Publishing and Linking WordNet using lemon and RDF

John P. McCrae*, Christiane Fellbaum† and Philipp Cimiano*

*University Bielefeld, †Princeton University
CITEC Building, Inspiration 1, 33615 Bielefeld; 35 Olden Street, Princeton
{jmccrae, cimiano}@cit-ec.uni-bielefeld.de, fellbaum@princeton.edu

Abstract

Each article must include an abstract of 150 to 200 words in Times 9 pt with interlinear spacing of 10 pt. The heading Abstract should be centred, font Times 10 bold. This short abstract will also be used for printing a Booklet of Abstracts containing the abstracts of all papers presented at the Conference.

Keywords: keyword A, keyword B, keyword C

1. Introduction

WordNet is one the first and still most widely used resources for natural language processing. In the time since the first version of WordNet was released many resources have been produced that represent complementary information(**?**; **?**) or extend the WordNet model to new languages(**?**; **?**). In contrast, in recent years we have seen the development of Web Technologies for the representation of language resources and in particular, the use of linked data. This has lead to a linguistic linked open data cloud, which is constructed by linking resources and publishing them on the web using RDF. Linked data, as proposed by Berners-Lee(**?**), has four main principles for publishing data: firstly, the use of URIs to identify objects; secondly, that these URIs should be resolvable; thirdly, that semantic information is returned, using standards such as RDF and finally, that links are provided to other resources. Chiarcos *et al.*(**?**) discuss the application of this to linguistic data and show that this model has notable advantages over standard approaches to data modelling, in particular they outline the following:

1. Representation: Graph-based models are a method that can represent any form of language resource.
2. Structural interoperability: By using RDF graphs and URIs datasets can be merged with no effort.
3. Federation: Multiple datasets can easily be drawn from different sources in the web and used together seamlessly.
4. Conceptual interoperability: Linking to common data category repositories allows common definitions to be inferred.
5. Ecosystem: Building on standards such as RDF, allows the use of common tools, including databases.
6. Dynamic import: Data on the web is not static and as such errors can be corrected after publication.
7. Expressivity: The use of other Semantic Web models allows the easy expression of metadata, provenance and ontological constraints on the data

In this paper we describe our experience in publishing WordNet following the linked data principles. While this is not the first version of WordNet to be published as linked data (**?**; **?**; **?**), this version has several advantages, firstly that it is well-linked to many resources, secondly it uses an open model in *lemon* and finally, that as it is integrated with the development of WordNet, and as such will be updated alongside future releases of WordNet.

2. Background

2.1. WordNet

2.2. lemon

2.3. Linguistic Linked Data

3. Representing WordNets with lemon

4. Linking WordNet

5. Related Work

6. Conclusion