

Proposta de Projeto de Doutoramento a Desenvolver no Âmbito do 1º Concurso para Atribuição de Bolsas de Investigação na Área de Engenharia Informática

1. Título do projeto

Título: Pesquisa e Recomendação Computacional de Conteúdo Noticioso

Palavras-chave: Mídia Social, Interesses dos Utilizadores, Recuperação de Informação

Referência: CEE_EI_FEUP1

2. Instituições envolvidas

Instituição onde o doutoramento será realizado: Faculdade de Engenharia da Universidade do Porto (FEUP)

Outras instituições participantes no projeto de investigação: Faculdade de Ciências Sociais (FCS) da Universidade de Timor-Leste (UNTL) e Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia (FECT) da UNTL.

3. Equipa de Orientação

Orientador: Sérgio Nunes (sergio.nunes@fe.up.pt)

4. Descrição do Projeto

O acesso a informação noticiosa mudou radicalmente com a adoção global da internet como plataforma de comunicação nas sociedades modernas. O ecossistema noticioso passou de um paradigma em que o acesso ao espaço público era controlado pelas organizações de mídia (por exemplo, jornais, televisões, estações de rádio), para um novo paradigma em que o espaço público foi transferido para o espaço online, onde o acesso é intermediado pelas plataformas de mídia social online (por exemplo, Facebook, Twitter, Google).

O papel dos anunciantes também mudou drasticamente. No passado, o anúncio era projetado e direcionado para segmentos de audiências (ou seja, perfis) e, mais importante, era direcionado ao conteúdo - por exemplo, uma secção sobre viagens mostraria anúncios relacionados com este tópico. Atualmente, a publicidade é direcionada ao indivíduo e é orientada pelo comportamento - por exemplo, os anúncios exibidos a uma pessoa dependem das interações dessa pessoa ao longo do tempo.

Nesse contexto, os modelos de negócios das plataformas de mídia social estão ancorados na maximização do tempo que os utilizadores gastam nessas plataformas. A chamada "economia da atenção". Assim, os algoritmos atuais de sugestão de conteúdo são orientados principalmente por esse critério - procuram maximizar o tempo dos utilizadores nestas plataformas. É nossa convicção que isso não está alinhado com os interesses dos utilizadores finais e que novos paradigmas de interação e recomendação precisam ser imaginados, estudados e propostos.

Com esta proposta, pretendemos abordar essa oportunidade e investigar como diferentes estratégias de acesso e recomendação de conteúdo noticioso podem ser projetadas para maximizar os interesses dos utilizadores e melhor alinhar-se com seus valores. Esta proposta enquadra-se na área de Recuperação de Informação, e tem ligação com as áreas de Jornalismo Computacional e Sistemas de Recomendação. O objetivo principal é contribuir para o

desenvolvimento de sistemas de pesquisa e de recomendação de conteúdo noticioso que incorporem simultaneamente características relacionadas com o conteúdo das notícias e com o comportamento dos utilizadores.

O trabalho terá acesso a dados do projeto <https://www.timornews.tl/>, um serviço noticioso Timorense disponível através da web e com cerca de 4 mil visitas diárias. O acesso a estes dados permitirá, por um lado estudar o comportamento dos utilizadores no acesso à informação e, por outro lado, estudar uma coleção de conteúdo noticioso. A colaboração com este projeto permitirá ainda construir e avaliar protótipos num ambiente real.

O projeto envolverá quatro fases principais, (1) a caracterização do comportamento de acesso à informação com recurso a estudo de registos de interação (logs) e com a realização de entrevistas; (2) a obtenção e caracterização da coleção de notícias; (3) desenvolvimento de algoritmos de pesquisa e recomendação de conteúdo usando uma coleção de referência para avaliação do desempenho; (4) implementação de protótipos para avaliação dos algoritmos e dos mecanismos de interação com o utilizador, com recurso a testes em ambiente real (testes A/B).

Espera-se que o trabalho aqui desenvolvido resulte em inovações ao nível dos algoritmos de ordenação e recomendação de conteúdos, incorporando simultaneamente características obtidas a partir dos conteúdos (p.e. texto das notícias), e características obtidas a partir do comportamento dos utilizadores (p.e. notícias consultadas antes). Esperam-se também resultados inovadores ao nível das métricas de otimização, privilegiando os interesses dos utilizadores em detrimento dos interesses das plataformas. Finalmente, esperam-se ainda inovações ao nível dos mecanismos de interação com os utilizadores, em particular dando possibilidade ao utilizador de controlar de forma transparente o algoritmo de ordenação ou recomendação.

5. Referências Bibliográficas

Ethan Zuckerman. *The Case for Digital Public Infrastructure*. Knight First Amendment Institute at Columbia University (2020). <https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure>

Fernando Diaz. *Integration of news content into web results*. In Proceedings of the Second ACM International Conference on Web Search and Data Mining, pp. 182-191. 2009. <https://dl.acm.org/doi/abs/10.1145/1498759.1498825>

Bernardo A. Huberman. *Social computing and the attention economy*. Journal of Statistical Physics 151, no. 1-2 (2013): 329-339. <https://link.springer.com/article/10.1007/s10955-012-0596-5>

Sérgio Nunes, and José Luís Devezas. *Social Media and Information Consumption Diversity*. News Information Retrieval (NewsIR) Workshop at European Conference on Information Retrieval (ECIR) (2018). <http://ceur-ws.org/Vol-2079/paper5.pdf>