zLinks: Semantic Framework for Invoking Contextual Linked Data

Michael K. Bergman
Zitgist LLC
Coralville, IA USA

mike@mkbergman.com

Frédérick Giasson
Zitgist LLC
Quebec City, Quebec Canada

fred@fgiasson.com

ABSTRACT

This first-ever demonstration of the new zLinks plug-in shows how any existing Web document link can be automatically transformed into a portal to relevant Linked Data. Each existing link disambiguates to its contextual and relevant subject concept (SC) or named entity (NE). The SCs are grounded in the OpenCyc knowledge base, supplemented by aliases and WordNet synsets to aid disambiguation. The NEs are drawn from Wikipedia as processed via YAGO, and other online fact-based repositories. The UMBEL ontology basis to this framework offers significant further advantages. The zLinks popup is invoked only as desired via unobtrusive user interface cues.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering, Query formulation; H.5.4 [**Hypertext/Hypermedia**]: Navigation; H.5.2 [**User Interfaces**]: Interaction styles.

Keywords

demo, Linked Data, zLinks, OpenCyc, Wikipedia, WordNet, YAGO, UMBEL.

1. INTRODUCTION

Linked Data [1] follows recommended practices for identifying, exposing and connecting data on the semantic Web. A robust Linked Open Data (LOD)[1] community has rapidly developed around the practice with literally billions of compliant data items now available.

A notable catalyst to the Linked Data movement has been DBpedia [2], which exposes Wikipedia data in best-practices format. It is appropriate that the flagship figure showing the interrelationships of many Linked Data sources has DBpedia positioned at its core [3].

However, the Linked Data community readily acknowledges that the existing semantics and basis for relating compliant datasets are relatively poor. Moreover, there presently are no techniques or methods for relating non-Linked Data to the rapidly growing storehouse of LOD-compliant datasets.

The newest release of the zLinks plug-in (see Figure 1) and its supporting server-side infrastructure directly addresses the issues of improved semantics for Linked Data matching and relating Linked Data with standard Web content. Our demonstration shows how normal hyperlinks in standard WordPress blogs can be automatically related to contextually relevant Linked Data.

This précis first describes the technical underpinnings to zLinks' semantic framework, then overviews the application and possible future directions.

2. SEMANTIC FRAMEWORK

W-O-W-Y is the term we apply to the semantic framework for relating a given hypertext link to its relevant Linked Data. The term is derived from the constituent resources of WordNet[2] (W), OpenCyc[3] (O), Wikipedia[4] (W) and YAGO [4] (Y). Via the WOWY framework, we first determine if the link refers to either a named entity (NE) or a subject concept (SC) as well as to disambiguate alternate senses. If an NE, the entity is also related to its parent subject concept; every link thus has a SC basis.

All canonical SCs are embedded in a subject structure ontology – the "backbone". Use of this ontology brings inference and other relationship advantages. These various semantic frameworks are described below.
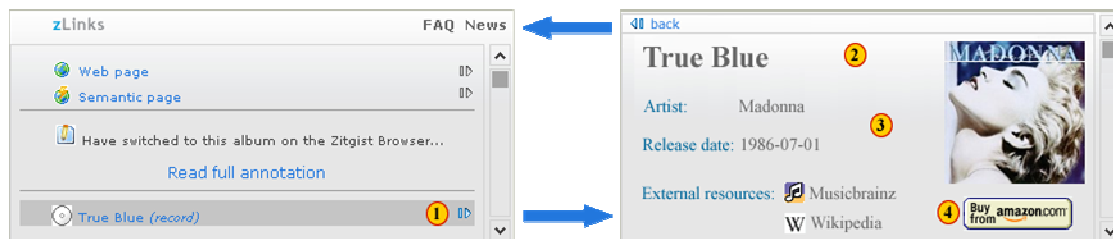

Figure 1. zLinks toggle popup [to be updated before final submission]

2.1 Subject Concepts

Subject concepts (SCs) are the core constituents to the framework. All SCs are based on existing concepts in OpenCyc, the open source version of the Cyc [5] knowledge base. SCs are the concrete, non-abstract topic-related classes within Cyc. About 22,000 of them were vetted from Cyc (paper in preparation). Aliases for these concepts as maintained by Cyc were combined with matching WordNet synsets to produce the SC disambiguation lexicon.

2.2 Named Entities

The named entities (NE) are drawn from Wikipedia as processed via YAGO, and other online fact-based repositories. NEs are the instances of the SC classes in the standard definition of the term[5]. NEs also have aliases for disambiguation purposes (such as the many ways to refer to the "United States").

Each NE is mapped to a governing SC for ontology purposes.

2.3 UMBEL Ontology

All of the SCs are expressed in the UMBEL (Upper-level Mapping and Binding Exchange Layer)[6] ontology. UMBEL is a lightweight structure of subject concepts and their semantic relationships. There is a direct overlap of UMBEL subject concepts to a subset of class concepts within OpenCyc.

Quick relations can be determined from UMBEL for a given SC; more involved inferencing can be directed to OpenCyc.

Thus, via these semantic relationships, other relations such as parent concepts, domains, various entity types, and similar relationships can be obtained once a given SC is identified.

3. APPLICATION

This semantic framework is applied on the server-side once a given standard link is processed. The client-side zLinks plug-in provides the user interface, initial link extraction and results reporting to the user.

3.1 Plug-in Design

The demo is provided as a standard WordPress[7] PHP plug-in, with many options parameterized.

3.2 Snippet Evaluation

A "snippet" of text based on a word window or sentence surrounding the target link is extracted, parsed, filtered and then submitted to the server for sense evaluation. We use a variant of a graph-based disambiguation algorithm suited for use with large knowledge lexicons [6].

This extraction process can also result in issued queries to standard Web search services.

3.3 User Cues

The semantic result of this link evaluation – the individual zLink – is presented to the user via a subtle, small icon. The results are only presented to the user after a mouseover with set delay, to ensure the popup is purposefully desired and unobtrusive.

---

4 http://en.wikipedia.org

5 See http://en.wikipedia.org/wiki/Named_entity_recognition

6 http://www.umbel.org

7 http://www.wordpress.org

3.4 Popup

A popup presents the contextual zLinks results (see Figure 1), with toggle views presented as substitution overlays to preserve screen real estate. Because of the multiple relations possible, there are multiple contextual overlays. The basic results paradigm is taken from Zitgist's related DataViewer for RDF data.[8]

4. FUTURE DIRECTIONS

This initial zLinks design is but a mere taste of the possibilities with Linked Data from the twin perspectives of additional relationships and presentation templates. We expect rapid developments in both areas.

A generalization of the plug-in architecture will enable extension to other user-content platforms. Still further extending this generalization to the Web browser would bring zLinks capabilities to every Internet user for all existing Web content.

The basic zLinks design also lends itself to incorporating additional sources of named entity lookups and Linked Data.

5. CONCLUSIONS

Linked Data has been a triggering event in the nascent emergence of the semantic Web. We expect to see similar innovations to zLinks – such as Wikify[9] – emerge to test out different paradigms and interfaces for how best to exploit a Web of Data.

An older zLinks prototype may be found at http://zlinks.zitgist.com/; an updated demo based on the version herein will be posted shortly after the presentation.

6. REFERENCES

[1] T. Berners-Lee. Design Issues–Linked Data. Published online, May 2007. http://www.w3.org/DesignIssues/LinkedData.html

[2] S. Auer, C. Bizer, J. Lehmann, G. Kobilarov, R. Cyganiak, and Z. Ives. DBpedia: A nucleus for a web of open data. In *Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), Busan, South Korea*, volume 4825 of *LNCS*, pages 715–728, November 2007.

[3] R. Cyganiak. The Linking Open Data dataset cloud. Figure published and maintained online; version December 2007. http://richard.cyganiak.de/2007/10/lod/

[4] F. Suchanek, G. Kasneci, and G. Weikum. Yago: a core of semantic knowledge. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 697–706, New York, NY, USA, 2007.

[5] D. B. Lenat. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38, no. 11, November 1995.

[6] R. Mihalcea and A. Csomai. Wikify! Linking documents to encyclopedic knowledge. In *Proceedings of the Sixteenth ACM conference on Conference on Information and Knowledge Management (CIKM '07)* , pp. 233-241, November 6-8, 2007.

---

8 http://dataviewer.zitgist.com

9 http://www.wikifyer.com/