Computer Science and Information Systems 14(1):25–50 DOI: 10.2298/CSIS151028031D

ResXplorer: Revealing Relations between Resources for Researchers in the Web of Data

Laurens De Vocht[1], Selver Softic[2],
Ruben Verborgh[1], Erik Mannens[1], and Martin Ebner[2]

[1] imec - IDLab, Ghent University,
Technologiepark-Zwijnaarde 15 9052 Ghent, Belgium.
{firstname.lastname}@ugent.be
[2] Institute for Information Systems and Computer Media, Graz University of Technology,
Inffeldgasse 16c 8010 Graz, Austria.
{firstname.lastname}@tugraz.at

Abstract. Recent developments on sharing research results and ideas on the Web, such as research collaboration platforms like Mendeley or ResearchGate, enable novel ways to explore research information. Current search interfaces in this field focus mostly on narrowing down the search scope through faceted search, keyword matching, or filtering. The interactive visual aspect and the focus on exploring relationships between items in the results has not sufficiently been addressed before. To facilitate this exploration, we developed ResXplorer, a search interface that interactively visualizes linked data of research-related sources. By visualizing resources such as conferences, publications and proceedings, we reveal relationships between researchers and those resources. We evaluate our search interface by measuring how it affects the search productivity of targeted lean users. Furthermore, expert users reviewed its information retrieval potential and compared it against both popular academic search engines and highly specialized academic search interfaces. The results indicate how well lean users perceive the system and expert users rate it for its main goal: revealing relationships between resources for researchers.

Keywords: Exploratory Search; Social Media; Digital Libraries; Research Collaboration Tools; Science 2.0; Web of Data; Linked Data

1. Introduction

Peer-reviewed research publications as well as related metadata from bibliography archives are widely available on the web. They offer a vast amount of information on related publications and can facilitate suggesting new contacts, collaborators, and interesting custom events. Usually the platforms supporting this information exchange expose a Web API that allows access to the structured content, or the information is present as Linked Data. Social media, such as Twitter and Facebook, are often used during scientific events. Researchers use them to comment and discuss about each other's work, and to exchange research related materials [19]. They also use Web collaboration tools like Mendeley or ResearchGate to present their scientific work. Such academic social networks have become wide-spread and can have millions of regular users [48]. Resources for research are not always easy to explore: they rarely come with strong support for identifying, linking, and selecting those that can be of interest for further investigation.

Personalized adaptation of the Web to the needs of researchers is the main vision of Research 2.0 [45]. Research 2.0 depicts using such Web 2.0 tools and principles in scientific research and learning. It is an application field of "Technology Enhanced Learning" which covers the entirety of learning and research with use of new media. It is an approach to science that maximally leverages information-sharing and collaboration tools and emphasizes the advantages of increased online collaboration between researchers. Therefore, we developed a personalized interactive Semantic Search environment based on our search infrastructure and data from diverse open Linked Data repositories including scientific publication archives and social media. The purpose of this research is to offer a set of tools and services which researchers can use to discover resources as well as to facilitate collaboration via the web. These tools and services are considered as mash-ups, APIs, publishing feeds, and specially designed interfaces based on social profiles [32,45]. These tools and services are in line with the principles of Research 2.0.

The first use cases, data architectures, mash-up concept studies and prototypes on aligning the social web with semantics in the context of research, we introduced in 2011 [14]. The data modeling concepts were discussed in [38,39] while we investigated the back-end used for ResXplorer, a framework for discovery of chains of links between resources [10]. The aligning and matching of research related semantic resources was the main scope of our work on dynamic alignment of scientific resources such as web collaboration tools and digital archives [15,40]. We introduced the first prototypes of the ResXplorer search interface at conferences in late 2013 [12] and 2014 [37]. One of the first live versions participated at the Semantic Web Challenge 2013 [16]. The goal of all these publications was to evolve the concept, demonstrate the interface and visualization, trigger discussion and gain insight on the exploration workflow.

We investigate how researchers can find the information they need to gain insight about the people, conferences or publications they want to explore, which can be formulated under the form of following *research questions*:

- How effective can the execution of context-sensitive search tasks in semantically annotated digital archives be facilitated by revealing relations between resources (i.e. adequately addressing the researcher's intent) and leveraging the links to social media and web collaboration tools?
- What is the contribution of exposing resources part of a revealed relation that the researcher was not aware of beforehand?
- For which kind of tasks does an increasing amount of user actions by the researcher positively influence the relevance and precision of the search results?
- Does the underlying approach excel in revealing relations in a certain context compared to state of the art? At what cost does it come in terms of required user actions and precision for more traditional search tasks?

We will test the following *hypotheses*:

- Interacting with the search results refines and improves the result set because interaction with the result set makes the information contained in the initial search query more specific.
- Each iteration enables the researcher to define more and more targeted queries, leading to more precise results.

We propose ResXplorer[3] to address these research questions by focusing exactly on

[3] `http://www.resxplorer.org`

exploring chains of links between these resources. ResXplorer is one of the first practical solutions combining the social web and the semantic web in an interactive search environment that visually emphasizes and represents the search context and results. We focus on showing how it is an example for the use of interactive visualizations to enable knowledge discovery in Linked Data, which can be invaluable to researchers. One of the use cases we support, focuses on the end-user usability of semantically enriched researcher profiles. In this use case, our prototype “ResXplorer” shows relations between researchers based upon the semantic analysis of researcher’s tweets and aligned with information about conferences and proceedings, users are presented how they are (indirectly) related based on their institutions, visited locations, and conferences they contributed to. As a measure of usability we investigated the ability of our search interface to support the construction of a good cognitive model of the underlying data and the relations within the data. Finally, we measured the effectiveness and productivity of the interface by checking to which extent end-users carry out knowledge-intensive and analytical tasks.

2. Related Work

The focus of this section is to discuss and pinpoint the efforts done in the field of interfaces for search and in the field of discovery of relevant content for researchers.

2.1. Alternatives without semantic graph in the back-end

Approaches not relying on semantic graphs typically recommend multi-faceted information to discover for example research, popular citations. One of these methods is *behavioral pattern discovery* in research (e.g., publishing papers, citing papers): the focus there lies in to understand user behavior and discover patterns to apply to for example web search, recommender systems and advertising.

Traditional methods usually consider the behaviors as simple user and item connections, or represent them with a static model. However, in practice, user behavior is dynamic: it includes correlations between the user and multiple types of objects and evolve over time. An example for the latter is the Flexible Evolutionary Multi-faceted Analysis (FEMA) framework for both behavior prediction and pattern mining [26]. FEMA utilizes a flexible and dynamic factorization scheme for analyzing human behavioral data sequences, which can incorporate various knowledge embedded in different object domains.

Related to this is the question on how possible author-keyword associations evolve over time. Rather than relying on graphs, some approaches manage to implement and apply this using *tensors* [26]. Working with graphs does not necessarily mean that they always require the entire knowledge base to be semantically annotated. Graph algorithms such as for *clustering* have been found to be a good way to identify experts in large knowledge bases where the semantics are being introduced only in the last stage of the analysis for uniquely identifying resources with Uniform Resource Identifiers (URIs). For example, related author profiles, which are initially parsed as unstructured, tree- or tabular structured data, can be dynamically clustered together via an author disambiguation process [25].

2.2. Interfaces for Research based on Linked Data

Our approach focuses on finding relationships between resources. It is a distinct example solution and implementation of an adaptive and intelligent web-based system [4] yet does not currently aim to compete in terms of feature completeness. A review on related search interfaces for science leads us to a few working solutions worth mentioning, despite the huge amount of published Linked Data especially publications meta data. A number of working solutions worth mentioning are: RKB Explorer, Faceted FBLP, and Bibbase.

RKB Explorer[4] [21], a visual browser which originated from the ReSIST[5] network of excellence, which unites within its realm many sources of scientific data. This visual browsing interface is based on categorised pre-selection and focuses on people, organisations, publications, and courses and materials. The search always centers around the selected category which makes the context based browsing less flexible but focused. Within the visualisation RKB Explorer evaluates relations of the first degree. In comparison to RKB Explorer our approach is more user and search centric rather than concept and context centric. In our interface, a user profile affects the pre-selection of search results. Users can configure the search context by executing searches for resources or by expanding one or more resources.

Another advanced research related effort is *Faceted DBLP search*[6]. The search approach in this case resides on DBLP++ data which enhances DBLP with additional keywords and abstracts as available on public web pages. It integrates facets on Time, Venues, Publications Years and Authors and delivers the results in various formats. These formats include: BibTeX, regular web pages, DOI identifiers, or RDF. Faceted DBLP offers a good flexibility in filtering and narrowing down the results as well as implementing basic syntactic query expansion based upon single word and whole phrase in an anonymous way. Retrieval is done by classic search engines and result selection is done by ranking without any possible relation to the user profile.

BibBase[7] [53] has an interface to leverage the personal publications into the Web of Data and integrates the retrieval of author publications with a small sample from Mendeley[8], DBLP[9] and Zotero[10].

2.3. Systems for Semantic Search

In recent years significant efforts have been made in semantic technologies and regarding the Semantic Search. An overview on related work in this area can be found in [46,47]. Semantic Web search engines like Hermes [44] have been developed which regarding initial intention and application is very closely related to the use case of our work. Whereas Hermes tries to translate the keywords into structured queries, our approach expands results using paths within the connected Linked Data graphs. ResXplorer uses the user's social profile as additional context information for these graphs.

4 http://www.rkbexplorer.com
5 http://www.resist-noe.org/
6 http://dblp.l3s.de/
7 http://bibbase.org
8 http://www.mendeley.com/
9 http://www.informatik.uni-trier.de/~ley/db/
10 http://zotero.org

ResXplorer shares the goal of search, data about research publications, and intended audience with Google Scholar (GScholar)[11], Microsoft Academic (MA) Search[12], ARnet Miner [42][13] and Falcons [6][14]. Besides Hermes also Falcons [7,6], Watson [9], SWSE [23] and Sindice [30] were developed for retrieving semantic data from the Web. These engines primarily rely on keywords as starting positions for defining and specifying queries. Hermes, Falcons and SWSE, also support more advanced querying capabilities, including basic SPARQL graph patterns. In general, the semantic matching frameworks within these semantic search engines reside on the approach of matching graph patterns against RDF data with, if applicable, inference run against the RDFS/OWL ontologies. This kind of semantic matching mechanism is also widely implemented by a wide range of RDF stores.

2.4. Vocabularies to annotate Social Media

Common vocabularies to annotate social media as Linked Data are: Friend of A Friend (FOAF)[15], Semantically Interlinked Online Communities (SIOC)[16] [3,2] and Dublin Core[17] [49]. FOAF describes the user profiles, their social relations and resources. SIOC is mostly combined with FOAF and Dublin Core for creating instances of web entries like blogs, microblogs, mailing list entries, forum posts, along with other entries from Web 2.0 platforms [43,38,14]. Passant et al. improved mapping social profiles with related content, such as via interlinking tags [33,34,35].

3. Interacting with Research and Social Media Data

Our approach collects and uses information from resources already explored by other, more experienced, researchers. This is especially interesting for cases when looking for the next practical piece of information or when trying to find a solution for a problem that requires out-of-the-box thinking (e.g., when forming the exact search query requires background knowledge of a domain unfamiliar to the researcher). The interaction and models in Figure 1 show how the researchers use this method.

Researchers can define and select their *intended* search goal over several iterations. When users are looking for new leads, they get an overview of possible objects of interest (like points of interest on a street map) by having their activities and contributions linked on social media and other platforms such as their own research publications profile.

3.1. Data Model

The Data Model has two spaces. It has a Linked Data space and an Entity space. The former is the representation of the data loaded into the model and the latter are the entities, each having a URI, a label, a type and a description consisting of one or more Linked Data triples. In this section we describe the two types of data that we model: Research Data and Linked Data extracted from social media.

[11] http://scholar.google.com
[12] http://academic.research.microsoft.com
[13] http://artnetminer.org
[14] http://ws.nju.edu.cn/falcons
[15] http://xmlns.com/foaf/spec/
[16] http://rdfs.org/sioc/spec/
[17] http://dublincore.org/documents/dcmi-terms/

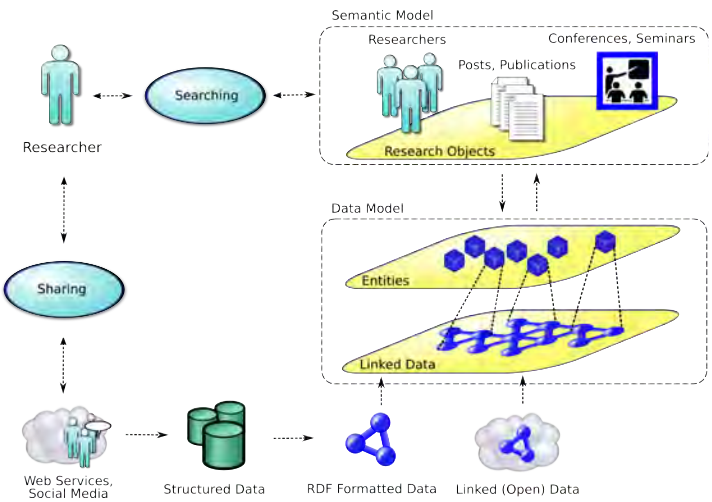

Fig. 1. The information researchers share via the Web services of research collaboration tools and social media is structured and transformed to RDF and interlinked with Linked Open Data. The resulting entities in the data model form the base for the semantic model. This process is outlined in Section **3.1**. When researchers search, they interact indirectly with the semantic model which we detail in Section **3.2**.

Research Data Research data is described as Linked Data using state-of-the-art vocabularies. We model research data with respect to their usage and wide popularity within the Semantic Web community, as well as to their applicability for the proposed use case. One of the modeling domains of interest are scientific events and their relatedness to bibliographical archives. The "Digital Bibliography and Library Project" (DBLP)[18][29] provides bibliographic information on major computer science journals and proceedings and indexes more than 2.3 million articles. Besides it also has many links to home pages of computer scientists. The Conference Linked Data (COLINDA) data set resolves this connection. COLINDA describes conferences using the Semantic Web for Research Communities (SWRC)[19] ontology [41]. Especially important for this decision was that DBLP Linked Data also applies this ontology to describe its resources.

Linked Data from social media We created an annotated set of extracted conference hash tags mentioned in tweets of researchers which would be associated with corresponding tweets and which can be used for further mining tasks like label based matching of scientific events in Linked Data sets e.g., COLINDA, DBLP. The motivation for linking data from social media: 'social data' as such, is manifold:

Link discovery It allows detecting and creating links between the users and the data they are exploring.
Timely context It enforces a timely and personalized context to the search.

[18] `http://dblp.l3s.de`
[19] `http://ontoware.org/swrc/`

Relationships It adds additional relationships between users and resources that are contained in the more static data and potentially introduce additional references to other Linked Open Data.

In our case, besides persons, locations, conferences and scientific publications, the researcher oneself is an important resource for the context of our search scenario. To triplify the personal profile of the researcher, we have used the FOAF vocabulary according to the common method within the community to describe the profile of a person or agent [18,34]. We obtained Twitter-related content from our tweets cache Grabeeter[20], which contains very useful set of more than a million tweets from couple of thousands of users in academia and the Twitter API[21] via a self developed "Semantic Profiling Framework". The necessity of creating our own framework emerged from the lack of solutions which turn microblog posts into RDF which met our needs. There have been some attemps like Semantic Tweet[22], Smesher[23] or SMOB[24] but they only partially met our needs. The other reason why we created own solution built on prior insights of Semantic Web community is that in this way we were also able to integrate the tweets and profiles from Grabeeter. The platform should be capable to integrate some other research related social platforms like e.g., Mendeley over the user accounts information. In this way, for instance, tweets can be aligned to publications at Mendeley in cases where users own both Twitter and Mendeley accounts.

To model micro blog content we used SIOC combined with the Dublin Core vocabulary to model the tweet text and information about it, such as post date and link to creator as recommended in earlier works on this topic [3,2]. For creator profiles we used FOAF to model and preserve the information. We used the Mendeley API[25] to enhance creator profiles, using the data from corresponding Mendeley profiles, with links to publications, scientific events, and persons found there.

3.2. Semantic Model

We introduce *Research Objects* [11], which center and group refined entities of extracted and integrated data in the Data Model and represent:

- Events: scientific conferences, seminars and/or lectures
- Publications: articles, reports, tutorials and/or posts
- Locations: both real-world and online (web pages, webinars)
- Concepts: topics, categories and/or classifications

Research objects enable and facilitate the use of research related information. The metadata that describes research objects facilitates searching and retrieving them.

[20] http://grabeeter.tugraz.at
[21] https://dev.twitter.com/
[22] http://datahub.io/dataset/semantictweet
[23] http://bnode.org/blog/2009/02/16/devx-article-about-semantifying-and-sparqling-twitter-streams
[24] http://apassant.net/2010/01/22/smob-v20/
[25] http://dev.mendeley.com/

Defining Research Objects A single *Research Object* can contain links to and information about an online tutorial, details about a seminar, links to fragments of related papers and tutors or people who are known to have contributed to the entities of this specific object. Researchers define a search query for their research and have it parsed by our system for identification in terms of the Semantic Model. Based on the research goal, we select the found research objects that are most closely related, after they have been refined from their representation in the Data Model. Publishing Linked Data into the cloud, however, does not ensure the required reusability, but the use of research objects in a semantic model should provide the reproducibility that enables validation of results [1]. We align the entities present in the Data Model with the registered activities of researchers by providing their profiles and feeds of social media. Researchers generate those by sharing and monitoring online activities such as blogs, (micro)posts, tags, shares and other resources.

Searching Research Objects We center searches around several research targets that a researcher wants to relate with another. We also combine related resources based on common links they share, such as being related to and containing more information about a Research Object. The users generate their own views by exploring and searching among the Research Objects in the model and can share or compare those with other researchers or earlier searches. All those views together lead to a personalized environment. This will boost interaction with and grouping of similar views and objects to bigger packages that ultimately lead to the discovery of even more relations. We customary map all objects for users based on their "researcher profile". We extract the research profile based on the content researchers monitor on social media or the resources they have shared over it. Most of the researchers today own a profile in a scientific or common social network like Twitter and Facebook[26] or on research related platforms like ResearchGate[27], Mendeley, or Google Scholar.

4. Interactive Search

In each search session users can combine keyword-based disambiguation combined with visual refinements through expansion queries on the semantic entities we recognize as facets. The idea behind their usage as facets is to offer always and at each step a complete understanding of why certain results are showed. We do not want to let the algorithm assume things about researchers preferences. Since it is meant to be an exploratory search, the point is to involve the researchers on the base of input-output principle in a guided approach through facets and three dimensions: shapes, colors and size. They choose what they want to see and get that result delivered. As a parallel process in the back-end, the engine discovers additional relations between the search results and presents them as alternatives to the already acquired information. In this way we automatically expand or narrow down the facet range available to the researcher. This leads the researchers through the data by offering them at each point in time exploration and involvement of new and already found items into the search. This section firstly describes the front-end and back-end, and illustrates the dynamics of the interactive search from a users point of view using an example.

[26] http://www.facebook.com/

[27] http://www.researchgate.net/

4.1. Remarks on significance of social media in interactive search

Social media Linked Data primarily contributes strongly to result and search personalization. In back-end it acts as nexus between conference and digital archive repositories (here COLINDA and DBLP) through the links pointing to this sources from social media profile data of the user. Further, it enhances the results with additional event and publication information from users's Twitter or Mendeley account. Since not every user has both accounts or any of them the involvement of this data happens through explicit user's authorization to ResXplorer to include his Twitter or Mendeley account into search process. The data is then extracted and brought into semantic form as well interlinked in background to offer a starting search configuration adapted to user. Missing the social media links primary results in shortages respectively insights on relations between the authors who share same scientific events and publications which are identifiable only via social content their accounts which are not necessary included in COLINDA or DBLP Linked Data.

4.2. Front-end

The decision process during the search is supported by a real-time keyword disambiguation to guide the researchers in expressing their research needs. We do this by allowing users to select the correct meaning from a drop down menu that appears below the search box. Presenting candidate query expansion terms in real-time, as users type their queries, can be useful during the early stages of the search [50].

Researchers can define and select their 'intended' search goal over several iterations. A combination various resources is then presented to the researchers. In case they have no idea which object or topic to investigate next, they get an overview of possible objects of interest (like points of interest on a street map). Researchers define a search query for their research and have it parsed by our system.

Based on the ability of humans to rapidly scan, recognize, recall images and detect changes in size, color, and shape, we aim to enhance the guidance of users during their search by using several visual aids of which the three most visible are:

1. **Shape:** We group sets of types in large groups and represent them using a shape. Types that cannot be assigned a group are grouped in a category 'Miscellaneous'. The shapes help the user to distinguish between the types of offered results
2. **Color:** Every entity has a type and associated unique color. For a certain result set the user gets an immediate impression of the nature of the found resources. Figure 2 depicts two different objects related to other objects and therefore have a different shape and size. On the left of the search interface there is a legend explaining the researcher the meaning of shapes and colors.
3. **Size:** Each entity is ranked according to novelty and relation to the context and enlarge those that should attract attention from the researcher first, this is shown in Figure 3. The novelty quantifies the degree of being new, original or unusual. Particularly in this context it entitles resources that are remarkable and differ from the others because of their direct relations with neighbours or their semantics (in terms of occurring predicates). A goal of the search is to explore information not seen before which makes it difficult to define an accurate search goal. Besides allowing to search specific entities, our visualization facilitates exploratory browsing. This is particularly useful when information seeking with unclear defined search targets [31].

Fig. 2. Different shape and color to distinguish types

Fig. 3. Different sizes to guide the user's focus

Figure 4 shows how researchers can track the history of their search: the explored relations are marked red and clearly highlight the context of a search. This is a good example of how our system adapts to the users and their environments. It shows one of the ways how to build a model of the goals and knowledge of an individual user[5], and the model is used throughout the interaction with the user. Researchers can click on a list of resources they have searched to focus the visualization. A screencast of the search interface is available online[28]. In this screencast, we show how researchers interact with the search interface and the above described visualization.

Fig. 4. A red line marks the explored relations in the visualized search context.

4.3. Back-end

The back-end supports the search process for tools like ResXplorer with two main tasks: discovering links between two resources and the ranking of found links and resources as shown in Figure 5. We find paths between resources for discovering links to match the seed input given by the researcher. With the delivery of first results, our engine expands the query and enhances the context by pathfinding and neighbours resolution within the "Everything is Connected engine" (EiCE) [10]. It uses the 'distance' to the first query as a

[28] http://youtu.be/tZU97BQxE-0

measure for ranking the result. The EiCE is used here to compute heuristically optimized minimum cost paths between pairs of researchers, publications, conferences or mixed pairs. The heuristics take into account the rarity of resources to avoid common resources (that have many in- or outgoing links) and the semantic relatedness between resources. Each time a user adds another resource to the results, the visualized path between the resources takes these factors into account.

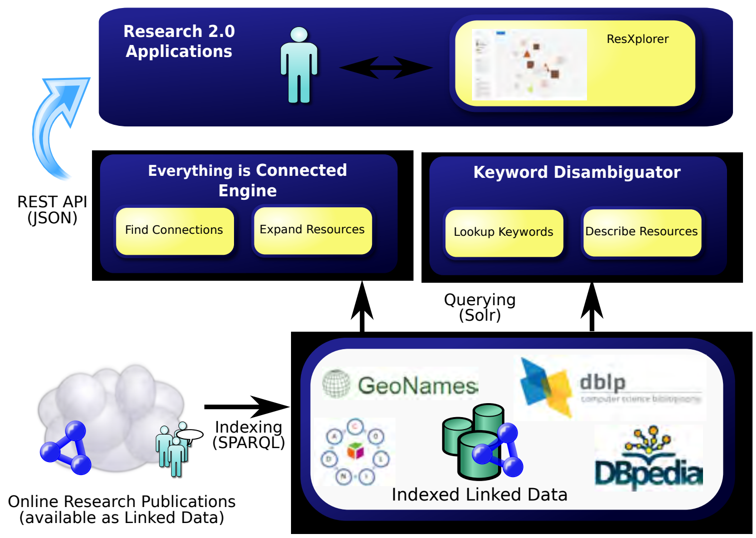

Fig. 5. Applications like ResXplorer use the Everything is Connected engine for finding relations between resources.

4.4. Example Illustrating the Dynamics

Each search starts within the search interface where a user can either login or query anonymously the Semantic Search Engine. Our search interface distinguishes between two types of queries: a query which consists of several keywords as seeds given as input and a profile-driven query, used as preset for further search, driven initially by user background information.

Except for the first step, the querying paradigm applies to the personalized search as well. The query in the figure illustrates the common case where a researcher enters the search process by entering simple keywords and tries to resolve the context of "finding useful resources from a certain conference".

1. One searches for a specific conference "Linked Data" and articles related to "WWW2012". Firstly, as Figure 6 shows, the visualization focuses first on the logged in user upon which the user can choose to expand on of the neighbouring resources.
   We note that the user changes the focus of the result view by clicking on a resource: the resource encircled with a mixed line. Within the simplified query progression process, entered keywords are first mapped towards the entities and properties in the index.

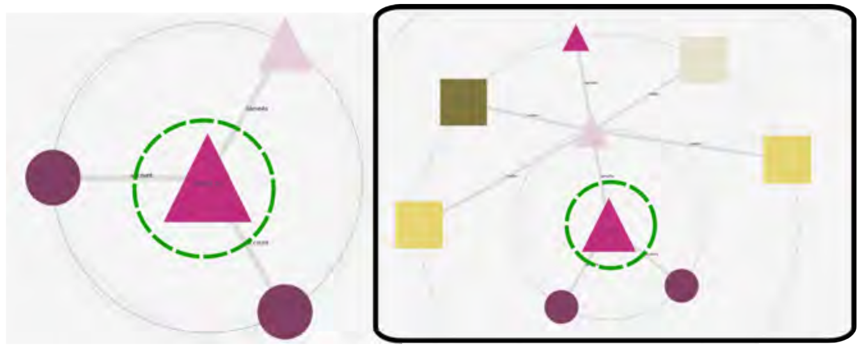

Fig. 6. User expands a direct neighbor after ResXplorer focuses on logged in user (encircled with dashes).

2. As a search result the engine delivers first set of links for each keyword entered, such as in Figure 7 for “Linked Data”.

Fig. 7. User searches for “Linked Data” and ResXplorer reveals the chain of links between the selected document (encircled with dots) and the user.

3. If available, the system also delivers the types of entities discovered in index. When the user searches for the next keyword “WWW2012”, relations to other already visualized resources are exposed as indicated in Figure 8.
4. By entering the location, for example “Germany”, one could narrow further the focus of the context by location. Each time a combination of various resources is visualized, the application suggests new queries: they are generally most useful for refining the system’s representation of the researcher’s need.
   In case they have no idea which entity to focus on or what topic to investigate next they get an overview of possible entities of interest, like points of interest on a street map. By profiling their activities and contributions on social media and other platforms such as their own research publications, the affinity with the proposed resources is enhanced.
5. With each further iteration the user can choose either one of two actions:
   – **Query Expansion:** The user expands the query space by clicking the results retrieved by initial keyword based search.The resolution of results is based upon

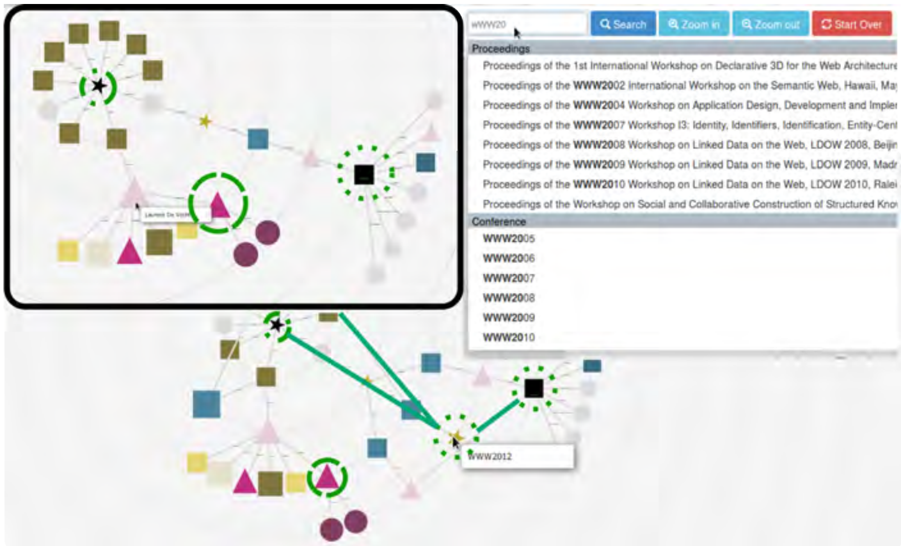

Fig. 8. After the user focused on a common resource of both and searched for "WWW2012", ResXplorer reveals relations to the selected conference "WWW2012" (encircled with dot-dashes).

the properties of Linked Data instances like *rdf:label*, *owl:sameAs*, *rdf:seeAlso*, *dc:title*, *dc:spatial* or *dc:description*. Those properties have been used in generation of Linked Data instances to preserve conference shortcuts (e.g. WWW2012), point to link of proceedings of a conference or ,to connect alternative link about it, as well to literally describe the venues of scientific events.

– **Additional Query Formulation:** Additional query expansion happens either through adding further keywords as well as through keyword combinations already entered where the back-end tries to deliver additional results based upon connection paths between the resources. What happens in return is that the engine tries to identify the terms that have been searched in the result space. In cases when they can be resolved by a Linked Data instance, the algorithm continues step by step looking via links to the neighbors of the instance to find a path to other terms identified by the engine as well. After a certain number of steps (here, seven) it terminates if it is unsuccessful.

5. Evaluation

We consider non-Linked Data researchers to be typical lean users of our system for the purposes of the Research 2.0 case and Linked Data specialists as expert users, especially because of their familiarity with the underlying graph structure. The primary target group of users of our system are non-Linked Data researchers, and in this particular research 2.0 use case, scientific researchers. They interact with research objects and not the more complex underlying Linked Data model. Another group of users which we must bear in mind are domain experts, as they are likely to have a very good understanding of data structure and content in their domain, and bring this knowledge to guide both browsing research and targeted searches. Based on the experience of our previous work [13][17], we selected experts and researchers in computer science and digital media as test group representatives. This group of people is also our target audience.

5.1. Methodology

Exploratory search represents a cognitively very intensive activity. Therefore conduction of searches should be possible with minimal interruptions. According to White et al. [52,51] : “Techniques such as questionnaires and interview techniques can be valuable tools, but one must be careful to include them in the experiment in such a way as to not interfere with their exploration”. We evaluated the tools in two ways: lean user tests and expert user reviews. These methods give us insight in how the users perceive the tools and show us quickly potential bottlenecks [22]. They also deliver us insight on how precise our solution performs in comparison to the existent state of the art solutions of industry as well as academia. All of the compared solutions target the same audience. They differ in implementation and interface design, but more importantly, they have more or less valuable stock of users. The choice of applied evaluation methodology was made by applying relevant aspects out of already existing achievements in this field introduced in [27,50,20] and adapting them to our specific use case. Since we want to offer a solution for research and learning purposes but also for wider community of users, a user centered methodology plays a decisive role in our evaluation process.

5.2. Implementation

We have developed the interface for search in a research linked data knowledge base combining the latest Linked Data technologies with an advanced indexing and path finding system. We build our implementation upon our earlier work using the “Everything is Connected” engine (EiCE) [10] EiCE and Web 2.0 technologies (such as JQuery and Django). The interface itself is a realization in HTML5 and Javascript making advanced use of JQuery UI in combination with the “Javascript Information Visualization Toolkit”[29]

5.3. Lean User Study

We have evaluated the effectiveness and productivity of our environment in this user study. The conducted user study focused on ResXplorer aimed at measuring the effectiveness, productivity and impact of the features of our environment. Sixteen test users were selected to participate to the observation and they received no information about the tool in advance. According to Faulkner [20] this number of test users is enough to reach nearly high level of certainty for finding the most of the existing usability problems.We asked the test-users to participate in a controlled experiment - to find a relevant person to contact or a conference to attend. They were asked to execute specific assignments and afterward to fill in a questionnaire with qualitative questions about their experience during the test. Furthermore we conducted a A/B testing survey among the users to measure the impact of the personalization and the EiCE features.

[29] http://philogb.github.io/jit/

Controlled Experiment. During the controlled experiment, users were asked to think aloud and their actions were recorded while an evaluator observed the comments and took notes. Each test took about 30 to 45 minutes. Their assignment was as follows:

Assignment The users had to mark all found resources relevant to them. Then, users could choose between three actions: searching, adding top related resources; this is done through disambiguated keyword based search on topics knowingly related to the initial search term e.g. choosing Tim Berners-Lee as initial keyword and WWW 2013 next related keyword in search, or expanding neighbors of found resources. In the last case they could chose between direct or indirect neighbors of the centrally focused node in the visualization. A 'top related' resource is the resource directly linked to the node in focus (centered) that shares the most common links with it.

Effectiveness measures how often a displayed result (R) related to a resource was marked relevant by the user (M).

$$\textbf{effectiveness} = E = \frac{\mid M \cap R \mid}{\mid R \mid} \tag{1}$$

Each action that delivered new resources to the result set resulted in an increase of quality of the result set.

Productivity measures this increase. The quality of a result set is the number of marked relevant resources compared to the total number of visualized resources. Productivity P_r measures the increase of effectiveness E_k after each test-user set of search actions in $A = \{a_1, ..., a_k, ...\}$:

$$\textbf{productivity} = P_r = \sum_{k \in A} \frac{E_k - E_{k-1}}{\mid A \mid} \tag{2}$$

where E_k is measured effectiveness after the action a_k.

The *E/P*$_r$ *Ratio* indicates the impact of the newly visualized nodes on the existing visualized resources. We get this ratio by dividing productivity by effectiveness. We learn from this ratio how beneficial the last action was to the search as it indicates the percentage of the relevant nodes contributed by the last search action.

$$\textbf{E/P}_r = \frac{P_r}{E_k} \tag{3}$$

The data in Figure 9 shows that adding a top related resource was not done often by the users and added only a couple of resources to the result set. However, it proved to be the most effective action as the users marked $^{13}/_{26}$ (50%) of the visualized resources relevant. The data in Figure 9 also shows an increase of 12% in productivity in based on an average over all test users. Adding top related resources resulted in a result set that contained 12% more relevant nodes as before adding top related nodes. The E/Pr ratio is 24% ($^{0.12}/_{0.5}$).

As Figure 10 indicates, searching for a resource was the most productive of all type of actions (+25%) and has an E/Pr ratio of 81% ($^{0.25}/_{0.31}$). This is remarkable as the user action effectiveness of searching is much lower than adding a top related resource on average over all test users. Adding top related resources resulted in a result set that contained +12% more relevant nodes as before adding top related nodes, even though it has higher effectiveness (50%). This means that the impact of each added resource when

Fig. 9. Overview of user actions used to visualize new resources to measure the effectiveness of the actions

searching is much bigger, because the quality of the result set was not relatively high at the moment users decided searching. On average less than 31% of the resources, which would result in an increase in productivity if of the newly added resources at least 31% was marked relevant according to users.

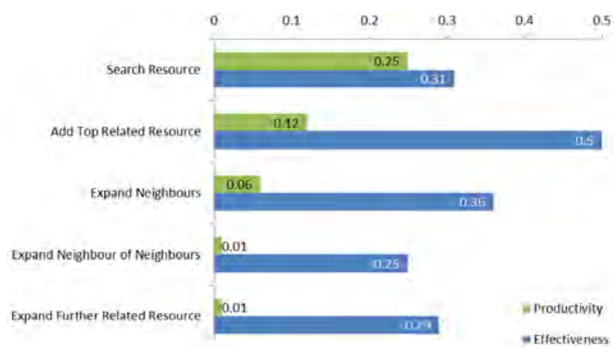

Fig. 10. User Action Effectiveness and Productivity results

The effectiveness of expanding resources $^{53}/_{166}$ (32%) is about the same as searching for a resource $^{54}/_{174}$ (31%). As the user actions resulted in about as many new resources in the case of searching and expanding, this is a very reliable comparison. Expanding the direct neighbors is the most productive (+6%) expansion. Expanding further related neighbours retains the quality of the result set and barely impacts it, but the productivity is still positive (+1%). Users mostly expanded direct neighbours, which leaded to 94 new resources compared to 72 expansions of indirect neighbours.

Feature Impact Survey. We conducted a survey among the users to measure the impact of the two most important features of ResXplorer: *personalization* (using social media data) and *pathfinding* (with EiCE). We presented the users screenshots of result sets in ResXplorer in A/B pairs, achieved with one of the features enabled or with both disabled/enabled. They were asked to rate on a Likert scale from -3 to $+3$, thus from more towards A to more towards B, which result set they preferred without knowing which one had which feature(s) enabled.

Fig. 11. Impact on the result set relevancy of ResXplorer features according to users.

Figure 11 shows disagreement or no clear positive impact for simple queries when EiCE is enabled and a rather negative impact when personalization is enabled for simple queries. The results are more positive where more than 60% of the users agrees that for complex queries the results when using the EiCE are preferred. For personalization the ratio is 45% positive against 36% negative, the bias is less positive here, but clearly better than the case for personalization with simple queries. When looking at enabling both features vs. disabling both features, nearly 66% prefers the results with both personalization en EiCE enabled and 56% in case of the simple queries.

The results are in line with previous studies we did on: (i) the dynamic alignment of social data with conference publication data [15]; and (ii) the usability study of the "Researcher Affinity Browser" [14]. All these findings back the emphasis at several places in the paper on the significance of pathfinding and social media (personalization) in interactive exploratory search.

5.4. Expert User Reviews

As already previously applied [27] for exploratory search we used a task based approach to obtain expert user reviews. The goal of the reviews is to compare ResXplorer against industry reference academic search interfaces and related academic projects, the state-of-the-art (SOTA). Two researchers – search interface experts – independently reviewed the performance of each of these search interfaces. They were familiar with all of the tools beforehand. We selected a set of six representative tasks supported by these systems for the reviews in Table 1.

We designed the search tasks optimized for the SOTA search engines and for ResXplorer and they are either simple (e.g. single fact or source) or complex (combinations of facts and sources). We outlined the a priori, thus before presenting it to the expert users, expected suitability of these tasks in Table 2.

In each of these tasks the experts had to indicate after each interaction by either a click or text input, how many relevant results they found. Their actions were recorded so that we could count the total number of actions for each task and the number of results after each action.

Table 1. List of tasks executed by the expert users.

Task	Description
T_1	Find proof that Chris(tian) Bizer is an author.
T_2	Find out three different people that know or are known by the person in T1 (e.g. co-authors).
T_3	Find out three different kinds of relations between the person in T1 and Chris(tian) Bizer.
T_4	Find three different conferences on the subject Artificial Intelligence.
T_5	Find at least two people that have a paper included in the proceedings in two consequent editions of the WWW (World Wide Web) Conference.
T_6	Find: (i) at least one publication that was presented in 2011 in a WWW workshop (co-)organized by Tim Berners-Lee (e.g. LDOW - Linked Data on the Web); and (ii) at least one publication with an author that relates this publication to both the '2011 publication and the ISWC Conference 2010.

Table 2. *A priori* optimal suitability of the search tasks.

	Straightforward	Complex
ResXplorer	T_3	T_6
Both	T_2, T_4	
SOTA	T_1	T_5

For each of the tasks we measured the *average precision* (between 0 and 1) and the *efficiency* (expressed as number of actions needed).

Average Precision measures the average of the search precision over all the required actions in certain task. Thereby the precision [36] corresponds in this case to the effectiveness of the kth search action as defined for the user evaluation:

$$\textbf{precision} = P_k = E_k \tag{4}$$

and the average precision over all actions A in certain task:

$$\textbf{average precision} = AP = \sum_{k \in A} \frac{P_k}{\mid A \mid} \tag{5}$$

However, the actions are different so a direct comparison for ResXplorer between the user action effectiveness and the precision measured here is not possible. It also would make no sense as the user tests focused on lean users while the experts are specialized in search interfaces.

Efficiency, expressed as the number of actions (N_x) when users perform a certain task (T_x). The lower the score, the less actions the experts needed to successfully complete the task.

To verify that the expert reviews are similar enough to be considered, we measured the inter-rater agreement among them. We selected therefore the *chance corrected agreement* (κ) measure [24] ($-1 < \kappa < 1$). The inter-rater agreement of the results between the experts is substantial ($\kappa = 0.61$ and F-measure 0.83) according to the Landis et al. scale [28]. The visualization in Figure 12 shows the mean average results for each of the tested search interfaces and indicates how well the expert reviews match.

Tables 3 and 4 display the results of the expert evaluations of ResXplorer in comparison to two industry references and three research projects in the same domain. In ARNet Miner and ResXplorer the autocomplete facilitated instant and precise matches. In Microsoft Academic Search, Google Scholar and Falcons the first page of results contained

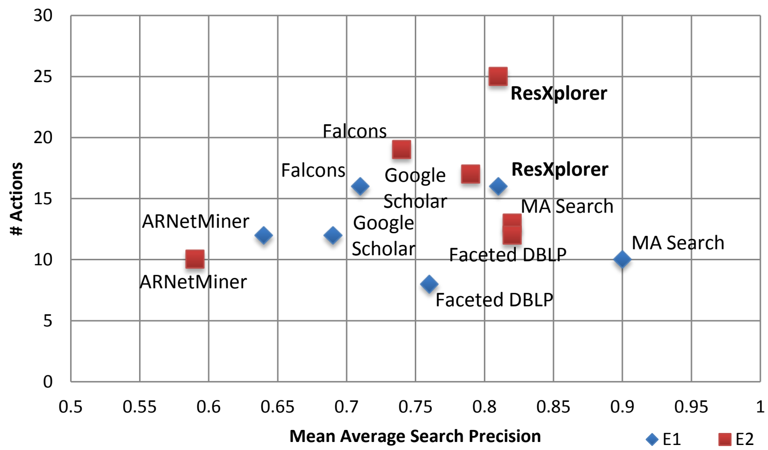

Fig. 12. The agreement between the experts on the ratings over all search interfaces combined is substantial.

Table 3. The search precision for getting the first search results returns all true positive matches except ArnetMiner returned 4 out of 5 false positives in T_1. ResXplorer is not as precise as the other interfaces for T_2 but excels in T_3. *(brighter = better)*

Effectiveness	T_1	T_2	T_3	T_4	T_5	T_6	Mean
Google Scholar	1.00	0.90	0.35	1.00	0.43	0.62	0.72
MA Search	1.00	1.00	0.63	1.00	0.90	0.64	0.86
Falcons	1.00	0.95	0.63	0.78	0.60	0.68	0.77
ResXplorer	1.00	0.84	0.84	0.70	0.39	0.80	0.76
ARNetMiner	0.60	1.00	0.81	0.74	0.20	0.49	0.64
Faceted DBLP	1.00	1.00	0.83	0.95	0.52	0.45	0.79

Table 4. An increased number of user actions does not always guarantee more precise (intermediate) results, but it does for ResXplorer, except in T_5. *(brighter = better)*

Efficiency	T_1	T_2	T_3	T_4	T_5	T_6	Sum
Google Scholar	1	1	2	2	5	3	15
MA Search	1	1	3	1	2	4	12
Falcons	1	2	3	3	3	6	18
ResXplorer	1	2	4	3	4	4	21
ARNetMiner	1	1	3	1	2	3	11
Faceted DBLP	1	1	3	1	2	3	10

the necessary results and Google Scholar and Microsoft Academic Search promoted the matching result as a suggestion on top of the list.

T3 is a non-direct relation finding task and that is the main goals of ResXplorer while T2 requires zooming in depth around a specific property of a person. ResXplorer intends to maintain the broad overview at all times during the search which induces some noise for task like T2.

In T4 the industry references beat the research engines. T4 requires skimming or filtering a list of conferences which is not supported in ResXplorer and in Falcons and

ArnetMiner not to the same degree as the industry references. Faceted DBLP also scores well for T4 thanks to the faceted search interface and tight DBLP link. For T4 required the Google Scholar interface scrolling through two pages to find three different conferences, many results of the same conference. Microsoft Academic Research allowed searching specifically for items of the type conference. That explains the highest rating here, all results were on the first page in contrast to Google Scholar. in Falcons the results were a little less accurate and did not allow searching specifically for conferences either. ResXplorer did, but as it did not provide a list but a limited set of entry points for exploration. This meant the search was repeated to find different entry points leading to a conference, in fact three times, each time to find a new conference. ARNet Miner provided a result view, containing distracting of widgets, not all material was relevant for the search. It included relatively many false positives to interpret but all results were found after one search action. The expert users judge the results presented in the *a priori* defined complex tasks having the most irrelevant results and they needed at least 2 actions in T5 and even 3 actions in T6 to resolve the search task. The highest effectiveness was found for MA Search in T5 and for ResXplorer in T6. In terms of efficiency Google Scholar required the most actions and in T5 and Falcons in T6.

6. Discussion

We observed that searching by keywords for resources increases the result set with the most new relevant resources, while it is on average as effective as expanding existing resources in the result set. The most effective user action was adding top-related nodes to the visualization.

ResXplorer is situated in the mid-range in terms of mean average search precision and requires relatively lots of action from the user. However, ResXplorer is best when the task consisted of relating resources that are not directly related or when at least the user is not aware of how they are related. That is precisely the goal we wanted to show with ResXplorer and the methods and techniques that drive it. Furthermore, this pinpoints, once again, to the importance and the need of user-centered evaluation concept within conducted measurements. The main concept of ResXplorer resides on the idea of an interactive search interface which leads the researcher through the process of expansion and exploration of results to the hidden implicit valuable information discoveries which are uncovered in such a process. The balanced choice of comparable solutions: two of them from industry (MA Search and Google Scholar) and three of them from research domain (ARNet Miner, Falcons and Faceted DBLP); this allows good positioning and qualitative reviewing of our solution.

Having visually more advanced solutions like MA Search and ARNet Miner and those with less search interface interactivity possibilities like Google Scholar, Faceted DBLP and Falcons we also want to cover the essential aspect in evaluation for user driven search applications which considers the visual representation and analysis of search results and interaction possibilities on search interface . In order to outline the differences between conventional search interfaces for scientific resources and our approach, we used a set of "Visual representation and analytics" based on guidelines identified by [8].

We compared the features of the search interfaces used in the expert evaluation as listed in table 5. We notice that industry references as MA Search and Google Scholar lack the interactivity with a visual representation, although MA Search for instance offers visual interfaces to the search results. On the other hand, ARNet Miner supports various visualizations based on data mining algorithms, like e.g. clustering, executed on the retrieved data in combination with the search results. Faceted DBLP features an interactive, all-round faceted search interface. Falcons Object Search [7] is considered as a keyword-based search engine for linked objects with extensive virtual documents indexed. Those documents consist from associated literals but also from the textual descriptions of associated links and linked objects. The results are ranked according to a combination of their relevance to the query and their popularity. Besides a classical list representation Falcons allows enhanced text based browsing of Linked Data as well filtering on concepts and relations.

Table 5. Comparision of functionality of different search interfaces for research.

Usability Criterion	ResXplorer	MA Search	GScholar	ARnet Miner	Falcons	Faceted DBLP
Query (forms / keyword)	●	●	●	●	●	●
Query (formal syntax)	◐	●	●	●	●	◐
View results as ordered list	○	●	●	●	●	●
Visual presentation	●	●	○	●	○	○
Interactively refine search	●	○	○	◐	●	◐
Combine and relate searches	●	○	○	○	○	○
Data overview	●	●	○	●	●	●
Detail on demand	●	●	●	●	●	●
Generic / Engine Reusable	●	?	?	◐	?	?
Support for scalability	●	●	●	●	●	●
Filtering	◐	●	●	●	●	●
History	●	◐	◐	○	○	○
View original source	●	●	●	◐	●	●
Feature coverage	●= full ◐= partially ○= none ? = uncertain					

With our search engine users can combine any searches and interact the results that exposes relationships between them. This is a feature not found in conventional search interfaces. It offers search for publications, as well as supports relation visualization on author level. We visually emphasize discovered types of entities and relations. In comparison to the current existing solutions we can use the snapshot of social content published by researchers on social media and collaborative platforms like Twitter and Mendeley to make a pre-set for exploratory search. This feature is unique to our solution. Furthermore, the method by which we generate context-based results differs from ARnet Miner because we do not rely on data mining and machine learning techniques to resolve the research related information. Our approach uses affinity based ranking derived from the social context and search process itself. We use graph based algorithms which perform independently of underlying Linked Data. In comparison to the existing search solutions, our interface is designed to visually explore the research space, rather than to support classical keyword based search. This exploration is based on personal preference and serendipity of information in the data set (publications, persons, events). This data is enhanced by additional information (e.g. venues of events) related to the search. Unlike Microsoft Academic Search and ARnet Miner our graph visualization is expandable and includes entities from Linked Data and description of relations between them. Since presets of the search reside on actualized social media content of the user our solution adapts better on changes of information and trends from social media. This aspect differs strongly from the conventional approaches mentioned here.

7. Conclusions and Future Work

We have presented a semantic model for searching resources in the Web of Data developed for scientific research. We have demonstrated and implemented the model with current state-of-the art technologies. The model uses research objects to represent semantically modeled research data. We explained how applying this model contributes to implementations for research use cases.

The result is a semantic search application providing both a technical demonstration and a visualization that could be applied in many other disciplines beyond Research 2.0. The main contribution of our work is, besides retrieving resources from Linked Data repositories, allowing researches to interactively explore relationships between the resources and entities like events or persons related to their work.

Compared to existing well established search interfaces with similar goals and the same target audience, ResXplorer is situated in the mid-range when compared to both industry and academic projects targeting the same use case, but requires more interaction of the users. However when a task consists of finding resources that connect a given statement, such as finding common items between two authors of an article, ResXplorer has relatively the highest search precision. Considering that the implementation of ResXplorer is still in the prototype phase, the potential of a visual and interactive search interface is well demonstrated and understood by the target users.

We will clearly distinguish between proposing new affinities between certain resources versus exploring the proposed resources in detail and explaining the motivation behind the affinities, where we characterize each affinity, between researchers and resources, by the amount of shared interests and other commonalities. To make ResXplorer more precise in classical search and retrieve scenarios, more accurate filters on the search keywords and results are crucial. We will analyze the efficiency further as a smaller number of actions does not always lead to the most efficient interface, certainly if it requires more thinking and judging from the users: more straightforward steps might be more efficient than less but more complicated steps. In this work we have focused on proposing novel affinities related to a personalized search context, but it is important to allow researchers to explore the proposed affinities more in depth. Further improvements on the ranking criteria should improve the precision of proposed affinities and the results even further.

Acknowledgments. The research activities described in this paper were funded by Ghent University, imec (former iMinds, Flemish interdisciplinary institute for technology), Graz University of Technology, the Department of Economy Science and Innovation Flanders (EWI), the Institute for Innovation and Entrepreneurship in Flanders (VLAIO), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

References

1. Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D., Owen, S., Newman, D., Sufi, S., Goble, C.: Why linked data is not enough for scientists. Future Generation Computer Systems 29(2), 599 – 611 (2013)
2. Breslin, J.G., Decker, S., Harth, A., Bojars, U.: SIOC: an approach to connect web-based communities. International Journal of Web Based Communities (IJWBC) 2(2), 133–142 (2006)
3. Breslin, J.G., Harth, A., Bojars, U., Decker, S.: Towards semantically-interlinked online communities. In: Gomez-Perez, A., Euzenat, J. (eds.) European Semantic Web Conference (ESWC). Lecture Notes on Computer Science, vol. 3532, pp. 500–514. Springer (2005)
4. Brusilovsky, P.: Methods and techniques of adaptive hypermedia. In: Adaptive hypertext and hypermedia, pp. 1–43. Springer (1998)
5. Brusilovsky, P., Peylo, C.: Adaptive and intelligent web-based educational systems. International Journal of Artificial Intelligence in Education 13(2), 159–172 (2003)
6. Cheng, G., Qu, Y.: Searching linked objects with falcons: Approach, implementation and evaluation. International Journal on Semantic Web and Information Systems (IJSWIS) 5(3), 49–70 (2009)
7. Cheng, G., Qu, Y.: Searching linked objects with falcons: Approach, implementation and evaluation. Int. J. Semantic Web Inf. Syst. 5(3), 49–70 (2009)
8. Dadzie, A.S., Rowe, M.: Approaches to visualising Linked Data: A survey. Semant. web 2(2), 89–124 (Apr 2011)
9. d'Aquin, M., Baldassarre, C., Gridinoc, L., Angeletou, S., Sabou, M., Motta, E.: Characterizing knowledge on the semantic web with Watson. In: Garcia-Castro, R., Vrandecic, D., Gmez-Prez, A., Sure, Y., Huang, Z. (eds.) EON. CEUR Workshop Proceedings, vol. 329, pp. 1–10. CEUR-WS.org (2007)
10. De Vocht, L., Coppens, S., Verborgh, R., Vander 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 (LDOW2013) (2013)
11. De Vocht, L., Deursen, D.V., Mannens, E., de Walle, R.V.: A semantic approach to cross-disciplinary research collaboration. Internation Journal of Emerging Technologies in Learning (iJET) 7(S2), 22–30 (2012)
12. De Vocht, L., Mannens, E., Van de Walle, R., Softic, S., Ebner, M.: A search interface for researchers to explore affinities in a Linked Data knowledge base. In: Proceedings of the 12th International Semantic Web Conference Posters & Demonstrations Track. pp. 21–24. CEUR-WS (2013)
13. De Vocht, L., Softic, S., Dimou, A., Verborgh, R., Mannens, E., Ebner, M., Van de Walle, R.: Visualizing collaborations and online social interactions at scientific conferences for scholarly networking. In: Proceedings of the Workshop on Semantics, Analytics, Visualisation: Enhancing Scholarly Data (SAVE-SD 15); 24th International World Wide Web Conference (2015)
14. De Vocht, L., 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. pp. 43:1–43:9. i-KNOW '11, ACM (2011)
15. De Vocht, L., Softic, S., Mannens, E., Ebner, M., Van de Walle, R.: Aligning web collaboration tools with research data for scholars. In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion. pp. 1203–1208. WWW Companion '14, International World Wide Web Conferences Steering Committee (2014)
16. De Vocht, L., Softic, S., Mannens, E., Van de Walle, R., Ebner, M.: Resxplorer : interactive search for relationships in research repositories. In: International Semantic Web Conference : Semantic Web Challenge, Abstracts. p. 8. Trentino, Italy (2013)
17. Dimou, A., De Vocht, L., Van Compernolle, M., Mannens, E., Mechant, P., Van de Walle, R.: A visual workflow to explore the web of data for scholars. In: Proceedings of the Companion

Publication of the 23rd International Conference on World Wide Web Companion. pp. 1171–1176. WWW Companion '14, ACM (2014)
18. Ding, L., Kleb, J., Mueller, W.: How the semantic web is being used: An analysis of FOAF documents. In: Proceedings of the 38th International Conference on System Sciences (2005)
19. Ebner, M., Reinhardt, W.: Social networking in scientific conferences - Twitter as tool for strengthen a scientific community. In: Cress, U., Dimitrova, V., Specht, M. (eds.) Learning in the Synergy of Multiple Disciplines, Proceedings of the EC-TEL 2009. Lecture Notes in Computer Science, vol. 5794. Springer, Berlin/Heidelberg (October 2009)
20. Faulkner, L.: Beyond the Five-User assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments, & Computers 35(3) (2003)
21. Glaser, H., Millard, I.C., Jaffri, A.: RKBExplorer.com: a knowledge driven infrastructure for linked data providers. In: Proceedings of the 5th European semantic web conference on The semantic web: research and applications. pp. 797–801. ESWC'08, Springer-Verlag, Berlin, Heidelberg (2008)
22. Graves, A.: Creation of visualizations based on Linked Data. In: Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics. p. 41. ACM (2013)
23. Hogan, A., Harth, A., Umbrich, J., Kinsella, S., Polleres, A., Decker, S.: Searching and browsing linked data with swse: The semantic web search engine. Web Semantics: Science, Services and Agents on the World Wide Web 9(4), 365 – 401 (2011)
24. Hripcsak, G., Rothschild, A.S.: Agreement, the F-measure, and reliability in information retrieval. Journal of the American Medical Informatics Association 12(3), 296–298 (2005)
25. Hristoskova, A., Tsiporkova, E., Tourw, T., Buelens, S., Putman, M., De Turck, F.: Identifying experts through a framework for knowledge extraction from public online sources. In: 12th Dutch-Belgian Information Retrieval Workshop (DIR 2012), Ghent, Belgium. pp. 19–22 (2012)
26. Jiang, M., Cui, P., Wang, F., Xu, X., Zhu, W., Yang, S.: Fema: Flexible evolutionary multi-faceted analysis for dynamic behavioral pattern discovery. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1186–1195. KDD '14, ACM, New York, NY, USA (2014)
27. Kraaij, W., Post, W.: Task based evaluation of exploratory search systems. In: SIGIR 2006 workshop, Evaluating Exploratory Search Systems (2006)
28. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. biometrics pp. 159–174 (1977)
29. Ley, M.: The DBLP computer science bibliography: Evolution, research issues, perspectives. In: String Processing and Information Retrieval. pp. 1–10. Springer (2002)
30. Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Stenzhorn, H., Tummarello, G.: Sindice.com: a document-oriented lookup index for open linked data. Int. J. of Metadata and Semantics and Ontologies 3, 37–52 (2008)
31. Pace, S.: A grounded theory of the flow experiences of web users. International journal of human-computer studies 60(3), 327–363 (2004)
32. Parra Chico, G., Duval, E.: Filling the gaps to know More! about a researcher. In: Proceedings of the 2nd International Workshop on Research 2.0. At the 5th European Conference on Technology Enhanced Learning: Sustaining TEL,. pp. 18–22. CEUR-WS (Sep 2010)
33. Passant, A., Bojars, U., Breslin, J.G., Decker, S.: The SIOC project: Semantically-interlinked online communities, from humans to machines. In: Padget, J.A., Artikis, A., Vasconcelos, W., Stathis, K., da Silva, V.T., Matson, E.T., Polleres, A. (eds.) Coordination, Organizations, Institutions, and Norms in Agent Systems V. Lecture Notes in Computer Science, vol. 6069, pp. 179–194. Springer (2009)
34. Passant, A., Bojars, U., Breslin, J.G., Hastrup, T., Stankovic, M., Laublet, P.: An overview of SMOB 2: Open, semantic and distributed microblogging. In: Cohen, W.W., Gosling, S. (eds.) ICWSM. pp. 303–306. The AAAI Press (2010)
35. Passant, A., Breslin, J.G., Decker, S.: Open, distributed and semantic microblogging with SMOB. In: Benatallah, B., Casati, F., Kappel, G., Rossi, G. (eds.) ICWE. Lecture Notes in Computer Science, vol. 6189, pp. 494–497. Springer (2010)

36. Powers, D.: Evaluation: From precision, recall and F-measure to ROC., informedness, markedness & correlation. Journal of Machine Learning Technologies 2(1), 37–63 (2011)
37. Softic, S., De Vocht, L., Mannens, E., Van de Walle, R., Ebner, M.: Finding and exploring commonalities between researchers using the resxplorer. In: Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration - First International Conference, LCT 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II. pp. 486–494 (2014)
38. Softic, S., Ebner, M., Mühlburger, H., Altmann, T., Taraghi, B.: Twitter mining #microblogs using #semantic technologies. 6th Workshop on Semantic Web Applications and Perspectives pp. 1–9 (2010)
39. Softic, S., Vocht, L.D., Mannens, E., Ebner, M., de Walle, R.V.: COLINDA: modeling, representing and using scientific events in the web of data. In: Proceedings of the 4th International Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2015) Co-located with the 12th Extended Semantic Web Conference (ESWC 2015), Protoroz, Slovenia, May 31, 2015. pp. 12–23 (2015)
40. Softic, S., Vocht, L.D., Mannens, E., de Walle, R.V., Ebner, M.: Finding and exploring commonalities between researchers using the resxplorer. In: Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration - First International Conference, LCT 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II. pp. 486–494 (2014)
41. Sure, Y., Bloehdorn, S., Haase, P., Hartmann, J., Oberle, D.: The SWRC ontology - semantic web for research communities. Progress in Artificial Intelligence pp. 218–231 (2005)
42. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: Arnetminer: extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 990–998. ACM (2008)
43. Tao, K., Abel, F., Gao, Q., Houben, G.J.: Tums: Twitter-based user modeling service. In: Garcia-Castro, R., Fensel, D., Antoniou, G. (eds.) ESWC Workshops. Lecture Notes in Computer Science, vol. 7117, pp. 269–283. Springer (2011)
44. Tran, T., Herzig, D.M., Ladwig, G.: Semsearchpro - using semantics throughout the search process. Web Semantics: Science, Services and Agents on the World Wide Web. 9(4), 349–364 (2011)
45. Ullmann, T.D., Wild, F., Scott, P., Duval, E., Vandeputte, B., Parra Chico, G.A., Reinhardt, W., Heinze, N., Kraker, P., Fessl, A., Lindstaedt, S., Nagel, T., Gillet, D.: Components of a research 2.0 infrastructure. In: Lecture Notes in Computer Science,. pp. 590–595. Springer (2010)
46. Uren, V., Lei, Y., Lopez, V., Liu, H., Motta, E., Giordanino, M.: The usability of semantic search tools: A review. The Knowledge Engineering Review. 22(4), 361–377 (Dec 2007)
47. Uren, V., Sabou, M., Motta, E., Fernandez, M., Lopez, V., Lei, Y.: Reflections on five years of evaluating semantic search systems. International Journal of Metadata, Semantics and Ontologies (IJMSO) 5(2), 87–98 (2010)
48. Van Noorden, R.: Online collaboration: Scientists and the social network. Nature News 512(7513), 126–130 (Aug 2014)
49. Weibel, S.: The Dublin Core: a simple content description model for electronic resources. Bulletin of the American Society for Information Science and Technology 24(1), 9–11 (1997)
50. White, R.W., Marchionini, G.: Examining the effectiveness of real-time query expansion. Information Processing & Management 43(3), 685–704 (2007)
51. White, R.W., Marchionini, G., Muresan, G.: Evaluating exploratory search systems: Introduction to special topic issue of information processing and management. Inf. Process. Manage. 44(2), 433–436 (Nov 2008)
52. White, R.W., Muresan, G., Marchionini, G.: Evaluating exploratory search systems. EESS 2006 p. 1 (2006)
53. Xin, R.S., Hassanzadeh, O., Fritz, C., Sohrabi, S., Miller, R.J.: Publishing bibliographic data on the semantic web using BibBase. Semantic Web 4(1), 15–22 (2013)

Laurens De Vocht obtained a Master of Science in Computer Science Engineering at the KULeuven in 2011 and specialized in human-computer interaction. After being active as software developer in an enterprise specialized in custom search engine solutions, he joined Ghent University iMinds in 2012 as a researcher and continues to focus on search in - and interactions with Linked Data. Laurens developed expertise in the standardization and the formalization of data semantics. He has been mostly active in the application areas: e-Government, Tourism, Technology Enhanced Learning (Web 2.0 for Science). Futhermore he has experience in implementing semantically driven software solutions for the industry, where at the moment he is investigating best-practices for linking data and distributing search queries in the context of medicine development.

Selver Softic holds a degree of Master of Science in Computer Science from Graz University of Technology and is currently a PhD candidate at Department for Educational Technology / Institute for Information Systems and Computer Media at Graz University of Technology. During the last decade, Selver participated as Senior Researcher at various European and national research projects on topics of e-learning, Social Networks Analysis, Technology Enhanced Learning, Semantic Web and Linked Data and Research 2.0. His main research focus deals with Knowledge Discovery in sparse semi-structured text fragments using semantic data modellig for the purposes of research and education.

Dr. Ruben Verborgh is a researcher in semantic hypermedia at Ghent University iMinds, Belgium and a postdoctoral fellow of the Research Foundation Flanders. He explores the connection between Semantic Web technologies and the Web's architectural properties, with the ultimate goal of building more intelligent clients. Along the way, he became fascinated by Linked Data, REST/hypermedia, Web APIs, and related technologies. He's a co-author of two books on Linked Data, and has contributed to more than 150 publications for international conferences and journals on Web-related topics.

Prof. dr. ing. Erik Mannens is CTO Data Scienece @ iMinds and Research Manager at Ghent University - Data Science Lab since 2005. Before joining iMinds, he was a software engineering consultant and Java architect for over a decade. His major expertise is centred around metadata modelling, semantic web technologies, linked open data, big data analytics, broadcasting workflows, iDTV and web development in general. Since 2008 Erik is paving the Open Data path in Flanders. He stood at the cradle of the first Hackatons and is a founding member of OKFN - Belgian Chapter. Since then, he is frequently invited as Open Data evangelist at national and international events. Furthermore, he currently actively participates in several W3C standardisation working groups.

Assoc. Prof. Phd Martin Ebner is currently head of the Department for Educational Technology at Graz University of Technology and therefore responsible for all university wide e-learning activities. He holds an Assoc. Prof. on media informatics and works also at the Institute for Information System Computer Media as senior researcher. His research strongly focuses on e-learning, mobile learning, learning analytics, social media, and Open Educational Resources. Martin holds a number of lectures in this area as well as workshops and talks at international conferences. Find more detailed information at http://www.martinebner.at.

Received: October 28, 2015; Accepted: June 25, 2016.