

Discovery and Construction of Authors' Profile from Linked Data (A case study for Open Digital Journal)

Atif Latif¹, Muhammad Tanvir Afzal², Denis Helic¹, Klaus Tochtermann¹, Hermann Maurer²

¹Institute for Knowledge Management, Know-Center
Graz University of Technology, Austria

atif.latif@student.tugraz.at, {klaus.tochtermann, dhelic} @tugraz.at

²Institute for Information Systems and Computer Media
Graz University of Technology, Austria
{mafzal, hmaurer} @iicm.edu

ABSTRACT

The Open access digital journals, motivated from open access movement, now play a vital role in the dissemination of scientific literature and author's information over the web. In digital journals' environment, a well-linked collection of electronic resources is of great importance especially in creating opportunities for collaborations between organization, institutions, and persons. Finding information about authors (authors' profiles) in a digital journal's environment is crucial to increase the overall productivity and unprecedented success. Inspired from Linked Open Data (LOD) initiative, we have developed a tool which can establish links between authors of digital journals with relevant semantic resources available in LOD. The proposed system is able to disambiguate authors and can: 1) locate, 2) retrieve, and 3) structure the relevant semantic resources. Furthermore, the system constructs comprehensive aspect oriented authors' profiles from heterogeneous datasets of LOD on the fly. This paper investigates the potentials of such an approach on a digital journal known as Journal of Universal Computer Science (J.UCS). It is our strong belief that this kind of applications can motivate researchers and developers to investigate different application areas where Linked Open Data can contribute, bring added value, and can take the idea of open access further.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: *Information filtering*; I.2.8 [Problem Solving, Control Methods, and Search]: *Heuristic methods*.

General Terms

Algorithms, Management, Design.

Keywords

Linked Data, Semantic Web, Digital Journals, Linked Data Mining, Concept Aggregation, DBpedia.

1. INTRODUCTION

Recently, the scientific community agreed on the requirements for the systems that provide access to published research papers, i.e. to a comprehensive collection that can be indexed, searched and linked efficiently [7]. The open access movement drives the scientific community to 'dynamic digital archive' [1] which reflects the visionary ideas of Roberts et al.: "*it ... will enable researchers to take on the challenge of integrating and interconnecting the fantastically rich, but extremely fragmented*

and chaotic, scientific literature" [1].

In digital journals' environment, a well-linked collection of electronic resources is of great importance in supporting number of task: 1) providing instant access to related resources 2) increasing knowledge visibility 3) supporting forthcoming research which is usually innovated based on exiting knowledge 4) creating opportunities for collaborations between organization, institutions, and persons. Modern digital journals provide such functionalities by intelligent linking of relevant resources with heterogeneous repositories [8][2].

In the context of current research, finding information about authors (authors' profiles) in a digital journal's environment is crucial to increase the overall productivity and unprecedented success. The discovery of authors' profiles helps in accomplishing the following task: 1) Users of digital journals need to search the research collaborators, 2) users need to search experts to seek guidance, and 3) journal administration want to explore new reviewers. All of this is not possible only by looking author's publication list alone, but one also needs other information of the authors. This information may include a short biography, research areas, co-author network, research projects, academic records, achievements, and geographical position of an author. Although much of this information may be acquired from the Web by exploring existing tools such as: search engines, citation indexes, and social networks. Finding task oriented information is bit of a challenge, simply due to 1) availability of the huge amounts of unstructured data and 2) wall gardened data repositories. Thus, users are often frustrated. Hence, there is a need to have a system which can retrieve, aggregate, structure information from diverse sources, and can present a coherent view of authors' profiles at one place. Furthermore, this information can be supplied in users' local focus and context.

In the past, various techniques/heuristics have been applied to find relevant resources using number of ways: 1) by exploiting metadata of resources [21], 2) by performing natural language processing on unstructured data, 3) by training the system on some machine learning algorithms [27], 4) by computing text similarity, and bibliographic analysis [21]. However, linking relevant resources with the help of all available metadata produce satisfactory results but most of the time, the results produced from unstructured data (whatever technique is used) is not up to the mark. In this scenario, The Semantic Web tries to structure data which can be processed intelligently by machines.

One of the most successful projects of the Semantic Web is Linked Open Data [9] which provides semantically rich datasets. These datasets can be explored for knowledge discovery and creating cross-references between relevant resources. The ever

growing Linked data cloud in its own narrates the success story of Linked open data movement. In aggregation, there are about 13.1 billion RDF triples which came cross from different practical, social, business and research domains. At present research community is changing its gear from techniques of opening authoritative data repositories to build applications which can make use/reuse of currently semantified data. The utmost goal is to make realization of the power of Linked Data in business, public and government as well as providing a medium to make the World Wide Web cleaner and connected. The semantified personal information present in different dataset of Linked Open data (Dbpedia, FOAF, SIOC, and DBLP) opens new horizon to make applications that can discover and link personal information.

Motivated from the success of Linked Data and new publishing paradigm [3], we establish links for authors of a digital journal with the semantically rich data sets of LOD. The process of creating cross-referencing consist of following sub-process 1) disambiguation of individuals, 2) retrieval of information resources 3) structuring the retrieved information and 4) presentation of structured resources. Although the data sets represented by Linked Data are semantically rich, but there are certain problems e.g. locating exact resource URI, lack of quality interlinking [10], heterogeneity in ontologies, and the lack of user interfaces for Linked Data consumption in scientific communities and business enterprises [11][4]. To locate intended resources URI, authors have presented an intelligent Keyword-URI mapping technique [5]. To retrieve and structure information resources, an innovative Concept Aggregation Framework [6] was employed. To present a coherent view of information, a user interface named as CAF-SIAL was developed¹.

Going further from this application, we have applied our past research experiences to establish a link between authors of a digital journal with their DBpedia mined profiles and their contributions (published work) crafted from Linked Data. We are confident that these types of applications can help in interconnecting the conventional web applications (e.g. Digital Journals) with third generation Web Semantic applications. This will bring added value for a digital journal from semantic domain and will increase information visibility of both mentioned applications domains.

2. RELATED WORK

In digital environments, finding relevant information is of great value for increasing the overall productivity. Finding relevant information is not a challenge only for naïve users but the experts (knowledge workers) also spend reasonable time in just searching for resources. According to an IDC white paper by Susan Feldman, knowledge workers spend 15% – 35% of their time searching for information. The success rate of such a search is 50% or less [22]. In 2006, DELOS, a network of excellence on digital libraries, prepared a reference manual entitled “Current Digital Library Systems: User Requirements vs. Provided Functionality” [19]. This document is based on surveys from digital library users. The Digital Libraries functions at the provider side identified by users as of highest priority include: 1) locating resources and 2) creating cross/references.

For the exploration of scientific information, currently users have different options. For example, users can search their required information using search engines, citation indexes, and socially

maintained digital libraries etc. Each of these applications has its own inherited problems. For example, exploring information using a search engine requires correct formulation of search queries. The experiments elaborates that a searcher quite frequently switched back n fourth between browsing and querying. The switching involves reformulation, reformatting of queries and change of search engines based on result set overview [23]. All of this exercise consumes reasonable time of a searcher.

Citation indexes have their own problems. In autonomous systems, one of the major problems is disambiguating individuals. For example, the automated approach employed by Google Scholar had identified name of places as actively cited authors [24]. The discovery of false-positive citations is another example of retrieving false information [25].

In the past research work, the task of finding information about individuals was accomplished by finding home pages of individuals. The home page finder systems use various techniques like neural net classifier [27], document content, anchor text, and URI-type classifications [26]. The above defined techniques, off course, service users to some extent, but as these technique were applied on unstructured text which makes it very difficult to get very good results. However, the availability of semantic datasets drives to knowledge explorations using semantic resources and their relationships. In this regards, one of the most successful and growing project of Semantic Web is Linked Open Data [9]. To consume information from the Linked Open Data cloud, different systems have been built. One of them is RKB explorer [28].

RKB Explorer is a Semantic web application which provides it's users (project members and other researchers) with a unified view of the momentous number of Linked Data sources regarding a selected domain. Underlying logic of this application works on the mediated huge ontology and various triple stores. This allows users to move seamlessly between the typical instances of general concepts (people, projects, publications, research areas) and other concepts in the application domain. The overall vision of the application was to have set of services and applications for both acquiring and publishing knowledge.

Most related work to our current research in presentation of the information is Freebase [29]. Freebase is an open database for people where community contributes for community knowledge built up. Freebase already contains millions of facts in dozens of categories spanning people, locations, books, movies, music, companies, science, sports, and more, and is growing daily. Major difference of freebase from Wikipedia is the underlying infrastructure about managing facts rather than full article. It's interesting use of the crowd intelligence in clustering related information into various informational aspects makes it unique in managing the heterogeneous structured data. On the other hand, our application uses semi-automated approach without relying on community services and also integrates the contributions (published work in conferences, Journals etc.) which are not covered in the Freebase. The brief introduction to our previous research is presented in the next section.

3. Concept Aggregation Framework

In our past research, we introduced a Concept Aggregation Framework [6] for structuring informational aspects of a *resource_type:person*.

¹ <http://cafsial.opendatahub.org/>

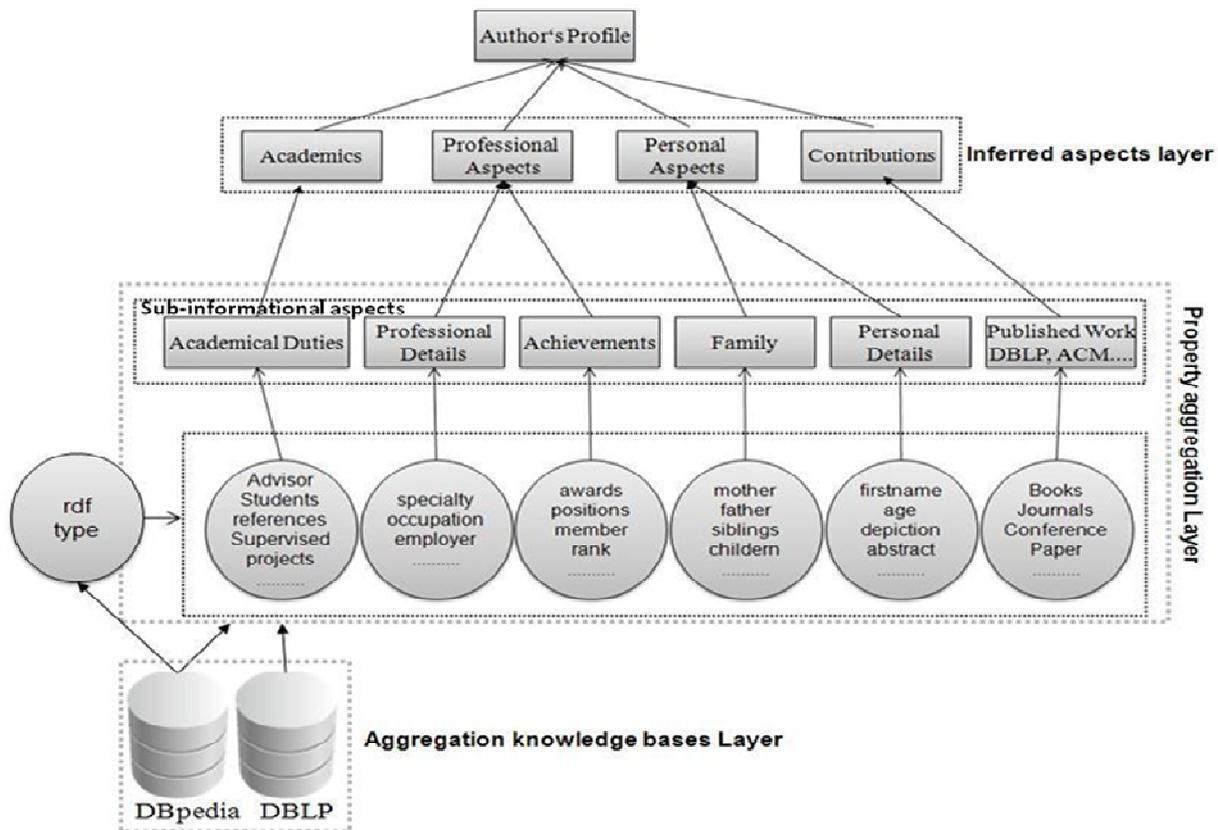


Figure 1. Concept Aggregation Framework.

This framework was able to organize various semantic resources of Linked Data into broad logical informational aspects representing an overall picture of a person. i.e. a *resource_type:person* was represented in the following informational aspects: personal, professional, social, and dark side. The Linked Data resources allowed us to explore semantic representations of resource properties and relationships to come up with such a framework.

The implemented algorithm was able to link a semantic resource-property-relationship with the most suitable informational aspect. To translate resources from Linked Data into informational aspects, a layered approach was used as explained in next paragraph and can be seen in figure 1. The framework has been applied in CAF-SIAL system [5][6][15] which is up and running since 2009. The proposed system also bridged a gap between semantic search and end users by hiding complex mechanics of the Semantic Web. The evaluations of the system showed a certain wavier of cognitive load from users in searching and presentation of information [5].

This paper is a continuity of the previous work by exploring a digital journal as an application area. Authors of a digital journal have been linked with Linked Data resources using Concept Aggregation Framework. In a scientific domain, the authors can be represented in four informational aspects like: 1) personal, 2) professional, 3) academics, and (4) their contributions (Published work). The figure 1 depicts an overall working of the framework. The framework is represented in three layers: 1) aggregation knowledge bases layer, 2) property aggregation layer, and 3) inferred aspect layer.

In the aggregation knowledge bases layer, all properties of a person (author) are extracted and aggregated from DBpedia. In property aggregation layer, retrieved properties are processed using a semi-automated technique. The relevant properties are filtered and linked with sub-informational aspects of an author as shown in figure 1. In our previous implementation, a set of experts manually tagged and assigned relevant properties to sub-informational aspects. However, in the current application, a semi-automated technique worked fine which is based on our previous research experiences and set of developed heuristics. In the inferred aspect layer, the sub-informational aspects are further linked to main informational aspects to get a comprehensive view of authors. More information about the Concept Aggregation Framework can be found in [5][6][15].

4. DATASETS

4.1 Journal of Universal Computer Science

The Journal of Universal Computer Science (J.UCS) is a high-quality electronic publisher that deals with all aspects of Computer Science [16]. J.UCS has been appearing monthly since 1995 with uninterrupted publications. According to the survey paper on electronic journals [17], J.UCS has incorporated innovative features such as the enabling of semantic and extended search and its annotative and collaborative features. It was one of the first electronic published journals to have implemented features such as personal and public-annotations, multi-format publications, multi categorization, etc. These features have made J.UCS a rather unique electronic journal. Readers of such high-quality electronic journals expect and anticipate highly sophisticated features, such as automatic reference analysis,

similarity search between documents and other features using knowledge management technology [18].

The J.UCS dataset provides the list of the authors who have published their work in any of the journal issues. The author ID maintained at J.UCS server along with first, middle and last name is tabulated in this dataset. In total 2593 authors from J.UCS were used for this experiment.

4.2 DBpedia

DBpedia, a semantic flip of Wikipedia is one of the biggest examples of Social Semantic Web. DBpedia is considered one of the most promising knowledge bases, having a complete ontology along with Yago classification [12]. It currently describes more than 2.9 million things, including at least 282,000 persons, 339,000 places etc [13]. The knowledge base consists of 479 million pieces of information (RDF triples). The openly available RDF dumps make DBpedia an interesting subject of study. There has been valuable work done on studying the reliability of Wikipedia URI's [14] that is a backbone of the DBpedia. This study suggests that the meaning of a URI stays stable approximately 93% of the time. Its heavy interlinking within the Linked Data cloud makes it a perfect resource to search URIs. For current experiments, we concentrated on the part of DBpedia that encompasses data about persons.

Two RDF dumps about personal information (*Persondata* and *Links to DBLP*) were selected to find relevant information of the J.UCS authors. These datasets are freely available in RDF dumps and can be downloaded from <http://wiki.dbpedia.org/Downloads34>.

4.2.1 Persondata

This dataset includes the information about persons extracted from the English and German Wikipedia, represented using the FOAF vocabulary.

4.2.2 Links to DBLP

Links between computer scientists in DBpedia and their contributions in the DBLP database are enlisted in a *same:as* relationship in this dataset. To follow the DBLP links, The D2R Server, a *semanitized* version of DBLP bibliography was accessed from Berlin and Hanover SPARQL endpoints. This D2R Server is based on the XML dump of the DBLP database. The DBLP database provides bibliographic information on major computer science journals and conference proceedings. The database contains more than 800.000 articles and 400.000 authors [20].

5. SYSTEM ARCHITECTURE

The architecture design of the application is depicted in figure 2. The proposed system is divided into four modules named such as: 1) Database and Triple Store Construction, 2) URI Acquisition, 3) Author URI Validation, and 4) Concept Mapping.

The database and triple store construction part discusses the data acquisition, manipulation of J.UCS data set and the process of converting RDF personal dataset into local triple store. The URI acquisition module describes how the URI of a J.UCS author is acquired from local triple stores and from remote semantic search services. Author's URI validation module encompasses the developed heuristic to validate the URI. Last module discusses the concept aggregation and presentation of the results.

5.1 Database & Triple Store Construction

For this application three datasets were used, one from J.UCS and other two from DBpedia. Along with these datasets, web service of Sindice (a semantic search engine) was also utilized when the local search fails.

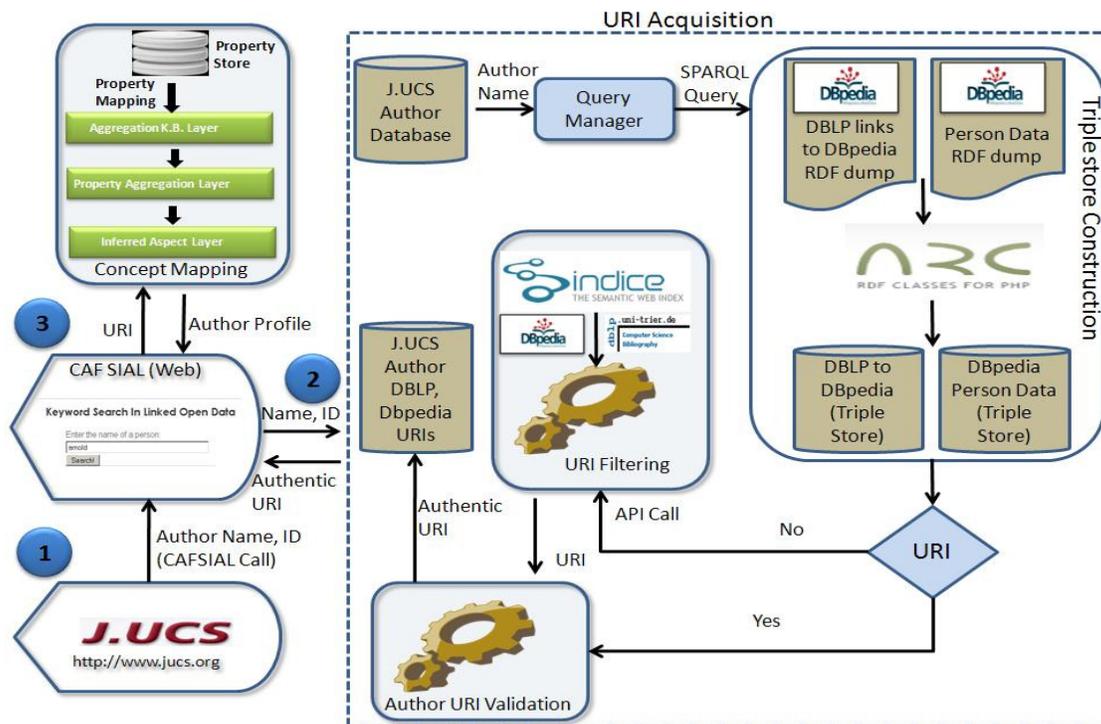


Figure 2. J.UCS to CAF-SIAL System Architecture.

The downloaded RDF dumps were converted into a Local triple store by using ARC2. ARC2 star triple store configuration gives a facility for querying at statement level. In DBpedia person triple store total 29,498 URIs are stored. The DBLP to DBpedia dataset contain information of 196 computer scientist.

5.2 URI Acquisition

A layered approach was employed to acquire the intended URI from local triple store and from Linked Data Cloud.

The layered approach is shown below:

- Direct matching of J.UCS authors with DBpedia Persondata dataset
- Direct matching of J.UCS authors with Links to DBLP dataset
- Querying and Filtering of URI from Sindice

Details can be found in the implementation section.

5.3 URI Validation

Our past research of CAF-SIAL system helped us to develop set of heuristics to validate the acquired resource URI. The developed heuristics helped us to remove URIs showing no data or representing a person from non-scientific domain. The details of heuristics can be seen in the implementation section.

5.4 Concept Mapping

The *concept mapping* module is responsible for mapping the retrieved set of properties onto concept aggregation framework. The *concept mapping* module receives an authenticated URI from *URI validation* module. This URI is used to retrieve available set of properties from LOD. These properties are passed to concept aggregation framework which organizes these properties in informational aspects of an author. The *concept mapping* module returns a comprehensive author profile represented in logically organized informational aspects. These aspects are further visualized to users. The visualization is shown in figure 3.

6. IMPLEMENTATION

The implementation process can be divided into three processes.

1. Locating Author DBpedia URI
2. DBpedia URI Authentication
3. Visualization

6.1 Locating Author DBpedia URI

This process can be divided into four steps. Overall these processes find the intended URI of the particular J.UCS author in DBpedia

6.1.1 J.UCS dataset Pre-Processing

Sometimes, the authors' names contain *umlauts* characters which need to be processed before matching. An automated script was written to remove such inconsistencies. Subsequently first, middle, and last names were concatenated to construct a full name for the matching process.

6.1.2 DBpedia Person Data direct matching with J.UCS Author Dataset

In this step, a complete author name, acquired from the previous step, was considered for matching in the *DBpedia Persondata* triple store. After this operation seven out of total authors were matched yielding in very low success rate.

6.1.3 DBpedia Links to DBLP direct matching

In this step, authors were matched with DBLP local triple store. The result of this matching was also not satisfactory as only eight more authors were found.

Due to inconsistencies in local triple store and old DBpedia dataset, the results were not satisfactory. To increase the discovery of URI, different online semantic Search engines were analyzed like Falcon [30], Swoogle [31]. Based on the up-to-date version, vast indexing corpus, and API access, the Sindice (Semantic Search Service) was selected for further matching of URIs.

6.1.4 Sindice Search Service

A web service was written to call the API of Sindice with the formulated query. The process executed iteratively for every unfound J.UCS author. In response, Sindice provided the list of the URI's which were further filtered out on the basis of DBpedia provenance. In the end direct matching of author full name with the DBpedia filtered out URI's was performed to pick the exact URI.

After this processing, 337 DBpedia URI's out of entire J.UCS author list were found giving substantial improvement to the results. These results were stored in a data table for further processing.

6.2 DBpedia URI Authentication

Verifying and disambiguating URI is an important part of this application. By manual de-referencing and inspection of the acquired URI's, we discovered some inconsistencies:

- URI of the respective author exists (wrongly indexed by Sindice) but with no information making it useless.
- Many ambiguous URI's which matched with exact name of the intended J.UCS author leading to wrong person.

To disambiguate authors, a set of heuristics were written as described below:

After inspection it was noted that there are certain kind of properties for a person type which can be exploited to disambiguate individuals. These properties are *dbpedia:Abstract* / *dbpedia:Comment* and *SKOS* categories. For example, *SKOS* categories and keywords, being used to represent the persons belonging to education profession are: "*computer science, computer scientist, professor, informatics, researcher*" etc. All of these constructing properties represent a person belonging to scientific community. Thus the persons having same names and belonging to different professions can easily be filtered.

An automated script was written to check the keywords in the abstract property and *SKOS* categories of the URI. After applying this script on 337 authors, 66 URIs were left. The remaining URI's were either bad links (showing no data) or were representing non-scientific persons.

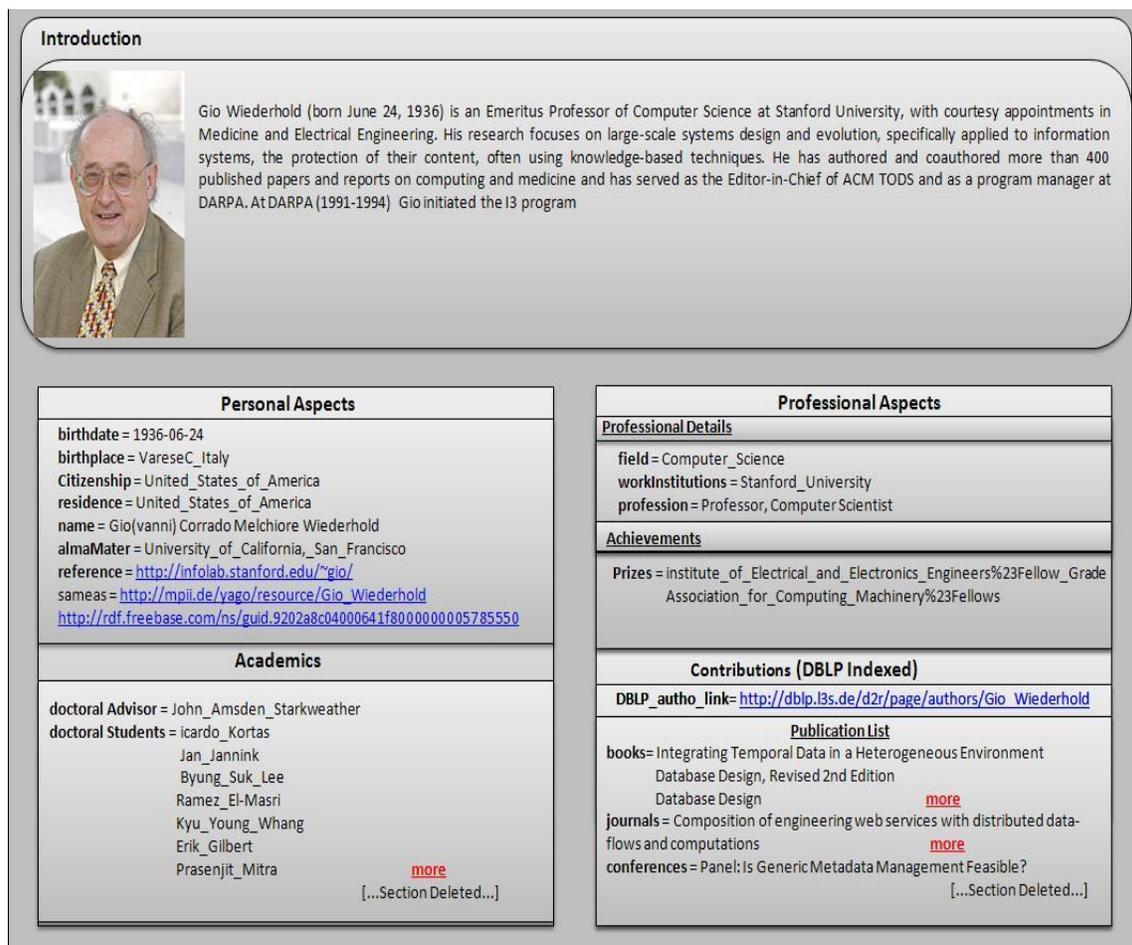


Figure 3. Case Study

In future implementations, we are also exploring to disambiguate authors having same full name and belonging to scientific community.

6.3 Visualization

The request for an author profile is instantiated from the Links into the Future functionality [8] page of JUCS. A hyper link is provided for the authors whose DBpedia or DBLP URI or both exist. When a user clicks on the hyperlink the CAF-SIAL application is invoked.

For example, a user clicks on *Gio Wiederhold*² author; it leads user to CAF-SIAL application where concept aggregation framework makes a conceptual representation of the available properties into different aspects of a person.

From the figure 3, it is obvious that a user will get instant information about the concerned person. A brief introduction of a person along with the picture is shown. Furthermore different aspects like personal, professional, academics, and contributions are shown to the user. From this coherent view, user gets a first overall impression about a person and can follow any hyperlink to see further details.

²http://www.jucs.org:8181/mashup/servlet/FutureLinks?url=/jucs_14_22/determining_software_investment_lag

7. LIMITATIONS

This application is dependent on readily availability of data from DBpedia and DBLP. The down time or maintenance issue is considered as limitations of this application.

8. FINDINGS

After this study, we figure out some recommendations which can help in making improvements in datasets and can benefit the ever growing data openness idea.

1. Users / Author to participate in the social semantic web (Wikipedia) to built up their professional's profiles adhering to the openness idea.
2. Semantic Search engines (Sindice) / DBpedia to clean the dataset from non existing URI's.
3. There are many profiles which are represented with the same name as of authors but representing other persons. So we envision that concrete steps need to be taken by DBpedia to identify ambiguous persons.

At the moment in DBpedia a unique Id is assigned as per title of the article which at certain point is good but if there are persons having similar name, he/she cannot be identified correctly and can lead to the problem of searching and name ambiguity. Assigning a URI on the

basis of profession or certain characteristic can be an option.

9. CONCLUSION AND FUTURE WORK

This paper is motivated to highlight the added value which can be drawn from rich semantic metadata repository of Linked Data for the real world application like Digital Journals. This paper investigates and highlights the potentials for digital publishing system (J.UCS) by intelligent manipulation of the semantic search services and datasets. This linking is helpful for different scenarios e.g.: for users who are searching research collaborators, for journal administration who want to assign new reviewers and for users who want to explore experts to seek guidance. A comprehensive profile of an author is structured and visualized at one place providing various opportunities for collaborations. This is helpful in getting deep insights of author's work, personal and professional life.

In the future, we will be expanding our corpus of searching by including the FOAF dataset as well as the micro-formats to extract the personal information, related blogs and articles. We will also include other paper indexing services like IEEE, ACM to get more entries/papers published by an author. We are also hopeful that the use of conventional crawling and mining techniques with semantic services can lead us to collect the additional and missing information of the authors. Keeping in mind limitation of our system, we are also planning to incorporate caching mechanism for saving the author profile data.

10. ACKNOWLEDGMENTS

This contribution is partly funded by the Know-Center and the Higher Education Commission of Pakistan. The Know-Center is funded within the Austrian COMET program — Competence Centers for Excellent Technologies — under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

11. REFERENCES

- [1] Roberts, R.J., Varmus, H.E., Ashburner, M., Brown, P.O., Eisen, M.B., Khosla, C., Kirschner, M., Nusse, R., Scott, M., Wold, B. Building A GenBank of the Published Literature. *Science*, 291 (5512), pp. 2318-2319. <http://www.sciencemag.org/cgi/content/full/291/5512/2318a>.
- [2] Afzal, M. T., Latif, A., Ussaeed, A., Sturm, P., Aslam, S., Andrews, K., Tochtermann, K., Maurer, H. Discovery and Visualization of Expertise in a Scientific Community. In: *Proceeding of International Conference of Frontiers of Information Technology*, Islamabad, Pakistan, 16-18, Dec. 2009.
- [3] Berners-Lee, T., Hendler, J. "Scientific publishing on the 'semantic web'". *Nature* 410, 1023-1024 (26 April 2001). <http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm>
- [4] Latif, A., Hoefler, P., Stocker, A., Ussaeed, A., Wagner, C. The Linked Data Value Chain: A Lightweight Model for Business Engineers. In: *Proceedings of International Conference on Semantic Systems*, pp. 568-576, Graz, Austria, 2-4, Sep. 2009.
- [5] Latif, A., Afzal, M.T., Ussaeed, A., Hoefler, P., Tochtermann, K. Harvesting Pertinent Resources from Linked Data, accepted in *Journal of Digital Information Management*.
- [6] Latif, A., Afzal, M.T., Ussaeed, A., Hoefler, P., Tochtermann, K. CAF-SIAL: Concept aggregation framework for structuring informational aspects of linked open data, In: *Proceedings of International Conference on Networked Digital Technologies*, pp. 100-105, Ostrava, Czech Republic, 28-31, Jul. 2009.
- [7] Hitchcock, S. M. *Perspectives in Electronic Publishing: Experiments with a New Electronic Journal Model*. PhD thesis, University of Southampton.
- [8] Afzal, M. T. Discovering Links into the Future on the Web, In: *Proceedings of Fifth International Conference on Web Information Systems and Technologies*, pp. 123-129, Lisbon, Portugal, 23-26, Mar. 2009.
- [9] Bizer, C., Heath, T., Ayers, D., Raimond, Y.: "Interlinking Open Data on the Web"; *Demonstrations Track at the 4th European Semantic Web Conference*, Innsbruck, Austria May 2007.
- [10] Jaffri, A., Glaser, H., Millard, I. URI Disambiguation in the Context of Linked Data, *Linked Data on the Web Workshop at the 17th International World Wide Web Conference*, Beijing, China (2008).
- [11] Heath, T. "How Will We Interact with the Web of Data?"; *IEEE Internet Computing*, vol. 12, no. 5, pp. 88-91, (2008).
- [12] YAGO: A Core of Knowledge, <http://www.mpiinf.mpg.de/yago-naga/yago/>
- [13] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives, Z. *DBpedia: A Nucleus for a Web of Open Data*. In *Proceedings of the 6th International Semantic Web Conference (Busan, Korea 2007)*. Springer.
- [14] Hepp, M., Siorpaes, K. and Bachlechner, D. Harvesting Wiki Consensus Using Wikipedia Entries as Vocabulary for Knowledge Management. *IEEE Internet Computing*. 11(5) pp.54-65 Sep 2007
- [15] Latif, A., Afzal, M.T., Hoefler, P., Ussaeed, A., Tochtermann, K. "Translating Keywords into URIS", accepted in the *2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, Seoul, Korea, 24-26 Nov. 2009.
- [16] Calude, C., Maure, H., Salomaa, A. *Journal of Universal Computer Science*, 0 (0), pp. 109-116, 1994.
- [17] Liew, C.L., Foo, S. *Electronic Documents: What Lies Ahead?*, In: *Proceedings of 4th International Conference on Asian Digital Libraries*, pp 88-105, Bangalore, India, 10-12, Dec. 2001.

- [18] Krottmaier, H. Links to the Future, *Journal of Digital Information Management* ,1 (1), pp. 3-8, 2003.
- [19] Candela, L., Castelli, D., Fuhr, N., Ioannidis, Y., Klas, C.-P., Pagano, P., Ross, S., Saidis, C., Schek, H.-J., Schuldt, H., Springmann, M. Current Digital Library Systems: User Requirements vs Provided Functionality, Deliverable D1.4.1, Mar. 2006, retrieved on 22, Oct. 2009.
- [20] Michael L. DBLP - Some Lessons Learned. *PVLDB* 2(2): 1493-1500 (2009)
- [21] Giles, C.L., Bollacker, K.D., Lawrence, S.: CiteSeer: An Automatic Citation Indexing System. In: 3rd ACM Conference on Digital Libraries, pp. 89–98. ACM Press, Pittsburgh (1998).
- [22] Feldman, S. The Hidden Costs of Information Work (IDC #201334, April 2006).
- [23] Hölscher C., Strube G. Web search behavior of internet experts and newbies. *Computer Networks*, Volume 33, Issues 1-6, June 2000, Pages 337-346 doi:10.1016/S1389-1286(00)00031-1.
- [24] Postellon, D.C. Hall and Keynes join Arbor in the citation indices. *Nature* 452, 282 (2008).
- [25] Price, G. Google Scholar Documentation and Large PDF Files, <http://blog.searchenginewatch.com/blog/041201-105511> (accessed 22, July 2009).
- [26] UPSTILL, T., CRASWELL, N., HAWKING, D. Query-Independent Evidence in Home Page Finding. *ACM Transactions on Information Systems*, Vol. 21, No. 3, July 2003, Pages 286–313.
- [27] Kennedy, A., Shepherd, M. Automatic Identification of Home Pages on the Web, *Proceedings of the 38th Hawaii International Conference on System Sciences* , 2005.
- [28] H. Glaser and I. C. Millard. Rkb explorer: Application and infrastructure. In *Proceedings of Semantic Web Challenge*, 2007.
- [29] Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250. ACM, 2008
- [30] Cheng, G., Ge, W., Qu, Y. Falcons: Searching and Browsing Entities on the Semantic Web. In: *Proceedings of 17th International World Wide Web Conference*, pp. 1101-1102, Beijing, China, 21-25 Apr. 2008.
- [31] Ding, L., Finin, T., Joshi, A., Pan, R., S. Cost, R., Peng, Y., Reddivari, P., C. Doshi, V., Sachs, J. Swoogle: A Search and Metadata Engine for the Semantic Web. In: *Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, pp. 652 - 659, Washington, D.C., USA, 8-13, Nov. 2004.