

Prémio Arquivo.pt

Descrição Sumária do Trabalho (documento-modelo)

Instruções gerais:

- Apresente de forma concisa os contributos do trabalho de acordo com os critérios de avaliação.
- Responda diretamente às questões colocadas.
- Use linguagem clara e sucinta.
- Pode incluir imagens e respetivas legendas.
- O tamanho máximo do ficheiro é de 10 MB.
- O limite de palavras é de 2 000.
- Submeta a versão final num dos seguintes formatos: DOCX ou PDF.

Identificação

- Título: Lupa Digital
- Área temática: Ciência de Dados, Metajornalismo, Literacia Mediática
- Candidato: Hugo Veríssimo
- Email: hugoverissimo21@gmail.com

Descrição do Trabalho

Descreva sucintamente o trabalho realizado.

A Lupa Digital [link](#) é um projeto desenvolvido no âmbito da competição Prémios Arquivo.pt 2025, focado em metajornalismo inteligente, com o objetivo de combater a parcialidade e tornar a informação acessível e transparente a qualquer pessoa.

Este projeto propõe transformar grandes volumes de informação jornalística arquivada em conhecimento útil e acessível. Com base nos milhões de páginas do Arquivo.pt e recorrendo a ferramentas inovadoras, como modelos de processamento de linguagem natural (NLP), a Lupa Digital torna possível a análise e exploração de entidades, eventos e temas ao longo do tempo, contribuindo para o reforço da literacia mediática.

Com notícias arquivadas desde 1998 até à atualidade, de 20 fontes de informação nacionais, a Lupa Digital, para além do desenvolvimento de um conjunto de dados com centenas de milhares de notícias e metadados referentes à sua fonte, data, perceção do sentimento e tópicos mencionados, consiste também numa aplicação web que permite ao utilizador explorar qualquer tópico à sua escolha.



Figura 1 - Captura de ecrã da página principal da Lupa Digital.

Esta aplicação web, após a pesquisa por parte do utilizador de um tópico do seu interesse, gera várias visualizações interativas sobre as aparições do mesmo nos media digitais portugueses, respondendo a perguntas tais como:

- Que fontes têm noticiado mais esse determinado tópico?
- Como têm evoluído as suas menções ao longo dos anos?
- Que outros tópicos se relacionam com este e como têm variado ao longo dos anos?
- Quais são as principais relações deste tópico e qual é a sua perceção?



Figura 2 - a) Captura de ecrã da página de resultados da Lupa Digital. b) Captura de ecrã da página Grafo da Lupa Digital.

Objetivos

Quais os principais objetivos do trabalho?

Os principais objetivos da Lupa Digital incluem: **METER SCREENSHOT PARA CADA BULLET POINT**

- Análise da abordagem dos media em relação a qualquer tópico

Permitir que qualquer utilizador, de forma simples e rápida, consiga pesquisar um tópico à sua escolha e, a partir dos resultados dessa pesquisa, consiga ter uma perceção em relação a como os media portugueses tendem a abordar esse tema e como é que esta abordagem tem vindo a mudar ao longo dos anos.

- Identificação e análise de relações entre tópicos

Revelar ligações entre qualquer par de tópicos, sejam eles entidades, eventos ou temas, e, por conseguinte, permitir ao utilizador ter uma perceção sobre a relação, no que toca a verificar quais as fontes de informação mais envolvidas na ponte entre os tópicos, o sentimento transmitido no contexto das notícias onde a relação está presente e como tem variado a ligação ao longo do tempo.

- Promoção de transparência e literacia mediática

Incentivar a análise crítica da informação, ajudando os utilizadores a compreender a evolução dos discursos e a influência dos media na opinião pública, de modo que os mesmos possam ter uma perceção imparcial através dos resultados apresentados pela Lupa Digital.

- Criação de uma ferramenta interativa e intuitiva

Desenvolver uma plataforma inovadora e acessível a qualquer pessoa, com uma interface simples e intuitiva, onde os utilizadores possam explorar dados relacionados com os media portugueses de forma visual e interativa.

Resultados Atingidos

Quais os principais resultados do trabalho?

Através da criação da Lupa Digital foram obtidos dois resultados principais: um conjunto de dados, disponível em acesso aberto, e uma aplicação web.

Atendendo ao conjunto de dados **link**, como necessidade para o desenvolvimento do trabalho, este foi criado a partir do tratamento de dados extraídos do Arquivo.pt. É um conjunto de dados com mais de três centenas de milhares de notícias, desde 1998 até à atualidade, de 20 fontes de informação portuguesas, nomeadamente AEIOU, Correio da Manhã, CNN Portugal, Diário de Notícias, Dinheiro Vivo, Expresso, IOL, Jornal de Notícias, Lusa, Jornal de Negócios, NiT,

Notícias ao Minuto, O Mirante, Observador, Público, Record, RTP, Sapo, SIC Notícias, e TSF, contendo metadados sobre cada uma delas, tais como a perceção do sentimento transmitido pela notícia e os tópicos mencionados na mesma. Este conjunto de dados para além de permitir futuros estudos, também permite a sua expansão futura com notícias e artigos mais recentes e com a adição de novas fontes de informação.

Quanto à aplicação web [link](#), sendo esta o foco principal da Lupa Digital, envolveu a criação de website que funciona como um motor de busca para qualquer tópico que um utilizador tenha interesse. Após a pesquisa por um tópico, o utilizador terá acesso a várias visualizações **CONCRETIZAR VISUALIZOES** e informações relacionadas com a pesquisa, podendo, de forma intuitiva, analisar criticamente as relações com outros tópicos, como é que a perceção do tópico pesquisado tem vindo a mudar e as divergências entre as menções de diferentes fontes de informação no que toca ao seu tópico.

Originalidade e carácter inovador

Quais os contributos inovadores que demonstram a originalidade do trabalho?

A Lupa Digital distingue-se pela sua abordagem inovadora no que toca ao metajornalismo, pelo facto de usar técnicas de automatização para o tratamento de texto e processamento de linguagem natural, em larga escala e de forma eficiente, permitindo uma exploração de uma maior quantidade de arquivos digitais e, desta forma, uma experiência interativa e dinâmica na análise de milhares de notícias preservadas no Arquivo.pt.

Para além disso, ao invés de uma simples pesquisa textual, o projeto proporciona várias visualizações gráficas sobre os tópicos pesquisados e as relações com os demais, permitindo que os utilizadores identifiquem conexões e padrões de forma intuitiva e transparente. Estas visualizações gráficas incluem um grafo interativo que revela como diferentes tópicos se interligam ao tópico pesquisado, e ao interagir com cada tópico permite que o utilizador obtenha variada informação **CONCRETIZAR?** sobre a relação em causa.

Este projeto não é apenas um motor de busca para notícias, mas também uma ferramenta de investigação e análise que ajuda a compreender como os media moldaram a perceção pública ao longo do tempo. A possibilidade de acompanhar a evolução de narrativas e identificar padrões entre fontes de notícias ou entre relações de tópicos tornam este projeto um recurso único para a literacia mediática.

PARA ALEM DISSO E A PUBLICACAO DO CONJUNTO DE DADOS. NUM FORMATO DE ACESSO ABERTO. CONTITUI UM DOS MAIORES CONJUNTO DE DADOS PUBLICO NO AMBITO DA MINERACAO DE DADOS DE NOTICIAS PORTUGUES (?)

CONTRIBUIÇÃO SIGNIFICATIVA PARA O CAMPO DO JORNALISMO DIGITAL E DA EDUCACAO MEDIATICA

Impacto social (aplicação e utilidade social)

Descreva como o seu trabalho poderá ser útil para os cidadãos?

O projeto Lupa Digital tem uma grande utilidade para os cidadãos, pelo facto de disponibilizar uma ferramenta inovadora que permite aos mesmos explorar e compreender a evolução de um tópico do seu interesse no contexto da informação jornalística com o passar dos anos, de forma rápida e intuitiva para o público em geral, INDEPENDENTEMENTE DOS SEUS CONHECIMENTOS TECNICOS.

Ademais, num contexto em que o acesso à informação e a literacia mediática são essenciais **DEVIDO A EXPOSIÇÃO NO DIA-A-DIA**, a Lupa Digital incentiva a compreensão mais profunda da informação e a análise crítica das fontes noticiosas, através da democratização do acesso a dados históricos a partir de uma simples pesquisa.

Atendendo à funcionalidade do grafo interativo, este ajuda a visualizar as conexões entre temas, entidades e eventos, permitindo aos utilizadores explorar padrões e relações que, de outra forma, poderiam passar despercebidos.

Por fim, a acessibilidade facilitada a uma grande quantidade de informação, para além de contribuir para o combate à desinformação, beneficia desde estudantes até cidadãos comuns interessados em várias áreas do saber (economia, história, política, ...), promovendo um maior envolvimento com a informação digital preservada.

Impacto científico (aplicação e utilidade científica)

Descreva como o seu trabalho poderá ser útil para investigadores?

Este projeto torna-se extensamente útil para investigadores e para a comunidade científica em geral, pelo facto de ter desenvolvido, para além de uma ferramenta inovadora para a análise de grandes volumes de notícias arquivadas, um conjunto de dados de centenas de milhares de notícias e informações associadas às mesmas, tal como a perceção do sentimento e os tópicos mencionados.

O desenvolvimento deste conjunto de dados deixa na mesa inúmeras possibilidades para estudos futuros, tais como estudos sobre temas específicos de forma mais aprofundada, sobre as diferentes fontes de notícias e as suas tendências de comunicação sobre determinados temas, como é que eventos históricos alteraram as narrativas dos meios de comunicação, entre outros.

Ademais, para investigadores da área de Processamento de Linguagem Natural (NLP) e Inteligência Artificial, a Lupa Digital também oferece um caso de estudo valioso, uma vez que combina técnicas avançadas de extração de entidades e análise semântica para estruturar informação proveniente de diversas fontes e em larga escala. Estes métodos podem ser

aplicados e aprimorados em pesquisas relacionadas com desinformação, viés mediático e análise de narrativas.

Relevância da utilização do Arquivo.pt

Exponha a necessidade de recorrer ao Arquivo.pt para obter recursos únicos para o trabalho realizado. Apresente exemplos de conteúdos preservados que utilizou.

A partir do uso da CDX-server API do Arquivo.pt foram extraídos mais de três milhões de URLs associados a diversas fontes de informação. Posteriormente, para cada um destes URLs, o Arquivo.pt foi utilizado de modo a permitir a extração do conteúdo associado ao URL, de forma a criar o conjunto de dados que serve de input à aplicação web criada. SEM O ARQUIVO NAO SERIA POSSIVEL FAZER UMA ANALISE TAO ABRANGENTE E COMPARATIVA DA COBERTURA NOTICIOSA AO LONGO DAS ULTIMAS DECADAS. ARQUIVO É UNICO COM VOLUME DE INFORMACAO SUSFIENTE ARQUIVADA PUBLICAMENTE, PERMITINDO RECUPERAR INFORMACAO DESCONTINUADA.

Isto demonstra a importância do Arquivo.pt no desenvolvimento deste projeto, sendo como uma fundação para a criação do conjunto de dados criado e por sua vez para a aplicação web.

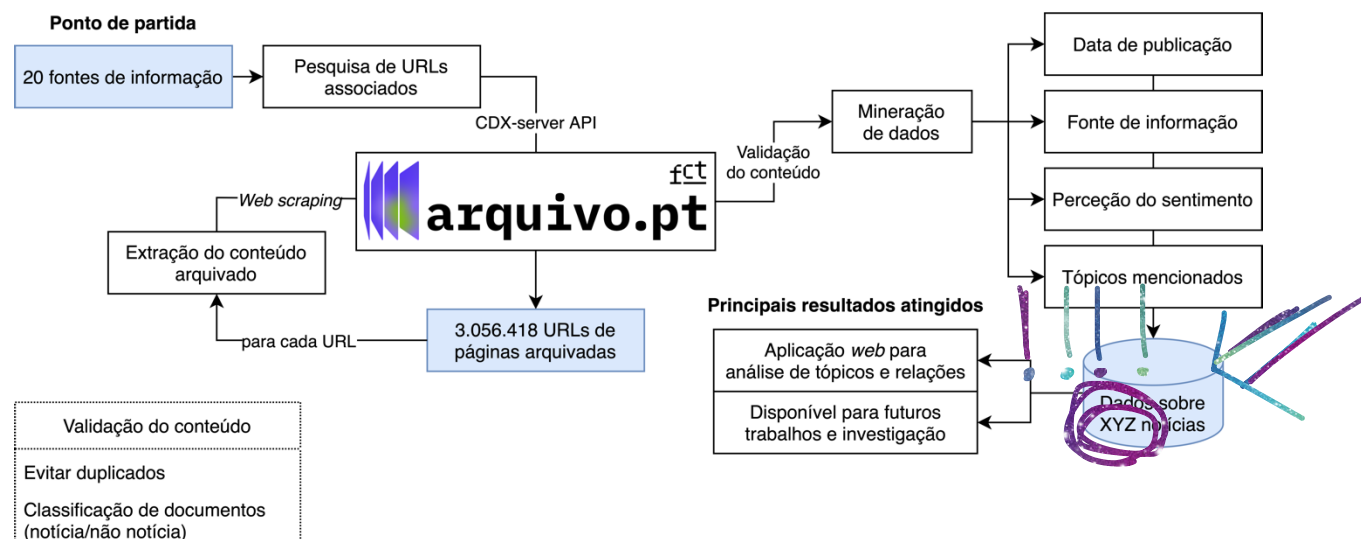


Figura 3 - Metodologia da extração de dados OU Esquema do fluxo de dados. ACERTAR QND DE NOTICIAS

Comentários adicionais

Inclua informação adicional que considere relevante para demonstrar o valor do trabalho.

não sei o que devo meter? PODES ACRESCENTAR QUE O TRABALHO FOI INICIADO EM CONTEXTO LETIVO, NO AMBITO DA DISCIPLINA DE FCD E POSTERIORMENTE

MELHORADO E APROFUNDADO. ISTO PARA DIZER QUE A EXPLORACAO DO AQUIVO PERMITIU O DESENVOLVIMENTO DE OUTRAS COMPETENCAS MAIS TECNICAS E QUE PODE SERVIR COMO MOTIVACAO PARA QUEM QUER DAR OS PRIMEIROS PASSOS NA AREA DE CIENCIA DE DADOS.

É TAMBÉM DE SALIENTAR O AGRADECIMENTO A EQUIPA TECNICA? DO ARQUIVO.PT PELA DOCUMENTACAO E ACESSO TOTAL E GRATUITO A API DO ARQUIVO SUJA DOCUMENTACAO SE ENCONTRA BEM ESTRUTURADA E INTUITIVA. E AO ARQUIVO PELA COLETA DE INUMEROS DADOS TORNANDO INUMEROS PROJETOS POSSIVEIS.

Recursos complementares

Liste referências disponíveis online que complementem o trabalho. Por exemplo, podem ser ligações para documentos, websites, imagens ou aplicações.

Apenas serão considerados recursos disponíveis em acesso aberto através da Internet. Utilize o seguinte formato para listar as referências:

- *Título, URL: descrição*
- *link para site*
- *github*
- *https://doi.org/10.5281/zenodo.15231163*
- *video*
- *https://natura.di.uminho.pt/download/sources/Dictionaries/wordlists/ - dicionario para fazer mapeamento das palavras para a sua escrita mais correta (lower case e afins)*
- *link para arquivo*
- *link cdx server api arquivo*
- *ESTILO:*
- *Site de exemplo com título, https://example.com*
descricao do site de exmplo tal tal
- *site 2*