

III FÓRUM DE PESQUISA DISCENTE 2021

PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO & ORGANIZAÇÃO DO CONHECIMENTO - PPGGOC/ECI/UFMG

Sessão: **Representação do Conhecimento**

O fluxo temporal de termos relevantes: uma análise em teses da UFMG de 2007 a 2018 nas ciências humanas¹

Luiz Antonio Lopes Mesquita 

Universidade Federal de Minas Gerais (UFMG) - Brasil

Célia da Consolação Dias 

Universidade Federal de Minas Gerais (UFMG) - Brasil

Renato Rocha Souza 

Universidade Federal de Minas Gerais (UFMG) - Brasil

Resumo: O objetivo geral desta pesquisa foi analisar se há uma variação temporal característica da distribuição de valores de termos relevantes ao longo do tempo da produção de textos que possa contribuir como um critério para o processo de sua indexação automática. Foram analisadas as teses de doutorado dos programas de pós-graduação (PPGs) da área de Ciências Humanas da UFMG, considerando-se 7 PPGs distintos, sendo cada um deles um corpus, com um total de 929 teses defendidas período de 12 anos, de 2007 a 2018. Os termos considerados foram todos os sintagmas nominais contidos nos próprios textos das teses. Cada sintagma nominal recebeu um valor associado à sua relevância como descritor de acordo com os critérios de frequência do termo na própria tese (*TF - Term Frequency*) e com o inverso da frequência de ocorrência do termo no total de teses de cada PPG (*IDF - Inverse Document Frequency*). As teses foram divididas em 12 grupos em cada PPG para o cálculo da data média de defesa das teses e da média de pontuação consolidada dos termos relevantes nas teses. Como resultados, identificou-se o comportamento característico de cada PPG através de um gráfico de dispersão do nível médio de pontuação de relevância ao longo do tempo. Para cada gráfico de cada um dos 7 PPGs foi adicionada uma linha de tendência, considerando seu respectivo R^2 , e feita sua análise específica. Todos os comportamentos de distribuição temporais foram caracterizados em equações e podem ser aplicados como critério para indexação automática.

Palavras-Chave: Recuperação da Informação Temporal. Indexação Automática. Sintagmas Nominais.

¹ Texto completo disponível em número especial da Revista Múltiplos Olhares em Ciência da Informação, link: <https://periodicos.ufmg.br/index.php/moci>

The temporal flow of relevant terms: an analysis in UFMG theses from 2007 to 2018 in human sciences²

Abstract: This research's general objective was to analyze if there is a temporal variation characteristic of the distribution of values of relevant terms over the time of the production of texts that can contribute as a criterion for the automatic indexing process. The doctoral theses of the graduate programs (PPGs) in Human Sciences at UFMG were analyzed, considering seven different PPGs, each of which is a corpus, with 929 theses defended in a period of twelve years, from 2007 to 2018. The terms considered were all the noun phrases contained in the texts of the theses. Each noun phrase received a value associated with its relevance as a descriptor according to the term frequency criteria in the thesis itself (TF - Term Frequency) and with the inverse of the frequency of occurrence of the term in the total of theses of each PPG (IDF - Inverse Document Frequency). The theses were divided into 12 groups in each PPG to calculate the average defense date of the theses and the average consolidated score of the relevant terms in the theses. As a result, each PPG's characteristic behavior was identified through a scatter plot of the average level of relevance score over time. For each graph of each of the 7 PPGs, a trend line was added, considering its respective R^2 , and its specific analysis was made. All temporal distribution behaviors were characterized in polynomial equations and applied as a criterion for automatic indexing.

Keywords: Temporal Information Retrieval. Automatic Indexing. Noun Phrase.

² The full text is available in a special issue of the Revista Múltiplos Olhares em Ciência da Informação, link: <https://periodicos.ufmg.br/index.php/moci>