Word sense disambiguation

Word sense disambiguation (WSD) deals with computationally resolving ambiguities in a text. Languages have many polysemous words i.e. words having more than one meaning or sense. WSD is the process of identifying the correct sense of a word in a particular context. For instance, consider the sentence, 'Ram is playing cricket in the park', where the word 'cricket' is ambiguous with two senses: 'a game' and 'an insect'. Here, the correct sense of 'cricket' is 'a game' as the word 'park' appears in its context.

WSD has been extensively studied in computational linguistics due to its importance in understanding the semantics of natural languages. It has a significant impact on various real-world applications including machine translation, sentiment analysis, information retrieval, text summarisation, etc. To build a WSD system, two important resources are required: (i) Sense repository like wordnet and (ii) Sense-annotated corpus.

Resource scarcity acts as a major bottleneck for developing a WSD system, as many languages lack these aforementioned resources. At the Centre for Indian Language Technology (CFILT), focus has been to examine resource scarcity and provide resource-conscious solutions for WSD. Some unsupervised WSD approaches have where WSD is performed without relying on a sense-annotated corpus. They are:

- Context-based bilingual WSD approach: This is a bilingual WSD approach, where two resource-deprived languages help each others' WSD using a context-based expectation maximisation (EM) formulation.
- Most frequent sense detection approach: This is a novel approach for finding the most frequent sense (MFS) of a word where the use of word embeddings is explored.

CFILT has collaborated with IBM Research for WSD related approaches which can be applied to any language provided the Wordnet in that particular language exists.

Our WSD approaches are available at www.cfilt.iitb.ac.in/wsd-demo/