

# ANNOTA: PLATFORM FOR CONDUCTING EXPERIMENTS IN DIGITAL LIBRARIES INFORMATION SPACE

Róbert MÓRO, Michal HOLUB,  
Jakub ŠEVCECH, Mária BIELIKOVÁ

*Slovak University of Technology in Bratislava  
Faculty of Informatics and Information Technologies  
Ilkovičova 2, 842 16 Bratislava, Slovakia  
{name.surname}@stuba.sk*

**Abstract.** Novel methods of navigation, search and exploration are needed in order to cope with the challenges posed by the volume of information present in the digital libraries. All these have to be tested and evaluated. For this purpose we built Annota<sup>1</sup>, a bookmarking and annotation service, which supports also the collaboration of the researchers. It combines a flexible and easily extensible domain and user model with an interface for setting up experiments. In addition, a dataset of research articles and the user interactions with them is provided.

## 1. Introduction

The main challenges of the current digital library systems include support of personalization and recommendation, support of exploratory search and navigation, use of semantics and Linked Data, as well as support of collaboration of the users (i.e. researchers).

These are only partially addressed by the existing systems, such as CiteSeer [7], CiteULike [6] or Mendeley [13], though they are very good at their area of specialization (crawling and indexing of freely available fulltexts on the Web, collaborative tagging and management of references in the personal collections of documents, respectively).

In addition, there is a need for a platform that would allow the researchers to easily and with minimal effort evaluate their proposed methods of navigation, search etc. As an attempt to address these challenges we designed a bookmarking and annotation service Annota, which serves also as a platform for conducting experiments.

---

<sup>1</sup><http://annota.fiit.stuba.sk>

## 2. Bookmarking and annotation: main scenarios

Annota comprises of two main components [1], [5]: a browser extension and a web interface. The main purpose of the browser extension is to allow the users to bookmark the articles in the digital libraries (though it is not limited to them; it can be used to bookmark any Web page) and annotate them using highlights, comments or tags. It also tracks and logs the activity of the researchers in the supported libraries, such as user queries, visited pages etc.

The bookmarks can be accessed with the help of the web interface, which provides the users with various methods of the bookmarks organization, search and navigation. It also supports collaboration of the users (researchers). We focused on the two main scenarios [1]:

- *Student – supervisor*: The supervisors can recommend their students important articles and track their progress. The students can generate reports of their activity and discuss the articles with their supervisors, as well as share articles that they find themselves.
- *Groups of researchers*: Whether it is for cooperation on the long-term project or an ad-hoc group of co-authors writing an article, the groups allow the researchers to share and comment the resources. The annotations of the researchers added to the documents can be also shared, thus supporting the collaborative annotation.

## 3. Evaluation support in Annota

The foundation of the evaluation support in Annota is formed by the domain and user models. They are represented as graphs; the user model is an overlay atop the domain model [2], [8]. The domain model includes all the activity of the users in Annota and in the supported digital libraries. The main entities are *paper*, *term* and *researcher* (the last added by the user model, see Fig. 1); the *paper-term relationship* reflects, how the terms are relevant for the paper. Analogically, the *researcher-term relationship* reflects the interests of the researchers.

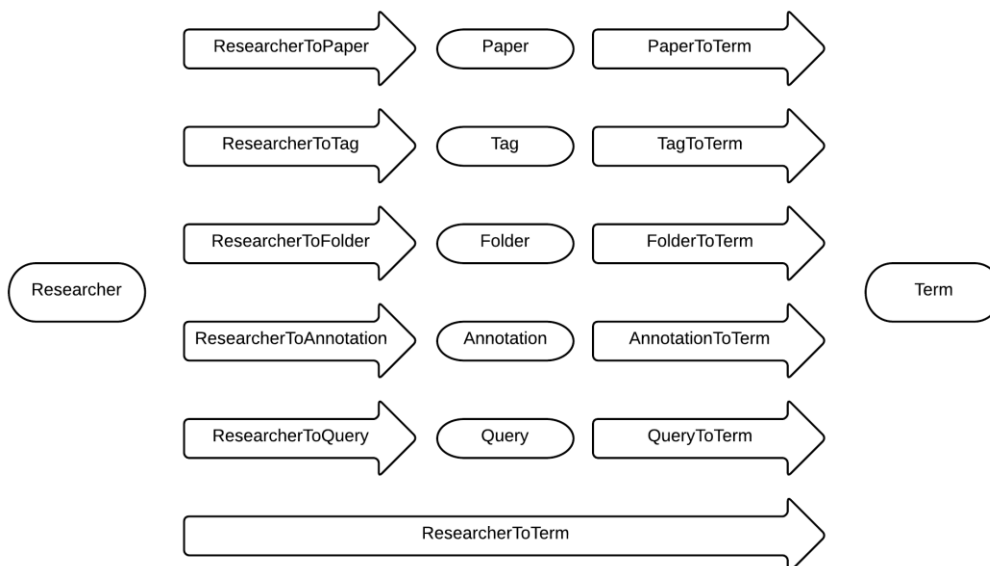


Figure 1. Conceptual design of the user model in Annota [8].

The main advantage are the flexibility and extensibility of both models. New relationships between the entities can be introduced into the model (e.g. when using or testing a new term extraction service). Various subsets of the existing relationships can be used to aggregate the user interests, thus allowing the different models to co-exist at the same time. This is very useful for evaluation purposes as we can compare results based on various models at once.

In addition, we support easy experimental setup including automatic creation of balanced groups of users based on their activities in Annota as well as A/B testing [2], [4].

All the collected data concerning the publications and the users is aggregated, anonymized and made freely available for research purposes in the form of a dataset<sup>2</sup> [2], [3], which now includes metadata about more than 50,000 papers, 230,000 authors and activity from 189 users. It is also published in the form of RDF which can be used for linking the entities in our dataset to the entities in the Linked Data cloud or for building semantic services.

#### 4. Conclusion

We have utilized the evaluation support in Annota as well as the dataset for evaluation of various methods of search, navigation or documents organization that we have proposed. In [11] we used the annotations as indicators of user interest when searching for related documents. In [9] we proposed a method of navigation using a cloud of terms, the content of which was personalized with the navigation history of the users. In order to support automatic and personalized organization of the documents in the researchers' personal libraries, we proposed a novel method of documents' organization in [10] using a facet tree. Lastly, we proposed a method for search query expansion in [12] by considering the context of the searchers' activities on the Web as well as in other applications (e.g. text processor).

By conducting these experiments, we showed the usefulness and the viability of the Annota platform. Besides its evaluation support, extensible domain and user models, Annota's main contribution is support of novice researchers based on the collaboration with their supervisors as well as on the provided methods of search, navigation and organization.

*Acknowledgement:* This work was partially supported by the Scientific Grant Agency of the Slovak Republic, grant No. VG1/0675/11 and VG1/0971/11 and by the Slovak Research and Development Agency under the contract No. APVV-0208-10. It was created also with the support of the Research and Development Operational Programme for the project "University Science Park of STU Bratislava", ITMS 26240220084, co-funded by the European Regional Development Fund.

#### References

*to other papers publishing the results that are summarized here*

- [1] Bieliková, M., Ševcech, J., Holub, M., Móro, R.: Annota – annotating the documents in the domain of digital libraries (in Slovak). In: *DATAKON '13: Proc. of the Annual Database Conference*. VŠB-Technická univerzita, Ostrava, (2013), pp. 143–152.

---

<sup>2</sup><http://annota.fiit.stuba.sk/dataset>

- [2] Holub, M., Móro, R., Ševcech, J., Lipták, M., Bieliková, M.: Annota: Towards enriching scientific publications with semantics and user annotations. *D-Lib Magazine*, (2014), to appear.
- [3] Holub, M., Ševcech, J., Móro, R., Bieliková, M.: Annota: Building a linked dataset of documents and annotations (in Slovak). In: *WIKT '13: Proc. of the 8th Workshop on Intelligent and Knowledge Oriented Technologies*. Centre for Inf. Tech., Košice, (2013), pp. 13–18.
- [4] Móro, R., Holub, M., Ševcech, J., Bieliková, M.: Annota: Sandbox for research in the fields of semantics-based search, organization and navigation in the domain of digital libraries (in Slovak). In: *Znalosti '14: Proc. of 13th Annual Conference of Znalosti*. (2014), pp. 88–91.
- [5] Ševcech, J., Bieliková, M., Burger, R., Barla, M.: Researcher activity tracking in digital library of scientific resources enriched by annotations (in Slovak). In: *WIKT '12: Proc. of the 7th Workshop on Intelligent and Knowledge Oriented Technologies*. STU Press, Bratislava, (2012), pp. 197–200.

#### Other references

- [6] Emamy, K., Cameron, R.: Citeulike: A researcher's social bookmarking service. *Ariadne*, (2007), vol. 51.
- [7] Li, H., Councill, I., Lee, W.C., Giles, C.L.: CiteSeerX: An architecture and web service design for an academic document search engine. In: *WWW '06: Proc. of 15th Int. Conf. on World Wide Web*. ACM Press, NY, (2006), pp. 883–884.
- [8] Lipták, M.: Researcher modeling in personalized digital library. Master's thesis, (2014).
- [9] Molnár, S., Móro, R., Bieliková, M.: Trending words in digital library for term cloud-based navigation. In: *SMAP'13: Proc. of the 8th Int. Workshop on Semantic and Social Media Adaptation and Personalization*, IEEE CS, (2013), pp. 53–58.
- [10] Móro, R., Bieliková, M., Burger, R.: Facet tree for personalized web documents organization. In: *WISE '14: Proc. of 15th Int. Conf. on Web Inf. Systems Engineering*, LNCS 8786. Springer, Berlin, (2014), pp. 372–387.
- [11] Ševcech, J., Móro, R., Holub, M., Bieliková, M.: User annotations as a context for related document search on the web and digital libraries. *Informatica*, (2014), vol. 38, no. 1, pp. 21–30.
- [12] Vnenk, L., Bieliková, M.: Searcher's activity in standalone and web applications as a source for search query expansion. In: *ENIC '14: Proc. of the 1st European Network Intelligence Conference*. IEEE, (2014), pp. 75–80.
- [13] Zaugg, H., West, R.E., Randall, D.L.: Mendeley: Creating communities of scholarly inquiry through research collaboration. *TechTrends*, (2011), vol. 55, pp. 32–36.