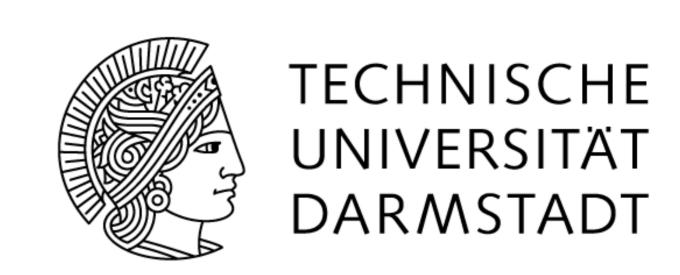


UBY UBY-LMF – A Uniform Model for Standardizing Heterogeneous LexicalSemantic Resources in ISO-LMF





Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek and Christian M. Meyer

http://www.ukp.tu-darmstadt.de/uby/

UBY-LMF in a Nutshell

Developing an LMF-based model for large-scale lexical-semantic resources (LSRs)

Motivation:

- Problem: Incompatible formats of (integrated) LSRs
- Solution: Using LMF for modeling LSRs
- Problem: Only few types of LSRs have been modeled in LMF so far
 - Solution: Creating a new, comprehensive lexicon model

Vision: A lexicon model which

- 1. Represents all information in various LSRs
- 2. Is extensible by further information types
- 3. Ensures structural and semantic interoperability

Contributions:

- 1. UBY-LMF: Lexicon model
- ISO Standard Lexical Markup Framework (LMF)
- Expert-built and collaboratively created LSRs
- Fine-grained modeling of information types
- Attributes and values refer to ISOCat data categories

2. UBY

- 9 LSRs in English and German
- 9 pairwise sense alignments between LSRs
- Confirms the applicability of the model
- 3. UBY-API:
 - Uniform access to LSRs in UBY via JAVA API

UBY-LMF UBY-LMF Model FormRepresentation WordForm PredicativeRepresentation grammaticalNumber writtenForm targetLexicalEntry: IDREF predicate: IDREF lanquageldentifier **grammaticalGender** exampleType nyphenation orthographyName isBreakBefore geographicalVariant verbFormMood TextRepresentation phoneticForm ⇔degree writtenText Context languageldentifier. Frequency orthographyName contextType LexicalEntry corpus geographicalVariant frequency |incorporatedSemArg: IDREF 🗠 RelatedForm generator transparentMeaning targetLexicalEntry: IDREF partOfSpeech targetSense: IDREF SyntacticBehaviou SenseRelation subcatFrame: IDREF SynsetRelation target: IDREF MonolingualExtRef subcatFrameSet: IDREF target: IDREF sense: IDREF relType externalSystem relName relName lexternalReference Lexicon SyntacticArq SubcategorizationFrame ConstraintSet Statement Synset parentSubcatFrame syntacticCat**ego**ry. f arammaticalFunction igaplus⇔subcatLabel statementType determiner SenseAxis: preposition LexemeProperty prepositionType LexicalResource l auxiliary sensel: IDREI syntacticProperty verbForm sense2: IDREI dtdVersion svnsetl: IDRER svnset2: IDREF senseAxisTy**pe** SemanticLabel optional GlobalInformation SubcatFrameSe SenseAxisRelatio quantification target: IDREF SvnSemArqMap synArg: IDREF parentSubcatFrame : IDREF relName semArg: IDREF Definition SubcatFrameSetElement definitionType arg1: IDREF element: IDREF Colored items mark LMF SemanticArg extensions SemanticPredicate PredicateRelation New LMF class ArqumentRelation New LMF class target: IDREF semanticRole arget: IDREF New LMF class coreType relType relType isincorporated ✓ lexicalized. relName New relationship relevantSemanticPredicate:IDREF perspectivalized

Extensions to LMF

New Classes

• SemanticLabel

- Subclass of Sense, SemanticPredicate, SemanticArgument
- Used for domain, style, register...

Frequency

- Attached to most classes
- Used to encode frequencies derived from corpora
- SubcatFrameSetElement
 - Subclass of SubcategorizationFrameSet
 - Used for linking frames with arbitrary cardinality

New Relationships

- SemanticArgument --- Definition
 - Allows to encode definitions of FrameNet frame elements
- MonolingualExternalRef --- Context
 - Allows to encode annotated corpus sentences
- SenseRelation --- FormRepresentation
 - Alleviates a special kind of ambiguity in Wiktionary
 Targets of relations and translations are word forms, not senses (as required by LMF)
 - ⇒ Relation has a null target, but an attached word form

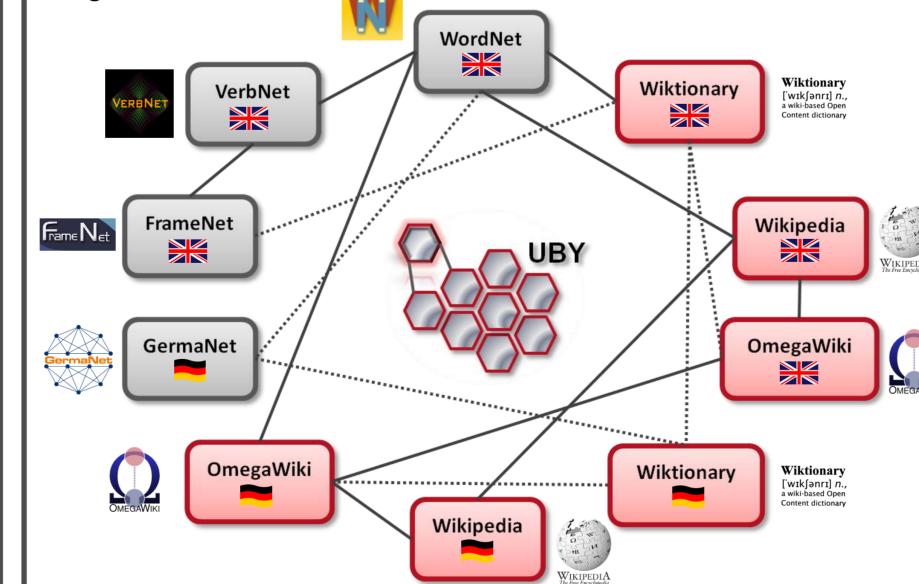
Novel Use of SenseAxis

Used for encoding cross-lingual and monolingual alignments between LSRs

Instantiation: UBY

10 UBY Lexicons represented in UBY-LMF

Solid lines: 9 sense alignments; Dotted lines: planned sense alignments



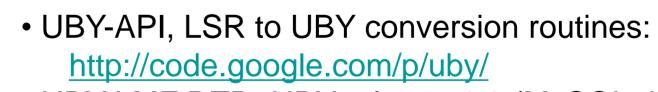
UBY Lexicon	Lexical Entry	Sense	Sense Relation
FrameNet 1.5	9,704	11,942	-
GermaNet 6.0	83,091	93,407	329,213
OmegaWiki-DE	30,967	34,691	60,054
OmegaWiki-EN	51,715	57,921	85,952
Wikipedia-DE	790,430	838,428	571,286
Wikipedia-EN	2,712,117	2,921,455	3,364,083
Wiktionary-DE	85,575	72,752	434,358
Wiktionary-EN	335,749	421,848	716,595
WordNet 3.0	156,584	206,978	8,559
VerbNet	3,962	31,891	-
UBY	4,259,894	4,691,313	5,570,100

Using UBY

UBY-API

- JAVA with Hibernate
- NLP application:
- Easy *swapping* of resources
 - Easy *combination* of resources

Open licenses for data and software:



 UBY-LMF DTD, UBY release 1.0 (MySQL database dump): www.ukp.tu-darmstadt.de/uby/

See also

UBY:

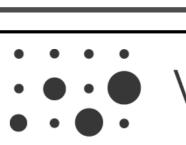
Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth:

Uby – A Large-Scale Unified Lexical-Semantic Resource, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), (to appear), April 2012. Avignon, France.

Subcat-LMF

Judith Eckle-Kohler and Iryna Gurevych: **Subcat-LMF – Fleshing out a standardized format for subcategorization frame interoperability**, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), (to appear), April 2012. Avignon, France.









Wiktionary
[wik[ənri] n.,

NLP