Everything is Connected: Using Linked Data for Multimedia Narration of Connections between Concepts*

Miel Vander Sande, Ruben Verborgh, Sam Coppens, Tom De Nies, Pedro Debevere, Laurens De Vocht, Pieterjan De Potter, Davy Van Deursen, Erik Mannens, and Rik Van de Walle

Ghent University - IBBT
Department of Electronics and Information Systems - Multimedia Lab
Gaston Crommenlaan 8 bus 201, B-9050 Ledeberg-Ghent, Belgium
firstname.lastname@ugent.be

Abstract. This paper introduces a Linked Data application for automatically generating a story between two concepts in the Web of Data, based on formally described links. A path between two concepts is obtained by querying multiple linked open datasets; the path is then enriched with multimedia presentation material for each node in order to obtain a full multimedia presentation of the found path.

1 Introduction

Interconnectedness is a major topic discussed in philosophy and physics, stating that everything in the world is connected in some way. A popular example is the Six degrees of Separation principle, introduced by S. Milgram [5]. It states that human networks grow in density despite the big physical distances, resulting in a maximal connection of five acquaintances between two people in the whole world. Between academia, this concept is often discussed for not being really scientifically proven [3]. However, with the Web of Data emerging, links between things become well defined and more known [2].

This demonstration relies on Linked Data to illustrate the above described principle (i.e. interconnectedness) by following machine-processable links within today's Web of Linked Data. The presented application will retrieve the connection (with intervening nodes) between two given concepts. Next, for each resource corresponding to a node in the found path, a media presentation is generated (e.g. use of a photo to illustrate a person). Finally, the found path is narrated through a full multimedia presentation, illustrating the intervening resources and their connections (i.e. the links). The aim of this demo is therefore threefold:

- bringing Linked Data closer to inexperienced users;
- a practical test of the link quality in today's Web of Data;
- investigate query/browse strategies for distributed path retrieval.

* The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

2 The story between two concepts

The goal of the demo is to automatically generate a story between two concepts, based on formally described links. We obtain such a story in three steps: path generation, node illustration, and path presentation.

2.1 Path generation

Path generation is based on querying Linked Data sources. In this demo, we consider the following linked datasets to obtain a path: DBpedia, Facebook, Freebase, Wiktionary, WordNet, CIA Factbook, Eurostat, GeoNames, MusicBrainz, linkedMDB, and New York Times. DBpedia serves as the interconnection hub between the other datasets; using Facebook adds a personalized touch to the application. Note that we use standard Named Entity and Disambiguation (NERD) techniques to link output from the Open Graph Protocol to DBpedia resources. Also note that we assume that the two input concepts (i.e. the start- and endpoint) are located within our predefined list of data sources.
Since we do not want to load all our datasets into one RDF store for scalability reasons, we introduce *progressive parallel link browsing*. The latter tries to find paths between two resources by combining the following tasks:
(1) find a path between the start resource and a DBpedia resource (DBpedia start resource);
(2) find a path between the end resource and a DBpedia resource (DBpedia end resource);
(3) find a path between the DBpedia start resource and the DBpedia end resource (either inside DBpedia or via another dataset).
This approach enables parallel execution by making the first task independent from the last. Also, since the nodes in the path are discovered sequentially, progressive playback is possible during discovery.
The choice of which links to follow is crucial in finding meaningful paths for the end-user. Therefore, a property prioritization mechanism is introduced. Certain properties (such as `rdfs:seeAlso`, `foaf:primaryTopic`, `dbpedia:wikiPageWikiLink`) contain little semantics in order to explain the link between two resources. Further, following `rdf:type` or `dcterms:subject` might result in links that are too general (e.g. both resources are `owl:Thing`s). However, when the type is specific enough (e.g., `Category:FC_Bayern_Munich_players`), it can be an interesting link. Therefore, a priority list is needed indicating which properties and classes/categories should be avoided during querying. Note that these 'blacklists' might still be used in fallback scenarios when the original attempt did not give any result.

2.2 Node illustration

Given a path between start and end resource, illustrations of the nodes need to be found in order to enable a path presentation (see further). These illustrations can be text, video, audio or images. A first attempt to provide a human-readable representation of a resource is to follow your nose: lookup the RDF properties of the resources and search for illustrative properties:

- text: `rdfs:label`, `dbpedia-owl:abstract`, `rdfs:comment`, ...;
- image: `foaf:depiction`, `dbpprop:picture`, `lode:illustrate`, ...;
- video/audio: `lode:illustrate`, ...

Alternatively, we use Web services to look up images (e.g. Flickr, Google Images, Bing Images, ImageNet), videos (e.g. YouTube, Vimeo), and audio (e.g. Grooveshark) to provide illustrative material for the resources.

2.3 Path presentation

As a final step, we present the retrieved path by means of a multimedia presentation with voice over. This requires a fluent narrated story, assembled from two types of data: topic descriptions and links. Topic descriptions are pieces of text fetched from certain RDF properties (see previous section) and give a short summary about the current topic. Links between nodes glue these pieces together into one story. However, narrating links is more difficult, since there is no descriptive text available. Different approaches are possible: from a simple display of the property label (e.g. `dbpedia-owl:birthDate` becomes 'has birth date') to a property-specific illustrator. The current version of the demo is limited to the simple property label approach.
Using the Acapela text-to-speech service[1], we are able to provide a voice over narrating the textual representation of nodes and links, making the multimedia presentation complete.

3 Implementation

The application runs on a classic client-server architecture where the path generation step is executed on the server, while node illustration and path presentation are performed at the client (i.e. in the browser). A number of performance optimizations were implemented:

- where possible, queries during the path generation are run in parallel;
- when parts of a path are found, these are already send back to the client;
- the client will run the node illustration tasks in the background (i.e. during presentation of the first nodes, information about the following nodes will be fetched).

4 Example scenario

In order to illustrate the idea, we will connect the dummy Facebook user *John Doe* to *Chicago*. Using the Facebook Open Graph protocol, we find out that he likes David Guetta (`ogp:userA ogp:likes dbpedia:David_Guetta`). By looking up `dbpedia:David_Guetta`, we find his birthplace to

[1] http://www.acapela-vaas.com

be Paris (dbpedia:David_Guetta dbpprop:birthPlace dbpedia:Paris). We derive from this resource that it is the location of the Eiffel tower (dbpedia:Eiffel_Tower dbpedia-owl:location dbpedia:Paris). One of the designers of the Eiffel tower was Émile Nouguier who died in 1898-02-20 (dbpedia:Emile_Nouguier dbpedia-owl:significantBuilding dbpedia:Eiffel_Tower; dbpprop:deathDate "1898-02-20"). The latter is also the birthdate of Jimmy Yancey, who died in Chicago (dbpedia:Jimmy_Yancey dbpprop:birthDate "1898-02-20"; dbpprop:deathPlace dbpedia:Chicago.).
A screencast of the Everything is Connected application is available at http://youtu.be/FawygFT5Brs. In this screencast, we show the resulting multimedia presentation for the above described example. Multimedia content showed in this example comes from Google Images, Google Maps, and YouTube.

5 Related Work

Concerning path generation, some work is done in the field of Social Network search. Given the *degrees of separation* principle, [1] studies changes made by today's popularity of social networks. New search techniques are introduced to browse dynamic Internet-based graphs, resulting in better memory-based search. Similar bidirectional search algorithms are introduced, but only consider the narrow Twitter data domain.
In [4], West et al. perform a large scale study on human navigation behaviour in Wikipedia. They examine the paths followed by users to get from one topic to another, by listing all the visited topics in between. From this extracted knowledge, a probabilistic model is derived for predicting what users are looking for. For characterising the types of topics a user will cross, a classification is introduced based on the connectedness of a topic.
For the Path presentation, QWiki[2] produces similar output. However, this tool is limited to turning any concept found on Wikipedia into a automatically narrated multimedia presentation based on the Wikipedia abstract.

References

1. Reza Bakhshandeh, Mehdi Samadi, Zohreh Azimifar, and Jonathan Schaeffer. Degrees of separation in social networks. In *SOCS*, 2011.
2. Chris Bizer, Anja Jentzsch, and Richard Cyganiak. State of the LOD Cloud. http://www4.wiwiss.fu-berlin.de/lodcloud/state/.
3. Judith Kleinfeld. Could It Be A Big World After All? The "Six Degrees of Separation" Myth. *Society*, 2002.
4. Jure Leskovec. Human wayfinding in information networks. In *In WWW-12*, 2012.
5. S. Milgram. The small world problem. *Psychology Today*, 61:60–67, 1967.

[2] Example for the topic Giraffe - http://www.qwiki.com/q/Giraffe