SCS Connector - Quantifying and Visualising Semantic Paths between Entity Pairs

Bernardo Pereira Nunes[1], José Herrera[1], Davide Taibi[2], Giseli Rabello Lopes[1], Marco A. Casanova[1], Stefan Dietze[3]

[1] Department of Informatics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro/RJ – Brazil, CEP 22451-900
{bnunes,jherrera,grlopes,casanova}@inf.puc-rio.br
[2] Italian National Research Council - Institute for Educational Technology, Palermo, Italy
davide.taibi@itd.cnr.it
[3] L3S Research Center, Leibniz University Hannover, Appelstr. 9a, 3016, Hannover, Germany
dietze@l3s.de

Abstract. A key challenge of the Semantic Web lies in the creation of semantic links between Web resources. The creation of links serves as a mean to semantically enrich Web resources, connecting disparate information sources and facilitating data reuse and sharing. As the amount of data on the Web is ever increasing, automated methods to unveil links between Web resources are required. In this paper, we introduce a tool, called *SCS Connector*, that assists users to uncover links between entity pairs within and across datasets. *SCS Connector* provides a Web-based user interface and a RESTful API that enable users to interactively visualise and analyse paths between an entity pair (e_i, e_j) through known links that can reveal meaningful relationships between (e_i, e_j) according to a semantic connectivity score (SCS).

Keywords: Semantic Connectivity Score, Graph Visualisation, Semantic Associations, Relationship Discovery, Semantic UI

1 Introduction

The adoption of Linked Data for publishing and interlinking structured data has brought a range of benefits to data providers and consumers such as data interoperability, reuse and sharing. However, as the amount of data on the Web is in constant growth and change, the links between Web resources become outdated and automated methods are required to create new links within and across datasets.

A key challenge in the provision of a well-interlinked graph of Web data lies in the identification and linkage of not only existing entities, but also in the interlinking of new entities. The linkage of entities provides data consumers with a richer representation of the data and the possibility of exploiting and uncovering information by traversing the Web of Data graph.

Over the past years most of the entity interlinking approaches has focused on recognising strict equivalences between entities through the creation of `owl:sameAs` links across datasets. Consequently, little attention has been drawn to identify related entities intra- and inter-datasets. For instance, by creating `skos:related` or `so:related` references between entities that are related at some extent.

In this paper, we present a tool, called *SCS Connector*, that is responsible for uncovering meaningful relationships between entity pairs within and across datasets. *SCS Connector* provides a Web-based user interface and a RESTful API that enable users to interactively visualise and analyse paths between an entity pair (e_i, e_j) through known links that can reveal some meaningful relationship between (e_i, e_j) according to a semantic connectivity score. The tool was developed based on our previous works [6, 5, 7] that introduced the *semantic connectivity score* (SCS) to measure the relatedness between entity pairs in reference datasets.

The remainder of the paper is organised as follows. Section 2 reviews the literature. Section 3 describes the most relevant features of the *SCS Connector* tool. Section 4 concludes the paper.

2 Related Work

RelFinder [4] is a tool that aims at finding relationships between a set of specified entities and providing a mechanism to explore the semantic links between entities. Similarly, OntoRelFinder [12] explores the semantic links between entity pairs but outperforms RelFinder since it relies on the schema paths to find the semantic links. Scarlet [10, 11] is another tool to identify semantic links that focuses on finding correspondences between entities belonging to a set of external ontologies. In a different perspective, Han et al. [3] focuses on finding related entities with respect to a given entity and semantic links.

The number of Web-based applications using semantic technologies to improve search, retrieval and recommendation of Web resources has dramatically increased. For instance, Passant [9, 8] introduces a recommender system that takes advantage of the semantic links between Web resources to recommend resources laterally related to a resource of interest to a given user. Another system is presented by Souvik et al. [1] that recommends movies based on a linear regression algorithm that uses a set of features to determine whether a movie is related (or not) to a given one.

SCS Connector differs from the related works outlined since it provides to the end-user a *semantic connectivity score* (SCS) (see Section 3.1) that measures how related an entity pair is. Furthermore, *SCS Connector* provides a user interface that allows users to explore the semantic paths between an entity pair as well as the adjacency matrix representing the entity pair graph, enabling users to compute new semantic measures to compare them against the SCS approach.

3 SCS Connector Tool

SCS Connector aims at finding and measuring semantic relationships between entity pairs. Uncovering semantic links between disparate entities open up new opportunities to link a variety of resources on the Web. In this section, we describe the most relevant features of *SCS Connector* along with technical explanations describing the strategies used to discover semantic paths between entity pairs.

3.1 Semantic Connectivity Score (SCS)

Let $G = (E, P)$ be an RDF graph, where E and P are a finite set of entities and properties, respectively. A property $p_i \in P$ represents a link between a pair of entities (e_i, e_j), where e_i and $e_j \in E$. Thus, given an entity pair (e_1, e_n), we say that they are related iff there exists at least one path connecting them. For instance, there is one path connecting (e_1, e_n) represented by $\rho_1(e_1, e_n) = \{(e_1, e_2), (e_2, e_3), \ldots, (e_{n-1}, e_n)\}$. Note that, as properties are often found in its inverse form [2], we consider the graph G as undirected, where $\rho_1(e_1, e_n) = \rho_1(e_n, e_1)$.

Thus, to quantify the relationships and possibly uncover semantic links between entity pairs, our tool uses the *semantic connectivity score* (SCS) previously introduced in [5]. The score function SCS between a pair of entities (e_i, e_j) is computed by equation 1 that considers the semantic paths found $\rho_1(e_i, e_j), \rho_2(e_i, e_j), \ldots, \rho_m(e_i, e_j)$, where m is the total number of paths having a defined maximum length. Although in [5] is established a maximum path length in which the score will be computed, in our tool, it can be freely specified by the user. A semantic relationship exists iff $SCS(e_i, e_j) > 0$. Moreover, if $e_i = e_j$, then $SCS(e_i, e_j) = 1$.

$$SCS(e_i, e_j) = 1 - \frac{1}{1 + (\sum_{l=1}^{\tau} \beta^l \cdot |paths_{(e_i,e_j)}^{<l>}|)} \tag{1}$$

where $|paths_{(e_i,e_j)}^{<l>}|$ is the number of paths of length l between entities e_i and e_j, τ is the maximum length of paths $\rho_k(e_i, e_j)$ considered (default is $\tau = 4$), and $0 < \beta \leq 1$ is a positive damping factor. The damping factor β^l is responsible for exponentially penalising longer paths. The final score is normalised to range between $[0, 1)$.

3.2 Finding Paths between Entity Pairs

To find paths between two specified entities, we adopted a similar preprocessing strategy applied by RelFinder [4]. Briefly, the preprocessing strategy computes the maximal connected subgraphs of G (see Section 3.1). Thus, only entities in the same connected subgraph have a (semantic) path linking them. In this case, SPARQL queries are issued to the RDF graph (in our case, DBpedia graph) to retrieve all semantic paths connecting two entities up to a maximum predefined length. Although, by default, the maximum path length is set to 4, the user can set this parameter to search for shorter/longer paths.

As a running example, suppose that a user wants to find paths between the entities `dbpedia:Barack_Obama` and `dbpedia:Michelle_Obama` and sets the maximum path length to 2. Thus, to find all paths up to a maximum length, the tool starts searching for direct links between the specified entities using the SPARQL queries below:

```
SELECT * WHERE {dbpedia:Barack_Obama ?p dbpedia:Michelle_Obama}
SELECT * WHERE {dbpedia:Michelle_Obama ?p dbpedia:Barack_Obama}
```

To find the paths with maximum length 2, *SCS Connector* issues the following queries:

```
SELECT * WHERE {dbpedia:Barack_Obama ?p1 ?e . ?e ?p2 dbpedia:Michelle_Obama}
SELECT * WHERE {dbpedia:Michelle_Obama ?p1 ?e . ?e ?p2 dbpedia:Barack_Obama}
SELECT * WHERE {dbpedia:Barack_Obama ?p1 ?e . dbpedia:Michelle_Obama ?p2 ?e}
SELECT * WHERE {?e ?p1 dbpedia:Barack_Obama . ?e ?p2 dbpedia:Michelle_Obama}
```

Note that `?e` and `?p1,?p2` are used to find semantic paths of length 2 and represents entities and properties, respectively. In this example, we omitted the clause FILTER from the SPARQL queries. However, *SCS Connector* also allows user to filter out some ontological classes from the paths linking an entity pair (e_i, e_j).

Finally, according to the existing paths between the entities (e_i, e_j), a $SCSscore$ is assigned representing how linked both entities are. If there is no direct links between the specified entity pair and the $SCSscore$ is above a given threshold, there is a high chance to exist a new link between the entity pairs.

3.3 Exposing the Paths between Entity Pairs

SCS Connector is a flexible cross-browser Web application implemented in PHP and Java to interactively visualise and analyse the semantic paths between entity pairs. The semantic paths are exposed through a *graph visualisation* and a *RESTful API* service.

Graph visualisation. The paths found (see Section 3.2) are rendered and displayed in the form of a graph, where the vertices and edges are represented by entities and properties, respectively. The paths are also represented and displayed as a squared adjacency matrix, where the entries of the matrix determines whether there is an edge linking entities or not. The $SCSscore$ quantifies how related the entity pair is.

RESTful API. SCS Connector also provides a RESTful API to support the development of new data interlinking approaches. The API is available in two formats: JSON and XML. The API can be accessed via REST at `http://lod2.inf.puc-rio.br/scs/similarities.json?entity1=db:Barack_Obama&entity2=db:Michelle_Obama`.

4 Conclusions

This paper introduced SCS Connector, a Web-based application to assist users on the discovery of semantic links between entity pairs. SCS Connector enables users to interactively visualise and analyse the semantic paths and also use the results obtained to develop further services. The tool is available at `http://research.ccead.puc-rio.br/scs`.

Acknowledgement. This work was partly supported by CNPq, under grants 160326/2012-5, 301497/2006-0, 475717/2011-2 and 57128/2009-9, by FAPERJ, under grants E-26/170028/2008 and E-26/103.070/2011.

References

1. S. Debnath, N. Ganguly, and P. Mitra. Feature weighting in content based recommendation system using social network analysis. In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 1041–1042, New York, NY, USA, 2008. ACM.
2. A. Graves, S. Adali, and J. Hendler. A method to rank nodes in an rdf graph. In *Proceedings of the 7th International Semantic Web Conference (posters and demos), Karlsruhe, Germany, October 28, 2008*, volume 401 of *CEUR Workshop*. CEUR-WS.org, 2008.
3. Y.-J. Han, S.-B. Park, S.-J. Lee, S. Y. Park, and K. Y. Kim. Ranking entities similar to an entity for a given relationship. In *Proceedings of the 11th PRICAI*, PRICAI'10, pages 409–420, Berlin, Heidelberg, 2010. Springer-Verlag.
4. J. Lehmann, J. Schüppel, and S. Auer. Discovering unknown connections - the dbpedia relationship finder. In *CSSW*, pages 99–110, 2007.
5. B. P. Nunes, S. Dietze, M. A. Casanova, R. Kawase, B. Fetahu, and W. Nejdl. Combining a co-occurrence-based and a semantic measure for entity linking. In *The Semantic Web: Semantics and Big Data (ESWC 2013)*, volume 7882 of *LNCS*, pages 548–562. Springer Berlin Heidelberg, 2013.
6. B. P. Nunes, R. Kawase, S. Dietze, D. Taibi, M. A. Casanova, and W. Nejdl. Can Entities be Friends? In *Proceedings of the WOLE Workshop in conjuction with the 11th ISWC*, volume 906 of *CEUR-WS.org*, pages 45–57, Nov. 2012.
7. B. P. Nunes, R. Kawase, B. Fetahu, S. Dietze, M. A. Casanova, and D. Maynard. Interlinking documents based on semantic graphs. *Procedia Computer Science*, 22(0):231 – 240, 2013. In Proc. of the 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - {KES2013}.
8. A. Passant. dbrec - music recommendations using dbpedia. In *International Semantic Web Conference (2)*, volume 6497 of *LNCS*, pages 209–224. Springer, 2010.
9. A. Passant. Measuring semantic distance on linking data and using it for resources recommendations. In *AAAI Spring Symposium: Linked Data Meets AI*, 2010.
10. M. Sabou, M. d'Aquin, and E. Motta. Exploring the semantic web as background knowledge for ontology matching. *J. Data Semantics*, 11:156–190, 2008.
11. M. Sabou, M. d'Aquin, and E. Motta. Relation discovery from the semantic web. In *Proceedings of the 7th ISWC (posters and demos)*, volume 401 of *CEUR Workshop*. CEUR-WS.org, 2008.
12. D. Seo, H. Koo, S. Lee, P. Kim, H. Jung, and W.-K. Sung. Efficient finding relationship between individuals in a mass ontology database. In *FGIT-UNESST*, pages 281–286, 2011.