ScholarLensViz: A Visualization Framework for Transparency in Semantic User Profiles

Felicitas Löffler[1], Valentin Wesp[1], Samira Babalou[1], Philipp Kahn[1], René Lachmann[1], Bahar Sateli[4], René Witte[4], Birgitta König-Ries[1,2,3] ⋆

[1] Heinz Nixdorf Chair for Distributed Information Systems
Institute for Computer Science, Friedrich Schiller University Jena, Germany
{firstname.lastname}@uni-jena.de
[2] Michael Stifel Center for Data-Driven and Simulation Science, Jena, Germany
[3] German Center for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany
[4] Semantic Software Lab, Concordia University, Montréal, Canada
{lastname@semanticsoftware.info}

Abstract. Personalized applications are a two-edged sword. They are convenient and assist users by keeping the focus on relevant topics, but they are often black boxes and users typically do not know why certain entries appear in their profile. As transparency and provenance are essential for researchers, in this paper, we introduce *ScholarLensViz*, a visualization component for scholarly user profiles displaying a scholar's research competences including the provenance. It also provides visualizations to inspect the diversity of a profile and to analyze the semantic similarity of the profile entries.

Keywords: Visualization, Provenance, Semantic User Profiles, LOD

1 Introduction

Scholarly user profiles are gaining increasing attention and are becoming more important for various application scenarios, such as expertise retrieval [10], search and recommendation of research articles [1] or research network analysis [6]. These intelligent systems support scholars in time-consuming daily tasks, such as data discovery and data reuse, by providing content tailored to a scholar's preferences. In most applications, scholars are permitted to inspect these collected user preferences but do not know where the information comes from. However, the justification of entries is an important issue in scholarly profiles. Users need explanations why a certain preference is presented.

In this work, we address these issues and introduce *ScholarLensViz* [5], a visualization framework to display scholarly profiles based on a semantic user model. In earlier work, Bakalov et al. [2] proved the feasibility of a transparent visualization of semantic user profiles. We further developed that idea and present a new graphical user interface with linkage to the LOD cloud. *ScholarLensViz* presents a scholar's competences and their provenance. We obtain the competences from publications and store them in an RDF graph, using a workflow being introduced in our previous work *ScholarLens* [8]. The system also enables the inspection of a profile's diversity and provides visualizations to explore the semantic similarity of the profile entries.

2 Related Work

Various scholarly applications have emerged that create user profiles out of different resources, for different purposes and with various visualizations. *Semantic Scholar* [1], a smart search over research articles, generates user profiles from publications based on paper heuristics. In contrast, *AMiner* [10] focuses on expertise retrieval, a research field that aims at finding experts on a certain topic. Following a topic modeling approach, data is collected from institutional websites, documents and conferences. The visualization provided includes radar diagrams for author's statistics and stacked area charts for a scholar's automatically extracted research interests. Full semantic approaches are *VIVO* [5] and *Scholia* [6]. While *VIVO* provides a framework, an ontology and a graphical user interface for representing and visualizing scholars and their research context within an organization, *Scholia* aims at visualizing scientific bibliographic information through *Wikidata*. Both applications provide visualizations for co-author networks and research topics a scholar is competent in. In addition, *Scholia* offers various entry points to explore not only authors and publications but also organizations, locations or projects and supports numerous display formats such as timelines, scatter charts, line chart or trees. Provenance information, e.g., explanations from which source a topic or research interest has been extracted, is less considered in current approaches for scholarly user profiles. Our system attempts to close that gap and to leverage transparency and provenance in the visualization of semantic user profiles.

3 Architecture

Fig. 1 presents the overall concept. A scholar's provided or selected publications are converted from PDF to XML and are added to *ScholarLens* [8] (orange-colored) that processes the documents and extracts competences. The user model comprises three core concepts: scholars, their competence topics and a set of scholarly documents. All scholars are instances of the User Model Ontology (UM) [3] `User` class. Users have `CompetenceRecords` containing the individual

[5] ScholarLensViz source code, `https://github.com/fusion-jena/ScholarLensViz`

`Competence` as a literal. These classes are used from the Competence Management Ontology (CM) [4]. In addition, we model a user's publications with the PUBO Ontology [9] that provides properties to describe relations between documents and semantic annotations generated by the pipeline. The final output is an RDF graph containing the competences and being visualized by *ScholarLensViz* (blue-colored). We separated the visualization (client-side) and the calls to the

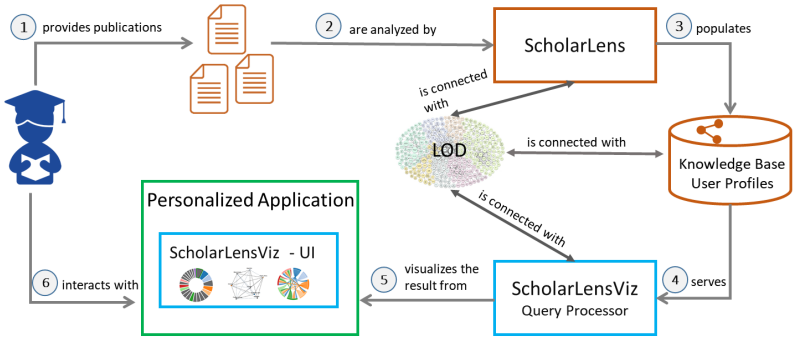

Fig. 1: Overall concept to construct scholarly profiles with ScholarLens and to visualize and integrate it in personalized applications with *ScholarLensViz*.

knowledge base (server-side). Thus, a server handles the SPARQL requests and provides HTTP methods to the client. This separation minimizes security risks such as sending direct SPARQL queries from the client to the knowledge base. Embedding *ScholarLensViz* into personalized applications, for instance as widget or standalone application, requires additional authentication mechanism to be handled by the integrating system (green-colored).

4 ScholarLensViz

ScholarLensViz provides three dialogs. (A) The start dialog aims to display a user's Top-25 ranked competences and their provenance. The elements of the pie chart (Fig. 2) represent a user's competences. The arc of the pie slices denotes the competence rank, and the color represents the category obtained from the Computer Science Ontology (CSO) [7] by matching the competence entries (which are presented as *DBpedia* entities) with CSO labels. The comment tab, becoming visible by clicking on a pie element, provides the competence description obtained from *DBpedia*. In order to justify the origin of the competences, we provide a provenance tab displaying the sentences from a user's publications and highlighting the relevant phrases that contain the competence and its rank, the paper title and, if available, the DOI.

In order to inspect the diversity of a user's profile and the semantic similarity between the profile entries, *ScholarLensViz* offers two further charts (Fig. 3).

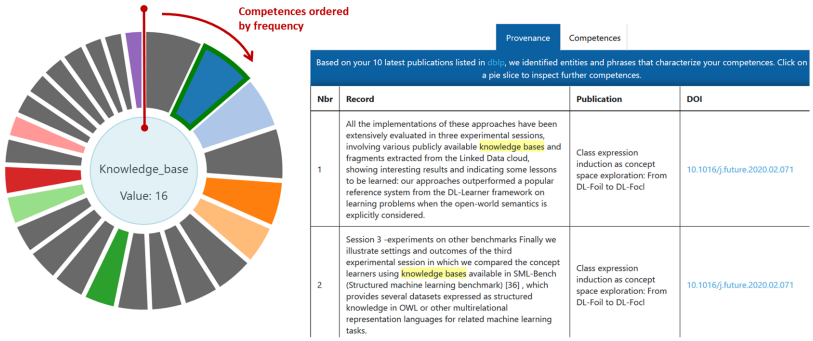

Fig. 2: The start dialog presents a scholarly user profile in a pie chart and displays the publications from which the competences have been extracted.

(B) A force-directed graph and (C) a chord chart visualizing the semantic similarity and relatedness between the competences computed with *Sematch* [11]. While semantic similarity is hierarchy-based, the semantic relatedness considers all relations between concepts [11]. We use the average of both values as both values influence the diversity of a profile. The more diverse a profile is, the less strong connections appear in the force-directed graph per competence. The chord chart allows a better analysis on the accumulated similarity values over the selected competences. The length of the outer arc reflects the number of similar connections a competence provides.

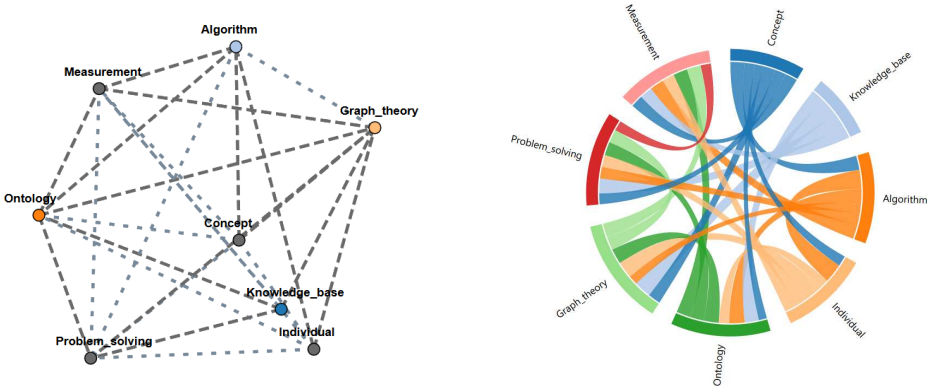

Fig. 3: Semantic similarity and relatedness of the profile entries are displayed in a force-directed graph (left) and a chord chart (right). Dotted lines in the force-directed graph denote a lesser (<0.4) semantic similarity, a dashed line a medium (0.4 – 0.7) similarity and solid (>0.7) lines a strong semantic similarity between the profile entries. The colors in the chord chart correspond to the colored competence entries.

5 Demonstration

In this demo[6], visitors can inspect the competences from known computer scientists working in the fields of Semantic Web. In addition, we added experts from other computer science domains as a basis for comparison. Users can select the competences and can change the thresholds for the semantic similarity.

Acknowledgements This work was supported by the German Academic Exchange Service (DAAD) within the scope of the PPP program and the Deutsche Forschungsgemeinschaft (DFG) within the scope of the GFBio project (229241684).

References

1. Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D., Dunkelberger, J., Elgohary, A., Feldman, S., Ha, V., Kinney, R., Kohlmeier, S., Lo, K., Murray, T., Ooi, H., Peters, M., Power, J., Skjonsberg, S., Wang, L., Wilhelm, C., Yuan, Z., van Zuylen, M., Etzioni, O.: Construction of the literature graph in semantic scholar. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (2018)
2. Bakalov, F., Meurs, M.J., König-Ries, B., Sateli, B., Witte, R., Butler, G., Tsang, A.: An approach to controlling user models and personalization effects in recommender systems. In: IUI. ACM, New York, NY, USA (2013)
3. Jovanovic, J., Gasevic, D., Siadaty, M.: Intelleo user model ontology. `http://intelleo.eu/ontologies/user-model/spec/` (March 2012)
4. Jovanovic, J., Siadaty, M., Gasevic, D., Milikic, N.: Intelleo competences ontology. `http://www.intelleo.eu/ontologies/competences/spec/` (2011)
5. Krafft, D., Cappadona, N., Caruso, B., Corson-Rikert, J., Devare, M., Lowe, B.: Vivo: Enabling national networking of scientists. In: Proceedings of the Web Science Conference 2010 (2010)
6. Nielsen, F.Å., Mietchen, D., Willighagen, E.: Scholia, scientometrics and wikidata. In: Blomqvist, E., Hose, K., Paulheim, H., Ławrynowicz, A., Ciravegna, F., Hartig, O. (eds.) The Semantic Web: ESWC 2017 Satellite Events. pp. 237–259. Springer International Publishing, Cham (2017)
7. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E.: The computer science ontology: A large-scale taxonomy of research areas. In: International Semantic Web Conference 2018, Monterey (CA), USA, 2018.
8. Sateli, B., Löffler, F., König-Ries, B., Witte, R.: Scholarlens: extracting competences from research publications for the automatic generation of semantic user profiles. PeerJ Computer Science **3**, e121 (Jul 2017)
9. Sateli, B., Witte, R.: Semantic representation of scientific literature: bringing claims, contributions and named entities onto the linked open data cloud. PeerJ Computer Science **1**, e37 (Dec 2015)
10. Wan, H., Zhang, Y., Zhang, J., Tang, J.: Aminer: Search and mining of academic social networks. Data Intelligence (2019)
11. Zhu, G., Iglesias, C.A.: Computing semantic similarity of concepts in knowledge graphs. IEEE Transactions on Knowledge and Data Engineering **29**(1) (2017)

[6] `https://dev.gfbio.org/scholar/`