Aligning Web Collaboration Tools with Research Data for Scholars

Laurens De Vocht[1]
laurens.devocht@ugent.be

Selver Softic[2]
selver.softic@tugraz.at

Erik Mannens[1]
erik.mannens@ugent.be

Martin Ebner[2]
martin.ebner@tugraz.at

Rik Van de Walle[1]
rik.vandewalle@ugent.be

[1]Ghent University - iMinds - Multimedia Lab
Ghent, Belgium

[2]Graz University of Technology
Graz, Austria

ABSTRACT

Resources for research are not always easy to explore, and rarely come with strong support for identifying, linking and selecting those that can be of interest to scholars. In this work we introduce a model that uses state-of-the-art semantic technologies to interlink structured research data and data from Web collaboration tools, social media and Linked Open Data. We use this model to build a platform that connects scholars, using their profiles as a starting point to explore novel and relevant content for their research. Scholars can easily adapt to evolving trends by synchronizing new social media accounts or collaboration tools and integrate then with new datasets. We evaluate our approach by a scenario of personalized exploration of research repositories where we analyze real world scholar profiles and compare them to a reference profile.

1. INTRODUCTION

Publication repositories and online journals all have search engines to help scholars find interesting resources. However, these approaches are often ineffective, mostly because scholars: (i) only look-up resources based, at best, on their topics or keywords, not taking into account the specific context and the scholar's profile; (ii) are restricted to resources from a single origin. Of course, aggregators exist that index resources from multiple sources. The challenge is therefore in matching research needs and contexts to opportunities from multiple, heterogeneous sources. In other words, we should make the most of the wealth of resources for research through relating and matching their scholar profile with the online available resources, publications and other scholar's profiles.

Usually scholars need a paid membership to get full access to journals' articles, the library 'paywall'. At the same time a growing number of "Open Journals" offer free online access to all their published works. Most prominent archives in this area are Directory of Open Access Journals[1] as well as Online Journals[2]. Many of these bibliographic archives provide APIs or are already published as Linked Data. Big national libraries are following this example. According to the Linked Open Data (LOD) Cloud stats[3] publication repositories are abundant[4]. Scholars have embraced Internet technologies in ways that broaden the scope of their research work beyond college walls and in ways reaching beyond data silos forced by libraries. Microblog platforms such as Twitter can be a useful way to expand their community even further by following others and sharing research interests.

We will describe in the following sections the model we developed for this purpose. We explain (i) which vocabularies used; (ii) the datasets selected for the implementation; (iii) our custom developed system for dynamic alignment of resources of social media, collaboration tools and selected datasets; and (iv) we evaluate the alignment and measure how well we can interlink conferences, publications and authors with scholar user profiles.

2. MODEL

We collect and use data from resources already explored by other 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 'outside-the-box' thinking (e.g., when formulating the exact search query requires background knowledge of a domain unfamiliar to the researcher). The model shows in Figure 1 how the researchers interact with the research data.

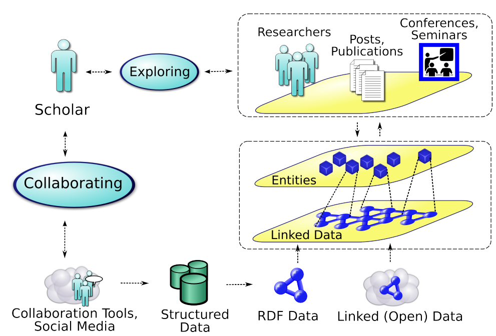

Figure 1: Interaction with research data.

[1]http://www.doaj.org/

[2]http://online-journals.org/

[3]http://stats.lod2.eu/

[4]http://lod-cloud.net/state/

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2579031.

Researchers can define and select their 'intended' search goal over several iterations. A combination various resources is then presented to the researchers. When scholars are looking for new information, they get an overview of possible objects of interest by having their activities and contributions on social media and other platforms such as their own research publications profiled. To achieve this, each scholar's profile sources needs to be aligned to the research data in the Linked Open Data cloud.

3. DATASETS

The datasets in our implementation, combine existing Linked Open Data sets: DBpedia[5], DBLP[6] and GeoNames[7] interlinked with research oriented datasets such as COLINDA[8] and a Social Linked Data set containing information about conferences and social profiles of the researchers from Twitter and Mendeley and the data they generated recently.

We used Twitter[9] data to profile scientists. Besides Twitter we used Mendeley[10], a popular example of a research publication and citation sharing tool, for linking with scientific resources. We used it to access the publications, tags and profile information of registered authors and link with the authors' social profiles. Table 1 highlights the statistics of the datasets (M = millions, G = gigabytes).

Dataset	Size (G)	#Triples/Rows	#Instances	#Literals
DBpedia	38	332M	27.1M	161M
DBLP (L3S)	12	95.2M	13.1M	17.5M
COLINDA	0.15	0.143M	0.016M	0.070M
Social LD*	0.06	0.041M	0.007M	0.015M

Table 1: Linked Data used within the search experiments.
* Average per user profile.

4. VOCABULARIES

We use common vocabularies to annotate social media as Linked Data provided by community and current research efforts in this area [9, 15, 13, 4]. We specifically applied: Friend of A Friend (FOAF)[11], Semantically Interlinked Online Communities (SIOC)[12], Semantic Web for Research Communities Ontology (SWRC)[13], and the Dublin Core[14].

FOAF describes the user profiles, their social relations and resources. We combined SIOC with FOAF and the Dublin Core for creating model instances of web entries like blogs, microblogs, mailing list entries and forum posts as well as other entries from collaboration tools [9, 2]. The SWRC [14] ontology was used to describe the academic resources and events with corresponding meta data in order to be compliant with research related Linked Data sets (COLINDA and DBLP). For tag binding we applied the Modular Unified Tagging Ontology (MUTO)[15] [8] as it combines the best approaches from earlier efforts on defining a tag ontology. MUTO instances bind hashtags from Twitter with entities in a user's context.

5. DYNAMIC RESOURCE ALIGNMENT

We combine three custom developed components for the alignment task: a *Profiler*, an *Interlinker* and an *Extractor*.

5.1 Profiler

The Profiler extracts the timeline and followers of the user's social account and then annotates them using the FOAF and SIOC vocabularies. We link their author's profile to DBLP based on publication title and the DOI (Digital Object Identifier) of each publication. Listing 1 shows how we combine these identifiers with all author names and use them to find matching author identifiers in DBLP for each publication. For each article in a Mendeley account linked to a subscribing researcher it checks the DOI and publication title in DBLP and retrieves the authors. If a match occurs, the articles are aligned using *owl:sameAs*. If all author names of the publication match, we interlink the Mendeley authors with the DBLP authors based on their URI's. Because users linked their Twitter and Mendeley when signing up, the profiler can link the author representation on DBLP with the author profile on Mendeley to the other social media accounts of the user and their contributions.

```
alignArticle(mendeleyArticle)
  title = find(mendeleyArticle, "dcterms:title")
  articleAuthors = aligner.getAuthors(title, article)
  foreach(articleAuthors -> (dblpArticle, authors))
    add(mendeleyArticle, "owl:sameAs", dblpArticle)
    foreach(authors -> (authorUri, authorName))
      add(articleUri, "dcterms:creator", authorUri)
      persons = find("foaf:name", authorName)
      foreach(persons -> person)
        add(person, "rdf:type", "foaf:Person")
        add(person, "owl:sameAs", uri)
```

Listing 1: Aligning research publications.

The original data from social media needed to generate user profiles reside mostly in-memory. After profiling, the original tweets are erased after at most seven days. The profiling and analysis results however are stored and indexed.

5.2 Interlinker

The Interlinker uses several steps to optimally align various sources. The first step is to define which Linked Data sets to use in which context, to identify the vocabularies in them and to define which resource to link with resources occurring in other datasets. When the dataset is not available as Linked Data, it selects a vocabulary to annotate the structured data directly. The case of social media is particularly interesting because social media content often consists of small posts and shares which we analyzed based on: URLs; hashtags and included mentions and tracking relations of the social accounts (e.g. followers).

After we have extracted the URLs, hashtags, entities and mentions from each post in social media, we link them to entities in the Linked Open Data Cloud. COLINDA is used for matching conference hashtags, GeoNames for locations, DBpedia for general concepts such as people, places as events. DBpedia is well-connected to GeoNames and DBLP which makes it a very valuable source for expanding the entities with more information about some common categories like cities and countries, people, or institutions.

[5] http://dbpedia.org
[6] http://dblp.l3s.de
[7] http://www.geonames.org/ontology/
[8] http://colinda.org
[9] http://www.twitter.com/
[10] http://www.mendeley.com/
[11] http://xmlns.com/foaf/spec/
[12] http://rdfs.org/sioc/spec/
[13] http://ontoware.org/swrc/
[14] http://dublincore.org/documents/dcmi-terms/
[15] http://muto.socialtagging.org/core

For each unique hashtag occurring in a microblogpost of a researcher: we firstly load the labels of the tag and load the interlink services from a configuration file; then we go over all the labels and services and look for a meaning for each of these labels on those services; finally for each found meaning we link the hashtag to the label using *muto:tagLabel* and to the URI of the meaning with *muto:tagMeans*.

This approach enriches tweets with Linked Data and is a way to achieve optimal meaning. Entities occurring in the resources shared via the tweets lead to the best results [1]. However, we have found in earlier research that the meanings of hashtags are consistent in many contexts [6], which makes us confident in using them for interlinking as well. The *rdfs:seeAlso* property connects the conferences from COLINDA with corresponding proceedings instances from DBLP Linked Data set. As usual in Linked Data community this property is used to link related and very relevant but not equivalent instances to the conferences. Linking in such manner . The *rdfs:label*, of each conference instance, matches the tags and hashtags from social media content and profiles of users. COLINDA instances also include *dcterms:spatial*[16] property for venues of conferences found in DBpedia. Conference web page links are generated with the *owl:sameAs* property. Data contained in COLINDA originates from WikiCfP[17] and Eventseer[18] and contains information about approximately 15000 conferences in the period from the year 2003 up to 2013.

Names (labels) of locations of conferences in COLINDA were used in CURL requests and SPARQL queries against DBpedia and Geonames to interlink these values from COLINDA over *dcterms:spatial*, *swrc: location* properties with the corresponding elements in DBpedia and Geonames instances. Included conference web page links were embedded into COLINDA instances using the *owl:sameAs* property.

5.3 Extractor

Each time when a certain source provides access to their structured content, the Extractor makes sure that the provided content is correctly converted conform our data model. Therefore it selects configured properties and maps them using the supported vocabularies and converts them to RDF.

6. EVALUATION

We are testing the aspects of our model and its implementation for making research data available through the interlinking of multiple data sources. The aligning of multiple data sources should improve the quality of the presented content. Test users noted this as an important criterion for improvement during earlier iterations [4]. Achieving this allows scholars a more refined and personalized access to heterogeneous sources for data they may find useful, one of the main challenges. To measure the quality of the linking we check three parameters: precision, sensitivity and accuracy of the linking applied to four types of resources: authors, friends, publications and hashtags. Each of these measures is a combination of *true positives* (TP), *false positives* (FP), *true negatives* (TN) and *false negatives* (FN) [10].

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

[16]http://purl.org/dc/terms/terms-spatial

[17]http://www.wikicfp.com

[18]http://eventseer.net

$$sensitivity = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{3}$$

We first describe the scenario to which the interlinked resources contributed, secondly we describe the user profiles part of this scenario and finally we present and discuss the measured results.

6.1 Scenario: Personalization

Users start by logging in with their Twitter accounts. Preferably scholars would authorize a social user account in which they often interact with the scientific community. After authorizing their accounts, users get access to a panel where they: manage configured repositories; browse a list to connect and disconnect account of available social media or collaboration tools and synchronize data from these accounts with the configured research data repositories.

They can then synchronize the latest social data (from Twitter) with the newest version of their public personal library (on Mendeley) and link it to the configured research data repositories. This synchronization happens client sides, after synchronization users can download their profiles' RDF which is automatically posted back to the server.

The goal of the scenario is to expose affinities, otherwise hidden proximities to or likings for specific resources, of the synchronized user profiles. It shows context-relevant relations for scholars based on common affinities using hashtags, mentions, people or conferences. The nature of our model enables the creation of personalized context as a starting point for further exploration of content made available through the interlinking process we described. The synchronization as pre-setup should enable scholars to explore content closely related to their interests more effectively if there is a sufficient number of accurate and precise links available.

6.2 User Profiles

Each used profile contains a Mendeley library and a Twitter feed. The libraries contain their bookmarked citations and publications, and the Twitter feed contains recent tweets that both they and the users they follow posted. We compared three different types of scholar profiles fitting the scenario:

1. An 'intense scholar profile' which uses all the tools efficiently and with a dense community of scholarly related people. We constructed this profile as 'Golden Profile' (*GP*). It is the only profile which we customly created for use a reference. The others are live profiles belonging to real users. It has a Mendeley containing publications only from the *Proceedings of the Linked Data on the Web Workshop* (2008-2012). The Twitter profile was created by adding the organizing committee of this workshop series and adding all Twitter recommended profiles to follow mentioning 'Linked Data' or 'Semantic Web' in their description.

2. Two 'typical scholar profiles' using these tools, but the Twitter account is not exclusively used for sharing academic resources for tweeting about conferences. One has a fairly large personal library (*UP1*) while the other has a small personal library (*UP2*). Both libraries contain a variety of publications, not all of these publications are indexed in DBLP.

3. A ‘basic profile’, making only use of Twitter, and this use is not limited to academic purposes either (*UP3*).

Characteristics for each of the profiles are listed in Table 2.

Characteristics	GP	UP1	UP2	UP3
Mendeley				
Articles in Personal Library	65	100	33	N/A
Twitter				
Following	30	245	258	N/A
Authors Following	21	35	140	N/A
Hashtags	21	26	18	22
Conference Hashtags	9	5	3	1

Table 2: An overview of the contents of each profile.

Listing 2 shows an interlinked article, person and tag. We see that the article’s authors are recognized in DBLP as well as the identifier of the article. An *owl:sameAs* connects the person representation with a link to the reference of the social account with the author profile. The example tag shown displays the *muto:tagMeans* property to link the conference hashtag with the URI of the conference on COLINDA.

6.3 Tags

For tag-entity linking we measured the accuracy by fraction conference tags and the sensitivity by the precision. In all cases it is clear that the GP delivers the best output (higher score is better). We also see in Figure 2 that UP1 has a slightly higher accuracy. UP1 also has the largest Mendeley library and used the most conference tags.

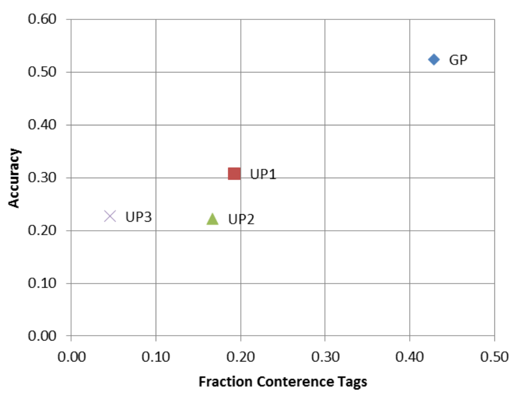

Figure 2: Accuracy by Fraction of tags which represent conferences: shows that higher fraction of conferences leads to better accuracy.

While all three UP’s have a much lower sensitivity than the GP, they have a considerably high precision, as Figure 3. The sensitivity for UP1 is better than UP2 for the same level of precision, this is due to the fact that UP1 has a slightly higher fraction of conference tags. Conference tags are better recognized than other tags, not surprising as the model is optimized for it.

6.4 Articles and Authors

When interlinking articles and authors, we considered the version of the article and author in the personal library of the user with the version available in DBLP. Obviously, except for the GP, not all publications are available in DBLP, so there are no TN in that case. In all these cases there are no

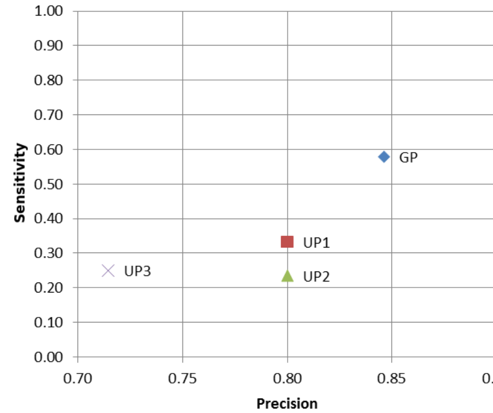

Figure 3: Sensitivity by precision: the precision and sensitivity of enity matching of the tags for the GP is as expected the highest.

FP, so precision is equal to 1. This is good and expected, as the links for articles and authors are based on the schema matching of the vocabularies rather than recurrences off the strings as is the case with the tags.

Figure 4 shows a relative high precision for authors in the UP1 and UP2 case compared to GP, spread is just above 20%. The spread with the article links is twice as high, GP’s library consisted of publications all in DBLP and was centered around the same community. UP1 and UP2 also have articles in their library not available in DBLP. Table 3

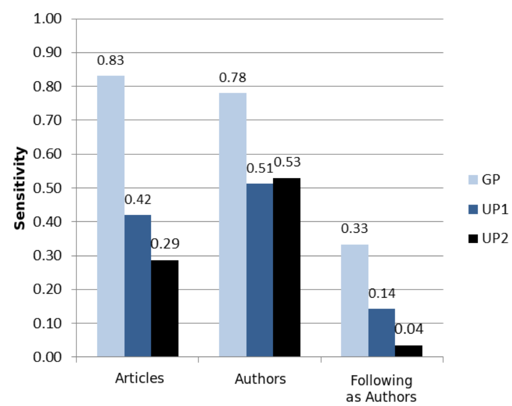

Figure 4: Sensitivity by type of resource linked. The sensitivity of linking entities for the GP is as expected the highest in all cases.

indicates that linking of followed users as authors has a bad sensitivity. This is because the personal library of the user which used to identify the link of the social profile of each other with their publications is limited by the scope of each user’s library. So it only contains a fraction of the available authors in DBLP. This is however normalized in the accuracy score, which takes into account the TN as well.

User	**Sensitivity**	**Accuracy**
GP	0.33	0.53
UP1	0.14	0.88
UP2	0.04	0.48

Table 3: Sensitivity and accuracy for linking followed users as authors: a high difference, especially for UP1 and UP2, because many of the followed users are not scholars or are unrelated.

```
<http://resxplorer.org/articles/6018551401> a swrc:Article ;
    rdfs:label "4th Linked Data on the Web Workshop ( LDOW2011 )" ;
    dc:creator <http://dblp.l3s.de/d2r/resource/authors/Christian_Bizer>,
        "Christian Bizer" ;
    dc:identifier "10.1145/1963192.1963323", "6018551401" ;
    dc:source <http://www.mendeley.com/c/6018551401/p/27542461/bizer-2011-4th-linked-data-on-the-web-workshop--
        ldow2011-/> ;
    owl:sameAs <http://dblp.l3s.de/d2r/resource/publications/conf/www/BizerHBH11> .

<http://resxplorer.org/people/timberners_lee> a foaf:Person ;
    rdfs:label "Tim Berners-Lee" ;
    dc:identifier "timberners_lee" ;
    owl:sameAs <http://dblp.l3s.de/d2r/resource/authors/Tim_Berners-Lee> ;
    foaf:account <http://resxplorer.org/accounts/timberners_lee> ;
    foaf:name "Tim Berners-Lee" .

<http://resxplorer.org/tags/www2010> a <http://rdfs.org/sioc/types#Tag> ;
    rdfs:label "www2010" ;
    dc:description "The World Wide Web Conference 2010, Raleigh, NC" ;
    muto:tagLabel "www2010" ;
    muto:tagMeans <http://colinda.org/resource/conference/WWW/2010> .
```

Listing 2: Excerpt from interlinked data of the GP

6.5 Discussion

We evaluated the aligned and interlinked scholar profiles and measured a relatively high accuracy when detecting conferences in tags and a promising sensitivity when interlinking articles and authors. It proves that the dynamic alignment of resources is useful for tools like ResXplorer, a radial graph interface for researchers[19]. Such tools make optimal use of the method and visualizes the aligned profiles and resources to allow exploration of the underlying research data.

The final interface supplied to users can give abundant and accurate information about scholars when the quality of the underlying alignment between datasets has a high accuracy and a minimum sensitivity. It does not necessarily have to be a high rate compared to the total number of resources as each single correct link builds a novel connection of interest to scholars.

7. RELATED WORK

Studies on the use of microblog platforms like Twitter within scientific communities [20] [5] have shown that researchers and scholars use Twitter to discuss and asynchronously communicate on topics during conferences and in their everyday work [11]. A survey of the use of Twitter for scientific purposes [7] showed that Twitter is not only a communication medium, but also a reliable source of data for scientific analysis and profiling tasks [13, 15]. In our earlier work on this subject, we built an interface [4] to allow scholars to browse their affinities such as interpersonal shared commonalities. Twitter users adopted hashtags to create threads of communication around a certain topic [6]. Hashtags can be suitable to link entities from microblogposts when combined with Linked Data [6, 16].

The efforts to make sharing scientific resources a reality occupied researchers in science and educational informational systems for a long time. The outcome of such quests lead to an increasing variety of heterogeneous technologies, schema, repositories and query mechanisms. This trend brings with it a constant growing amount of publicly available Linked Data including scientific repositories. Within the research community commercial digital libraries like Association for Computer Machinery) Digital Library[21] started to publish their archives in the LOD Cloud providing, in this special case, more than 12 million triples. Parallel to the commercial scientific content providers some academic institutions as well as the most famous public libraries, such as Library of Congress[22], British National Library[23] and Bibliothèque Nationale de France[24], provided their public Linked Data. Besides the initiative of big digital and national libraries, the efforts made by the scientific community like bootstrapping the eScience assets from the Open Archives Initiative - Object Reuse and Exchange (OAI-ORE) project [3] into the Web of Data are worth mentioning.

Currently only a limited number of works describe semantic modeling of data from social platforms. In [12] authors applied semantic modeling to different social platforms in common contexts and evaluated the potentials of reasoning on such an infrastructure. According to the authors even a small amount of data yields good results with simple reasoning and delivers very precise matches. Passant et al. improved mapping social profiles with related content, such as via interlinking the content tags [9]. Semantic modeling for Twitter data has been applied by [13] identifying hashtags as good resolvers for the retrieval of information and a solid interlinking base for the Linked Data Cloud. Similar use of semantic modeling of Twitter users was introduced on service level [15] and confirmed the benefits of this approach. These findings have been extended by the work on the "Researcher Affinity Browser" [4], as a prototype of Research 2.0 mash-ups based upon a personal semantic model from Twitter connected with the Linked Data set COLINDA, allowing researchers to find and identify colleagues with the same or similar affinities and to track scientific events they visited.

[19] http://www.resxplorer.org
[20] http://www.twitter.com
[21] http://acm.rkbexplorer.com/
[22] http://id.loc.gov
[23] http://bnb.data.bl.uk
[24] http://data.bnf.fr

8. CONCLUSIONS AND FUTURE WORK

We presented a new approach for dynamic alignment of research data from social media and collaboration tools with Linked Open Data for scholars and researchers. We were able to match resources from researchers based on their personal library and contributions on social media. This achievement is essential for the effective realization of a tool to facilitate the personalized exploration of heterogeneous data sources containing both research data and social data. Both providers of research data, through opening up their data to a broader audience, and scholars, through actively using collaboration tools and social media, will benefit.

Our preliminary results indicate sensitivity, precision and accuracy when linking tags, authors and articles to conferences. Conference tags are better recognized than other tags, this is not surprising because we optimized our model for this task. We have never obtained false positives when interlinking authors and articles. When we interlink followed users on Twitter as authors, we encountered a high amount of negatives. All found links of users as authors were correct but there is room for reducing false negatives.

Future research will focus on how to determine the efficiency of the model and a user evaluation of the environment involved. The environment needs enough incentives for the users to remain synchronized. We will also improve the accuracy of the interlinking by processing the contributed links that weren't immediately recognized. We want to interlink the user's personal library with the libraries of other users. This will allow links to be made to social and research data beyond a single user's scope. This should lead to more fine-grained details facilitating scholars to obtain a more sophisticated selection and linking of contributed resources based on previous assessments and explored links.

9. ACKNOWLEDGEMENT

The research activities described in this paper were funded by Ghent University, 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.

10. REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Semantic enrichment of Twitter posts for user profile construction on the Social Web. In *The Semanic Web: Research and Applications*, volume 6644 of *Lecture Notes in Computer Science*, pages 375–389. Springer Berlin Heidelberg, 2011.

[2] J. G. Breslin, S. Decker, A. Harth, and U. Bojars. SIOC: an approach to connect Web-based communities. *International Journal of Web Based Communities (IJWBC)*, 2(2):133–142, 2006.

[3] H. V. de Sompel, C. Lagoze, M. L. Nelson, S. Warner, R. Sanderson, and P. Johnston. Adding eScience assets to the Data Web. *CoRR*, abs/0906.2135, 2009.

[4] L. De Vocht, S. Softic, M. Ebner, and H. Mühlburger. 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 '11, pages 43:1–43:9, New York, NY, USA, 2011. ACM.

[5] M. Ebner, T. Altmann, and S. Softic. @twitter analysis of #edmedia10 - is the #informationstream usable for the #mass. *Form@re - Open Journal per la formazione in rete*, 11(74), 2011.

[6] D. Laniado and P. Mika. Making sense of Twitter. In P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, I. Horrocks, and B. Glimm, editors, *International Semantic Web Conference (1)*, volume 6496 of *Lecture Notes in Computer Science*, pages 470–485. Springer, 2010.

[7] J. Letierce, A. Passant, J. Breslin, and S. Decker. Understanding how Twitter is used to widely spread scientific messages. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, 2010.

[8] S. Lohmann, P. Díaz, and I. Aedo. MUTO: the modular unified tagging ontology. In C. Ghidini, A.-C. N. Ngomo, S. N. Lindstaedt, and T. Pellegrini, editors, *I-SEMANTICS*, ACM International Conference Proceeding Series, pages 95–104. ACM, 2011.

[9] A. Passant, J. G. Breslin, and S. Decker. Open, distributed and semantic microblogging with SMOB. In B. Benatallah, F. Casati, G. Kappel, and G. Rossi, editors, *ICWE*, volume 6189 of *Lecture Notes in Computer Science*, pages 494–497. Springer, 2010.

[10] D. Powers. Evaluation: From precision, recall and F-measure to ROC., informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

[11] W. Reinhardt, M. Ebner, G. Beham, and C. Costa. How people are using Twitter during conferences. In *Hornung-Prähauser, V., Luckmann, M.(Hg.): 5th EduMedia conference, Salzburg*, pages 145–156, 2009.

[12] M. Rowe. Interlinking distributed social graphs. In C. Bizer, T. Heath, T. Berners-Lee, and K. Idehen, editors, *Proceedings of the 6th International Workshop on Linked Data on the Web*, volume 538 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2009.

[13] S. Softic, M. Ebner, H. Mühlburger, T. Altmann, and B. Taraghi. @twitter mining# microblogs using# semantic technologies. *6th Workshop on Semantic Web Applications and Perspectives*, pages 1–9, 2010.

[14] Y. Sure, S. Bloehdorn, P. Haase, J. Hartmann, and D. Oberle. The SWRC ontology - Semantic Web for Research Communities. *Progress in Artificial Intelligence*, pages 218–231, 2005.

[15] K. Tao, F. Abel, Q. Gao, and G.-J. Houben. TUMS: Twitter-based user modeling service. In R. Garcia-Castro, D. Fensel, and G. Antoniou, editors, *ESWC Workshops*, volume 7117 of *Lecture Notes in Computer Science*, pages 269–283. Springer, 2011.

[16] P. Thonhauser, S. Softic, and M. Ebner. Thought Bubbles: a conceptual prototype for a Twitter based recommender system for research 2.0. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, i-KNOW '12, pages 32:1–32:4, New York, NY, USA, 2012. ACM.