

A Step towards the Arabic DBpedia

Haytham Al-Feel, Ph.D
Faculty of Computers &
Information
Information Systems
Department
Fayoum University, Egypt

ABSTRACT

DBpedia is considered as one of the recently emerging hot topics in the field of Semantic Web due to its importance in adding structure to Wikipedia, the largest online encyclopedia worldwide covering 285 Wikipedia languages. Thought to be a step towards the internationalization of DBpedia, this paper describes the best practices and the efforts performed as part of the work in the deployment of the Arabic DBpedia as a new chapter will be added to the available multilingual DBpedia chapters which can support the maintenance, extraction and enriching of DBpedia especially for those users who their native language is Arabic.

General Terms

Semantic Web, Information retrieval.

Keywords

DBpedia, Arabic DBpedia, Ontology, Wikipedia, Data Extraction Framework, Infobox templates.

1. INTRODUCTION

Wikipedia was established formally on 15 January 2001 as an open source online encyclopedia in one single place [1] which depends on the collaboration in authoring, editing, publishing and maintenance of textual content in several domains and languages by contributors and volunteers worldwide [2] contains various types of information represented as wiki-markup such as geo-coordinates, infobox templates, images, categorization and others [3]. Since that date till now, Wikipedia has become more popular due to its huge knowledge base that is updated all time by many users worldwide and become the 6th most visited website all over the web [4]. Wikipedia editions are available in 285 languages by total number of articles around 22,092,942 [1], however, it has faced many problems, such as; most of Wikipedia content is unstructured, it doesn't provide other ways of search rather than the textual search, and it doesn't have the capability to support sophisticated queries.

Tim Berners Lee has defined the Semantic Web as it is not just the way that help machines and people to work in cooperation, it also adds structure and meaning to the web content which can facilitate the automation of tasks and ask of sophisticated queries [5] [6]. But, unfortunately, the total picture of the Semantic Web is not completed till now. There are some endeavors and projects attempting to use the Semantic Web technologies to structurally and semantically enrich the web. One of these projects is the DBpedia project, which is a result of collaboration between the Freie University of Berlin, the University of Leipzig and the Open Link software. DBpedia project, which started in 2007, aims to extract structured information from Wikipedia as RDF triples to facilitate doing of sophisticated queries for semantic applications [7][3].

This paper describes the efforts performed as part of the work in the deployment of the Arabic chapter of DBpedia as a new multilingual chapter hopefully to be added to available chapters such as the German [8], Greece [9], Korean [10] and others. This study is a step towards the internationalization of DBpedia to help internet users whose native language is Arabic to ask sophisticated queries in their native language. Moreover, it facilitates the performance of those applications which deal with the Arabic language and work as an infrastructure for new applications which may change the web nature.

This paper is organized as follows. Section 2 gives an overview of the DBpedia project. Section 3 explains the process of creating and mapping of infoboxes templates of the Arabic pages and the difficulties faced. Section 4 shows the DBpedia extraction process including the architecture of the DBpedia Knowledge Framework. Finally, Conclusion and suggested Future Work are discussed in Section 5.

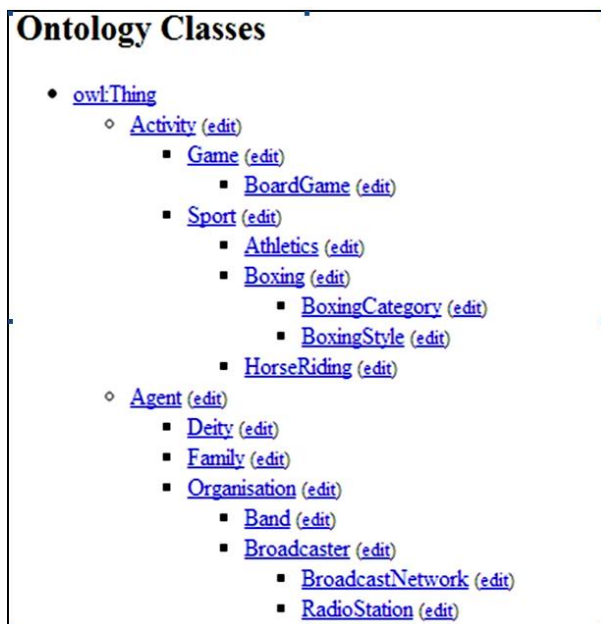
2. INTERNATIONALIZATION OF DBPEDIA

Currently at the time of writing of this paper, DBpedia dataset describes over 3.64 million things across 97 languages providing 1 billion pieces of information in the form of RDF triples with an open license [3] covering various domains, such as; persons, organizations, software, films, books, etc.. [11][12]. There is an increasing number of publishers, contributors and data providers who publish their data on DBpedia assuming that DBpedia will improve and enhance the data quality and search in Wikipedia textual content according to DBpedia's capability to take out structured information from Wikipedia and to update it regularly if any changes occur in Wikipedia. An early version of DBpedia was in English. DBpedia Information Extraction Framework (DIEF)[11] assists DBpedia scope to become wider and cover other Wikipedia languages via adding new other chapters. The available release of DBpedia at the time of writing of this paper covers 15 localize DBpedia chapters that include Catalan, German, Greek, Spanish, French, Irish, Croatian, Hungarian, Italian, Dutch, Polish, Portuguese, Russian, Slovene, Turkish languages in addition to English.

Unfortunately, the Arabic language was not available in these localized chapters of DBpedia. However, the number of internet users who speak Arabic is 86,077,806, around 3.8% of internet users worldwide [13]. This number is relatively large if compared to the numbers of users who speak other languages. On the other hand, this large number of users has not been reflected actually on the total number of articles available in the Arabic edition, which is around 179,044 article of Wikipedia content, if compared to the total number of articles, which is around 22,175,193 available in 285 language. The Arabic edition has been ranked as the 25th

Through this section we will describe briefly the DBpedia ontology, the idea beyond the infobox template, the SPARQL language which is considered a protocol and query language used to retrieve data from DBpedia via end points such as Virtuso[14] and Snorql [15].

DBpedia Ontology is manually created depending on the English chapter of Wikipedia infoboxes. The number of classes that form the ontology increased from 170 classes [3] to 359 classes [16] in addition to the growing of number of properties representing different relations with domain and range from 720[3] to 1,775 [16]



2.2 Infobox Template

Fig 2: Part of the description of the University infobox [17]

SPARQL is considered the query language for the Semantic Web and is used as an end point to query the DBpedia knowledgebase [3] to retrieve triples which can help in the development of applications based on the structured information of DBpedia. To query information from DBpedia is not a sophisticated process, it is like other query languages, the only difference in the form of the statement that depends on the idea of representing data on triples. SPARQL became a standard in 2008.[18] and from that time it is used in most of Semantic Web applications due to its flexibility to retrieve information from the web of data.









```
SPARQL:
```

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT ?s, ?x WHERE {
  ?s dbpedia-owl:country :Egypt.
  ?x rdfs:label "Adel Emam"@en.}
```

Results:

SPARQL results:

S	X
:11'09%2201_September_11 	:Adel_Emam 
:Ain_Sukhna 	:Adel_Emam 
:Al-Azhar_University 	:Adel_Emam 
:Egyptian_Navy 	:Adel_Emam 

3. CREATION & MAPPING OF ARABIC TEMPLATES IN WIKIPEDIA

28

Wikipedia that has been extracted to DBpedia. Beyond the rendered infobox, there is a syntax describes the features appear to the user in the infobox. Infobox consists of pairs that represent attribute name and its value. Not all attributes have values, and this may go back to several reasons, one of them is related to the unavailability of the value of the attribute to the author of the Wikipedia infobox who may not be the same author of the article. An attribute may not have any meaning in the article context perhaps is another reason, for example; an attribute of chancellor in the university template does not have any meaning in the Arabian Universities according to the unavailability of this position there, on the other hand; it is an important attribute inside the infobox template for the European Universities. In addition to the above mentioned reasons, another important reason is the inconspicuous meaning of the attribute corresponding to a bad choice or the wrong expression for the attribute name by the creator of the main template [19]. In the Arabic Edition of Wikipedia, the infobox is placed on the top left hand side of the Wikipedia page, opposite to the direction of the English edition; attributes can be written in Arabic or English, but it is preferable to be written in Arabic, but the attribute values should be written in Arabic. DBpedia uses article names for creating identifiers; these names are extracted from the URL of Wikipedia article via <http://ar.wikipedia.org/wiki/Name> and the URI of that resource in DBpedia became <http://ar.dbpedia.org/resource/Name>.

There are two ways of extraction generally used in DBpedia described in detail in [3]. They are the Generic Infobox Extraction and the Mapping-Based Infobox extraction. In the Generic Infobox Extraction; all attributes and their values in Wikipedia infoboxes are extracted in the following: the article will be the subject, the infobox attribute concatenated with the namespace of <http://dbpedia.org/property/> will be a predicate, while the value of the infobox attribute in Wikipedia is considered an object.

On the other hand, the Mapping-Based Infobox Extraction is used to cover the shortcomings of the first method and it depends on the mapping of Wikipedia templates to ontology created to collect the most common properties used in Wikipedia.

Arabic DBpedia, like other chapters of DBpedia, has some problems in mappings; one of these problems is the unavailability of infoboxes for many Arabic articles. For example, in the list of the Egyptian museums which considered value pages in the cultural heritage there is only 4 out of 50 museums have infoboxes which is a relatively low percentage, 8%, of the whole list. For that reason; many infoboxes for various articles were built, but this needs more efforts from contributors and the community to finish as many as can to increase the amount of data extracted from the Arabic Wikipedia feeding the Arabic chapter in DBpedia.

Because of the importance of the infobox template to DBpedia, and because the work here is considered the first attempt in the deployment of the Arabic DBpedia, this paper will explain simply the various steps needed for mapping from scratch. The process begins with the creation of infobox as an effort to increase the number of infoboxes in articles written in Arabic language.

To check the mapping of an infobox template, all what we need is to take the infobox name concatenated with the URI of the Arabic chapter in the mapping statistics, for example: mappings.dbpedia.org/index.php/mapping_ar:معلومات_جامعة

If the mapping to the ontology has been done before, a template created including the mapping to the class inside the ontology in addition to the mapping of various properties will be displayed as in Figure 4.

If the Wikipedia infobox has not been mapped before, the mappings page in DBpedia can be made via mapping the infobox name and attributes to the suitable class properties inside the ontology as shown in Figure 5.

Mapping ar:معلومات جامعة

This is the mapping for the Wikipedia template [معلومات جامعة](#). Find usages of this Wikipedia template [here](#).
Test this mapping [with](#) some example Wikipedia pages. Check which properties are not mapped yet [here](#).
[Read more about mapping Wikipedia templates.](#)

Template Mapping (help)

map to class [University](#)

Mappings

Property Mapping (help)

template property [الاسم](#)
ontology property [foaf:name](#)

Property Mapping (help)

template property [الشعار](#)
ontology property [motto](#)

Property Mapping (help)

template property [التأسيس](#)
ontology property [foundingDate](#)

Property Mapping (help)

template property [الرئيس](#)
ontology property [chairman](#)

Fig 4: Mapping of "معلومات جامعة" infobox template to the corresponding properties in the DBpedia ontology

```
{{TemplateMapping
| mapToClass = University
| mappings =
{{PropertyMapping | templateProperty = الاسم | ontologyProperty
foaf:name }}
{{PropertyMapping | templateProperty = الشعار | ontologyProperty
motto }}
{{PropertyMapping | templateProperty = التأسيس | ontologyProperty
= foundingDate
}}
{{PropertyMapping | templateProperty = الرئيس | ontologyProperty
= chairman }}
{{PropertyMapping | templateProperty = الناظر | ontologyProperty
= chancellor }}
{{PropertyMapping | templateProperty = المدير | ontologyProperty
= president }}
{{PropertyMapping | templateProperty = النائب | ontologyProperty =
vicePresident }}
```

Fig 5: Mapping syntax of "معلومات جامعة" infobox template

Each infobox is mapped to a class and each attribute is mapped to a property in the DBpedia ontology[20]. Infobox is considered a flag indicating that this Wikipedia page belong to a particular class in the DBpedia ontology[20] and this look in the attribute mapToClass; while the second mappings attribute represents the mapping of attributes to properties inside the DBpedia ontology[20] [21].

Arabic language is considered as non-Latin language that contains characters appears as symbols via web browsers according to ASCII, for that reason users are advised to use Mozilla Firefox or Google Chrome when dealing with the Arabic DBpedia.

Before the work done in this paper, the Arabic chapter of DBpedia has not been discussed yet and there was no Arabic chapter in DBpedia. After the creation of an Arabic

```
{{Infobox university
| name                =
| native_name         =
| image_name          =
| image_size          =
| image_alt           =
| caption             =
| latin_name          =
| motto               =
| motto_lang          =
| mottoeng            =
| established          = <!-- {{Start date|YYYY}} -->
| closed              = <!-- {{End date|YYYY}} -->
| type                =
| affiliation          =
| endowment            =
| budget              =
| officer_in_charge   =
| chairman             =
| chancellor           =
```

From the left side menu, the Arabic template for that infobox can be selected, and the attributes of "جامعة الفيوم" can be filled by its values, while other unknown attributes can be left blank till other contributors fulfill them later.

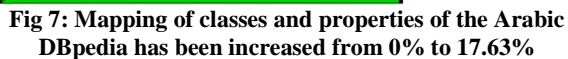
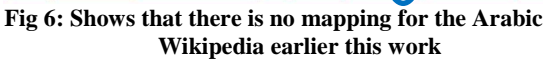


Fig 9: Part of the infobox for the University infobox in Arabic

Or ["ar.wikipedia.org/wiki/Template:جامعة"](https://ar.wikipedia.org/wiki/Template:جامعة)
["قالب:جامعة"](https://ar.wikipedia.org/wiki/قالب:جامعة)

One of the problems in the Arabic Wikipedia, is the occurrence of different names for the same attribute, such as the birth date attribute which appears in various infoboxes by different names: one time as "تاريخ الميلاد" another time as "ولد" "في", third time as "تاريخ الولادة" and fourth time as "ولد" as in Figure 10. But these conflicts for all these attributes can be mapped easily to birthdate in DBpedia ontology <http://dbpedia.org/ontology/birthDate> and by this way we can jump over different names for the same attribute



Fig 10: Shows different names for the birth date attribute in different infoboxes

Another problem happened in the Arabic DBpedia can be concluded as the occurrence of various infobox names referring to the same concept such as:

"معلومات فنان موسيقى" "صندوق معلومات فنان موسيقى" "فنان موسيقى"

This problem can be solved by mapping the four concepts to the same equivalent ontology class, which is the Musical Artist class in the DBpedia ontology. Inconsistency of names between the infobox and its template considered also one of the problems in the Arabic chapter of DBpedia, for example, there is a template called "جامعة" while the infobox name is called "معلومات جامعة" and for the ease of use, we recommend that both the infobox name and the template name should have the same name.

In most of the Arabic templates, when Geo-names are placed in the infobox, they create problems in the formatting of templates, but when they are situated outside the infobox, they work without any problem in Wikipedia, but in this case they became unusable to DBpedia.

On the other hand, If the infobox includes more than one words, such as: "معلومات فنان موسيقى", the different words in the infobox name should be concatenated with each other by underscore to be used in the URL of the Wikipedia article or in the URI of DBpedia mapping.

4. THE DBPEDIA EXTRACTION

The goal of DBpedia is to extract the Wikipedia's structured content, organize it by a way to be easily accessible by users and applications. Extractors extract all Wikipedia attributes inside infoboxes and templates and put them on properties in the "http://dbpedia.org/property/" namespace. Names of properties in the ontology reflect attribute names in the infoboxes [16].

The extraction process in DBpedia can be made via two ways: the first one is a Dump Based Extraction, while the second can be through the live extraction. The extraction manager, which is the main actor in the DBpedia framework, manages the extraction of the Wikipedia articles collected by the page collection either in dump based or in the live extraction and passes them to different extractors. On the other hand, parsers support the extractors by converting the values into different units, splitting markup into list and determining data types and Geo-coordinates [3] [22].

4.1 Dump Based Extraction

The Dump Based Extraction depends on the Wikipedia dumps published in SQL for all editions and updated monthly and collected in DBpedia via the page collection of Wikipedia database. Wikipedia database is considered the source for the text of different articles in different languages and editions while the destination for those articles is the N triple serializer which pass the triples to the N-triple dumps to virtuoso triple store.

4.2 The Live Extraction

It is the second type of extraction in DBpedia; it depends on the live update of DBpedia according to any changes or modifications done in Wikipedia due to the privilege given to the DBpedia project to access the Wikipedia Open Archives Initiative protocol for Metadata Harvesting (OAI – PMH) live feed that instantly reports all Wikipedia changes [3] [22] [23]. Before the live extraction, the state of Wikipedia was difficult to be reflected in a precise way [22]. The live Wikipedia page collection is the source of article's text in different languages as a case in the Dump Based Extraction; while the SPARQL update is a destination that store the extracted RDF triples in the triple store. When a wiki page is processed, the extractors work to extract the infobox inside it, after the extractors complete the extraction of the new extracted triples; these triples are inserted into a triple store, which is called virtuoso, instead of the old triples and at the same time these new triples are placed in the N-Triple Dumps [22].

4.3 Architecture of the DBpedia Knowledge Extraction Framework

The architecture of the extraction framework will be described in this section showing tasks of various extractors and modifications that can facilitate the extraction of articles written in the Arabic language.

- **Label Extractor:** This extractor is used to extract the article's title in various Wikipedia editions and place them into `rdfs:label`. An example for the label extractor is the page of the famous Egyptian actor Adel Emam in Wikipedia that has this URL:

"http://en.wikipedia.org/wiki/Adel_Emam". This article is available in 14 different language, however, Adel Emam's page in DBpedia via "dbpedia.page/Adel Emam" has only 5 values for the `rdfs:label` for only 5 languages, not for all the 14 languages used for publishing this article. The only 5 extracted languages are those languages which have localized chapters in DBpedia on their local language and have different extractors, one of which is the label extractor which extract the title of an article in various Wikipedia chapters and set it on `rdfs:label` in the DBpedia ontology.

- **Abstract Extractor:** Short abstract is extracted from the Wikipedia and placed in `rdfs:comment` in DBpedia. This short abstract is taken from the first paragraph of the Wikipedia page; not only from the English edition, but from all other editions written by different languages. On the other hand there is a property in DBpedia called `dbpedia-owl:abstract` includes the long abstract of different editions of the Wikipedia page.



Fig. 11: Different labels for Adel Emam's page via
rdfs:label

- **Infobox Extractor:** This extractor is responsible for extracting all the properties available in infoboxes with a clean-up of some of the values of these properties.
- **Disambiguation Extractor:** This extractor works with words that have homonyms meaning that may refer to several web resources. When a searching done for a homonyms words, Wikipedia return a page that has several links to the word which has different meaning; and this is what we call disambiguation. This is extracted in DBpedia in a property called dbpedia-owl:wikiPageDisambiguates.
- **Wiki Page Extractor:** Is an extractor used to extract links to the corresponding articles in Wikipedia [24].
- **Interlingua Links Extractor:** Links that connect an article in one language to other editions of Wikipedia are used to extract content of these editions and use them to enrich labels, short and long abstracts.
- **Geo Extractor:** Most of infoboxes related to places in Wikipedia have Geo-Coordinates. The Geo extractor is responsible for the extraction of the geographic coordinates.
- **Page Link Extractor:** This extractor is used to extract links between different related articles in Wikipedia and place them in dbpedia-owl:wikiPageWikiLink property.
- **Home Page Extractor:** This extractor extracts the website for the homepage of that article and places it on foaf:homepage.
- **Article Categories Extractor:** Each Wikipedia article is categorized according to its topic of interest. These categories are extracted to DBpedia properties via dct:subject, category and skos vocabulary[25] which simply represent categories as concepts via SKOS:concepts. On the other hand, category relations are represented via skos:broader[3].
- **Image Extractor:** It extracts the links pointing to the image in articles and replaces these URI into foaf:depiction.
- **Redirects Extractor:** This extractor is used to extract the redirection of Wikipedia to other articles via synonymous terms [3] and place them into dbpedia-owl:wikiPageRedirects.
- **External Links Extractor:** It is different than the homepage extractor; this extractor extracts links to external web pages related to a concept[12]. If an article contains links to other web resources, the external link extractor will extract these links and will set them in a property called dbpedia-owl:wikiPageExternalLink.

- **Revision ID Extractor:** This extractor extract the ID of the Wikipedia revision.
- **Pnd Extractor:** The Pnd number is a record created by the German National Library used to identify.
- **Person Data Extractor:** This extractor is used to extract the person name, date and place of birth from the infobox and represent them using FOAF vocabulary [12][24].

Table 1. Availability of Extractors for the Arabic Language

Type of Extractor	Availability for the Arabic Language
Infobox	√
Person Data	X
Label	√
Pnd	X
Disambiguation	√
Wiki Page	√
Page Link	√
Geo	√
Article Categories	X
Image	X
External Links	X
Homepage	X
Redirect	√
Abstract	√
Interlingua Links	√

5. CONCLUSION & FUTURE WORK

Till the time of writing of this paper Arabic DBpedia stills a prototype. This paper represents the experiences in mapping of infobox templates and properties in Wikipedia to the ontology manually created in DBpedia as a base layer in the DBpedia extraction framework. This paper focuses on the Arabic DBpedia as a new chapter hopefully to be added to other international chapters in the near future. Arabic DBpedia dataset needs to be improved and well managed to increase the number of extracted triples written in the Arabic language as a step towards Arabic Semantic Web applications which are rarely corresponding to the internet users who their native language is Arabic. In addition to that contributors and volunteers are invited to join the work in the Arabic chapter of DBpedia that requires more efforts to finish as many as can of the mapping of Arabic content

6. ACKNOWLEDGMENTS

The author wish to thank the AG Corporate Semantic Web; Freie Universität Berlin; especially Prof. Adrian Paschke, and Alexandru Todor. In addition to M. Morsey from

University of Leipzig for their continuous support in this work. This work also partially supported by the Erasmus Mundus fund project.

7. REFERENCES

- [1] Wikipedia Encyclopedia.[Online].Available: http://en.wikipedia.org/wiki/History_of_Wikipedia .[Accessed: September.2, 2013].
- [2] S. Auer and J. Lehmann, "What Have Innsbruck and Leipzig in Common? Extracting Semantics from Wiki Content," in ESWC 2007, pp. 503-517.
- [3] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann., "DBpedia-A Crystallization Point for the Web of Data," Journal of Web Semantics, vol. 7, pp. 154-165, September 2009.
- [4] Alexa.[Online].Available:<http://www.alexa.com/topsites> [Accessed: September.9, 2013].
- [5] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web" in Scientific American, May 2001, pp.29-37.
- [6] T. Berners-Lee, "Artificial Intelligence & the Semantic Web, " World Wide Web Consortium, 2006.
- [7] Aprosio, A. P., C. Giuliano, et al. Automatic Mapping Of Wikipedia Templates for Fast Deployment of Localized DBpedia datasets. ISWC 2013.
- [8] DBpedia: A Multilingual Cross-Domain Knowledge Base Pablo N. Mendes , Max Jakob, Christian Bizer, Web Based Systems Group, Freie Universität Berlin,Germany,2011.
- [9] D. Kontokostas, C.Bratsas, S. Auer and S. Hellmann, I. Antoniou,and G. Metakides, "Internationalization of Linked Data.The Case of the Greek DBpedia Edition." ,Journal of Web Semantics January 10, 2012.
- [10] Kim, E. Kyung, Weidl, M., Choi, K.-S. and Auer, S, Towards a Korean DBpedia and an Approach for Complementing the Korean Wikipedia based on DBpedia. In: Lecture Notes in Computer Science, vol. 575. CEUR-WS, London, UK. pp. 12-21.
- [11] Sebastian Hellmann, Claus Stadler, et al. (2012). "The German DBpedia: A Sense Repository for Linking Entities." Linked Data in Linguistics, Springer.
- [12] C. Bizer. DBpedia 3.7 released, including 15 Localized Editions. [Online]. Available: <http://blog.dbpedia.org/2011/09/11/dbpedia-37-released-including-15-localized-editions/>.Accessed: September.7, 2013].
- [13] Internet World Statistics,".[Online]. Available: <http://www.Internetworldstats.com/stats19.htm>[Accessed : September.10, 2013].
- [14] SPARQL. [Online]. Available: <http://dbpedia.org/sparql> [Accessed: September. 5, 2013].
- [15] Snorql. [Online]. Available: <http://dbpedia.org/snorql> [Accessed: September. 8, 2013].
- [16] DBpedia. [Online] Available: <http://dbpedia.org/About> [Accessed: September. 6, 2013].
- [17] Wikipedia [Online].Available: <http://en.wikipedia.org/wiki/University>.[Accessed: September. 6, 2013].
- [18] World Wide Web Consortium.[Online]. Available: http://www.w3.org/2009/sparql/wiki/Main_Page[Access ed: September. 5, 2013].
- [19] D. Lange, C. Böhm, and F. Naumann, "Extracting Structured Information from Wikipedia Articles to Populate Infoboxes," in The 19th ACM international conference on Information and knowledge management, pp. 1661-1664.
- [20] Sebastian Hellmann, Claus Stadler, et al. (2009). DBpedia Live Extraction OTM '09 Proceedings of the Confederated International Conferences, Springer-Verlag Berlin, Heidelberg. pp. 1209 – 1223
- [21] Pablo N. Mendes, Max Jakob, et al. (Sept. 7-9, 2011). DBpedia Spotlight: Shedding Light on the Web of Documents. I-SEMANTICS , 7th Int. Conf. on Semantic Systems, Graz, Austria.
- [22] M.Morsey, S. Auer, J. Lehmann,C.Stadler and S.Hellmann, "Dbpedia and the Live Extraction of Structured Data from Wikipedia,"In Program:electronic library and information systems,Vol 46(2012),p.27.
- [23] [C. Lagoze, H. V. d. Sompel, M. Nelson, and S. Warner, "The Open Archives Initiative Protocol for Metadata Harvesting,"in 2008.[Online].Available <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm> [Accessed: September. 5, 2013].
- [24] DBpedia Extractor.[Online].Available: <http://wiki.dbpedia.org/DeveloperDocumentation/Extractor> . [Accessed: September. 5, 2013].
- [25] SKOS.[Online].Available:<http://www.w3.org/2004/02/skos> . [Accessed: June. 25, 2013].