

# URI Identity Management for Semantic Web Data Integration and Linkage

Afraz Jaffri, Hugh Glaser and Ian Millard

Dependable Systems and Software Engineering Group  
School of Electronics and Computer Science  
University of Southampton  
{a.o.jaffri,hg,icm}@ecs.soton.ac.uk

**Abstract.** The Semantic Web vision involves the production and use of large amounts of RDF data. There have been recent initiatives amongst the Semantic Web community, in particular the Linking Open Data activity and our own ReSIST project, to publish large amounts of RDF that are both interlinked and dereferenceable. The proliferation of such data gives rise to millions of URIs for non-information resources such as people, places and abstract things. Frequently, different data providers will mint different URIs for the same resource, giving rise to the problem of coreference. This paper describes the phenomenon of coreference, where it occurs in other disciplines and how it is relevant to the Semantic Web. We propose a ‘Consistent Reference Service’ for URI identity management and describe how this is being used in the infrastructure of a scalable Semantic Web system.

## 1 Introduction

The Semantic Web is growing and evolving. The increased adoption of RDF and OWL as knowledge representation formats is enabling the production of Semantic Web systems that can manage, manipulate and display data in novel ways [1, 2]. However, there are also those who believe that the Semantic Web is merely a dream that will never be fulfilled [3]. In order to encourage the adoption of Semantic Web technologies there has been an increased amount of activity in providing a linked data backbone that can be used to bootstrap the Semantic Web [4]. The Linking Open Data project [5] has been a catalyst for such activity, producing data sources that expose their knowledge as RDF and assert links between datasets. Current information sources include Geonames, DBLP, MusicBrainz, The CIA Factbook and US Census data.

### 1.1 Linking Open Data

The production of the first tutorial on how to link Open Data [6] means that many more information providers are likely to make their knowledge available. Such activity will allow a formidable mass of knowledge to be used by Semantic Web applications. The linked data methodology has also introduced the use of additional

techniques to publish Semantic Web data, such as using HTTP 303 redirects to dereference URIs about non-information resources, which have already allowed a new breed of Web browser to be built that can analyse and explore linked data [7].

The first set of data that is being used as a base for all subsequent data linkage is the DBpedia [8] dataset. The DBpedia dataset reportedly contains over 91 million RDF triples and has knowledge covering over one million concepts. The knowledge has been extracted from Wikipedia info boxes that appear on Wikipedia pages. Consequently there have been over one million URIs created corresponding to each Wikipedia page that contains an info box. DBpedia URIs take the form `http://dbpedia.org/resource/resourceName` where *resourceName* is the name of a Wikipedia article. DBpedia has a lightweight ontology that has predicates derived from infobox data such as *name*, *placeofbirth*, *placeofdeath* and *capital*. There are also predicates used from other ontologies that link into the dataset including *foaf:page*, *rdfs:label* and *geonames:featureCode*.

## 1.2 The Problem of Coreference

The explosion in the number of information sources being exposed as RDF has also led to an explosion in the number of URIs used to identify different entities. It is often the case that data in different repositories will hold information regarding identical entities. For example, DBpedia, Geonames, the CIA Factbook and Eurostat all have different URIs for the same country. In another context, Citeseer, DBLP, IEEE and the ACM have different URIs for the same authors and papers. The responsibility of giving an information or non-information resource a URI lies with the data provider and they will assign URIs based on the web domain over which they have control.

The multiplicity of URIs leads to the problem of *coreference*, where different URIs are used to describe the same entity. On an open Semantic Web this presents a problem when there is a need to link together knowledge from disparate information providers. The present approach, used by the Linking Open Data community, is to use various equivalence mining techniques in order to assert *owl:sameAs* relations between entities that are considered to be the same [9]. DBpedia has, for example, made an assertion that: `<http://dbpedia.org/resource/Berlin>` is `<owl:sameAs>` `<http://sws.geonames.org/2950159/>`. In this paper we will argue why this is not the best approach for dealing with coreference, and propose a system for dealing with consistent reference across multiple knowledge bases. Section 2 describes the problem of coreference and how it fits into the current Web architecture, Section 3 gives the implementation details of our solution, Section 4 presents an application that is using the system and Section 5 gives a summary and presents areas for future work.

## **2 Coreference and URI Identity**

The term ‘coreference’ is used in the field of linguistics to define the situation where different terms are used to describe the same referent. This is often done using words such as ‘he’, ‘she’, ‘we’, ‘them’ or ‘it’. On the Semantic Web we use the term coreference to define the situation where different URIs are used to describe the same non information resource. This section gives a brief description of coreference in information science and databases and then goes on to discuss the meaning of a URI and the importance of giving coreference due importance in Web architecture.

### **2.1 Coreference in Information Science**

The problem of coreference within the field of information science has existed for many years and the solution to the problem is based around the use of controlled vocabularies. Such a solution is possible because of the closed world nature and human processing characteristics of a library system. The most popular closed vocabulary is the Library of Congress Subject Headings (LCSH). This vocabulary gives a defined and precise meaning to each subject in the vocabulary which contains over 280 000 terms. Thus it is not possible for people to make their own subject headings, keyword descriptions or tags as is prevalent on the Web today.

A more relevant case of coreference occurs in digital libraries when the author of a publication has to be disambiguated. There are many authors who share the same name and matters are made more complex by the use of initials, different naming formats and spelling errors. For example, the author ‘Hugh Glaser’ could be represented with his full name or by using ‘H. Glaser’, or ‘Glaser, H.’. The task of author disambiguation is an active area of research in information science and many solutions have been proposed. Some solutions use Web based searches in order to determine if one author is the same as another [10, 11]. Such techniques will be needed on the Semantic Web if a consistent web of data is to be created. The ReSIST project [12] has published RDF data from institutions such as the ACM, Citeseer, IEEE and DBLP. Amongst these datasets the problem of author, as well as paper, disambiguation has proved a challenging task. Each repository has its own naming scheme for authors and publications and the overlap between each repository is significant. In order to have a Semantic Web that provides the scalability to cope with such inconsistencies the issue of coreference will have to be addressed. Our proposed solution for this problem is given later in the paper.

### **2.2 Coreference in Databases**

Within the database community the problem of coreference is referred to as record linkage. The need for record linkage arises when records or files from different databases need to be joined or merged. Each database could have duplicate records of the same person or thing which, when amalgamated, would make the data inconsistent or ‘dirty’ [13].

Record linkage has a well defined mathematical theory as proposed by Fellegi and Sunter [14]. The theory is based on records referring to the same entity having a number of characteristics in common. If  $a$  and  $b$  are elements from populations A and B and some elements are common to A and B then two disjoint sets can be created. The first set, M, is the set of elements that represent identical entities and the second set, U, is the set of elements representing different entities. Now if  $\alpha(a)$  and  $\beta(b)$  refer to records from databases A and B respectively and each record has  $k$  characteristics then a comparison vector,  $\gamma$  is defined that contains the coded agreements and disagreements on each characteristic:

$$\gamma[\alpha(a), \beta(b)] = \{\gamma^1[\alpha(a), \beta(b)], \dots, \gamma^k[\alpha(a), \beta(b)]\} \quad (1)$$

The theory then gives the probability of observing a specific vector given  $(a,b) \in M$  as:

$$M(\gamma) = \sum_{(a,b) \in M} P\{\gamma[\alpha(a), \beta(b)]\} \cdot P[(a,b) | M] \quad (2)$$

This theory can also be used for identifying the probability of having identical resources on the Semantic Web. If A and B are two RDF graphs which have URIs  $a$  and  $b$  identifying the same resource such that  $a \in A$  and  $b \in B$ , then when the following graphs of each URI are made:

$$\text{CONSTRUCT } \{ \langle \text{URI-1} \rangle \text{ ?ap ?ao} \} \text{ WHERE } \{ \langle \text{URI-1} \rangle \text{ ?ap ?ao} \} \quad (3)$$

$$\text{CONSTRUCT } \{ \langle \text{URI-2} \rangle \text{ ?bp ?bo} \} \text{ WHERE } \{ \langle \text{URI-2} \rangle \text{ ?bp ?bo} \} \quad (4)$$

$\alpha(a)$  can be substituted by the result ?ao and  $\beta(b)$  can be substituted by the result ?bo. The characteristics  $k^i$  can then be substituted by the union of the result of ?ap and ?bp. The equations (1) and (2) will then hold with these substituted values. The result only shows the probability of two URIs referring to the same entity and can only be used as a basis for coreference resolution as it has been used in the database community for the same purpose.

The problem with this theory is that it is assumed that there is at least one characteristic that is in common across the different graphs. In the Semantic Web this is not always the case as different graphs will have different predicates for a particular resource and there may be little or no overlap between predicates. For example the URIs <http://sws.geonames.org/2950159> and <http://dbpedia.org/resource/Berlin> are URIs for Berlin that each have over 100 predicates, yet only one of them is in common. There is also the problem that predicates in the Semantic Web are not like database field names that have common names. Therefore, ontology matching techniques need to be used in order to match predicates between different resources, so that characteristics (instances) can then be matched.

There are thus more advanced algorithms that need to be developed in order to identify identical URIs on the Semantic Web. However, the problem is not just of finding equivalences, but what to do when a candidate equivalence is found. This problem is discussed in Section 3.

### 2.3 Coreference in The Semantic Web

The subject of coreference on the Semantic Web has been raised previously [15], but it was not pressing or therefore studied for many years because of the lack of real scalable RDF data that was freely available. However, the Linking Open Data project and our own ReSIST project are highlighting the need to have some form of URI management system. For example, the following are all URIs for Spain:

```
http://dbpedia.org/resource/Spain
http://www4.wiwiss.fu-berlin.de/factbook/resource/Spain
http://sws.geonames.org/2510769/
http://www.daml.org/2001/09/countries/fips#SP
http://www4.wiwiss.fu-berlin.de/eurostat/resource/countries/Espa%C3%B1a
```

These URIs come from 5 different sources. There are also at least 9 URIs for Hugh Glaser that originate from 6 different sources:

```
http://acm.rkbexplorer.com/rdf/resource-P112732
http://citeseer.rkbexplorer.com/rdf/resource-CSP109020
http://citeseer.rkbexplorer.com/rdf/resource-CSP109013
http://citeseer.rkbexplorer.com/rdf/resource-CSP109011
http://citeseer.rkbexplorer.com/rdf/resource-CSP109002
http://dblp.rkbexplorer.com/rdf/resource-27de9959
http://europa.eu/People/#person-0ff816fa
http://resist.ecs.soton.ac.uk/wiki/User:hugh_glaser
http://www.ecs.soton.ac.uk/info/#person-00021
```

We have grouped these URIs together because we believe they all refer to the same non-information resource. However, the standard way of dealing with such a plethora of URIs is to use *owl:sameAs* to link between them. The semantics of *owl:sameAs* mean that all the URIs linked with this predicate have the same identity [16], this means that the subject and object must be the same resource. The major disadvantage with this approach is that the two URIs become indistinguishable even though they may refer to different entities according to the context in which they are used. For example, consider the case where a person has a URI at one institution and then moves to another institution that provides another URI. If the person makes an *owl:sameAs* link between them then it will not be possible to differentiate between the person as they were at the first institution and the person as they are at the second institution. The knowledge about the person at institution 1 and institution 2 effectively become merged so, for example, the addresses would not be able to be separated.

Even worse, an incorrect equivalence can cause other incorrect equivalences to be inferred. For example, we found that one of the project investigators (Tom Anderson) had extra information which appeared plausible, but was not correct. We finally tracked this information down to DBLP, where the two Tom Andersons had been conflated.

We subscribe to the belief that the meaning of a URI may change according to the context in which it is used [17]. For example the URIs that refer to Spain given above could refer to ‘Spain the political entity’, or ‘Spain the geographic location’, or ‘Spain the football team’. Some people would be happy to use each URI interchangeably

because they do not care about the precise definition, whereas others will want a URI that specifically matches their intended meaning. There is a requirement to have some form of a system that deals with URIs about the same resource that are not exactly identical. The semantics of *owl:sameAs* are too strong and other alternatives like *rdfs:seeAlso* do not fit the intended purpose. Such a requirement is vital if data is to be cleanly linked together in a consistent fashion. The next section details our initial attempt to handle URI management called the Consistent Reference Service (CRS).

### 3 The Consistent Reference Service

The Consistent Reference Service (CRS) has been created in order to manage coreference between the millions of URIs that are accumulating on the Semantic Web. This section will describe the concept of a *bundle* that groups together URIs referring to the same resource, and also describe the implementation and architecture of the CRS.

#### 3.1 URIs and Bundles

The CRS service has been implemented as both an RDF knowledge base and a relational database with RDF export. The CRS sits in the Semantic Web as any other knowledge base or database would. Each data provider maintains one or more CRSs for their own knowledge. In the ReSIST project there are over 15 repositories each with their own CRS.

The CRS introduces the concept of a *bundle* to group together resources that have been deemed to refer to the same concept within a given context. Different bundles may be used to group together URIs of the same resource in different contexts. For example, there may be a bundle containing all of the URIs about a person in the context of institution 1; and another bundle containing all of the URIs about the same person in the context of institution 2. Each CRS can use different algorithms to identify equivalent resources. For example, the algorithms to detect equivalence amongst authors are different from the algorithms used to detect equivalence between countries. To begin with, each URI in a repository has its own bundle in the CRS. When an equivalence is detected the bundles containing the URIs are merged together to create a new bundle. In this way successive iterations group together larger bundles, with each bundle having an anonymous URI.

The concept of a bundle is defined as a class in a coreference ontology used by the CRS. There is also a database schema that maps onto the ontology. Every resource that is defined as being of *rdf:type coref:Bundle* can have the following properties:

*coref:hasCanonicalReference* – One URI in a bundle can be made to be the canonical representation i.e. the preferred URI that one should use.

*coref:hasEquivalentReference* – The URIs in a bundle are grouped together using this predicate.

*coref:updatedOn* – The date of the last update to the bundle.

To illustrate let us take the example of the URIs referring to Hugh Glaser in the previous section. If we assume that we want to group together all the URIs that Citeseer has referring to Hugh then the triples asserted in RDF/XML format would look like:

```
<rdf:RDF xmlns:coref=http://www.resist.ecs.soton.ac.uk/ontology/coref#
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <coref:Bundle
    rdf:about="http://www.rkbexplorer.com/crs/coref#bundle1">
    <coref:hasEquivalentReference rdf:resource=
      "http://citeseer.rkbexplorer.com/rdf/resource-CSP109020"/>
    <coref:hasEquivalentReference rdf:resource=
      "http://citeseer.rkbexplorer.com/rdf/resource-CSP109013"/>
    <coref:hasEquivalentReference rdf:resource=
      "http://citeseer.rkbexplorer.com/rdf/resource-CSP109011"/>
    <coref:hasEquivalentReference rdf:resource=
      "http://citeseer.rkbexplorer.com/rdf/resource-CSP109002"/>
    <coref:hasCanonicalReference rdf:resource=
      "http://citeseer.rkbexplorer.com/rdf/resource-CSP109002"/>
  </coref:Bundle>
</rdf:RDF>
```

The bundle mechanism provides an easy method to manage URI identities without having to incorporate expensive inference mechanisms. When dereferencing a resolvable URI the RDF document returned contains additional predicates identifying CRS services that may provide further information regarding the resource. If the user wishes, then they can assert explicitly *owl:sameAs* or *rdfs:seeAlso* links between the equivalent URIs. The next section will look at how the CRS is used in conjunction with multiple knowledge bases and how bundles can be linked to other open data.

### 3.2 The CRS and Web Architecture

There have been many discussions in the Semantic Web community regarding the actual meaning of a URI. Does it refer to a sequence of bits? A Web page? A concept? These questions and others arising from the URI identity crisis [18] are outside the scope of this paper. We will use the definitions as given by the W3C Technical Architecture Group (TAG) to show how a coreference mechanism can be included in the current Semantic Web infrastructure.

The CRS can be treated as any other knowledge base, in that it contains knowledge about a particular URI. Our infrastructure implements the current best practice on how to serve linked data [6] and uses cool URIs [19]. As an example we will use the URI <http://southampton.rkbexplorer.com/id/person-21> to represent the non-information resource, ‘Hugh Glaser’. When a request is given to the server for a description of the URI, an HTTP 303 redirect is issued to one of two locations, depending on the accept headers sent by the client. If the requested content is application/rdf+xml then the server will generate an RDF description detailing the properties of the requested URI by issuing SPARQL CONSTRUCT queries to the

appropriate knowledge base. The resulting description is cached and the 303 is issued to <http://southampton.rkbexplorer.com/description/person-21>. However, if the accept header is set to text/html then a 303 See Other is returned identifying an html description of Hugh Glaser at <http://southampton.rkbexplorer.com/browse/peson-21>. The server architecture conforms to the latest httpRange-14 [20] recommendation of the TAG that involves HTTP 303 Redirects from the URI of a non-information source to an RDF or HTML information resource.

Each URI that is maintained by an institution will also have its own CRS. This CRS can be termed the ‘home’ CRS of a URI. The home CRS will provide a level of trust over what URIs it considers to be the same because it is the sole provider of that URI. To find all possible equivalences for a URI the following algorithm can be performed:

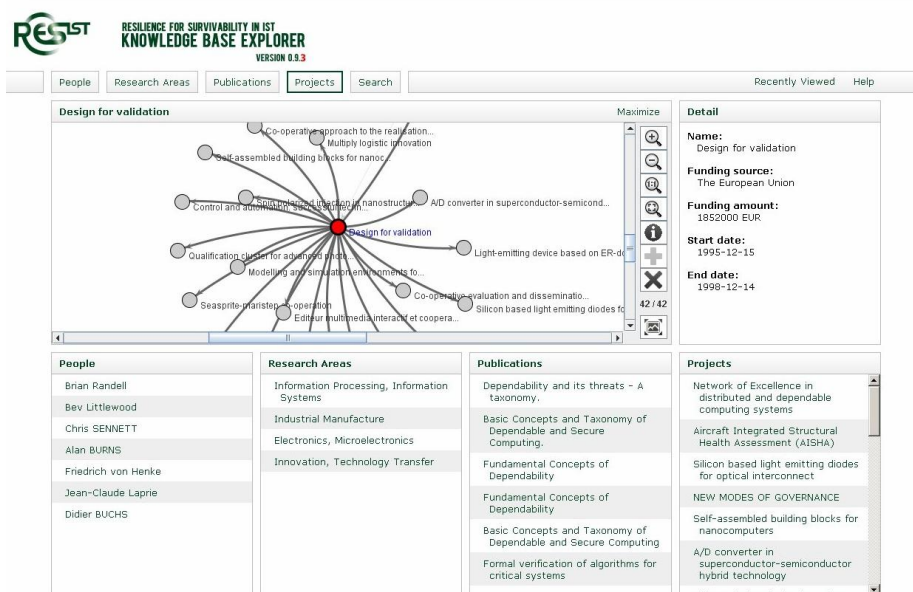
```
findEquivalence(URI u) {
  Dereference u;
  while (u coref:hasCRS a) {
    add a to equivalences;
    if(equivalences contains a)
      break;
  }
  findEquivalence(a);
}
```

Finding all equivalences is entirely at the discretion of the application wishing to process the results of the search. If only one CRS is required, then only one iteration is necessary. Computing equivalences in this manner gives a considerable amount of flexibility in choosing duplicate URIs for a resource. Taking the URI management into a separate layer without fixing *owl:sameAs* links is an efficient and controllable way to manage coreference between URIs. The next section will describe an application that has been built using the CRS infrastructure.

#### **4 A CRS Application: The Resilience Knowledge Base Explorer**

Resilience Knowledge Base (RKB) Explorer is a Semantic Web application that is able to present unified views of a significant number of heterogeneous data sources regarding a given domain. We have developed an underlying information infrastructure that utilises the CRS architecture outlined in Section 3. Our current dataset totals many tens of millions of triples, and is publicly available through both SPARQL endpoints and resolvable URIs. To realise the synergy of disparate information sources we are using the CRS system and have devised an architecture to allow the information to be represented and used.





**Fig. 1.** The figure above shows the single window interface of the faceted browser available at <http://www.rkbexplorer.com/explore/>

Figure 1 shows the user interface for the RKB Explorer. The main pane shows a chosen concept and related concepts of the same type that the system has identified as being related. In this figure, the ReSIST Project itself is under consideration, with its details on the right, and related projects are shown around it. These are chosen according to the relative weight given to ontological relationships, and the number of those relationships to each concept. The weight of the lines gives a visual ranking. They represent a project 'Community of Practice' (CoP) for the project. Clicking on a resource will show the detail for it, while double-clicking will add the CoP for the new resource to the pane. This will then allow a user to see how different projects are related, and see the projects that provide linkage between them.

The panes in the lower half of the display show the related people, research areas, publications and projects, identified by similar ontologically informed algorithms, and are ranked by decreasing relevance. Thus the lower right-hand pane gives a list of the related projects found in the main pane, while the lower left-hand pane shows those people involved in the currently selected project.

The RKB Explorer is based on the implementation described in Section 3. The CRS system manages the URIs for each knowledge base. There are many URIs from each knowledge base that refer to the same resource, for example there are hundreds of the same authors and papers in different knowledge bases, such as the ACM, IEEE and DBLP. Managing these millions of URIs has led to increased scalability and performance benefits as compared with taking an *owl:sameAs* approach. The RKB

Explorer is being expanded and integrated with existing linked data and it is envisioned that the CRS system behind the explorer will also follow the same route.

## 5 Summary and Future Work

This paper has shown that managing URI identities at Semantic Web scale is a serious issue and problem. The main aim is to stimulate debate and discussion about how URI identities need to be managed and dealt with. The issue is becoming all the more relevant with the linked data that is now available online as RDF. This has caused a real debate on how to manage coreference in the Semantic Web [21]. Our proposed solution is meant to highlight the issues involved in building a URI management system and how such a system can be used to handle millions of triples of RDF data.

Our future plan is to distribute all our knowledge bases such that the ownership is delegated to a third party, together with an accompanying CRS or several CRSs. The next stage of research will then highlight exactly what issues need to be resolved when linking and reasoning about such a highly distributed system. The idea is to be able to mimic the stability and robustness of the document Web to the data Web in order to make the Semantic Web a truly scalable system.

## References

1. Shadbolt, N. R., Gibbins, N., Glaser, H., Harris, S. and schraefel, m. c. (2004) CS AKTive Space or how we stopped worrying and learned to love the Semantic Web. *IEEE Intelligent Systems*.
2. Lei, Y., Uren, V.S. & Motta, E.. (2006) SemSearch: a search engine for the Semantic Web. Proceedings EKAW 2006, pp.238-245, Podebrady, Czech Republic.
3. Shirky, C., 2001. The Semantic Web, Syllogism, and Worldview [online], [http://www.shirky.com/writings/semantic\\_syllogism.html](http://www.shirky.com/writings/semantic_syllogism.html) [15 Feb, 2007]
4. Suchanek, F.M., Kasneci, G. & Weikum, G. 2007. YAGO:A Core of Semantic knowledge. Proceedings International WWW Conference 2007, ACM Press, pp.697-706, Banff, Alberta, Canada.
5. Linking Open Data Project <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
6. Bizer, C., Cyganiak, R. & Heath, T., How to Publish Linked Data on the Web, [online], <http://sites.wiwi.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/> [20 July 2007]
7. Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanara, R., Hollenbach, J., Lerer, A. & Sheets, D., 2006. Tabulator: Exploring and Analyzing Linked Data on the Web. Proceedings 3<sup>rd</sup> International Semantic Web User Interaction Workshop. Athens, Georgia, USA.
8. DBpedia [online] <http://dbpedia.org/docs> [1 July 2007]
9. Equivalence Mining and Matching Frameworks [online] <http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/EquivalenceMining> [1 July 2007]
10. Yang, K., Jiang, J., Lee, H. & Ho, J. Extracting Citation Relationships from Web Documents for Author Disambiguation, Technical Report No. TR-IIS-06-017, Institute of Information Science, Academia Sinica, Taipei, Taiwan, December 2006.

11. Fan-Tan, Y., Yen-Kan, M. & Lee, D. Search Engine Driven Author Disambiguation, Proceedings 6<sup>th</sup> ACM/IEEE-CS Joint Conference on Digital Libraries, pp.314-315, ACM Press, New York.
12. Resilience for Survivability in Information Society Technologies (ReSIST) Network of Excellence. <http://resist-noe.eu>
13. Hernandez, M. & Stolfo, S. Real-world Data is Dirty: Data Cleansing and the Merge/Purge Problem, Data Mining and Knowledge Discovery, 2(1), pp.9-37, Kluwer, Hingham, MA, USA
14. Fellegi, I.P. & Sunter, A.B. A Theory for Record Linkage, Journal of the American Statistical Association, 64(328), pp.1183-1210, December 1969
15. Alani, H., Dasmahapatra, S., Gibbins, N., Glaser, H., Harris, S., Kalfoglou, Y., O'Hara, K. and Shadbolt, N. (2002) Managing Reference: Ensuring Referential Integrity of Ontologies for the Semantic Web. In Proceedings of 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02), pp. 317-334, Sigenza, Spain.
16. Bechofer, S., Van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Schneider, P.F. & Stein, L.A. OWL Web Ontology Language Reference, Technical Report, W3C, [online] <http://www.w3.org/TR/owl-ref/>
17. Booth, D. URIs and the Myth of Resource Identity, Proceedings of the Workshop on Identity, Meaning and the Web (IMW06) at WWW2006, Edinburgh, Scotland.
18. Halpin, H. Identity, Reference and Meaning on the Web, Proceedings of the Workshop on Identity, Meaning and the Web (IMW06) at WWW2006, Edinburgh, Scotland.
19. Berners-Lee, T. Cool URIs Don't Change, [online] <http://www.w3.org/Provider/Style/URI>
20. Fielding, R., W3C Technical Architecture Group mailing list, June 18, 2005. [online] <http://lists.w3.org/Archives/Public/www-tag/2005Jun/0039>
21. W3C Mailing List Discussion Thread, Terminology Question Concerning Web Architecture and Linked Data, <http://lists.w3.org/Archives/Public/semantic-web/2007Jul/0049.html>