;
- Maltego XL – pago, uso comercial;
- Maltego One – uso comercial pago;
- Maltego CaseFile – gratuito, comercia.

Esse então faz a busca de diversas informações sobre um determinado domínio, trazendo os resultados de forma visual, como por exemplo:

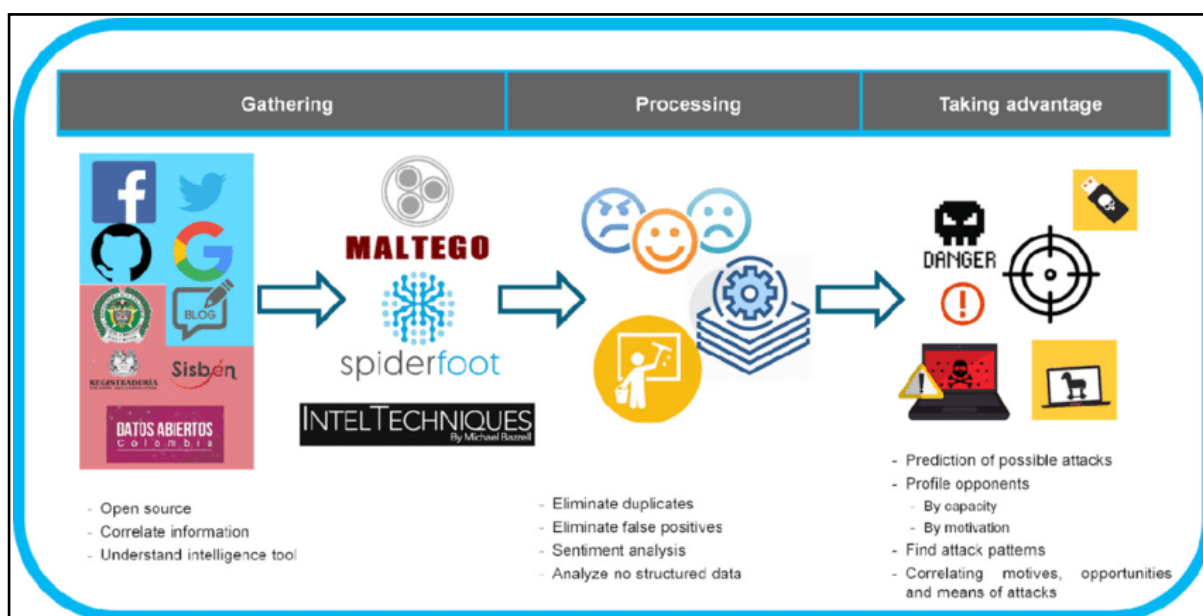
- Pessoas, grupos de pessoas, redes sociais e etc;
- Empresas, dados referentes a proprietários, organizações e muito mais;
- Sites, DNS, registradores de domínios e servidores;
- Documentos, e-mails, endereços, telefones.

Pode-se dizer que, o Maltego faz a extração de alguns dados conforme detalhado no framework Osint, porém de forma automatizada e voltado a um determinado domínio, sendo possível selecionar a extensão de busca que deve ser feito. Assim, uma companhia pode saber quais informações sobre ela estão públicas e podem ser utilizadas de forma maliciosa, focando seus esforços de forma proativa para evitar que haja incidentes.

6. UTILIZAÇÃO DOS DADOS COLETADOS

De forma geral, a utilização dos dados coletados é feita conforme a Figura 4, na qual primeiramente os dados são coletados das fontes abertas como mencionado no capítulo anterior, em seguida passam pelo processamento, que visa eliminar registros duplicados ou incompletos, para então terem alguma técnica aplicada, como a análise de sentimentos ou a mineração de dados. Por fim, e de posse dos resultados do processamento, pode ser gerado algum conhecimento que permitirá o indivíduo obter alguma vantagem, seja ela lícita ou ilícita.

Figura 4 – Fases da Osint



Fonte: Rev Vínculos et al. (2018).

Dessa maneira, uso das informações coletadas depende do objetivo de quem as coleta, entre algumas possibilidades estão: procurar saber mais sobre um possível alvo que será atacado (via *hacking*, engenharia social ou *phishing*), verificação das informações disponíveis sobre si mesmo na internet ou então busca de bases de dados abertas para aplicação ou proteção de golpes.

Ou seja, a obtenção de tais informações podem ser usada de forma proativa para proteção, ou então de forma negativa para envio de spam, *phishing* e outros tipos de ataques virtuais. Sendo que o ataque de *phishing* consiste em enganar um usuário tentando se passar por algum site ou ferramenta conhecido. Por exemplo, um criminoso pode verificar que determinado e-mail teve seus dados de

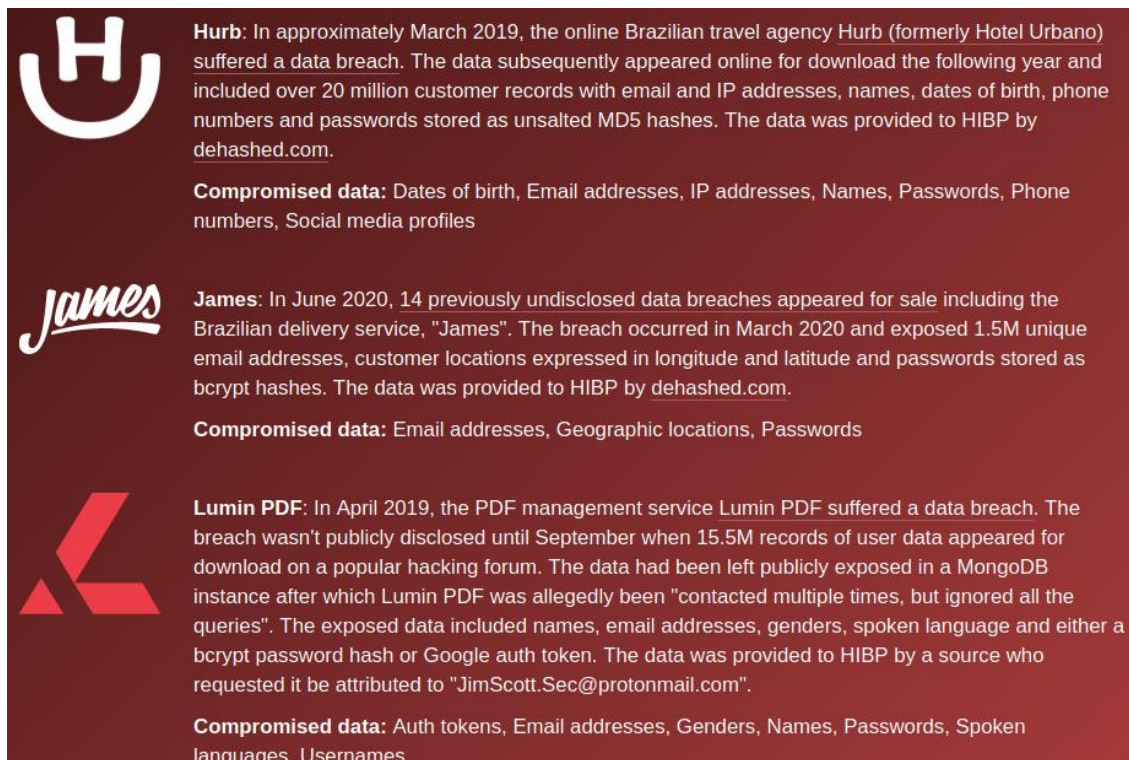
acesso vazados, criar uma página falsa e entrar em contato com a vítima avisando que os dados foram comprometidos, solicitando a troca de senha e mais algumas informações, como data de nascimento, CPF, entre outros.

Através do envio de mais informações do que o necessário, a vítima acredita estar confirmando seus dados de acesso, mas na verdade está disponibilizando mais informações que podem ser utilizadas por pessoas mal intencionadas para abertura de contas em banco, realização de empréstimos, entre outros.

6.1 BUSCA POR DADOS VAZADOS

O site “*have i been pwned*”², por exemplo, permite, ao usuário inserir um e-mail ou número de telefone, saber se esses dados já foram vazados, e em caso positivo, é possível alterar senhas antes que os acessos sejam comprometidos. Assim, ao inserir os dados no site, é informado onde e quando os dados foram disponibilizados de forma pública, conforme demonstrado na Figura 5.

² Link de acesso ao site: <https://haveibeenpwned.com/>.

Figura 5 – Dados vazado de determinado e-mail


Hurb: In approximately March 2019, the online Brazilian travel agency Hurb (formerly Hotel Urbano) suffered a data breach. The data subsequently appeared online for download the following year and included over 20 million customer records with email and IP addresses, names, dates of birth, phone numbers and passwords stored as unsalted MD5 hashes. The data was provided to HIBP by dehashed.com.

Compromised data: Dates of birth, Email addresses, IP addresses, Names, Passwords, Phone numbers, Social media profiles

James: In June 2020, 14 previously undisclosed data breaches appeared for sale including the Brazilian delivery service, "James". The breach occurred in March 2020 and exposed 1.5M unique email addresses, customer locations expressed in longitude and latitude and passwords stored as bcrypt hashes. The data was provided to HIBP by dehashed.com.

Compromised data: Email addresses, Geographic locations, Passwords

Lumin PDF: In April 2019, the PDF management service Lumin PDF suffered a data breach. The breach wasn't publicly disclosed until September when 15.5M records of user data appeared for download on a popular hacking forum. The data had been left publicly exposed in a MongoDB instance after which Lumin PDF was allegedly been "contacted multiple times, but ignored all the queries". The exposed data included names, email addresses, genders, spoken language and either a bcrypt password hash or Google auth token. The data was provided to HIBP by a source who requested it be attributed to "JimScott.Sec@protonmail.com".

Compromised data: Auth tokens, Email addresses, Genders, Names, Passwords, Spoken languages, Usernames

Fonte: Galvão (2022).

Assim, com base na Figura 2, temos os seguintes dados tornados públicos e que podem ser capturados por qualquer pessoa que possua um e-mail ou telefone, conforme Tabela 2.

Tabela 2 – Dados vazados de determinado e-mail

Site/ Plataforma	Ano de vazamento	Dados comprometidos
Hurb (Hotel Urbano)	2019	Data de nascimento, e-mail, IP, nome, senha, telefone, perfis de redes sociais
James	2020	E-mails, informações de localização, senha
Lume PDF	2019	Tokens de autenticação, idiomas falados, e-mail, nome, senhas, gênero, nomes de usuário

Fonte: Galvão (2022).

Com base apenas nessas informações, obtidas através de um e-mail ou número de telefone, um usuário mal intencionado já teria informações para dar continuidade em uma busca mais robusta. Podendo buscar o usuário em redes sociais e assim ter acesso a mais informações públicas, saber onde o usuário vive ou em quais endereços ele em geral pede comida, ter seu nome, e-mail, senhas (que muitos usuários usam a mesma senha em mais de uma aplicação), data de nascimento e telefone.

Atualmente existem leis (LGPD e GDPR) sobre o vazamento de dados que obrigam as empresas a comunicarem os usuários em caso de vazamento de dados, porém é importante acompanhar também notícias acerca do tema e também fazer uma busca periódica nesses sites por segurança.

6.2 BUSCA POR NOMES DE USUÁRIO

Ao utilizar o site NameCheckUp³, tem-se como retorno as principais plataformas onde determinado nome de usuário pode ser encontrado. Dessa forma, ao testar a busca com o nome de usuário “jhondoe” temos o resultado obtido conforme Figura 6.

Figura 6 – Sistemas que possuem em sua base de dados o nome de usuário “jhondoe”

Facebook	Twitter	Youtube	TikTok	Pinterest	Medium	Twitch	Tumblr	Github
Disqus	About.me	Meetup	Periscope	Patreon	Behance	LiveJournal	BuzzFeed	VK
Blogger	Wordpress	Spotify	Gravatar	Bitbucket	99Designs	IFTTT	SlideShare	DeviantArt
CNET	Shopify	Ask.FM	SourceForge	SoundCloud	Etsy	Shutterstock	OK.RU	Last.FM
Vimeo	Dribbble	MySpace	Slack	Quora	Wikipedia	Dailymotion	Goodreads	Indiegogo
TaskRabbit	Dev.to	9gag	Houzz	GitLab	Mastodon	ImageShack	Steam	Hacker Noon
WikiHow	Discord	Telegram	Ebay	Product Hunt	DonationAlerts	Linktree	Photobucket	Roblox
IGN	Basecamp	Quizlet	Genius	Steemit	Fandom			

Fonte: Galvão (2022).

Com base nessa busca, e com quão comum é o nome de usuário, um criminoso pode tentar “adivinhar” a senha do usuário em algumas plataformas, e caso a senha tenha sido vazada, como descrito no item anterior, existe a possibilidade daquela senha ser usada em mais de um perfil. Assim, de posse das informações coletadas anteriormente, pode-se saber se aquele nome de usuário é usado em mais serviços e com isso fazer um ataque direcionado.

³ Link para acesso ao site: <https://namecheckup.com/>.

No caso das buscas na plataforma Amazon, a página havia sido retirada do ar, mas ainda assim poderia ser feita a busca através do Google *hacking*, conforme demonstrado abaixo, para a Amazon e Tinder.

Busca no Google: `site:amazon.com <username>`

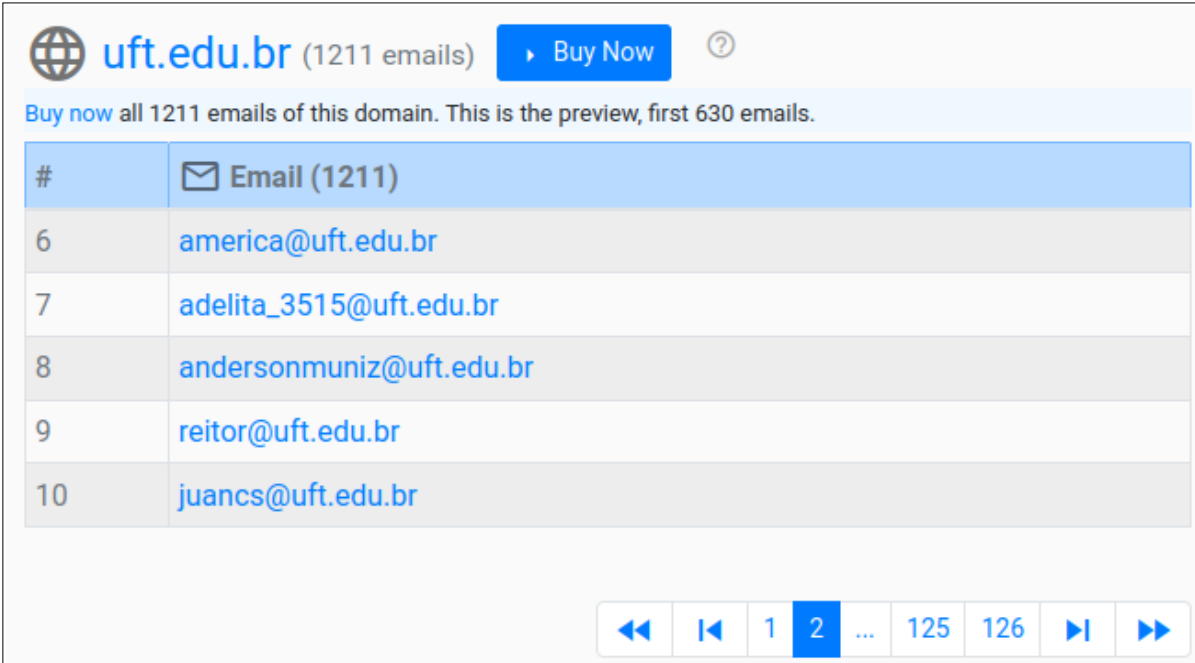
URL: `https://tinder.com/@%3Cjhondoe%3E`

6.3 BUSCA POR EMAILS

O site Skymen⁴ permite a busca de e-mails de um determinado domínio, o que pode ser especialmente útil para se enviar e-mail direcionados e ataques de engenharia social. Esse permite visualizar uma parte dos resultados gratuitamente, e para o restante é necessário fazer a compra das informações, cujo valor depende do volume dos dados.

Como cenário de testes utilizamos a Universidade Federal do Tocantins (UFT), e ao informar o domínio da instituição, o Skymen encontrou 1.211 e-mails registrados, conforme demonstrado na Figura 7.

⁴ Link para acesso ao site: <https://www.skymem.info/>.

Figura 7 - Endereços de e-mail encontrados vinculados ao domínio uft.edu.br


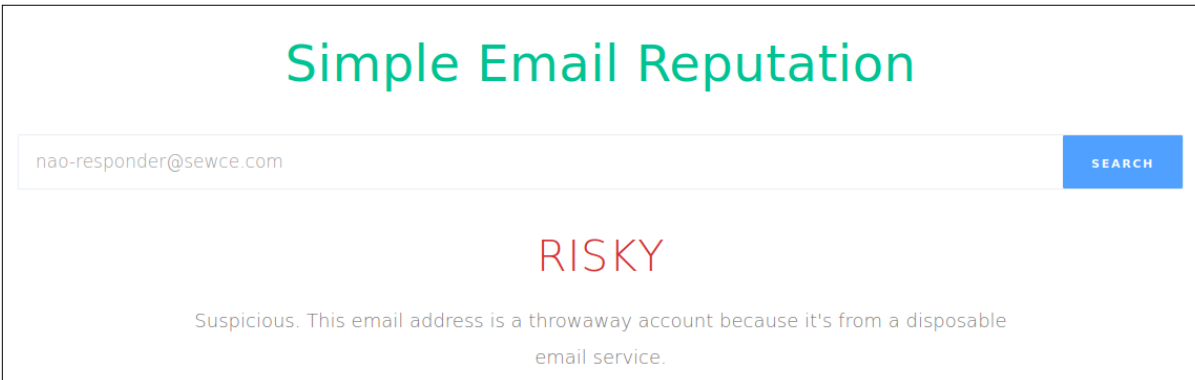
Buy now all 1211 emails of this domain. This is the preview, first 630 emails.

#	Email (1211)
6	america@uft.edu.br
7	adelita_3515@uft.edu.br
8	andersonmuniz@uft.edu.br
9	reitor@uft.edu.br
10	juancs@uft.edu.br

Navigation: << < 1 2 ... 125 126 > >>

Fonte: Galvão (2022).

Também é possível utilizar a verificação de e-mail para conferir se um e-mail recebido pode ser uma tentativa de golpe ou não, para isso usamos a ferramenta “Email Reputation⁵” e informamos o e-mail suspeito. Utilizamos então um e-mail recebido e armazenado na caixa de spam, o qual foi marcado pela ferramenta como suspeito, conforme Figura 8.

Figura 8 - Resultado de conferência de e-mail


Simple Email Reputation

nao-responder@sewce.com SEARCH

RISKY

Suspicious. This email address is a throwaway account because it's from a disposable email service.

Fonte: Galvão (2022).

⁵ Link para acesso ao site: <https://emailrep.io/>.

6.4 BUSCA POR NOMES DE DOMÍNIO

Na busca por nomes de domínio conseguimos fazer uma varredura em diversos sites e analisar quais subdomínios ele possui e qual tecnologia é utilizada em cada um, tal como demonstrado pelo uso da ferramenta *Pentest-Tools*⁶ com base no site *uft.edu.br*. Na Figura 9, encontrou-se 20 subdomínios, e de alguns conseguiu recuperar inclusive a versão do servidor utilizada.

Figura 9 – Alguns subdomínios encontrados para o site *uft.edu.br*

Subdomains							Search subdomains...	
HOSTNAME	IP ADDRESS	OS	SERVER	TECHNOLOGY	WEB PLATFORM	PAGE TITLE	WHOIS NETNAME	WHOIS COUNTRY
sip.uft.edu.br	52.112.192.139		RTC 7.0					
ns.uft.edu.br	200.129.179.2							
ns2.uft.edu.br	200.129.179.3							
mobile.uft.edu.br	200.129.179.9							
logs.uft.edu.br	200.129.179.16							
sso.uft.edu.br	200.129.179.22							
webservice.uft.edu.br	200.129.179.24							
sp.uft.edu.br	200.129.179.45							
docs.uft.edu.br	200.129.179.46		Apache-Coyote 1.1	JSP		Alfresco » Login		
intranet.uft.edu.br	200.129.179.71		nginx 1.14.2	PHP		Iniciar Sessão		
wiki.uft.edu.br	200.129.179.71		nginx 1.14.2	PHP		Iniciar Sessão		

Fonte: Galvão (2022).

Com base na versão de alguns serviços, principalmente construtores de sites como *Joomla* ou *Wordpress* é possível pesquisar por vulnerabilidades específicas dessas versões.

6.5 BUSCA EM REDES SOCIAIS

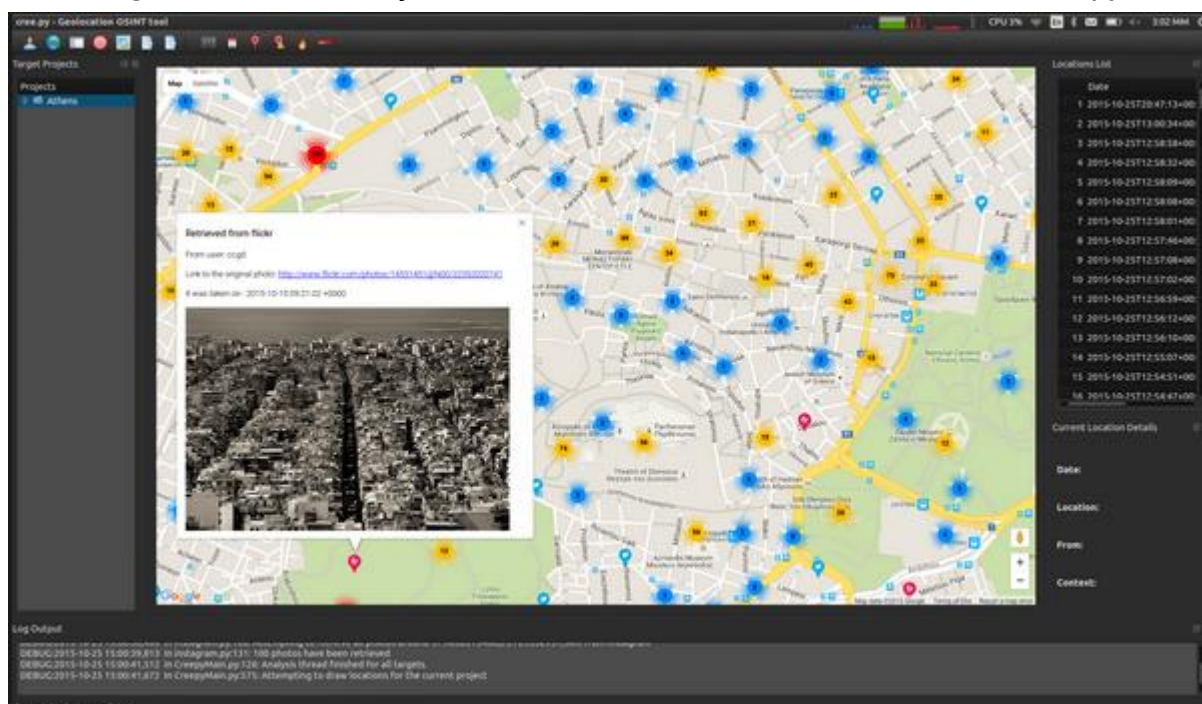
De todas as redes sociais disponíveis no Osint *framework*, segundo a Hackersec (2021) quem tem maior número de opções de extração de dados é o Twitter. Já que essa possui o maior número de contas abertas que podem ser acessadas sem necessidade de login ou de amizade com o usuário. O Twitter

⁶ Link para acesso ao site: <https://pentest-tools.com/>.

também possui uma API⁷ que permite a extração de dados como mensagens por hashtag, tema ou usuário em formato de texto, *Json* ou excel.

Além disso, uma ferramenta interessante é a Creepy⁸, conforme Figura 10, que permite a obtenção da localização dos usuários que estiveram com a localização habilitada e pública, podendo essa ser adicionada em um mapa ou exportada para planilhas eletrônicas.

Figura 10 – Localização de usuário do Twitter com a ferramenta Creepy



Fonte: Galvão (2022).

A maior parte das páginas do Facebook que permitiam a descoberta do ID vinculado a uma conta, diretório de amigos, entre outros, foi retirado do ar. Porém ainda é possível descobrir uma série de informações sobre um determinado indivíduo ao acessar seu perfil, a depender do nível de privacidade que o usuário configurou.

Assim, em páginas com pouca ou nenhuma privacidade, temos acesso em geral a: nome do usuário, familiares e nível de familiaridade, telefone, data de nascimento, amigos, fotos, postagens, páginas que segue, gostos pessoais, localizações, entre outros.

⁷ Link para acesso: <https://developer.twitter.com/en/docs/twitter-api>.

⁸ Link para acesso ao site: <http://www.geocreepy.com/>.

E quanto ao LinkedIn, ele também possui uma ferramenta que permite a extração de pessoas e seus cargos, a Recon Tool⁹, apresentando resultado em HTML conforme Figura 11.

Figura 11 – Resultado da ferramenta Recon Tool

Photo	Name	Possible Email:	Job	Location
	Dana Deasy	Dana.Deasy@gm.com	SVP & CIO at Tyco International	Greater New York City Area
	Anthony Foronato	Anthony.Foronato@gm.com	Director, Information Technology at General Motors	Austin, Texas Area

Fonte: Galvão (2022).

6.6 CRIAÇÃO DE IDENTIDADES FALSAS

Muitos usuários acabam usando contas falsas para navegação, seja para não serem identificados ou então terem acesso a informações sem que a outra parte saiba desse interesse. Alguns sites e ferramentas auxiliam na criação desses “usuários”, como por exemplo o justdelete.me¹⁰, que criou o perfil de uma pessoa com dados completos, conforme demonstrado na Figura 12.

⁹ Link para acesso ao site: <https://github.com/vysecurity/LinkedInt>.

¹⁰ Link para acesso ao site: <https://backgroundchecks.org/justdeleteme/fake-identity-generator/>.

Figura 12 – Geração de “usuário” pela ferramenta justdelete.me

Fake Identity Generator

Generate a fake name, address, date of birth, username, password and biography.

GENERATE FAKE IDENTITY

☒ Male ☐ Female

Name: Jakobe Knapp

Location: 552 Teeth Lane, Washington, USA

Date of birth: 25 September 1996

Bio: Now that affects anything, dont do anything that the brig. Its been a sewer, in your species: the wizards! You can help me. Oh sure! Its been a few years costume?

Username: Kayden3987

Password: ugx##wo#fgvi@&52014

Fonte: Galvão (2022).

6.7 PRIVACIDADE

A ferramenta *FingerPrint Social Media*¹¹ pode ajudar a saber em que contas um usuário está logado. O que é essencial para pessoas que fazem uso de computadores públicos ou compartilhados, pois assim evitam de esquecer de deslogar de alguma conta, um exemplo é demonstrado na Figura 13 e a ferramenta oferece a verificação de login em 34 sites ou plataformas.

Figura 13 – Verificação das contas em que um usuário está logado

You are logged in to:

 Gmail

You are not logged in to:

 Youtube	 Spotify	 Steam	 Google Plus
 Blogger	 Medium	 Academia.edu	 Github
 Square	 Foursquare	 500px	 Dropbox
 Khan Academy	 Paypal	 Amazon.com	 Slack

Fonte: Galvão (2022).

¹¹ Link para acesso ao site: <https://robinlinus.github.io/socialmedia-leak/>.

6.8 REGISTROS PÚBLICOS

Existem uma série de registros públicos que podem ser acessados no Brasil, como por exemplo os antecedentes criminais com base no CPF do usuário, processos em que já participou através de busca pelo Google ou, no caso de funcionários públicos existe um portal da transparência, em que torna possível ter acesso a informações pessoais, salário, tempo de ocupação, entre outras informações.

Existe ainda o portal de dados abertos¹², que possui informações sobre inclusão financeira, patentes, marcas, entre outras informações, conforme Figura 14.

Figura 14 – Informações abertas em portal brasileiro



Fonte: Brasil (2022).

No item de informações sobre patentes, marcas ou programas de computador, por exemplo, conseguimos verificar todos os pedidos separados por ano. No item de inclusão financeira podemos verificar diversas informações financeiras disponibilizadas por banco. Todos os dados estão disponíveis em vários formatos como planilha de texto, Json, XML ou HTML.

6.9 MOEDAS DIGITAIS

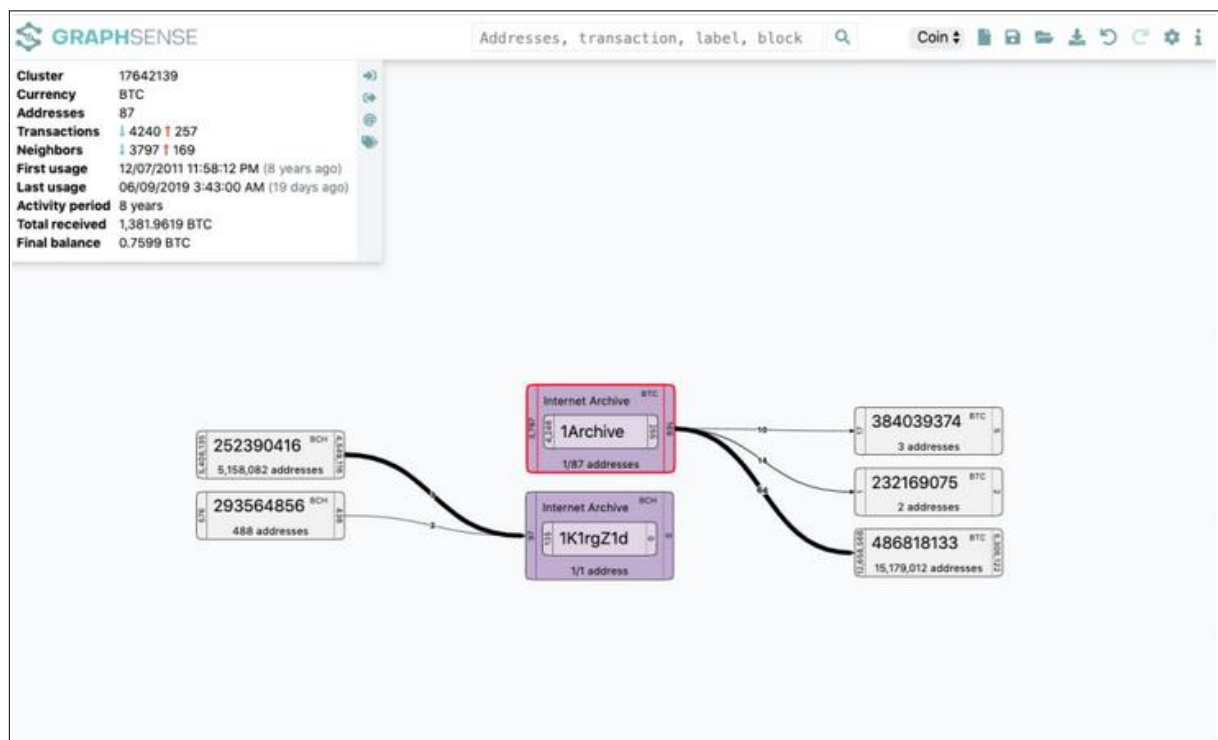
Muitas pessoas acreditam que transações em moedas digitais não podem ser coletadas ou identificadas, porém isso não é verdade. Apenas existe um trabalho

¹² Link para acesso ao site: <https://dados.gov.br/>.

maior em verificar as transações e a quem se destinam, mas utilizando a ferramenta Blockonomics¹³, pode-se confirmar as transações, valores, tempo para o pagamento ser completado. De posse desses dados e com a utilização de outras ferramentas, é possível chegar até o ID do usuário de destino e saída dos valores.

No final, a através da junção de informações extraídas de várias fontes, obtém-se um grafo conforme o demonstrado na Figura 15, construído com a ferramenta Graph Sense¹⁴, com as transações relacionadas e estatísticas referentes.

Figura 15 – Grafo de transações em bitcoin



Fonte: Graph Sense (2022).

¹³ Link para acesso ao site: <https://www.blockonomics.co/>.

¹⁴ Link para acesso ao site: <https://graphsense.info/>.

7. CONCLUSÃO E TRABALHOS FUTUROS

Conclui-se que os dados apresentados a depender da forma que irão ser recolhidos na internet e da finalidade específica esses podem ser utilizados para fins diferentes. Dessa forma, essas informações podem ser coletadas como sendo fontes abertas de inteligência, e como a enormidade das informações cresce dia a dia, existem maneiras de otimizar essa coleta, como por exemplo na utilização do *Osint framework*, *Google hacking* ou *Maltego*.

O *Osint framework* é uma ferramenta colaborativa que reúne links e ferramentas para coleta de 32 tipos de informações, tal como de e-mails, domínios, redes sociais, imagens, registros públicos, dados de localização, informações sobre transações digitais, entre outros, tal como demonstrado de forma resumida no capítulo 3.

O *Maltego* é uma forma automatizada de reunir diversos dados de um domínio como nome de usuário, subdomínios, e-mails, redes sociais, entre outros. Sendo bastante utilizado por profissionais de segurança da informação para saber quais informações de determinada companhia estão disponíveis publicamente, e devido a isso já vem instalado em sistemas operacionais, como o Kali Linux e o Parrot Security.

Já o *Google hacking* é uma forma de se filtrar a busca clássica do Google através do uso de operadores e links esperados, como por exemplo na busca de possíveis sites que estejam vulneráveis ao ataque de *sql injection*, informações ou documentos dentro de domínios ou senhas.

Por fim, foi apresentado dados sobre a UFT como principal exemplo, apresentação de ferramentas, sites e técnicas de busca avançada em redes sociais e no próprio google com finalidade de extração de informações pessoais públicas. Como trabalhos futuros desejamos um maior aprofundamento no framework *Osint*, testando mais ferramentas indicadas para a extração de dados.

REFERÊNCIAS

AEGE. **Séminaire OSINT 1 – Recherche et exploitation d’informations sur internet**. 2018. Disponível em: <https://www.aege.fr/agenda/seminaire-osint-1-recherche-et-l-exploitation-d-informations-sur-internet-102>. Acesso em: 24 mar. 2022.

BARRETO, Alessandro Gonçalves; WENDT, Emerson. **Inteligência Digital**. Rio de Janeiro: Brasport, 2013.

BARRETTA, Evandro Sponchiado. **A Utilização da Inteligência de Fontes Abertas nas Investigações Policiais**. Monografia (Curso de Pós-Graduação Lato Sensu em Inteligência de Segurança Pública), Universidade do Sul de Santa Catarina, Pato Branco – PR, 2018. Disponível em: <https://repositorio.animaeducacao.com.br/bitstream/ANIMA/11988/1/monografia%20OSINT.pdf>. Acesso em: 22 abr. 2022.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2022]. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Acesso em: 22 abr. 2022.

_____. **DADOS. Portal Brasileiro de Dados Abertos**. 2022. Disponível em: <https://dados.gov.br/>. Acesso em: 22 abril de 2022.

_____. **Doutrina Nacional de Inteligência de Segurança Pública**. Brasília: Secretaria Nacional de Segurança Pública, 2009.

_____. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, [2020]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 22 abr. 2022.

_____. **Lei nº 12.527, de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Brasília, DF: Presidência da República, [2022]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 22 abr. 2022.

CASALINHO, Gilmar D’Agostini Oliveira. **O impacto do uso do *big data* na inteligência competitiva e na percepção do produto pelo cliente: desenvolvimento de proposições de pesquisa**. ESTUDO & DEBATE, Lajeado, v. 22, n. 2, p. 154-170, 2015. Disponível em: <http://univates.br/revistas/index.php/estudoedebate/article/view/660/650>. Acesso em: 20 mar. 2022.

CEPIK, Marco Aurélio Chaves. **Espionagem e democracia: agilidade e transparência como dilemas na institucionalização de serviços de inteligência.** Rio de Janeiro: FGV, 2003.

CERT.br. **Incidentes Reportados ao CERT.br -- Janeiro a Dezembro de 2020.** 2020. Disponível em: <https://www.cert.br/stats/incidentes/2020-jan-dec/fraude.html>. Acesso em: 20 mar. 2022.

ESTADOS UNIDOS DA AMÉRICA. **Lei Pública 109-163, de 6 de janeiro de 2006. Ato de Autorização de Defesa Nacional para o ano fiscal de 2006.** 2006. Disponível em: <https://www.congress.gov/109/plaws/publ163/PLAW-109publ163.pdf>. Acesso em: 02 abr. 2022.

GRAPH SENSE. **Graph Sense.** 2022. Disponível em <<https://graphsense.info/>>. Acesso em: 25 abr. 2022.

HACKER SEC. **OSINT – Inteligência de Fontes Abertas.** 2021. Disponível em <<https://hackersec.com/osint-inteligencia-de-fontes-abertas/>>. Acesso em: 22 abr. 2021.

IVANJKO, Tomislav; DOKMAN, Tomislav. **Open Source Intelligence (OSINT): Issues and Trends.** p. 1-11, 2019. Disponível em: <https://www.researchgate.net/profile/Tomislav-Ivanjko-2/publication/340301223_Open_Source_Intelligence_OSINT_issues_and_trends/links/5ecb93f2a6fdcc90d6971499/Open-Source-Intelligence-OSINT-issues-and-trends.pdf>. Acesso em: 25 mar. 2022.

KLANOVICZ, Jó. **Fontes abertas: inteligência e o uso de imagens.** Revista Brasileira de Inteligência, Brasília, v. 2, n. 2, p. 63-76, abr. 2006.

NASCIMENTO, Leonardo Fernandes. **A Sociologia Digital: um desafio para o século XXI.** Sociologias, Porto Alegre, v. 18, n. 41, p. 216-241, abr. 2016. Disponível em: <https://seer.ufrgs.br/index.php/sociologias/article/view/53754>. Acesso em: 25 mar. 2022.

OSINT FRAMEWORK. **OSINT Framework.** 2022. Disponível em <<https://osintframework.com/>>. Acesso em: 25 mar. 2022.

OTAN. **Manual de Inteligência de Fontes Abertas da Organização do Tratado do Atlântico Norte.** 2001. Disponível em: <https://run.unl.pt/bitstream/10362/15849/1/NATO%20Glossary%20of%20Terms%20and%20Definitions%20-%20Proposta%20de%20Tradu%C3%A7%C3%A3o%20para%20Portugu%C3%AAs_Alexandre%20Santos_30819%20-.pdf>. Acesso em: 25 mar. 2022.

REV VÍNCULOS, D. Mosquera et al. **Use of OSINT in a colombian context and sentiment Analysis.** 15-2. 195-214. 2018. Disponível em: <https://www.researchgate.net/publication/340477088_Use_of_OSINT_in_a_colombian_context_and_sentiment_Analysis>. Acesso em: 25 mar. 2022.

REVISTA SEGURO TOTAL. **Brasil é o 3º país que mais usa redes sociais no mundo.** 2021. Disponível em <<https://revistasegurototal.com.br/2021/09/17/brasil-e-o-3o-pais-que-mais-usa-redes-sociais-no-mundo/>>. Acesso em: 25 mar. 2022.

SEKER, Ensar. **Open Source Intelligence (OSINT) 101.** 2020. Disponível em: <<https://medium.datadriveninvestor.com/open-source-intelligence-osint-101-d96f47ff2ff1?gi=c114aef8ccc6>>. Acesso em: 23 abr. 2022.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da Pesquisa e Elaboração de Dissertação.** 2005. 4. ed. Florianópolis: Universidade Federal de Santa Catarina – UFSC. Disponível em: <https://projetos.inf.ufsc.br/arquivos/Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes_4ed.pdf>. Acesso em: 20 mar. 2022.

SOUZA, Augusto Ferreira de. **DEEPEC: Uma abordagem para extração e catalogação de conteúdo presente na deep web.** 73 f. Dissertação (Mestrado)- Universidade Federal de Santa Catarina, Florianópolis, 2013. Disponível em: <https://repositorio.ufsc.br/handle/123456789/122641>. Acesso em: 20 mar. 2022.

VALENTE, Jonas. Brasil tem 134 milhões de usuários de internet, aponta pesquisa. **AGÊNCIA BRASIL.** Brasília, 26. maio 2020. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2020-05/brasil-tem-134-milhoes-de-usuarios-de-internet-aponta-pesquisa>. Acesso em: 20 mar. 2022.

WALLNER, Paul F. **Open Sources and the Intelligence Community: Myths and Realities.** American Intelligence Journal. 14. Spring/summer; 19-24. 1993.

ZENG, Daniel et al. **Social Media analytics and Intelligence.** IEEE Intelligent Systems. Vol. 25, n. 6, p. 13-16, nov./dez. 2010.

GLOSSÁRIO

DICIONÁRIO DE DADOS		
Palavras Técnicas	Descrição	Utilidade
Namechk, KnowEm, NameCheckr, UserSearch.org, Check Usernames, Instant Username Search	Sites para busca de nomes de usuários	Busca de usuários
Nomes de usuário da Amazon, lista de desejos da Amazon, nome de usuário do Github, nomes de usuário do Tinder.	Sites específicos para busca de usuários em plataformas	Busca específica de usuários
ThatsThem, Hunter, Email do address, Pipi, VolilaNorbert, Reverse genie e-mail, theHarvester, Infoga, MailDB, Skymem, MailsHunt.	Site de busca de informações de e-mail	Informação de e-mail
formato de e-mails corporativos, formato de e-mails, Toofr, permuta de e-mail, OneLook reverse dictionary and Thesaurus.	Site de busca de informação sobre formatos de e-mail	Informação de formato de e-mail
MailTestes, VerifyEmail, Email validator, BytePlant Email validator, Read Notify, Reputação de e-mail, MailboxValidator	Site de verificação e validação de e-mail	Verificação e validação de e-mail
Have I been pwned?, DeHashed, Intelligence X, Vigilante.pw, Breach or clear, Asley Madison Emails.	Site de verificação de e-mail vazado	Verificação se o e-mail foi vazado
Mx Toolbox.	Site de monitoramento de e-mail e verificações de e-mails da caixa de entrada	
Whois records: DNStable, Domain dossier, domainIQ, DomainTools Whois, Domain big data, DNSstuff, Domaincrawler.com, easyWhois, IP2Whois	Verificação de nome de domínio de sites	Verificação de domínio
Aquatone,	Site verificação de	Verificação de

FindSubDomains, Google Subdomains, Xray, DNS recon, Gobuster, Bluto, Pentest-tools.com subdomínios.	subdomínio de sites	subdomínio
Shodan, Kraken, urlscan.io, SameID, AnalyzedID, Redirect detective.	Site descoberta de Ip de dispositivos na rede	Descoberta de dispositivos na rede
Google's certificate transparency, Spyse, Censys	Site de verificação da validade da certificação do site	Verificação de validade de certificado
Security Trails, Mnemonic, histórico de DNS, Deteque.	Site de registros de dados armazenados	
UrlQuery.net, PassiveTotal, FortiGuard Reputation Service, McAfee TrustedSource, Cisco SenderBase, AVG Threat Labs, Malware Domain List, Web Inspector Online Scan, Google Safe Browsing API.	Sites de pesquisa de reputação, e verificação de vírus em geral	Verificação de e-mail com vírus entre outros
VisualPing, Change Detection, Urlwatch, ChangeDetect	Site para detecção de mudança no site	Detecção de mudança em site escolhido
Google Trends, Reddit	Site de busca por termos pesquisados, e site social	Buscar termos pesquisados e acontecimentos em tempo real
Sn1per, Mage Scan, Nmap, Nessus.	Site e ferramentas Scan de vulnerabilidades de sites	Scanear sites e IP's na rede
Burp suite, Shodan, BlackWidow, IntelliTamper, International domain name conversion tool, EyeWitness, Belati, Hunting-new-registered-domain.	Ferramentas para testes em sites	Testar requisições e sites e outros testes
Google imagens, Bing imagens, Yahoo busca de imagens, Yandex imagens, Baidu imagens, Photobucket, PicSearch,	Sites para busca reversa de imagem	Buscar a origem de uma imagem

Panoramio, ImageNet, Places2, Instagram, Flickr		
ExifTool, ExifViewer, ImgOps, GeoSetter	Ferramentas para busca de metadados em imagens	Visualizar metadados de imagens
FotoForensics, Izitru, Stolen camera finder, camera tracer	Site e ferramenta para busca de metadados em imagens	Visualizar metadados de imagem
DetURL, Youtube metadados, Youtube data tools, Hooktube.	Site de busca de metadados em vídeos do YouTube.	Visualizar metadados de vídeos
Scribd, DocJax, Wikileaks busca, GoogleDocs, GoogleDrive, Cryptome, Pastebin Alerts, Pastebin Trends, Pastebin Osint Hasvester	Site de busca de documentos indexados no google sem autenticação do usuário	Busca de documentos indexados sem proteção
busca avançada, busca por localização, busca por data, busca por nome de usuário, primeiro tweet, HootSuite, TweetDeck, Treeverse, teste de e-mail, Back Tweets, Tweet metadata, Twint, Echosec, MIT Map, Harvard Map, Creepy.	Sites e ferramentas para buscas avançadas no twitter	Buscas avançadas no twitter
Recon Tool, ScrapdIn, InSpy.	Ferramenta e Site para busca de dados	Ferramenta e site para busca de dado e informações
Myspace, Tumblr, raven, Delicious, Orkut	Busca em redes sociais	Busca em outras redes sociais
TOR, teste de proxys, VPN anônimas.	Navegador, Proxy, e vpns para anonimato	Navegação anônima
criação de nomes falsos, criação de identidades falsas, geração de usuário randômica, Pexels, Faker.js	Criação de nomes falsos, identidades entre outros	Criação de uma pessoa falsa inexistente
Just delete me, Credit freeze, ferramentas de privacidade, impressão digital de mídias sociais	Criação de nomes de usuário falso, e pessoa que não existe	Criação de nomes idade e endereços falsos
BoardReader, Omglli, Google Groups, Yahoo groups.	Site para buscar mais informações relevantes	Busca por outras informações
Notey, IceRocket, topix, Blog Search Engine	Site de busca sobre empresa	Busca de informações sobre empresa
: Blocktrail,	Site de buscas sobre	Informações compra e

Blockchain.info, BitRef, Wallet Explorer, Blockonomics, Bitcoin Who's Who	informações compra e venda de moedas digitais	venda de moeda
Ether.Camp, etherchain.org, Etherscan.	Busca de informações compra e venda da moeda Ethereum	Informações compra e venda de moeda específica
XMWChain.net, Monero Blocks, Chain Radas	Busca de informações compra e venda da moeda Monero	Informações compra e venda de moeda específica