

MESTRADO
MULTIMÉDIA - ESPECIALIZAÇÃO EM TECNOLOGIAS

ESTUDO PARA O DESIGN DE UM OBSERVATÓRIO MEDIA ONLINE

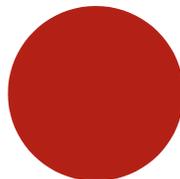
Jéssica Pereira da Silva

M

2018

FACULDADES PARTICIPANTES:

**FACULDADE DE ENGENHARIA
FACULDADE DE BELAS ARTES
FACULDADE DE CIÊNCIAS
FACULDADE DE ECONOMIA
FACULDADE DE LETRAS**





Estudo para o Design de um Observatório Media Online

Jéssica Pereira da Silva

Mestrado em Multimédia, Universidade do Porto

Orientador: Sérgio Sobral Nunes

Coorientador: Tiago Devezas

Junho, 2018

© Jéssica Pereira da Silva, 2018

**Estudo para o Design de um Observatório
Media Online**

Jéssica Pereira da Silva

Mestrado em Multimédia, Universidade do Porto

Abstract

The Internet is currently one of the most used means of accessing news and information. The storage of data provided by the different online channels is increasingly facilitated, which allows the creation of large databases. However, analyzing these collections can be complicated and a time-consuming job. In this case, turning that data into visualizations would make this process easier.

The main objective of this study is to make an analysis that supports the design of an online media observatory - *MediaViz*. With this analysis it will be possible to design a platform that presents the data visually clear to its users.

In order to reach such objectives a non-experimental study was developed. Firstly, an analysis grid was developed and it was applied to the different existing platforms. This grid was then used as a basis for developing a questionnaire that was sent to people representing the target audience (media students, researchers and journalists). The data collected are related to the demography and the application of the developed grid.

The results regarding the application of the grid in the different platforms show that, regarding the levels of analysis, the platforms Europe Media Monitor, MediaCloud, PageOneX, MediaViz and PickaNews stand out in comparison to the rest. Here, the most commonly used levels of analysis are: the collection, the term, and the phrase. As far as the analysis dimensions are concerned, the Europe Media Monitor, MediaCloud, PageOneX and GDELT Project platforms give priority to frequency, temporal and geographical dimensions. The results of the questionnaire reveal that the most important levels of analysis are the sources and the documents. Regarding the dimension of analysis, the participants of the questionnaire gave greater importance to topics and non-geographic entities.

This study presents data relevant to the design of a data visualization platform

that allows its user to reach its objectives and to use the platform regularly. This data will be used for the future development of the *MediaViz* platform but may also be used in the development of other platforms. Regarding the future development of the platform, it will be important to make usability studies, design wireframes, test these wireframes with users representing the target audience and then develop the high fidelity prototype.

Keywords: data visualization, online media observatory, analysis grid

Resumo

A Internet é, atualmente, um dos meios mais utilizados para aceder a notícias e informação. O armazenamento dos dados disponibilizados pelos diferentes canais online é cada vez mais facilitado, o que permite a criação de grandes bases de dados. No entanto, a análise desses dados pode ser um trabalho complicado e demorado. Neste caso, transformar esses dados em visualizações facilitaria esse processo.

Este estudo tem como objetivo primordial fazer uma análise que sustente o design de um observatório media online - *Media Viz*. Com essa análise será possível desenhar uma plataforma que apresente visualmente os dados de forma clara para os seus utilizadores.

Para atingir tais objetivos foi desenvolvido um estudo não experimental. Primeiramente, foi desenvolvida uma grelha de análise e esta foi aplicada às diferentes plataformas já existentes. De seguida, esta grelha serviu de base para o desenvolvimento de um questionário que foi enviado para pessoas representativas do público-alvo (estudantes da área dos media, investigadores e jornalistas). Os dados recolhidos estão relacionados com a demografia e com a aplicação da grelha desenvolvida.

Os resultados relativos à aplicação da grelha nas diferentes plataformas revelam que, quanto aos níveis de análise, as plataformas Europe Media Monitor, MediaCloud, PageOneX, MediaViz e PickaNews destacam-se comparativamente às restantes. Aqui, os níveis de análise mais utilizados são: a coleção, o termo e a frase. Já quanto às dimensões de análise, destacam-se as plataformas Europe Media Monitor, MediaCloud, PageOneX e GDELT Project que dão primazia às dimensões da frequência, temporal e geográfica. Os resultados do questionário revelam que os níveis de análise mais importantes são as fontes e os documen-

tos. Quanto à dimensão de análise, os participantes do questionário deram maior importância aos tópicos e às entidades não geográficas.

Este estudo apresenta dados relevantes para o design de uma plataforma de visualização de dados que permita que o seu utilizador atinja os seus objetivos e que utilize a plataforma regularmente. Estes dados serão utilizados para o futuro desenvolvimento da plataforma *MediaViz* mas também poderá ser utilizado no desenvolvimento de outras plataformas. Relativamente ao futuro desenvolvimento da plataforma, será primordial fazer estudos de usabilidade, proceder ao desenho de wireframes, testar esses wireframes com utilizadores representativos do público-alvo e desenvolver, depois, o protótipo de alta fidelidade..

Palavras-chave: visualização de dados, observatório media online, grelha de análise

Agradecimentos

Agradeço aos meus orientadores, Sérgio Nunes e Tiago Devezas, por todo o acompanhamento, por todas as sugestões, motivação e disponibilidade.

Agradeço e dedico este trabalho, de forma especial, aos meus pais e ao meu irmão por todo o apoio ao longo destes anos, sem eles nada seria possível.

A ti e para ti, melhor amigo, estejas tu onde estiveres, meu tio Luciano.

Jéssica Silva

Conteúdo

1	Introdução	1
1.1	Motivação e Objetivos	3
1.2	Estrutura do documento	4
2	Visualização de Informação	6
2.1	Principais Características	6
2.2	Técnicas	10
3	Fluxos de Notícias	19
3.1	Europe Media Monitor	20
3.2	Pickanews	22
3.3	Google Trends	23
3.4	Media Cloud	26
3.5	PageOneX	34
3.6	The GDELT Project	36
3.7	Comparação	40
3.8	Plataforma MediaViz	41
4	Métodos	45
4.1	Desenho do Estudo	45
4.2	Grelha de Análise	47
4.2.1	Aplicação da grelha	55
4.2.2	Análise comparativa	68
4.2.3	Síntese	72
4.3	Questionário	72
4.3.1	Participantes	72

4.3.2	Instrumentos	74
4.3.3	Procedimentos	75
4.3.4	Análise dos dados	76
4.3.5	Caraterização da amostra	76
4.3.6	Síntese	78
5	Resultados	79
5.1	Resultados do questionário	79
5.1.1	Níveis de análise	79
5.1.2	Dimensões de Análise	81
5.2	Síntese dos resultados e discussão	88
6	Conclusões	93
6.1	Contribuições	95
6.2	Trabalho Futuro	95
A	Grelha de análise	100
B	Personas	102
C	Questionário	107
D	E-mail enviado ao público-alvo	113
E	Lista de contactos	117
E.1	Unidades de investigação	117
E.2	Redações de Órgãos de Comunicação Social	118

Lista de Figuras

2.1	Classificação das técnicas de visualização. Adaptado de Keim (2001) .	11
2.2	Notas dos alunos. Adaptado de Vaz e de Carvalho (2004) .	13
2.3	Importância dos atributos visuais (do Nascimento e Ferreira, 2012).	14
2.4	Marcas visuais mais comuns (do Nascimento e Ferreira, 2012).	15
2.5	Modelo de referência para visualização de informação. Adaptado de Vaz e de Carvalho (2004) .	16
3.1	Página inicial da plataforma da Europe Media Monitor.	20
3.2	Ferramentas da plataforma Europe Media Monitor.	21
3.3	Página inicial da plataforma Pickanews.	22
3.4	Resultados da Plataforma Pickanews.	22
3.5	Página inicial da plataforma do Google Trends.	23
3.6	Tendências de pesquisa diárias.	24
3.7	Tendências de pesquisa em tempo real.	25
3.8	Google Trends - Comparação de termos.	26
3.9	Página inicial das ferramentas da plataforma MediaCloud.	27
3.10	Dados obtidos através da ferramenta Explorer	28
3.11	Dados obtidos através da ferramenta Topic Mapper.	30
3.12	Dados obtidos através da ferramenta Source Manager.	32
3.13	Dados obtidos através da ferramenta Source Manager.	33
3.14	Página inicial da plataforma PageOneX.	34
3.15	Criação de novo tópico através da plataforma PageOneX.	35
3.16	Página inicial da plataforma The GDELT Project.	36
3.17	Possibilidades de análise do The GDELT Project.	36
3.18	Ferramenta do Serviço de Análise GDELT.	38

3.19	Ferramenta Google BigQuery	39
3.20	Ferramenta de exportação dos dados.	39
3.21	Página inicial da plataforma MediaViz.	41
3.22	Pesquisa de Palavras-chave na Plataforma.	42
3.23	Pesquisa de Fontes na Plataforma.	43
3.24	Pesquisa no NewsMap da Plataforma.	43
3.25	Pesquisa de Artigos na Plataforma.	44
4.1	Primeira versão da grelha de análise.	49
4.2	Segunda versão da grelha de análise.	51
4.3	Terceira versão da grelha de análise.	52
4.4	Aproveitamento percentual das potencialidades jornalísticas da Internet.	74
4.5	Sexo dos inquiridos.	76
4.6	Faixa etária dos inquiridos.	76
4.7	Profissão dos inquiridos.	77
4.8	Instituição dos inquiridos.	77
4.9	Área científica de formação dos inquiridos.	78
5.1	Importância dos níveis de análise.	81
5.2	Importância das dimensões de análise - Coleção.	82
5.3	Importância dos tipos de entidades.	83
5.4	Importância das dimensões de análise - Fonte.	84
5.5	Importância das dimensões de análise - Documento.	85
5.6	Importância das dimensões de análise - Frase.	86
5.7	Importância das dimensões de análise - Termos.	88

Lista de Tabelas

2.1	Notas dos alunos. Adaptado de Vaz e de Carvalho (2004)	12
3.1	Comparação das diferentes plataformas.	40
4.1	Níveis de análise dos dados.	53
4.2	Dimensões de análise dos dados.	54
4.3	Níveis de análise utilizados por todas as plataformas.	70
4.4	Dimensões de análise utilizadas por todas as plataformas.	71
5.1	Resultados obtidos através da análise dos dados.	90
5.2	Níveis de análise utilizados por todas as plataformas.	92
5.3	Dimensões de análise utilizadas por todas as plataformas.	92

Abreviaturas e Símbolos

CSV	Comma-separated values
GDELT	Global Database of Events, Language, and Tone
GKG	Global Knowledge Graph
RSS	Really Simple Syndication
URL	Uniform Resource Locator

Capítulo 1

Introdução

A web é, hoje em dia, um dos meios preferenciais e mais utilizados para aceder a notícias e informações. O Anuário da Comunicação OberCom de 2016-2017 reporta que 67.4% da população portuguesa tem acesso e utiliza a Internet ([OberCom, 2017](#)). Um outro estudo, feito pelo Instituto Reuters para o Estudo do Jornalismo, demonstra que a televisão continua a ser o meio preferencial para o consumo e consulta de notícias (54.5%), mas a Internet aparece logo de seguida (31.5%) e, por último, surge a rádio (4.7%) ([Newman et al., 2017](#)). No caso das notícias existem diversos meios online, como websites de informação e blogs, que utilizam feeds RSS permitindo um acesso fácil e rápido às notícias - na íntegra ou parcialmente. Os conteúdos disponibilizados por estes canais de publicação automática podem ser facilmente armazenados usando ferramentas informáticas. Além do conteúdo propriamente dito, como o título e o sumário da notícia, os artigos publicados por esta via incluem normalmente outros dados, como, por exemplo, o URL, a data e hora de publicação, o autor, a categoria, ou o nome da fonte. Após o armazenamento, é possível realizar diversas operações sobre estes dados e, desta forma, diferentes tipos de análise.

Nestes casos, a exploração visual da informação pode ter um papel fundamental e vantajoso. Transformar essa informação em visualizações providenciaria intuições mais rápidas e eficazes, permitindo uma melhor perceção dos dados ([Keim, 2001](#)). A visualização de informação permite que o utilizador seja capaz de observar gran-

des quantidades de dados e, rapidamente, identificar relações, estruturas e padrões (Manovich, 2011). O tempo necessário para a análise diminui quando comparado com texto. Isto é, a exploração visual permite analisar grandes quantidades de dados de forma mais eficiente. Um exemplo ilustrativo é a experiência de comparar um conjunto de dados em formato de tabela e a sua representação por meio de gráficos de barras ou linhas. Através dos gráficos é possível verificar e, quase de imediato, retirar algumas informações, como a evolução dos dados. Já com uma tabela, é necessário fazer comparações que se tornam mais demoradas e nem sempre eficazes (Vaz e de Carvalho, 2004). No entanto, a escolha da representação deve sempre assentar nos objetivos de análise. Isto porque, por exemplo, para análise individual de dados, as tabelas são mais eficazes, comparativamente, com os gráficos.

Neste trabalho, pretende-se fazer um estudo para o desenho de uma plataforma web que permita aos utilizadores a exploração visual dos dados noticiosos de forma eficiente e interativa. Para proporcionar uma boa experiência de utilização, e assim potenciar a aceitação e sucesso da plataforma, o utilizador deve ser o foco principal. A aplicação de técnicas de visualização de informação e o desenvolvimento de uma plataforma com bom grau de usabilidade podem ajudar a alcançar esses objetivos.

De um modo geral, a usabilidade prende-se com três características: eficácia, eficiência e satisfação, todas elas relacionadas com a experiência do utilizador (Koivunen e May, 2002). Usabilidade pode ser definida como a facilidade com que os utilizadores aprendem a utilizar uma plataforma e se conseguem realizar as tarefas pretendidas (Carvalho, 2002).

O desenvolvimento de uma plataforma web é um trabalho iterativo que deve ser várias vezes testado e otimizado. A análise das necessidades dos utilizadores e a realização de testes com os mesmos levará a que se obtenha uma plataforma com um bom grau de usabilidade.

A principal área de estudo deste trabalho é a visualização de informação. As notícias serão o objeto de análise e é fundamental perceber de que forma os dados para essa análise podem ser apresentados ao utilizador. Desta forma, pretende-se

investigar o que pode ser feito nesse campo e o que já foi feito nas plataformas existentes. Através dos resultados obtidos e das conclusões retiradas pretende-se sustentar o desenvolvimento de uma plataforma que, por um lado, tire partido de técnicas de visualização de dados para permitir a interação e exploração de dados noticiosos e, por outro, vá de encontro às necessidades e expectativas dos utilizadores, proporcionando uma boa experiência de utilização.

1.1 Motivação e Objetivos

Quando o principal objetivo é transmitir informação aos utilizadores é necessário perceber quais as melhores formas de o fazer, pois se tal não acontecer a análise da informação pode não ser eficiente e a experiência do utilizador não será positiva. Tal pode levar a que o utilizador abandone a plataforma e procure outra que vá de encontro às suas expectativas. No caso desta dissertação, o objetivo é estudar a melhor forma de informar o desenho de uma plataforma web que permita visualizar um conjunto de notícias e dados associados, de forma a facilitar a sua exploração em diferentes dimensões, ou seja, não só a nível do conteúdo, mas também das suas propriedades temporais, por exemplo. A plataforma já existe e tem, atualmente, mais de 6 milhões de artigos recolhidos. A plataforma em questão - *MediaViz* (Devezas et al., 2016), permite o acesso a uma grande quantidade de artigos recolhidos dos feeds RSS de dezenas de meios online. Os dados recolhidos pela plataforma incluem, entre outros, data, hora, fonte, título e resumo do artigo. O público-alvo desta plataforma é, primeiramente, investigadores da área dos media e de outras ciências sociais, assim como entusiastas da área da comunicação. O objetivo é perceber quais as melhores opções ao nível visual e ao nível da informação a apresentar para, posteriormente, ser efetuado o desenho da plataforma de forma a que esta tenha utilidade para os utilizadores. Neste contexto, o estudo dos métodos de visualização de informação que são mais eficazes e a perceção das expectativas e necessidades dos utilizadores são ferramentas fundamentais para obter dados que permitirão desenhar uma plataforma eficaz e acessível, tornando a

experiência do utilizador agradável.

Desta forma, os principais objetivos deste estudo são:

1. Identificar as plataformas existentes neste campo e quais as ferramentas e informações que disponibilizam aos seus utilizadores;
2. Identificar os critérios que os utilizadores reconhecem ser os fundamentais para obterem uma boa experiência aquando da navegação neste tipo de plataformas;
3. Identificar as informações fundamentais para os utilizadores, como por exemplo: fontes, entidades, temas mencionados, entre outros.

A partir dos dados recolhidos no estudo feito para o estado da arte e, através da análise dos dados quantitativos resultantes das respostas às questões anteriores, serão tiradas conclusões fundamentais para o possível desenvolvimento do observatório media online *Media Viz*.

Será, desta forma, possível verificar o que já foi implementado, que ferramentas já estão ao dispor do utilizador e quais poderão ser as possibilidades no desenvolvimento da nova plataforma.

Para além disso, procura-se responder à seguinte questão qualitativa:

- Como poderão os resultados provenientes deste estudo ajudar na construção de um observatório media online que permita a interação e exploração visual de dados noticiosos e responda às necessidades e expectativas dos utilizadores, proporcionando uma boa experiência de utilização?

1.2 Estrutura do documento

Este trabalho está dividido em seis capítulos, entre os quais o atual - [Capítulo 1](#). No [Capítulo 2](#) e no [Capítulo 3](#) é apresentado o estado da arte, trabalhos relacionados e ferramentas relativas ao tema visualização de informação. Já no [Capítulo 4](#) são referenciados os métodos deste estudo, as técnicas e ferramentas usadas para a recolha de dados. A análise e discussão dos dados recolhidos são apresentados

no [Capítulo 5](#). O [Capítulo 6](#) apresenta as conclusões, limitações, contribuições do trabalho e o que poderá ser realizado no futuro.

Capítulo 2

Visualização de Informação

2.1 Principais Características

O desenvolvimento de algumas tecnologias levou a que grande parte da sociedade, em todo o mundo, tivesse acesso à Internet e, por sua vez, produzisse informação digital. Segundo [Manovich \(2011\)](#), os desenvolvimentos ocorridos entre 1990 e 2010 – digitalização, diminuição dos preços, expansão da capacidade dos dispositivos, o aumento do conteúdo gerado pelo utilizador e a globalização – levaram ao aumento exponencial da disponibilidade de conteúdos, principalmente digitais. Tal facilitou a pesquisa e partilha dos mesmos.

No mesmo contexto, o projeto “*How much information?*”, da Universidade da Califórnia, estimou, em 2001, que cerca de 1 exabytes de dados (1 milhão de terabytes) seriam gerados anualmente em todo o mundo ([Keim, 2001](#)). Em 2010 estimou-se que se produziria, apenas nesse ano, 1.200 exabytes de informação ([Heer et al., 2010](#)). Informações mais atuais indicam que se produz 2.5 quintilhões de bytes em dados, em média, todos os dias, segundo uma pesquisa feita pela Domo¹. [Keim \(2001\)](#) refere que, uma das explicações parte do facto do avanço das tecnologias permitir que os sistemas computacionais armazenem grandes quantidades de dados. Estes dados são registados, capturados e armazenados automaticamente através de sistemas e sensores de monitorização.

¹<https://www.domo.com/learn/data-never-sleeps-5>

No entanto, nem todos os dados armazenados são importantes, sendo necessário distinguir quais são potencialmente vantajosos e quais não terão qualquer importância (Keim, 2001).

Isto porque, a busca por informações específicas que sejam do interesse do utilizador, na enorme quantidade de dados disponíveis, é complexa (Vaz e de Carvalho, 2004). Torna-se, assim, fundamental encontrar maneira de explorar, relacionar e comunicar esses dados de forma a que sejam bem utilizados (Heer et al., 2010).

Os dados que se apresentam através de texto são menos acessíveis e práticos, pois, segundo Keim (2001), apenas podem ser visualizados 100 itens dos dados em questão, algo que o autor refere ser equiparável a uma gota no oceano quando lidamos com dados que contêm milhões de itens. A sua exploração é demorada e, geralmente, tornam-se desnecessários e as bases de dados sem relevância, pois não há seleção.

É aqui que a exploração visual dos dados tem um papel fulcral. O principal objetivo da mesma é providenciar intuições através da visualização dos dados, ou seja, apresentá-los numa forma visual, permitindo uma melhor perceção (Keim, 2001). Desta forma, o utilizador será capaz de visualizar padrões, tendências e discordâncias havendo um estímulo dos sistemas visuais, facilitando a análise e compreensão (Heer et al., 2010).

Assim, os dados em texto podem ser substituídos, através de técnicas de visualização de dados, por visualizações perceptíveis que reduzem o trabalho cognitivo necessário para realizar algumas tarefas (Keim, 2001).

Card et al. (1999) afirmam que a visualização de dados pode ser vista como “*O uso de representações visuais, interativas e suportadas por computador, de dados abstratos para ampliar a cognição.*”

Resumindo, a visualização de informação pode ser vista como o processo de traduzir o mundo em números e depois visualizar a relação entre esses (Manovich, 2011). Já que possibilita reunir grandes quantidades de dados em apenas uma imagem (Vaz e de Carvalho, 2004). Essa exploração dos dados é feita por um humano que alia a sua flexibilidade, criatividade e conhecimento geral com a

capacidade de armazenamento, cada vez maior, que existe nos dias de hoje (Keim, 2001).

A observação dos dados deve obter o mesmo resultado que uma observação direta – uma observação do mundo – em que somos capazes de detetar, através dos nossos olhos, padrões, estruturas e relações. Assim, o utilizador será capaz de observar conjuntos de dados e, rapidamente, constatar relações, estruturas e padrões (Manovich, 2011). Sendo que, “*Visualizar é algo feito naturalmente, já que a visão humana é o sentido com maior capacidade de captação de informações por unidade de tempo; é rápido e paralelo, além de ser treinado para reconhecer padrões.*” (Vaz e de Carvalho, 2004).

“*A visualização de informações e a mineração de dados podem ajudar a lidar com o fluxo de informações.*” (da Costa, 2017). Desta forma, a estreita relação da visualização de dados com a computação gráfica, interfaces homem-computador e mineração de dados permite a transformação das informações em imagens, utilizando a perceção visual para as analisar e compreender (Vaz e de Carvalho, 2004).

Das várias definições que se encontram sobre prospeção de dados destaca-se a seguinte: “*Mineração de Dados é a análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados.*” (Hand et al., 2001). Assim, com a seleção feita através da prospeção de dados é possível, através da exploração visual dos mesmos, transmiti-los de forma mais eficaz. As principais vantagens deste tipo de exploração de dados são (Keim, 2001):

- O enorme potencial que apresenta no que diz respeito à exploração de bases de dados de grandes dimensões;
- A sua utilidade quando há poucas informações sobre os dados e os objetivos de exploração são vagos;
- A facilidade em lidar com dados heterogéneos;
- Os dados tornam-se mais intuitivos e permite maior envolvimento entre os dados e o utilizador.

Segundo [Vaz e de Carvalho \(2004\)](#), este tipo de representações podem ser “*utilizadas por uma comunidade leiga e diversificada, que possui diferentes níveis de educação, capacidades e necessidades.*” Para além disso, as imagens podem ser entendidas por diversas pessoas, mesmo que não partilhem a mesma língua, o que não ocorre com informações textuais ([Ward et al., 2010](#)).

[Keim \(2001\)](#) afirma que toda esta capacidade de analisar melhor os dados leva a que a sua exploração seja mais rápida e os resultados obtidos são, normalmente, melhores, quando comparados com análise de dados textuais. Por conseguinte, há um maior grau de confiança nos mesmos e um aumento do desenvolvimento de técnicas de exploração visual da informação.

O mesmo autor defende que, a capacidade de visualização de dados permite que o utilizador ganhe uma maior perceção sobre esses, possibilitando-lhe tirar conclusões mais eficazes e gerar hipóteses. A verificação dessas hipóteses pode ser feita, de igual modo, através da exploração visual. Os objetivos e hipóteses geradas, previamente, pelo utilizador, podem ser alteradas através do envolvimento com a exploração.

O processo de visualização de informação divide-se em 3 passos ([Shneiderman, 1996](#)):

1. Visão geral;
2. Zoom e filtro;
3. Detalhes a pedido.

Desta forma, o utilizador identifica padrões interessantes, através da visão geral, e foca-se em um ou mais desses, através do zoom e filtro, tentando, depois, ter acesso a mais detalhes sobre os dados que está a analisar, através dos detalhes a pedido, sendo também apresentados subconjuntos de dados de interesse para o utilizador.

Já [Ward et al. \(2010\)](#) afirmam que o processo de visualização divide-se em cinco fases:

1. **Modelação de dados:** os dados que vão ser visualizados devem ser estruturados para facilitar a sua visualização.
2. **Seleção de dados:** envolve a identificação do subconjunto dos dados que poderão ser visualizados. Pode ser controlado pelo utilizador ou por métodos algorítmicos.
3. **Dados para mapeamento visual:** transformação dos dados para entidades gráficas. Antes do mapeamento, é feito o processamento dos dados de forma a organizar a sua apresentação.
4. **Ajuste de parâmetros visuais:** a transformação dos dados para entidades gráficas conduzem a diversas tomadas de decisão acerca dos atributos da visualização como as cores, o tamanho, os sons, a iluminação, etc.
5. **Renderização ou geração da visualização:** o processamento do produto final.

2.2 Técnicas

A evolução das tecnologias, o aumento da produção de dados e a importância da análise dos mesmos levou ao desenvolvimento de técnicas para tornar a sua visualização mais eficiente e eficaz. A grande dimensão dos dados leva a que as técnicas tradicionais de visualização – sistema de coordenadas x-y, gráficos de linhas e pontos, e histogramas – se tornem obsoletas, pois estão limitadas a conjuntos de dados de pequenas dimensões (Keim, 2001).

1. Representação:

O foco principal das técnicas de visualização de dados deverá ser representar informações tirando partido da capacidade de perceção humana, tornando a visualização eficaz e compreensível (Vaz e de Carvalho, 2004).

Segundo Heer et al. (2010) as principais questões que se devem colocar aquando da transformação dos dados para uma visualização são: quais as

questões a colocar, quais os dados mais importantes e quais as melhores configurações visuais (posição, tamanho, forma e cor).

Os mesmos autores afirmam que, o principal desafio é tornar as visualizações bem estruturadas e adequadas aos dados a serem representados.

As técnicas de representação de dados podem ser classificadas segundo três critérios, como visto na [Figura 2.1](#):

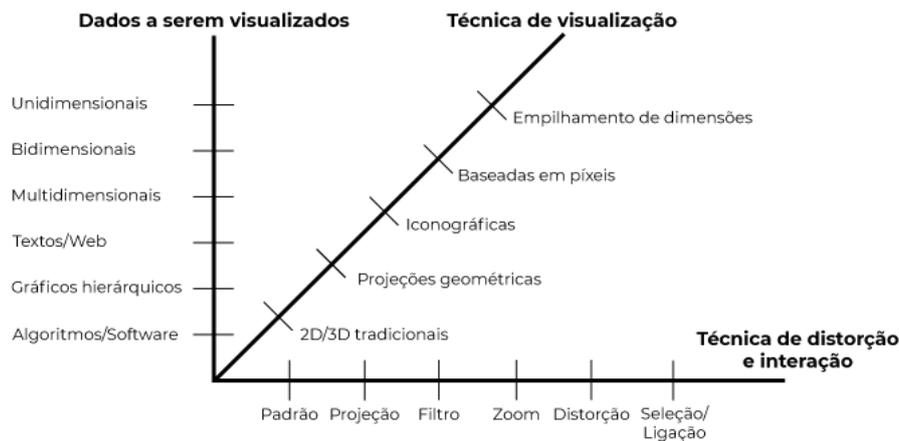


Figura 2.1: Classificação das técnicas de visualização. Adaptado de [Keim \(2001\)](#).

- (a) **Tipo de dados a representar:** podem ser unidimensionais (dados temporais), bidimensionais (mapas geográficos), multidimensionais (tabelas de relação), texto e hipertexto (artigos e documentos web), hierarquias e gráficos (chamadas telefônicas e websites), algoritmos e software (operações de depuração).
- (b) **Técnicas:** podem ser inúmeras. Estas apenas diferem na forma como organizam e representam os dados no ecrã.
- (c) **Métodos de interação e distorção:** permitem ao utilizador interagir diretamente com o que está a visualizar, ou seja com os dados. As técnicas mais utilizadas no contexto da interação e da distorção são o filtro, o zoom e a linkagem. Possibilitam que o utilizador altere a visualização de forma dinâmica, tendo em conta os seus objetivos.

2. Classificação:

do Nascimento e Ferreira (2012) afirmam que os critérios a ter em conta para classificação dessas técnicas são: a forma como é explorado o substrato visual, as marcas e as propriedades visuais do desenho e as características dos dados a serem utilizados – sendo que existem gráficos que se adequam melhor a determinado tipo de dados em detrimento de outros.

A avaliação das técnicas é feita através da comparação. Aqui, tem-se em conta a capacidade que cada uma delas apresenta, no que toca à representação de determinado tipo de dados. Algumas delas foram desenvolvidas para representar tipos específicos de dados, e outras são mais gerais (Keim, 2001).

3. Escolha:

Segundo Vaz e de Carvalho (2004), a escolha da melhor técnica parte do tipo de dados a ser representado e do processo de análise do utilizador. Isto é, qual a melhor opção para tornar a perceção dos dados quase imediata. Partindo desse princípio, o autor dá-nos dois exemplos práticos. Primeiramente, são dispostas as notas dos alunos numa tabela (Tabela 2.1), de seguida, dispostas num gráfico de linhas (Figura 2.2). A segunda opção mostrou ser mais adequada, neste caso. Era possível, quase de imediato, retirar algumas conclusões como a evolução do aluno e a comparação entre alunos.

Tabela 2.1: Notas dos alunos. Adaptado de Vaz e de Carvalho (2004).

	1º Aluno	2º Aluno	3º Aluno	4º Aluno
1º Bimestre	5.0	7.0	6.0	8.5
2º Bimestre	5.5	8.0	7.0	7.0
3º Bimestre	6.0	6.5	8.5	6.5
4º Bimestre	7.0	4.0	9.5	5.5

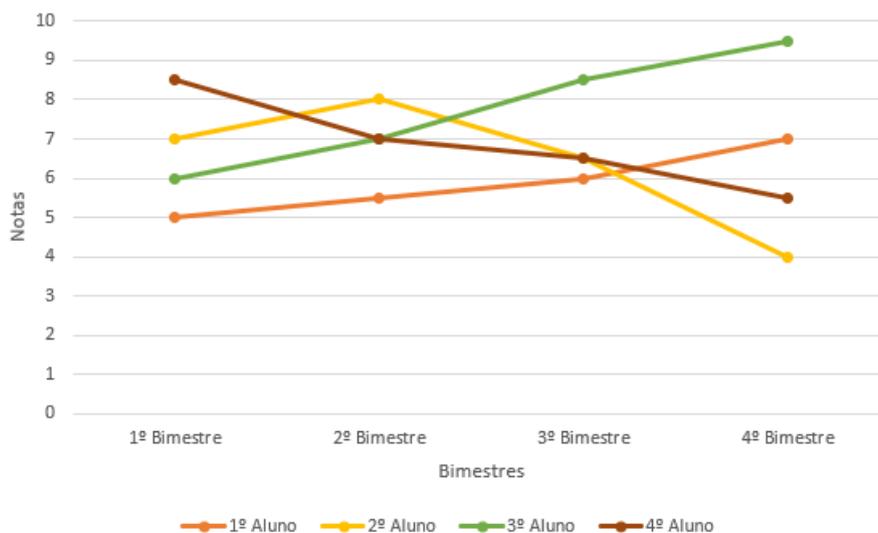


Figura 2.2: Notas dos alunos. Adaptado de [Vaz e de Carvalho \(2004\)](#).

Assim, podemos verificar que a escolha de uma técnica para uma visualização dos dados mais eficaz é imprescindível.

Como já referido anteriormente, o principal objetivo da escolha de técnicas de visualização é tornar a visualização e análise da informação mais eficaz. Assim, podemos identificar dois conceitos essenciais: a efetividade e a expressividade ([Vaz e de Carvalho, 2004](#)).

A expressividade refere-se à apresentação dos dados que apenas são do interesse do utilizador não prejudicando a interpretação. Já a efetividade define que a visualização deve ser clara, ou seja, de fácil e rápida compreensão por parte do utilizador e induzir a uma quantidade inferior de erros, comparativamente a outras formas de visualização.

Segundo os autores, para obter tais características é necessário:

- Colocar dados suficientes na visualização de forma a contextualizar as informações mais importantes;
- Associar atributos de maior importância dos dados a atributos visuais de maior clareza facilitando a perceção dos dados e dos padrões existentes ([Figura 2.3](#));

- Procurar a simplicidade, permitir diferentes formas de interação com a visualização.

Percepção	Dados Quantitativos	Dados Ordinais	Dados Nominais
Maior	Posição	Posição	Posição
^	Comprimento	Densidade	Croma de cor
	Ângulo	Saturação de cor	Textura
	Inclinação	Croma de cor	Conexão
	Área	Textura	Envolvimento
	Volume	Conexão	Densidade
	Densidade	Envolvimento	Saturação de cor
	Saturação de cor	Comprimento	Forma
	Croma de cor	Ângulo	Comprimento
	Textura	Inclinação	Ângulo
	Conexão	Área	Inclinação
v	Envolvimento	Volume	Área
Menor	Forma	Forma	Volume

Figura 2.3: Importância dos atributos visuais (do Nascimento e Ferreira, 2012).

O processo de desenvolvimento de uma visualização de dados eficaz assesta em três etapas (Figura 2.5):

- (a) **Transformar os dados brutos em tabelas de dados:** transformar o conjunto de dados brutos em relações lógicas mais estruturadas facilitando a sua compreensão. Este processo pode envolver a eliminação de dados redundantes, errados ou incompletos, tal como a inclusão de novas informações, como análises estatísticas como o caso de médias, desvio padrão, moda, etc.
- (b) **Modificar essas tabelas para estruturas visuais, através do mapeamento visual:** o mapeamento visual transforma os dados em estruturas visuais representando visualmente os dados em uso. Estas estruturas são compostas de três partes:
 - i. **Substrato espacial:** espaço para a visualização, normalmente representado por eixos. Há quatro tipos elementares de eixos:
 - U = eixo não estruturado (ou sem eixo);

- N = eixo nominal (região dividida em sub-regiões);
 - O = eixo ordenado (região dividida em sub-regiões onde a ordem das mesmas tem importância);
 - Q = eixo quantitativo (região que tem uma métrica, como valores inteiros ou reais).
- ii. **Marcas visuais:** símbolos gráficos utilizados para representar os itens de dados (Figura 2.4).

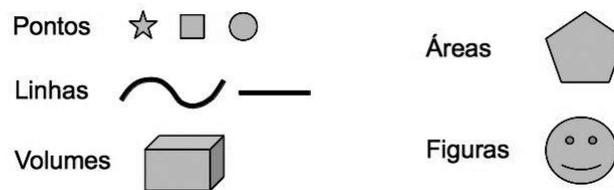


Figura 2.4: Marcas visuais mais comuns (do Nascimento e Ferreira, 2012).

- iii. **Propriedades gráficas:** atributos gráficos que caracterizam as marcas visuais. Algumas das propriedades gráficas são: posição da marca, forma, cor, tamanho, comprimento, volume, orientação, cor, brilho, textura, animação, tempo, etc. Algumas destas características são utilizadas para destacar atributos dos dados.
- (c) **Transformar as visualizações,** modificando e estendendo as estruturas visuais interativamente. Estas transformações podem ser feitas de três formas:
- **Testes de localização:** permitem obter informações adicionais sobre um item da tabela de dados;
 - **Controles de ponto de vista:** permite ampliar (zoom), reduzir e deslocar a imagem oferecendo diferentes visões;
 - **Distorção:** visa criar visualizações do tipo foco + contexto ou seja, ampliações de uma região específica em detrimento de outra.

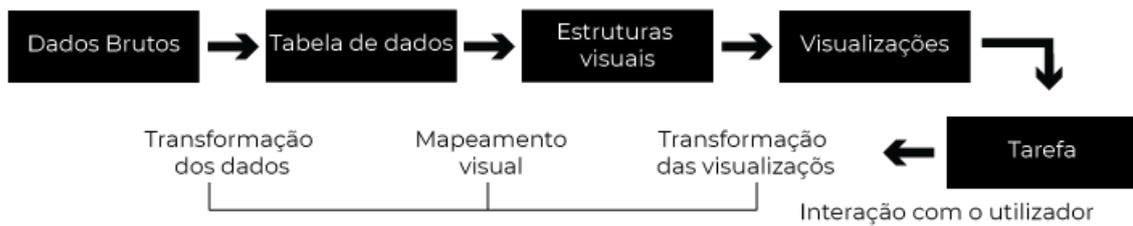


Figura 2.5: Modelo de referência para visualização de informação. Adaptado de [Vaz e de Carvalho \(2004\)](#).

4. Técnicas:

De entre as técnicas de visualização referidas por [Vaz e de Carvalho \(2004\)](#) podemos identificar:

- **Bifocal display:** aqui as informações são apresentadas em três áreas distintas. Havendo uma hierarquização da informação: a área central é aquela que contém a informação em destaque e as informações adicionais são apresentadas nas laterais. A principal vantagem é o acesso a grandes quantidades de informação. A principal desvantagem é que por ser bidimensional não é fácil integrar detalhes e contextos.
- **Flip zooming/ foco + contexto:** as informações são apresentadas de forma sequencial. Esta forma de representação apresenta uma visão geral dos dados a serem visualizados destacando uma região de interesse através de uma ampliação. A restante informação normalmente serve como contexto da informação central. As principais vantagens são: a não deformação do texto em foco, contextualização da informação principal através das informações que não se encontram em foco, de fácil implementação. No entanto, quando o número de informações a ser representado é grande esta técnica pode não ser eficaz comprometendo a boa visualização das informações. Os principais exemplos de técnicas do tipo Foco+Contexto são o Browser Hiperbólico, o Perspectiva Wall.
- **Perspective Wall:** feita através de uma visualização 3D permite a visualização de muitas informações sequenciais apresentando uma região

perspetivada. As principais vantagens desta visualização são: utilização eficiente do espaço, transição suave entre as informações e mantém sempre o contexto.

- **Hiperbolic tree / Browser hiperbólico:** permite a representação de hierarquias através de um desenho radial. A informação de maior interesse é apresentada num nó central e a restante informação (detalhes ou contexto) é apresentado em nós que vão diminuindo de tamanho até serem suprimidos (informações não tão importantes). O ponto central (de maior interesse inicialmente) pode ser modificado selecionando outro nó ou arrastando-o para o centro. As principais vantagens é a apresentação de mais informação, navegação mais efetiva e mudança rápida de foco.

5. Interação e Distorção:

O principal objetivo da interação é permitir a exploração de diferentes cenários, levando a uma melhor perceção dos dados. Desta forma, o utilizador é capaz de explorar os dados, interativamente, através de ferramentas que lhe são fornecidas, tendo acesso, algumas vezes, a novos detalhes sobre os mesmos. Existem diversas técnicas de interação (Yi et al., 2007):

- **Seleção:** marcar algo de interesse: permite ao utilizador selecionar aquilo que é do seu interesse mantendo um foco no mesmo e permitindo destacar, excluir ou modificar tal elemento.
- **Navegação/Exploração:** mostrar outra coisa: permitem ao utilizador examinar subconjuntos de dados. Algumas das técnicas utilizadas são a movimentação da câmara ou alteração do foco através de várias seleções.
- **Reconfiguração:** mostrar de forma diferente: providencia diferentes perspetivas dos dados alterando a representação espacial.
- **Codificação:** apresentar uma representação diferente: permite alterar a representação visual dos dados incluindo a aparência visual como

cor, tamanho, forma, tipo de gráfico, entre outros aspetos revelando diferentes recursos.

- **Resumo/elaboração:** apresentar mais ou menos detalhe: permite ajustar o nível de abstração dos dados apresentados. Permite alterar de uma visão geral para uma visão mais detalhada.
- **Filtro:** apresentar algo condicionalmente: permite alterar a quantidade de dados apresentados baseando-se em alguma condição. O utilizador indica a condição que pretende e apenas os dados que preenchem o critério utilizado é que são apresentados.
- **Conexão:** apresentar conteúdos relacionados: técnicas de interação que permitem evidenciar associações e relações entre dados que já são apresentados ou apresentar novos dados que sejam relevantes.

Para além das técnicas já referidas, temos ([Vaz e de Carvalho, 2004](#)):

- **Brushing:** é um exemplo de interação com transformação de dados. Possui um método interativo que permite selecionar na tela dados pontuais específicos ou subconjuntos de dados, possibilitando a identificação de suas características comuns ou a investigação de seus efeitos sobre as relações entre variáveis consideradas relevantes.
- **Pivot Tables:** é uma interação com mapeamentos visuais. Permite ao utilizador manipular o mapeamento de dados para linhas e colunas.
- **Direct Selection:** é uma interação com transformações de visualizações, que permite selecionar e destacar objetos gráficos individualmente ou em grupos.

As técnicas de distorção preservam uma visão geral dos dados, mas permitem a visualização de informação com mais detalhe. Apresentando dados com pouco detalhe e dados com muito detalhe ([Keim, 2001](#)).

Capítulo 3

Fluxos de Notícias

O desenvolvimento de visualizações de dados pode trazer bastantes desafios, principalmente quando lidamos com dados que mudam diariamente e são atualizados regularmente. Aqui podemos falar da representação visual de fluxos de notícias.

Segundo [Novak et al. \(2015\)](#) existem três desafios nesse campo:

- A enorme quantidade de artigos leva a que alguns não possam ser mostrados na íntegra ou apresentar os seus detalhes;
- O uso da linha temporal é inadequado. Já que as notícias mais antigas perdem a sua importância, ou são substituídas por novas informações que, por vezes, são completamente diferentes. Para além disso, guardar todas as notícias torna-se inviável, tanto a nível de espaço como de desempenho;
- É produzida muita redundância, já que diferentes fontes produzem conteúdos sobre assuntos iguais.

Neste âmbito, foi feito um levantamento de plataformas de fluxos de notícias que servirão de base para este estudo. A análise destas plataformas permitirá obter informação relativa aos dados que são disponibilizados aos utilizadores e de que forma esses são apresentados. Irá potenciar, de igual forma, a análise da existência de padrões ou discrepâncias entre plataformas. A seleção das plataformas partiu da análise de literatura, mais precisamente de [Novak et al. \(2015\)](#), que refere as

plataformas Europe Media Monitor, PlickaNews e Google Trends. As plataformas MediaCloud, PageOneX e GDELT Project foram selecionadas através de pesquisas feitas relativas a plataformas desta área.

3.1 Europe Media Monitor

Europe Media Monitor¹ é um agregador de notícias e um sistema de análise que recolhe informação de fontes noticiosas de mais de 70 idiomas. Este extrai a informação e determina o que está a ser noticiado e produz apresentações visuais intuitivas. Recolhe cerca de 280 000 relatórios por dia. O principal foco é fornecer atualizações frequentes e quase em tempo real de tópicos de interesse do utilizador. Este é atualizado de 10 em 10 minutos, 24 horas por dia. Na página inicial é apresentado um gráfico com as dez histórias mais citadas no momento, por língua (que pode ser alterada do lado direito da plataforma) (Figura 3.1).



Figura 3.1: Página inicial da plataforma da Europe Media Monitor. Captura de ecrã de <http://emm.newsbrief.eu>

O primeiro link de cada uma das notícias redireciona-nos para uma página interna onde podemos encontrar a evolução do tema, e as diferentes fontes que noticiaram esse assunto (Figura 3.2 (a)). O segundo link redireciona-nos para a fonte original. Os dados apresentados de cada notícia são: o título, a fonte, o número de artigos publicados sobre a mesma, a data, o número de fontes, a

¹<http://emm.newsbrief.eu>

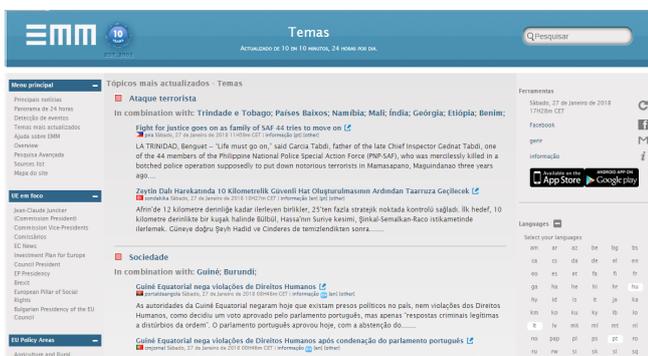
classificação e o início do texto da notícia. No lado esquerdo da plataforma são apresentados vários temas em destaque. É possível, também, verificar quais as dez histórias mais noticiadas nas últimas 24 horas (Figura 3.2 (b)), os eventos mais falados e os temas mais atualizados (Figura 3.2 (c)) e fazer pesquisa por termos (Figura 3.2 (d)). Podemos obter atualizações através de subscrição.



(a) Página com a evolução do tema e diferentes fontes que noticiaram o assunto. Captura de ecrã de <http://emm.newsbrief.eu/NewsBrief/moreclusteredition/pt/rtp-4a37da88b7f96bf3b5d129bd65aeaaf2.20180127.pt.html>



(b) Página com os temas mais atualizados nas últimas 24 horas. Captura de ecrã de http://emm.newsbrief.eu/NewsBrief/clusteredition/pt/24hrs_pt.html



(c) Página com os temas e eventos mais atualizados. Captura de ecrã de <http://emm.newsbrief.eu/NewsBrief/categoryedition/pt/Themes.html>



(d) Página de pesquisa com a palavra-chave “Portugal”. Captura de ecrã de <http://emm.newsbrief.eu/NewsBrief/dynamic?language=pt&page=1&edition=searchresults&option=&atLeast=portugal>

Figura 3.2: Ferramentas da plataforma Europe Media Monitor.

3.2 Pickanews

PickaNews² é um motor de pesquisa multimídia europeu que tem como fontes a televisão, a rádio, a web e imprensa (Figura 3.3).

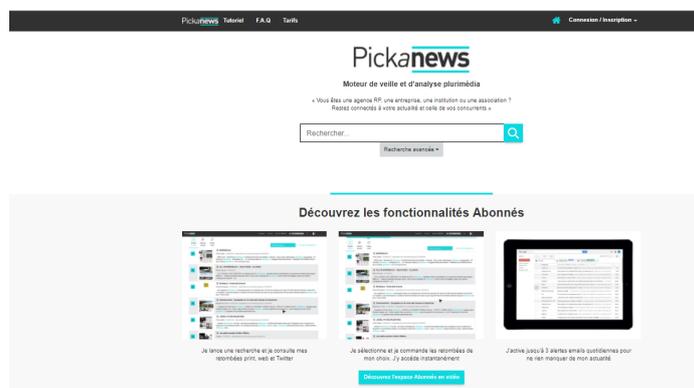


Figura 3.3: Página inicial da plataforma Pickanews. Captura de ecrã de <https://www.pickanews.com/>

É possível verificar o número de referências, a evolução do termo ao longo do tempo, a repartição pelos diferentes tipos de mídia que tem como fonte e as fontes principais onde o termo foi mencionado (Figura 3.4). A pesquisa é gratuita, no entanto o serviço de alerta, a visualização completa e de indicadores são pagos através de subscrição. Aqui, a pesquisa é feita através de um ou mais termos, possibilitando a colocação de algumas restrições.

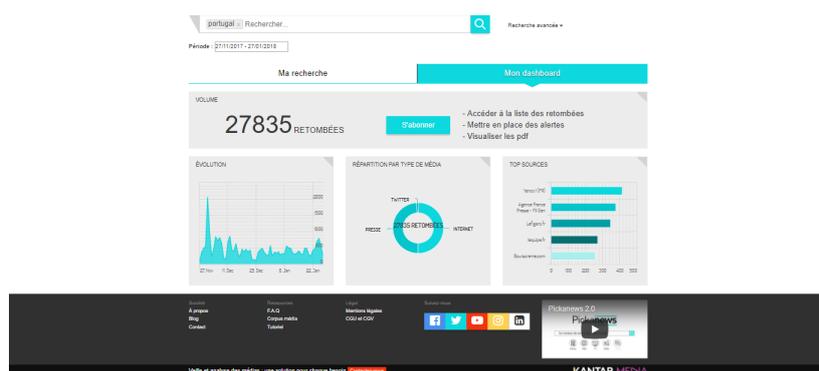


Figura 3.4: Resultados da Plataforma Pickanews. Captura de ecrã de <https://www.pickanews.com/find?q=portugal&ee=&nw=&ww=>

²<https://www.pickanews.com/>

3.3 Google Trends

A Google Trends³ é uma plataforma (Figura 3.5) pertencente à Google Inc. e é utilizada para fazer pesquisa de tendências. Mostra com que frequência um termo específico foi inserido na pesquisa do Google, em relação ao volume de procura total em várias regiões do mundo e em várias línguas. Permite aos utilizadores alterar a região, o período de tempo, a categoria e o tipo de pesquisa. Enquanto os agregadores de notícias procuram os termos em fontes noticiosas, o Google Trends faz a sua análise através das consultas feitas no Google.

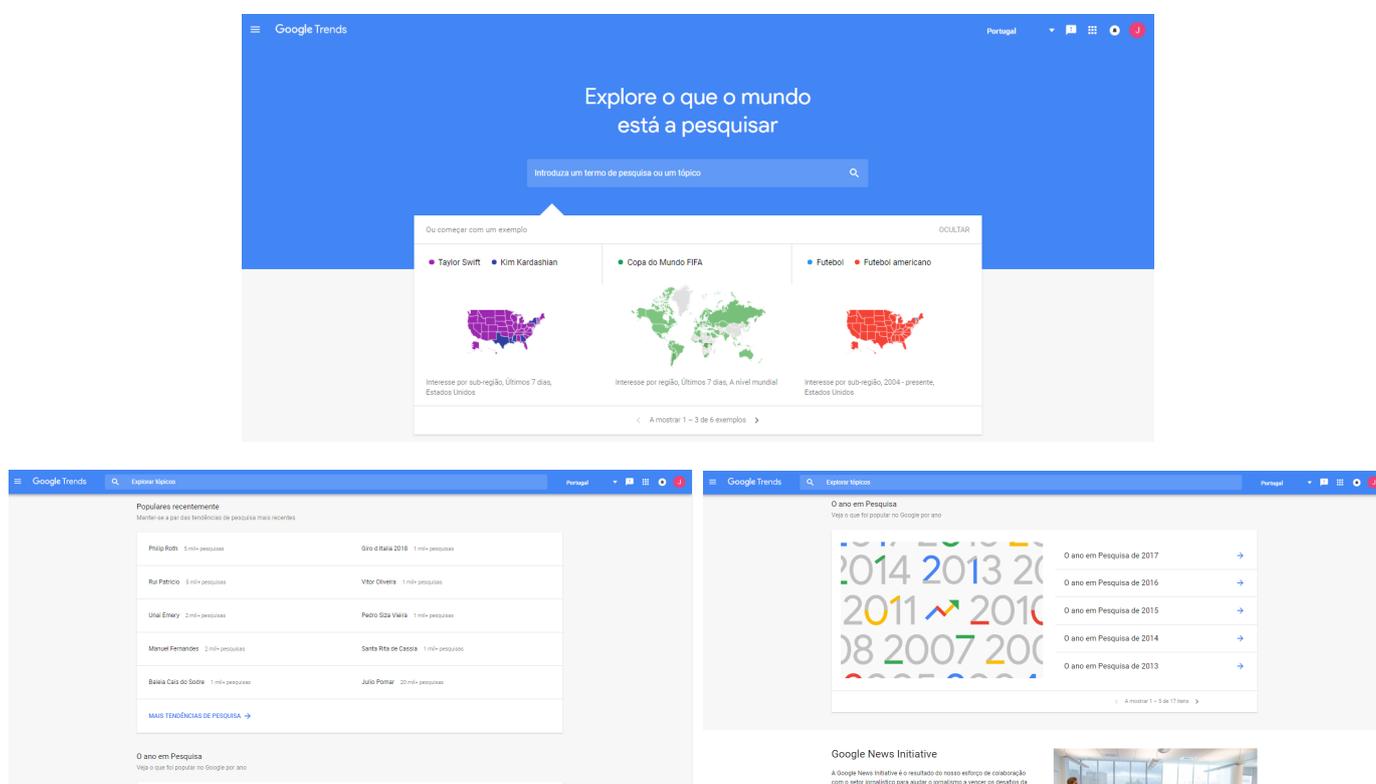


Figura 3.5: Página inicial da plataforma do Google Trends. Captura de ecrã de <https://trends.google.pt/trends/>

Na página inicial é possível fazer a pesquisa ou pesquisar os exemplos apresentados pela plataforma, ver as tendências de pesquisa mais recentes e ver os assuntos populares em cada ano (“O ano em Pesquisa”), bem como acesso ao “Google News

³<https://trends.google.pt/trends/>

Initiative”⁴. As histórias e assuntos mais populares são referentes ao país onde nos encontramos, mas esta seleção pode ser alterada na mesma página. Se acedermos ao link de uma das tendências somos redirecionados para uma página interna onde temos acesso a: tendências de pesquisa diárias e tendências de pesquisa em tempo real, onde o local geográfico pode ser novamente alterado.

- **Tendências de pesquisa diárias:** aqui são apresentadas as tendências de pesquisa nos vários dias mostrando o termo, o número de pesquisas, o título, a fonte e link para a mesma e a hora a que foi publicado o artigo (Figura 3.6).
- **Tendências de pesquisa em tempo real:** aqui são apresentadas as tendências de pesquisa nas últimas 24 horas. Temos acesso a: termos pesquisados, título, fonte e link para a fonte, hora de publicação, evolução do tema, notícias e consultas relacionadas (Figura 3.7).

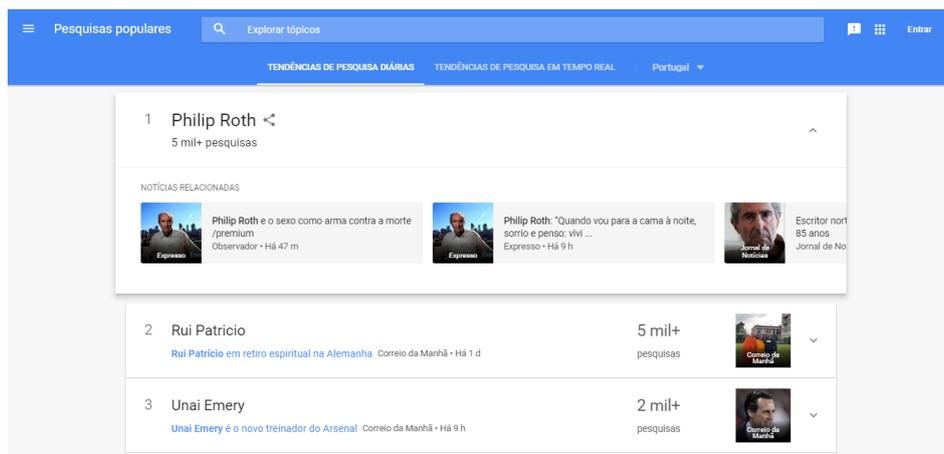


Figura 3.6: Tendências de pesquisa diárias. Captura de ecrã de <https://trends.google.pt/trends/trendingsearches/daily?geo=PT>

⁴<https://newsinitiative.withgoogle.com/training/lessons?tool=Google%20Trends&image=trends>

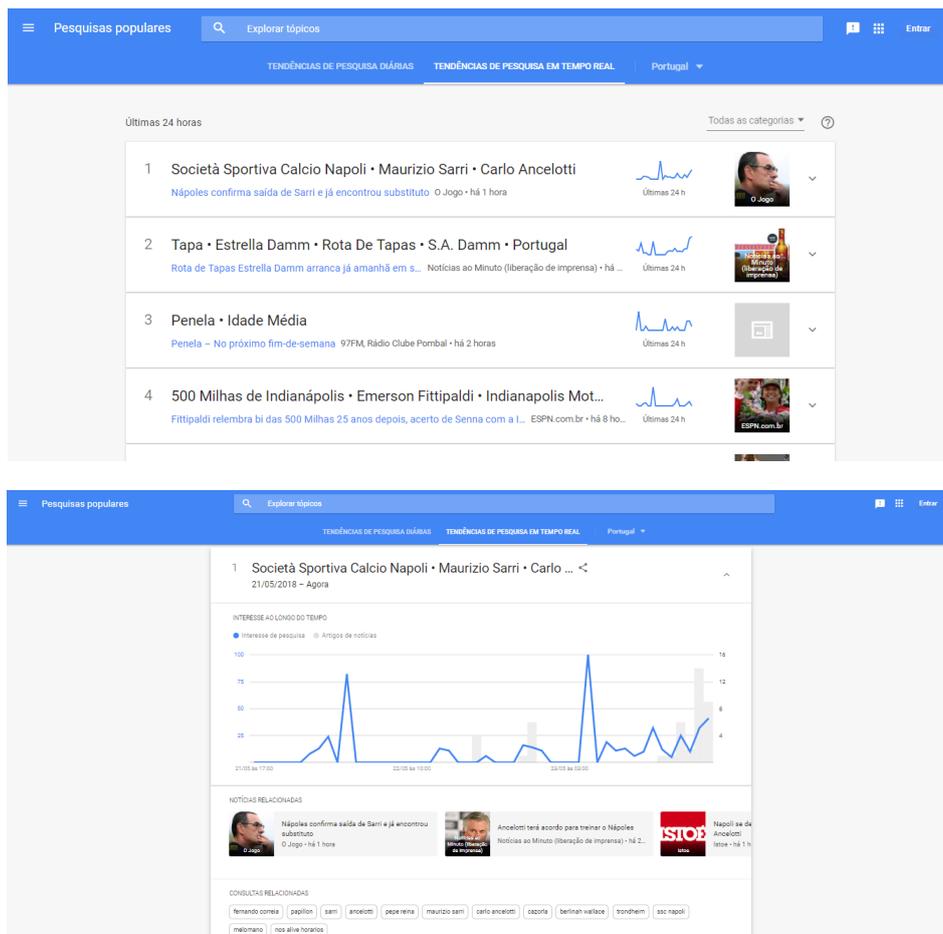


Figura 3.7: Tendências de pesquisa em tempo real. Captura de ecrã de <https://trends.google.pt/trends/trendingsearches/realtime?geo=PT&category=all>

Através da página de tendências temos acesso à página com dados de análise aos termos aí apresentados. Esta página apresenta o interesse ao longo do tempo referente ao termo selecionado, o interesse por região/sub-região/cidade e tópicos e consultas relacionadas. Estes mesmos dados são apresentados quando fazemos a pesquisa de uma palavra-chave única.

Para além disso, a plataforma permite fazer comparação entre termos. Aqui temos acesso a diversas informações como: o interesse ao longo do tempo (evolução) dos termos em pesquisa, a análise detalhada do interesse por região/sub-região/cidade comparando os termos, e a mesma análise individualmente. Esta análise é apresentada através de um mapa das sub-regiões e através de barras. São também apresentadas as consultas relacionadas com cada um dos termos (Figura 3.8).

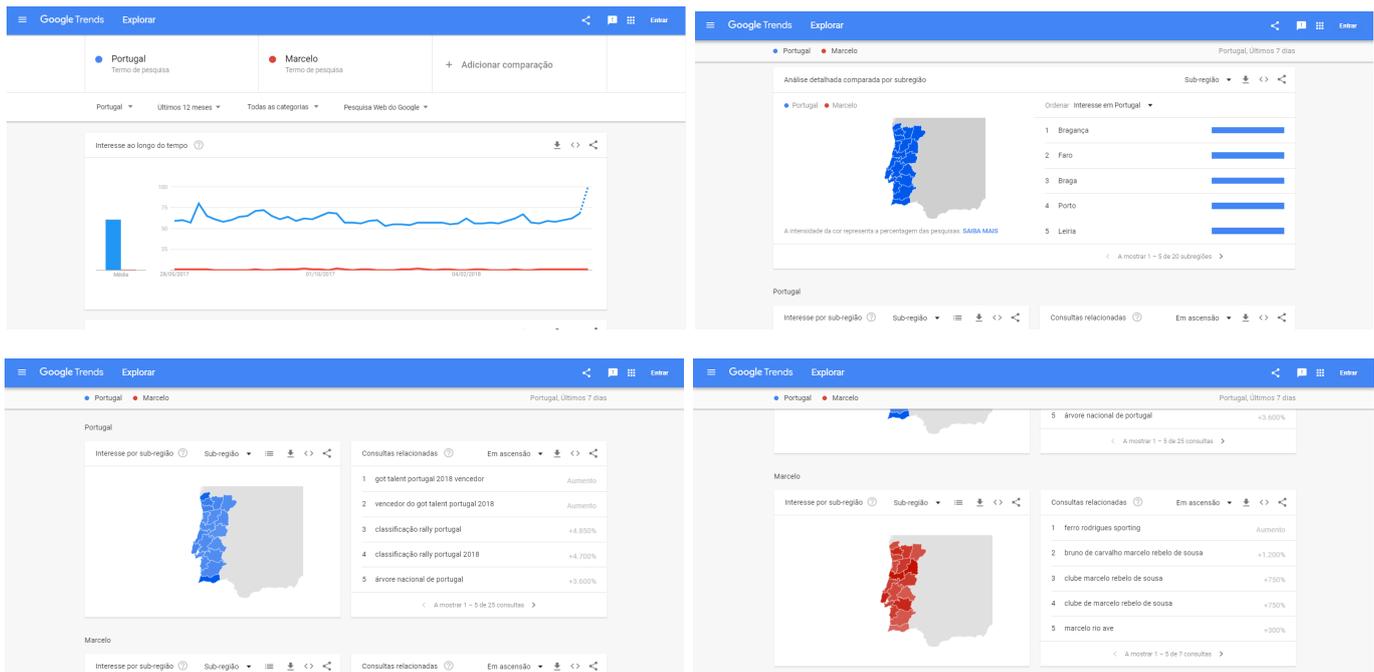


Figura 3.8: Google Trends - Comparação de termos. Captura de ecrã de <https://trends.google.pt/trends/explore?geo=PT&q=Portugal,Marcelo>

3.4 Media Cloud

A plataforma Media Cloud⁵ (Figura 3.9) é de código aberto e tem como objetivo o estudo dos ecossistemas dos media. A busca é feita em blogs, websites e jornais online e a informação é guardada. Os utilizadores podem pesquisar por palavras-chave e obter a resposta a questões como: onde, quando e com que frequência o termo é referenciado. Possibilita também a análise de vários termos em simultâneo permitindo uma análise comparativa. A plataforma dispõe de três ferramentas: Explorer, Topic Mapper e Source Manager.

⁵<https://mediacloud.org/>

Media Cloud is an open-source platform for media analysis.

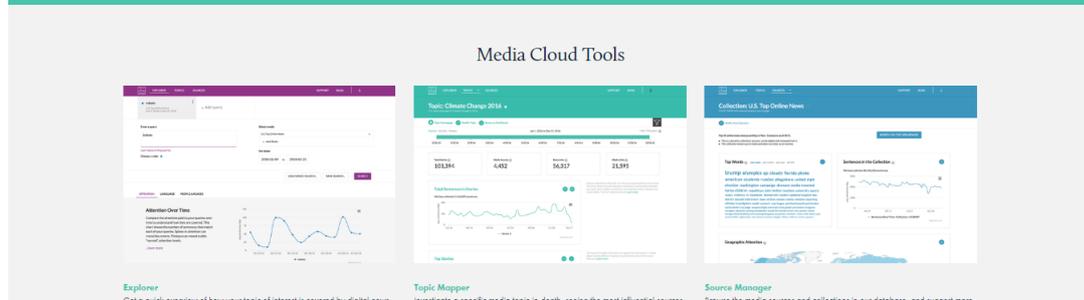
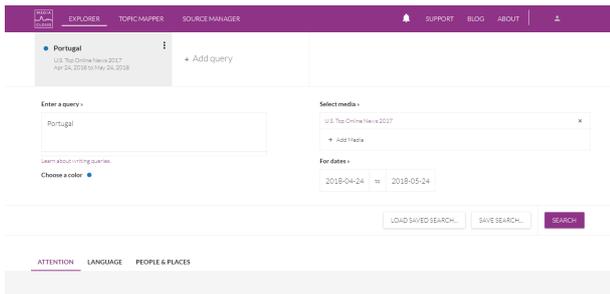
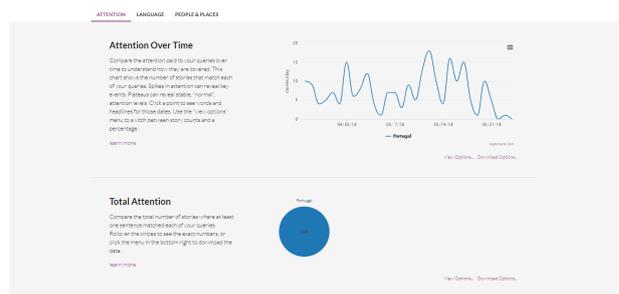


Figura 3.9: Página inicial das ferramentas da plataforma MediaCloud. Captura de ecrã de <https://mediacloud.org/>

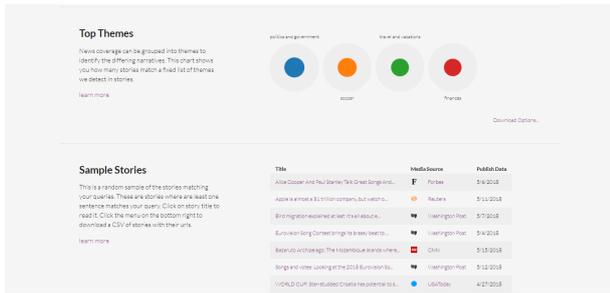
- **Explorer:** permite a pesquisa de nomes, locais, eventos de entre as mais de 500 milhões de histórias e disponibiliza algumas pesquisas já feitas. Os resultados obtidos estão divididos em três temas: a atenção (“Attention”) ao longo de tempo, a língua (“Language”) e pessoas & locais (“People&Places”). É possível alterar a fonte da informação (de entre os disponibilizados), a data de pesquisa e guardar a pesquisa (Figura 3.10 (a)). Na atenção ao longo de tempo obtemos a evolução do tópico em estudo, o total de atenção, o top de temas relacionados com a pesquisa e algumas histórias que incluem o tópico pesquisado (Figura 3.10 (b) e (c)). Na língua obtemos um top de palavras relacionadas (Figura 3.10 (d) e (e)). Por último, no tema pessoas & locais temos acesso ao top de organizações e pessoas que são referenciadas juntamente com os tópicos pesquisados e a cobertura geográfica (Figura 3.10 (f), (g) e (h)).



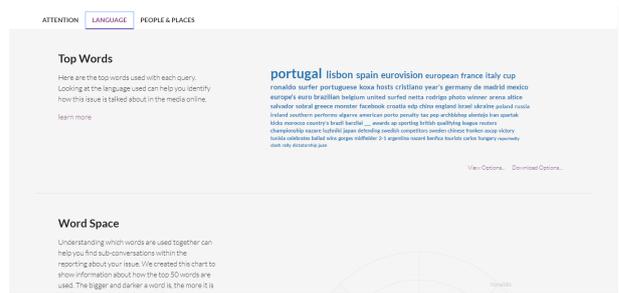
(a) Página com as definições de pesquisa.



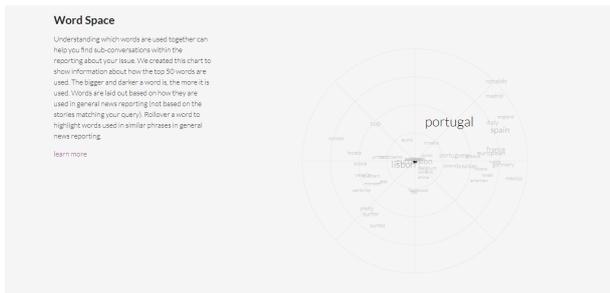
(b) Secção "Attention".



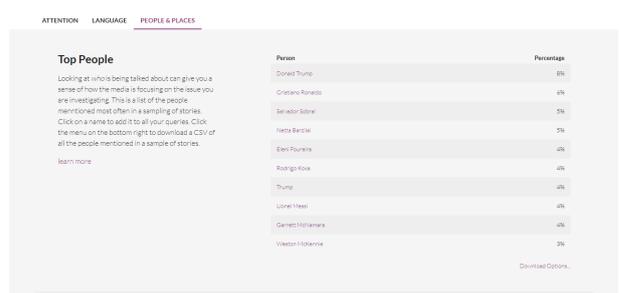
(c) Secção "Attention".



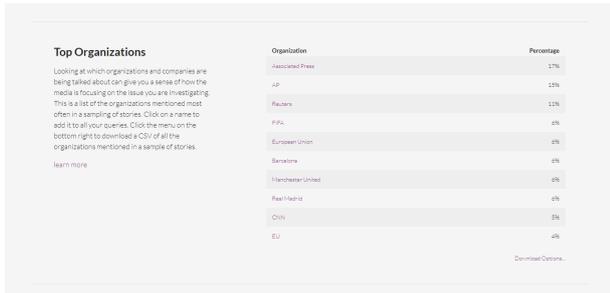
(d) Secção "Language".



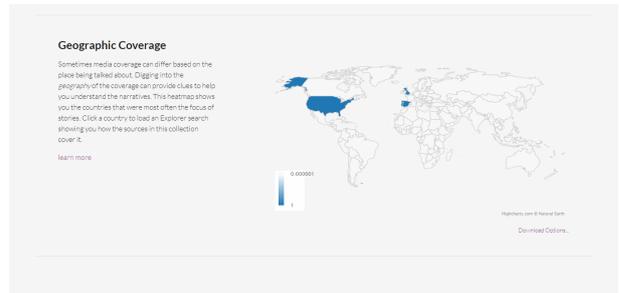
(e) Secção "Language".



(f) Secção "People & Places".



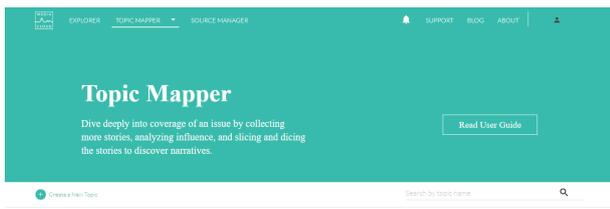
(g) Secção "People & Places".



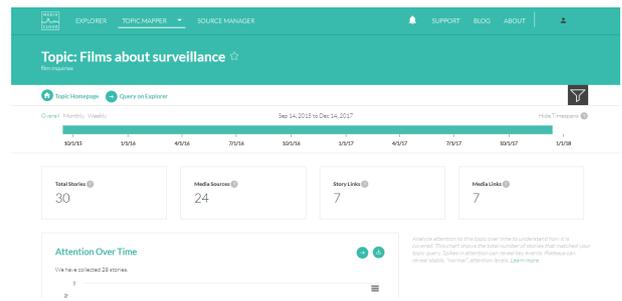
(h) Secção "People & Places".

Figura 3.10: Dados obtidos através da ferramenta Explorer com a palavra-chave "Portugal". Captura de ecrã de [https://explorer.mediacloud.org/#/queries/search?q=\[{"label":"Portugal","q":"Portugal","color":"%22%231f77b4%22","startDate":"2018-04-24","endDate":"2018-05-24","sources": \[\], "collections": \[58722749\]}\]&auto=true](https://explorer.mediacloud.org/#/queries/search?q=[{)

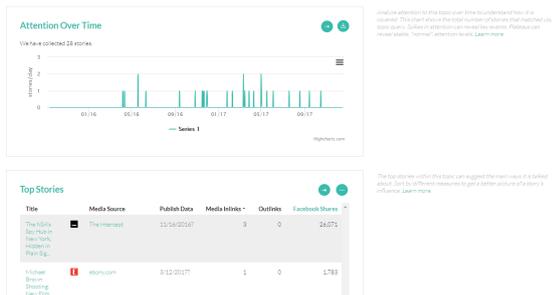
- **Topic Mapper:** Permite a criação ou exploração de tópicos. Na exploração podemos visualizar o número total de histórias, o número de fontes, o número de links entre as histórias e media, a evolução do tópico ao longo do tempo e as histórias que referenciam esse termo – título, fonte, data de publicação, links -, a lista de fontes, os 5 temas relacionados, as palavras mais referenciadas juntamente com o tópico, o top 10 das pessoas e organizações que são mencionadas nos artigos referentes ao tópico em pesquisa, a atenção geográfica, possibilidade de obter o gráfico de ligação entre as fontes e gráfico com as fontes que utilizam as mesmas palavras bem como, dados adicionais (Figura 3.11).



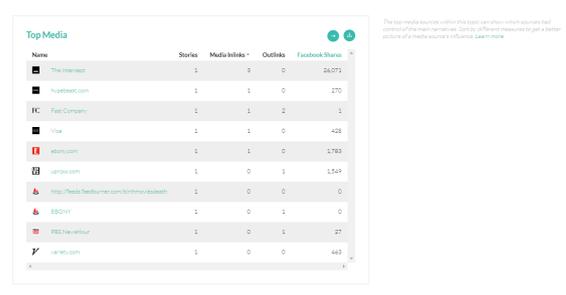
(a) Página da pesquisa.



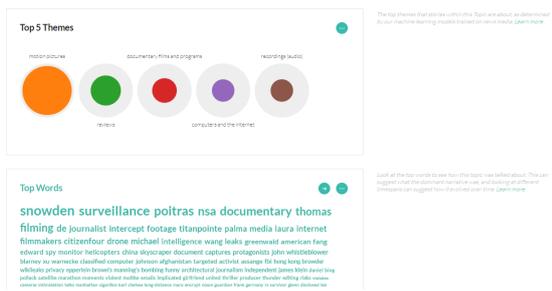
(b) Dados gerais.



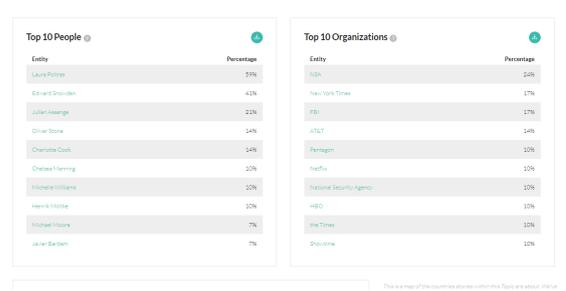
(c) Total de frases e top de histórias.



(d) Top de fontes.



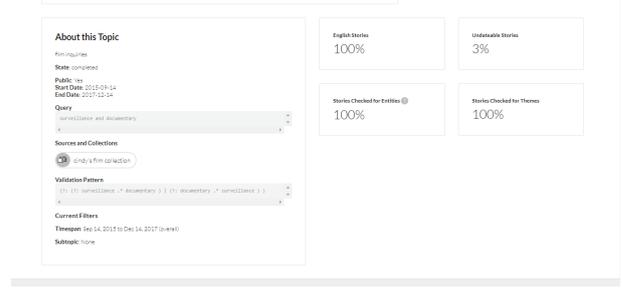
(e) Top de temas e palavras.



(f) Top de pessoas e organizações.



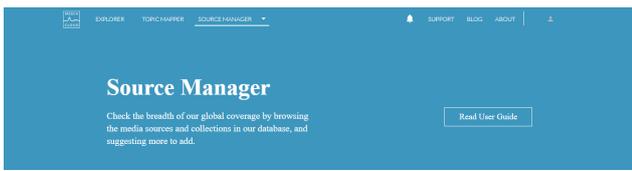
(g) Atenção geográfica.



(h) Acerca do tópico.

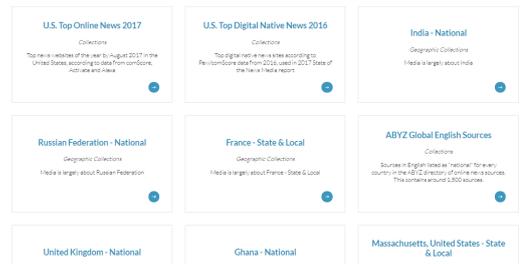
Figura 3.11: Dados obtidos através da ferramenta Topic Mapper com o tema “Films about surveillance”. Captura de ecrã de <https://topics.mediacloud.org/#/topics/1948/summary?focusId&q&snapshotId=2127×panId=178980>

- **Source Manager:** aqui é possível sugerir novas fontes e fazer dois tipos de pesquisa: pesquisa individual de fontes e pesquisa relativa à coleção de fontes da plataforma ([Figura 3.12](#) (a) e (b)).
 - **Pesquisa de coleção:** aqui obtemos dados relativos à lista de fontes que compõem essa coleção e ao conteúdo que essas disponibilizam. Os dados acerca da lista de fontes são: lista e representação comparativa das fontes, metadados associados e coleções similares ([Figura 3.12](#) (c) a (e)). Relativamente ao conteúdo obtemos: a evolução de recolha de artigos no último ano, as palavras mais utilizadas e a atenção geográfica ([Figura 3.12](#) (f) a (h)).
 - **Pesquisa de fonte individual:** aqui obtemos dados acerca da fonte e do conteúdo disponibilizado pela mesma. Relativamente aos dados acerca da fonte temos: o total de histórias, desde quando começou a ser utilizada esta fonte, número de coleções onde está inserida e link para as mesmas, número de histórias por semana, percentagem de histórias que incluem entidades, número de histórias que se inserem em temas, país e estado de publicação, língua primária, país de foco e tipo de media ([Figura 3.13](#) (a) e (b)). Os dados referentes ao conteúdo disponibilizado são: evolução do número de histórias publicadas ao longo do último ano, palavras mais utilizadas e atenção geográfica ([Figura 3.13](#) (c) e (d)).

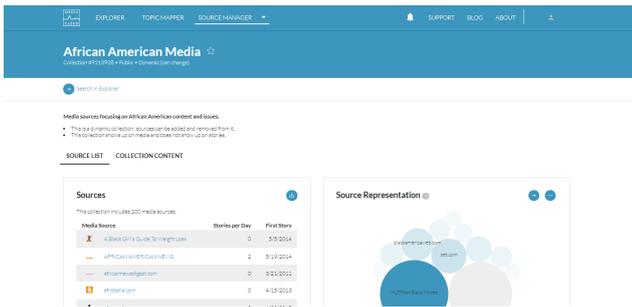


(a) Página de pesquisa.

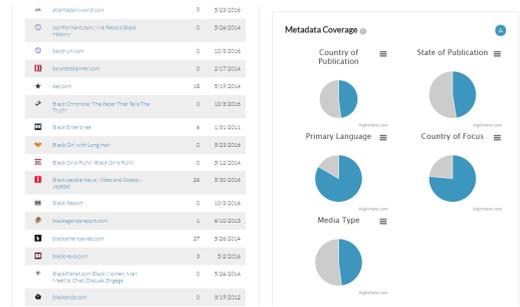
Featured Collections



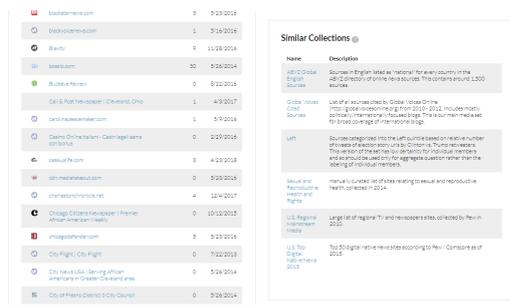
(b) Coleções de fontes populares.



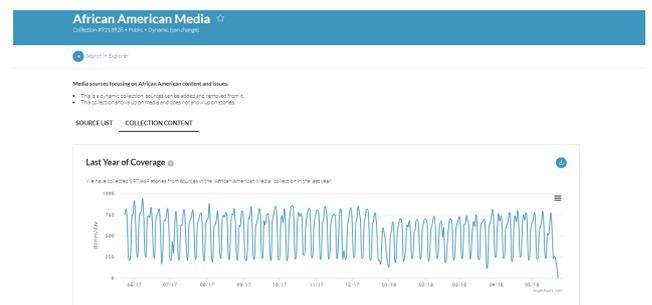
(c) Lista de fontes: lista e representação.



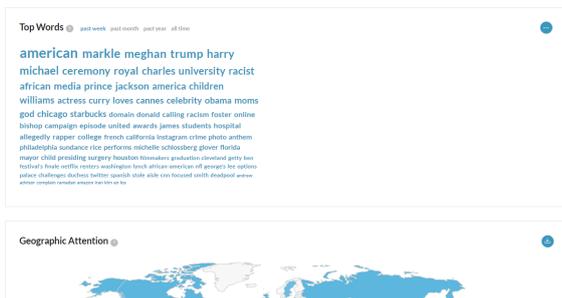
(d) Lista de fontes: metadados.



(e) Lista de fontes: coleções similares.



(f) Conteúdo: evolução de recolha no último ano.



(g) Conteúdo: palavras mais utilizadas.



(h) Conteúdo: atenção geográfica.

Figura 3.12: Dados obtidos através da ferramenta Source Manager com a coleção de fontes “African American Media”. Captura de ecrã de <https://sources.mediacloud.org/#/collections/9213928>

3.5 PageOneX

A plataforma PageOneX⁶ é de código aberto (Figura 3.14). O principal objetivo da plataforma é analisar quais as notícias que são referenciadas na página frontal dos jornais, demonstrando as prioridades dos mesmos. Conta, atualmente, com mais de 600 fontes de 50 países, maioritariamente de Espanha, França e EUA.

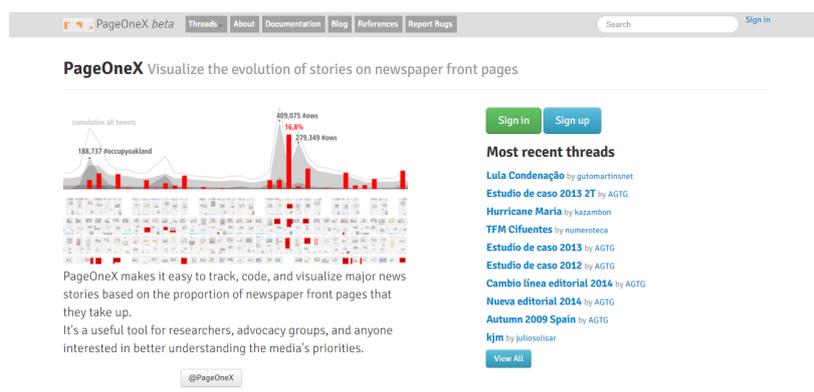


Figura 3.14: Página inicial da plataforma PageOneX. Captura de ecrã de <http://pageonex.com/>

É possível ter acesso, na página inicial, aos tópicos mais recentes (Figura 3.14). Na mesma página, cada utilizador, após fazer o login, tem a possibilidade de criar um novo tópico. Aqui, o utilizador terá de colocar diversos dados, tais como: nome do tópico, data de início e data de fim, categorias, descrição do tópico e selecionar as fontes nas quais pretende fazer a pesquisa, poderão ser adicionados colaboradores e criadas categorias de análise (Figura 3.15 (a) a (c)). Estando preenchidos os dados, a plataforma redireciona-nos para outra página. Nesta página, é possível ter acesso às páginas frontais das fontes selecionadas e proceder à codificação - seleção das áreas que se referem ao tópico em pesquisa (Figura 3.15 (d)). Feita a codificação, é possível ter acesso ao gráfico de barras que demonstra a percentagem de atenção de cada jornal (Figura 3.15 (e)).

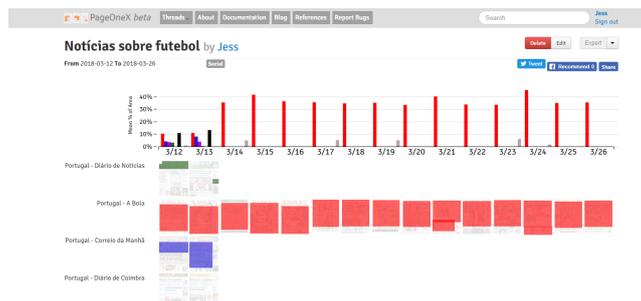
⁶<http://pageonex.com/>

(a) Dados necessários para a criação de Novo tópico.

(b) Dados necessários para a criação de Novo tópico.

(c) Dados necessários para a criação de Novo tópico.

(d) Codificação da páginas frontais de casa jornal.



(e) Gráfico de barras com os dados.

Figura 3.15: Criação de novo tópico através da plataforma PageOneX. Captura de ecrã de <http://pageonex.com/threads/new>

3.6 The GDELT Project

A plataforma The GDELT Project⁷ recolhe dados em tempo real, atualizando as informações de 15 em 15 minutos, faz a monitorização das notícias, a nível televisivo, impresso e online, de todo o mundo contando com mais de 100 línguas (Figura 3.16). Os dados recolhidos são: as pessoas envolvidas, os locais, as organizações, temas, fontes, emoções, frequências, citações, imagens e eventos. A base de dados é de acesso gratuito e aberto e é possível fazer o download dos dados, visualizá-los através do serviço de análise do GDELT ou analisá-los com o Google BigQuery (Figura 3.17).



Figura 3.16: Página inicial da plataforma The GDELT Project. Captura de ecrã de <https://www.gdeltproject.org/>

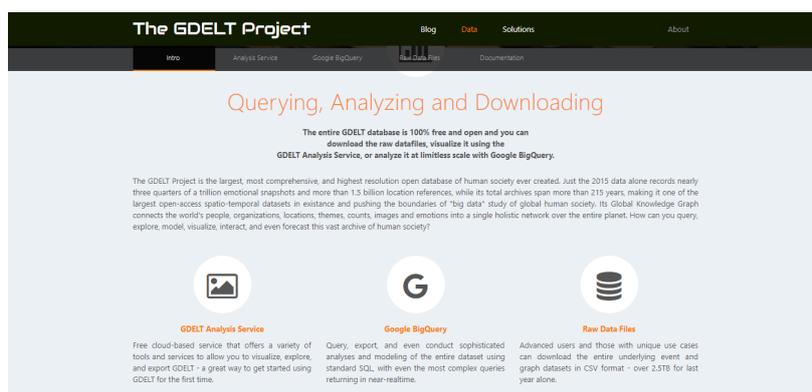


Figura 3.17: Possibilidades de análise do The GDELT Project. Captura de ecrã de <https://www.gdeltproject.org/data.html>

⁷<https://www.gdeltproject.org/>

1. **Serviço de Análise GDELT**⁸: fornece um conjunto de ferramentas que permite a exploração, visualização e exportação dos dados. Apresenta 14 ferramentas distintas para diferentes tipos de dados (Figura 3.18):

– **Eventos:**

Exportador: exportação de sub-conjuntos em que os dados se referem a eventos;

Rede Geográfica: rede geográfica que relaciona eventos e cidades e disponibiliza, de igual modo, imagens e arquivos georeferenciados;

Mapa de calor: mapa de calor interativo com a distribuição dos eventos;

Linha de tempo: linha de tempo que representa a intensidade;

Mapa do tempo: arquivo .KML do Google Earth que possibilita a visualização das alterações ao longo do tempo e do espaço e permite a exportação dos dados;

– **GKG:** Global Knowledge Graph é uma rede que recolhe tudo o que se está a passar no mundo (o que está a acontecer, o seu contexto, quem está envolvido e os sentimentos).

Rede: diagrama em rede que apresenta os maiores influenciadores;

Mapa de calor: mapa de calor interativo que representa a afinidade espacial/padrões espaciais;

Linha de tempo - Países: linha de tempo que permite entender padrões temporais entre países- a linha Y é o dia e X é o país;

Exportador: a pesquisa é feita em toda a base de dados do GKG e os resultados que forem de encontro à pesquisa são disponibilizados;

Rede Geográfica: rede geográfica que relaciona cidades e pontos de referência próximos à pesquisa e disponibiliza, de igual modo,

⁸<https://www.gdeltproject.org/data.html#gdeltprojectanalysiservice>

imagens e arquivos georeferenciados;

Nuvem de palavras: nuvem de palavras relacionada com a pesquisa;

Linha de tempo - Tom/Emoções: linha de tempo que permite entender padrões emocionais - tom positivo a tom negativo - ao longo do intervalo de pesquisa;

Linha de tempo: linha de tempo que representa a intensidade e permite entender padrões temporais;

Linha de tempo - Temática: linha de tempo que representa a intensidade e permite entender padrões temporais relativamente a cada dia e os temas recolhidos pelo GKG;

Em cada uma das possibilidades, o utilizador terá de preencher alguns campos de pesquisa. Os que são mais utilizados são: o intervalo de pesquisa e as palavras-chave. Os resultados são enviados por e-mail.



Figura 3.18: Ferramenta do Serviço de Análise GDELT. Captura de ecrã de <https://www.gdeltproject.org/data.html#gdeltanalysisservice>

2. **Google BigQuery**⁹: devido à enorme quantidade de dados que este projeto recolhe, foi necessário disponibilizar estes mesmos dados no Google BigQuery. Desta forma, os utilizadores são capazes de consultar e exportar dados quase em tempo real, sendo que a base é atualizada

⁹<https://www.gdeltproject.org/data.html#googlebigquery>

de 15 em 15 minutos (Figura 3.19). Este tipo de pesquisa funciona através de SQL e requer mais conhecimento por parte do utilizador. Para facilitar a pesquisa, são disponibilizados alguns exemplos.



Figura 3.19: Ferramenta Google BigQuery. Captura de ecrã de <https://www.gdeltproject.org/data.html#googlebigquery>

3. **Arquivos de dados brutos**¹⁰: é possível fazer o download dos dados recolhidos pela plataforma em formato CSV. No entanto, é necessário ter um vasto conhecimento na área da análise de dados devido à enorme quantidade de dados que são disponibilizados (Figura 3.20).

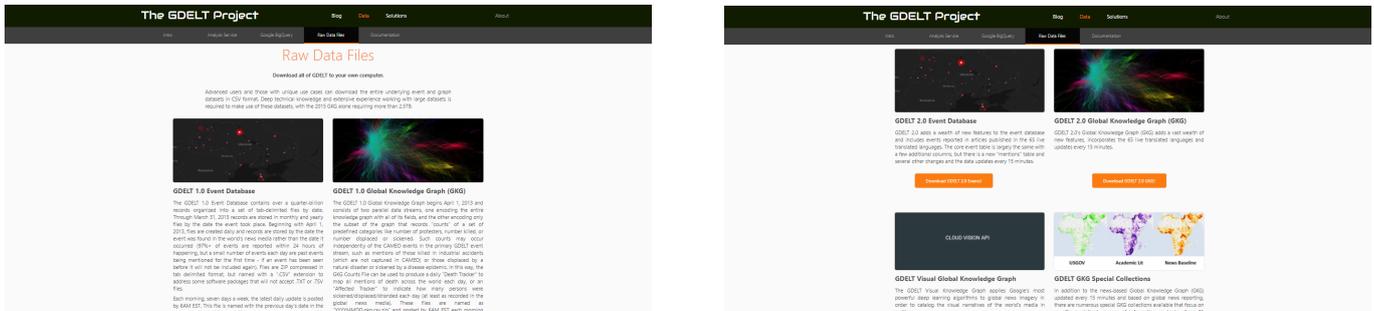


Figura 3.20: Ferramenta de exportação dos dados. Captura de ecrã de <https://www.gdeltproject.org/data.html#rawdatafiles>

¹⁰<https://www.gdeltproject.org/data.html#rawdatafiles>

3.7 Comparação

Comparando as diferentes plataformas acima referidas é possível verificar as diferentes características de cada uma. Desta forma, a [Tabela 3.1](#) apresenta as principais diferenças entre as mesmas.

Tabela 3.1: Comparação das diferentes plataformas.

Caraterísticas	Media Cloud	Europe Media Monitor	Pickanews	Google Trends	PageOneX	GDELT
Gratuito	x	x		x	x	x
Conta para aceder a todas as ferramentas	x		x		x	
Redirecionamento para fontes	x	x		x		
Comparação de dois ou mais termos	x			x		
Pesquisa de dois ou mais termos			x			x
Dados por região	x			x		x
Apresentação de tópicos relacionados				x		
Evolução do termo/notícia	x	x	x	x		x
Apresentação de tendências		x		x		
Notícias por língua/país		x		x		

Da mesma forma, é possível verificar que as diferentes plataformas, no que diz respeito à representação de informação, utilizam gráficos - maioritariamente de linhas e barras, mapas geográficos e nuvens de palavras para apresentar informações adicionais. A interação entre a plataforma e o utilizador é garantida através de links entre as páginas, seleção de informação e a navegação que permitem obter informação mais detalhada e de sub-conjuntos de dados. Relativamente à plataforma PageOneX, esta apresenta características bem diferentes das restantes no que toca à representação visual da informação. Apesar disso, no que toca à análise da informação e de notícias, é uma plataforma que apresenta uma ferramenta bastante interessante e importante, possibilitando visualizar o destaque dado pelas capas dos jornais a determinadas notícias em detrimento de outras.

3.8 Plataforma MediaViz

A plataforma MediaViz¹¹ foi criada por Tiago Devezas, Sérgio Nunes e Maria Teresa Rodriguez da Faculdade de Engenharia da Universidade do Porto. A plataforma recolhe, continuamente, conteúdo partilhado por um conjunto de fontes nacionais e internacionais, mais de 68 fontes, e torna os dados desses conteúdos acessíveis. Os dados são apresentados através de visualizações interativas, permitindo a sua análise e exploração (Devezas et al., 2015).

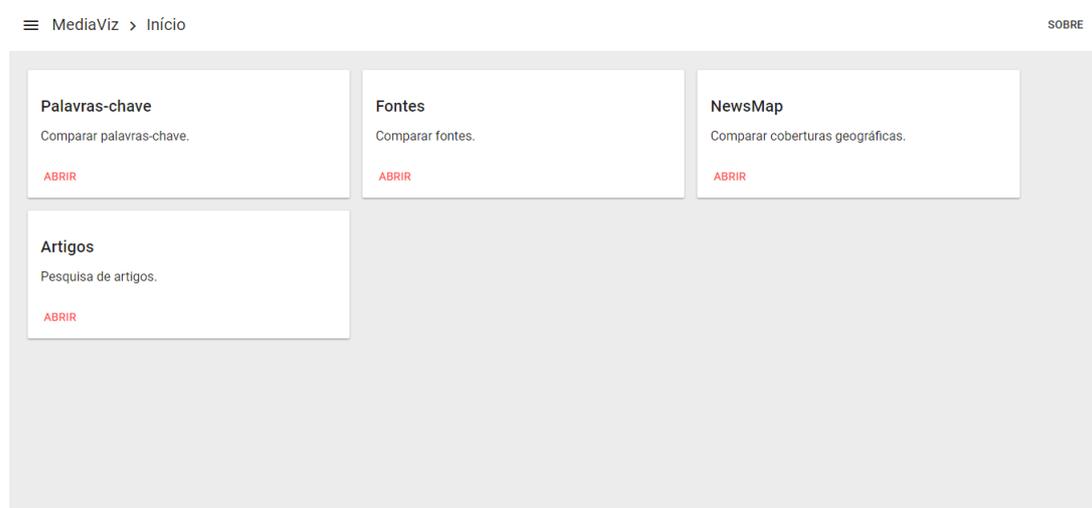


Figura 3.21: Página inicial da plataforma MediaViz. Captura de ecrã de <http://irlab.fe.up.pt/p/mediaviz/>

A plataforma permite vários tipos de pesquisa:

1. **Palavras-chave:** permite a análise de uma determinada palavra-chave ou a sua comparação com outras palavras. Os resultados são apresentados através de um gráfico de linhas que demonstra a evolução dos termos ao longo do tempo. São dadas ao utilizador as possibilidades de introduzir as palavras-chave pretendidas, alterar as fontes de pesquisa de entre as disponíveis, escolher o tipo de dados apresentados (totais ou percentuais) e o intervalo de pesquisa (Figura 3.22). Ao clicar num ponto do gráfico o utilizador é

¹¹<http://irlab.fe.up.pt/p/mediaviz>

redirecionado para a página 'Artigos' referente ao período e palavra-chave selecionada.

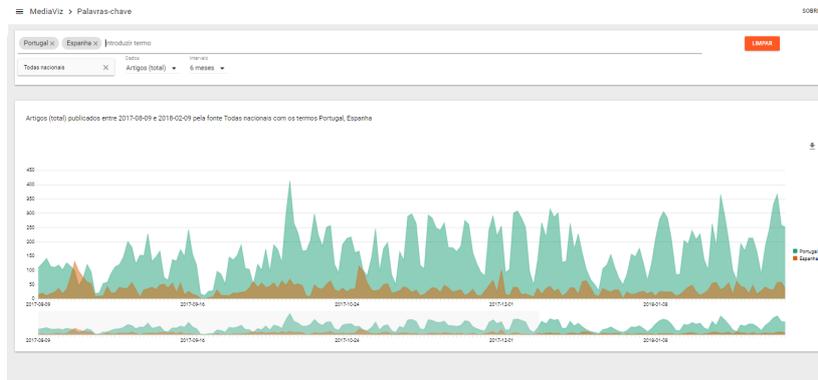


Figura 3.22: Pesquisa de Palavras-chave na Plataforma MediaViz. Captura de ecrã de <http://irlab.fe.up.pt/p/mediaviz/#/chronicle?source=Todas%20nacionalidades&ince=2017-08-09&until=2018-02-09&data=articles&keywords=Portugal,Espanha>

2. **Fontes:** permite obter o total de artigos publicados por uma coleção de fontes ou fontes individuais, possibilitando a comparação entre as mesmas. São dadas ao utilizador as possibilidades de escolher o conjunto de fontes ou fontes individuais, introduzir palavras-chave, escolher o tipo de dados apresentados (totais ou percentuais), o intervalo de pesquisa e o ciclo (Figura 3.23). Ao colocar palavras-chave a plataforma vai pesquisar o total de artigos de cada uma das fontes que incluam determinada palavra-chave. Ao clicar sobre um ponto do gráfico o utilizador é redirecionado para a página 'Artigos' onde serão apresentados os artigos publicados no dia selecionado e pela fonte pesquisada.

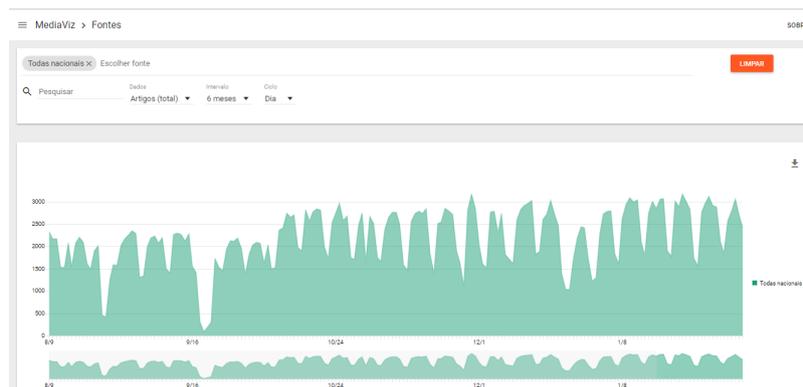


Figura 3.23: Pesquisa de Fontes na Plataforma MediaViz. Captura de ecrã de <http://irlab.fe.up.pt/p/mediaviz/#/sources?since=2017-08-09&until=2018-02-09&cycle=day&data=articles&sources=Todas%20nacionais>

3. **NewsMap:** permite obter o total de artigos publicados por uma coleção de fontes ou fontes individuais que incluam a palavra-chave referente a cada um dos países do mapa - mapa mundo -, ou dos distritos, quando a pesquisa é feita no mapa de Portugal. São dadas ao utilizador as possibilidades de escolher o mapa (mundo ou Portugal), o conjunto de fontes ou fontes individuais, introduzir uma palavra-chave e escolher o intervalo de pesquisa (Figura 3.24). Quando uma zona do mapa é clicada o utilizador é redirecionado para a página 'Artigos' onde serão disponibilizados os artigos que tenham sido publicados pela(s) fonte(s) selecionada(s), que tenham a palavra-chave referente ao distrito/país selecionado e que inclua - se aplicado - a palavra-chave escolhida pelo utilizador.

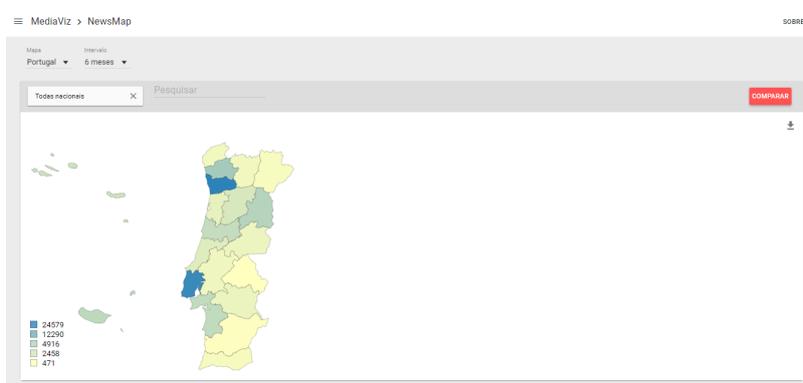


Figura 3.24: Pesquisa no NewsMap da Plataforma MediaViz. Captura de ecrã de <http://irlab.fe.up.pt/p/mediaviz/#/newsmap>

4. **Artigos:** permite obter os artigos publicados por determinada fonte. São dadas ao utilizador as possibilidades de introduzir as palavras-chave, escolher o conjunto de fontes ou fontes individuais e escolher o intervalo de pesquisa (Figura 3.25). Aqui são disponibilizados diversos dados: título do artigo, resumo, fonte, data, hora, link para a fonte e mapa com a evolução da palavra-chave na última semana.

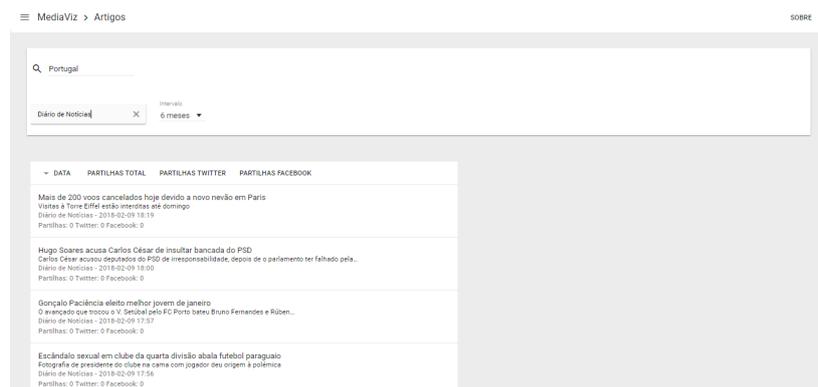


Figura 3.25: Pesquisa de Artigos na Plataforma MediaViz. Captura de ecrã de <http://irlab.fe.up.pt/p/mediaviz/#/articles?since=2017-08-09&until=2018-02-09&source=Di%C3%A1rio%20de%20Not%C3%ADcias>

Capítulo 4

Métodos

Neste capítulo é descrita a abordagem metodológica do estudo, os participantes e as suas características. Para além disso, são apresentados os instrumentos de recolha e análise de dados.

4.1 Desenho do Estudo

Este estudo apresenta um desenho não experimental sendo que não existe manipulação da variável independente, não existe um grupo de controlo, nem medidas múltiplas. Foram desenvolvidas duas personas representativas do público-alvo e utilizado um instrumento quantitativo (questionário) de forma a obter dados que possibilitassem a satisfação desse público. Para estruturar o questionário, e para que a sua implementação fosse vantajosa para o trabalho, foi elaborada uma grelha de análise. Esta grelha de análise consiste no agrupamento de informações acerca da representação visual dos dados. Essa grelha permitiu a estruturação de um gráfico que agrupa as opções visuais mais adequadas para cada tipo de informação.

- **Grelha:** o objetivo primário da grelha era agrupar, num só local, toda a informação que foi recolhida através dos seis estudos feitos às diferentes plataformas descritas no [Capítulo 3](#). Ao longo do tempo, a grelha foi sofrendo algumas alterações para que esta pudesse ser utilizada em projetos futuros e

por pessoas que se interessem pelo mesmo tema. Desta forma, era necessário obter algo que fosse claro e que agrupasse toda a informação necessária. A evolução da grelha vai ser apresentada na próxima secção - [Secção 4.2](#). As principais distinções feitas em termos de informação foram relativamente aos níveis de análise e às dimensões de análise e entre informação diretamente recolhida - como fonte, título, palavras -, e informação processada - entidades e tópicos.

- **Personas:** segundo [Cooper et al. \(2007\)](#), as personas são modelos de utilizadores que representam distintos grupos de comportamentos, atitudes, aptidões, objetivos e motivações observadas e identificadas durante a fase de pesquisa e devem ser baseadas em pessoas reais. Neste contexto, foram desenvolvidas duas personas tendo em conta o público-alvo da plataforma. [Mendes \(2014\)](#) afirma que o principal objetivo é perceber quais os objetivos dos utilizadores para com o uso do sistemas e os seus comportamentos e atitudes na realização dos mesmos. Aqui, o objetivo primário era obter informações mais específicas sobre as necessidades, capacidades e limitações dos futuros utilizadores da plataforma. As personas primárias são investigadores da área dos media e as personas secundárias são entusiastas dessa mesma área (como estudantes). Desta forma é possível, antecipadamente, perceber como deve ser feita a comunicação com o público-alvo, como desenvolver o design centrado neste público e perceber de que forma cada um dos utilizadores poderá utilizar a plataforma. As duas personas são apresentadas no [Apêndice B](#). As personas permitiram direcionar o questionário para pessoas que se enquadrassem no público-alvo da plataforma.
- **Questionário:** a principal função do questionário foi a obtenção de informação relativamente aos níveis e dimensões de análise que o público-alvo da plataforma considera mais importantes e fulcrais. Os resultados obtidos através do mesmo podem ser vistos na [Secção 4.3](#) e o questionário pode ser visto no [Apêndice C](#).

4.2 Grelha de Análise

A necessidade de estudar e perceber o que já foi previamente feito e entender o que pode ser melhorado ou o que de novo pode ser implementado, é fulcral para o desenvolvimento de qualquer projeto. Neste caso, a criação de uma plataforma - *MediaViz* - que seja útil e vantajosa para o público-alvo deve assentar num estudo rigoroso. Este estudo tem dois pontos fundamentais:

1. **Plataformas já existentes:** testar, analisar e descrever as plataformas que já existem, apresentadas no [Capítulo 3](#).
2. **Necessidades dos utilizadores:** cada plataforma deve ser criada e adaptada tendo em vista o público-alvo. A análise das suas necessidades, capacidades e limitações é fulcral para que este utilize regularmente a plataforma e não, simplesmente, a use uma vez e a abandone.

Neste contexto, surgiu a necessidade de agrupar, num só local, as informações que foram recolhidas através dos seis estudos de caso - *Europe Media Monitor*, *Pickanews*, *Google Trends*, *Media Cloud*, *PageOneX*, *The GDELT Project*.

A criação de uma grelha que agregasse informação transversal a todas as plataformas e que servisse de base para o futuro desenvolvimento da plataforma *MediaViz* ou outras plataformas era essencial.

De entre os objetivos do desenvolvimento da grelha podemos destacar:

1. Perceber qual a informação disponibilizada e análise feita pelas plataformas já existentes;
2. Possibilitar a análise da informação que os futuros utilizadores da plataforma valorizam;
3. Permitir estudar o input dos potenciais utilizadores e quais as funcionalidades necessárias na plataforma.

Este foi um processo iterativo na medida que, para chegar à versão final da grelha, esta sofreu diversas alterações. Essas alterações serão descritas de seguida.

1. **Primeira fase:** o primeiro passo foi analisar uma das plataformas em estudo e agregar a informação disponibilizada pela mesma numa grelha. A plataforma escolhida foi a *MediaCloud*. A divisão inicial da informação foi feita consoante o tipo de pesquisa. Esta plataforma, como anteriormente visto, disponibiliza três ferramentas que permitem a pesquisa através de palavras-chave, por tópicos ou por fontes. Cada uma dessas ferramentas apresenta informações/resultados distintos e esses são representados através de gráficos. Cada um desses gráficos permite, ou não, uma interação por parte do utilizador. No entanto, a divisão feita inicialmente, tendo em conta o tipo de pesquisa, não mostrou ser a mais correta, pois nem todas as plataformas apresentam as mesmas funcionalidades. Desta forma, a informação passou a ser distinguida e avaliada através do tipo de informação: temporal, numérica e semântica, como pode ser visto na [Figura 4.1](#). Foram também agregadas informações relativamente a possíveis alterações ao nível das opções visuais. Esta primeira abordagem serviu para a verificar a estrutura da plataforma em estudo e perceber que tipo de dados e informação é que são disponibilizados e como.

Tipologia	Vista/Resultados	Representação	Possibilidades	Interação
Temporal	Evolução do termo/tópico	Gráfico de linhas		Obtenção de dados com hover: data, palavra e média diária.
		Gráfico de barras		Zoom
				Possibilidade de comparação com sub-tópicos
				Linkagem para página com mais dados
Numérica	Total de artigos	Gráfico de bolhas	Mapa de árvore com percentagens	Obtenção de números exatos com hover
	Dados gerais	Quadros com valores		Nenhuma
	Representação das fontes (%)	Gráfico de bolhas		Obtenção de percentagem com hover
	Top de fontes	Gráfico de barras		Linkagem para página com dados da fonte
		Lista		Nenhuma
		Linkagem para página com mais dados		
Semântica	Temas relacionados	Gráfico de bolhas	Ávore de palavras	Obtenção de números exatos com hover
	Histórias/artigos	Lista		Linkagem para página da fonte
	Palavras mais referenciadas (top)	Nuvem de palavras ordenada		Alteração da visualização
				Ocultação de palavras
			Seleção sobre a palavra para a adicionar à pesquisa	
	Palavras usadas em conjunto (espaço de palavras)	Nuvem de palavras	Infográfico/gráfico com ligações	Obtenção de melhor percepção da palavra no hover
			Ávore de palavras	
	Pessoas e organizações referenciadas (top)	Lista		Seleção sobre a palavra para a adicionar à pesquisa
	Cobertura geográfica	Mapa Mundo		Obtenção do número de artigos/percentagem com hover
Cobertura de metadados (dados adicionais)	Gráfico de pizza	Obtenção de percentagens com hover		
Artigos mais relevantes	Texto + foto	Linkagem para a fonte		

Figura 4.1: Primeira versão da grelha de análise em que a informação é avaliada segundo o tipo de informação.

2. **Segunda fase:** após a análise de apenas uma das plataformas era necessário avaliar as restantes. Isto porque, se é fundamental perceber tudo o que já foi feito anteriormente e se um dos principais objetivos é criar uma grelha que se adapte a todas as plataformas, é necessário ver o que há de comum e de distinto entre as mesmas. A discussão voltou a prender-se com a forma como a informação da grelha seria avaliada. Os níveis temporal, numérico e semântico não pareciam ser os mais apropriados neste caso. Desta forma, partimos para a análise da forma como a informação é tratada e quais os níveis de representação. Isto é, de que forma é que os dados são tratados para, posteriormente, serem apresentados ao utilizador. Foram então especificadas duas dimensões: dimensão de análise e dimensão da informação.

A dimensão de análise diz respeito à análise que é feita dos dados, que se distingue entre:

- **Frequência**
- **Tempo**
- **Espaço**

Para além da dimensão de análise temos a dimensão da informação ou seja, que tipo de informação é recolhida, tratada e apresentada. Não seria possível agregar numa grelha concisa todo o tipo de informação que é disponibilizado. Desta forma, foi necessário perceber quais eram os elementos de análise através dos quais, essa informação era recolhida. Foi assim possível distinguir sete níveis:

- **Coleção**
- **Documento**
- **Tópico**
- **Frase**
- **Termo**
- **Entidade**
- **Fonte**

Para perceber melhor o que era representado em cada uma das dimensões de informação, foram disponibilizadas as propriedades inerentes a essas dimensões. Isto é, informação mais detalhada sobre os dados que essas dimensões representam. A grelha final é apresentada de seguida - [Figura 4.2](#). Esta também pode ser vista no [Apêndice A](#) em tamanho maior.

Dimensão da Informação	Propriedade	Dimensão da análise			Plataformas										
		Frequência	Tempo	Espaço	Representação gráfica	MediaCloud			Europe Media Monitor	Pickanews	Google Trends	PageOneX	GDELT Project		
						Explorer	Topic Mapper	Source Manager							
Coleção	Total de artigos	x			Gráfico de bolha	x									
	Temas relacionados	x			Número		x			x					
	Artigos	x			Gráfico de bolhas	x	x								
	Artigos mais recentes	x			Lista	x	x			x					
	Temas mais atualizados	x			Gráfico de linhas					x					
Documento	Capas de jornais	x	x		Lista					x					
	Tópico	Evolução ao longo do tempo	x	x	Gráfico de linhas	x	x	x		x		x			
Gráfico de áreas											x				
Gráfico de barras														x	
Linha de tempo															x
Diagrama de rede															x
Frase	Total de frases	x	x		Gráfico de linhas		x	x							
					Diagrama de rede										x
Termo	Palavras mais utilizadas	x			Nuvem de palavras	x	x	x					x		
	Palavras mais utilizadas em conjunto	x			Nuvem de palavras	x									
Entidade	Pessoas	x			Lista	x	x								
					Diagrama de rede										x
	Organizações	x			Lista	x	x								
					Diagrama de rede										x
	Locais	x		x	Mapa	x	x	x			x				
					Rede Geográfica										x
					Mapa de calor										x
Fonte	Total	x			Número		x								
					Lista			x	x				x		
	Mais referenciadas	x													
	Representação	x			Gráfico de barras										
	Metadados	x			Gráfico de bolhas				x						
Tipo de fonte				Gráfico de pizza				x							
				Gráfico de rosca						x					

Figura 4.2: Segunda versão da grelha de análise em que a informação é avaliada segundo a dimensão de análise e dimensão da informação.

No entanto, esta grelha apresenta, logo à partida, algumas desvantagens como:

- **Falta de resumo:** não é sucinta, sendo que apresenta informação demasiado detalhada e das diferentes plataformas;
- **Falta de flexibilidade:** o principal objetivo do desenvolvimento desta grelha é a possibilidade de adaptação a diferentes plataformas e esta versão não o permite;
- **Difícil análise:** a quantidade de informação apresentada torna a sua análise complicada.

3. **Terceira fase:** a distinção entre dimensão de análise e dimensão de informação tornou-se ambígua. Desta forma, níveis de análise seria um termo que se aplicaria melhor neste caso. Nesta fase foram definidos que os dois níveis que iriam distinguir a informação a disponibilizar na grelha seriam então:

- **Níveis de análise;**

- **Dimensão da análise.**

A necessidade de elaboração da grelha e inclusão de demasiada informação era, ainda nesta fase, algo que era discutido. Nesta versão foram colocados os níveis anteriormente descritos e as propriedades intrínsecas aos níveis de análise. Isto é, que dados eram apresentados.

Nível de análise	Propriedade	Dimensão da análise		
		Frequência	Tempo	Espaço
Fontes	Documentos	Total de documentos		Referência geográfica
		Total de frases		
		Dados associados (metadados)		
	Representação	Total de fontes		
		Relação relativamente ao total		
		Referenciados		
Entidades	Referenciadas			
Tópicos	Associados			
Documentos	Conteúdo	Titulo		
		Data de publicação		
		Fonte		
		Termos utilizados		
	Total de frases			
	Entidades	Referenciadas		
Tópicos	Associados			
Frases	Conteúdo	Fonte		
		Data de publicação		
		Termos utilizados		
	Entidades	Referenciadas		
Tópicos	Associados			
Termos	Representação	Evolução		
	Agrupamento	Termos utilizados em conjunto		
	Entidades	Referenciadas		
	Tópicos	Associados		

Figura 4.3: Terceira versão da grelha de análise em que a informação é avaliada segundo o nível de análise e as suas propriedades e dimensão da informação.

4. **Fase final:** ao longo deste processo de desenvolvimento da grelha foi perceptível que a quantidade de informação apresentada era algo que seria fulcral para a correta análise e aplicação da mesma. Grande parte da informação, desde a primeira versão, foi retirada. Isto porque, a principal função da mesma é aplicar-se em cada um dos projetos e a partir daí, tornar-se flexível e sustentar aplicações diferentes. Desta forma, foram realizadas duas grelhas finais.

A primeira grelha diz respeito aos níveis de análise, como pode ser visto na [Tabela 4.1](#).

Tabela 4.1: Níveis de análise dos dados.

Níveis de análise	Descrição
Coleção	o objeto de estudo é a coleção de dados
Fonte	o objeto de estudo são as fontes de origem dos documentos
Documento	o objeto de estudo é o conjunto de documentos integrais
Frase	o objeto de estudo é o conjunto de todas as frases existentes na coleção
Termo	o objeto de estudo é o conjunto de todos os termos existentes na coleção

Esta grelha apresenta os objetos de estudo que cada uma das plataformas pode utilizar para recolher informações. Aqui, estes níveis de análise podem adaptar-se a todas as outras e futuras plataformas e estarem, ou não, presentes consoante o objetivo da mesma.

Estes níveis de análise permitem, de igual forma, que o utilizador esteja ciente de qual é o objeto de estudo e configurar, se possível, consoante o seu objetivo na pesquisa, o elemento que está a ser estudado. Por exemplo, se a plataforma o permitir, o utilizador poderá ser capaz de fazer pesquisas apenas relacionadas com as fontes de origem dos documentos ou relacionadas com o conjunto de termos associados à sua pesquisa.

Neste contexto, para além da descrição de cada um dos níveis como apresentado na grelha, é necessário clarificar cada uma das dimensões assim:

- **Coleção:** composta pelos artigos/documentos recolhidos através das diversas fontes;
- **Fonte:** fonte ou conjunto de fontes a que a plataforma recorre;
- **Documento:** diz respeito ao objeto de estudo da plataforma ou seja, qual o objeto através do qual a plataforma retira os seus dados. Como

por exemplo: artigo/documento integral, capa de jornal, dados do artigo (fonte, título, data de publicação), entre outros;

- **Frase:** conjunto de frases que estão incluídas nos documentos da coleção em análise;
- **Termos:** conjunto de termos que estão incluídos nos documentos da coleção em análise;

A segunda grelha diz respeito às dimensões de análise. Isto é, que tipo de análise é feita aos dados recolhidos e apresentados, como pode ser visto na [Tabela 4.2](#).

Tabela 4.2: Dimensões de análise dos dados.

Dimensões de análise	Descrição
Frequência	análise dos valores totais associados ao objeto de estudo
Temporal	análise da evolução do objeto de estudo ao longo do tempo ou de ciclos temporais
Geográfica	análise do objeto de estudo em função da geografia (locais mencionados)
Entidades não geográficas	análise das entidades mencionadas (pessoas e organizações)
Tópicos	análise dos temas mencionados

Esta grelha permite perceber que tipo de análise pode ser feita aos diferentes tipos de dados. Desta forma, é possível entender quais as opções e de que forma é que essas se adaptam aos dados em questão.

Aqui foi necessário fazer distinção entre dimensão de análise geográfica e entidades não geográficas. Isto porque, normalmente o termo entidades está relacionado com pessoas, organizações e locais. Mas, como a análise geográfica é bastante utilizada e um meio fácil de obter uma grande quantidade de informações, optou-se por destacar esta dimensão e tratá-la de forma independente.

4.2.1 Aplicação da grelha

4.2.1.1 Europe Media Monitor

- **Níveis de análise:**

1. **Coleção:** a coleção desta plataforma é composta por artigos/documentos recolhidos em mais de 20 mil feeds RSS e páginas HTML durante 24h e em mais de 70 línguas, recolhendo cerca de 280 mil artigos por dia ([European Commission, 2018](#)). A plataforma apresenta o top 10 de artigos mais falados nesse momento ou nas últimas 24 horas e a evolução de determinado assunto;
2. **Fonte:** ao nível das fontes de origem dos documentos a plataforma apresenta a fonte dos artigos e a lista de fontes a que recorre;
3. **Documento:** não é feita nenhuma análise ao nível dos documentos, sendo feita apenas a linkagem para o documento original;
4. **Frase:** é possível fazer uma pesquisa através de frases (conjunto de termos) e obter os resultados que vão de encontro à pesquisa. Esta pesquisa permite seleccionar qual a língua em que o artigo foi escrito, o país de origem e a fonte específica, bem como o período de tempo;
5. **Termo:** é possível efetuar a pesquisa de determinados termos com as mesmas opções de pesquisa e é possível fazer uma pesquisa por artigos que incluam o termo pesquisado.

- **Dimensões de análise:**

1. **Frequência:** esta plataforma apresenta a frequência de publicação de artigos referentes aos assuntos mais falados e permite ver os artigos em questão;
2. **Temporal:** a evolução de determinado assunto é apresentada através de gráficos de linhas que permitem fazer comparações;

3. **Geográfica:** ao serem apresentados os artigos publicados sobre determinado assunto é apresentado um mapa mundo que referencia o local onde os eventos ocorreram. Também são apresentados os países referenciados nos artigos dando acesso a informação como: notícias sobre o país, fontes desse país, nível de alerta, citações extraídas, temas associados, entidades mais associadas e o número diário de artigos, bem como acesso aos artigos recolhidos de fontes desse país;
4. **Entidades não geográficas:** são apresentadas apenas as pessoas referenciadas nos artigos de determinado assunto e apresentada a percentagem relativa a cada uma delas. Através da referência a pessoas é possível obter dados como: frases associadas à pessoa em questão, títulos (cargos), outras entidades associadas e relacionadas, relações e artigos onde é referenciado;
5. **Tópicos:** esta plataforma apresenta os temas mais atualizados.

- **Opções visuais:**

- Mapas;
- Gráfico de linhas;
- Gráfico de barras;
- Blocos de texto;
- Tabelas;
- Gráfico de relações.

4.2.1.2 PickaNews

- **Níveis de análise:**

1. **Coleção:** a coleção desta plataforma é composta por artigos/documentos recolhidos em meios como a televisão, a rádio, a web e a imprensa. A coleção de artigos é recolhida de sites de notícias e de informação generalistas e/ou profissionais - ([Pickanews](#)). Ao nível da coleção é apresentado o número de artigos referentes ao termo pesquisado;
2. **Fonte:** ao nível das fontes de origem dos documentos a plataforma apresenta o top de fontes que publicaram sobre o assunto e a repartição pelo tipo de media;
3. **Documento:** não é feita nenhuma análise ao nível dos documentos;
4. **Frase:** esta plataforma permite fazer uma pesquisa através de frases (conjunto de termos) e obter os resultados que vão de encontro à pesquisa;
5. **Termo:** é apresentado o total de resultados referentes ao(s) termo(s) pesquisado(s) e a sua evolução.

- **Dimensões de análise:**

1. **Frequência:** esta plataforma apresenta o total de referências à palavra(s)-chave pesquisada(s), a percentagem de publicação de cada um dos tipos de media e de cada uma das fontes através de um gráfico circular e de um gráfico de barras, respetivamente;
2. **Temporal:** a evolução de determinado assunto/palavra-chave é apresentada através de um gráfico de linhas;
3. **Geográfica:** não é feita nenhuma análise relativamente a referências geográficas;
4. **Entidades não geográficas:** não é feita nenhuma análise a entidades;
5. **Tópicos:** não é feita nenhuma análise referente a tópicos.

- **Opções visuais:**

- Gráfico de linhas;
- Gráfico de barras;
- Gráfico circular;
- Blocos de texto;

4.2.1.3 Google Trends

- **Níveis de análise:**

1. **Coleção:** a coleção desta plataforma é composta por artigos/documentos recolhidos através do sistema de pesquisa da Google. Ao nível da coleção são apresentados, na página inicial, os artigos mais pesquisados na plataforma, diariamente e em tempo real. E, na pesquisa, são apresentados os tópicos e pesquisas relacionadas;
2. **Fonte:** não é feita nenhuma análise ao nível das fontes de origem dos documentos, sendo estas apenas referenciadas;
3. **Documento:** não é feita nenhuma análise ao nível dos documentos, sendo feita apenas a linkagem para o documento original;
4. **Frase:** esta plataforma permite fazer uma pesquisa através de frases (conjunto de termos) e obter os resultados que vão de encontro à pesquisa;
5. **Termo:** a plataforma permite a pesquisa de termos e a comparação entre os mesmos. Esta dá acesso ao interesse demonstrado ao longo do período de tempo selecionado, bem como o interesse por região e assuntos e consultas relacionadas com a pesquisa.

- **Dimensões de análise:**

1. **Frequência:** esta plataforma apresenta o total de pesquisas com determinada(s) palavra(s)-chave, o interesse de pesquisa por região/sub-

região/cidade, bem como o aumento de interesse das consultas e tópicos relacionados;

2. **Temporal:** o interesse ao longo do tempo é apresentado através de um gráfico de linhas que pode apresentar diversos períodos de tempo como: última hora, últimas 4 horas, últimos 7 dias, últimos 5 anos e até selecionar datas específicas;
3. **Geográfica:** é feita uma análise por região, sub-região ou cidade que apresenta o interesse por determinada palavra-chave. Quando a pesquisa é feita com mais do que um termo é apresentado o interesse por cada termo individualmente, mas também é apresentado um gráfico comparativo das palavras-chave pesquisadas;
4. **Entidades não geográficas:** não é feita nenhuma análise a entidades;
5. **Tópicos:** esta plataforma apresenta os assuntos e tópicos relacionados que estão em ascensão, ou os mais referenciados.

- **Opções visuais:**

- Gráfico de linhas;
- Gráfico de barras;
- Mapas;
- Listas.

4.2.1.4 MediaCloud

- **Níveis de análise:**

1. **Coleção:** a coleção desta plataforma é composta por mais de 837 milhões de artigos/documentos recolhidos das mais de 50 mil fontes de media online ([The Berkman Center for Internet and Society, 2009](#)). No que diz respeito à coleção de dados, esta plataforma permite fazer pesquisas ao nível de palavras-chave e obter dados relativos a essa pesquisa;

2. **Fonte:** esta plataforma apresenta uma ferramenta específica para a análise ao nível das fontes - *Source Manager*. Aqui é possível fazer pesquisa de fontes individuais ou de coleções de fontes que existem na plataforma;

- **Pesquisa individual:** Quando fazemos a pesquisa de uma fonte individual é possível obter dados a dois níveis: dados acerca da fonte e dados acerca do conteúdo publicado pela mesma. No que se refere aos dados acerca da fonte obtemos: total de histórias, data de início de cobertura, número de coleções em que se insere, total de histórias por semana, histórias que contêm entidades e temas, país e estado de publicação, língua primária, país em que se foca e tipo de media que divulga, bem como, a lista das coleções em que se insere. Já referente ao conteúdo temos: a evolução da publicação, as palavras mais utilizadas e a atenção geográfica, ou seja, os países mais referenciados.

- **Pesquisa de coleção:** quando fazemos pesquisa de uma coleção de fontes obtemos, também, dados a dois níveis: listas de fontes e dados relativos ao conteúdo publicado pelas fontes. Relativamente à lista de fontes obtemos: a lista e representação comparativa de cada uma das fontes, metadados associados e coleções similares. Relativamente ao conteúdo obtemos: a evolução de publicação durante o

último ano, as palavras mais utilizadas e a atenção geográfica;

3. **Documento:** não é feita nenhuma análise ao nível dos documentos;
4. **Frase:** esta plataforma tem no seu arquivo mais de 5 bilhões de frases que foram analisadas através dos artigos recolhidos das suas fontes. Ao efetuar uma pesquisa na plataforma é possível obter o número de histórias/artigos que incluem, pelo menos, uma frase que referencie a palavra(s) que está a ser pesquisada. Esta plataforma permite, também, fazer uma pesquisa através de frases (conjunto de termos) e obter os resultados que vão de encontro à pesquisa;
5. **Termo:** a MediaCloud permite ao utilizador fazer pesquisa de palavras-chave e obter dados relativos a essa pesquisa. Entre esses dados podemos encontrar: a evolução da atenção dada aos termos em pesquisa, no período de tempo e fontes selecionadas, os temas em que a palavra se insere, e alguns dos artigos que referenciam esses termos. São ainda apresentadas as palavras mais utilizadas em conjunto com cada um dos termos pesquisados. Esta análise é individual e comparativa relativamente aos termos.

- **Dimensões de análise:**

1. **Frequência:** ao nível de análise de frequências, a plataforma faz esta análise em diversos dados como: total de artigos, a atenção dada a cada um dos termos, nos temas, nas palavras utilizadas em conjunto e nos metadados das coleções de fontes. No que diz respeito ao total de artigos, esta plataforma permite obter o total de artigos, num determinado dia, que referenciam a palavra pesquisada. De igual modo, a atenção dada a cada um dos termos em pesquisa é apresentada através de círculos de frequência, que referem o total de artigos em que, pelo menos, uma frase inclui a palavra pesquisada. Também a análise ao nível dos temas é apresentada através da frequência, neste caso, percentagens. A análise feita ao nível das palavras utilizadas em conjunto

é feita através de nuvens de palavras que permitem ver quais as palavras que se destacam e a relação entre as mesmas. Já os metadados e a representação das fontes nas coleções são apresentados através de gráficos circulares;

2. **Temporal:** esta plataforma apresenta a atenção ao longo do tempo através de um gráfico de linhas que permite analisar a evolução ao longo do tempo e a comparação entre termos;
3. **Geográfica:** esta plataforma apresenta um mapa mundo que representa a percentagem de histórias que incluem, não só, o termo em pesquisa, mas também, referências geográficas (países);
4. **Entidades não geográficas:** ao nível das entidades esta plataforma apresenta análise ao nível das pessoas e organizações referenciadas. Esta análise é apresentada através de uma tabela que inclui a entidade e a percentagem de participação, para cada um dos termos;
5. **Tópicos:** esta plataforma apresenta, como já referido, as palavras que são utilizadas em conjunto e são analisados os temas que se relacionam com os termos em pesquisa, como por exemplo: política e governo, futebol, viagens e férias, finanças, entre outros. No entanto, esta plataforma apresenta uma ferramenta específica para a pesquisa de tópicos - *TopicMapper*. Aqui é possível analisar um tópico em específico. Os dados obtidos nesta pesquisa são: total de artigos, total de fontes e links associados, a atenção dada ao tópico no período de tempo selecionado, os artigos mais importantes, as fontes, os temas relacionados, as palavras mais utilizadas e as entidades mais referenciadas, a atenção geográfica e dados associados.

- **Opções visuais:**

- Gráfico de linhas;
- Gráficos de frequências;

- Gráficos circulares;
- Tabelas;
- Nuvens de palavras;
- Mapa mundo.

4.2.1.5 PageOneX

- **Níveis de análise:**

1. **Coleção:** a coleção desta plataforma é composta por artigos/documentos recolhidos através de mais de 600 fontes de 50 países. Ao nível da coleção, são apresentadas as páginas frontais dos jornais às quais podemos fazer uma análise;
2. **Fonte:** através da codificação das capas dos jornais é possível fazer uma comparação entre os assuntos publicados por cada uma das fontes selecionadas;
3. **Documento:** nesta plataforma o documento de análise são as capas dos jornais e é possível fazer uma codificação das mesmas;
4. **Frase:** esta plataforma permite fazer a codificação das capas dos jornais, onde poderão ser analisadas as frases aí apresentadas;
5. **Termo:** tal como no caso das frases, é possível codificar apenas os termos presentes nas capas dos jornais e obter dados acerca das fontes que utilizam mais determinado termo na capa, por exemplo.

- **Dimensões de análise:**

1. **Frequência:** fazendo a análise a cada uma das capas das fontes selecionadas é nos dada a percentagem da área ocupada pelo assunto que estamos a pesquisar;
2. **Temporal:** é apresentada a evolução dessa percentagem no período de tempo que selecionarmos para a nossa análise;

3. **Geográfica:** poderá ser feita uma análise ao nível geográfico se fizermos uma codificação da capa que vá de encontro a esse objetivo. No entanto, não será gerado nenhum mapa geográfico, sendo mais uma análise ao nível dos termos que se relacionam com identidades geográficas;
4. **Entidades não geográficas:** poderá ser feita uma análise a entidades se codificarmos apenas as entidades (termos) que queremos;
5. **Tópicos:** poderá ser feita uma análise ao nível dos termos se fizermos uma codificação da capa que vá de encontro a esse objetivo.

- **Opções visuais:**

- Gráfico de barras;
- Codificação das capas dos jornais.

4.2.1.6 GDELT Project

- **Níveis de análise:**

1. **Coleção:** a coleção desta plataforma inclui dados recolhidos em tempo real a nível televisivo, impresso e online, de todo o mundo, em mais de 100 línguas. Esta coleção é composta por: pessoas, locais, organizações, temas, fontes, emoções, frequências, citações, imagens e eventos ([The GDELT Project, 2013](#)). Esta coleção de dados é disponibilizada através de 14 ferramentas;
2. **Fonte:** não é feita nenhuma análise relativamente às fontes de origem dos dados, estas apenas são disponibilizadas quando os dados são transferidos. Uma das ferramenta - GKG Exporter-, permite fazer uma pesquisa por palavra-chave e obter a lista de fontes que publicaram dados relativos a essa pesquisa;
3. **Documento:** não é feita nenhuma análise ao nível dos documentos;

4. **Frase:** esta plataforma permite fazer uma pesquisa através de frases (conjunto de termos) em cada uma das ferramentas e obter os resultados que vão de encontro à pesquisa;
5. **Termo:** nesta plataforma uma das possibilidades dada aos utilizadores é a pesquisa por determinada palavra ou exclusão de dada palavra e obter os dados pretendidos em cada ferramenta. Neste campo, a plataforma tem uma grande quantidade de gráficos que apresentam, de forma diferente e dinâmica, a informação pesquisada pelo utilizador.

- **Dimensões de análise:**

1. **Frequência:** grande parte das ferramentas disponibilizadas pela plataforma apresentam dados de frequência em diferentes formas de visualização;
2. **Temporal:** ao nível temporal a plataforma apresenta várias ferramentas:
 - **Event Timeline:** permite compreender padrões temporais nos registos que correspondem aos critérios pesquisados pelo utilizador;
 - **GKG Country Timeline:** permite compreender padrões temporais nos registos que correspondem aos critérios pesquisados pelo utilizador relativamente a países;
 - **GKG Tone Timeline:** permite analisar o “tom” (entre positivo e negativo) relativo à pesquisa;
 - **GKG Timeline:** permite entender padrões temporais na pesquisa;
 - **GKG Thematic Timeline:** possibilita compreender padrões temáticos;
3. **Geográfica:** a nível geográfico a plataforma apresenta as seguintes ferramentas:
 - **Event Geographic Network:** permite entender as ligações existentes entre as cidades no que diz respeito a eventos e relativamente à pesquisa efetuada;

- **Event e GKG Heatmap:** possibilita entender padrões espaciais na pesquisa;
- **Event TimeMapper:** permite visualizar padrões relativamente a eventos a nível espaço-temporal;
- **GKG Geographic Network:** permite entender as ligações espaciais relativas à pesquisa.

4. **Entidades não geográficas:** a plataforma apresenta a ferramenta GKG Network que permite obter dados relativos às pessoas, organizações, temas, países, cidades e divisões administrativas relacionadas com a pesquisa feita pelo utilizador. Estes dados são apresentados através de um diagrama de rede;
5. **Tópicos:** ao nível dos tópicos a plataforma disponibiliza a ferramenta GKG Word Cloud que permite obter uma nuvem de palavras relacionadas com a pesquisa feita.

- **Opções visuais:**

- Tabelas;
- Diagrama de rede geográfico;
- Mapas de calor;
- Linhas de tempo;
- Mapas Google Earth;
- Diagramas de rede;
- Nuvem de palavras;
- Gráfico de linhas.

4.2.1.7 MediaViz

- **Níveis de análise:**

1. **Coleção:** a coleção desta plataforma é composta por artigos/documentos recolhidos de várias fontes desde dezembro de 2014, agregando mais de 5 milhões de artigos. A plataforma permite a pesquisa por palavras-chave, uma ou várias, e dá-nos acesso aos artigos e número de artigos, publicados pelas fontes selecionadas, no período de tempo selecionado;
2. **Fonte:** ao nível das fontes de origem dos documentos a plataforma *MediaViz* permite fazer uma comparação entre fontes ou conjuntos de fontes ao nível do total de artigos, num determinado intervalo de tempo. Para além disso, permite fazer uma comparação ao nível de palavras-chave. Esta plataforma inclui dezenas de fontes distintas. Para além de meios online portugueses, inclui também vários meios internacionais;
3. **Documento:** a plataforma permite obter alguns dados relativos, por exemplo, ao número de documentos/artigos publicados em determinado período de tempo, com determinadas palavras-chave e de determinada(s) fonte(s);
4. **Frase:** esta plataforma permite fazer uma pesquisa através de frases (conjunto de termos) e obter os resultados que vão de encontro à pesquisa;
5. **Termo:** esta plataforma permite fazer pesquisa por termo ou termos que se encontram na sua coleção de dados. Esta apresenta a evolução dos artigos publicados que incluem esse(s) termo(s) nas fontes e período de tempo selecionados. Estes dados são apresentados através de um gráfico de linhas que permite fazer comparações e dá acesso aos artigos em questão.

- **Dimensões de análise:**

1. **Frequência:** esta plataforma analisa as fontes tendo em conta a frequência

com que publicam e permite ver os artigos publicados em determinado dia;

2. **Temporal:** a evolução de publicação de determinada fonte ou de determinada palavra-chave é apresentada através de gráficos de linhas. Aqui é possível selecionar vários intervalos de tempo para análise: dias, ciclos de 24 horas, semanal e mensal;
3. **Geográfica:** o mapa mundo e mapa de Portugal permite obter dados relativamente à(s) palavra(s)-chave pesquisada(s) e a cobertura geográfica ao nível dos países e dos distritos portugueses;
4. **Entidades não geográficas:** não é feita nenhuma análise referente a entidades;
5. **Tópicos:** não é feita nenhuma análise referente a tópicos.

- **Opções visuais:**

- Mapas;
- Gráfico de linhas;
- Blocos de texto.

4.2.2 Análise comparativa

Após a aplicação e análise individual das grelhas em cada uma das plataformas é possível obter, através da [Tabela 4.3](#) e da [Tabela 4.4](#), uma análise comparativa referente aos níveis e dimensões de análise utilizados por cada uma das plataformas em estudo.

- **Níveis de análise:** relativamente aos níveis de análise é possível verificar, através da [Tabela 4.3](#), que as plataformas que se destacam são: Europe Media Monitor, MediaCloud, PageOneX, MediaViz e PickaNews. De entre estas, a PageOneX é aquela que mais se distancia das outras no que diz respeito aos dados disponibilizados. Isto porque, apesar de utilizar todos os níveis

de análise como objeto de estudo, esta apenas permite uma codificação da capa dos jornais recolhidas através das suas fontes e esta codificação tem de ser feita pelo utilizador. A PickaNews, no meu entender, é a plataforma que disponibiliza menos informação ao utilizador, talvez por ser necessário subscrição e pagamento. As outras três plataformas - Europe Media Monitor, Media Cloud e MediaViz -, apresentam mais semelhanças. A MediaCloud destaca-se das outras devido aos dados disponibilizados ao utilizador.

As restantes plataformas, Google Trends e GDELT Project apresentam menos níveis de análise. No caso da GDELT Project esta é uma plataforma com uma grande base de dados, que pode ser utilizada posteriormente pelo utilizador para análise e apresenta diversas ferramentas de visualização. A principal distinção da Google Trends relativamente às outras ferramentas é a fonte dos dados, sendo que a sua fonte são as pesquisas feitas no Google. Apesar disso, esta tira bastante partido das opções de visualização de dados.

Através da mesma grelha é possível verificar que, quanto aos níveis de análise mais utilizados temos:

1. **Coleção:** utilizada por todas as plataformas pois todas elas têm uma coleção de dados;
2. **Termo:** todas as plataformas permitem a pesquisa, por parte do utilizador, a partir de termos e conjunto de termos. Sendo os dados obtidos referentes a essa pesquisa;
3. **Frase:** todas as plataformas permitem a pesquisa através de conjunto de termos e os dados obtidos vão de encontro a essa pesquisa;

Os níveis de análise menos utilizados são: a fonte e o documento. As plataformas que tiram mais partido da análise das fontes é a MediaCloud, que apresenta uma ferramenta própria - *Source Manager*, e a MediaViz, que permite fazer comparação relativamente ao número de publicações de cada fonte ou conjunto de fontes. As restantes plataformas, grande parte das vezes, apenas fazem referência à fonte dos documentos. Relativamente ao nível de

análise referente aos documentos todas as plataformas apresentam diferentes documentos como objeto de estudo, mas apenas a plataforma PageOneX e a MediaViz permitem uma análise a esse nível.

Tabela 4.3: Níveis de análise utilizados pelas plataformas em análise.

Níveis de análise	Europe Media Monitor	PickaNews	Google Trends	MediaCloud	PageOneX	GDELT Project	MediaViz
Coleção							
Fonte							
Documento							
Frase							
Termo							

- **Dimensões de análise:** quanto às dimensões de análise é possível verificar, através da [Tabela 4.4](#), que as plataformas que se destacam são: Europe Media Monitor, MediaCloud, PageOneX e GDELT Project. Aqui devemos, novamente, distanciar a PageOneX das restantes plataformas, pois os resultados obtidos para cada uma destas dimensões depende, diretamente, da codificação e dos objetivos do utilizador. Já as outras plataformas são as que, no geral, se destacam e apresentam mais ferramentas e dados ao utilizador. A Europe Media Monitor e a MediaCloud apresentam esses dados e ferramentas diretamente no navegador, já a GDELT Project envia os dados por e-mail. Das restantes plataformas, destacamos a MediaViz e a Google Trends que tiram bastante partido dos dados e da representação gráfica dos mesmos. Estes dois não apresentam análise relativamente às entidades, e no caso da MediaViz, também relativamente aos tópicos. A plataforma que tira menos partido dos dados é a PickaNews, talvez pelo motivo já anteriormente referido.

Desta forma, é possível verificar que as dimensões de análise mais utilizadas são:

1. **Frequência:** utilizada por todas as plataformas. Todas elas apresentam valores totais referentes ao objeto de estudo;
2. **Temporal:** todas as plataformas permitem obter dados relativos à evolução do objeto de estudo ao longo do tempo ou de ciclos temporais. Esta análise é, geralmente, apresentada através de um gráfico de linhas;
3. **Geográfica:** com exceção da plataforma PickaNews, todas as outras plataformas permitem obter dados relativamente aos locais mencionados. Esta análise é sempre apresentada através de mapas;

As dimensões de análise menos utilizadas são: as entidades não geográficas e os tópicos. Este tipo de informação é, como anteriormente referido, informação processada. Ou seja, não é diretamente recolhida através dos feeds RSS e precisa de outro tipo de tratamento. Esta pode ser uma das justificações para a falta de análise a estes níveis. No entanto, será possível, através do questionário realizado, perceber se estas dimensões são importantes para os utilizadores e se seria uma boa aposta para a plataforma que irá ser, posteriormente, desenvolvida.

Tabela 4.4: Dimensões de análise utilizadas pelas plataformas em análise.

Dimensões de análise	Europe Media Monitor	PickaNews	Google Trends	MediaCloud	PageOneX	GDELT Project	MediaViz
Frequência							
Temporal							
Geográfica							
Entidades não geográficas							
Tópicos							

4.2.3 Síntese

Esta secção apresentou a aplicação da grelha nas diferentes plataformas e a análise dessa aplicação. Essa análise individual permitiu fazer uma análise comparativa. Por sua vez, foi possível obter dados relativamente a:

- Existência de padrões;
- Existência de discrepâncias.

Sumariamente, é possível verificar que os níveis de análise mais utilizados são a coleção, o termo e a frase e, quanto às dimensões de análise temos a frequência, a temporal e a geográfica.

Quanto às plataformas é possível verificar que a PageOneX é que se distancia mais das outras plataformas quanto às funcionalidades e resultados obtidos. Das restantes, é possível destacar quatro delas: Europe Media Monitor, a MediaCloud, a GDELT Project e a MediaViz. De entre estas, aquelas que tiram melhor partido das possibilidades inerentes à visualização de dados são a MediaCloud, a GDELT Project e a MediaViz. De referir que, como dito anteriormente, a MediaViz é a plataforma que será alvo de desenho à posteriori deste projeto, mas já existe com uma versão inicial.

4.3 Questionário

4.3.1 Participantes

A amostra deste estudo é constituída essencialmente por investigadores, jornalistas e estudantes da área dos media. O grau de participação no questionário foi baixo, não obtendo os resultados esperados. Foram recolhidas, no total, treze respostas. A amostra deste estudo é composta por treze participantes (7 mulheres, 5 homens, 1 preferiu não responder). Esta amostra divide-se em seis grupos etários 15-24 (1), 25-34 (3), 35-44 (2), 45-54 (6), 55-64 (1) e +65 (0). Os participantes são

jornalistas, investigadores e estudantes do Programa Doutoral em Media Digitais e do Mestrado em Ciências da Comunicação da Universidade do Porto.

Os participantes do questionário foram recrutados através do e-mail institucional da Universidade do Porto para três tipos de público:

- **Investigadores:** a distribuição dos e-mails foi feita pelo e-mail da Universidade do Porto e foram contactadas diversas unidades de investigação das diferentes universidades. O e-mail enviado e a lista de contactos podem ser consultados no [Apêndice D](#) e no [Apêndice E](#).
- **Jornalistas:** a distribuição dos e-mails foi feita pelo e-mail da Universidade do Porto e foram contactadas diversas redações, diretores de redação e jornalistas. O e-mail enviado e a lista de contactos pode ser consultado no [Apêndice D](#) e no [Apêndice E](#).
- **Estudantes:** foi utilizado o serviço da Universidade do Porto para chegar a estudantes do Programa Doutoral em Media Digitais e estudantes do Mestrado em Ciências da Comunicação. O e-mail enviado pode ser consultado no [Apêndice D](#).

A identidade dos participantes não foi revelada, sendo que o questionário é anónimo, e não receberam qualquer incentivo à resposta.

4.3.1.1 Método de Amostragem

Este estudo aplica uma amostra não-probabilística por conveniência, sendo que não é utilizada qualquer seleção aleatória e o envio do formulário é feito a conjuntos de pessoas acessíveis. Juntamente com o questionário, foi feito um recrutamento para obter uma lista de contactos que esteja disponível para possíveis testes de utilizadores no futuro.

A seleção dos contactos para o envio dos e-mails foi feita tendo em conta os objetivos do projeto e o público-alvo. Através do artigo “Cibermeios portugueses: 10 anos de lenta evolução” ([Ana Isabel Reis, Fernando Zamith, Helder Bastos, 2017](#)) foi possível obter dados relativamente ao aproveitamento do potencial do

jornalismo online no caso dos jornais portugueses e fazer uma seleção de potenciais redações para contacto. Desta forma, foi feita a seleção dos meios que deviam ser contactados através da análise da [Figura 4.4](#) e foram adicionados outros meios que se enquadravam no projeto. A lista final pode ser consultada na [Apêndice E](#).

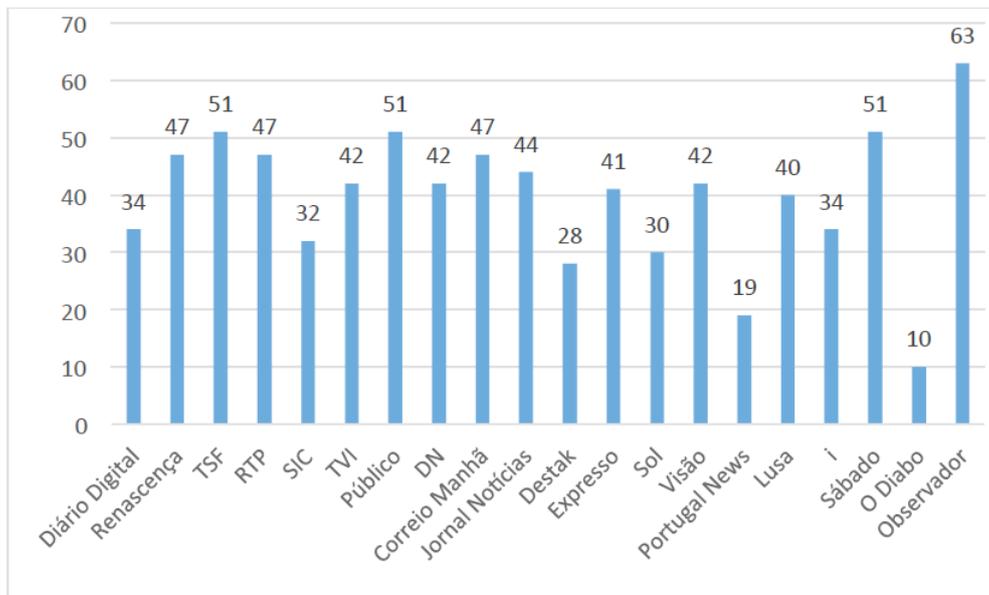


Figura 4.4: Aproveitamento percentual das potencialidades jornalísticas da Internet pelos cibermeios generalistas nacionais portugueses em novembro de 2016 ([Ana Isabel Reis, Fernando Zamith, Helder Bastos, 2017](#)).

4.3.2 Instrumentos

O instrumento utilizado para todo o processo foi a aplicação e análise da grelha e esta serviu de base para o desenvolvimento de um questionário em Português. O questionário foi realizado através do Google Forms e pode ser encontrado na [Secção 4.3](#). Este é constituído por três secções e contém, no total, 14 questões. Os dados recolhidos na primeira secção foram dados demográficos (5 questões). A seguinte secção era dedicada aos níveis de análise dos dados e continha 2 questões (uma com escala de Likert e outra aberta). A última secção abordava as dimensões de análise e apresentava 6 questões com escala de Likert e uma questão aberta. Este questionário foi realizado especificamente para este estudo e possibilitou a recolha de informação para posterior design e prototipagem da plataforma.

A seguir são apresentadas as variáveis investigadas em cada uma das secções do questionário:

- **Demografia** - idade, sexo, profissão, instituição e área científica de formação - 5 questões.
- **Níveis de análise** - importância dos níveis de análise, opinião - 2 questões.
- **Dimensões de análise** - importância de cada uma das dimensões de análise - 6 questões e uma questão de opinião.

Foi utilizada uma escala de Likert com cinco níveis. Esses níveis foram: “Nada importante”, “Pouco importante”, “Moderadamente importante”, “Importante” e “Muito importante”. As duas questões de opinião foram colocadas para que o utilizador pudesse deixar sugestões e comentários. Foi realizado um teste piloto com três utilizadores. Através dos testes foi possível melhorar o questionário e torná-lo mais perceptível.

4.3.3 Procedimentos

O questionário foi enviado através de e-mail. No e-mail foi incluído o link para o questionário, o contexto do estudo, os objetivos, recrutamento para posteriores testes de utilizador, o tempo necessário para preencher o questionário (aproximadamente) e um agradecimento. Foram incluídas descrições e exemplos para que os termos em estudo fossem claros para os participantes e evitar discrepâncias. As respostas foram recolhidas entre as datas 11/05/2018 e 11/06/2018 e guardadas automaticamente através da plataforma Google Forms. Os dados foram posteriormente analisados.

A partir dos dados recolhidos e da análise feita, foi possível obter algumas respostas quanto às funcionalidades, requisitos e informações necessárias na plataforma. No entanto, devido ao grau de participação, não é possível fazer generalizações, apenas descrever os resultados.

4.3.4 Análise dos dados

A análise dos dados é meramente descritiva. Os dados foram guardados através da Google Forms numa folha de cálculo do Excel. Essa organização dos dados permitiu obter gráficos que facilitaram o processo de análise. Os principais objetivos eram obter respostas relativamente aos níveis e dimensões de análise mais importantes para os utilizadores.

4.3.5 Caraterização da amostra

A amostra é constituída por treze elementos, 7 (53,8%) são mulheres, 5 (38,5%) são homens e 1 (7,7%) não quis revelar o sexo - [Figura 4.5](#). Em termos etários temos, 6 (46,2%) com idade compreendida entre 45 a 54 anos, 3 (23,1%) com idade entre 25 a 34 anos, 2 (15,4%) com 35 a 44 anos, 1 (7,7%) com idades entre 55 a 64 anos e o mesmo valor para a faixa etária de 15 a 24 anos - [Figura 4.6](#).

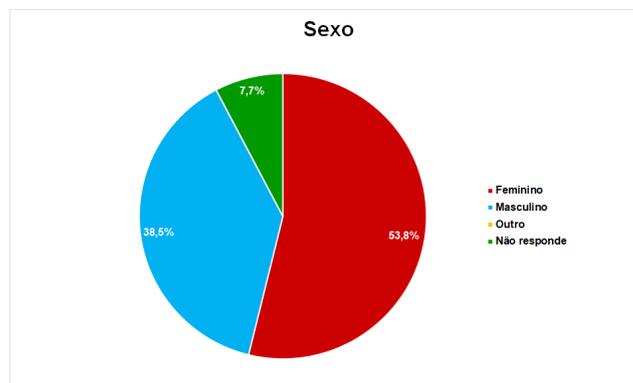


Figura 4.5: Sexo dos inquiridos do questionário.

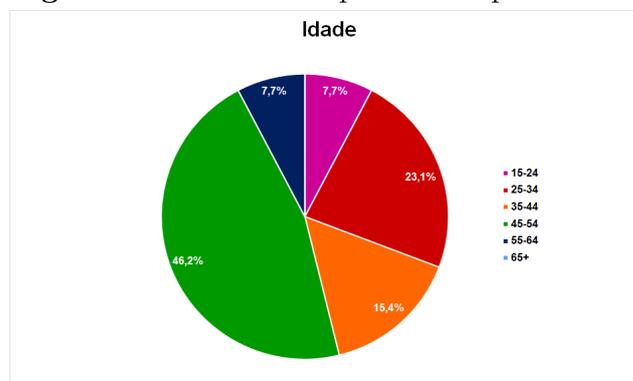


Figura 4.6: Faixa etária dos inquiridos do questionário.

No que diz respeito à profissão temos: 4 (30,8%) jornalistas, 3 (23,1%) estudantes de Mestrado, 2 (15,4%) estudantes de Doutoramento, o mesmo valor para professores universitários e para professores - [Figura 4.7](#). Relativamente à instituição temos, como podemos ver na [Figura 4.8](#), 4 (33,3%) inquiridos pertencentes à Faculdade de Letras da Universidade do Porto (FLUP), 2 (16,7%) ao Jornal de Notícias, e depois temos 1 (8,3%) inquirido para cada uma destas instituições: IPAMAIA, TVI, Universidade do Porto, E-Konomista, ISMAI e FEUP. Um dos inquiridos não revelou a instituição à qual pertence.

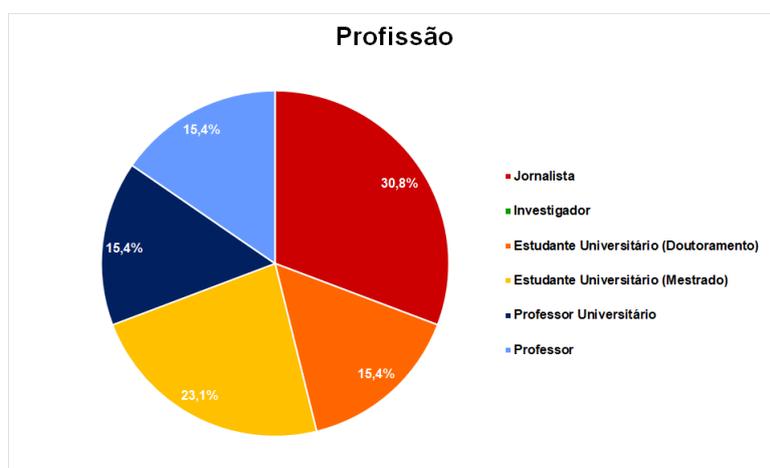


Figura 4.7: Profissão dos inquiridos do questionário.

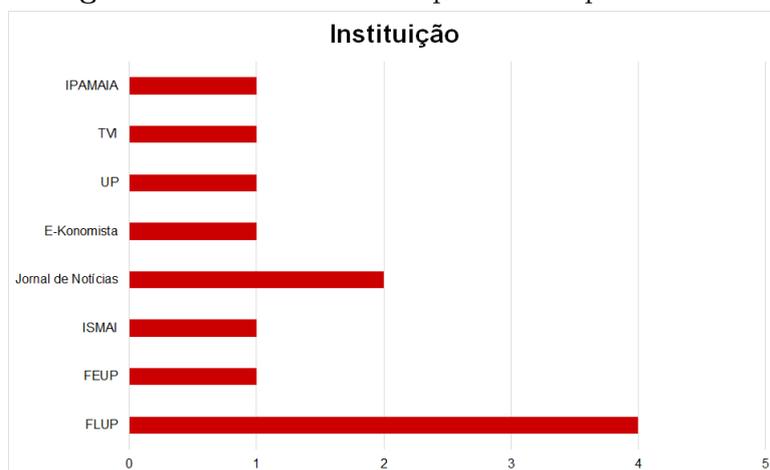


Figura 4.8: Instituição dos inquiridos do questionário.

Quanto área científica de formação, 7 (53,8%) dos inquiridos estão ligados à área de Ciências da Comunicação, e depois temos 1 (7,7%) para cada uma das outras áreas: Ciência da Informação, Informática, Multimédia, Interação Humano-Computador, Humanidades e Economia - [Figura 4.9](#).

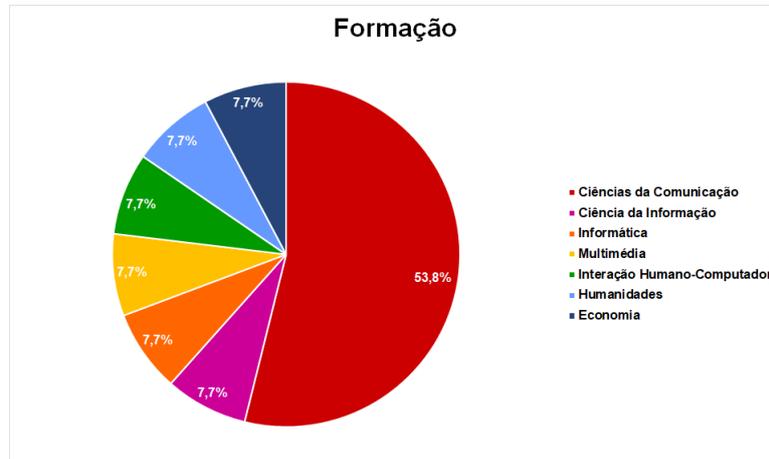


Figura 4.9: Área científica de formação dos inquiridos do questionário.

4.3.6 Síntese

Neste capítulo foram apresentadas as metodologias utilizadas neste estudo não experimental. A grelha de análise desenvolvida foi aplicada às seis plataformas já existentes e estudadas no [Capítulo 3](#) e à plataforma inicial *MediaViz*. O questionário teve como base a grelha e foi possível recolher respostas de treze participantes. No questionário foram abordadas as medidas das grelhas: níveis e dimensões de análise. Os dados recolhidos foram analisados através da elaboração de gráficos.

Estes dois elementos permitirão obter conclusões relativamente às opções que podem ser tomadas no futuro desenvolvimento e melhoramento da plataforma *MediaViz*. Perceber quais as necessidades do público-alvo e quais são as ofertas de que este já dispõe é essencial para criar uma plataforma inovadora e eficaz.

Capítulo 5

Resultados

Este capítulo apresenta os resultados provenientes da análise dos dados do questionário que medem a importância dos níveis e dimensões de análise. Os gráficos apresentam no eixo vertical os níveis ou dimensões de análise e a importância está codificada como apresenta a legenda em baixo.

O eixo horizontal apresenta a percentagem relativa às respostas em ordem crescente de importância - do nível 1 “Nada importante” ao nível 5 - “Muito importante”. Os valores da escala de Likert foram transformados em valores numéricos de forma a possibilitar o cálculo da média. Desta forma, o valor 1 corresponde a “Nada importante”, 2 a “Pouco importante”, 3 a “Moderadamente importante”, 4 a “Importante” e 5 “Muito importante”. Apenas as questões de resposta aberta não tiveram total participação, todas as outras foram respondidas por todos os participantes.

5.1 Resultados do questionário

5.1.1 Níveis de análise

No que diz respeito aos níveis de análise foi desenvolvida apenas uma questão com uma escala de Likert que representa o grau de importância de cada um dos níveis e foi colocada uma questão de opinião/sugestão aberta.

Nesta questão o nível de análise “Coleção” não foi inserido pois este engloba todos os dados pertencentes aos restantes níveis e o principal objetivo era perceber, em termos de objeto de estudo, qual seria o mais importante e quais os dados que seriam mais relevantes de filtrar através da coleção de dados.

Como podemos verificar através da [Figura 5.1](#), a fonte é o nível de análise com maior importância para os participantes ($M=4.77, DP=0.6$), seguindo-se o documento com média igual a 4.54 ($DP=0.66$). Cerca de 73% de participantes avaliaram estes dois níveis como “Muito importante”. O nível referente ao conjunto de frases que constituem a coleção obteve uma média de 3.69 ($DP=0.94$) tendo uma avaliação de “Importante”. Por último, os termos demonstraram ser o nível menos importante obtendo uma média de 3.46 ($DP=1.05$), apresentando uma avaliação de nível 3 (“Moderadamente importante”).

Desta forma, podemos dizer que, se tivermos de ordenar cada um dos níveis de análise por grau de importância, temos:

1. Fonte;
2. Documento;
3. Frase;
4. Termo.

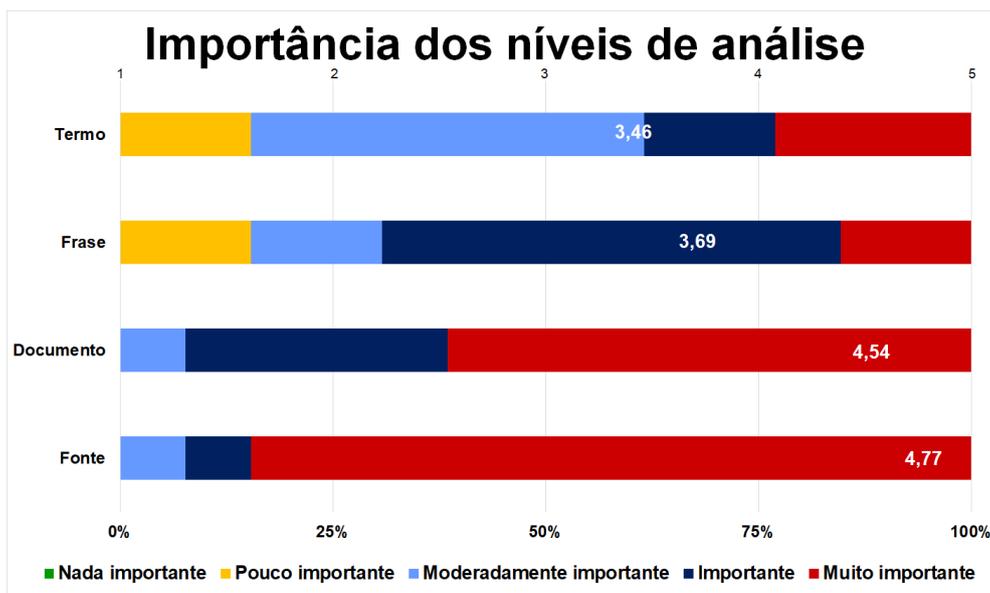


Figura 5.1: Importância dos níveis de análise.

Relativamente à questão aberta não foram recolhidas quaisquer respostas.

5.1.2 Dimensões de Análise

Para a análise ao nível das dimensões foram desenvolvidas seis questões com escala de Likert para avaliar a importância de cada uma das dimensões e uma questão de opinião/sugestão aberta. As seis questões relacionam os níveis e as dimensões de análise.

A primeira questão diz respeito à importância das dimensões de análise relativamente ao estudo de uma coleção de documentos noticiosos. Através da [Figura 5.2](#) é possível verificar que todas as dimensões apresentam, em média, uma avaliação próxima ao nível 4 (“Importante”). As entidades não geográficas são a dimensão com maior média ($M=4.15, DP=0.69$), seguindo-se os tópicos ($M=4.00, DP=0.71$), a frequência ($M=4.00, DP=0.82$) e a dimensão temporal ($M=3.92, DP=0.76$). Por último, temos a dimensão geográfica ($M=3.62, DP=0.87$).

A distribuição do grau de importância pelas diferentes dimensões de análise, no que diz respeito à coleção de documentos, mostrou-se bastante equilibrada. No entanto, podemos listar a importância de cada uma das dimensões da seguinte forma:

1. Entidades não geográficas;
2. Tópicos;
3. Frequência;
4. Temporal;
5. Geográfica.

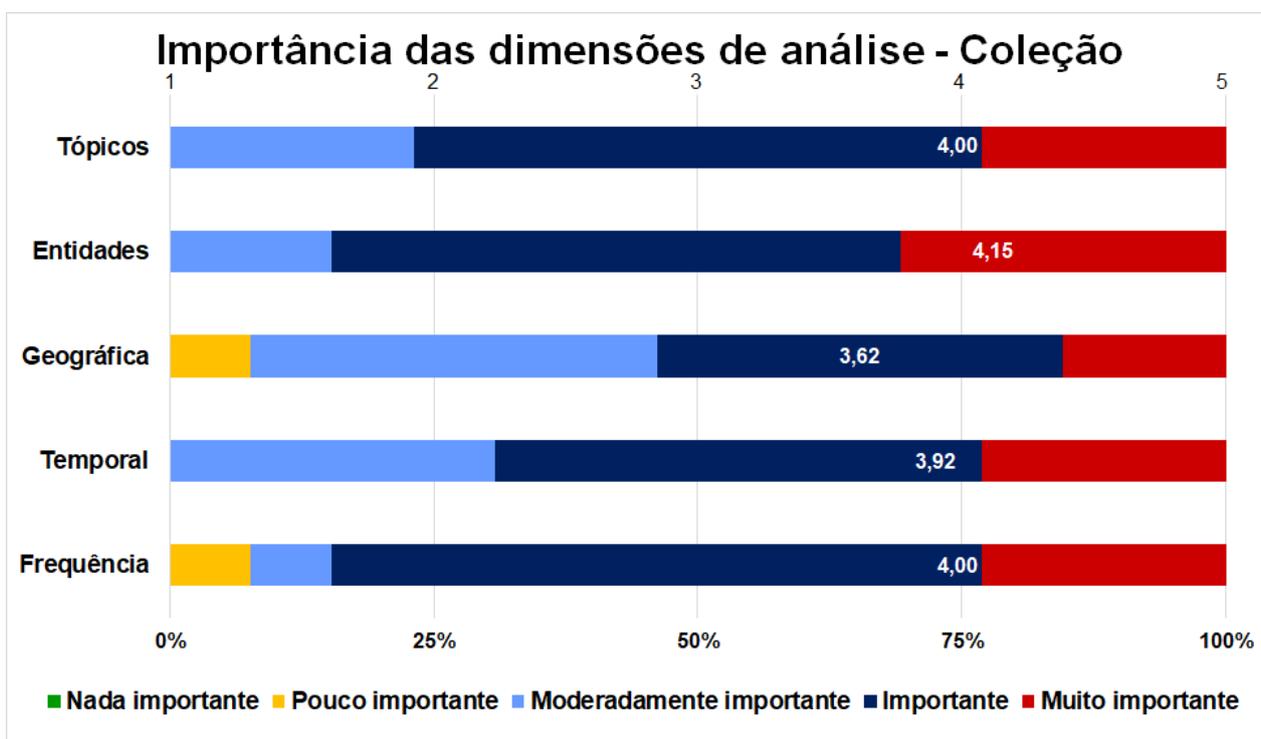


Figura 5.2: Importância das dimensões de análise quando o objeto de estudo é a coleção.

A segunda questão está relacionada com a importância dos tipos de entidades que é possível obter. A [Figura 5.3](#) apresenta os resultados obtidos. É possível constatar que as organizações e as pessoas apresentam resultados iguais em todos os graus apresentando uma média de 4,31 (DP=0.63), correspondendo ao grau de importância “Importante”. Os locais apresentam uma média de 3.46 (DP=0.78), dispondo de uma avaliação de nível 3 (“Moderadamente importante”). Desta forma, ao nível da importância das entidades temos:

1. Pessoas e organizações;
2. Locais.

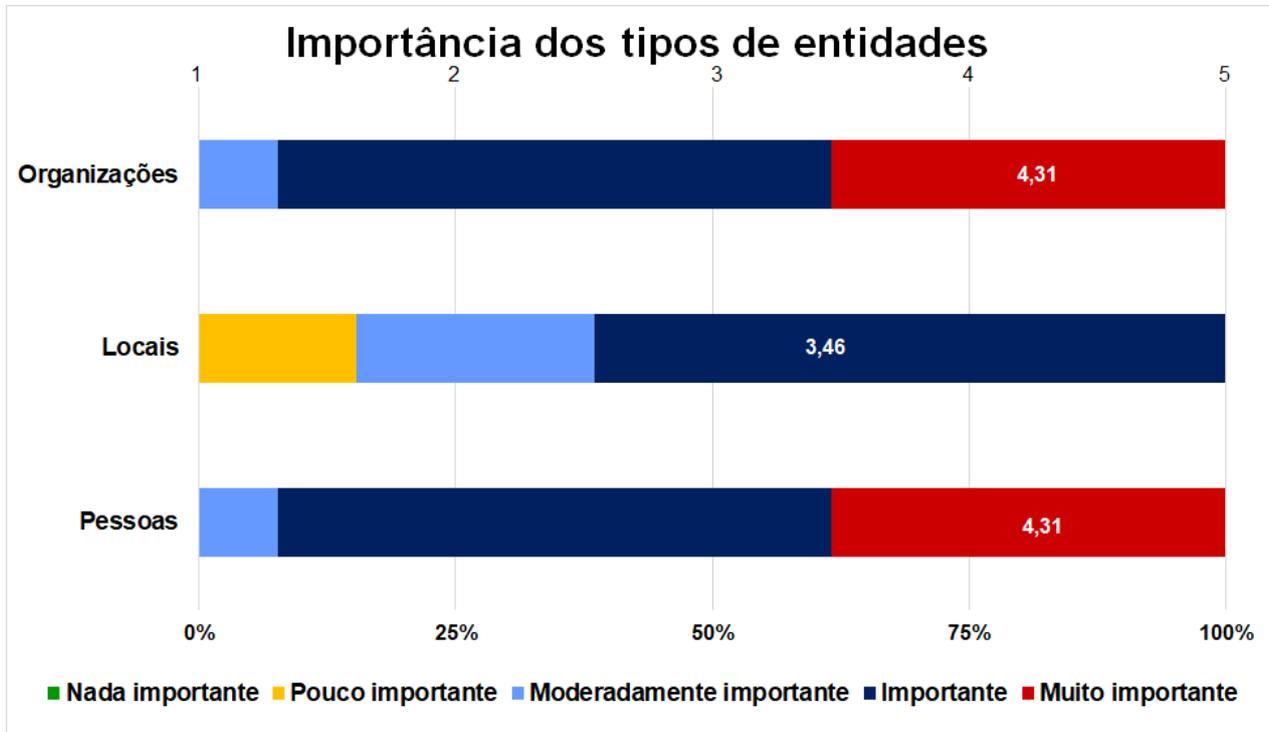


Figura 5.3: Importância dos tipos de entidades.

A terceira questão relaciona as dimensões de análise e as fontes de origem dos documentos. É possível verificar que, todas as dimensões apresentam, em média, uma avaliação próxima ao nível 4 (“Importante”). A dimensão de análise frequência apresenta uma média superior às restantes ($M=4.08, DP=0.86$), seguindo-se as entidades não geográficas ($M=3.92, DP=0.49$), a dimensão temporal ($M=3.77, DP=0.73$) e os tópicos ($M=3.69, DP=0.75$). Por último, temos a dimensão geográfica, sendo a menos importante para os participantes ($M=3.46, DP=0.78$).

Desta forma, podemos organizar as dimensões de análise, ao nível de análise das fontes, da seguinte forma:

1. Frequência;
2. Entidades não geográficas;

3. Temporal ;
4. Tópicos;
5. Geográfica.

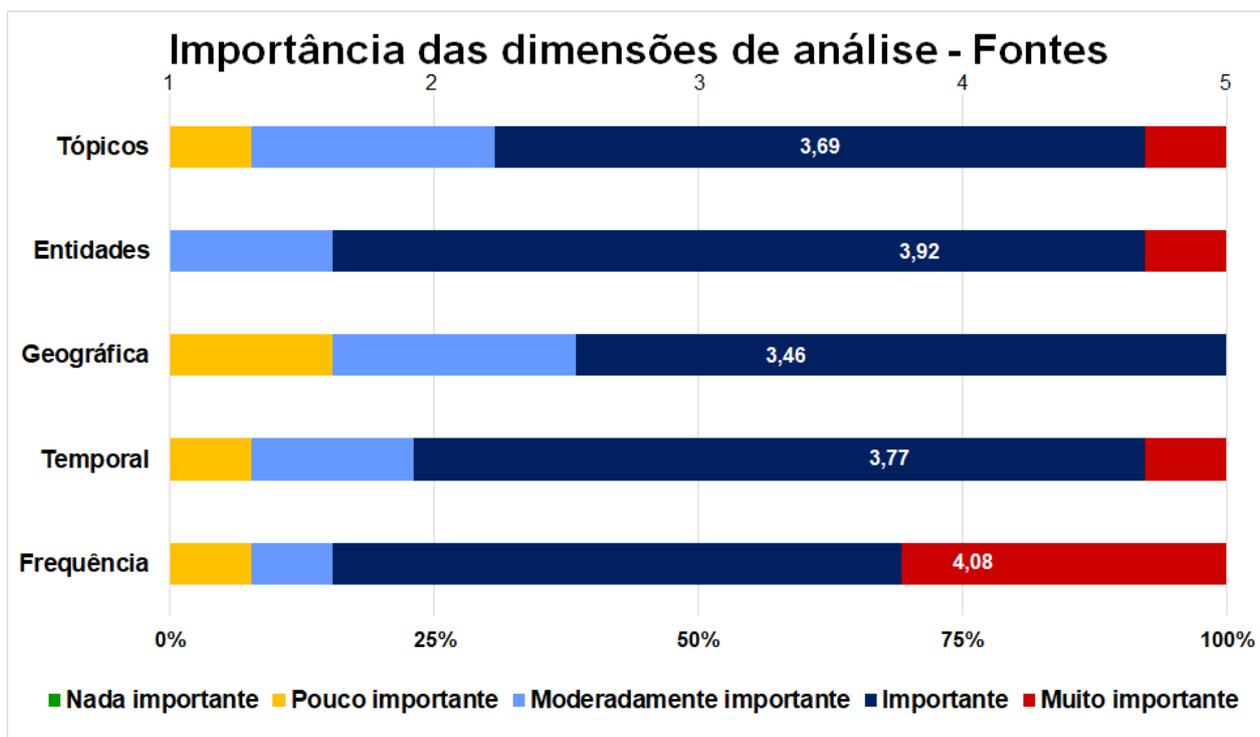


Figura 5.4: Importância das dimensões de análise quando o objeto de estudo é a fonte de origem dos documentos.

Por sua vez, a quarta questão está relacionada com a análise dos documentos que existem na coleção de dados e as dimensões de análise. Através da [Figura 5.5](#) é possível analisar os resultados obtidos no questionário. Assim sendo, é possível constatar que os tópicos, as entidades não geográficas, a frequência e a dimensão temporal foram avaliadas, em média, como “Importante”. Os tópicos obtiveram uma média de 4.23 (DP=0.6), as entidades não geográficas de 4.00 (DP=0.58), seguindo-se a frequência com 3.85 (DP=0.69) e a dimensão temporal com 3.54 (DP=0.88). Por último, a dimensão geográfica apresenta uma média de 3.23 (DP=1.01) tendo sido atribuída a escala de “Moderadamente importante” e sendo a dimensão menos importante para os participantes.

Desta forma, podemos organizar as dimensões de análise no que diz respeito ao objeto de estudo ser o documento da seguinte forma:

1. Tópicos;
2. Entidades não geográficas;
3. Frequência;
4. Temporal;
5. Geográfica.

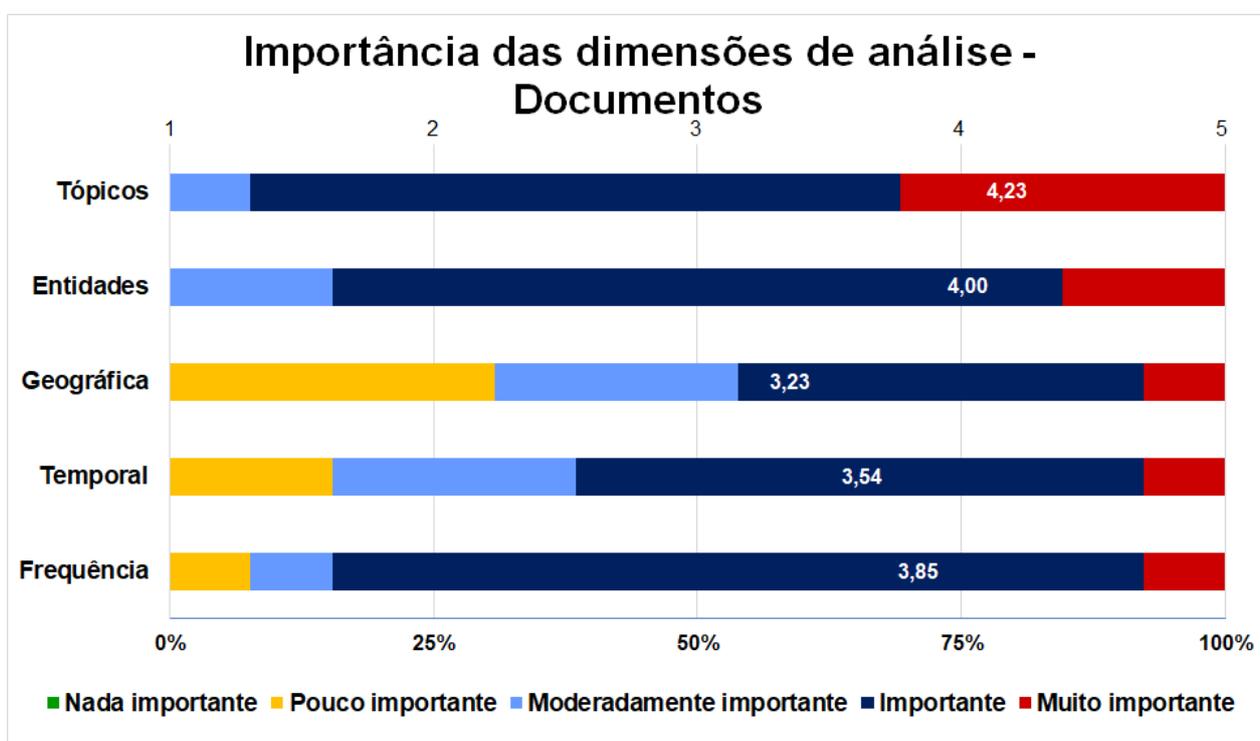


Figura 5.5: Importância das dimensões de análise quando o objeto de estudo são os documentos da coleção de dados.

A quinta questão diz respeito às dimensões de análise e às frases existentes na coleção de dados. Através da [Figura 5.6](#) é possível analisar os dados obtidos nesta questão. Verificamos que as dimensões dos tópicos, das entidades não geográficas e da frequência apresentam, em média, uma avaliação na escala 4 (“Importante”). Os tópicos (DP=0.77) e as entidades não geográficas (DP=0.65)

apresentam uma média igual ($M=3.62$), mas um desvio padrão diferente. Já a dimensão da frequência apresenta uma média de 3.54 ($DP=0.66$). Por último, a dimensão temporal ($M=3.38, DP=0.77$) e geográfica ($M=3.00, DP=0.71$) apresentam uma avaliação na escala 3 (“Moderadamente importante”).

A organização da importância de cada uma destas dimensões, no que diz respeito à análise das frases existentes na coleção, pode ser feita da seguinte forma:

1. Tópicos;
2. Entidades não geográficas;
3. Frequência;
4. Temporal;
5. Geográfica.

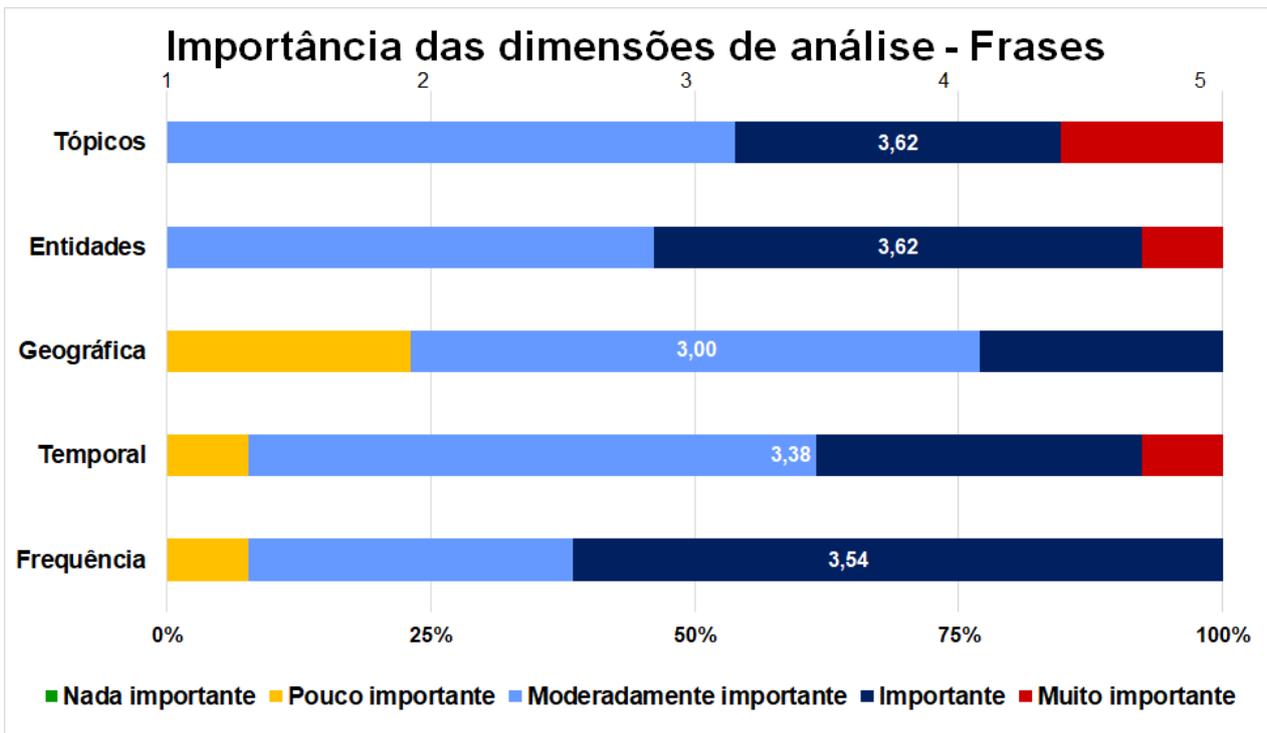


Figura 5.6: Importância das dimensões de análise quando o objeto de estudo são as frases existentes na coleção de dados.

A sexta questão, e última referente às dimensões de análise, relaciona essas dimensões e a análise dos termos existentes na coleção de dados. A [Figura 5.7](#) apresenta os resultados obtidos através desta questão. Desta forma, podemos constatar que os tópicos ($M=3.85, DP=0.89$), as entidades não geográficas ($M=3.77, DP=0.60$) e a frequência ($M=3.69, DP=0.63$) foram avaliados de forma similar, obtendo, por parte dos participantes, uma escala de “Importante”. Por último, as dimensões temporal ($M=3.15, DP=0.80$) e geográfica ($M=3.08, DP=0.64$) foram avaliadas pelos participantes como “Moderadamente importante” sendo, comparativamente às anteriores, menos importantes. De referir que, a dimensão temporal foi a única que obteve uma resposta no grau de importância “Nada importante”, em relação a todas as questões.

Esta análise permite organizar as dimensões da seguinte forma:

1. Tópicos;
2. Entidades não geográficas;
3. Frequência;
4. Temporal;
5. Geográfica.

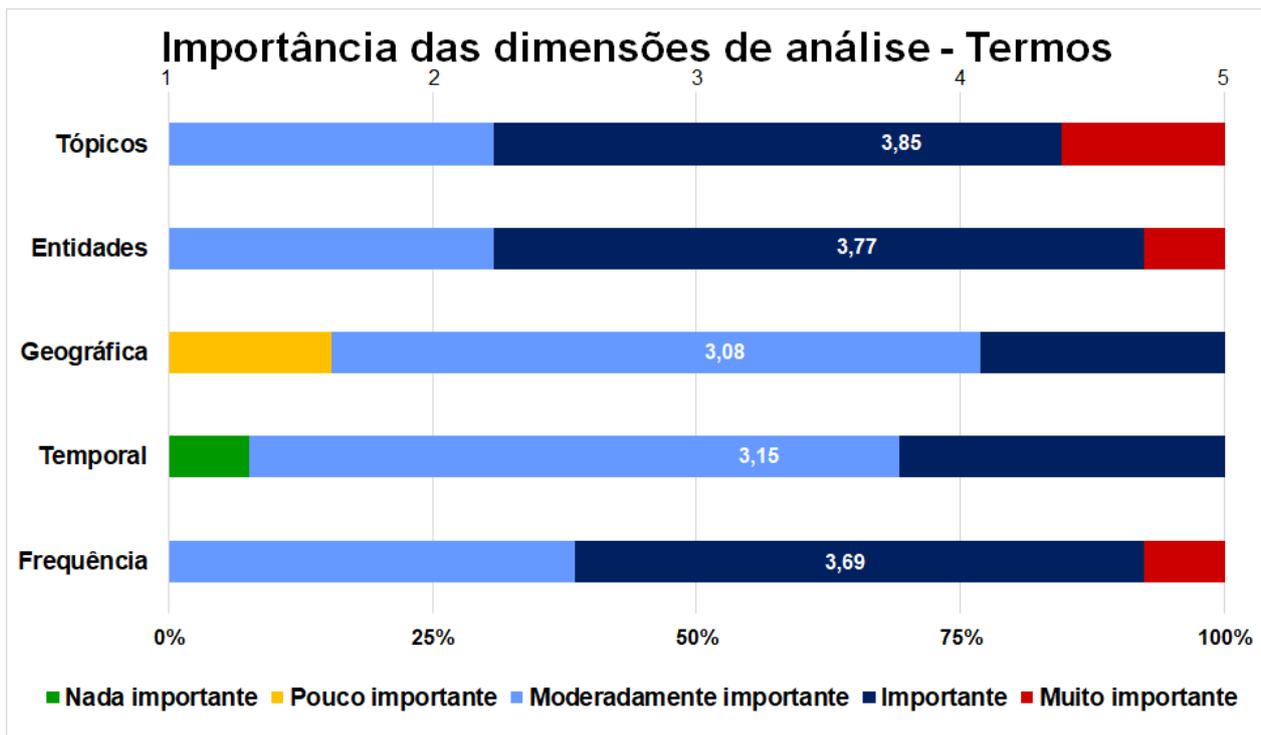


Figura 5.7: Importância das dimensões de análise quando o objeto de estudo são os termos existentes na coleção de dados.

Por último, a questão de opinião teve duas respostas:

“O inquérito tem perguntas que dependem muito dos objetivos e objetos de análise. Ou seja, as respostas tiveram de ser dadas sem ter essa informação.”

“O grau de importância depende do objetivo da busca e da análise, por isso, não é possível fazer uma classificação abstrata.”

5.2 Síntese dos resultados e discussão

Neste capítulo foram apresentados os resultados obtidos através da análise das respostas dadas pelos participantes ao questionário. Foram assim avaliadas a importância dos níveis de análise e a importância das dimensões de análise nesses mesmos níveis. Através da [Tabela 5.1](#) é possível verificar as conclusões retiradas através da análise dos dados. Esta tabela apresenta os níveis de análise do mais importante para o menos importante (de cima para baixo) e as dimensões (da esquerda para a direita). Em relação aos níveis de análise podemos afirmar que

as fontes, ou seja, o conjunto de fontes de origem dos documentos, é bastante importante para os participantes do questionário, assim como os documentos. Relativamente às dimensões de análise verificamos que, em relação à coleção de dados, as entidades não geográficas são um elemento fundamental para os inquiridos. Isto é, é importante a análise das entidades não geográficas mencionadas na coleção de dados que é o objeto de estudo. Mais especificamente, no que diz respeito às entidades não geográficas, os participantes dão primazia às pessoas e às organizações. Relativamente ao nível de análise visto como sendo o mais importante pelos participantes, fonte, as dimensões de análise mais importantes são a frequência e as entidades não geográficas. Quanto à análise dos documentos integrais da coleção, os participantes preferem uma análise relativa aos temas (tópicos) e às entidades não geográficas mencionadas. O mesmo acontece quando o objeto de estudo são as frases ou os termos que constituem a coleção de dados.

Podemos então verificar que os tópicos e as entidades não geográficas foram as dimensões de análise que obtiveram, por parte dos participantes, mais interesse. Sendo que, apenas na análise das fontes foi dada primazia à frequência. Em todos os outros níveis temos os tópicos e as entidades não geográficas como as dimensões de maior importância. Para além disso, no nível inferior de importância temos a análise geográfica. Isto vai de encontro aos dados obtidos na segunda questão, em que os participantes deram maior importância às pessoas e organizações, comparativamente aos locais mencionados.

No que diz respeito à questão de opinião, já era algo esperado sendo que, o questionário tem de abordar em geral os objetivos do utilizador, e não seria possível fazer algo mais objetivo, pois teríamos de deixar níveis ou dimensões de análise de parte. Isto também porque, todos os possíveis utilizadores da plataforma em estudo, e que será desenhada, terão, certamente, objetivos e objetos de análise diferentes.

Tabela 5.1: Resultados obtidos através da análise dos dados.

Níveis de análise	Dimensões de análise				
Coleção	Entidades não geográficas	Tópicos	Frequência	Temporal	Geográfica
Fonte	Frequência	Entidades não geográficas	Temporal	Tópicos	Geográfica
Documento	Tópicos	Entidades não geográficas	Frequência	Temporal	Geográfica
Frase	Tópicos	Entidades não geográficas	Frequência	Temporal	Geográfica
Termo	Tópicos	Entidades não geográficas	Frequência	Temporal	Geográfica

Como referido no [Capítulo 1](#) os principais objetivos deste estudo são:

1. Identificar as plataformas existentes neste campo e quais as ferramentas e informações que disponibiliza aos seus utilizadores;
2. Identificar os critérios que os utilizadores reconhecem ser os fundamentais para obterem uma boa experiência aquando da navegação neste tipo de plataformas;
3. Identificar as informações fundamentais para os utilizadores, como por exemplo: fontes, entidades, temas mencionados, entre outros.

Desta forma, mais do que fazer uma análise à grelha desenvolvida e aos resultados obtidos através do questionário, torna-se fundamental cruzar estes dois elementos. Esta nova análise permitirá perceber de que forma as plataformas existentes (ponto 1 dos objetivos) respondem, ou não, aos objetivos dos utilizadores (ponto 2 e 3).

Fazendo uma comparação entre a escala de importância dada a cada um dos níveis de análise ([Tabela 5.1](#)) e as ferramentas que foram analisadas ([Tabela 5.2](#)) podemos verificar que: a plataforma PageOneX e a MediaViz preencheriam todos os requisitos, no entanto, como foi referido na análise comparativa, a plataforma PageOneX apenas apresenta e permite a codificação das capas dos jornais. Depois desta plataforma, encontramos a Europe Media Monitor, a PickaNews e a MediaCloud. O único nível que falta em todas estas plataformas é o documento. Por

último, temos o Google Trends e a GDELT Project. É importante referir que, o segundo nível de análise mais importante para os participantes do questionário (Fonte), não é alvo de análise em duas das sete plataformas (Google Trends e GDELT Project), já o terceiro nível de análise (Documento) apenas é analisado na plataforma PageOneX e na MediaViz.

Relativamente às dimensões de análise ([Tabela 5.3](#)), é possível verificar que, retirando a plataforma PageOneX pelos motivos já referidos, de entre as outras plataformas, podemos destacar a Europe Media Monitor, a MediaCloud e a GDELT Project. Desta forma, estas seriam as plataformas que melhor corresponderiam às necessidades dos utilizadores. Pessoalmente, destacaria a plataforma MediaCloud pela facilidade de obtenção de resultados, pelas ferramentas que disponibiliza aos utilizadores, pela forma como transmitem visualmente os dados e pelas inúmeras possibilidades de pesquisa desde através de palavras-chave, conjunto de termos, por fontes, entre outras. A GDELT Project é, também, uma ferramenta que disponibiliza bastantes dados aos utilizadores, mas as visualizações implicam mais trabalho da parte do utilizador. É também importante referir que, ainda relativamente às dimensões de análise, é possível verificar que três das sete plataformas em análise, PickaNews, Google Trends e MediaViz, não apresentam qualquer análise relacionada com as entidades não geográficas mencionadas no objeto de estudo. Se formos verificar a tabela que resume o grau de importância dado pelos participantes do questionário, verificamos que as entidades não geográficas são a segunda dimensão mais importante. Também os tópicos não merecem qualquer atenção por parte das plataformas PickaNews e MediaViz.

Tabela 5.2: Níveis de análise utilizados pelas plataformas em análise.

Níveis de análise	Europe Media Monitor	PickaNews	Google Trends	MediaCloud	PageOneX	GDELT Project	MediaViz
Coleção							
Fonte							
Documento							
Frase							
Termo							

Tabela 5.3: Dimensões de análise utilizadas pelas plataformas em análise.

Dimensões de análise	Europe Media Monitor	PickaNews	Google Trends	MediaCloud	PageOneX	GDELT Project	MediaViz
Frequência							
Temporal							
Geográfica							
Entidades não geográficas							
Tópicos							

Capítulo 6

Conclusões

A necessidade de criar uma plataforma que facilite a interpretação de dados noticiosos por parte do utilizador é a base deste estudo. Desta forma, a exploração visual dos dados é a área principal em estudo pois esta permite que o utilizador seja capaz de visualizar padrões, tendências e discordâncias, facilitando a análise e compreensão (Heer et al., 2010). Nesse sentido, foi criada uma grelha de análise que permitiu agregar informação relativa às plataformas já existentes e de que forma estas transmitem os dados aos seus utilizadores. Com este elemento foi possível responder ao primeiro objetivo do estudo. Após a análise dessa grelha, foi desenvolvido um questionário que permitiu avaliar os níveis e dimensões de análise fundamentais para o utilizador e, conseqüentemente, responder ao segundo e terceiro objetivo do projeto. Esse questionário foi preenchido por treze participantes nos quais estão incluídos jornalistas, investigadores e estudantes da área dos media. As conclusões retiradas através da aplicação da grelha nas diferentes plataformas já existentes, e através do questionário, servirão para o futuro desenho do observatório media online MediaViz.

Em relação aos níveis de análise, foi possível verificar que os níveis mais importantes são: o conjunto de fontes e o conjunto de documentos que fazem parte da coleção de dados. No que diz respeito às dimensões de análise temos, como mais importantes para os participantes do questionário, os tópicos e as entidades não geográficas. É fundamental referir que, estas duas dimensões não apresentam qual-

quer análise na versão atual da plataforma MediaViz. Desta forma, os resultados obtidos através do questionário permitem verificar que essa análise é importante para os utilizadores e deve, conseqüentemente, fazer parte da reformulação que será feita futuramente. A dimensão menos importante para os participantes é a geográfica. Algo que vai de encontro à pergunta colocada no questionário referente à importância de cada tipo de entidade em que foi dada preferência às pessoas e organizações, em detrimento dos locais.

Estes resultados permitem visualizar uma hierarquia relativamente aos níveis e dimensões de análise mais importantes para os utilizadores. Sendo importantes para os utilizadores, é fulcral que estes estejam incluídos na nova plataforma.

Os dados recolhidos, a análise dos mesmos e o estudo feito no [Capítulo 2](#) e [Capítulo 3](#) permitirão, no futuro, desenvolver uma plataforma que vá de encontro às necessidades do público-alvo, que forneça ferramentas para este atingir os seus objetivos e que forneça, num só local, informação relevante para o utilizador, sem que este tenha de saltar entre plataformas para realizar as suas tarefas. Para além da visualização da informação, a usabilidade será outra área que deverá ser estudada e aplicada nessa plataforma. Pois, para que uma interface seja aceite pelo utilizador é necessário que esta seja fácil de aprender a usar, fácil de utilizar e provocar satisfação no mesmo (Carvalho, 2002). Na medida que, em termos de funcionalidade, o software pode estar corretamente concebido, mas se a usabilidade do mesmo não for boa, o utilizador irá rejeitá-lo (Carvalho, 2002).

É de referir que este estudo apresenta algumas limitações como a utilização de um método de amostragem não probabilística que limita a generalização dos resultados. O questionário foi feito especialmente para este estudo e, portanto, não foi validado e não teve a participação esperada. No entanto, acredito que as descobertas aqui feitas possibilitarão o desenho de uma plataforma que vá de encontro às necessidades do utilizador e estes dados poderão ser utilizados não só para a plataforma MediaViz, mas também adaptados a outras plataformas.

6.1 Contribuições

Este estudo apresenta contribuições ao nível do estudo das plataformas existentes para a análise de dados, apresenta uma grelha de análise que pode ser aplicada a outras plataformas e, por último, apresenta informação relevante para o desenho da plataforma MediaViz. A primeira contribuição permite, não só no âmbito deste estudo, mas de outros que poderão surgir, perceber o que já foi feito ao nível das plataformas que permitem a visualização de dados. A segunda, permite analisar essas plataformas ou outras que poderão surgir. E, por último, o principal objetivo deste estudo, apresenta informação que permitirá desenhar a plataforma MediaViz de forma a que os seus utilizadores a utilizem regularmente e se sintam satisfeitos com a mesma.

6.2 Trabalho Futuro

Relativamente aos trabalhos futuros podemos incluir:

- Estudo de usabilidade;
- Desenho de wireframes;
- Testes de utilizadores;
- Melhoramento das wireframes (se necessário);
- Protótipo de alta fidelidade;
- Testes de utilizadores e melhoramento;
- Implementação.

Sedo este estudo a sustentação do desenho da plataforma MediaViz, os próximos passos acima referidos dizem respeito ao desenho dessa mesma plataforma, tendo em conta as descobertas aqui feitas.

Bibliografia

Pedro Jerónimo (org.) Ana Isabel Reis, Fernando Zamith, Helder Bastos. Cibermeios portugueses: 10 anos de lenta evolução. In *V Congresso Internacional de Ciberjornalismo - Livro de Atas*, number May, pages 24–37, 2017. ISBN 9789899819924.

Stuart K. Card, Jock D. Mackinlay, e Ben. Shneiderman. *Readings in information visualization : using vision to think*. Morgan Kaufmann Publishers, 1999. ISBN 1558605339. URL <https://dl.acm.org/citation.cfm?id=300679>.

Ana Carvalho. Testes de Usabilidade: exigência supérflua ou necessidade? In *Actas do 5º Congresso da sociedade Portuguesa das Ciências da Educação.*, pages 235–242. Sociedade Portuguesa de Ciências da Educação, Lisboa, 2002. URL <http://www.lits.dei.uminho.pt/tu.pdf>.

Alan Cooper, Robert Reimann, e David Cronin. *About Face 3: The essentials of interaction design*. 2007. ISBN 9780470084113.

Felipe Garcia da Costa. *Visualização de dados e a sua importância na era do big data*. PhD thesis, 2017. URL <http://www.lume.ufrgs.br/handle/10183/169109>.

Tiago Devezas, Sérgio Nunes, e María Teresa Rodríguez. MediaViz: An Interactive Visualization Platform for Online Media Studies. *Proceedings of the 2015 International Workshop on Human-centric Independent Computing - HIC '15*, pages 7–11, 2015. doi: 10.1145/2808469.2808474. URL <http://dl.acm.org/citation.cfm?doid=2808469.2808474>.

Tiago Devezas, José Devezas, e Sérgio Nunes. Exploring a large news collection using visualization tools. In *CEUR Workshop Proceedings*, volume 1568, pages 48–53, 2016.

- Hugo A. D. do Nascimento e Cristiane B. R. Ferreira. Uma introdução à visualização de informações. *Visualidades*, 9(2):13–44, aug 2012. ISSN 2317-6784. URL <http://www.revistas.ufg.br/index.php/VISUAL/article/view/19844>.
- European Commission. EMM Overview, 2018. URL <http://emm.newsbrief.eu/overview.html>.
- David Hand, Heikki Mannila, e Padhraic Smyth. *Principles of Data Mining*. 2001. ISBN 9780262082907.
- Jeffrey Heer, Michael Bostock, e Vadim Ogievetsky. A tour through the visualization zoo. *Communications of the ACM*, 53(6):59, jun 2010. URL <http://portal.acm.org/citation.cfm?doid=1743546.1743567>.
- Daniel A. Keim. Visual exploration of large data sets. *Communications of the ACM*, 44(8):38–44, 2001. ISSN 00010782.
- Marja-Ritta Koivunen e Matt May. Exploring Usability Enhancements in W3C Process, 2002. URL <http://www.w3.org/2002/Talks/0104-usabilityprocess/>.
- Lev Manovich. Media Visualization: Visual Techniques for Exploring Large Media Collections. *K. Gates, Media studies futures. Blackwell. Retrieved from http://softwarestudies.com/cultural_analytics/Manovich.Media_Visualization.web*, (June):1–21, 2011. doi: 10.1002/9781444361506. URL <http://manovich.net/index.php/projects/media-visualization-visual-techniques-for-exploring-large-media-collections>.
- Luís Mendes. *Avaliação De Usabilidade Em Sistemas Web - Desktop*. PhD thesis, Universidade do Porto, 2014. URL <https://repositorio-aberto.up.pt/bitstream/10216/74600/2/31920.pdf>.
- Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, David AL Levy, e Leis Nielsen. Reuters Institute Digital News Report 2017. Technical report, 2017.

URL https://reutersinstitute.politics.ox.ac.uk/sites/default/files/DigitalNewsReport2017web_0.pdf?utm_source=digitalnewsreport.org&utm_medium=referral.

Petra Kralj Novak, Miha Grcar, e Igor Mozetic. Analysis of financial news with NewsStream. Technical report, IJS-DP-11892, 2015.

OberCom. Anuário da Comunicação 2016 – 2017 – Obercom. Technical report, 2017. URL <https://obercom.pt/wp-content/uploads/2018/06/2017-OBERC-OM-Anuario-da-Comunicacao.pdf>.

Pickanews. FAQ Pickanews. URL <https://www.pickanews.com/faq>.

B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343. IEEE Comput. Soc. Press, 1996. ISBN 0-8186-7508-X. doi: 10.1109/VL.1996.545307. URL <http://ieeexplore.ieee.org/document/545307/>.

The Berkman Center for Internet and Society. Media Cloud, 2009. URL <https://explorer.mediacloud.org/#/home>.

The GDELT Project. The GDELT Story, 2013. URL <http://www.gdeltproject.org/about.html>.

Fernando Rosa Vaz e Cedric Luiz de Carvalho. Visualização de Informações. Technical report, Universidade Federal de Goiás, 2004. URL http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-04.pdf.

Matthew Ward, Georges G. Grinstein, e Daniel Keim. *Interactive data visualization : foundations, techniques, and applications*. CRC Press, 2 edition, 2010. ISBN 9781482257373. URL <https://www.crcpress.com/Interactive-Data-Visualization-Foundations-Techniques-and-Applications/Ward-Grinstein-Keim/p/book/9781482257373>.

Ji Soo Yi, Youn ah Kang, John T. Stasko, e Julie A. Jacko. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 13(6):1224–1231, 2007. doi: 10.1109/TVCG.2007.70515.

Apêndice A

Grelha de análise

Apêndice B

Personas



João Ribeiro

“Sempre estive ligado à área dos média. Desde pequeno que sempre acompanhei as notícias. Roubava sempre o jornal ao meu pai”

Idade: 31 anos;

Nacionalidade: Portuguesa;

Curso: Ciências da Informação da Universidade do Porto;

Profissão: Investigador;

Interesses: Jornalismo, investigação, música, desporto e cinema.

Caraterísticas pessoais

O jornalismo acompanha-o desde pequeno e a resposta à pergunta “O que queres ser quando fores grande?” foi sempre a mesma. Seguir a área da comunicação sempre foi o sonho do João e a Universidade do Porto foi a sua escolha. Realizou a licenciatura, o mestrado e o doutoramento nesta instituição e encontra-se agora na área da investigação. Os jornais foram sempre a sua primeira escolha, mas com o evoluir dos tempos o João não se deixa ficar por aí. O computador e a Internet tornaram-se nos seus melhores aliados ao longo do curso. E, atualmente, encontrar ferramentas que lhe facilitem o trabalho e que o complementem é, para ele,

“uma vatangem enorme, tornando o trabalho não só mais rápido, mas também credível”.

Objetivos e Desafios

Encontrar ferramentas que sejam credíveis e que complementem o trabalho de investigação sempre foi algo complicado para o João. Há diversas ferramentas, mas nem todas são fidedígnas e têm acessos limitados.

Objetivos finais:

- Criar análises credíveis, sustentadas e eficazes;
- Ter ferramentas de trabalho acessíveis e de fácil utilização;
- Encontrar todas as ferramentas num só sítio facilitando-lhe o trabalho;
- Obter dados atualizados.



João Ribeiro

Objetivos experienciais:

- Facilidade e satisfação ao utilizar a plataforma para análise de dados;
- Sentir que consegue controlar a plataforma e obter os resultados pretendidos;
- Focar-se nos objetivos pretendidos.

Objetivos de vida:

- Ser melhor no que faz;
- Aprender cada vez mais sobre a área.

Cenários

- O João tenta sempre procurar as melhores e mais atualizadas ferramentas que o ajudem a encontrar dados e informações atualizadas, credíveis e de fácil manipulação e acesso. Durante uma pesquisa encontra a plataforma MediaViz. Como é

uma ferramenta recente decide investigar e fazer alguns testes. A principal preocupação do João é conseguir obter dados atualizados. Na plataforma este consegue ter de imediato acesso às notícias mais reportadas nas últimas 24h e em constante atualização.

- Para uma das suas investigações o João precisa de saber a quantidade de artigos e a evolução ao longo do tempo de diversas palavras-chave ligadas a determinado tema. Decide fazer a pesquisa na plataforma e obtém os resultados pretendidos. Tem acesso à evolução das palavras-chave ao longo do tempo pretendido e ao número de artigos em determinado dia. Ao longo da análise dos dados chega à conclusão que seria interessante obter palavras-chave associadas às que previamente selecionou e acaba por perceber que a plataforma dispõe dessa ferramenta.

- Uma colega do João está a fazer uma investigação que cruza artigos que incluam a referência a alguma localidade e outra palavra-chave, pretendendo verificar qual das regiões é mais referenciada. Como utilizador já assíduo da plataforma MediaViz ele sabe que uma das ferramentas disponibilizadas permitirá à sua amiga obter os resultados pretendidos e ainda informação complementar, partilhando a plataforma.



Catarina Ferreira

“Nos dias de hoje o jornalismo trabalha de uma forma diferente. A análise dos dados e da informação nem sempre é correta, muitas vezes serve para beneficiar a venda das notícias.”

Idade: 19 anos;

Nacionalidade: Portuguesa;

Curso: Ciências da
Comunicação da

Universidade do Minho;

Profissão: Estudante;

Interesses: Jornalismo,
multimédia, causas humanitárias,
análise de dados e viagens.

Caraterísticas pessoais

A Catarina está a frequentar a Licenciatura em Ciências da Comunicação na Universidade do Minho, encontra-se no segundo ano. O interesse pela comunicação surgiu à pouco tempo.

“Andava pelo Facebook e reparei que muita da informação divulgada não era credível ou nem sempre divulgavam a informação mais pertinente”.

Diz que está no curso para tentar mudar alguma coisa e fazer aquilo que designa de “*bom jornalismo*”. A sua primeira opção seria sempre o curso de Gestão. Analisar e perceber dados foi algo que sempre a cativou. A matemática, que para muitos é um bicho de sete cabeças, era para Catarina algo fascinante.

“Se conseguir aliar os dois mundos, o Jornalismo e a análise de dados, serei uma pessoa realizada”.

Objetivos e Desafios

Obter informação credível e que possa ser analisada.

Objetivos finais:

- Criar análises credíveis, sustentadas e eficazes;
- Obter várias informações acerca de um tema,
- Obter grandes quantidades de dados passíveis de análise.



Catarina Ferreira

Objetivos experienciais:

- Facilidade e satisfação ao utilizar a plataforma para análise de dados;
- Sentir que consegue controlar a plataforma e obter os resultados pretendidos;
- Facilidade na manipulação da visualização dos dados.

Objetivos de vida:

- Praticar bom jornalismo;
- Aliar o jornalismo e a análise de dados.

Cenários

- A Catarina está a frequentar o curso de Ciências da Comunicação e para a unidade curricular de Jornalismo e Informação necessita de recolher dados sobre o total de artigos publicados pelas diversas fontes

online existentes. Para não ter de analisar cada uma das fontes individualmente a Catarina procura na Internet se existe alguma plataforma que a possa ajudar. Ao fazer a pesquisa, a Catarina encontra a plataforma MediaViz. Esta verifica que a plataforma permite obter o total de artigos publicados nas últimas 24 horas por cada fonte e, para além disso, consegue verificar a mesma informação para um intervalo de tempo selecionado por ela. Como gosta de investigar e analisar dados, esta começa a explorar as ferramentas que a plataforma disponibiliza e conclui que lhe permitirá obter um trabalho bastante completo, já que fornece informações complementares.

- A Catarina está a fazer scroll no seu Facebook quando se depara com uma notícia de última hora. Abre a notícia mas esta ainda apresenta poucos detalhes. Decide então ir à plataforma MediaViz e verificar se existem mais fontes a reportar esse assunto. Entra na página inicial e consegue ver que o assunto tem sido noticiado por diversas fontes. Esta tem acesso ao número de artigos e quais as fontes que os divulgaram. Através da mesma plataforma consegue ter ligação para os artigos e verificar se existem mais informações acerca do noticiado.

Apêndice C

Questionário

Design de um Observatório Media Online

No âmbito de um projeto de dissertação do Mestrado em Multimédia da Universidade do Porto, agradeço a sua colaboração no preenchimento do seguinte questionário. Os dados recolhidos serão utilizados no desenvolvimento do protótipo de uma plataforma para análise de dados noticiosos - plataforma MediaViz.

A MediaViz é uma plataforma que recolhe e apresenta de forma interativa dados noticiosos. As principais vantagens desta plataforma passam por:

- . Acesso a dados de artigos de diversas fontes nacionais e internacionais;
- . Exploração visual e interativa dos mesmos;
- . Facilidade de análise, comparação e compreensão dos dados.

Nesta plataforma, o utilizador será capaz de efetuar pesquisas através de palavras-chave e obter uma coleção de dados relacionada com a pesquisa. Por exemplo, será possível obter dados como: o uso de palavras-chave ao longo do tempo, as entidades mais mencionadas, os padrões de publicação de fontes noticiosas, entre outros.

As respostas são confidenciais e serão apenas usadas para os efeitos já descritos. Qualquer dúvida/sugestão sobre o questionário ou sobre o trabalho que está a ser desenvolvido poderão contactar para o e-mail up201607847@fe.up.pt

Obrigada pela participação!

Jéssica Silva

***Obrigatório**

1. 1. Idade? *

Selecione apenas uma opção.
Marcar apenas uma oval.

- 15 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- +65

2. 2. Sexo? *

Selecione apenas uma opção.
Marcar apenas uma oval.

- Masculino
- Feminino
- Outro
- Não respondo

3. 3. Profissão? *

Marcar apenas uma oval.

- Jornalista
- Investigador
- Estudante Universitário (Doutoramento)
- Estudante Universitário (Mestrado)
- Outra: _____

4. 4. Instituição? *

5. 5. Área científica de formação? *

Marcar apenas uma oval.

- Ciências da Comunicação
- Ciência da Informação
- Informática
- Multimédia
- Outra: _____

I. Níveis de análise - granularidade

Na seguinte questão valorize o nível de análise consoante a escala apresentada.

Os níveis de análise dividem-se em quatro, sendo o seu agrupamento denominado de "coleção":

- Fonte: as fontes de origem dos documentos.
- Documento: o conjunto dos documentos integrais.
- Frase: o conjunto de todas as frases existentes nos dados.
- Termo (incluindo multi-palavra): o conjunto de todos os termos existentes nos dados.

Por exemplo:

Imagine que está a fazer uma pesquisa na plataforma MediaViz e utiliza a expressão de pesquisa "Marcelo Rebelo de Sousa". Dentro da coleção de dados disponíveis sobre essa palavra-chave, qual a análise que, para si, terá mais importância?

- Fonte: análise ao nível das fontes que publicaram sobre esse tema;
- Documento: análise relativamente aos documentos que incluem a palavra-chave;
- Frases: análise às frases que incluem o termo pesquisado;
- Termo: análise ao conjunto de termos presentes na expressão pesquisada.

6. 1. Na análise de uma coleção de documentos noticiosos, qual a importância que atribui a cada um dos seguintes níveis de análise (granularidade da análise) para o estudo e exploração dessa coleção? *

Marcar apenas uma oval por linha.

	Nada importante	Pouco importante	Moderadamente importante	Importante	Muito importante
Fonte	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Documento	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Frase	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Termo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. 2. Se tiver alguma questão/sugestão poderá indicar aqui:

II. Dimensões de análise

Na seguinte questão valorize a dimensão de análise consoante a escala apresentada.

As dimensões de análise são:

- Frequência: estatísticas descritivas (total, média, etc) sobre os dados (excluindo dimensão temporal).
- Temporal: análise das propriedades ao longo do tempo ou de ciclos temporais.
- Geográfica: análise das propriedades em função da geografia (locais mencionados).
- Entidades: análise das entidades mencionadas no conteúdo (pessoas, locais, organizações).
- Tópicos: análise dos temas mencionados no conteúdo.

Por exemplo:

Tomando como exemplo a expressão de pesquisa anterior - "Marcelo Rebelo de Sousa" - imagine que esta a fazer uma pesquisa na plataforma e indique que tipo de dimensão de dados é mais relevante, na sua opinião.

- Frequência: total de fontes, documentos ou frases com esta expressão de pesquisa;
- Temporal: evolução da expressão de pesquisa ao longo do tempo nas fontes, documentos ou frases;
- Geográfica: locais mencionados em associação à expressão de pesquisa;
- Entidades: pessoas e organizações mencionadas em associação à expressão de pesquisa;
- Tópicos: temas associados à expressão de pesquisa (política, social, entretenimento, etc).

8. 1. Na análise de uma *coleção* de documentos noticiosos, qual a importância que atribui a cada dimensão de análise (perspetiva de análise) para o estudo e exploração dessa coleção? *

Marcar apenas uma oval por linha.

	Nada importante	Pouco importante	Moderadamente importante	Importante	Muito importante
Frequência	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Temporal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geográfica	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entidades	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tópicos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. 2. No contexto da dimensão de análise focada nas *entidades*, qual a importância que atribui a cada um dos tipos de entidades que é possível obter? *

Marcar apenas uma oval por linha.

	Nada importante	Pouco importante	Moderadamente importante	Importante	Muito importante
Pessoas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Locais	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Organizações	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. **3. Para a análise ao nível das *fontes*, qual a importância que atribui a cada uma das seguintes dimensões de análise? ***

Marcar apenas uma oval por linha.

	Nada importante	Pouco importante	Moderadamente importante	Importante	Muito importante
Frequência: total de fontes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Temporal: volume de publicação ao longo do tempo e por ciclo (diário, semanal, etc).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geográfica: a origem das fontes, os locais mencionados por fonte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entidades: as entidades mencionadas por fonte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tópicos: os tópicos mencionados por fonte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11. **4. Para a análise ao nível dos *documentos*, qual a importância que atribui a cada uma das seguintes dimensões de análise? ***

Marcar apenas uma oval por linha.

	Nada importante	Pouco importante	Moderadamente importante	Importante	Muito importante
Frequência: total de documentos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Temporal: distribuição dos documentos ao longo do tempo e por ciclo (diário, semanal, etc).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geográfica: a origem de cada documento, os locais mencionados por documento.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entidades: as entidades mencionadas por documento.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tópicos: os tópicos mencionados por documento.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

12. 5. Para a análise ao nível das *frases*, qual a importância que atribui a cada uma das seguintes dimensões de análise? *

Marcar apenas uma oval por linha.

	Nada importante	Pouco importante	Moderadamente importante	Importante	Muito importante
Frequência: total de frases.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Temporal: distribuição das frases ao longo do tempo e por ciclo (diário, semanal, etc).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geográfica: os locais mencionados por frase.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entidades: as entidades mencionadas por frase.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tópicos: os tópicos mencionados por frase.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

13. 6. Para a análise ao nível dos *termos*, qual a importância que atribui a cada uma das seguintes dimensões de análise? *

Marcar apenas uma oval por linha.

	Nada importante	Pouco importante	Moderadamente importante	Importante	Muito importante
Frequência: total de termos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Temporal: distribuição dos termos ao longo do tempo e por ciclo (diário, semanal, etc).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geográfica: os locais referidos nos termos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entidades: as entidades referidas nos termos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tópicos: os tópicos associados aos termos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

14. 7. Se tiver alguma questão/sugestão poderá indicar aqui:

Apêndice D

E-mail enviado ao público-alvo

- **Investigadores:**

Boa tarde.

Sou aluna do Mestrado em Multimédia da Universidade do Porto e, no âmbito do projeto de dissertação intitulado “Design e Prototipagem de um Observatório Media Online”, venho, por este meio, pedir a sua colaboração no preenchimento e disseminação de um questionário. O questionário visa informar a prototipagem da plataforma MediaViz. Esta plataforma recolhe dados de diversas fontes online nacionais e internacionais e apresenta-os de uma forma dinâmica e interativa, através de visualizações.

Algumas das principais características desta plataforma são:

- . Acesso a dados de artigos de diversas fontes nacionais e internacionais;
- . Exploração visual e interativa dos mesmos;
- . Facilidade de análise, comparação e compreensão dos dados.

Por exemplo, será possível, através de uma pesquisa na plataforma, visualizar, de forma interativa, dados relativos a artigos de diversas fontes como: a evolução do termo pesquisado, os termos mais utilizados, as fontes que publicaram mais sobre determinado assunto, entre outros.

As respostas ao questionário permitirão perceber quais os níveis e dimensões

de análise mais relevantes para investigadores, que fazem parte do público-alvo da plataforma. Assim, e caso seja possível, pedia-lhe para, por favor, reencaminhar este email para todos os elementos do grupo de investigação. O preenchimento do questionário tem uma duração de 10/15 minutos.

Para além disso, após a prototipagem, irão ser realizados testes de utilizadores para possíveis correções. Desta forma, peço que se tiver disponibilidade para participar num teste de utilizador, presencialmente ou remotamente, que me indique como resposta a este email.

Link para o questionário: <https://goo.gl/forms/YGak7A9vZI0gALc23>

Agradeço a atenção e a colaboração,

Jéssica Pereira da Silva

Mestrado em Multimédia, Universidade do Porto

● **Jornalistas:**

Boa tarde.

Sou aluna do Mestrado em Multimédia da Universidade do Porto e, no âmbito do projeto de dissertação intitulado “Design e Prototipagem de um Observatório Media Online”, venho, por este meio, pedir a sua colaboração no preenchimento e disseminação de um questionário.

O questionário visa informar a prototipagem da plataforma MediaViz. Esta plataforma recolhe dados de diversas fontes online nacionais e internacionais e apresenta-os de uma forma dinâmica e interativa, através de visualizações.

Algumas das principais características desta plataforma são:

- . Acesso a dados de artigos de diversas fontes nacionais e internacionais;
- . Exploração visual e interativa dos mesmos;
- . Facilidade de análise, comparação e compreensão dos dados.

Por exemplo, será possível, através de uma pesquisa na plataforma, visualizar, de forma interativa, dados relativos a artigos de diversas fontes como:

a evolução do termo pesquisado, os termos mais utilizados, as fontes que publicaram mais sobre determinado assunto, entre outros.

As respostas ao questionário permitirão perceber quais os níveis e dimensões de análise mais relevantes para jornalistas, que fazem parte do público-alvo da plataforma. Assim, e caso seja possível, pedia-lhe para, por favor, reen-caminhar este email para todos os elementos da redação.

O preenchimento do questionário tem uma duração de 10/15 minutos.

Para além disso, após a prototipagem, irão ser realizados testes de utilizadores para possíveis correções. Desta forma, peço que se tiver disponibilidade para participar num teste de utilizador, presencialmente ou remotamente, que me indique como resposta a este email.

Link para o questionário: <https://goo.gl/forms/YGak7A9vZI0gALc23>

Agradeço a atenção e a colaboração,

Jéssica Pereira da Silva

Mestrado em Multimédia, Universidade do Porto

- **Estudantes:**

Bom dia.

Sou aluna do Mestrado em Multimédia da Universidade do Porto e, no âmbito do projeto de dissertação intitulado “Design e Prototipagem de um Observatório Media Online”, venho, por este meio, pedir a sua colaboração no preenchimento e disseminação de um questionário.

O questionário visa informar a prototipagem da plataforma MediaViz. Esta plataforma recolhe dados de diversas fontes online nacionais e internacionais e apresenta-os de uma forma dinâmica e interativa, através de visualizações.

Algumas das principais características desta plataforma são:

- . Acesso a dados de artigos de diversas fontes nacionais e internacionais;
- . Exploração visual e interativa dos mesmos;

. Facilidade de análise, comparação e compreensão dos dados.

Por exemplo, será possível, através de uma pesquisa na plataforma, visualizar, de forma interativa, dados relativos a artigos de diversas fontes como: a evolução do termo pesquisado, os termos mais utilizados, as fontes que publicaram mais sobre determinado assunto, entre outros.

As respostas ao questionário permitirão perceber quais os níveis e dimensões de análise mais relevantes para os utilizadores da plataforma.

O preenchimento do questionário tem uma duração de 10/15 minutos.

Para além disso, após a prototipagem, irão ser realizados testes de utilizadores para possíveis correções. Desta forma, peço que se tiver disponibilidade para participar num teste de utilizador, presencialmente ou remotamente, que me indique como resposta a este email.

Link para o questionário: <https://goo.gl/forms/YGak7A9vZI0gALc23>

Agradeço a atenção e a colaboração,

Jéssica Pereira da Silva

Mestrado em Multimédia, Universidade do Porto

Apêndice E

Lista de contactos

E.1 Unidades de investigação

CIC.Digital FLUP

CIC.Digital Lisboa

ICNOVA - Instituto de Comunicação da NOVA

CITEI - Centro de Investigação e, Tecnologias e Estudos Intermédia

CECS - Centro de Estudos de Comunicação e Sociedade

ICS - Instituto de Ciências Sociais da Universidade do Minho

Departamento de Ciências da Comunicação da Universidade do Minho

LabCom.IFP - Unidade de Investigação Científica da Faculdade de Artes e Letras

da Universidade da Beira Interior

ECSE - Escola Superior de Comunicação Social

CECC - Centro de Estudos de Comunicação e Cultura

E.2 Redações de Órgãos de Comunicação Social

Público

Jornal de Notícias

Diário de Notícias

Expresso

A Bola

Record

Jornal Económico

Jornal de Negócios

Correio do Minho

Diário do Minho

E-Konomista

Rádio Renascença

RTP

SIC

TVI

Observador

ECO.PT

Dinheiro Vivo