

Identifying relevant keywords in Scientific Collaboration Networks

Thiago M. R. Dias¹, * Gray F. Moita¹

¹ CEFET-MG - Federal Center for Technological Education of Minas Gerais, Av. Amazonas, 7675, Nova Gameleira, 30510-000, Belo Horizonte, MG, Brazil, thiagomagela@gmail.com and gray@dppg.cefetmg.br

Key Words: *Extraction and data integration, Information Retrieval, Scientific Collaboration.*

1. Introduction

The scientific production and publication have increased considerably in recent years and may highlight the Internet as the primary factor distribution of this large volume of publications.

For Stroele et al. [1], scientific social networks are specific types of social networks that represent social interactions originating in the academic environment. These interactions usually occur through the publication of scientific articles, academic guidelines and the development of research projects. Various goals can lead to the study of scientific collaboration networks, such as recommendation of new collaborators, intensifying the collaboration, ranking of groups or individuals, or identifying groups and their characteristics.

Based on growing interest in understanding the scientific collaboration networks, the focus of this paper is to explore repositories containing data from scientific publications available on the web and generate networks of keywords these publications aiming to identify the most relevant keywords from a set words and highlight those that have the greatest impact on the network.

For this, techniques of social network analysis are adopted to be able to identify the keywords that have a higher number of citations, links and degree of centrality, and thus understand how these keywords are linked to others. Given this, it is possible to carry out recommendations of keywords to a particular researcher or research groups.

2. Development

For the extraction of data to be used in the construction of the keywords network, the architecture extraction and integration developed by Dias et al. [2] was used. This architecture extracts data from the Lattes Platform. The Lattes Platform was conceived to integrate the information systems of Brazilian federal agencies, optimising the Science and Technology (S&T) management process from the standpoint of both the user as well as promotion agencies and institutions of education and research

The extracted data is formatted and stored in XML files that contain all the scientific production of the users of the platform. These data are updated by the users and are freely accessible.

Subsequently, the identification algorithm proposed by Dias and Moita [3] is used to generate the keywords network. The network is generated by analyzing the keywords of each article of each researcher, generating a clique formed by the analyzed article linked words, and this is then inserted in the graph by the juxtaposition of keywords. Figure 1.

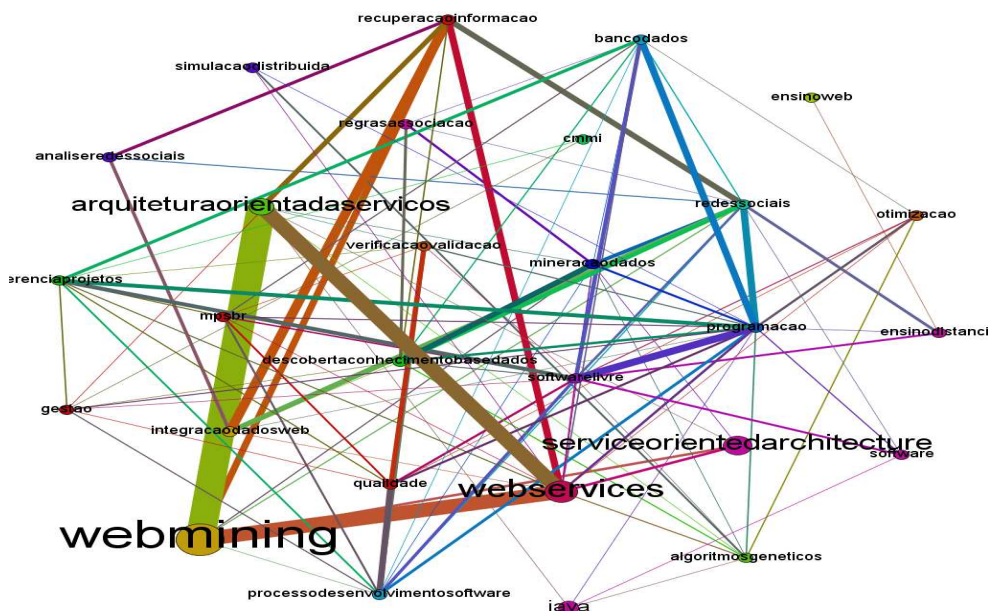


Figure 1. Example Network Keywords

The network shown corresponds to a group of researchers from Federal Center for Technological Education of Minas Gerais – Brazil. Thus, it is possible to identify with the analysis of keywords extracted from articles published by the group, which terms are most common in publications and how they are linked.

3. Results

As a result of the identification of keywords network process, it is possible to analyze which are the most relevant words within the network and how they are connected. Given this, it is possible to extract various information that can be used for the recommendation process keywords that are evolving or being poorly adopted. Therefore, various types of recommendation can be made.

REFERENCES

- [1] V. Stroele, G. Zimbrão and J. M. Souza, Análise de Redes Sociais Científicas: Modelagem Multi-relacional. In: I Brazilian Workshop on Social Network Analysis and Mining (BraSNAM), 2012, Curitiba, PR, Brazil.
- [2] T. M. R. Dias, G. F. Moita, T. Moreira, L. Santos and P. M. Dias, Modelagem e Caracterização de Redes Científicas: Um Estudo Sobre a Plataforma Lattes. In: II Brazilian Workshop on Social Network Analysis and Mining (BraSNAM), 2013, Maceio, AL, Brazil
- [3] T. M. R. Dias and G. F. Moita, Extração e Modelagem de Redes de Colaboração Científica. In: Conferência IADIS Ibero-Americana WWW/Internet, 2013, Porto Alegre, Brazil.