



MultiWordNet ®

ITC-irst, via Sommarive 18, Povo - Trento, Italy

MultiWordNet is a multilingual lexical database including information about English and Italian words. It is an extension of WordNet 1.6, a lexical database for English developed at the Princeton University. MultiWordNet contains information about the following aspects of the English and Italian lexica:

- lexical relations between words
- semantic relations between lexical concepts
- correspondences between Italian and English lexical concepts
- semantic fields (domains)

The basic lexical relationship in MultiWordNet is *synonymy*. Groups of synonyms are used to identify lexical concepts, which are also called *synsets*. Here is an example of an Italian synset:

```
{elaboratore, computer, cervello_elettronico, calcolatore}
```

Synsets are the most important unit in MultiWordNet. Many interesting information are attached to them, such as *semantic fields* and *semantic relationships*. For example, the above synset belongs to the "Computer Science" semantic field and has three different semantic relationships:

```
has_hyponym {macchina}
```

```
has_hyponym {calcolatore_analogico}, {calcolatore_digitale}, etc.
```

```
has_part {microchip, chip}, etc.
```

Cross language correspondence is defined between synsets as well:

```
{elaboratore, computer, cervello_elettronico, calcolatore}
```

```
corresponds_to
```

```
{computer, data_processor, electronic_computer, information_processing_system}
```

MultiWordNet can be used for a variety of NLP tasks including:

- *Information Retrieval*: synonymy relations are used for query expansion to improve the recall of IR; cross language correspondences between Italian and English synsets are used for Cross Language Information Retrieval.
- *Semantic tagging*: MultiWordNet constitutes a large coverage sense inventory which is the basis for semantic tagging, i.e. texts are tagged with synset identifiers.
- *Disambiguation*: Semantic relationships are used to measure the semantic distance between words, which can be used to disambiguate the meaning of words in texts. Also semantic fields have proved to be very useful for the disambiguation task.
- *Ontologies*: MultiWordNet can be seen as an ontology to be used for a variety of knowledge based NLP tasks.
- *Terminologies*: MultiWordNet constitutes a robust framework supporting the development of specific structured terminologies.

Release Information

The release 1.39 of MultiWordNet is currently available. It includes information about 58,000 Italian word meanings and 32,700 synsets (in correspondence with the English equivalents). It also includes a labelling of all WordNet 1.6 synsets with domain labels. Work on MultiWordNet is going on.

System Description

MultiWordNet data are provided as a dump of a MySQL database, together with Java API if required. From the next release, an html-based interface using Apache/php/MySQL will also be available.

Licences

Research and commercial licences (with restrictions) are available.

Research licences with unlimited duration are granted to Academic, non-profit institutions. Private companies involved in a research project can also obtain a research licence. In this case, the licence allows for the usage of MultiWordNet only within the framework of the research project and its duration is limited to that of the project itself. Evaluation and Commercial licences are granted to all the others. The following table illustrates the costs of the licences in Euros:

<i>Licence</i>	<i>MWN (with WN Domains)</i>	<i>WordNet Domains only</i>
Internal use by a commercial organisation*	6,000	600
Evaluation licence (3 month)	500	100
Research use by an academic institution*	250	free

* For relicensing or research use by a commercial organization, please contact us directly.

Contact Persons

Emanuele Pianta, Luisa Bentivogli, Bernardo Magnini
e-mail: pianta@itc.it, bentivo@itc.it, magnini@itc.it

Web Address

Further information and MultiWordNet on-line are available at <http://multiwordnet.itc.it>