

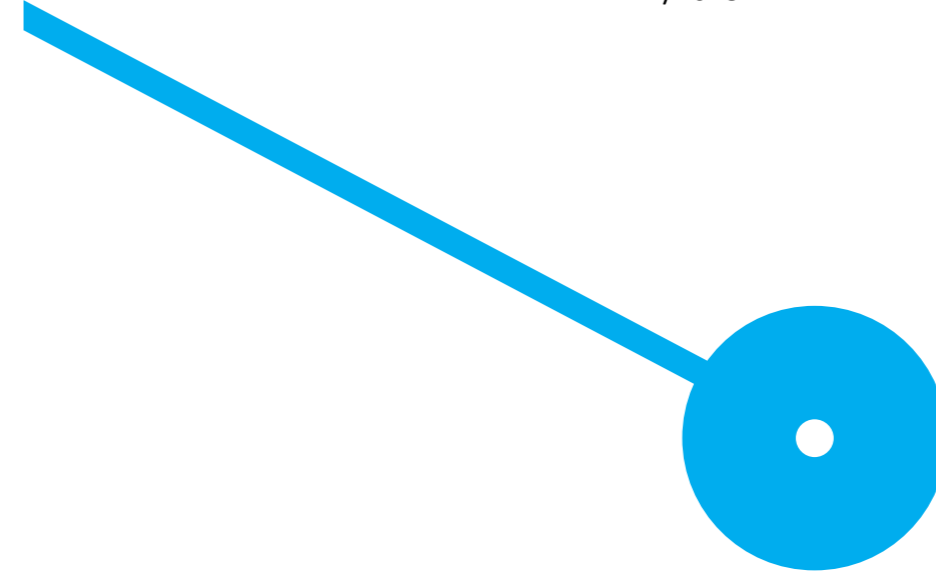
TExtractor: Ferramenta OSINT baseada na
extração e análise de dados áudio/vídeo
António Alberto Marinho Magalhães

4/2018

António Alberto Marinho Magalhães. TExtractor: Ferramenta OSINT baseada na
extração e análise de dados áudio/vídeo

TExtractor: Ferramenta OSINT
baseada na extração e análise de
dados áudio/vídeo
António Alberto Marinho Magalhães

4/2018



TExtractor: Ferramenta OSINT baseada na extração e análise de dados áudio/vídeo

António Alberto Marinho Magalhães
ESTG-IPP
8110244@estg.ipp.pt

Tese apresentada à Escola Superior de Tecnologia e Gestão
para obtenção do grau de Mestre em Engenharia Informática
realizada sob a orientação do
Prof. Doutor João Paulo Magalhães
Professor e investigador na Escola Superior de Tecnologia e Gestão (IPP) 2017

Agradecimentos

A realização deste projeto de mestrado foi um longo caminho, no qual pude contar com apoios e incentivos que de uma ou outra forma facilitaram esta minha caminhada. Ao meu orientador, Professor Doutor João Paulo Magalhães, pela orientação no trabalho, a sua disponibilidade e a pela revisão do documento. A um amigo, Gil Mendes, que esteve sempre pronto a ajudar nas dificuldades com que me deparei no uso da framework Stellar, framework da sua autoria. A todos os amigos e companheiros de caminho, principalmente os que mais diretamente trabalhei, pois sem o seu contributo tudo seria mais difícil. Aos meus familiares, pelo apoio, paciência e compreensão, porque sei que muitas vezes estive ausente. Por ultimo, a todos quantos de alguma forma contribuíram para que atingisse este objetivo.

Resumo

O ciber ataques têm aumentado tanto em número como em sofisticação. Detetar e prevenir a ocorrência desses ataques é um trabalho árduo que requer uma visão holística de segurança envolvendo pessoas, processos e tecnologias. Uma das iniciativas de prevenção e deteção de ciber ataques passa pela definição de um projeto de ciber inteligência. Estes projetos incluem um ciclo de ações e promovem o conhecimento sobre os atores maliciosos, modos de operação e vulnerabilidades em antecipação aos ciber ataques. No âmbito da ciber inteligência uma das referências basilares à construção de um programa é o OSINT (Open Source Intelligence). Trata-se da recolha de informação em fontes abertas que se junta à informação interna e promove um melhor conhecimento sobre as ciber ameaças e formas de mitigação.

Neste trabalho propomos uma ferramenta OSINT baseada na extração e análise de dados áudio/vídeo. Trata-se de uma ferramenta, sobre a forma de uma aplicação Web, capaz de extrair texto a partir de fontes áudio/vídeo e procurar nessa extração informação que possa indiciar o planeamento ou ocorrência de ciber ataques. Na base da ferramenta consta um extrator de áudio e um comparador de texto. O extrator e o comparador foram selecionados após a realização de um estudo experimental, também descrito neste trabalho. O estudo experimental consistiu na análise de vários extratores e comparadores considerando três tipos diferentes de dados de *input* e três idiomas base. Os resultados obtidos no estudo comparativos revelaram uma percentagem de similaridade entre 60% a 70% para ficheiros áudio contendo discursos e quando o idioma base é Inglês. Para *inputs* áudio/vídeo com maiores níveis de ruído e outros idiomas que não o Inglês os resultados são francamente inferiores. Ainda que a percentagem de similaridade não seja elevada, todo e qualquer contributo de automatização para a deteção e prevenção de ciber ataques é relevante e neste âmbito a aplicação tem um papel a desempenhar.

Palavras chave- *Ciber ameaças, Ciber inteligência, Ciber segurança, OSINT*

Abstract

Cyber attacks have increased in both number and sophistication. Detecting and preventing the occurrence of such attacks is a hard work that requires a holistic view of security involving people, processes, and technologies. One of the initiatives to prevent and detect cyber attacks involves the creation and maintenance of a cyber intelligence program. Such programs include a cycle of actions to promote knowledge about malicious actors, modes of operation and the vulnerabilities in anticipation of cyber attacks. One of the basic references to the construction of a such program is OSINT (Open Source Intelligence), i.e. the collection of information from open sources to be combined with internal information promoting a better knowledge about cyber threats and ways of mitigation.

In this work we propose an OSINT tool based on the extraction and analysis of audio/video data. It is a tool, in the form of a Web application, capable of extracting text from audio/video sources and searching for information that may indicate the planning or occurrence of cyber attacks. At the base of the tool, there is an audio extractor and a text comparator. The extractor and the comparator were selected after an experimental study, also described in this work. The experimental study consisted of the analysis of several extractors and comparators considering three different types of input data and three languages. The results obtained in the comparative study revealed a similarity percentage between 60% and 70% for audio files containing speeches in English. The other results are downright lower. Although the percentage of similarity is not high, any contribution of automation to the detection and prevention of cyber attacks is relevant and in this scope the application has a role to play.

Keywords- *Cyber threats, Cyber intelligence, Cyber security, OSINT*

Índice

1	Introdução	1
1.1	Proposta e objetivos do trabalho	4
1.2	Estrutura do documento	5
2	Estado da Arte	7
3	Proposta de trabalho	17
3.1	Ferramentas de extração áudio/vídeo e metodologia para um estudo comparativo	17
3.1.1	Extratores	18
3.1.1.1	Speech To text	18
3.1.1.2	Web Speech API	19
3.1.1.3	Speechlogger	20
3.1.1.4	Speechnotes	20
3.1.2	Comparadores	20
3.1.2.1	Copyleaks	21
3.1.2.2	wDiff	21
3.1.3	Metodologia de análise	23
3.2	Plataforma a implementar: TExtractor	24
4	Estudo comparativo das ferramentas	27
4.1	Ambiente de Testes	27
4.2	Estudo experimental	28
4.3	Apresentação da análise de resultados	28
4.4	Discussão sobre os resultados	33
5	Implementação da plataforma TExtractor	35
5.1	Implementação do TExtractor	36
5.1.1	Mongodb	37
5.1.2	NodeJS	37
5.1.3	Frameworks	38
5.1.3.1	Stellar	38
5.1.3.2	Vue.js	38
5.1.3.3	Vuetify.js	39
5.1.4	Integração das ferramentas	39
5.2	Apresentação da Aplicação	40
5.2.1	Funcionalidades	40
5.2.1.1	Registo e autenticação	40
5.2.1.2	Extrações	42
5.2.1.3	Nova extração	43

5.2.1.4	Análise da extração	44
5.2.2	Conclusão	45
6	Conclusão	47

Lista de Figuras

2.1	Top 5 de ataques com sucesso	9
2.2	Top 5 de agentes mais eficazes	10
2.3	Conhecimento humanos versus Ferramentas disponíveis para ciber ataques	11
2.4	OSINT framework	13
3.1	Texto de origem	21
3.2	Texto extraído	21
3.3	wDiff output example	22
3.4	Imagem elucidativa da plataforma da proposta da plataforma	24
4.1	Resultados comparação copyleaks	29
4.2	Resultados da comparação wDiff	30
4.3	Resultado da tradução copyleaks	32
4.4	Resultado da tradução wDiff	33
5.1	Imagem elucidativa da plataforma	36
5.2	Formulário de registo	41
5.3	Formulário de login	41
5.4	Interface inicial sem extrações	42
5.5	Interface de extrações	43
5.6	Introdução de <i>palavras chave</i>	44
5.7	Documento com as <i>palavras chave</i>	45

Lista de Tabelas

2.1 Ameaças e motivação	8
2.2 Ameaças e motivação	12
3.1 Comparação das Ferramentas Relacionadas	19
3.2 Análise dos Comparadores de ficheiros	23
4.1 Resultados da comparação copyleaks	29
4.2 Resultados da comparação wDiff	30
4.3 Resultados da comparação copyleaks (tradução)	31
4.4 Resultados da comparação wDiff (tradução)	32
6.1 Análise à eficácia da solução final - TExtractor	48

Lista de Siglas e Acrónimos

API	Application Programming Interface
ASR	Automatic Speech Recognition
BSON	Binary JSON
CIA	Central Intelligence Agency
CSS	Cascading Style Sheets
DOM	Document Object Model
DOC	Microsoft Word File (file extension)
FBI	Federal Bureau of Investigation
FLAC	Free Lossless Audio Codec
GEOINT	Geospatial Intelligence
HD	Hard Drive
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HUMINT	HUMAN INTelligence
I/O	Input/Output
ICON	Intelligent Computer Optimised Navigation
IETF	Internet Engineering Task Force
IMINT	Imagery Intelligence
JS	JavaScript
JSON	JavaScript Object Notation
KwS	Keyword Spotting
macOS	computer operating system for Apple
MASINT	Measurement and signature intelligence
MIDI	Musical Instrument Digital Interface
MP3	MPEG-1 Audio Layer-3
MPEG	Moving Picture Experts Group

MVC Model-View-Controller
NPL Natural Language Processing
OCR Optical Character Recognition
OPUS Opus is unmatched for interactive speech and music transmission over the Internet
OSINT Open Source Intelligence
REST Representational State Transfer
RFC Request for Comments
SIEM Security information and event management
SIGINT SIGnal INTelligence
SRT SubRip Text
SVM Support Vector Machine
TIP Threat Intelligence Platform
TCP Transmission Control Protocol
TXT Text file
UI User Interface
URL Uniform Resource Locator
WAV Waveform Audio File Format
WEB World wide web
wDiff Word Difference Finder

Glossário

C4.5: é um algoritmo criado por **Ross Quinlan**, com o objetivo de criar árvores de decisão. É uma extensão de outro algoritmo criado antes, também pelo mesmo autor, o **ID3**. As árvores de decisão criadas pelo **C4.5** podem ser usadas para classificação, por esta razão este algoritmo é muitas vezes apelidado de classificador estatístico [2] e [43].

C: é uma linguagem de programação. **C++:** o mesmo que *C*.

Cyber intelligence: recruta especialistas em segurança *TI* e implementa meios técnicos para proteger a infraestrutura de uma organização ou propriedade intelectual [4].

Cyber Threat Intelligence: é o conhecimento baseado em evidências, incluindo o contexto, mecanismos, indicadores, implicações e conselhos adicionais sobre uma ameaça ou perigo, existente ou emergente, para ativos que podem ser usados para informar decisões [7].

Ciber segurança: é a tentativa de proteção do sistema, de terceiros e de potenciais estragos no seu hardware, software ou informação, assim como o redirecionamento ou corte dos serviços disponibilizados [6].

Deep web: é o conteúdo de bases de dados e outros *web services*, que por uma qualquer razão não podem ser indexados pelos motores de busca convencionais [8].

Hidden Markov Model: è o modelo mais importante de todos os modelos de aprendizagem de máquina na área de processamento de voz. É um modelo ou classificador sequencial, a sua função é associar uma classe a cada unidade da uma sequência [10].

hackers: é um especialista em informática, qualificado, que usa os conhecimentos técnicos para superar problemas. embora possa ser um programador, o termo associou-se a alguém com conhecimentos técnicos que usa *bugs* ou *exploits* para invadir sistemas informáticos [9].

Hacking: é uma tentativa de explorar o sistema de um computador ou de uma rede privada. Ou seja, é o acesso não autorizado ou controle sobre os sistemas de segurança da rede de computadores com um propósito ilícito.

hidden Internet: o mesmo que *Deep web*.

Interface: é o hardware ou software desenhado para passar informação entre dispositivos de hardware, software, dispositivos e software, ou entre dispositivos e o utilizador.

Malware: é a abreviatura de *malicious software*, software malicioso, desenvolvido para obter acesso e/ou danificar o computador sem conhecimento do proprietário [19].

Motor V8: é o motor **JavaScript** de alta performance, código aberto, desenvolvido em **C++**.

Naive Bayes: é um conjunto supervisionado de algoritmos de aprendizagem baseados no teorema de **Bayes** assumindo que os atributos são independentes entre si [15] e [16].

Open Source: designa uma fonte aberta, de acesso publico, que qualquer pessoa pode modificar e partilhar [18].

OSINT: é o conjunto de dados recolhidos a partir de fontes disponíveis publicas para ser usados no contexto do conhecimento [23]. Também pode ser a *frameWork OSINT* que se pode descrever como um conjunto de ferramentas para a obter dados de fontes abertas publicas [24].

Streaming: é a transmissão continua, também conhecida por fluxo de dados, frequentemente utilizada para distribuir conteúdo de multimédia [31].

Threat Intelligence Platform: é uma disciplina emergente que ajuda as organizações a agregar, correlacionar e analisar informação ameaçadora de várias fontes, em tempo real para suportar ações de defesa [32].

Web 2.0: designa a segunda geração de comunidades e serviços disponibilizados pela internet. o termo "2.0" vem da industria de software, onde as novas versões são incrementadas (na parte numérica). Como o software, a nova geração da *web* trouxe novas funcionalidades que não estavam disponíveis no passado [39].

WebSocket: é um protocolo de comunicação de computador, disponibiliza um canal de comunicação full-duplex através de uma única ligação **TCP**. O protocolo WebSocket é um standard definido pela **IETF** no **RFC** 6455 em 2011 e a **API WebSocket** [41].

Capítulo 1

Introdução

Todos os anos Lori Lewis e Chadd Callahan da empresa Cumulus Media fazem a representação gráfica do volume de dados que circula na Internet num minuto. No gráfico de 2017 verificam-se, entre vários itens, que em 60 segundos são visualizados 87 mil horas de conteúdos no Netflix, ouvidas 40 mil horas de musica no Spotify e visualizados 4,1 milhões de vídeos no Youtube [1]. Para a maioria das pessoas a visualização e partilha de conteúdos áudio e vídeo é meramente lúdica. Para as empresas é vista como um canal para promoção e expansão do negócio. Para outros mais ligados a uma temática designada por Inteligência Digital estes conteúdos constituem uma fonte de informação que potencia a análise de dados sobre novas perspectivas. A visualização de imagens para deteção automática de fenómenos estranhos/anómalos, como por exemplo o inicio de um fogo florestal ou um assalto e a análise de sentimentos feita a partir de vídeos e trechos de áudio são exemplos de análises que cada vez mais assumem um papel ativo.

No âmbito da Segurança nacional e internacional, o conceito de Inteligência ou Informações não é novo, mas com o crescimento da dimensão cibernética há a necessidade de endereçar novos desafios que nos são colocados pelo volume, velocidade e variedade criado no mundo digital. A ciber Inteligência ou ciber Informações são encaradas tanto pelas forças armadas como pelas empresas como um ativo a desenvolver por forma a conhecer e antecipar decisões de nível estratégico, tático e operacional. A obtenção da informação pode ser feita a partir de fontes privadas ou de fontes abertas, sendo este último um modelo muito adotado e com uma designação própria: Open Source Intelligence (OSINT). O OSINT define-se por um modelo de inteligência que visa encontrar, selecionar e adquirir informações de fontes públicas (disponíveis publicamente) e analisá-las para que junto com outras fontes possam produzir um conhecimento. O OSINT junta-se a outras fontes de informação muito utilizadas no âmbito militar e que se designam por HUMINT (Inteligência de fontes humana), SIGINT (Inteligência de sinais) e IMINT (Inteligência de imagens). No entanto o OSINT é a fonte que tem vindo a ganhar expressão no âmbito da ciber Inteligência e poderá ser usada tanto para fins de ciber segurança como para o desenvolvimento dos negócios sendo que neste caso a área designa-se normalmente por *business intelligence*. No âmbito deste trabalho explorarmos o OSINT para alargar o âmbito do processo de ciber Inteligência à recolha e análise de informação em fontes de vídeo e áudio.

Antes de passar ao detalhe do projeto apresenta-se o papel do OSINT no âmbito do processo de ciber Inteligência.

São muitas as áreas que utilizam OSINT, como por exemplo a área tecnológica, militar, comunicação e económica. O objetivo é reunir inteligência com sentido através do processamento da informação, utilizando processos como a análise e avaliação. A inteligência obtida a partir do OSINT advém de vários *inputs* de informação, tais como jornais, revistas, áudio e vídeo retirados da Internet. Por este motivo, os dados recolhidos tem, de acordo com [54], de satisfazer dois requisitos. Um é o fato da fonte e conteúdo analisado ter de ser verificado, o outro, a inteligência gerada com a análise e a avaliação da informação tem que ser útil e com sentido para o seu propósito.

O FBI, define o OSINT como toda a informação disponível que pode ser extraída de fontes como jornais, televisão, *websites*, entre outras. Há 20 anos atrás, a falta de informação era uma dificuldade, hoje o desafio é outro, passa por saber que informação consultar ou se realmente vale apenas consultar essa informação. Como consequência o uso de software capaz de recolher e analisar dados em fontes OSINT tornou-se uma atividade estratégica para segurança interna. O recurso a OSINT é muito abrangente. Os governos que tentam adquirir toda e qualquer informação tem os seus próprios processos de uso do OSINT. As pessoas efetuam pesquisas através da Internet para encontrar notícias de mercados, competições, entre outras, estando na prática a adquirir e selecionar informação OSINT. Empresas da área da ciber segurança monitorizam fontes OSINT para antecipar ataques informáticos ou detetar a sua ocorrência no menor espaço de tempo possível. É também importante que cada individuo tenha a noção que, por exemplo, ao comparar ofertas de férias, voos, hotéis, está a dar uso ao OSINT [22] e a alimentar informação que poderá ser usada por terceiros (e.g. publicidade dirigida, períodos de ausência para assaltos, locais que irá frequentar).

Para a CIA, as informações não precisam de ser secretas para serem valiosas. Existe uma oferta infinita de informações nas revistas que se lê, nos blogs ou nas transmissões a que se assiste e esta informação contribui para a compreensão do mundo. De acordo com a CIA [21], a comunidade da área de inteligência refere-se a esta informação como OSINT, desempenhando esta um papel essencial ao dar á comunidade da segurança nacional não só uma visão completa como também um contexto, a um custo relativamente baixo.

Tal como referido anteriormente, o conceito de inteligência, ou seja, recorrer e explorar informação aberta não é novo. O mesmo já era usado antes da era digital. Na prática pode-se dizer que governos e organizações desde sempre usaram esta ferramenta para suportar algumas das suas ações e decisões [44]. Na literatura é referido que o serviços norte-americano Foreign Broadcast Information Service (FBIS) foi pioneiro no OSINT. Iniciou as atividades em finais dos anos 30 e teve como função analisar os noticiários internacionais captados por rádio durante a Guerra Fria e monitorizar publicações oficiais provenientes da União das Repúblicas Socialistas Soviéticas (URSS). Após a guerra fria a área abrandou a atividade, mas com os atentados contra o World Trade Center e o Pentágono em 2001 retomou-se o papel e a importância da utilização das fontes abertas.

No âmbito do OSINT a captura de dados e análise tem sofrido evoluções. Considerando o volume, variedade e velocidade a que a informação é criada o maior desafio passa por consumir essa informação e produzir análises em tempo útil. A área de análise de dados no âmbito do Big Data tem aqui um papel importante. Em [49] o papel da análise dos dados OSINT é referido e descrito um propósito mais recente do OSINT que consiste em sugerir aos utilizadores o desenho das suas próprias inferências sobre o significado e relevância da informação, mapeando-as num sistema de regras de ordem superior desenvolvendo a sua própria solução para problemas únicos. O utilizador não só deve ser capaz de localizar informação baseada no seu problema, como também ver o seu relacionamento com outra informação.

Ainda, e no âmbito do processo de ciber inteligência é de referir que muita da informação relacionada com ciber ataques circula por canais alternativos dificultando o processo de aquisição. A utilização de redes cifradas como é o caso da rede TOR e de websites que garantem o anonimato dos autores é extremamente utilizada pelos atores maliciosos. Estes canais de informação são referidos como **Deep Web** ou *hidden Internet* [20] e requerem mecanismos específicos para a coleta de informação (e.g. ser convidado para participar em fóruns, intervir junto dos canais de comunicação suportado por fundamento legal).

A captura e análise de dados OSINT pode ser usada tanto para detetar situações que já ocorreram, como é o caso de um *data leak* publicado num fórum específico (e.g. pastebin) como para acompanhar as atividades dos atores maliciosos detetando atempadamente o planeamento de ataques dirigidos à organização ou organizações. Este último âmbito enquadra-se num conceito, também apresentado sob a forma de atividade profissional, que se designa por caça às ameaças (threat hunting). A caça às ameaças está focada na procura de ameaças e anomalias dentro das redes e sistemas das organizações, para tal são monitorizados e analisados os *logs*, quer de forma automática como humana. Este conceito assenta na procura de ameaças, numa altura em que ainda não existem sinais específicos de que a segurança foi comprometida. Existem várias áreas a levar em consideração, como a análise de *logs* de autenticação e atividade interna bem como a recolha de dados externa em linha com um contexto. O sucesso da caça depende da análise e a correlação entre eventos e obviamente da experiência do analista. Para trabalhar os dados, são usadas ferramentas e técnicas de análise como as descritas em [58]. Uma das formas de prevenir o ataque é inspecionando e mitigando as vulnerabilidades de segurança com base em informação obtida, por exemplo, através do OSINT. Para suportar a atividade têm sido criadas ferramentas, vulgarmente designadas por TIP (Threat Intelligence Platforms). Algumas destas plataformas agregam dados externos recolhidos em fontes OSINT e dados internos criando uma visão holística sobre ameaças e explorando os indicadores de compromisso existentes, potenciando atividades de mitigação antes que o ataque real ocorra. Estas ferramentas podem também categorizar as ameaças identificadas em quatro tipos. Estes tipos, de acordo com [54], são: estratégico, isto é, as informações de alto nível sobre a alteração do risco; tático, dando indicações sobre as metodologias e ferramentas usadas pelos atacantes; operacional, no qual são dados detalhes específicos de um ataque; técnico, no qual são apresentados por exemplo os indicadores de um *malware* específico.

Nos últimos anos temos assistido a um crescimento no número de soluções de ciber segurança. Mas infelizmente o número e sofisticação dos ataques também tem crescido, tal como revela em [60], evidenciando a complexidade deste jogo em que o “gato” corre permanentemente atrás do “rato”. Os atacantes mudam frequentemente as suas táticas, técnicas e procedimentos dificultando o processo de prevenção proactiva contra os ciber ataques. Melhorar os processos de recolha e análise de dados e alargar o leque de fontes de dados é o que resta para melhor lidar com as ameaças e desenvolver capacidades defensivas que permita uma rápida reação perante a ocorrência de um ataque. Neste trabalho focamos o alargamento das fontes de dados no âmbito OSINT, nomeadamente na análise automática de fontes vídeo e áudio para extração e análise de dados. Na próxima seção são apresentados mais detalhes sobre a proposta de trabalho bem como os objetivos a atingir com o mesmo.

1.1 Proposta e objetivos do trabalho

Os ciberataques têm crescido tanto em volume como em complexidade. Manter os negócios protegidos é um desafio constante que passa pelo desenvolvimento de capacidades tecnológicas, procedimentais e pessoais. No âmbito das capacidades tecnológicas são várias as soluções que ao longo do tempo têm sido adotadas para mitigar os ataques. Firewalls, sistemas anti-spam, sistemas anti-virus, SIEMs são exemplos dessas soluções, porém verifica-se que apesar da existência das mesmas continuam a verificar-se ciber ataques causando impacto económico e de reputação às organizações.

A recolha e análise de dados sobre ataques ocorridos no passado bem como o estudo sobre planeamento de ciberataques são fundamentais para reduzir a probabilidade de ameaça. Dados internos e dados externos correlacionados e associados a um contexto constituem assim uma fonte de informação muito importante. Se os dados internos são já tratados de alguma forma, a recolha de dados externos e o correlacionamento de atividades é ainda insuficiente. A recolha e análise de dados OSINT é vista como um meio de recolha e análise de ciber ameaças e ciber ataques sobre a perspetiva externa. Data leaks publicados em fóruns da área, mensagens trocadas nas redes sociais, nos fóruns e em sites específicos são exemplos de informação que poderão ajudar as organizações a detetar de forma mais rápida a ocorrência de um ciber ataque bem como o planeamento de ataques dirigidos à organização. Neste âmbito existem inúmeras ferramentas que recolhem informação de páginas Web, de documentos PDF e Office e procuram palavras chave ou padrões que sugerem a ocorrência ou planeamento de ciber ataques. Alargar o leque, a profundidade e a capacidade de analisar os eventos é uma área de melhoria continua permitindo às organizações adquirir uma maior inteligência sobre as ciber ameaças, desenvolver capacidades de deteção e proteção e canalizar os seus esforços para os eventos de acordo com uma ordem de criticidade.

Os ciber criminosos tem conhecimento das atividades desenvolvidas na área da ciber defesa e ciber segurança e procuram meios alternativos para contornar as barreiras criadas. Neste âmbito a comunicação oral e vídeo suportada pelas tecnologias de informação funciona como

mais um canal de comunicação entre os ciber criminosos, particularmente na fase de preparação das ameaças, sejam elas no âmbito do terrorismo, ciber terrorismo ou ciber ataques que visam o roubo ou destruição de informação relevante ou afetar da capacidade de organizações.

Neste trabalho propomos uma ferramenta para a recolha e análise de dados áudio e vídeo em fontes OSINT. Considerando a frequência com que se partilham arquivos de áudio/vídeo, a criação de uma ferramenta capaz de recolher e analisar de forma automática este tipo de dados levanta vários desafios. O primeiro desafio prende-se com a forma de análise e neste âmbito iremos explorar a transcrição automática de áudio/vídeo para texto. Depois sobre o texto extraído será feito um motor de pesquisas que visa encontrar referencias a *palavras chave* pré-definidas nesta fase por um analista. Outros desafios se levantam no que diz respeito à precisão das ferramentas de extração de texto a partir de fontes áudio e vídeo. A precisão é no âmbito do projeto um aspeto relevante e por isso será merecedor de um estudo aprofundado, nomeadamente a análise da:

- taxa de acerto da extração de texto a partir de fontes de áudio/vídeo considerando diferentes velocidades de speech e várias ferramentas de comparação entre o original e o resultado;
- taxa de acerto da extração de texto a partir de fontes de áudio/vídeo considerando diferentes idiomas e várias ferramentas de comparação entre o original e o resultado;
- taxa de acerto considerando a tradução automática do texto para uma linguagem base (inglês) e utilizando diferentes tradutores.

Será com base nos resultados obtidos pelo estudo que irá ser proposta a aplicação de recolha e análise de dados em fontes OSINT. Considerando que os ‘hackers’ falam sobre os seus ‘alvos’ antes dos ataques, publicam vídeos, usam fóruns, redes sociais entre outros canais, uma ferramenta como a que se propõe, possibilita a deteção de ataques em fase de preparação. É ainda importante referir que uma aplicação como a que se propõe não é exclusiva da recolha e análise de dados no âmbito da ciber segurança. Uma ferramenta deste tipo poderá ser usada por exemplo para a monitorização de uma marca ou automatizar o processo de clipping, isto é, de identificar referências a marcas ou produtos em fontes de áudio (e.g. rádio) ou vídeo (e.g. televisão ou Youtube).

1.2 Estrutura do documento

Este documento encontra-se dividido em capítulos. No capítulo 2 são analisados alguns trabalhos e ferramentas existentes no âmbito da recolha e análise de dados áudio e vídeo. São também apresentados trabalhos relacionados com a área da ciber segurança com especial destaque para o OSINT. No terceiro capítulo, é apresentada a proposta de trabalho que engloba uma metodologia de análise à eficiência das ferramentas de extração automática de texto a partir de diferentes fontes áudio/vídeo e os requisitos para o desenvolvimento da ferramenta. No capítulo 4 são

apresentados os resultados da eficiência das ferramentas de extração e análise de dados. A apresentação da solução implementada neste projeto é apresentada no quinto capítulo. Este capítulo apresenta as funcionalidades da aplicação desenvolvida, assim como, as ferramentas utilizadas para o seu desenvolvimento. O sexto e último capítulo, contém as conclusões deste trabalho e indica linhas de trabalho futuro.

Capítulo 2

Estado da Arte

A evolução digital ocorre a passos largos e abrange organizações e indivíduos que dependem dessa mesma evolução para melhorar os seus negócios e qualidade de vida. Infelizmente a evolução também se faz sentir em coisas menos positivas. A exposição a ciber ameaças é uma dessas coisas. Existem vários tipos de ameaças, as mais preocupantes no âmbito deste trabalho são as falhas técnicas e as que comprometam a informação das organizações. Estas ameaças podem ter várias origens, como acidental, ambiental, deliberada ou negligente, podem vir de um grupo ou apenas de um indivíduo. Tipicamente as ações destes agentes passam por obter acesso e usar ativos das organizações sem que estejam autorizados a tal, assim como divulgar informação confidencial ilicitamente e fazer alterações a recursos sem autorização. Os ataques informáticos, podem ser vistos como um negócio, isto porque se o ataque surtir efeito, o resultado vale dinheiro. Pedir resgate em troca da libertação de informação (ransomware), vender contas de email, vender números de cartões de crédito e vender registos de saúde, entre outros, são alguns dos exemplos referido pelo autores em [56]. A conjugação entre o resultado das ações maliciosas (ciber ataques) com a facilidade de efetuar ataques incrementa o nível de ameaça assim como as tentativas de ataque. Esta realidade reforça a necessidade de preparar, construir e incrementar a defesa, estar sempre alerta porque a realidade atual assim o exige.

Atualmente, a maior parte das equipas de segurança informática tem de lidar com ataques uma vez que estão a aumentar significativamente. Para acabar, mitigar ou evitar estes ataques as equipas de segurança informática têm que pensar como os atacantes e perceber como os ataques se desenrolam com o objetivo de os conter numa fase inicial. De acordo com a *cyber kill chain* proposta por Lockheed Martin existem sete passos envolvidos num ataque típico. Em primeiro lugar é feito um *reconhecimento* dos alvos, que pode ser feito através de uma pesquisa nas redes sociais. Após encontrar os alvos, o atacante usa um canal de comunicação para construir o ataque distribuindo uma *ferramenta de acesso remoto* através de código dissimulado. Ao ser *entregue*, este alvo passa também a ser uma ponte entre os seus contactos, fazendo dele próprio um ponto de distribuição para o ataque. A fase seguinte é a *exploração* e poderá ocorrer por exemplo através do click num link levando à *instalação* das ferramentas de acesso remoto e ao controlo do alvo por parte do atacante. A partir deste ponto, o atacante tem a possibilidade

de recolher ou destruir informação pertencente à organização (ou seus clientes) [56][5]. Muito devido às estruturas das organizações atuais, quando um atacante é detetado, se acontecer, já será durante o sexto ou sétimo passo da *cyber kill chain*, o que já é demasiado tarde, uma vez que o atacante já terá acesso ao sistema e possivelmente alguma informação sensível do seu lado.

As motivações que levam a estes ataques são variadas. As que tem tido mais expressão são as relacionadas com o ganho financeiro, mas existem outras, tal como apresentado na Tabela 2.1.

Ameaças e motivação		
Atacante	Motivos	Alvos
<ul style="list-style-type: none"> • Estado 	<ul style="list-style-type: none"> • Vantagens económicas • Vantagens políticas • Vantagens militares 	<ul style="list-style-type: none"> • Segredos comerciais • Informação sensível • Infraestruturas críticas
<ul style="list-style-type: none"> • Crime organizado 	<ul style="list-style-type: none"> • Lucro imediato • Recolha de informação para venda 	<ul style="list-style-type: none"> • Sistemas de pagamento • Informação pessoal • Informação de saúde • Informação de cartão de crédito
<ul style="list-style-type: none"> • Ativistas 	<ul style="list-style-type: none"> • Influencia política • Alteração social • Pressionar mudança de negócio 	<ul style="list-style-type: none"> • Informação de negócio sensível • Informação de pessoas chave • Informação de parceiros de negócio
<ul style="list-style-type: none"> • Colaboradores 	<ul style="list-style-type: none"> • Vantagem pessoal • Vingança profissional • Patriotismo 	<ul style="list-style-type: none"> • Vendas, negócio e estratégia de mercado • Operações de negócio • Informação pessoal

TABELA 2.1: Ameaças e motivação

Na Figura 2.1 é possível verificar os cinco tipos de ataques com mais sucesso. Os dados foram obtidos em [56]. Parte deste sucesso deve-se à falta consciencialização das organizações e

respetivos colaboradores, assim como à falta ferramentas e procedimentos adequados às ciber ameaças.

O que muitos dos especialistas e empresas já perceberam e argumentam é que a abordagem tradicional já não é suficientemente eficaz para lidar com a escalada das ciber ameaças e travar os ciber ataques da atualidade. Os atacantes trabalham de forma mais sofisticada do que as organizações alvo dos seus ataques, revendo frequentemente as suas ferramentas, técnicas e procedimentos de ataque evitando dessa forma padrões de ataque. Outro desafio tem a ver com a sofisticação das ferramentas. A evolução das ferramentas de ataque é mais rápida e avançada, do que as ferramentas usadas pelas organizações para se protegerem de possíveis ataques.

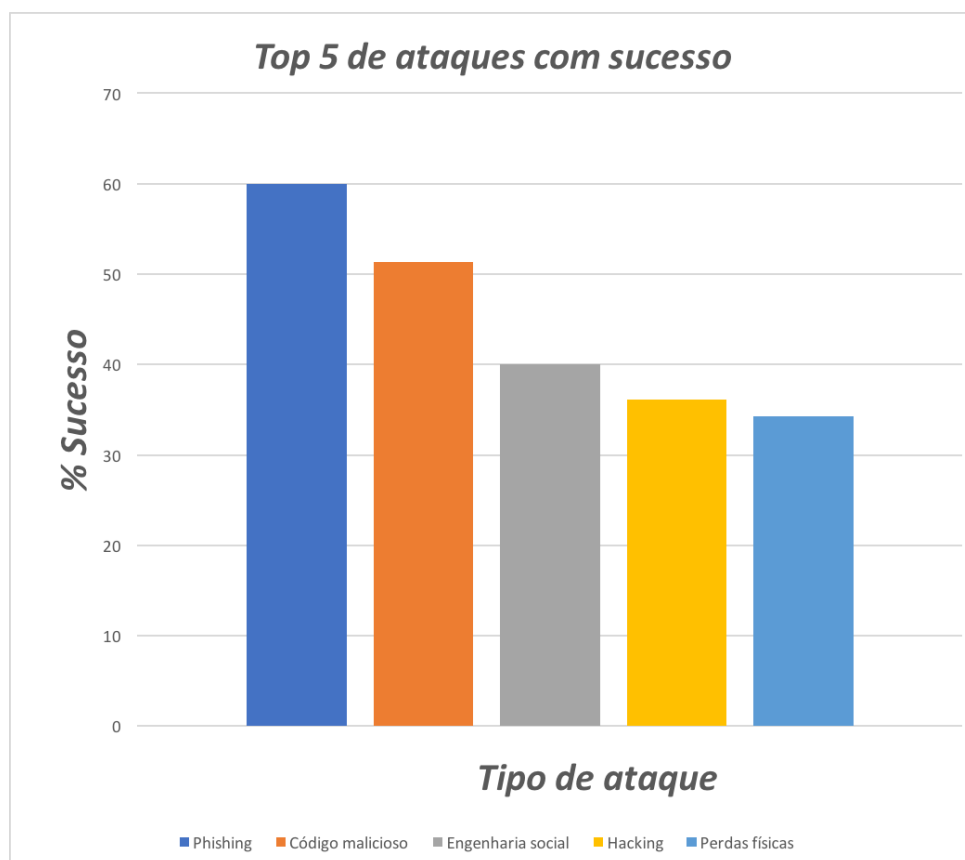


FIGURA 2.1: Top 5 de ataques com sucesso

Tal como as ameaças, os atores maliciosos podem ser de vários tipos, podem ser empregados, organizações criminosas, empresas concorrentes, causas naturais e/ou humanas (intencionais ou não). As ameaças têm as suas causas e motivações. As causas refletem geralmente a insatisfação ou a procura de lucro. Um trabalhador insatisfeito, uma organização criminosa ou mesmo uma empresa concorrente, podem ver uma oportunidade de lucro ao explorar uma falha. São também muitas as motivações dependendo do âmbito da ameaça, como por exemplo, o ganho financeiro, a vantagem competitiva, política ou económica, o domínio militar ou apenas para alimentar o ego, mostrar que é capaz. Na Tabela 2.1 já foi feita uma apresentação dos atores e das motivações. A Figura 2.2 por sua vez representa os cinco principais agentes de ameaça (atores maliciosos),

e o seu peso nos ataques efetuados. Como podemos verificar pelo gráfico, os ataques com maior percentagem de sucesso são os efetuados propositadamente, que são lançados por um qualquer agente. Os valores apresentados resultam do estudo efetuado em [56].

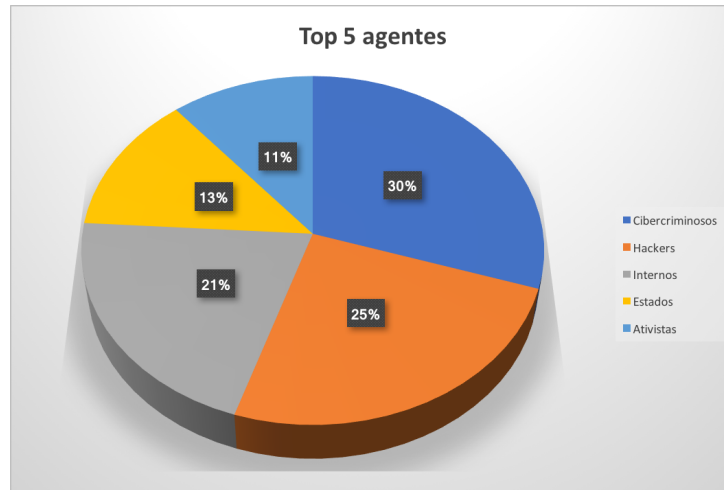


FIGURA 2.2: Top 5 de agentes mais eficazes

Algo que sofreu uma evolução tremenda na área da ciber segurança tem a ver com a quantidade e sofisticação das ferramentas. As ferramentas ao dispor podem ser usadas tanto por agentes do bem que de forma proactiva auditam os seus sistemas e aplicações, ou pelos atores criminosos aumentando o seu alcance e rapidez no ataque. Como referido pelo autor [56], em 1980 era necessária mais formação por parte do atacante do que atualmente, isto porque, com a evolução e disponibilidade de ferramentas de ataque na Internet, qualquer pessoa, com pouca formação na área pode despoletar um ataque. Como podemos verificar na Figura 2.3, com o avançar do tempo, a necessidade de conhecimento por parte do atacante é proporcional ao crescimento da qualidade das ferramentas desenvolvidas para tal, ou seja, cada vez se desenvolvem mais ferramentas e mais sofisticadas, diminuindo a necessidade de conhecimento do atacante. Tal situação faz com que o número de atores maliciosos seja maior.

Num contexto em que as ciber ameaças tem crescido, existe a necessidade de uma preparação para retaliar tal tendência. A ciber inteligência, ou seja, o conhecimento útil e que permite atuar sobre as ciber ameaças, é uma das abordagens para ajudar esta retaliação. Não existe um consenso sobre a definição de inteligência no contexto de conhecimento digital, mas o seu núcleo é a informação que precisa de ser recolhida, correlacionadas produzindo um conhecimento mais profundo sobre os atores maliciosos, o seu modo de atuação, os riscos existentes e forma de os mitigar.

A informação só por si não tem valor significativo, precisa ser tratada e selecionada, outro requisito para se poder estar preparado para uma eventual ameaça é a aquisição de conhecimento,

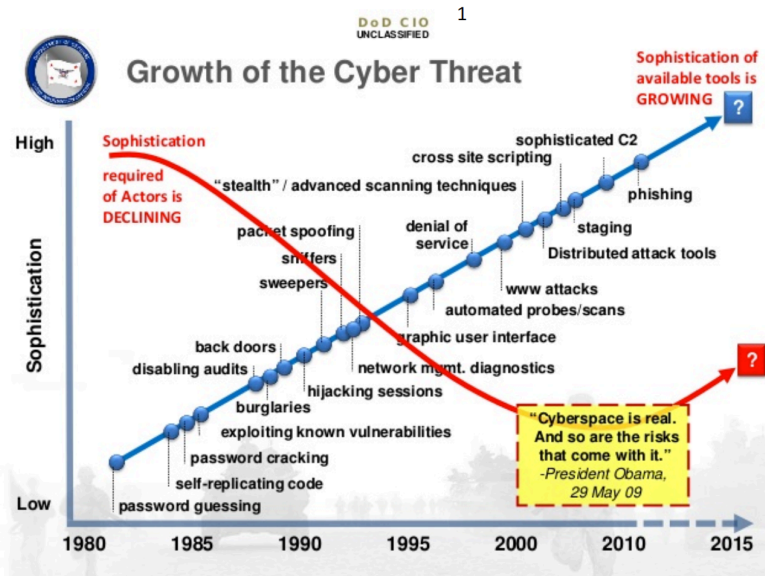


FIGURA 2.3: Conhecimento humanos versus Ferramentas disponíveis para ciber ataques

a *ciber inteligência*, ou seja, recolha e tratamento de informação. Existem três tipos de conhecimento, o conhecimento estratégico, de operações e o tácito. O conhecimento estratégico diz respeito às tendências das ameaças ou potenciais ameaças, a longo prazo, pelo seu lado o conhecimento tácito avalia as capacidades da ameaça focando-se nos pontos fracos, fortes e nas reais intenções da ameaça. O conhecimento de operações, é o conhecimento em tempo real ou perto disso, deriva dos meios técnicos.

Para a produção de inteligência/informação, o próprio conhecimento tem o seu ciclo de vida, como se descreve na Tabela 2.2.

A informação, dentro do ciclo de vida do conhecimento, está dividida em seis áreas e pode ser recolhida em:

- Sistemas eletrónicos, geralmente de forma fraudulenta (SIGINT).
- Fontes abertas como o rádio, jornais e Internet (OSINT).
- Humanas de forma consciente ou involuntária (HUMINT).
- Imagens (IMINT).
- Satélites, drones e outras fontes que acompanham a atividade relacionada à segurança em torno do planeta. Muitas vezes associado com IMINT (GEOINT).
- Fontes que não encaixam na área SIGINT ou IMINT, ex.: radiofrequência (MASINT).

Um exemplo de ferramenta que se enquadra na área da recolha e análise de informação é a **watson-tone-analyzer**. Esta ferramenta analisa o texto numa conversação, permitindo detetar

Ciclo do Conhecimento	
<ul style="list-style-type: none"> • Planeamento 	<ul style="list-style-type: none"> • Determinar requisitos de conhecimento
<ul style="list-style-type: none"> • Aquisição 	<ul style="list-style-type: none"> • Planear a aquisição de informação de várias fontes e através de vários métodos • Procurar conhecimento de outras agências • Reunir toda a informação possível
<ul style="list-style-type: none"> • Processamento 	<ul style="list-style-type: none"> • Traduzir, transcrever e descriptar informação • Avaliar a confiança da informação • Se necessário converter a informação para um formato legível
<ul style="list-style-type: none"> • Análise 	<ul style="list-style-type: none"> • Combinar informações diferentes para identificar informação colateral e padrões • Interpretar o significado de qualquer conhecimento desenvolvido
<ul style="list-style-type: none"> • Tomada de decisão • Divulgação 	<ul style="list-style-type: none"> • O conhecimento acabado toma muitas formas, dependendo das necessidades do decisor e os requisitos de comunicação • Os vários níveis de urgência são geralmente estabelecidos pela organização ou comunidade do conhecimento
<ul style="list-style-type: none"> • Feedback 	<ul style="list-style-type: none"> • O feedback é recebido do decisor e os requisitos revistos

TABELA 2.2: Ameaças e motivação

sentimentos a partir do *input*. Já é disponibilizado pela ferramenta alguns tipos de tons que representam outros tantos sentimentos, como por exemplo a frustração, tristeza, satisfação, excitação, educação, má educação e simpatia. A ferramenta disponibiliza um serviço que deteta os tons e referentes sentimentos acima referidos em ambos os participantes de uma conversação. A ferramenta foi treinada em conversações mantidas no **twitter** [37].

Uma ferramenta que se baseia na recolha de informação para criação de conhecimento é a *framework* OSINT (Figura 2.4). O foco principal da *framework* é a recolha de informação através

de ferramentas livres e fontes publicas com a intenção de ajudar as pessoas a encontrar estes recursos(informação) de forma grátis. Esta ferramenta tem um leque bastante alargado de funcionalidades, como podemos verificar na figura 2.4. Ao clicar nos itens com os pontos preenchidos a cor estes estendem funcionalidades relacionadas com o item. A maioria das funcionalidades são disponibilizadas de forma gratuita, mas também disponibiliza funcionalidades desenvolvidas por terceiros, que por sua vez tem algumas pagas. A ideia da *framework* é recolher e disponibilizar informação de fontes abertas e de forma gratuita.

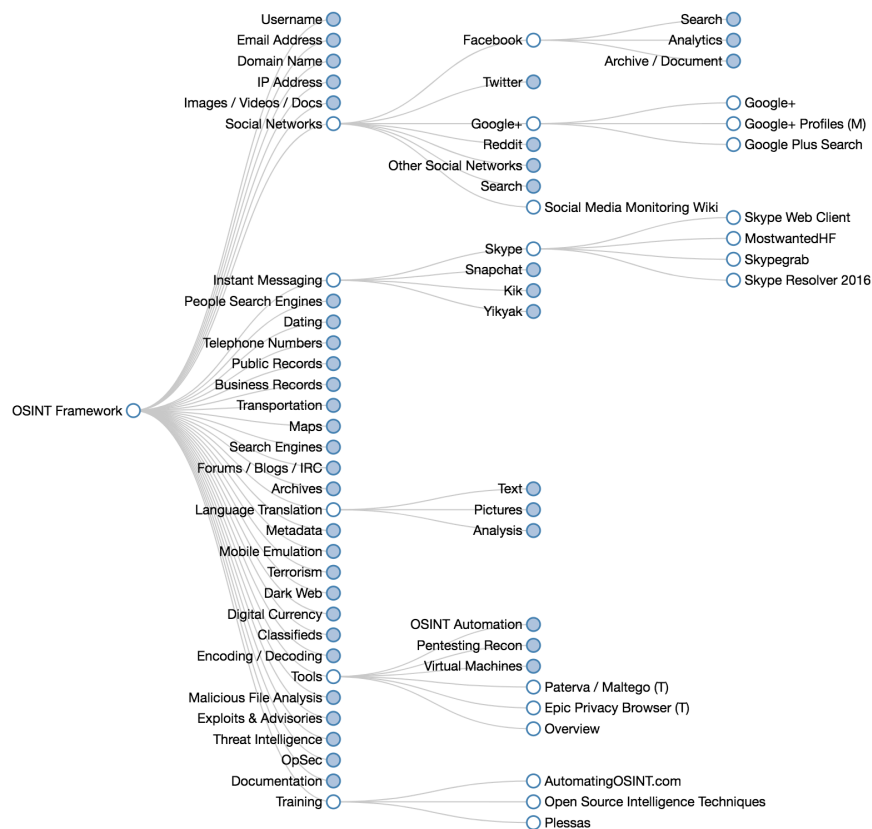


FIGURA 2.4: OSINT framework

A recolha de informação de fontes abertas, onde também se englobam arquivos de áudio/vídeo, certamente é uma mais valia para o OSINT. A recolha deste tipo de dados vem disponibilizar informação em quantidade e possivelmente qualidade, bastando para tal a sua recolha e tratamento. Considerando o propósito do projeto, isto é, a criação de uma ferramenta para análise automática de áudio/vídeo em fontes abertas, centraremos a discussão do estado da arte na extração e análise deste tipo de dados.

Atualmente é partilhado conteúdo áudio e vídeo em enormes quantidades, como tal, é uma fonte de informação apetecível. Existem várias técnicas de análise deste tipo de informação. Sistemas capazes de reconhecer automaticamente objetos em imagens, de modelar a expressão labial de uma pessoa a um *input* áudio são exemplos do que é feito na área. Uma forma simples de analisar o áudio/vídeo consiste na extração de texto a partir do mesmo. Existem alguns trabalhos na

área do reconhecimento e extração de texto a partir de fontes de áudio e vídeo. É uma área enorme, varia desde a classificação de conteúdo de vídeo [51], ao suporte médico [63], detecção de sentimentos [53] e detecção de violência [48], entre outros.

A comunicação é muito importante para a satisfação e o desenvolvimento do ser humano. Um facto, é que o nosso modo de comunicar está a mudar e cada vez mais e existe a necessidade de interagir de forma rápida. A *Web 2.0* disponibiliza essa facilidade, desta forma, as pessoas incrementaram significativamente a sua interação *online* de forma tão frequente que se tornou uma rotina diária para quase toda a gente. Entre várias coisas, são partilhados conteúdos de áudio e vídeo, os quais, só por si, disponibilizam muita informação. O trabalho apresentado em [51] classifica texto baseado em conteúdo de vídeo partilhado em ‘sites’ (**Classifying text-based video content for online vídeo sharing sites**). A *framework* desenvolvida neste trabalho conta com funcionalidades baseadas em técnicas de classificação (**C4.5**, **Naive Bayes** e **SVM**) e em três tipos extração de texto (lexical, sintático e conteúdo). De acordo com [51] a classificação além de fornecer uma forma de classificar vídeos conforme os interesse dos utilizadores com uma precisão de 87,2 %, também é muito útil para identificar comunidades online envolvidas na partilha de vídeos .

Naturalmente, também na área da saúde já existem trabalhos do género, um exemplo a interface **MammoClass** [63]. Nesta área existe a necessidade de trabalhar rapidamente e toda a ajuda automatizada é importante. Trata-se de uma ferramenta **Web** que permite a entrada de um conjunto de variáveis referentes aos resultados de uma mamografia, com as quais verifica a probabilidade de um achado benigno ou maligno. Esta ferramenta obriga a que o radiologista introduza um valor para cada variável da ferramenta para obter a probabilidade do resultado da mamografia, a interface desenvolvida neste trabalho, permite que o radiologista faça um relatório falado em detrimento de um escrito. Após esta ação, a interface **MammoClass** extrai os valores das variáveis e usa uma ferramenta de análise para reconhecer o texto. Os resultados dos relatórios relatados usando a interface **MammoClass** mostram que as mesmas variáveis são extraídas para ambos os tipos de entrada: ditado ou texto digitado.

A análise de sentimentos através do reconhecimento automático de áudio é uma área de investigação emergente onde os sentimentos exibidos são detetados através de áudio. É uma área menos explorada comparando com a detecção de sentimentos através de texto. Há trabalhos nesta área que usam transcrições de sistemas de reconhecimento de fala e classificadores de sentimentos baseados em texto para processar a essas transcrições. Em [53] é apresentado um trabalho que propõe uma nova arquitetura, que conjuga o **KwS** (*keyword spotting*) e um classificador de sentimentos baseado em texto para determinar automaticamente palavras que representem sentimentos comportamentais, que posteriormente serão usadas para atualizar/expandir a lista de termos do **KwS**. É desenvolvido um sistema híbrido que utiliza um sistema automático de reconhecimento de voz (**ASR**) e um sistema de processamento de linguagem natural (**NPL**) baseado em técnicas de análise de sentimentos para detetar sentimentos em áudio. É também definido um modelo máximo de sensação de entropia para reduzir a lista de *palavras-chave* enquanto o modelo permanece efetivo. As fontes de dados usadas para testes foram o **Youtube.com**.

Na *Internet* são compartilhados muitos vídeos, com diferentes temas e objetivos, onde o teor do seu conteúdo também difere igualmente. A verdade é que parte desses vídeos compartilhados mostram muita violência e é importante que possam ser filtrados. Em [48] é apresentada uma abordagem para detetar conteúdo violento nos vídeos compartilhados. Esta abordagem usa uma fusão de três modos: áudio, imagem em movimento e dados de texto. O problema é tratado como uma classificação binária - conteúdo violento ou não violento. O conteúdo de áudio é testado pelas frequências naturais de arquivos de áudio frequentemente encontradas em fluxos de vídeo, divididas por classes. No que diz respeito aos dados de texto, foi elaborado um dicionário de 500 palavras-chave composto por palavras frequentemente relacionadas com o racismo, abuso de drogas, palavrões, sexismo e crime em geral. A classe é fornecida pelo número total de palavras dividido pela soma de todas as ocorrências das palavras-chave. De acordo com os resultados apresentados a precisão geral do sistema atingiu 82%, usando dados reais de fluxos de vídeo do **Youtube**.

O sistema **Android** também está presente nesta área. Em [62] é apresentado um trabalho que faz a conversão de voz para texto usando a plataforma **Android**. Com processos, algoritmos e métodos modernos é possível reconhecer o texto e processar sinais de voz mais facilmente. Este sistema recebe a voz a processar através do microfone do dispositivo, o texto obtido pode ser armazenado num ficheiro. Existe um leque alargado de opções de entrada de dados, por exemplo, o sistema pode disponibilizar a utilização a pessoas com deficiência visual, auditiva ou mesmo fisicamente incapacitadas. A plataforma disponibiliza a oportunidade de criar uma mensagem falada, que converte para texto, criando assim uma forma de possibilitar a pessoas com deficiência audiovisual poder enviar uma mensagem de texto. Este sistema está adaptado ao idioma Inglês e tira partido do **Hidden Markov Model** para construir o texto a partir do áudio melhorando a sequência das palavras e frases.

Existem também trabalhos na variante de extração de texto a partir de fontes de vídeo, estes trabalhos embora com o mesmo objetivo, a extração de texto, não comungam do mesmo princípio uma vez que a extração é feita da imagem e não do áudio do vídeo. Para o reconhecimento do texto são usadas técnicas de **OCR**, estas técnicas disponibilizam formas de detetar texto em imagens. O texto extraído é depois usado para ajudar a catalogar e procurar vídeos [61] [66].

Capítulo 3

Proposta de trabalho

Este capítulo está dividido em duas fases que se complementam. A primeira fase consiste na seleção de ferramentas capazes de extrair texto a partir de fontes áudio/vídeo e ferramentas que permitam a análise da extração para encontrar as palavras chave definidas. A análise é fundamental para medir a taxa de acerto na extração do texto e ao longo deste capítulo será apresentada a metodologia de análise a realizar. A segunda fase do trabalho consiste na apresentação da plataforma a implementar e que terá como *input* as ferramentas analisadas na primeira fase.

3.1 Ferramentas de extração áudio/vídeo e metodologia para um estudo comparativo

Como já referido, para a implementação da plataforma, existe a necessidade de análise das ferramentas. Assim, ao longo desta secção é feita uma descrição geral de ferramentas de extração de texto a partir de fontes áudio/vídeo e ainda de ferramentas de análise do texto extraído para um estudo comparativo.

A apresentação das ferramentas está dividida em dois grupos. O primeiro grupo designa-se por “Extratores” e diz respeito a ferramentas que permitem a extração automática de texto a partir de fontes áudio/vídeo. O segundo grupo designa-se por “Comparadores” e corresponde a um conjunto de ferramentas que permite a comparação entre um texto original ou um conjunto de palavras chave e o resultado da extração de texto obtida pelos “Extratores”. Os “Extratores” e os “Comparadores” foram encontrados através de pesquisa na literatura e na Web. Foram encontradas quatro ferramentas de extração e duas de comparação de texto, que são enumeradas e descritas a seguir.

3.1.1 Extratores

Os “Extratores” são ferramentas que permitem, dado um input de áudio/vídeo traduzir o seu conteúdo para texto. Da pesquisa efetuada foram encontradas quatro ferramentas principais:

- **Speech to Text** - <https://speech-to-text-demo.mybluemix.net>
- **Web Speech API Demo** - <https://www.google.com/intl/en/chrome/demos/speech.html>
- **Speechlogger** - <https://speechlogger.appspot.com/en/>
- **Speechnotes** - <https://speechnotes.cc>

Todas estas ferramentas, possibilitam a extração de texto a partir de fontes de áudio/vídeo, as quatro são bastante idênticas no objetivo (a extração de texto), pese, no entanto, o facto de diferirem nas funcionalidades que disponibilizam. Na tabela 3.1 são identificadas as funcionalidades disponibilizadas por cada uma delas.

3.1.1.1 Speech To text

A ferramenta Speech To Text [26] ou dito de outra forma, o serviço IBM Watson Speech to Text usa a capacidade de reconhecimento de um discurso (áudio/vídeo) para a sua conversão para texto. O serviço tem suporte para vários idiomas, árabe, espanhol, francês, português, japonês e mandarim.

Recebe o áudio a transcrever através do microfone ou através do *upload* de um ficheiro de áudio. É permitida a entrada de áudio em alta ou baixa frequência, assim com a introdução de raiz de um conjunto de *palavras chave* para procura no ficheiro transcrito. O resultado da transcrição conjuga o texto extraído, as palavras chave encontradas e o número de palavras que não foi capaz de extrair automaticamente apresentando possíveis alternativas.

É possível aceder às funcionalidades deste serviço através de três formas diferentes, uma interface **WebSockets** ou **HTTP**, podendo nesta ultima opção uma interface **REST** ou de forma assíncrona. Também é possível personalizar o idioma de forma a ajusta-lo ao domínio e ambiente. No trabalho [46] foi usado este serviço para ajudar a avaliar a inteligibilidade da fala em pessoas com a doença de Parkinson, como estas pessoas tem dificuldade em falar devido à reduzida coordenação dos músculos que controlam a respiração, fonação, articulação e prosódia. O principal propósito é o de ajudar a avaliação profissional desta dificuldade. Outro trabalho em que esta ferramenta foi de muita importância, foi no desenvolvimento de uma plataforma para facilitar a comunicação entre pessoas de audição normal e audição diminuída, ou mesmo em que ambas tem audição diminuída [64]. O serviço Speech To Text é usado em muitos outros trabalhos, principalmente no que a traduções diz respeito, desde a tradução multilingue, á tradução para possibilitar a interação e conversação de pessoas de culturas diferentes, este serviço é bastante usado com podemos verificar em [65] [67] [47] [45] [14].

Comparação de extratores					
Item	Extrator	Speech to Text	Web speech API	Speech Logger	Speech Notes
API		✓	✓		
Suporta frequências diferentes		✓			
Exporta para ficheiro.txt (TXT)				✓	✓
Exporta para ficheiro.doc (DOC)				✓	✓
Exporta legendas (SRT)				✓	
Exporta para Google translate (DOC)				✓	
Encontra palavras chave		✓			
Inclui tradução				✓	
Multilingue		✓	✓	✓	✓
Importa ficheiro de HD		✓		✓	
SWebSocket traffic		✓			
Envio de email			✓	✓	✓
Imprime				✓	✓
Grava ficheiro para HD				✓	✓
Partilha no Google+		✓			
Partilha no Twitter		✓			
Registo horário da tradução				✓	
Usa microfone para entrada de audio		✓	✓	✓	✓
Upload para o Google drive				✓	✓
Conta palavras		✓			
Formatos de ficheiros suportados					
(FLAC)		✓			
(OPUS)		✓			
(WAV)		✓			
(MP3)		✓			
(MPEG)		✓			

TABELA 3.1: Comparação das Ferramentas Relacionadas

3.1.1.2 Web Speech API

A Web Speech API [40] é disponibilizada pela Google e requer a utilização do Google Chrome, versão 25 ou mais recente. Apesar de ser muito referenciada na literatura a utilização da ferramenta está limitada pela própria Google. A versão grátis da API é limitada a 50 pedidos por dia. O texto é extraído a partir do áudio que é recolhido através do microfone. A API foi projetada para permitir a entrada de discurso de forma breve ou de forma contínua, os resultados fornecidos são uma lista de hipóteses de transcrição, juntamente com outras informações relevantes. A possibilidade de combinação gramatical e o uso de um conjunto gramatical alargado evita a necessidade de um segundo reconhecimento do áudio para posterior extração, esta API suporta ambas as especificações.

3.1.1.3 Speechlogger

O **Speechlogger** [27] é uma aplicação **WEB** de reconhecimento e tradução instantânea de voz. Esta aplicação usa as tecnologias de extração de texto a partir de fontes de áudio da **Google**, é completamente livre e não precisa de registo. A ferramenta possibilita também a edição de texto e disponibiliza a pontuação automática. Uma vez que usa a tecnologia de reconhecimento da **Google**, esta aplicação à semelhança da Web Speech necessita do **Google Chrome** versão 25 ou superior. Esta ferramenta, além de extrair texto de áudio, permite a sua tradução instantânea, que não só apresenta em formato de texto como também a transcreve para áudio. O **Speechlogger** é uma ferramenta de grande utilidade, devido à sua capacidade de extração de texto a partir de fontes áudio, permite criar legendas em filmes no idioma desejado, assim como trabalhar como um tradutor. Esta capacidade de extração e tradução instantânea é uma mais valia não só para a aprendizagem de um idioma estrangeiro, como também para servir de interprete entre duas ou mais pessoas que utilizem idiomas diferentes para manter uma conversa. Uma grande utilidade, é o ato de permitir a utilização de telefones a pessoas surdas, bastando para tal direcionar o áudio do telefone para a aplicação **Speechlogger**.

3.1.1.4 Speechnotes

A SpeechNotes [28] é um bloco de notas *online* habilitado a receber áudio como entrada. A ferramenta faz uso de tecnologias de reconhecimento de voz de forma a disponibilizar maior produtividade e conforto ao utilizador. Disponibiliza três métodos para efetuar a pontuação, por fala, usando o teclado e clicando num dos items disponibilizados na lista de pontuação. Após a extração é dada a possibilidade de confirmar o texto ditado, através de uma ferramenta de leitura embebida no extrator. Esta ferramenta, além de extrair texto de áudio, também possibilita a criação de áudio a partir de texto. As funcionalidades testadas foram apenas as disponibilizadas de forma gratuita, sendo a ferramenta mais completa na versão paga.

3.1.2 Comparadores

Após a extração do texto a partir do áudio/vídeo é necessário comparar texto para identificar as ocorrências das palavras chave a pesquisar. Para o efeito existem várias ferramentas que facultam a comparação entre textos, no entanto a pesquisa efetuada recaiu sobre ferramentas capazes de comparar largas quantidades de texto apresentado de forma simplificada os resultados obtidos para a comparação. Foram identificadas duas ferramentas muito referenciadas na área:

- **Copyleaks** - <https://copyleaks.com/compare>
- **wDiff** - <https://www.gnu.org/software/wdiff/>

Estas duas ferramentas são descritas nas próximas subsecções.

3.1.2.1 Copyleaks

O *copyleaks* [3] apresenta o resultado na forma de duas caixas de texto. Em cada caixa figura um texto, o ficheiro de origem e o que se quer comparar, ambos os textos tem as palavras encontradas marcadas a cor diferente. Esta ferramenta procura semelhanças no texto, não palavra a palavra, mas por blocos de palavras iguais em ambos os textos. Para ser reportada semelhança nos textos tem que existir blocos não só com as mesmas palavras como também tem que estar na mesma posição. Nas figuras 3.1 e 3.2 são ilustradas as duas caixas nas quais se vê o texto que está marcado a amarelo, ou seja o texto semelhante nos dois arquivos e cuja a ordem se mantém entre os textos.

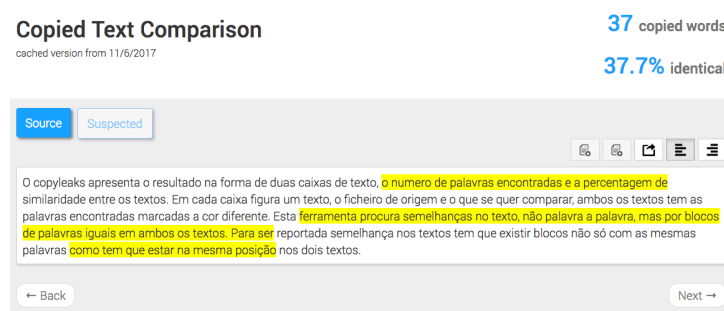


FIGURA 3.1: Texto de origem

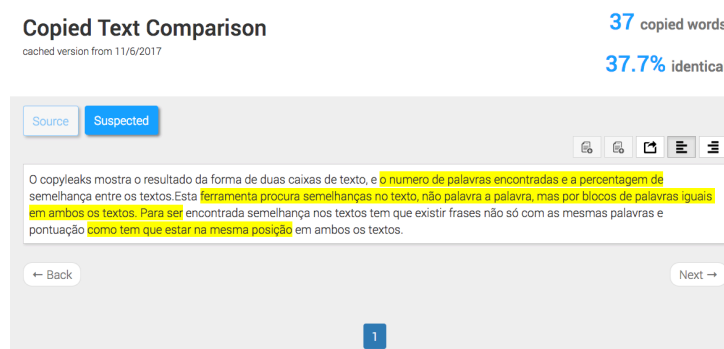


FIGURA 3.2: Texto extraído

3.1.2.2 wDiff

O *wDiff* (GNU *wDiff*) 1.2.2 [38], é uma ferramenta local, isto é, executa na própria máquina. A comparação é baseada na palavra e o resultado é apresentado em formato de texto. Durante o processo de comparação a ferramenta cria dois ficheiros temporários e executa a comparação de cada palavra de um por cada linha de outro. O resultado é um novo texto onde figuram as alterações referentes ao texto origem. Esta ferramenta pode ser usada diretamente a partir da linha de comandos, e dispõe de opções para alterar o *output* do resultado. Uma dessas opções é

-s, e permite disponibilizar como resultado uma estatística referente à comparação dos textos. Os dados estatísticos apresentados são os seguintes:

- Número de palavras - Total de palavras encontradas em cada um dos ficheiros;
- Palavras comuns - Total de palavras iguais encontradas nos dois ficheiros;
- Palavras apagadas - Total de palavras que estão no ficheiro original, mas não estão no ficheiro de resultado;
- Palavras inseridas - Total de palavras que estão no ficheiro de resultado, embora não constem do ficheiro original;
- Palavras alteradas - Total de palavras que foram alvo de troca, isto é, foram inseridas outras no lugar das originais.

Com a exceção do número de palavras, todos os outros itens tem dois valores, o número de palavras e a percentagem em relação ao número de total de palavras no ficheiro. A ferramenta `wDiff` considera as repetições para a contagem de palavras, ou seja conta a mesma palavra tantas vezes quantas esta existir no texto.. No cálculo da similaridade, faz todo o sentido que as palavras repetidas sejam contadas, uma vez que a comparação é feita palavra a palavra entre os dois arquivos.

Na Figura 3.3 é um ilustrado um exemplo do *output* obtido com a ferramenta `wDiff`.

```
ve after curricular in+} several [-stones inside and sewed-] {+break vico veu again+} the belly [-aga  
in. When the-] {+when+} wolf [-awoke, he-] {+woke up+} saw [-the hunter and fled, full-] {+grow doris  
fled chair+} of [-fear.\nLittle Red Riding Hood hugged her grandmother and promised-] {+the same is+} that {+turchinov primary  
open suava liter retains what not that most jubilation will be his mind+} she [-would never disobey  
her mother again. She, the grandmother and-] {+is carried in+} the [-hunter-] {+middle that sepror+}  
ate the cake and [-the honey, glad that everything went well.-] {+one in her luiz+} Doc1_Orig_PT - t  
ranslated.rtf/TXT.rtf: 735 words 162 22% common 30 4% deleted 543 74% changed  
Transcri_doc_1_16Khz - translated.rtf: 701 words 162 23% common 66 9% inserted 473 67% changed
```

FIGURA 3.3: `wDiff` output example

Ao contrário da ferramenta `copyleaks` o `wDiff` não revela a percentagem de similaridade entre o original e o resultado da extração. O propósito da ferramenta Por forma a poder comparar as duas ferramentas definiu-se o valor de similaridade entre ficheiros de forma a que este corresponda à percentagem de palavras comuns entre os ficheiros. Tal similaridade é dada pela

Equação 3.1.

$$\text{Porcentagem Similaridade} = \frac{\text{Palavras Comuns} * 100}{\text{Numero de palavras}} \quad (3.1)$$

Na Tabela 3.2 pode-se ver as funcionalidades dos dois comparadores de ficheiros apresentados.

Comparação de extratores		
Item	Extrator	
Palavras inseridas	✓	
Palavras apagadas	✓	
Palavras alteradas	✓	
Palavras comuns	✓	✓
Total de palavras (origem)	✓	
Total de palavras (extração)	✓	
Similaridade		✓

TABELA 3.2: Análise dos Comparadores de ficheiros

3.1.3 Metodologia de análise

Considerando que o objetivo deste trabalho é implementar uma solução capaz de encontrar palavras chave em conteúdos áudio/vídeo obtidos em fontes abertas, os extratores e comparadores desempenham um papel importante. Neste âmbito e antes de passar à fase de implementação propriamente dita é fundamental fazer uma análise sobre as ferramentas apresentadas.

A análise das ferramentas consistirá em avaliar a percentagem de similaridade entre um ficheiro origem conhecido e um ficheiro de texto obtido através dos extratores. Para efeitos de obtenção da percentagem de similaridade serão usados os comparadores descritos. O estudo comparativo contempla formatos áudio/vídeo de três fontes diferentes (musica, audiobook e texto) e com três idiomas diferentes (Espanhol, Inglês e Português). Os resultados a obter permitirão observar:

- a taxa de acerto dos diferentes extratores e comparadores considerando diferentes idiomas base;
- a taxa de acerto dos diferentes extratores e comparadores considerando diferentes velocidades de discurso;
- a necessidade de traduzir idiomas para um idioma standard que apresente maiores níveis de acerto;
- a fiabilidade de uma solução de identificação automática de palavras chave em fontes áudio/vídeo.

Considerando os resultados, o objetivo da análise é selecionar quais as ferramentas a integrar na ferramenta **OSINT** baseada na extração e análise de dados áudio/vídeo: TExtractor.

3.2 Plataforma a implementar: TExtractor

Na segunda fase do trabalho será implementada a plataforma TExtractor. O objetivo desta plataforma é a extração de texto a partir de fontes de áudio/vídeo e posterior pesquisa por *palavras chave*. As fontes podem ser arquivos de áudio/vídeo, assim como um **URL** do **YouTube**. O diagrama lógico da plataforma é ilustrado na Figura 3.4.

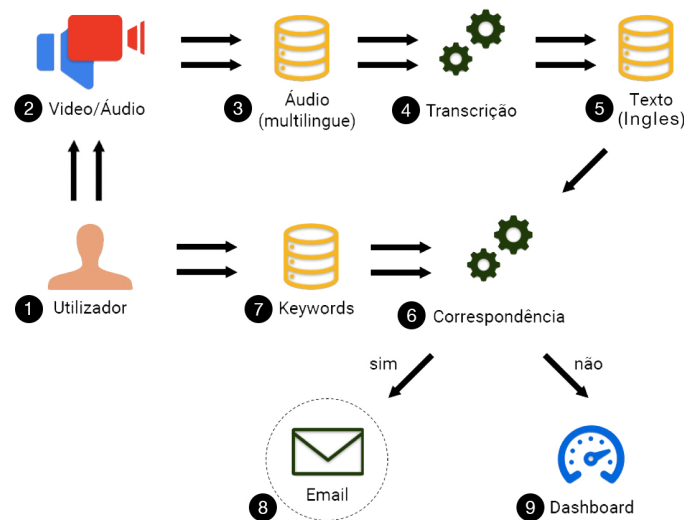


FIGURA 3.4: Imagem elucidativa da plataforma da proposta da plataforma

A plataforma contempla uma sequência de ações identificada pelos números incluídos na Figura 3.4. Essa sequência é a seguinte:

1. Utilizador, insere ficheiros/**URL** de áudio/vídeo e *palavras chave*.
2. Ficheiros/**URL** áudio/vídeo inseridos pelo Utilizador.
3. Armazenamento de ficheiros (vários idiomas).
4. Extração de texto dos arquivos de áudio/vídeo.
5. Tradução do texto para inglês.
6. Verificar se existe correspondência com as *palavras chave* predefinidas.
7. Armazenamento das *palavras chave*.
8. Se existem *palavras chave* no texto, é enviado email de alerta.

9. É dada possibilidade de fazer nova extração ou visualizar extrações já efetuadas num dashboard da aplicação.

Nesta primeira versão da plataforma a introdução dos arquivos na plataforma será efetuada através do *upload* de ficheiros ou pela indicação de um **URL** relativo a um vídeo **YouTube**. A indicação da origem dos dados bem como do conjunto de *palavras chave* a pesquisar é da responsabilidade do utilizador. O email será usado para envio de alertas sobre a ocorrência de *palavras chave* nos dados fonte e é dada a possibilidade de acesso ao um dashboard através do qual se poderá ver o estado da extrações e resultados obtidos. A implementação da plataforma tirará partido das ferramentas de extração e comparação apresentadas. Em função dos resultados obtidos pela análise das ferramentas serão definidas as ferramentas a integrar na plataforma.

Capítulo 4

Estudo comparativo das ferramentas

Neste capítulo apresentamos o estudo comparativo realizado para avaliar a taxa de acerto das ferramentas de extração de texto a partir de fontes áudio/vídeo. Ao longo do mesmo será descrito o ambiente de testes e os resultados obtidos para as diferentes combinações de extratores, comparadores, idiomas e *speech* do *input*.

4.1 Ambiente de Testes

Os testes realizados foram executados num sistema `macOS` tirando partido da `MIDI` para proceder à extração do áudio. A `MIDI` [50] cria o canal de "alimentação" dos extratores com os ficheiros fonte. Esta ferramenta permite utilizar computadores e instrumentos musicais para criar e ouvir música.

O `macOS` tem as seguintes *specs* e foram utilizadas as funcionalidades:

- MacOS Sierra Version 10.12.5 (16F73)
- Ferramenta de discurso do macOS
- Midi Interface Versão 3.1 (3.1)
- Goole Chrome Version 58.0.3029.110 (64-bit)

Para efeitos de estudo serão usados três tipos de dados audiobook, música e texto em três tipos de linguagem cada (Espanhol, Inglês e Português). Para cada um dos casos existirá o ficheiro origem em modo texto que contém o texto integralmente escrito. Este texto será usado pelos dois comparadores (*wDiff* e *copyleaks*) para averiguar a capacidade de extração de cada um dos extratores (Speech to Text, Web Speech API, Speechlogger e Speechnotes). Os resultados serão mostrados por percentagem de similaridade.

Em função dos resultados serão escolhidas as ferramentas (um extrator e um comparador de texto) mais eficazes para utilizar no desenvolvimento da plataforma.

4.2 Estudo experimental

Para o desenvolvimento deste estudo, foram usadas várias ferramentas que foram descritas na secção 3.1. Todas elas foram escolhidas de forma aleatória, sem qualquer tipo de rigidez quanto aos seus requisitos, apenas se levou em conta que fossem ferramentas com licença *open source* ou pelo menos fornecesse um meio de teste mesmo que limitado.

Para obter resultados consistentes, a análise das ferramentas selecionadas é feita a partir de três tipos de fontes abertas de áudio/vídeo, são elas: Audiobook, Música e Texto. Cada uma das fontes existe em três idiomas diferentes: Inglês, Português e Espanhol.

Cada extrator extraiu a mesma fonte três vezes, o que dá um total de cento e oito extrações. A Equação 4.1. representa o número de testes e combinações que levam a tal número de extrações.

$$\text{Extrações} = 3 \text{ Transcrições} * 4 \text{ Extratores} * 3 \text{ Idiomas} * 3 \text{ Fontes} \quad (4.1)$$

4.3 Apresentação da análise de resultados

Na Tabela 4.1 são apresentados os resultados dos extractores Speech to Text, Web Speech API, Speechlogger e Speechnotes quando comparados através do **copyleaks**. As percentagens representam a taxa de acerto, isto é, o número de palavras encontradas entre o texto original e o texto obtido pela extração automática com cada uma das ferramentas a dividir pelo número total de palavras. Como se verifica pela tabela os resultados são pouco satisfatórios. Só existem quatro valores acima dos 50% de acerto.

É de salientar que tais resultados podem ser consequência da forma de funcionamento do **copyleaks**. Este apenas conta as palavras se estas estiverem na mesma frase e na mesma posição. Porém verifica-se que apesar de válido este não é o único motivo, porque, quando é feita a extração de um arquivo de áudio, se o arquivo for música, a extração é fraca ou mesmo nula. Pelo resultado verifica-se que os extractores não lidam bem com barulho de fundo, uma vez que estão formatados para extrair discursos em ambientes limpos, isto é, isentos de ruído. Nos ficheiros de música a voz não é totalmente clara, assim como a pronuncia e mesmo a fluidez de discurso varia muito, o que dificulta o reconhecimento da voz.

Os valores apresentados na Tabela 4.1 e no gráfico da Figura 4.1, não são animadores. A partir dos mesmos não parece viável implementar um sistema com esta configuração. Estes valores, mostram que os extractores não funcionam muito bem ao extrair texto de um áudio de música. Ainda que com resultados baixos, os melhores valores obtidos verificam-se na análise de texto e nas extrações em inglês.

Copleaks melhores resultados				
Item	Extrator	Inglês	Português	Espanhol
		Música		
Speech to Text		0.00%	0.00%	0.00%
Web speech API		0.00%	6.90%	0.00%
Speechlogger		0.00%	0.00%	0.00%
Speechnotes		0.00%	0.00%	0.00%
Texto				
Speech to Text		74.60%	0.00%	0.00%
Web speech API		22.40%	9.60%	8.70%
Speechlogger		17.60%	55.60%	15.80%
Speechnotes		22.60%	5.30%	16.90%
AudioBook				
Speech to Text		31.10%	0.00%	11.80%
Web speech API		18.90%	6.20%	14.80%
Speechlogger		100.00%	64.00%	14.40%
Speechnotes		17.80%	4.80%	15.10%

TABELA 4.1: Resultados da comparação **copleaks**

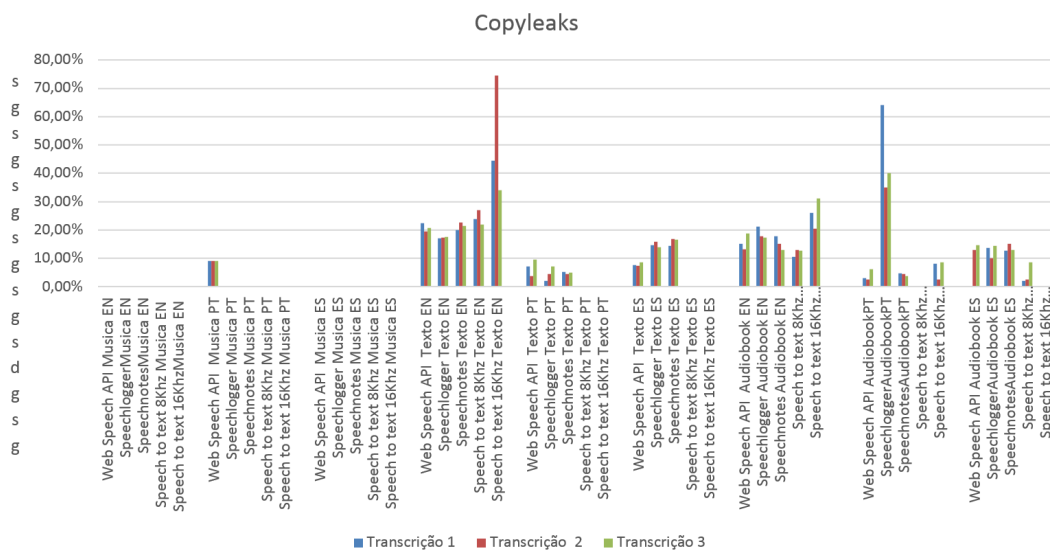


FIGURA 4.1: Resultados comparação **copleaks**

A análise que se segue mostra o resultado obtido para os mesmos dados, mas desta vez comparados com recurso à ferramenta **wDiff**. Os resultados são apresentados na Tabela 4.2. Como se pode ver pela tabela os resultados são melhores. Para este comparador todas as palavras encontradas contam, independentemente da sua posição no texto. O único requisito para que a palavra seja contada é que esteja presente em ambos os ficheiros. Este é o elemento mais diferenciador relativamente ao anterior e por isso atribui-se a tal a melhor *performance* para a comparação de texto em termos absolutos. É certo que tais resultados poderão ser diferentes do apresentado e neste âmbito será necessário executar estudos mais aprofundados para determinar se a ordem das palavras é realmente importante, isto é, se não distorce a mensagem original.

wDiff melhores resultados				
Extractor	Extração(%)	Inglês	Português	Espanhol
		Música		
Speech to Text		5.64%	15.13%	23.9%
Web speech API		4.97%	13.16%	1.06%
Speechlogger		5.37%	6.58%	3.01%
Speechnotes		2.28%	3.95%	0.35%
Texto				
Speech to Text		72.30%	23.54%	11.18%
Web speech API		67.47%	45.48%	34.76%
Speechlogger		60.38%	40.89%	58.20%
Speechnotes		64.55%	45.85%	49.59%
AudioBook				
Speech to Text		68.46%	19.54%	52.57%
Web speech API		63.42%	43.45%	56.19%
Speechlogger		62.56%	34.93%	51.51%
Speechnotes		63.08%	40.96%	56.19%

TABELA 4.2: Resultados da comparação **wDiff**

Como se pode verificar, não só pelos valores da Tabela 4.2 como também no gráfico da Figura 4.2, os valores obtidos já são mais consistentes entre os extratores. Continuam a predominar valores altos nas extrações em inglês e o texto extraído, a parte de música continuam a ser os mais baixos. Isto reforça a ideia que os extratores não estão preparados para fazer extração de áudio com valores de ruído de fundo alto, como é o caso da música.

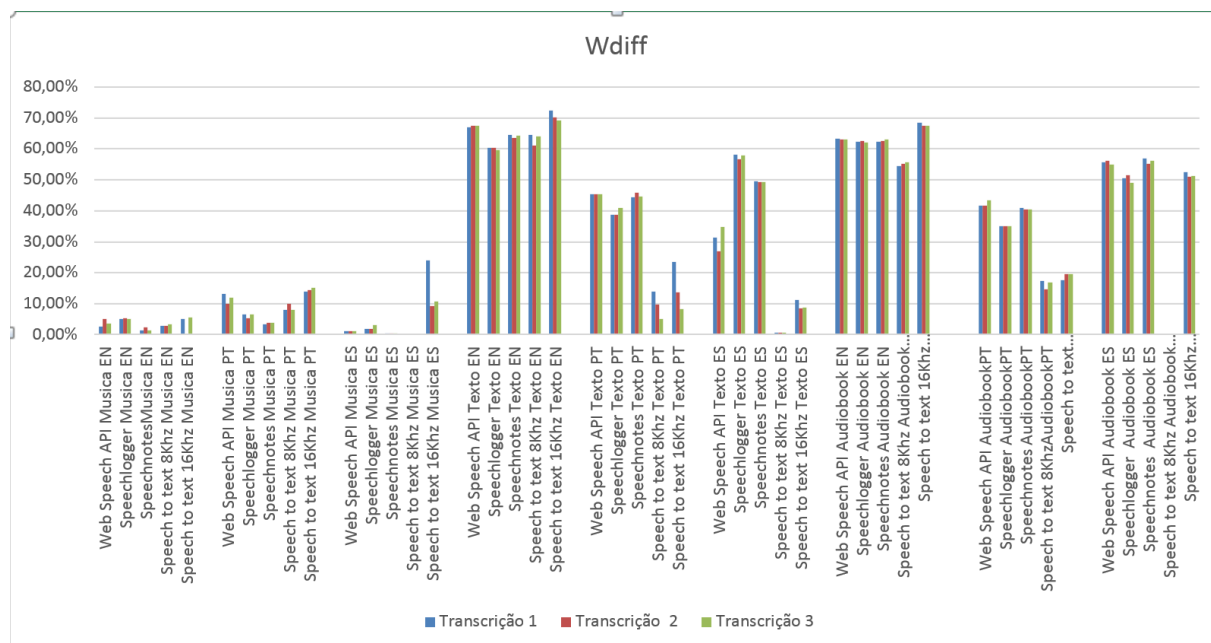


FIGURA 4.2: Resultados da comparação **wDiff**

Para confirmar se realmente se mantêm os valores altos em inglês para as extrações em outros idiomas, fez-se outro teste que passa por traduzir todos os arquivos fonte e extraídos para inglês, e assim verificar se confirmam a tendência de valores mais altos. Após a extração de um arquivo o mesmo vai ser traduzido, assim como a sua fonte, de forma a utilizar as ferramentas de comparação para verificar se existe um aumento na sua similaridade.

Todas as fontes e extrações nos idiomas Português e Espanhol foram traduzidos para Inglês utilizando para tal o tradutor *online* disponibilizado pela **Google**. Após estas traduções, foram efetuadas novamente as comparações entre a fonte e a extração, ambas traduzidas, utilizando as mesmas ferramentas de comparação, o **copyleaks** e o **wDiff**.

Na Tabela 4.3 são apresentados os resultados obtidos após a tradução das extrações para uma língua universal, o Inglês. O comparador utilizado foi o **copyleaks**.

Copyleaks traduzido melhores resultados		
Extractor \ Extração(%)	Português	Espanhol
Música		
Speech to Text	0.00%	0.00%
Web speech API	0.00%	0.00%
Speechlogger	0.00%	0.00%
Speechnotes	0.00%	0.00%
Texto		
Speech to Text	0.00%	0.00%
Web speech API	5.20%	7.40%
Speechlogger	0.00%	6,10%
Speechnotes	2.10%	14.00%
AudioBook		
Speech to Text	0.00%	6.10%
Web speech API	0.00%	10.00%
Speechlogger	0.80%	2.10%
Speechnotes	0.00%	6.1%

TABELA 4.3: Resultados da comparação **copyleaks**(tradução)

De acordo com os resultados apresentados na Tabela 4.3 e tal como na primeira comparação, com fontes e arquivos extraídos no idioma de origem, estes testes, agora com fontes e arquivos extraídos traduzidos para Inglês, não foram obtidos valores que fortalecessem uma questão inicial de que as comparações em Inglês trariam melhores resultados. Estes valores vêm sim fortalecer a ideia de que os comparadores são coerentes, independentemente dos idiomas, pois os resultados não diferem muito dos resultados das comparações nos idiomas de origem. No caso do **copyleaks** como se pode verificar na Tabela 4.3 e no gráfico da Figura 4.3, pode-se confirmar que os resultados ficam significativamente abaixo dos resultados obtidos nas comparações nos idiomas de origem.

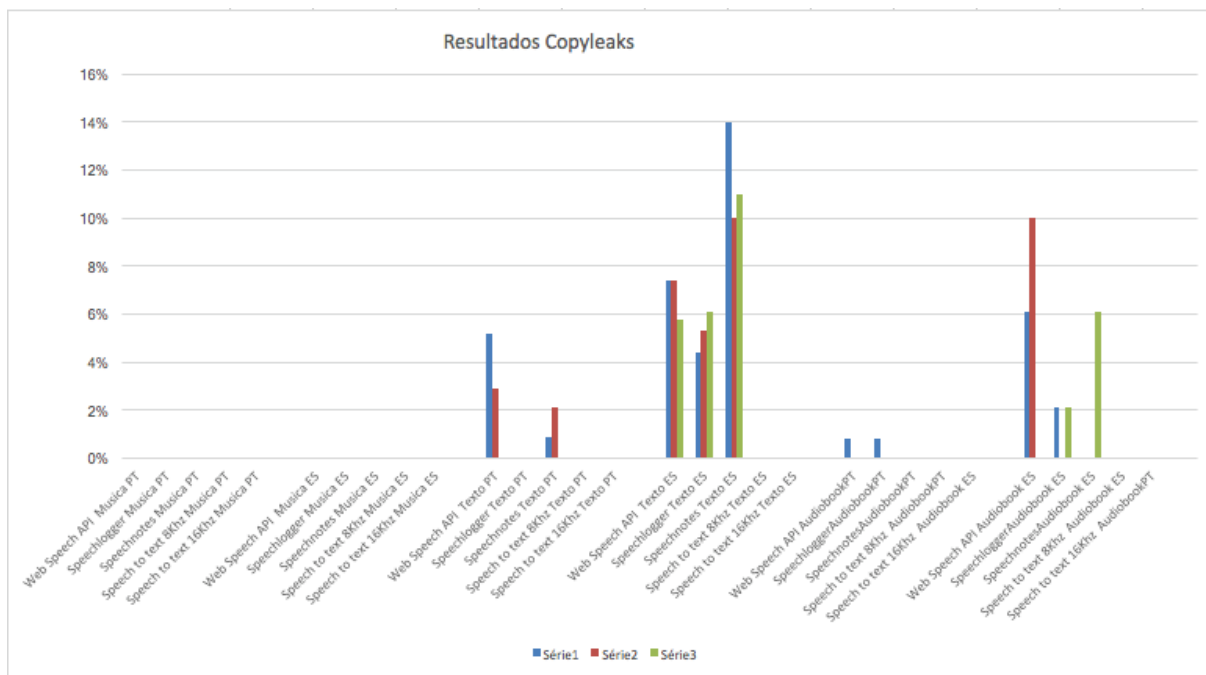


FIGURA 4.3: Resultado da tradução copyleaks

A mesma análise foi efetuada com o **wDiff**. Os resultados obtidos são apresentados na Tabela 4.4 e no respetivo gráfico da Figura 4.4. O **wDiff** obteve valores mais altos que o **copyleaks**, mas os valores não diferiram dos obtidos na primeira comparação de arquivos extraídos e fontes no idioma de origem.

wDiff traduzido melhores resultados		
Extractor \ Extração(%)	Português	Espanhol
Música		
Speech to Text	16.56%	5.76%
Web speech API	8.59%	5.57%
Speechlogger	4.91%	3.61%
Speechnotes	4.29%	3.07%
Texto		
Speech to Text	22.04%	11.76%
Web speech API	44.90%	45.85%
Speechlogger	37.09%	50.17%
Speechnotes	43.54%	56.59%
AudioBook		
Speech to Text	19.39%	52.57%
Web speech API	36.28%	46.71%
Speechlogger	30.71%	43.78%
Speechnotes	34.93%	48.11%

TABELA 4.4: Resultados da comparação **wDiff** (tradução)

De acordo com os resultados verifica-se que após a tradução dos arquivos para um idioma base (Inglês) e efetuando uma nova comparação mantém-se a tendência da primeira. Os valores da

extração de arquivos de *texto* e os arquivos *audiobook* apresentam percentagens de similaridade mais elevadas (intervalo entre os 60% e os 70%). Relativamente aos ficheiros de música as percentagens não vão além dos 20%. Uma visão integrada dos resultados obtidos com o **wDiff** é apresentada na Figura 4.4.

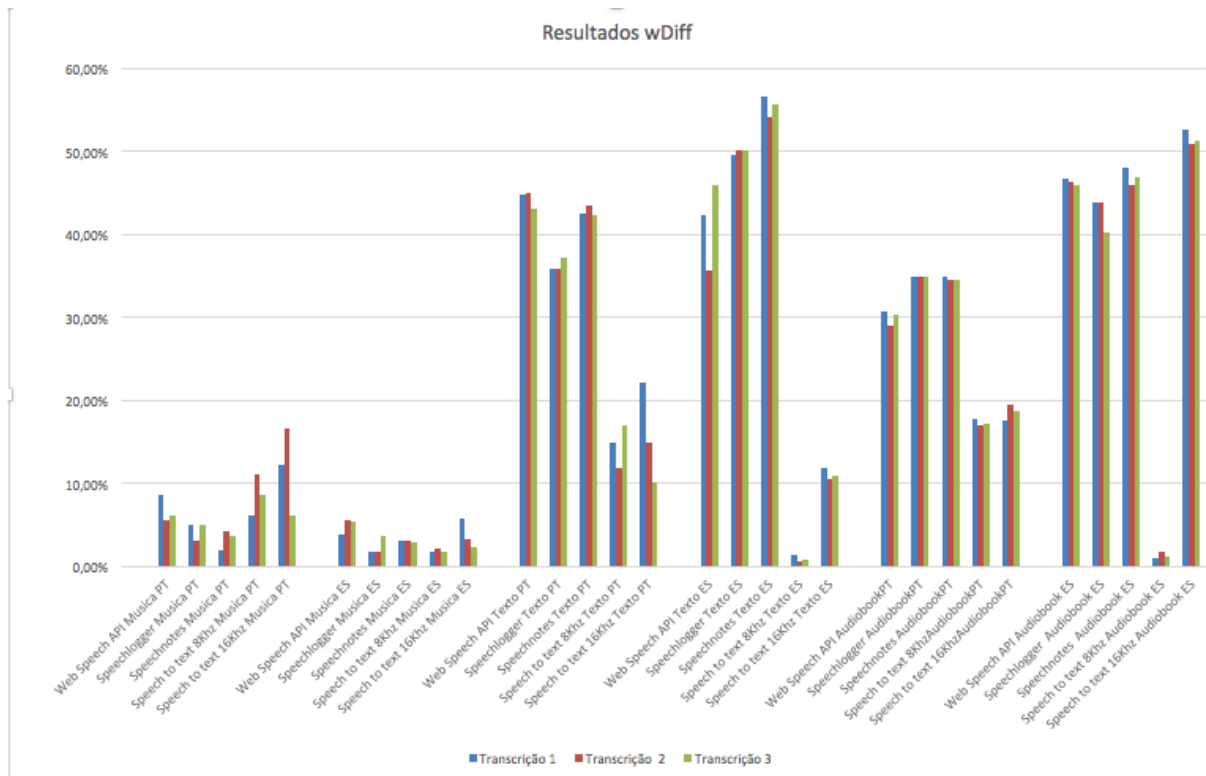


FIGURA 4.4: Resultado da tradução **wDiff**

4.4 Discussão sobre os resultados

Através dos resultados obtidos e apresentados ao longo deste capítulo, é possível verificar que independentemente dos idiomas de origem, a comparação dos arquivos de musica apresentam valores muito baixos em ambos os comparadores. Tal como já foi referido, o principal motivo será a forma como os extratores lidam com o ruído de fundo, sendo eles baseados em tecnologias de reconhecimento de voz, nenhum deles tem filtros que possibilitem a separação da voz dos ruídos de fundo. Esta, que pode ser considerada uma lacuna na extração de texto através de áudio/vídeo, uma vez que por vezes as fontes a extrair não são isentas de ruído de fundo. Outra possibilidade para os baixos valores nos resultados obtidos através da extração de musica pode ser pelo facto de haver alterações de voz na pronuncia e/ou dicção. Esta situação não está considerada no âmbito de estudo, uma vez que o trabalho propõe a extração e posterior comparação e não a forma como funcionam os extratores.

Outra situação verificada tem a ver com o facto dos valores das extrações e comparações efetuadas no idioma Inglês serem mais altos, mais uma vez esta situação deriva da forma como os extratores foram construídos. Foi possível confirmar esta situação, pois após a tradução dos

arquivos fonte e extraídos, ambos comparadores obtiveram resultados similares aos anteriores. Embora os resultados tenham ficado um pouco abaixo, estes vêm confirmar a consistência dos comparadores independentemente do idioma, o facto de serem relativamente mais baixos pode ser uma perda através do processo de tradução, que carece, contudo, de análise mais profunda.

Quanto aos comparadores, é de salientar os resultados obtidos, como se pode verificar através da Tabela 4.1 e do gráfico ilustrado na Figura 4.1 referentes ao **copyleaks**, este apresenta resultados muito inferiores ao **wDiff**, apresentados na Tabela 4.2 e no gráfico ilustrado na Figura 4.2.

Estes resultados foram obtidos através de várias extrações e comparações como também já foi referido tem o intuito de servir de base à escolha das ferramentas para a implementação do sistema proposto.

Embora nem sempre o **Speech to text 3.1.1.1** apresentasse os valores mais altos nos idiomas Português e Espanhol, foi o escolhido para a execução do sistema proposto. Esta escolha foi suportada fundamentalmente pelos resultados observados na extração e comparação de fontes de dados em inglês, pesando também as funcionalidades e a existência de uma **API** que facilita a sua adoção.

Capítulo 5

Implementação da plataforma TExtractor

O TExtractor é uma aplicação idealizada para a monitorização de *palavras chave* em conteúdos áudio/vídeo tirando partido da transcrição desses conteúdos para texto. Trata-se de uma ferramenta útil para a monitorização de marcas, entidades e também para a deteção de atividades criminosas planeadas com recurso a áudio/vídeo (e.g. planeamento de uma ação ciberterrorista).

Além da utilidade também é um objetivo da aplicação ser fácil de utilizar. Tal como apresentado na proposta de plataforma, para utilizar a aplicação é apenas necessário escolher o canal a monitorizar e inserir uma lista de palavras chave a monitorizar (e.g., nome da marca, nomes de pessoas, endereços de email, domínios Internet, linguagem usada em atividades maliciosas). Sempre que essas palavras forem pronunciadas, independentemente do idioma deverá ser detetada a sua ocorrência e emitidos alertas. Uma aplicação deste tipo promove a deteção antecipada de ataques.

A plataforma e respetiva aplicação não se limita a área da *cyber security*. É na prática um sistema transversal a qualquer área em que as organizações ou entidade necessitem saber se são referenciadas ou não nos canais abrangidos pela aplicação. As empresas investem muito dinheiro na área publicitária e contratam serviços de *clipping* para observar a execução da publicidade. O *clipping* tende a ser uma atividade manual e uma ferramenta como a que aqui se propõe poderá auxiliar o processo de recolha dessa informação, libertando o humano para outras tarefas/funções.

Esta ferramenta é de grande importância para o **OSINT**. Acrescenta valor na aquisição de dados, uma vez que aumenta a amplitude das fontes de dados e fornece informação útil a definição de informações/inteligência.

5.1 Implementação do TExtractor

A Figura 5.1 dá a conhecer o design lógico da plataforma a criar. Tal pressupõe um conjunto de módulos que permitem a recolha de dados, o seu armazenamento, a transcrição de áudio/vídeo para texto, a comparação entre o texto traduzido e a lista de palavras chave a monitorizar, interfaces de visualização e gestão e ainda sistemas de alertas para os utilizadores.

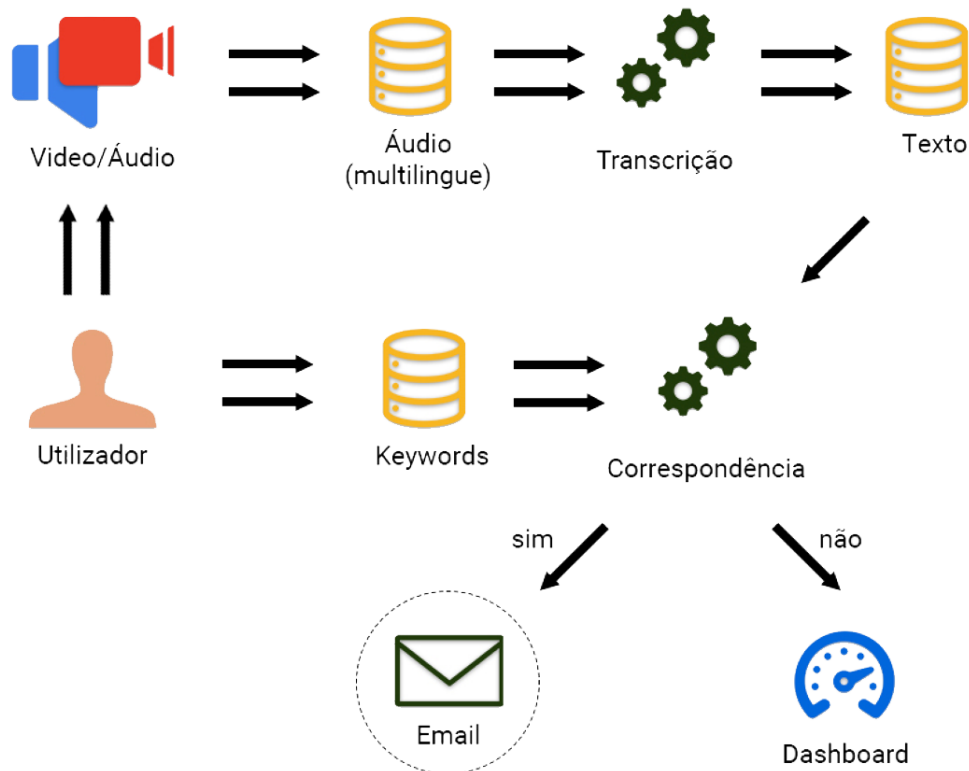


FIGURA 5.1: Imagem elucidativa da plataforma

Uma das decisões fundamentais numa plataforma como a que se propõe tem a ver com o nível de confiança que se poderá ter na mesma para a função que desempenha. No capítulo anterior foi apresentado o resultado de um estudo comparativo que considerou diferentes tipos de input e diferentes comparadores para identificar ocorrências de palavras chave. Os melhores resultados obtidos variam os 60% e os 70% de acerto. Não é propriamente um valor elevado, mas se considerarmos uma redução no esforço humano para estas análises para a deteção antecipadas de ciber ataques, a sua utilidade ganha expressão.

Considerando a utilização da aplicação por uma comunidade e ciber segurança, tomou-se a opção de desenvolvimento de uma aplicação Web. Tal aplicação é suportada por um conjunto de ferramentas que serão descritas nas próximas subsecções.

5.1.1 MongoDB

O **MongoDB** [12] é uma base de dados popular e frequentemente aplicada para desenvolvimento de projetos baseados em Node.js, isto porque os documentos são criados e armazenados em arquivos **BSON**. Estes arquivos são de formato binário **JSON** que suporta todos os dados de tipo **JS**, o que permite transferir dados entre servidores e aplicações **Web** com um formato legível para humanos. É também uma melhor opção quando se necessita de grande capacidade de armazenamento e velocidade, pois oferece maior eficiência e confiabilidade.

Um dos principais benefícios oferecidos pelo **MongoDB** é o uso de esquemas dinâmicos que eliminam a necessidade de pré-definir a estrutura, como campos ou tipos de valor. Este modelo permite a representação de relacionamentos hierárquicos, o armazenamento de matrizes e a capacidade de alterar a estrutura de registos simplesmente adicionando ou excluindo campos. Esta solução *NoSQL* tem incorporada a replicação para uma melhor escalabilidade e alta disponibilidade [13].

Esta aplicação propõe-se a manipular arquivos de grande dimensão, como já foi referido, o **Mongodb** não só permite alojar arquivos grandes, como também disponibiliza respostas rápidas a pedidos para encontrar esses arquivos. Além de permitir armazenar de forma fácil arquivos de vídeo, tem ainda a vantagem sobre o *filesystem* de poderem ser grandes arquivos, permitir a sua consulta e em caso de necessidade poder replicar esses arquivos [12]. O **MongoDB** é orientado para documentos, usa um documento que substitui o conceito "linha" em base de dados relacionais. Uma das principais ideias do **MongoDB** é que as operações podem ser entregues ao cliente, são transferidas do servidor para o cliente [55] [59].

5.1.2 NodeJS

A escolha do **Node** [17] assenta não só no facto de ser a plataforma de desenvolvimento da *framework* **Stellar** que por sua vez é responsável pelo desenvolvimento da **API** da aplicação, mas também pelo facto de funcionar de forma assíncrona, permitindo executar várias instancias ao mesmo tempo enviando a resposta sempre que o processo termine. Como as extrações podem ser longas, esta ferramenta permite correr várias instancias e devolver o resultado sempre que esteja pronto sem comprometer os outros processos, o que veio permitir que o utilizador do sistema desenvolvido tenha informação dos processos a correr [17]. É uma das melhores *plataformas* conhecidas que disponibiliza o melhor ambiente de desenvolvimento em *JavaScript* para o lado do servidor. Foi lançado em 2009 por **Ryan Dahl** como um projeto de código aberto. É baseado no "motor **V8**", uma implementação *Google*, que compila o código *JavaScript* para código de máquina nativo antes da execução em vez usar tecnologias mais tradicionais como a interpretação de *bytecode* ou a compilação do programa inteiro para código de máquina. O "V8" [33] e o **Node** são implementados maioritariamente em **C** e **C++** o que permite melhor performance e baixo consumo de memória. [52]. O **Node** visa suportar processos de servidor de longa duração. Ao contrário da maioria dos outros ambientes modernos, um processo do **Node**, não conta com o *multithreading* a execução simultânea da lógica de negócio, é baseado num

evento assíncrono **I/O** [52]. O **Node** é uma tecnologia que implementa um modelo baseado em eventos para criar aplicações de rede, já existe uma série de grandes serviços de Internet, como o *PayPal* e *LinkedIn* que foram migrados para **Node**, tendo como resultado melhorias consideráveis no desempenho e facilidade no desenvolvimento [57].

5.1.3 Frameworks

Neste sistema foram usadas algumas *frameworks*, para o desenvolvimento do *frontEnd* foram usadas duas, o **Vue** [35] e o **Vuetify** [36] que são descritas na secção 5.1.3.2 e na 5.1.3.3, respetivamente. Para desenvolvimento do *backend* foi utilizado o **Stellar** [30] cujos pormenores serão apresentados na secção 5.1.3.1. O uso destas *frameworks* teve por objetivo agilizar e simplificar o desenvolvimento da aplicação referente ao sistema proposto.

5.1.3.1 Stellar

O Stellar [30] é uma framework Web baseada em ações para a criação de APIs Web de forma fácil. Esta foi concebida para corrigir as principais falhas das frameworks atuais tendo em vista um novo paradigma, tem como principais objetivos facilidade na manutenção, elevada escalabilidade e performance. Esta framework consegue suportar múltiplos protocolos em simultâneo, permitindo partilhar o mesmo código base entre funcionalidades que necessitem diferentes tipos de ligações, seja em tempo real (em websocket ou **TCP**), ou então através do protocolo **HTTP**, segundo (ou não) o padrão RESTfull **REST** [25]. Esta encontra-se toda desenvolvida em ES6 e algumas funcionalidades com JavaScript ES7, tudo isto, correndo por cima de Node.JS [30]. Esta framework, além de ser simples, moderna e desenhada de raiz para a realidade de hoje e de futuro, faz-se acompanhar por uma exaustiva documentação de todas as funcionalidades disponibilizadas.

O Stellar não funciona como a maioria das frameworks que recorrem ao modelo **MVC** [14]. Para estruturar a sua arquitetura, o Stellar usa uma filosofia baseada em ações em que todas as funcionalidades correspondem a uma ação. Uma ação, é um pequeno bloco de lógica isolado e altamente testável e pode ser ligado a outras ações de forma a criar ações mais complexas, sem que a complexidade de leitura do código seja aumentada. Um ponto forte é a capacidade de gerar documentação de forma completamente automática [29].

5.1.3.2 Vue.js

O **Vue** [35] é uma *framework* fácil de integrar, focada apenas na camada de visualização, para construção de interfaces **Web**. É uma *framework* acessível, versátil e de alta performance. Acessível, porque se tem conhecimento intermédio em **HTML**, **CSS** e **JavaScript** está pronto a começar a desenvolver aplicações [34]. Versátil, porque é uma *framework* simples e adaptável em camadas incrementais, o que permite lidar com aplicações de pequena ou grande dimensão. É capaz de alimentar aplicações desde que seja combinada com outras bibliotecas de suporte.

A alta performance deve-se ao facto de se poder fazer otimizações com pouco esforço e a uma incrível rapidez do virtual **DOM** [35]. A primeira versão da *framework* apareceu a publico em fevereiro de 2014, desde então a comunidade tem crescido imenso e já há algumas grandes empresas a usar esta *framework*, como por exemplo, o **Facebook**, **Xiaomi**, **Alibaba**, **WizzAir**, **EuroNews**, **Grammarly**, **Gitlab** e a **Laracasts** [42].

5.1.3.3 Vuetify.js

O **Vuetify** [36] é fácil de aprender e fácil de lembrar, uma vez que possibilita criar componentes e propriedades com os nomes desejados. Tem um conjunto de componentes agradáveis e funcionais. O **Vuetify** tem um ciclo de atualização consistente, permitindo que obtenha correções de erros e melhoras com mais frequência. A equipe de desenvolvimento está empenhada em fornecer a melhor experiência que se pode ter. Em caso de necessidade é possível utilizar o *chat* de suporte onde se encontram os membros da nossa comunidade que continua em largo crescimento. Desenvolver com **Vuetify** significa que nunca se estará sozinho. É possível obter ajuda equipe de desenvolvimento e outros membros da comunidade no servidor oficial, imediatamente.

O **Vuetify** é desenvolvido exatamente de acordo com as especificações do *Material Design*. Cada componente é trabalhado manualmente para poder oferecer as melhores ferramentas **UI** possíveis para o seu próximo projeto. O desenvolvimento não pára nos componentes principais delineados nas especificações do Google. Através do apoio dos membros da comunidade e patrocinadores, estes componentes são projetados, desenvolvidos e disponibilizados componentes adicionais [36].



5.1.4 Integração das ferramentas

O conjunto de ferramentas apresentadas na secção 5.1 tornaram possível a implementação da plataforma, simplificando o desenvolvimento da mesma. Usou-se o **Node** por ser não só usado no desenvolvimento da *framework Stellar*, mas porque também disponibiliza formas simples e de rápida implementação para efetuar uma de ligação e interação com o **Mongob**. Por sua vez o **Mongob** já tem embebido o armazenamento de ficheiros de grandes tamanhos, um requisito, devido à natureza dos ficheiros com a plataforma terá que lidar, por norma serão ficheiros grandes. A interface do utilizador foi desenvolvida com recurso às plataformas **Vue** e **Vuetify**, estas *frameworks* foram de grande valor, a *framework Vuetify* disponibiliza componentes totalmente desenvolvidas o que retirou a necessidade do seu desenvolvimento e respetivos estilos. O desenvolvimento da interface gráfica utilizando o **Vue** possibilitou a reutilização de componentes e como é uma *framework* de desenvolvimento simples, agilizou o desenvolvimento da plataforma. Todas as ferramentas utilizadas no desenvolvimento da plataforma tinham por objetivo facilitar e agilizar o desenvolvimento de forma a utilizar o menor tempo possível. Esta necessidade de ganhar tempo advém da necessidade de um maior esforço no estudo e análise efetuado na primeira fase do trabalho.

5.2 Apresentação da Aplicação

A aplicação denominada de **TExtractor**, está disponível online através do [URL](http://www.textractor.xyz) www.textractor.xyz. A aplicação permite extrair texto de arquivos de áudio/vídeo através do *upload* de ficheiros ou de um **URL** para por exemplo um vídeo do **Youtube**.

Inicialmente é pedido ao utilizador que se registe e faça login de forma a possibilitar à aplicação obter um endereço de email, este serve única e exclusivamente para se identificar perante a aplicação e é usado para notificar o utilizador. O utilizador apenas será notificado caso sejam encontradas *palavras chave* nas extrações efetuadas pelo mesmo. As *palavras chave* são solicitadas sempre que seja iniciado um processo de extração, assim como o idioma da extração a fazer. Por padrão o idioma a transcrever será o inglês, no caso das *palavras chave* não corresponderem ao idioma do arquivo, estas dificilmente serão encontradas.

A introdução das *palavras chave* são feitas no espaço reservado para tal como se pode ver na Figura 5.6, podendo ser eliminadas ao clicar no [X] da *palavra chave* referente (). No canto superior esquerdo existe o **ICON** , ao clicar neste **ICON** o menu lateral é escondido ou mostrado, conforme a situação do mesmo. A ideia de esconder o menu lateral propõe-se a disponibilizar mais espaço para a análise do texto. Neste menu figuram as funcionalidades principais, clicando em qualquer uma delas, o utilizador é redirecionado para a *vista* da funcionalidade em questão.

Apenas são guardadas as extrações que contenham uma ou mais *palavras chave*, todas as extrações guardadas ficam disponíveis para consulta com as *palavras chave* sombreadas a amarelo. Ao passar o cursor por cima é disponibilizada informação com o nível de confiança com o qual a palavra foi transcrita, assim como o período temporal em que foi referida (dentro do arquivo extraído). Existem dois estados em que as extrações podem estar: processadas ou em processamento. As processadas são as extrações que foram feitas e foram analisadas quanto à existências de *palavras chave*. Em processamento estão as extrações que ainda se encontram em processamento. A aplicação permite ao utilizador apagar arquivos.

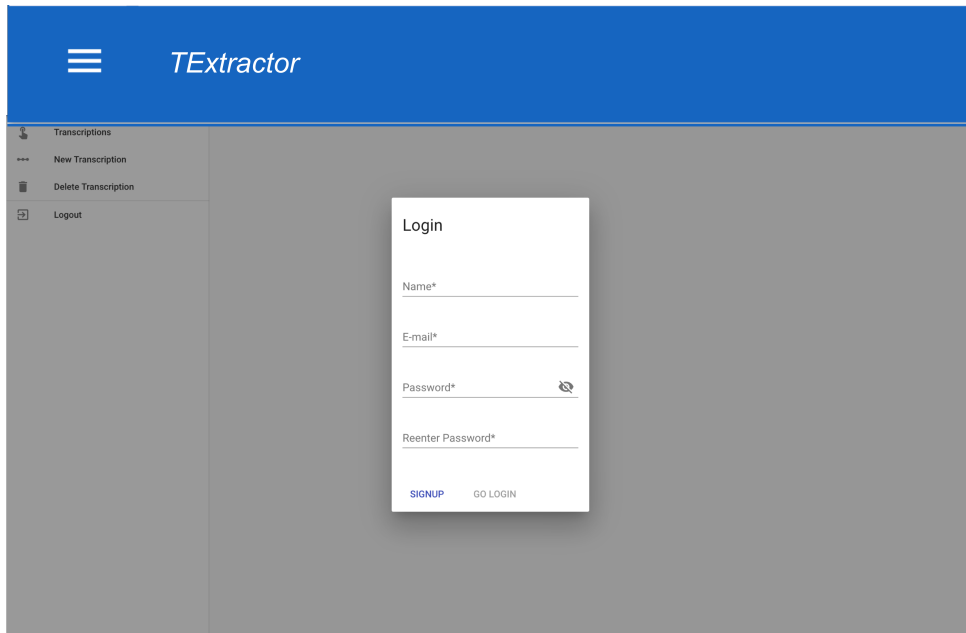
5.2.1 Funcionalidades

Nesta secção são apresentadas as funcionalidades da aplicação, assim como a descrição do seu funcionamento. São apresentadas imagens da aplicação referente a cada funcionalidade.

5.2.1.1 Registo e autenticação

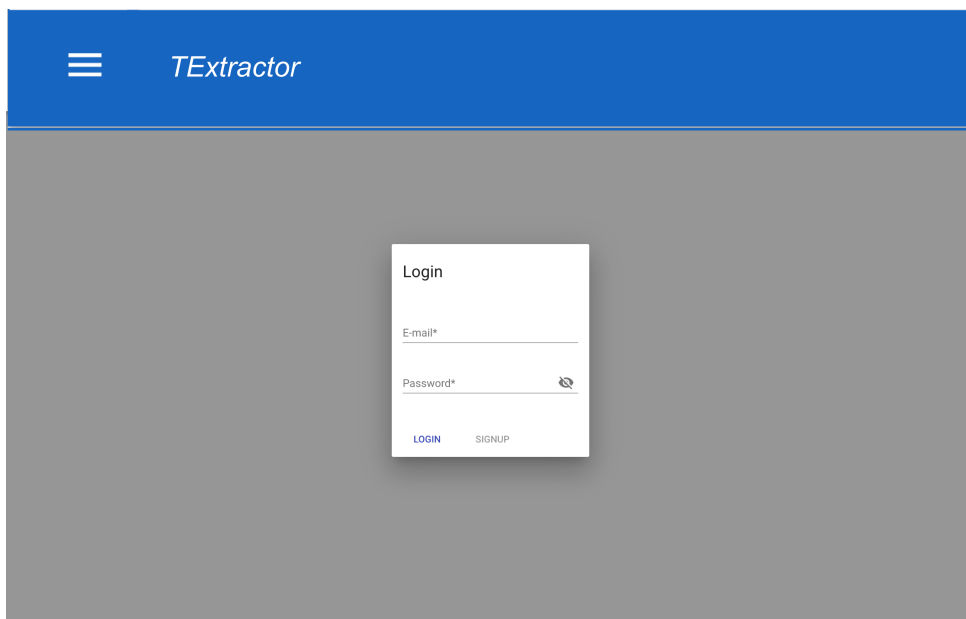
Por uma questão de segurança, a maioria das aplicações tem um sistema de registo e autenticação, também esta está provida de tais operações. A principal razão para se ter desenvolvido e incorporado estas funcionalidades na aplicação, foi para que vários utilizadores possam utilizar a aplicação ao mesmo tempo e apenas receberem como resultado os processos por eles iniciados. Desta forma cada utilizador apenas tem acesso ao resultado das suas extrações. A aplicação

está provida de um sistema de envio de email, no caso de existir uma ou mais *palavras chave* na extração, é enviado um email ao utilizador. Esta ação carece da identificação do utilizador, o que é feito através das credenciais de acesso, onde figura o email, que é de preenchimento obrigatório ao fazer o registo. A pagina de entrada da aplicação disponibiliza estas duas funcionalidades, como apresentado nas figuras 5.2 e 5.3.



The screenshot shows the TExtractor application interface. At the top, there is a blue header with a hamburger menu icon and the text "TExtractor". Below the header, a sidebar menu is visible with the following items: "Transcriptions", "New Transcription", "Delete Transcription", and "Logout". The main content area is a light gray background. In the center, there is a white modal form titled "Login". The form contains the following fields: "Name*", "E-mail*", "Password*" (with an eye icon for visibility), and "Reenter Password*". At the bottom of the form, there are two buttons: "SIGNUP" and "GO LOGIN".

FIGURA 5.2: Formulário de registo



The screenshot shows the TExtractor application interface. At the top, there is a blue header with a hamburger menu icon and the text "TExtractor". Below the header, the main content area is a light gray background. In the center, there is a white modal form titled "Login". The form contains the following fields: "E-mail*" and "Password*" (with an eye icon for visibility). At the bottom of the form, there are two buttons: "LOGIN" and "SIGNUP".

FIGURA 5.3: Formulário de login

5.2.1.2 Extrações

O foco da aplicação são as extrações, como tal, após um registo válido e o login tenha sido efetuado com sucesso, é apresentado ao utilizador uma de duas *vistas*, uma lista com, ou sem documentos como se pode verificar nas figuras 5.4 e 5.5 respetivamente.

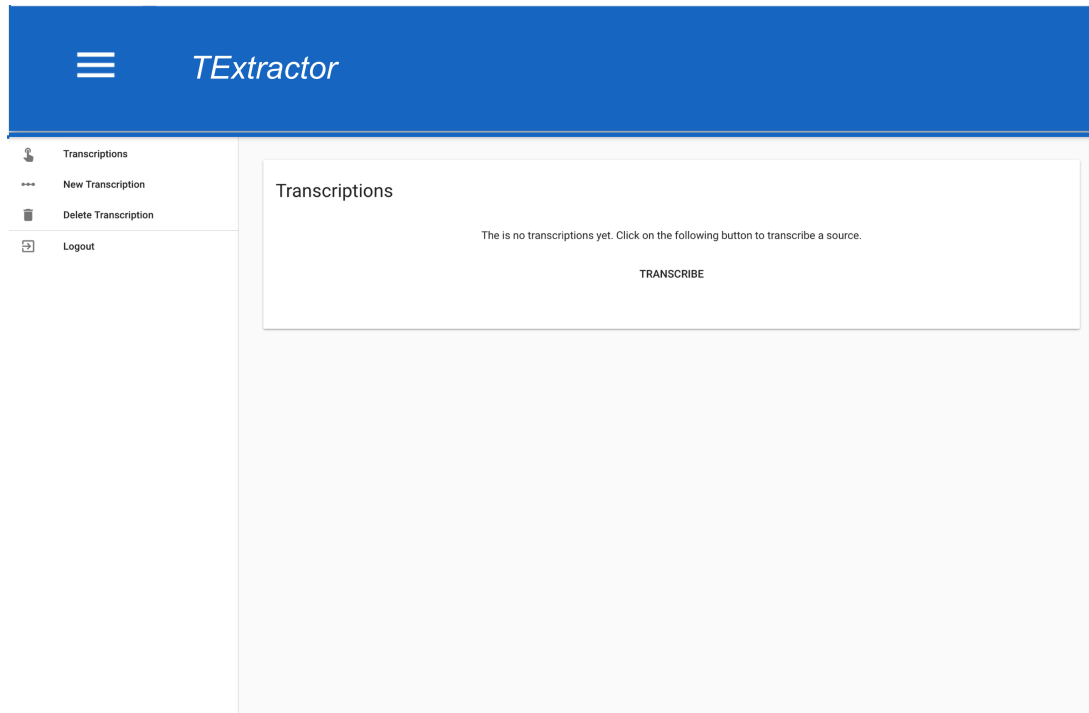


FIGURA 5.4: Interface inicial sem extrações

As *vistas* ilustradas nas figuras 5.4 e 5.5 diferem apenas da atividade do utilizador. Caso o utilizador já tenha usado a aplicação para extrair arquivos e não os tenha apagado, terá um ou mais itens na sua lista de extrações (Figura 5.5). A *vista* sem elementos (Figura 5.4), é apresentada aos utilizadores que ainda não efetuaram qualquer extração, ou a utilizadores que apagaram as suas extrações por completo. A *vista* da Figura 5.5 é a principal, uma vez que é nela estão presentes todos os processos, fala-se em todos os processos uma vez que nela podem figurar além dos processos terminados, todos os outros que se encontrem ainda em execução.

Para que se possam identificar todos os processos da lista, através dos seus estados, cada processo está identificado pelo nome, por baixo deste figura o resultado do processo e à direita um **ICON** [11] que identifica o seu estado. Os processos podem ter dois estados, *em processamento* ou *processado*, os processos no estado *em processamento* dividem-se em dois grupos, processados e com *palavras chave* encontradas ou sem *palavras chave* encontradas. Além da descrição que aparece por baixo do nome do arquivo ser diferente, também o **ICON** de cada processo é diferente. O **ICON** ⓘ identifica processos terminados em que o arquivo extraído contém *palavras chave*, o **ICON** ✓ representa os processos terminados que não contém *palavras chaves* e o **ICON** ⏸

representa processos que ainda estão a decorrer, não terminaram. A Figura 5.5 ilustra a listagem com os elementos referidos.

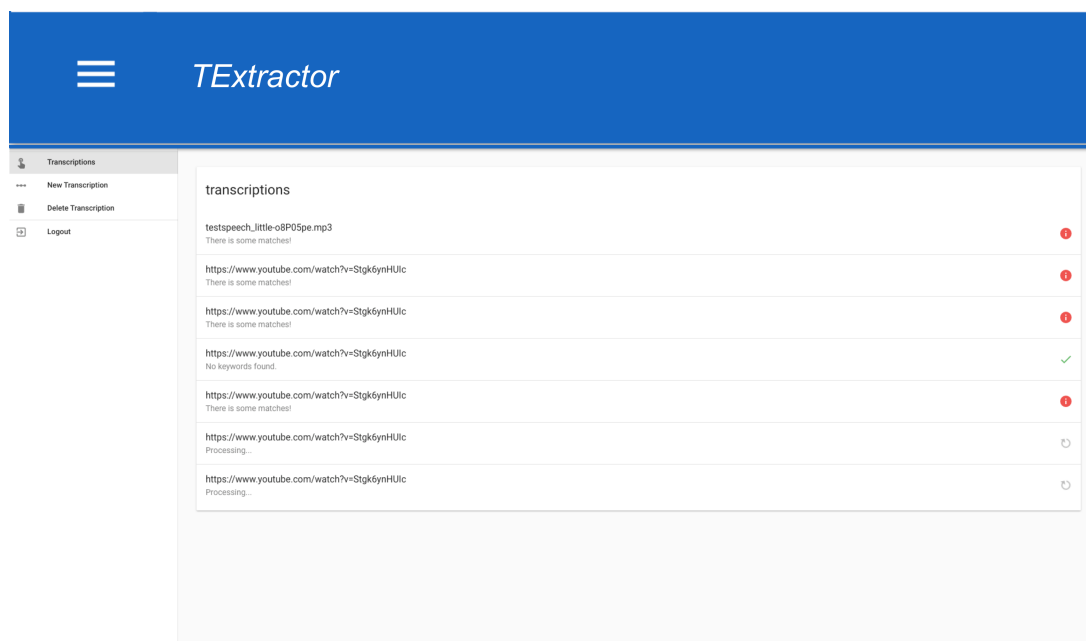


FIGURA 5.5: Interface de extrações

5.2.1.3 Nova extração

Ao clicar no separador **New transcription** é apresentada uma *vista* ao utilizador, onde este pode optar por fazer a extração de um arquivo. O utilizador terá que efetuar o *upload* de um arquivo ou indicar o **URL** do mesmo. Levando em consideração o âmbito da aplicação, não faria sentido fazer uma extração sem a introdução de *palavras chave*, figura 5.6, uma vez que o objetivo é a pesquisa destas dentro do arquivo extraído. Assim, em ambas as *vistas* é solicitado ao utilizador que introduza uma ou mais *palavras chave*. É disponibilizada a opção de subtração de *palavras chave*, para tal apenas é necessário clicar no "X" que encontra na *tag* da *palavra chave*. Caso o utilizador não insira um **URL** ou não fizer o *upload* de um ficheiro, na *vista* correspondente, é devolvida uma mensagem de alerta e a aplicação não inicia sem a correção necessária.

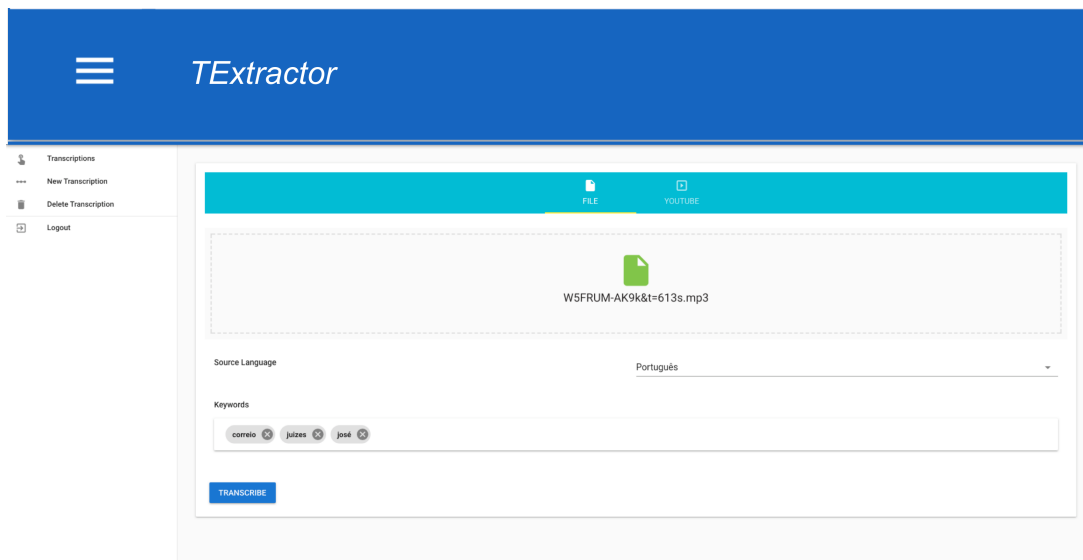


FIGURA 5.6: Introdução de *palavras chave*

5.2.1.4 Análise da extração

A aplicação desenvolvida, além de enviar um email ao utilizador *logado*, como já foi referido na secção 5.2.1.1 também disponibiliza uma análise ao arquivo extraído, para tal, o utilizador apenas necessita clicar no arquivo a analisar na lista da *vista* em questão (Figura 5.5). Esta funcionalidade apresenta ao utilizador o arquivo de texto com as *palavras chave* sombreadas. Ao passar o cursor sobre as palavras sombreadas é disponibilizada informação referente a essa *palavra chave* (Figura 5.7). Entre a informação disponibilizada consta o nível de confiança da extração da palavra assim como o registo temporal dentro do texto de quando a palavra foi proferida. O nível de confiança é disponibilizado pelo extractor, este encontra-se no intervalo [0-1] e refere-se ao nível de confiança com que a palavra foi extraída, ou seja a percentagem de certeza do algoritmo de extração quanto á extração da palavra do arquivo fonte.



<https://www.youtube.com/watch?v=Stgk6ynHUIc&t=76s>

e **hoje** na minha **poesia** com rapadura vou falar de **saudade saudade** que talvez seja o sentimento mais poético que existe mas até do que o amor porque é nem todo mundo sentiu amor à vida mas **saudade** sim por isso que eu disse um poema em cima de moto que lhes quer saber quanto custam a **saudade** tem amor queira bem e viva ausente diz assim a **saudade** de alguém que foi embora de um amigo me um amor de um parente de alguém que não está mais entre a gente com o peito do adoentado há uma chora efeito gripe que de noite só piora não é bem mais do que vinte dor de dente juliana não é do cabra mais valente sem sentir pena dó nem piedade quer sair **saudade** tem amor queira bem ele ausente tanto amor no meu peito estocado esperando por você quem já passou **saudade** quer se despediu o vez por outra me pergunto agoniado se a **saudade** mora mesmo no passado porque **hoje** o olho mais para trás do que para frente para lembrar que jacintha infelicidade quer saber quanto custa **saudade** hor é queira bem vive ausente a **saudade** a **saudade** observando a minha dor me levou para mesa de cirurgia sem ao menos aplicar anestesia segurou meu coração e arrancou nessa hora até **saudade** choro pensar bem no todo o mal que faz para gente viu seu nome gravado em ferro quente e deu remorso do tamanho da crueldade quer saber quanto custam a **saudade** tem amor se a **saudade** além de ferir mata se com certeza já teria falecido armas nem morto eu teria esquecido nem a morte dava fim a esse impasse cada vez que a minha alma se lembrasse cep adeus de forma insistente ele desse um jeitinho diferente para juntar nas dois por toda eternidade porque tem a saber quanto custa uma **saudade** tem amor teira bem e viva ausente a

Created on Today at 9:57 AM

FIGURA 5.7: Documento com as *palavras chave* encontradas

5.2.2 Conclusão

Resumidamente, como exposto no conteúdo da secção [5.2](#), esta aplicação tem por objetivo receber um arquivo de áudio/vídeo ou um [URL](#) e um conjunto de *palavras chave* fornecido pelo utilizador, o qual também seleciona o idioma para o qual o arquivo vai ser extraído. Preenchidos estes requisitos, a aplicação faz a extração do conteúdo do arquivo fonte para texto. Este arquivo de texto se contiver uma ou mais *palavras chave* será guardado para posterior análise caso seja desejável, se o utilizador preferir apagar o arquivo, também tem essa funcionalidade à sua disposição. Sempre que a aplicação encontra *palavras chave* no arquivo extraído é enviado um email de alerta para o utilizador, com a lista de *palavras chave* encontradas, para o endereço de email com que que está “logado”. A aplicação possibilita várias extrações em simultâneo e permite ao utilizador acompanhar o desenrolar dos processos e verificar o estado dos mesmos, como referido na secção [5.2.1.2](#).

Capítulo 6

Conclusão

As ciber ameaças existem há muitos anos, mas é nestes últimos que temos assistido a um maior desenvolvimento das mesmas e a um maior envolvimento das organizações na gestão dessas ameaças. Se antes a área da segurança era considerada de IT, ela começa cada vez mais a ser vista como uma área transversal que requer o suporte da gestão de topo por forma a ser endereçada da melhor forma. Infelizmente é uma área muito difícil, na qual os “maus” tendem a estar sempre um passo à frente dos “bons”. Para lidar com o problema devem ser movidos esforços coordenados e capazes de potenciar a adoção das melhores tecnologias e processos bem como educar e sensibilizar as pessoas para o problema. Só assim se estará mais preparado para lidar com as técnicas, tecnologias e procedimentos usados pelos atores maliciosos.

No que diz respeito às tecnologias utilizadas para lidar com as ciber ameaças temos assistido a uma grande evolução. Não se tratam de tecnologias que substituam outras, mas sim um conjunto que endereça um problema multifacetado. Firewalls, sistemas anti-spam, sistemas antivírus, SIEMs, IDS são exemplos dessas tecnologias. Mais recentemente e percebendo que para melhor endereçar o problema é necessária a recolha e análise de informação interna, externa e de contexto, surgiram plataformas de threat intelligence (TIP - Threat Intelligence Platforms). Estas plataformas alimentam-se de informação, muita da qual fornecida/disponível online, fazendo por isso recolha ativa de dados em fontes abertas e gerando inteligência a partir das mesmas, designando vulgarmente por **OSINT**. De entre as fontes comuns de informação temos páginas Web, blogs, redes sociais, documentos online.

Neste trabalho apresentamos a implementação de uma ferramenta para a recolha e análise de dados áudio e vídeo em fontes **OSINT**. A ferramenta designa-se por TExtractor e trata-se de uma aplicação Web que permite monitorizar a ocorrência de palavras chave em fontes áudio e vídeo relacionado com atividades ciber criminosas. A TExtractor contempla assim a extração de texto a partir do áudio/vídeo, a indexação da informação em bases de dados, a pesquisa de palavras chaves pré-definidas e ainda o envio de alertas sempre que são encontradas as palavras chave. Na base da sua implementação está um estudo de eficácia das soluções de extração de texto a partir de fonte áudio/vídeo. Esse estudo foi descrito no capítulo 3 e apresentados os resultados no capítulo 4.

Realizada uma análise final à ferramenta verificou-se que os resultados de eficácia na transcrição e detecção de palavras chave vão de encontro aos apresentados no capítulo 4. Como se pode ver na Tabela 6.1, os valores das extrações efetuadas pela aplicação são similares aos valores obtidos pelas ferramentas na fase de testes. A tabela mostra três extrações de um audiobook e de uma música considerando três idiomas diferentes. Em linha com os resultados dos testes, a ferramenta tem maior precisão na detecção de áudio/vídeo relatado num tom moderado, em inglês e sem ruídos de fundo. Quando a frequência do áudio é elevada existe muito “ruído”, como é o caso da música, onde as ferramentas revelam maior dificuldade no reconhecimento/extração de texto.

TExtractor			
	Inglês	Português	Espanhol
AudioBook			
Extração 1	64.10%	22.25%	55.74%
Extração 2	61.95%	22.87%	53.93%
Extração 3	61.95%	22.45%	55.74%
Musica			
Extração 1	5.03%	13.73%	7.75%
Extração 2	5.56%	9.15%	7.75%
Extração 3	5.56%	10.46%	6.71%

TABELA 6.1: Análise à eficácia da solução final - **TExtractor**

Através da Tabela 6.1 é também possível verificar que os valores agrupados por tipo (fontes) e idioma da extração, são bastante parecidos sendo a amplitude da sua diferença de cerca de 3%. A diferença entre estes valores e os valores da análise da ferramenta não se trata de uma incoerência, o facto do extractor ser o mesmo e os ficheiros fonte também, poderia levar a pensar em tal situação. Esta diferença é da responsabilidade do extractor, ou seja, os valores de similaridade obtidos formam sempre diferentes, a cada extração a ferramenta extrai um texto com uma leve diferença, isto devido ao algoritmo utilizado pela ferramenta. Isto revela que a ferramentas de extração de texto a partir do áudio/vídeo são consistentes. A taxa de palavras devidamente transcritas oscila entre os 60% e os 70% para áudio/vídeo sem ruído e em frequência moderada. A melhoria desta taxa está estritamente relacionada com as ferramentas utilizadas. No que diz respeito à detecção de palavras chaves feita pelo TExtractor, verificamos que, estando todas as palavras chaves transcritas, a ferramenta foi capaz de as encontrar na sua totalidade.

Tratando-se de um trabalho baseado na eficácia de outras ferramentas é fundamental dizer que o TExtractor terá a ganhar com a melhoria dessas mesmas ferramentas. Ainda assim, considerando os resultados obtidos, considera-se que o TExtractor é uma ferramenta poderosa que poderá ser usada para monitorizar a ocorrência de palavras chave em conteúdos áudio/vídeo alertando o utilizador para a sua ocorrência. Considerando o volume de informação gerado e consumido ter um automatismo que faça 60% a 70% do trabalho manual é uma mais valia.

Como trabalho futuro é fundamental referir a importância da melhoria das ferramentas de extração de texto a partir de fontes áudio/vídeo. No âmbito do TExtractor é também importante alargar a sua área de aplicação explorando, por exemplo, a sua utilização em contextos de

streaming direto. Uma funcionalidade que estava contemplada, mas por falta de tempo não foi testada e implementada tem a ver com a tradução automática do áudio/vídeo para um idioma mais universal. Face aos resultados verifica-se que tal funcionalidade deverá ser implementada. Uma tradução automática com precisão do idioma original do áudio/vídeo para inglês irá com certeza contribuir para um aumento da precisão dos extratores de texto e consequentemente uma maior cobertura das palavras chave a pesquisar.

O TExtractor foi desenvolvido para atuar como ferramenta **OSINT**, isto é, poder receber dados de fontes áudio/vídeo abertas e a partir dessa informação ser capaz de detetar sinais de ciberameaças. Importa, no entanto, dizer que a sua área de aplicação pode ser muito mais vasta. Uma ferramenta deste tipo poderá ser usada por exemplo para monitorizar a promoção de marcas ou produtos em canais áudios (e.g. rádio) ou vídeo-áudio (e.g. televisão ou canais online).

Bibliografia

- [1] "60 seconds on internet 2017", [online] <https://www.inverse.com/article/33612-here-s-everything-that-happens-in-one-minute-on-the-internet>. [acedido em 21 11 2017].
- [2] "c4.5", [online] available: https://en.wikipedia.org/wiki/c4.5_algorithm. [acedido em 22 10 2017].
- [3] "copleaks", [online] <https://copleaks.com/compare>. [acedido em 28 11 2017].
- [4] "cyber inteligenca", [online] https://en.wikipedia.org/wiki/cyber_threat_intelligence. [acedido em 29 10 2017].
- [5] "cyber inteligenca", [online] <https://www.tripwire.com/state-of-security/security-data-protection/introduction-cyber-intelligence/>. [acedido em 19 11 2017].
- [6] "cyber seguranca", [online] https://en.wikipedia.org/wiki/computer_security. [acedido em 29 10 2017].
- [7] "cyber threat inteligenca", [online] <https://www.tripwire.com/state-of-security/security-data-protection/cyber-threat-intelligence/>. [acedido em 29 10 2017].
- [8] "deep web", [online] available: <https://www.quora.com/what-is-the-deep-dark-web-and-how-do-you-access-it>. [acedido em 29 10 2017].
- [9] "hacker", [online] available: <https://en.wikipedia.org/wiki/hacker>. [acedido em 29 10 2017].
- [10] "hidden markov models", [online] available: <https://web.stanford.edu/~jurafsky/slp3/9.pdf>. [acedido em 23 10 2017].
- [11] "icon", [online] available: <http://www.abbreviations.com/term/187457>. [acedido em 14 10 2017].
- [12] "mongodb", [online] available: <https://www.mongodb.com/blog/post/storing-large-objects-and-files-in-mongodb>. [acedido em 9 10 2017].
- [13] "mongodb vs mysql comparison: Which database is better?", [online] available: <https://hackernoon.com/mongodb-vs-mysql-comparison-which-database-is-better-e714b699c38b>. [acedido em 9 10 2017].
- [14] "mvc", [online] <https://blog.codinghorror.com/understanding-model-view-controller/>. [acedido em 28 11 2017].

- [15] "naive bayes", [online] available:http://scikit-learn.org/stable/modules/naive_bayes.html. [acedido em 22 10 2017].
- [16] "naive bayes", [online] available:https://www.maxwell.vrac.pu-rio.br/9947/9947_5.pdf. [acedido em 22 10 2017].
- [17] "node.js", [online] available: <https://nodejs.org/en/>. [acedido em 9 10 2017].
- [18] "opensource", [online] available: <https://opensource.com/resources/what-open-source>. [acedido em 29 10 2017].
- [19] "opensource", [online] available: <https://us.norton.com/internetsecurity-malware.html>. [acedido em 29 10 2017].
- [20] "osint", [online] available: <https://brightplanet.com/2013/04/what-is-osint-and-how-can-your-organization-use-it/>. [acedido em 28 10 2017].
- [21] "osint", [online] available: <https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html>. [acedido em 28 10 2017].
- [22] "osint", [online] available: <http://www.expertsystem.com/what-is-osint/>. [acedido em 28 10 2017].
- [23] "osint", [online] https://en.wikipedia.org/wiki/open-source_intelligence. [acedido em 29 10 2017].
- [24] "osintframework", [online] <http://osintframework.com/>. [acedido em 29 10 2017].
- [25] "restfull", [online] available: https://en.wikipedia.org/wiki/representational_state_transfer. [acedido em 15 10 2017].
- [26] "speech-to-text", [online] <https://speech-to-text-demo.ng.bluemix.net/>. [acedido em 28 11 2017].
- [27] "speechlogger", [online] <https://speechlogger.appspot.com/en/>. [acedido em 28 11 2017].
- [28] "speechnotes", [online] <https://speechnotes.co/>. [acedido em 28 11 2017].
- [29] "stellar", [online] available: <https://stellar-framework.com/guide/actions.html>. [acedido em 15 10 2017].
- [30] "stellar overview", [online] available: <https://stellar-framework.com/guide/>. [acedido em 15 10 2017].
- [31] "streaming", [online] <https://pt.wikipedia.org/wiki/streaming>. [acedido em 12 11 2017].
- [32] "threat intelligence platform", [online] https://en.wikipedia.org/wiki/threat_intelligence_platform. [acedido em 29 10 2017].
- [33] "v8", [online] available: <https://developers.google.com/v8/v8>. [acedido em 10 10 2017].

- [34] "vue.js: a 3-minute interactive introduction", [online] available: <https://medium.freecodecamp.org/learn-basic-vue-js-crash-course-guide-vue-tutorial-e3da361c635>. [acedido em 8 10 2017].
- [35] "vue.js", [online] available: <https://vuejs.org/>. [acedido em 8 10 2017].
- [36] "vuetify.js", [online] available: <https://vuetifyjs.com/>. [acedido em 9 10 2017].
- [37] "watson tone analyser", [online] <https://www.ibm.com/blogs/watson/2017/04/watson-tone-analyzer-7-new-tones-help-understand-customers-feeling/>. [acedido em 19 11 2017].
- [38] "wdiff", [online] <https://www.gnu.org/software/wdiff/>. [acedido em 28 11 2017].
- [39] "web 2.0", [online] <https://techterms.com/definition/web20>. [acedido em 29 10 2017].
- [40] "web speech api demo", [online] <https://www.google.com/intl/en/chrome/demos/speech.html>. [acedido em 28 11 2017].
- [41] "websocket", [online] available: <https://en.wikipedia.org/wiki/websocket>. [acedido em 15 10 2017].
- [42] "will vue.js become a giant like angular or react?", [online] available: <https://10clouds.com/blog/vuejs-angular-react/>. [acedido em 8 10 2017].
- [43] Hanane EZZIKOURI Mohammed ERRITALI Badr HSSINA, Abdelkarim MERBOUHA. A comparative study of decision tree id3 and c4.5.
- [44] C. Best. Osint, the internet and privacy. In *2012 European Intelligence and Security Informatics Conference*, pages 4–4, Aug 2012.
- [45] M. Chen, X. Xu, X. Yu, and X. Zhu. The src-b speech-to-text systems for oc16 chinese-english mix-asr challenge. In *2016 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*, pages 89–94, Oct 2016.
- [46] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi. Assessment of speech intelligibility in parkinson #x2019;s disease using a speech-to-text system. *IEEE Access*, 5:22199–22208, 2017.
- [47] Y. H. Ghadage and S. D. Shelke. Speech to text conversion for multilingual languages. In *2016 International Conference on Communication and Signal Processing (ICCSPP)*, pages 0236–0240, April 2016.
- [48] T. Giannakopoulos, A. Pikrakis, and S. Theodoridis. A multimodal approach to violence detection in video sharing sites. In *2010 20th International Conference on Pattern Recognition*, pages 3244–3247, Aug 2010.
- [49] Michael Glassman and Min Ju Kang. Intelligence in the internet age: The emergence and evolution of open source intelligence (osint). *Computers in Human Behavior*, 28(2):673 – 682, 2012.

- [50] Robert Guerin. *MIDI POWER! The Comprehensive Guide*. Course Technology PTR, 2009.
- [51] Chunneng Huang, Tianjun Fu, and Hsinchun Chen. Text-based video content classification for online video-sharing sites. *Journal of the American Society for Information Science and Technology*, 61(5):891–906, 2010.
- [52] W. Jiang, B. Zhou, and M. Zhang. Architecture analysis and implementation of 3d theatre display system based on node.js. In *2015 International Conference on Network and Information Systems for Computers*, pages 496–499, Jan 2015.
- [53] L. Kaushik, A. Sangwan, and J. Hansen. Automatic sentiment detection in naturalistic audio. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, PP(99):1–1, 2017.
- [54] S. Lee and T. Shon. Open source intelligence base cyber threat inspection framework for critical infrastructures. In *2016 Future Technologies Conference (FTC)*, pages 1030–1033, Dec 2016.
- [55] L. Liang, L. Zhu, W. Shang, D. Feng, and Z. Xiao. Express supervision system based on nodejs and mongodb. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pages 607–612, May 2017.
- [56] Allan Liska. *Building an Intelligence-Led Security Program*. Syngress Media, 2014.
- [57] A. Maatouki, J. Meyer, M. Szuba, and A. Streit. A horizontally-scalable multiprocessing platform based on node.js. In *2015 IEEE Trustcom/BigDataSE/ISPA*, volume 3, pages 100–107, Aug 2015.
- [58] M. N. S. Miazi, M. M. A. Pritom, M. Shehab, B. Chu, and J. Wei. The design of cyber threat hunting games: A case study. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–6, July 2017.
- [59] M. M. Patil, A. Hanni, C. H. Tejeshwar, and P. Patil. A qualitative analysis of the performance of mongodb vs mysql database based on insertion and retrieval operations using a web/android application to explore load balancing #x2014; sharding in mongodb and its advantages. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pages 325–330, Feb 2017.
- [60] S. Patra, N. C. Naveen, and O. Prabhakar. An automated approach for mitigating server security issues. In *2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pages 1075–1079, 2016.
- [61] Wei Qi, Lie Gu, Hao Jiang, Xiang-Rong Chen, and Hong-Jiang Zhang. Integrating visual, audio and text analysis for news video. In *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, volume 3, pages 520–523 vol.3, 2000.
- [62] B. Raghavendhar Reddy and E. Mahender. Speech to text conversion using android platform. *International Journal of Engineering Research and Applications (IJERA)*, 3:253–258, January-Febr.

- [63] R. S. Rocha, P. Ferreira, I. Dutra, R. Correia, R. Salvini, and E. Burnside. A speech-to-text interface for mammoclass. In *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1–6, June 2016.
- [64] E. Saranya, B. B. Sam, and R. Sethuraman. Speech to text user assistive agent system for impaired person. In *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, pages 221–226, Aug 2017.
- [65] R. Shadiev, B. L. Reynolds, Y. M. Huang, N. Shadiev, W. Wang, R. Laxmisha, and W. Wan-
napipat. Applying speech-to-text recognition and computer-aided translation for supporting
multi-lingual communications in cross-cultural learning project. In *2017 IEEE 17th Inter-
national Conference on Advanced Learning Technologies (ICALT)*, pages 182–183, July
2017.
- [66] Jae-Chang Shim, C. Dorai, and R. Bolle. Automatic text extraction from video for content-
based annotation and retrieval. In *Proceedings. Fourteenth International Conference on
Pattern Recognition (Cat. No.98EX170)*, volume 1, pages 618–620 vol.1, Aug 1998.
- [67] B. Tombaloğlu and H. Erdem. A svm based speech to text converter for turkish language.
In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, pages
1–4, May 2017.