Finding and Exploring Commonalities between Researchers Using the ResXplorer

Selver Softic[2], Laurens De Vocht[1],
Erik Mannens[1], Rik Van de Walle[1], and Martin Ebner[2]

[1] Ghent University - iMinds, Multimedialab
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
{laurens.devocht,erik.mannens,rik.vandewalle}@ugent.be
[2] Graz University of Technology, IICM - Institute for Information Systems and Computer Media
Inffeldgasse 16c, 8010 Graz, Austria
{selver.softic,martin.ebner}@tugraz.at

Abstract. Researcher community produces a vast of content on the Web. We assume that every researcher interest oneself in events, persons and findings of other related community members who share the same interest. Although research related archives give access to their content most of them lack on analytic services and adequate visualizations for this data. This work resides on our previous achievements[1,2,3,4] we made on semantically and Linked Data driven search and user interfaces for Research 2.0. We show how researchers can find and visually explore commonalities between each other within their interest domain, by introducing for this matter the user interface of "ResXplorer", and underlying search infrastructure operating over Linked Data Knowledge Base of research resources. We discuss and test most important components of "ResXplorer" relevant for detecting commonalities between researchers, closing up with conclusions and outlook for future work.

1 Introduction

"ResXplorer"[1] is aggregated interface for search and exploration of the underlying Linked Data Knowledge Base. Data within originates from Linked Data repositories DBLP(L3S)[2] which is a bibliography of computer science conference proceedings, COLINDA[3] containing information about up to 15000 conferences in the time range from 2003 up to 2013, DBPedia[4] common knowledge encyclopedia and Open Linked Data repository with geographical information named GeoNames[5]. Schematic structure of Linked Data Knowledge Base contains graphs of different

[1] http://www.resxplorer.org
[2] http://dblp.l3s.de/
[3] http://colinda.org
[4] http://dbpedia.org
[5] http://geonames.org

P. Zaphiris and A. Ioannou (Eds.): LCT 2014, Part II, LNCS 8524, pp. 486–494, 2014.

semantic entities represented as RDF (Resource Description Framework)[6] data model instances indexed and searchable by Apache Solar[7] interface. Functionality for keyword based finding of commonalities between research related artifacts (persons, publications and conferences) is an extension consisting of our earlier work on module for path finding between resources in semantic entity graphs, which is a part of the "Everything is Connected" engine (EiCE) [3], and interface solutions for exploration of Linked Data research repositories [4] based on Web 2.0 technologies.

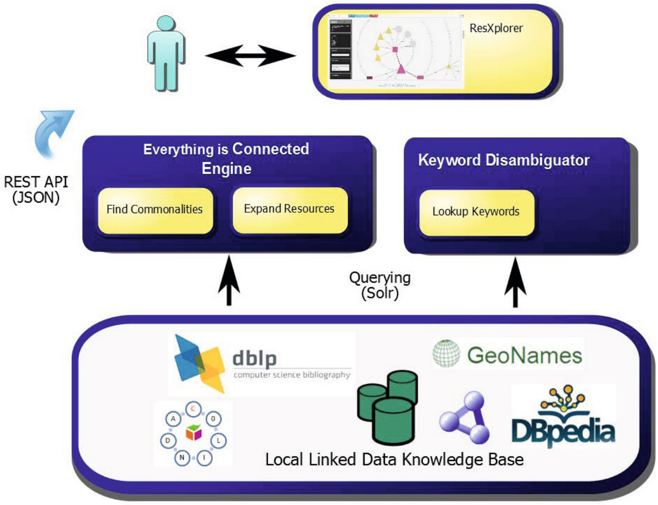

Fig. 1. ResXplorer concept for finding scholar artifacts necessary to reveal the commonalities

2 Finding Commonalities

As first step, a real-time keyword disambiguation via semantic entities from Linked Data Knowledge Base guides researchers by expressing their needs. Researcher select the desired meaning from a type-ahead drop down menu. Figure 2 shows the type-ahead expansion of results as disambiguation for “Laurens De Vocht” as “Agent” an entity which describes person or organisation in the Linked Data Knowledge Base. Expansion of results for entered terms happens in real-time. This feature is especially useful, during the early stages of the search as reported in [5].

Whole process around finding commonality is shown in figure 1. In behind the back-end (EiCE engine) connects the resources and ranks them according to the entered context. At the same time background modules also fetch neighbour links which match the selected suggestion. As result, choice of various resources is then presented to the researchers.

[6] http://www.w3.org/RDF/

[7] https://lucene.apache.org/solr/

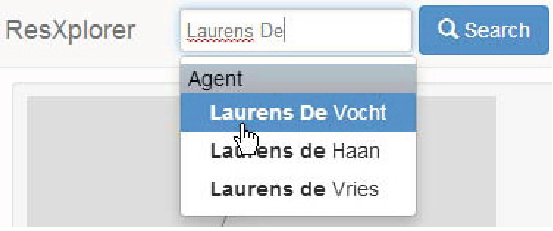

Fig. 2. Mapping of keywords

3 Visual Exploration

The visualization emphasizes commonalities by showing, on a radial map [6], how the current focused entity relates to the other found entities. It adopts the concept of affinity appropriately expressed in visual terms as a spatial relationship: proximity [7]. We additionally express the amount of unexpectedness as *novelty* of a resource in each particular search context. A typical example of such situation is in the Figure 3.

Features like color, shape and size of the items enhance user guidance during the exploration process [4]. The user expands the query space by clicking the results retrieved by the first keyword based search. Additional query expansion happens either through adding further keywords as well as through keyword combinations already entered where the back-end (EiCE engine) tries to deliver extra results based upon connection paths between the resources.

4 Evaluation of the Back-end

4.1 Setup

For evaluation of the module responsible to find commonalities, we defined a set of ten queries shown in table 1 consisting from the name pairs of authors of this paper knowing that they will deliver results, and that author profiles already exist in the DBLP bibliography archive. This set of queries is selected for reason to easier determinate relevance of results. Measurement of recall is left out intentionally because of the size of search space (hundreds of millions of potentially relevant resources).

4.2 Measures

Definition represented in equation (1) expresses precision as combination of *true positives* (TP), *false positives* results. Links discovered along traversing path of algorithm which lead to scientific resources (publications, persons and events) relevant for one of the both authors represent true positives. All other unresolvable or repeating links are false positives.

$$precision = \frac{TP}{TP + FP} \tag{1}$$

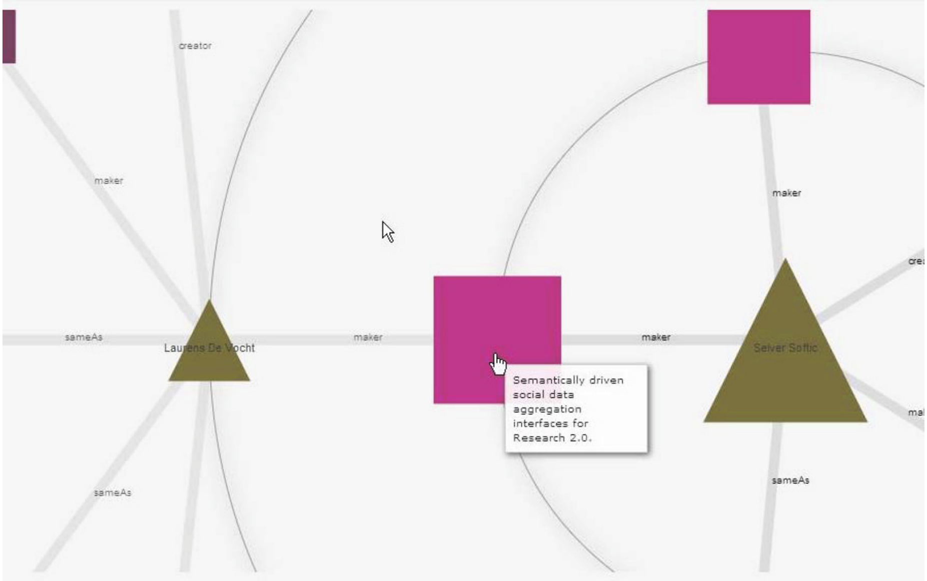

Fig. 3. Visual representation of commonality between *Laurens De Vocht* and *Selver Softic* based on common publications such as the highlighted "*Semantically...Research 2.0*"

4.3 Preliminary Results

Table 2 summarizes preliminary results of our tests. We measured precision of retrieved commonalities, path length between the two resources entered as terms of the query, and total count of discovered commonalities per query. The precision values range from **0.7** up to **0.95**. This precision rate is unexpectedly high even we knew that test queries represent authors who know and work with each other. These results are partly influenced by the well-connectedness of graph structures in the Linked Data Knowledge base. Path lengths are very short as expected and range from **2** up to **4** hops. Total count of detected commonalities ranges from **4** up to **11** except in query *Q10*. The explanation for this outlier is that relation in *Q10* is the strongest one because of the length of common period of collaboration between those two researchers and the number of together published works. They also have a bigger social network of collaborators which allows finding more alternative connection paths within semantic graphs than in the case of other queries. Evaluation of precision versus the path lengths in figure 4 reveals that; there is no linear dependency between the path lengths and precision. At least in our evaluation, results with shorter path lengths reach in average better precision then the ones with long paths.

Figure 6 shows that changes of total number of retrieved commonalities does not have any immediate significant impact on the precision score. This is not

Table 1. Set of queries, for finding of commonalities between researchers

Query	Keywords
Q1	Selver Softic, Laurens De Vocht
Q2	Selver Softic, Erik Mannens
Q3	Martin Ebner, Selver Softic
Q4	Martin Ebner, Laurens De Vocht
Q5	Erik Mannens, Martin Ebner
Q6	Erik Mannens, Laurens De Vocht
Q7	Laurens De Vocht, Rik Van De Walle
Q8	Rik Van De Walle, Selver Softic
Q9	Rik Van De Walle, Martin Ebner
Q10	Rik Van De Walle, Erik Mannens

Table 2. Precision, path length, commonalities count along the detection path for test queries

Query	Precision	Path length	Commonalities
Q1	0,75	2	4
Q2	0,86	4	7
Q3	0,78	2	9
Q4	0,75	2	4
Q5	0,82	4	11
Q6	0,83	2	6
Q7	0,83	2	6
Q8	0,7	4	10
Q9	0,7	4	10
Q10	0,95	3	37

surprising since the precision depends directly on the ratio of true positives and false positives.

For sure, most interesting finding reveals figure 6 where path lengths face the total counts of detected commonalities. The results depicted here discount the assumption that the length of a path traversed by algorithm within a graph structure which is well-connected implies inductively the increase of detected commonalities by each new hop. Even the outlier in the *Q10* proves this assumption wrong. This confirms once again the latter findings that solely quality of the detected commonality links determinate the precision and do not correlate strongly with changes of path lengths and total count of discovered commonalities. This finding is potentially influenced by the specific form of data graph structures in the Linked Data Knowledge Base, however this assumption is not confirm able with current results.

Quantitative reasons for the high precision are visible in figure 7 where total count of detected commonalities faces the count of true positives and false positives. The count of true positives almost correlates with the total count of commonalities which is a strong indicator for high precision.

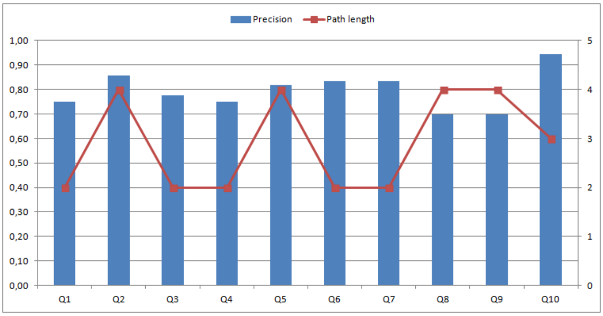

Fig. 4. Precision vs. Path lengths

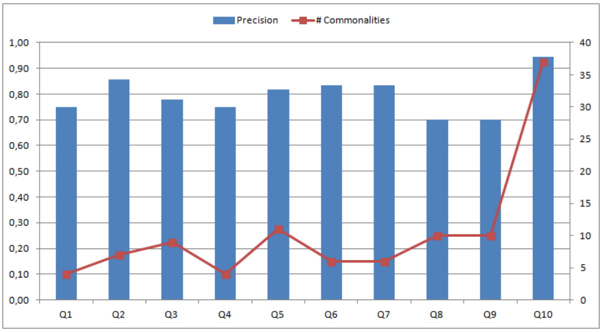

Fig. 5. Precision vs. total count of Commonalities

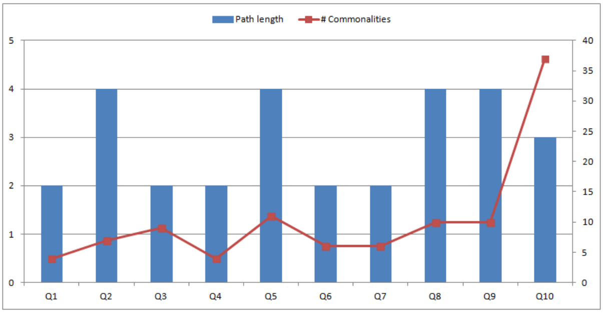

Fig. 6. Path lengths vs. total count of Commonalities

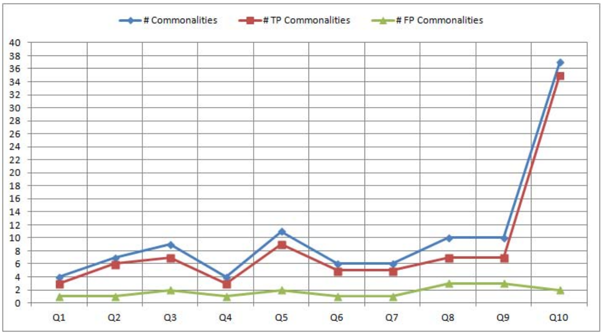

Fig. 7. Total count of Commonalities vs. TP Commonalities vs. FP Commonalities

5 Usability

We conducted a short survey on *percieved usefulness* based on the *Technology Acceptance Model (TAM)* [8] with 31 researches where users judged the usage of "ResXplorer" on a Likert-Scale with values (Strongly Disagree, Disagree, Undecided, Agree, Strongly Agree). The result of the evaluation shows the figure 8 and table 3.

Table 3. Preliminary results of the short survey on the *percieved usefulness*

ResXplorer			
What is the main goal of ResXplorer?	Goal	Score	Variance
1. [To explore]	**Explore**	**4.12**	1.61
2. [To discover]	Discover	3.88	1.86
3. [To search]	Search	3.71	1.10
4. [To analyse]	Analyse	3.18	1.78
5. [To clarify]	Clarify	3.12	1.74
6. [To tell stories]	**Tell stories**	**2.47**	1.70

The primary goal according to test-users for "ResXplorer" is to explore. According to the users, "ResXplorer" is not intended to tell stories. The users are unsure whether "ResXplorer" is more suited to analyse or to clarify. Highest score, at the moment it also has relatively low variance. Biggest variance and most averaged score goes above or below **2.5**. This is an indicator that users recognised the exploration as intention of the system.

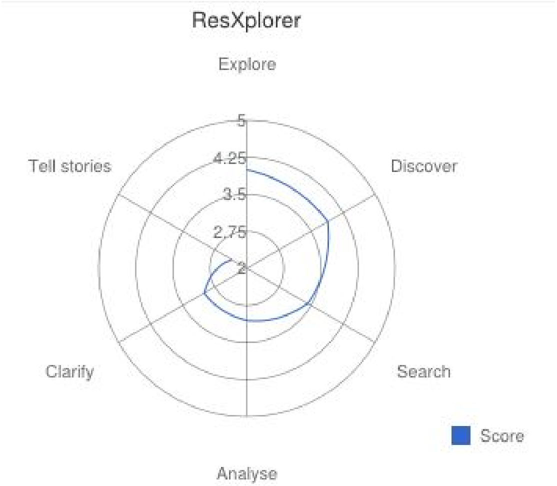

Fig. 8. Results of the short survey on *usefulness*

6 Conclusion

The main contribution of our work is allowing researches to interactively explore relations between the resources and entities like events, places, publications or persons related to their work and discover commonalities between them. Preliminary tests on "ResXplorer" back-end show that module for finding commonalities reaches high precision which does not depend from the length of search path, and the count of found commonality links but only from their quality and relevance. We also observed that longer traversed paths does not necessary mean implicitly bigger amount of discovered commonalities. All these findings lead us to assumption that underlying data is well-prepared and well-connected as well, and offers a variety of potentially interesting and useful resources for researchers. Conducted short survey on the "precieved usefulness" approved the 'ResXplorer" as exploration interface. In the future we want to extend the usability survey with aspects about the *ease of use*. Further we are aiming to extend our precision measurement on bigger test set to verify initially achieved good results. Moreover, we also want to test the assumptions about the quality and well connectedness of data in used Linked Data Knowledge Base.

Acknowledgement. The research activities that have been described in this paper were funded by Ghent University, iMinds (Interdisciplinary institute for Technology) a research institute founded by the Flemish Government, Graz University of Technology, the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

References

1. Vocht, L.D., Softic, S., Ebner, M., Mühlburger, H.: Semantically driven social data aggregation interfaces for research 2.0. In: Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, i-KNOW 2011, pp. 43:1–43:9. ACM, New York (2011)
2. De Vocht, L., Van Deursen, D., Mannens, E., Van de Walle, R.: A semantic approach to cross-disciplinary research collaboration. International Journal of Emerging Technologies in Learning (iJET) 7(S2), 22–30 (2012)
3. De Vocht, L., Coppens, S., Verborgh, R., Van der Sande, M., Mannens, E., Van de Walle, R.: Discovering meaningful connections between resources in the web of data. In: Proceedings of the 6th Workshop on Linked Data on the Web, LDOW (2013)
4. Vocht, L.D., Mannens, E., de Walle, R.V., Softic, S., Ebner, M.: A search interface for researchers to explore affinities in a linked data knowledge base. In: International Semantic Web Conference (Posters & Demos), pp. 21–24 (2013)
5. White, R.W., Marchionini, G.: Examining the effectiveness of real-time query expansion. Information Processing & Management 43(3), 685–704 (2007)
6. Yee, K.-P., Fisher, D., Dhamija, R., Hearst, M.: Animated exploration of dynamic graphs with radial layout. In: Proceedings of the IEEE Symposium on Information Visualization, INFOVIS (2001)
7. Pintado, X.: The affinity browser. In: Object-oriented Software Composition, pp. 245–272. Prentice Hall (1995)
8. Davis, F.: A Technology Acceptance Model for Empirically Testing New End-user Information Systems: Theory and Results. Massachusetts Institute of Technology (1985)