Technical Report

UMBEL Ontology

Volume 2

Subject Concepts and
Named Entities Instantiation

TR 08-07-16-A2

July 2008

Acknowledgements

The UMBEL project would like to thank Zitgist LLC for its generous donation of time and resources in programming and writing the documentation for this Technical Report.

UMBEL would also like to thank Cycorp for its support and preparation of a more current OWL version of the OpenCyc knowledge base. We would especially like to thank Larry Lefkowitz for his internal advocacy and answering many questions.

The Cyc Foundation, notably Mark Baltzegar, has been instrumental in helping to guide us through OpenCyc and to share with us many Foundation resources and projects currently in progress. The effort to date would not have been possible without this assistance.

<div align="right">

– *Michael K. Bergman*, editor
– *Frédérick Giasson*, editor

</div>

UMBEL (Upper Mapping and Binding Exchange Layer) *is a lightweight ontology structure for relating Web content and data to a standard set of subject concepts. Its purpose is to provide a fixed set of reference points in a global knowledge space. These subject concepts have defined relationships between them, and can act as binding or attachment points for any Web content or data.*

Connecting to the UMBEL structure thus provides **context** *to Web data. In this manner, Web data can be linked, made interoperable, and more easily navigated and discovered. The project Web site is at* http://www.umbel.org*.*

UMBEL defines "subject concepts" as a distinct subset of the more broadly understood concept *such as used in the SKOS RDFS controlled vocabulary or formal concept analysis or the very general concepts common to some upper ontologies. Subject concepts are a special kind of concept: ones that are concrete, subject-related and non-abstract. We further contrast these with named entities, which are the real things or instances in the world that are members of these subject concept classes. The UMBEL "backbone" is this set of reference subject concepts.*

Table of Contents

THIS VERSION AND RELATED VOLUMES

This technical documentation is for UMBEL version 0.70, released July 16, 2008. This is the first release of the ontology.

The UMBEL ontology is documented in two volumes:

- *UMBEL Ontology, Vol. 1: Technical Documentation*, **TR 08-07-16-A1**, that overviews the ontology schema, vocabulary and use, and
- *UMBEL Ontology, Vol. 2: Subject Concepts and Named Entities Instantiation*, **TR 08-07-16-A2**, which is an explanation of the N3 files in the ontology distribution.

In addition, the ontology documentation itself is also accompanied by a three-volume series describing the selection and vetting of UMBEL's 20,000 subject concepts from OpenCyc.[1] [†] These three volumes are:

- *Distilling Subject Concepts from OpenCyc, Vol. 1: Overview and Methodology*, **TR 08-07-16-B1**, the basic introduction and explanation of terminology and the distillation process
- *Distilling Subject Concepts from OpenCyc, Vol. 2: Files Documentation*, **TR 08-07-16-B2**, the listing and description of the various files accompanying this process, and
- *Distilling Subject Concepts from OpenCyc, Vol. 3: Appendices*, **TR 08-07-16-B3**, supporting materials and detailed backup.

UMBEL INSTANTIATION FILES BUNDLE

This document describes the files that are part of the UMBEL Ontology. These files serialize the description of instances of subject concepts, abstract concepts and named entities. They also outline al the existing subject concepts and abstract concepts.

Some of the files help people to visualize the structure of UMBEL by using Cytoscape. Other files describe the ontology and its links between external ontologies classes and UMBEL subject concepts.

[†] All references are provided in the concluding Endnotes section.

Access Directory

All files for the current bundle distribution of UMBEL may be accessed under this root directory:

- http://umbel.org/ontology/

All archived bundle distributions of UMBEL may be accessed under this form of root directory:

- http://umbel.org/ontology/v*nnn*.

Where v*nnn* represents the version number of the archived distribution (such as "v070").

Naming Conventions

- All files names that belong to the file bundle begin with "umbel_"
- The files that end with ".n3" are RDF files serialized in N3[2]
- The files that end by ".csv" are comma-separated values
- The version of the files within the bundle are denoted using "_v*nnn*"; for example: "_v050", "_v064" and "_v070"
- Lists of Subject Concepts are denoted by "_subject_concepts_"
- Lists of Abstract Concepts are denoted by "_abstract_concepts_", and
- Lists of Named Entities are denoted by "_nes_".

Ontology File

- *umbel.n3* – describes all of the classes and properties created for the UMBEL Ontology; that is, the basic UMBEL vocabulary and formal specification.

Concepts Files

- *umbel_subject_concepts_vnnn.csv* – lists all of the subject concepts that create the UMBEL structure. This file is intended for manual or human review of the concepts
- *umbel_abstract_concepts_vnnn.csv* – lists all of the abstract concepts that create the UMBEL structure. This file is intended for manual or human review of the concepts
- *umbel_subject_concepts_vnnn.n3* – instantiates all of the subject concepts that belong to UMBEL using the RDF/N3 serialization
- *umbel_abstract_concepts_vnnn.n3* – instantiates all of the abstract concepts that belong to UMBEL using RDF/N3 serialization.

Named Entities Files

- *umbel_yago_nes_vnnn.*.n3* – instantiates all of the named entities that belong to an UMBEL named entities dictionary, in RDF/N3, with the source being the Yago[3] dataset
- *umbel_john_peel_session_nes_vnnn.n3* – instantiates all of the named entities that belong to a UMBEL named entities dictionary, in RDF/N3, with the source being the John Peel Sessions[4] dataset.

External Linkages Files

- *umbel_external_ontologies_linkage_vnnn.n3* – lists all of the triples that link UMBEL subject concepts to external ontologies classes.

Cyctoscape Files

Please note that Appendix G in *Distilling Subject Concepts from OpenCyc, Vol. 3: Appendices*, **TR 08-07-16-B3**, provides installation and use tips for Cytoscape.

- *umbel_cytoscape_vnnn.csv* – lists all the nodes and arcs to import into Cytoscape[5] to visualize the UMBEL graph
- *umbel_cytoscape_vnnn.cys* – a pre-prepared input file to Cytoscape that includes a force-directed layout of the UMBEL subject concept graph; note the slight change in the file extension; this is the file that most should use unless you want to re-build from scratch within Cytoscape.

USAGE

The N3 files are intended to be indexed in a triple store. The first two CSV files are intended to be loaded into OpenOffice Calculator or Microsoft Excel. The *Cytoscape* CSV file is intended to be loaded in Cytoscape.

ENDNOTES

[1] http://opencyc.org

[2] http://www.w3.org/DesignIssues/Notation3.html

[3] http://www.mpi-inf.mpg.de/~suchanek/downloads/yago/

[4] http://dbtune.org/bbc/peel/

[5] http://www.cytoscape.org/