References


[42] http://www.cfilt.iitb.ac.in/


Appendix A

Definitions of Terms

This appendix lists the terms frequently used in this thesis.

**Computational Linguistics**: Computational linguistics is an interdisciplinary field concerned with the statistical or rule-based modelling of natural language from a computational perspective.

**Corpus**: A corpus or text corpus is a large and structured set of texts. They are used to do statistical analysis and hypothesis testing, checking occurrences or validating linguistic rules within a specific language territory.

**Lexical Item**: A lexical item is a single word, a part of a word, or a chain of words that forms the basic elements of a language's lexicon.

**Lexicon**: A lexicon is a language's inventory of lexemes. The word "lexicon" derives from the Greek λεξικόν (lexicon), neuter of λεξικός (lexikos) meaning "of or for words".

**Linguistics**: Linguistics is the scientific study of language. There are broadly three aspects to the study, which include language form, language meaning, and language in context.

**Morphosyntactic**: The study of grammatical categories or linguistic units that have both morphological and syntactic properties.
**Morpheme**: A meaningful linguistic unit consisting of a word, such as *man*, or a word element, such as *-ed* in *walked*, that cannot be divided into smaller meaningful parts.

**Morphology**: A branch of linguistics that studies and describes patterns of word formation, including inflection, derivation, and compounding of a language.

**Part of speech tagging**: In corpus linguistics, part of speech tagging also called grammatical tagging or word-category disambiguation is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition, as well as its context.

**Tagset**: The set of tags used for annotation in a particular language in a particular corpus.

**Tagged Corpus**: The text corpus in which all the lexical items are annotated with its proper part of speech tag is known as tagged corpus. They are used to do statistical analysis and hypothesis testing, checking occurrences or validating linguistics rules within a specific language territory.
Appendix B

List of Publications


Appendix C

Participation in Conferences and workshops

Conferences


2. “National Conference on Current trends in Computer Science 2010” organized by Department of Computer Science, Assam University Silchar from 22nd to 24th February 2010.

3. “National Conference on Computational Intelligence and Signal Processing CISP 2001” organized by Assam Don Bosco University, Guwahati on 2nd March 2011.


6. “1st National Conference on Research & Higher Education in Information Technology (RHEIT-2013)” held on 4th -5th February, 2013 in the Department of Information Technology, Assam University, Silchar.

Workshops

1. “National Workshop on Neural Networks and Applications” held at Gauhati University, Guwahati jointly organized by ISI Kolkata and Department of Statistics, Gauhati University from 15th to 17th December, 2009.
2. “2nd national Workshop on Indo-WordNet” held at Shillong jointly organized by Department of Computer Science, Assam University, Silchar, IIT Bombay, Gauhati University and Manipur University from 12th to 14th April, 2010.

3. “National Workshop on Language Teaching, Testing and Evaluation” organized by Department of Linguistics, Assam University Silchar from the 23rd to 26th of February, 2001 at Assam University, Silchar.

4. “National Workshop on Introduction to Natural Language Processing” from 4th to 8th March, 2013 organized by the Linguistic Data Consortium for Indian Languages, Central Institute of Indian Languages, Mysore in collaboration with the Department of Linguistics, Assam University, Silchar.