

Prémios Arquivo.pt 2018

Memória Descritiva de trabalho

Identificação

- Título: Arquivo de Opinião
- Área temática: Ciências da Computação e Ciência Política
- Candidato: Miguel Won
- Email: miguel.won@inesc-id.pt

Originalidade e carácter inovador

O Arquivo de Opinião (www.arquivodeopinio.pt) é uma aplicação web que disponibiliza ao utilizador um repositório digital de artigos de opinião, publicados entre 2008 e 2016, nos principais agentes de media de Portugal. A aplicação permite ao utilizador a realização de buscas por palavra ou expressão, autor, período temporal ou jornal.

Não existe em Portugal ferramenta equivalente dedicada ao sector da opinião política. Neste sentido, esta aplicação destaca-se pela inovação de produzir uma ferramenta digital, que além de preservar uma memória histórica da opinião escrita, também disponibiliza a sua consulta a qualquer utilizador.

Em termos técnicos, o processo de construção desta aplicação exigiu a implementação de múltiplas técnicas *state-of-the-art* de Processamento de Língua Natural, desde processos de etiquetagem morfosintática a reconhecimento automático de entidades. Este processamento permitiu a identificação de palavras-chave relevantes e entidades mencionadas em cada artigo. A cada autor também foi associada uma página que contém a sua nuvem de palavras, bem como as entidades mais mencionadas nos seus artigos. Esta é outra componente inovadora desta aplicação, pois permite ao utilizador informar-se sobre tendências, temáticas, pessoas, locais ou organizações que obtiveram mais atenção por parte do autor.

Ao mesmo tempo também foram indexadas mais de 30 mil palavras-chave, e disponibilizada uma ferramenta que permite a consulta da frequência de ocorrência, por período temporal e por agente de media. Por último, também foram implementados modelos de língua com o objetivo de calcular os chamados “word embeddings”, que permitem a consulta de palavras-chave relacionadas.

Em resumo, o Arquivo de Opinião é uma ferramenta digital original, que reúne informação previamente dispersa e desorganizada, e por isso de difícil acesso, mas também informação adicional como a identificação de entidades mencionadas nos textos.

Impacto social (aplicação e utilidade social)

A opinião escrita é reflexo e parte essencial da opinião pública, e também um corpo de influência da mesma. A sua presença é uma componente essencial nos media noticiosos e frequentemente marca o tom e a agenda da discussão pública.

Tal como o acesso público à informação é peça fundamental das chamadas sociedades liberais, também o acesso e acompanhamento da discussão política, em particular a opinião escrita, é essencial no debate político de qualquer democracia.

A existência de um repositório digital da opinião escrita permite não só o seu acompanhamento em retrospectiva, bem como a preservação da memória histórica dos debates e lógicas argumentativas em volta dos múltiplos confrontos políticos. A consulta dos assuntos discutidos neste campo da opinião, dos autores relevantes ou mesmo os termos mais usados, é um excelente contributo para a memória política.

A própria possibilidade de consulta de contraditório ao longo do tempo, permite também uma melhor agilização do escrutínio necessário a qualquer debate político. Este escrutínio e a preservação da memória escrita torna-se ainda mais essencial no contexto do problema atual da desinformação existente no mundo digital. Acresce ainda o facto de frequentemente o político ativo de hoje contribuiu, ou contribui, para o debate através da sua coluna de opinião. Neste sentido, este repositório online torna-se assim também numa ferramenta para mais e melhor transparência do debate político e dos seus atores envolventes.

Impacto científico (aplicação e utilidade científica)

A aplicação web aqui apresentada fornece aos investigadores das áreas da Ciência Política, Comunicação ou Historiadores, um portal online em que podem realizar consultas a esta componente da discussão pública, i.e., a coluna de opinião dos media noticiosos. Não só lhes é permitido consultar por período temporal, mas também filtrar por termos ou expressões, jornais ou autor.

Trabalhos científicos poderão ser realizados com base nos indicadores mostrados pela aplicação web, mas também a partir do uso da base de dados recolhida. O facto de ter sido constituído uma base de dados limpa e ordenada, abre a possibilidade de desenho de novos estudos que necessitem de análise em grande escala dos textos recolhidos. Este tipo de estudos insere-se dentro do recente paradigma *text as data*, que tem vindo a ganhar força nos últimos anos. Dentro deste contexto, esta nova base de dados também permite desenvolvimento de técnicas de computação na área de Processamento de Língua Natural, aplicadas ao contexto político, como

por exemplo, desenvolvimento de deteção de padrões argumentativos, análise de sentimento ou mesmo determinação automática de posicionamento no quadro político.

Relevância da utilização do Arquivo.pt

A recolha dos dados dos artigos de opinião, i.e., autor, título, texto principal, data de publicação, etc. foi realizada através da programação de processos automáticos de recolha de informação online. A fonte de recolha de informação depende da sua disponibilidade nas respetivas páginas locais (as páginas nos diferentes agentes noticiosos), em particular o quanto cada página disponibiliza a nível temporal. Assim, ao passo que, por exemplo, a página online do jornal Público disponibiliza informação sobre os seus artigos desde o período inicial considerado (2008), o mesmo já não acontece com jornais como o Diário de Notícias ou Jornal i. Por essa razão foi usado o arquivo histórico do portal arquivo.pt, através das suas APIs, para encontrar e descarregar o código fonte que serviu de publicação para maioria dos artigos armazenados nesta aplicação web. Daqui resultou que dos 85530 artigos recolhidos, 38703 artigos foram obtidos recorrendo à API do arquivo.pt

Comentários adicionais

Esta aplicação web encontra-se no início do seu desenvolvimento e está a ser planeada uma fase seguinte que terá como objetivo a incorporação de mais agentes de media, dados mais recentes e dados das redes sociais, como métricas de partilhas e reações aos artigos, em conjunto com informação das páginas públicas dos autores. Este tipo de indicadores permitirá um acompanhamento temporal e mais presente da discussão pública.

Recursos complementares

Endereço eletrónico do Arquivo de Opinião:

- www.arquivodeopinião.pt