Open Research Online

The Open University's repository of research publications and other research outputs

Supporting Springer Nature Editors by means of Semantic Technologies

Conference or Workshop Item

oro.open.ac.uk

Supporting Springer Nature Editors
by means of Semantic Technologies

Francesco Osborne[1], Angelo Salatino[1], Aliaksandr Birukou[2],
Thiviyan Thanapalasingam[1], Enrico Motta[1]

[1] Knowledge Media Institute, The Open University, MK7 6AA, Milton Keynes, UK
{francesco.osborne,angelo.salatino,thiviyan.thanapalasingam,enrico.motta}
@open.ac.uk
[2] Springer-Verlag GmbH, Tiergartenstrasse 17, 69121 Heidelberg, Germany
aliaksandr.birukou@springer.com

Abstract. The Open University and Springer Nature have been collaborating since 2015 in the development of an array of semantically-enhanced solutions supporting editors in i) classifying proceedings and other editorial products with respect to the relevant research areas and ii) taking informed decisions about their marketing strategy. These solutions include i) the Smart Topic API, which automatically maps keywords associated with published papers to semantically characterized topics, which are drawn from a very large and automatically-generated ontology of Computer Science topics; ii) the Smart Topic Miner, which helps editors to associate scholarly metadata to books; and iii) the Smart Book Recommender, which assists editors in deciding which editorial products should be marketed in a specific venue.

Keywords: Scholarly Data, Ontology Learning, Bibliographic Data, Scholarly Ontologies, Data Mining, Conference Proceedings, Metadata, Classification.

1   Classifying Proceedings using Semantic Technologies

Correctly classifying proceedings and other editorial products in terms of the relevant research areas is critical to facilitate their discovery and to allow editors to take informed decisions about where to market them. Traditionally, this process has been handled manually by experienced editors, leading to high costs and slow throughput. In this short paper, we present a number of solutions informed by Semantic Technologies, which we have developed to address this issue in the context of a collaboration between Springer Nature (SN) and the Knowledge Media Institute (KMi) of The Open University. These solutions include: i) the Smart Topic API, ii) the Smart Topic Miner, and iii) the Smart Book Recommender.

Purely syntactic solutions, which extract frequent keyphrases from a set of documents, have shown to be limited for classifying conference proceedings [1], since they typically return a very large and unwieldy distribution of terms. The Smart Topic API addresses this issue by mapping these terms to concepts in the *Computer Science Ontology (CSO)* and returning a human-friendly number of structured topic descriptors. CSO is a large scale and granular ontology of research topics that has been created automatically, by running the Klink-2 algorithm [2] on the Rexplore dataset [3]. This consists of about 16 million publications in the field of Computer Science. CSO includes about 17K topics, which are linked by 70K semantic relationships. The mappings from keywords to concepts take into account both synonyms and sub-areas of a topic – e.g., all documents associated to terms such as "semantic technologies", "linked data", "RDF", "OWL" will be also tagged with the topic "Semantic Web". The result is a balanced topic distribution, which is both accurate in terms of topic annotation and also easy to understand and edit for the user.

The Smart Topic API, which is also available for research purposes, currently supports two web applications. The first is the Smart Topic Miner (STM)[1] [1], which helps editors in understanding and classifying proceedings i) by suggesting, for each proceeding, both a structured set of relevant research topics drawn from CSO, as well as a set of codes drawn from the SN Classification for Computer Science, and ii) by providing editors with an environment to make sense of the proposed annotations and edit them if necessary. The second application is the Smart Book Recommender[2] [4], which takes as input the proceedings of a conference and returns books, journals and other proceedings that are likely be of interest for its attendees. It does so by computing the similarity of SN editorial products over the vectors of semantic topics returned by the Smart Topic API.

2 Business Value

The STM tool has been routinely used by the SN Computer Science editorial team for classifying conference proceedings since January 2017. It is being applied to the Lecture Notes in Computer Science (LNCS) and other computer science series (LNBIP, CCIS, IFIP-AICT, LNICST), which publish about 780 volumes each year.

STM halves the time needed for classifying proceedings from 20-30 to 10-15 minutes. The benefits are especially evident when classifying complex multi-volume conferences, such as the European Conference on Computer Vision (ECCV[3]). The perceived value of STM can be roughly described as "it helps one to get 80-90% of topics correct very quickly". Indeed, while the classification of proceedings has traditionally been performed only by very experienced editors, thanks to STM it is now possible for assistant editors to perform the task, thus distributing the load and reducing costs. In addition, the adoption of a controlled vocabulary (in terms of the CSO ontology) makes the process more robust and facilitates the identification of related editorial products.

In the future, we plan to integrate the STM tool with the SN Linked Open Data portal[4], which describes Springer Nature conferences [5]. This will allow users to formulate complex queries that take advantage of the granular topic taxonomy provided by CSO.

In conclusion, we believe that this project offers an excellent example of how the use of ontologies and other semantic technologies can be effectively deployed in an organization to make workflows more robust and reduce costs.

References

1. Osborne, F., Salatino, A., Birukou, A., Motta, E.: Automatic Classification of Springer Nature Proceedings with Smart Topic Miner. In International Semantic Web Conference 2016 (pp. 383-399). Springer. (2016)
2. Osborne, F., Motta, E.: Klink-2: integrating multiple web sources to generate semantic topic networks. In International Semantic Web Conference 2015 (pp. 408-424). Springer. (2015)
3. Osborne, F., Motta, E. and Mulholland, P.: Exploring scholarly data with Rexplore. In International Semantic Web Conference 2013 (pp. 460-477). Springer. (2013)
4. Osborne, F., Thanapalasingam, T., Salatino, A., Birukou, A., Motta, E.: Smart Book Recommender: A Semantic Recommendation Engine for Editorial Products. In International Semantic Web Conference 2017 (Poster Track).
5. Bryl, V., Birukou, A., Eckert, K., Kessler, M.: What's in the proceedings? Combining publisher's and researcher's perspectives. In SePublica 2014 (2014)

---

[1] Demo available at http://rexplore.kmi.open.ac.uk/STM2_demo/
[2] Demo available at http://rexplore.kmi.open.ac.uk/SBR-demo
[3] https://link.springer.com/conference/eccv
[4] http://lod.springer.com/