

## 1. Title

Domain-specific hyponym relation, Release 1.0

## 2. Authors

Jun Liu, Bifan Wei, Jian Ma, Chenchen Wang

## 3. Data Type

Text

## 4. Data sources

All the hypernym words and hyponym words of hyponym relations come from the article titles of English Wikipedia, and all relations between them are annotated by domain experts manually in terms of a strict process.

## 5. Applications

- 1) Domain-specific hyponym relation extraction;
- 2) Domain-specific faceted taxonomy learning.

## 6. Languages and Dialects

English [eng]

## 7. Narrative Description

- 1) A brief description of the corpus

This corpus provides more than 5000 hyponym relations of five domains, including *Data mining*, *Computer network*, *Data Structure*, *Euclidean Geometry* and *Microbiology*.

Table 1. Number of terms and hyponym relations

| Dataset                   | # Terms | # Hyponym Relations |
|---------------------------|---------|---------------------|
| <b>Data mining</b>        | 278     | 364                 |
| <b>Computer network</b>   | 336     | 399                 |
| <b>Data structure</b>     | 315     | 578                 |
| <b>Euclidean geometry</b> | 455     | 690                 |
| <b>Microbiology</b>       | 1,028   | 3,533               |

- 2) A brief explanation of the hyponym relation

A hyponym relation is a word sense relation that is an *IS-A* relation (with the more general word sense being the *hypernym* and the other the *hyponym*). A hyponym relation can be defined more formally as: Noun X, Noun Y, *Hyponymy* (X,Y) iff  $F(X)$

entails, but is not entailed by  $F(Y)$ , where,  $Y$  is the *hyponym*,  $X$  is the *hypernym* and  $F(-)$  is a sentential function satisfied by  $X$  and  $Y$ . It is in an inverse relationship with a hypernym relation and can be instantiated as a hyponymy relation mention in a sentence. For example, *dog* is a hyponym of *animal* and *binary tree* is a hyponym of *tree structure*.

### 3) Suitable applications

The corpus of *Domain-specific hyponym relations* aims to provide useful domain-specific hyponym relations in a domain (e.g., *Data structure*). The article pages corresponding to the hypernym or hyponym words can be acquired from the English Wikipedia to help the analysis of hyponym relations. The potential application of domain-specific hyponym relations are described as follows.

- It can be used for the ground truth of domain-specific hyponym relation extraction, domain-specific taxonomy learning and ontology learning.
- It can also be more effectively used for query result organization in faceted search.
- It can be utilized for knowledge organization and automated reasoning in knowledge-rich applications.

## 8. Data

Each XML file in this corpus provides hyponym relations of one domain. In the XML file, we show the name of a term in the tag of name. We also give the corresponding article's URL of English Wikipedia. Generally, the term is the title of the article. In the tag of relation, we show a hyponym relation, including the names of the hyponym word and the hypernym word.

## 9. Related papers

The corpus has been utilized in several academic papers listed below.

- [1] B. Wei, J. Liu, J. Ma et al., "Motif-based hyponym relation extraction from Wikipedia hyperlinks," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2014. (Accepted)
- [2] B. Wei, J. Liu, J. Ma et al., "DFT-Extractor: a system to extract domain-specific faceted taxonomies from Wikipedia," in *Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, Brazil, 2013*, pp. 277-280.
- [3] B. Wei, J. Liu, J. Ma et al., "MOTIF-RE: motif-based hypernym/hyponym relation extraction from Wikipedia links," in *Proceedings of the 19th International Conference on Neural Information Processing, Doha, Qatar, 2012*, pp. 610-619.
- [4] B. Wei, J. Liu, Q. Zheng et al., "A survey of faceted search," *Journal of Web engineering*, vol. 12, no. 1-2, pp. 041-064, 2013.

## **10. Licensing notice**

The corpus is licensed under CC-BY-SA.

<http://creativecommons.org/licenses/by-sa/3.0/>