

Towards Linked Open Scientific Communication: Sharing, Analyzing, Exchanging

Sahar Vahdati

Informatik III, Universität Bonn, Germany

Vahdati@uni-bonn.de

Abstract

Many systems have emerged to help researchers seeking for relevant information about the research domain of their interest. Although with their support scientific documents are easily available by now, scientists need more and more advanced support while searching for different kinds of bibliographical information. This includes finding most important research trends, major papers, auspicious approaches, established conference series as well as the search for the most active groups for a specific research topic, similar approaches, partner candidates active in a certain domain, related research schools, etc. Answering such queries is still a hard and time-consuming task for researchers. We believe that having a comprehensive system with recommender services based on Linked Open Data can help to support researchers in this respect. In this paper, we address the problem and present a rough overview of our approach over one particular use-case and possible analysis over citation graphs using enriched metadata of publications.

1 Introduction

Scientific progress can be accelerated by sharing as much information and experience as possible among researchers from different, yet related disciplines [3]. Currently available systems (e.g, ACM Digital Library, DBLP, CiteSeerX, SpringerLink, Google Scholar or Microsoft Academic Search) are supporting researchers in this process to a certain extent. Most of these systems are based on keyword search methods which retrieve a list of highly matched results from a large pool of information. Although they have been optimized with more facilities, they still lack support for researchers looking for information other than documents. Therefore, new researchers still need to do bibliographical analysis before they can really start their research. Even more experienced researchers are not guarded in the mass of publications, and it is not easy for them as well to find relevant information. Sample scenarios could be: “A research group applying for a large project needs to find partners belonging to active research schools in a special field”, “A new researcher in a community needs to know about the development of other related research groups over time to find what information has been inherited from a scientific works of one school to another”. “New recommendations for scientific documents not for newly published ones but for the missed ones during the past activities”. “Researchers in the role of reviewers that needs to determine the novelty and quality of papers under review”. The questions above cannot be answered by a single search rather depends on the process of building one’s personal knowledge graph of a research community which needs years of effort. Our vision is to have a system supporting scientific communication that automatically answers this kind of questions. The central assumption that motivates our research is that such questions can be answered in an automated way by exploiting the connections between different entities, including publications, persons, events, organizations and groups. Linked Open Scientific Communication is our approach to provide the realization of our vision as a combination of Linked Data, Semantic Web using enriched metadata, graph databases and incremental update methods.

2 Techniques

The initial step towards this vision is conceptually modeling of the application domain and analyzing the current situation of available systems supporting researchers in the level of communication via scientific document exchange [4]. In our previous work [5], the results of queries over a local collection of papers are evaluated indicating that such a system is able to answer the bibliographical questions which researchers might encounter. In that work, an algorithm is presented distinguishing between different kinds of reference relationships between papers. The idea is to extend the algorithm using more similarity metrics based on the metadata extracted from the scientific publications such as title, author, keywords and reference similarity, frequency of certain references, etc. We aim at realizing our vision over reusable linked data-sets of scientific publications such as DBLP¹. Using Linked Data that allows efficient distribution of data using the web standards, the challenges regarding handling the size of data is minor issue [1]. Therefore, it becomes much easier to make as many links as possible between different existing datasets as scientific resources. This means low cost computations through a large network of scientific communications. Backbones of this approach are using Linked data technologies with concepts of URIs as names for things, querying the data using standard methods based on RDF and SPARQL, visualizing results to illustrate them in a better understandable way. Therefore, the key components towards a new culture in scientific communication would be:

- sharing as much relevant information as possible among researchers
- having comprehensive bibliographical analysis to identify different components
- easily exchanging newly achieved results in addition to bibliographical metadata

3 Future work and perspectives

We have presented the idea of a system desirable for many researchers offering bibliographical analysis. We plan to implement different visualization illustrating results for different possible queries in the future. Using linked data technology, we will have the possibility to reuse existing applications to improve the implementation of our analysis algorithm. In particular, link discovery tools such as Silk or LIMES [2] could help to improve our similarity algorithm. Looking to the close future, we will study the usage of linked data vocabularies, linked data sets and tools which may practically improve the service that we are planning to provide.

References

- [1] G. Kahn and D. B. MacQueen. Linked open science - communicating, sharing and evaluating data, methods and results for executable papers. In *Procedia Computer Science 4*, 2011.
- [2] Axel-Cyrille Ngonga Ngomo and Sören Auer. Limes: A time-efficient approach for large-scale link discovery on the web of data. In *22nd IJCAI - Volume 3*, pages 2312–2317, 2011.
- [3] Petric, Ingrid, and et al. Literature mining method rajolink for uncovering relations between biomedical concepts. *Biomedical Informatics.*, 2(42), 2009.
- [4] Sahar Vahdati. A state-of-the-art report about systems and tools for supporting scientific communities. Master’s thesis, Department of Computer Science III, University of Bonn, 2013.
- [5] Sahar Vahdati, Andreas Behrend, Gereon Schüller, and Rainer Manthey. A flexible system for a comprehensive analysis of bibliographical data. In *4th International Conference WEBIST*, 2014.

¹<http://datahub.io/dataset/13s-dblp>