

Mineração Textual como ferramenta para identificar o entendimento sobre Direito Autoral dos usuários do Twitter

Guilherme Pinto¹ (aluno autor); Ariadne Chloe Furnival¹ (orientador)
1-Universidade Federal de São Carlos

Objetivo

Compreender como os usuários do Twitter utilizam a plataforma para obter e compartilhar informações relacionadas ao Direito Autoral, a partir do conteúdo das postagens e respostas disponíveis no site.

Introdução

Esta pesquisa tem foco na intersecção entre dois campos temáticos da Ciência da Informação, a saber: o Comportamento Informacional e o uso de Redes Sociais como meios para busca de informações relevantes ao usuário inserido neste contexto digital. Há diversas pesquisas que visam compreender quais os objetivos dos usuários do Twitter quando usam a plataforma, que passam por educação, pesquisas científicas, discussões políticas, de saúde, culturais, sociais, entre outras¹.

Para tanto levaremos em consideração pesquisas anteriores que versem sobre a busca de informações dentro do Twitter², levando em consideração os aspectos éticos da utilização destes de dados abertos³.

Materiais e Métodos

Para obter os dados dentro do Twitter, usamos o API da plataforma. Nele é possível inserir *queries* a fim de recuperar postagens feitas de acordo com os parâmetros estabelecidos⁴. Todos os parâmetros de buscas avançadas estão disponíveis, como: palavras-chave, hashtags, idioma, localização e data.

Segundo Blake⁵, o processo de Mineração Textual pode ser entendido como a identificação de padrões originais, interessantes e compreensíveis dentro de uma coleção de textos. Para descobrir tais padrões utilizamos o software R, escolhido por ser um software que contém recursos para análise de Mineração Textual e por ser gratuito.

Os dados foram coletados durante o mês de Agosto de 2017. Utilizamos filtros para captar apenas resultados em língua portuguesa e que apresentassem uma das palavras-chave escolhidas, como vemos na Tabela 1.

Tabela 1: Quantidade de resultados por palavra-chave

Palavra-chave	Nº de resultados
Acesso Aberto	113
Copyleft	12
Copyright	1533
Creative Commons	78
Direito Autoral	498
Direitos Autorais	5700
Domínio Público	433
Total	8367

Fonte: Produzido pelos autores

Resultados preliminares

Entendemos cada *tweet* capturado como um documento único e o conjunto de todos os documentos o corpus deste trabalho.

A partir da coleção de *tweets* coletados é possível analisar diversos aspectos dos documentos. Como exemplo, criamos uma nuvem de palavras (Figura 1) que mostra as palavras mais frequentes de todos os documentos.

Figura 1: Nuvem de palavras



Fonte: Produzido pelos autores

Alguns pacotes de funções do R que serão utilizados neste trabalho são: “tm”, “tidytext”, “fpc”, “wordcloud”, “topicmodels” e “cluster”, que auxiliarão no pré-processamento do texto, na análise de cluster, modelagem por tópicos e construção de gráficos. Estes pacotes, bem como outros que podem ser relevantes, são descritos nos livros *Text Mining with R*⁶ e *Text Mining in Practice with R*⁷ que utilizaremos como base em nossas análises.

Conclusão

Nesta pesquisa pretendemos investigar o comportamento dos usuários do Twitter no que diz respeito a buscar e compartilhar informações relacionadas ao Direito Autoral, mediante a captação e análise de *tweets* sobre o assunto.

A plataforma Twitter e seu conceito de *microblogging* fornecem uma oportunidade interessante para estudar tal comportamento e as tendências dos usuários dentro deste ambiente específico. A enorme quantidade de textos produzidos e disseminados na plataforma impossibilita uma leitura simples como método para determinar seus conteúdos. Desse modo, as ferramentas de Mineração Textual oferecem um método automatizado capaz de extrair estes conteúdos e facilitar a análise.

Referências

1. ELSWEILER, D.; HARVEY, M. Engaging and Maintaining a Sense of Being Informed: Understanding the Tasks Motivating Twitter Search. *Journal of the Association for Information Science and Technology*, v. 66, n. 2, p. 264–281, 2015.
2. JAVA, A.; FININ, T.; SONG, X.; TSENG, B. *Why We Twitter: Understanding Microblogging Usage and Communities*. Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007. Anais...San Jose, USA: Springer, 2007
3. ZIMMER, M. “But the data is already public”: on ethics of research in Facebook. *Ethics and Information Technology*, v. 12, p. 313-325, 2010.
4. ZIMMER, M.; PROFERES, N. J. A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, v. 66, n. 3, p. 250-264, 2014.
5. BLAKE, C. Text Mining. *Annual Review of Information Science and Technology*, v. 45, n. 1, p. 123–155, 2011.
6. SILGE, J.; ROBINSON, D. *Text Mining with R: A Tidy Approach*. [s.l.] O’Reilly Media, 2017.
7. KWARTLER, T. *Text Mining in Practice with R*. Hoboken, NJ: John Wiley & Sons, 2017.