

SEMESTER -III

MAHATMA GANDHI UNIVERSITY

M.A. PROGRAMME IN SANSKRIT VEDANTA

COURSE XIV- Core XIV Computational Sanskrit for Vedanta Natural Language Processing (NLP)

Contact Hrs: 90

Course code SA040305 Credits: 4

Outcome of the course:

This course provides a general introduction to Natural Language Processing with a particular emphasis on language models related to Sanskrit and other Indian languages, morphology, parsing, syntax, semantics, information extraction, information retrieval, text processing and Natural Language Tool Kit.

OBJECTIVES OF THE COURSE:

1. Understand the core tasks in NLP.
2. Understand and analyse natural languages and be familiar with the theoretical frameworks.
3. Appreciate the relationship between linguistic representations and their computational applications.
4. Identify NLP tools and understand the logic behind the tools.
5. Know how to use the different NLP applications.

COURSES OUTLINE:

Module I:

1. Introduction and brief history of NLP
2. MT (Machine Translation) in India and abroad.
3. Linguistic issues in NLP
4. Morpheme, word, sentence and paragraphs.
5. Morphology – Morphological analysers and generators.

Module II:

1. Input method
2. Transliteration schemes.

3. Indian scripts
 - a. Nature and structure.
 - b. Indian scripts on computers.
4. Encoding schemes – ASII, ISCII, Unicode.
5. XML Tagging, Tagging schemes.

Module III:

1. Introduction to Shabdabodha systems.
 - a. Vyakarana.
 - b. Nyaya.
 - c. Mimamsa.
2. Introduction to Navya Nyaya Paribhasha.
 - a. Language.
 - b. Metalanguage.
 - c. Artificial language.
 - d. Restricted language.

Module IV:

1. Machine translation – Various approaches – Paninian approach.
2. Karaka relations.
3. Computational concepts in Ashtadhyayi.
4. Introduction to Sanskrit computational tools.

Recommended readings:-

1. Sanskrit Parsing- Based on the theories of Sabdabodha – Prof AmbaKulkarni.
2. **Bharati, A., Chaitanya, V., Sangal, R. and Ramakrishnamacharyulu, K.V.,1995.** *Natural language processing: a Paninian perspective.* New Delhi, Prentice-Hall of India.
3. **Haspelmath, M. and Sims, A., 2013.** *Understanding morphology.* Routledge.
4. **Jurafsky, D. and Martin, J.H., 2014.** *Speech and language processing (Vol. 3).* London, Pearson.
5. **Grune, D. and Jacobs, C.J., 2008.** *Parsing Techniques-A Practical Guide,* David Gries and Fred P. Schneider, Eds.
6. **Kübler, S., McDonald, R. and Nivre, J., 2009.** *Dependency parsing.* Synthesis Lectures on Human Language Technologies, 1(1), pp.1-127.
7. **PavankumarSatuluri, 2015.** *Sanskrit Compound Generation: With a Focus on the Order of Operations,* Thesis, University of Hyderabad.
8. **Anil Kumar, 2012.** *An Automatic Sanskrit Compound Processing,* Thesis, University of Hyderabad