P G Diploma Course in Computer Science and NLP
Code – PGDCSN

Total Credits = 42

Semester 1		
पत्र कोड	पाठ्यवस्तु	क्रेडिट
PGDCSN01	व्याकरण**,** भाषाविज्ञान एवं शाब्दबोध **- 1 (Introductory)** • **Unit I** - Paninian Concepts • **Unit II** - Phonology, Morphology, word formation • **Unit III** - Sentence Analysis (Karaka tagging and analysis) • **Unit IV** - Akanksha, yogyata and sannidhi concepts	क्रेडिट 4
PGDCSN02	**Natural Language Processing - 1 (Introductory)** • **Unit I** - Introduction to NLP • **Unit II** - Morpheme, Word, Sentence etc .. • **Unit III** - Tagging schemes and Morphological Analysis • **Unit IV** - Rule based MT vs Machine learing based MT	क्रेडिट 4
PGDCSN03	**Computer Programming - 1 (Introductory)** • **Unit I** - Basics of Computer Application / Fundamental / Structure • **Unit II** - Introduction to Programming • **Unit III** - Execute through Command line • **Unit IV** - Regular expressions	क्रेडिट 4
PGDCSN04	**Machine Learning - 1 (Introductory)** • **Unit I** - Introduction to Artificial Intelligence, Unsupervised vs Supervised Learning, Feature Representation for NLP • **Unit II** - Unsupervised Learning: Clustering, Matrix Decompositions, Supervised Learning paradigms for NLP: Text Classification, Sequence Labeling, Sequence to Sequence • **Unit III** - Text Classification using K-Nearest Neighbour, Naive Bayes Model, Sequence Labeling using Hidden Markov Models • **Unit IV** - Machine Translation as Sequence to Sequence, Evaluation	क्रेडिट 4
PGDCSN05	**Project - 1**	क्रेडिट 5

	Semester 2	
PGDCSN06	व्याकरण, भाषाविज्ञान एवं शाब्दबोध - 2 (Advanced learning) • **Unit I** - Morphological analysis and generation • **Unit II** - Compound analysis and generation • **Unit III** - How to resolve word sense dissambiguation • **Unit IV** - Deep Paninian concepts	क्रेडिट 4
PGDCSN07	**Natural Language Processing - 2 (Advanced learning)** • **Unit I** - Introduction to various Tools • **Unit II** - Introduction to various Tools • **Unit III** - Introduction to various Tools • **Unit IV** - Introduction to various Tools	क्रेडिट 4
PGDCSN08	**Computer Programming - 2 (Advanced learning)** • **Unit I** - Data structure • **Unit II** - Programming with Python • **Unit III** - Web designing (HTML and XML) • **Unit IV** – Writing small script using python	क्रेडिट 4
PGDCSN09	**Machine Learning - 2 (Advanced learning)** • **Unit I** - Introduction to neural networks • **Unit II** - Word Representations • **Unit III** - Neural Language Modeling, Recurrent Neural Networks and NLP applications • **Unit IV** - Contextualized word representations and pretraining, Pretrained language models for Indian languages	क्रेडिट 4
PGDCSN10	**Project - 2**	क्रेडिट 5

Semester 1   =   4 main course X 4 Credits = **16 Credits**
                      1 Project work X 5 Credits = **5 Credits**
                      Total = **21 Credits**
Semester 2   =   4 main course X 4 Credits = **16 Credits**
                      1 Project work X 5 Credits = **5 Credits**
                      Total = **21 Credits**
Sub Total   =   21 X 21 = **42 Credits**

NOTE : For each of the paper, we describe the objectives and outcomes and the topics likely to be covered. The syllabus, as well as the reading material and reference list is only indicative.

P G Diploma in

Computer Science and Natural Language Processing

1. Introduction and objectives of the course (पाठ्यक्रम की प्रस्तावना एवं उद्देश्य)

To train Sanskrit scholars in the emerging field of Sanskrit Computational Linguistics, showing the relevance of Indian grammatical theories to the field of Computational Linguistics, thereby bridging the gap between the past and the present. This programme belongs to an emerging and high-tech area of Natural Language Processing. The past few years have seen considerable growth in the field of Sanskrit Computational Linguistics. This course aims at disseminating this knowledge to Sanskrit scholars. To deliver better technology, there is a need to train the manpower so that they can take up the challenges thrown open by the newly emerging field and provide meaningful solutions. During the course, the students will be exposed to the basics of logic, computer programming and NLP in addition to exposure to linguistics and strengthening their base of Vyakarana.

In this course, we look at the relevance of language sciences from an Information Technology perspective. This course will introduce the following modules a) Segmentation. b) Word Analysers and Generators c) Sentential analyser and generator followed by the introduction to various Resources such as a) Lexical Resources b) Annotated corpus with a focus on Sanskrit and the Indian contribution to the language sciences.

2. Qualification (पाठ्यक्रम अध्ययन की योग्यता)

This is a one-year course. This has 42 credits in total. The minimum qualification to get admission into this course is 50% in MA in Sanskrit from a recognized board (good knowledge in Vyakarana is necessary)

3. Pedagogy (पाठ्यक्रम की शिक्षाप्रविधि)

There will be classroom teaching lectures/tutorials followed by assignments or practical sessions (whichever is necessary at that time). Workshops will be arranged whenever it is needed. Learning materials (research papers, teaching materials, PowerPoint presentations, and videos of valuable lectures) will be provided. Discussions among experts and students will also be conducted. In case experts are not available offline we will arrange video conferencing for their talks/lectures.

4. Consistency (सङ्गति)

India's unique contribution to the world is the theories of Language Sciences. Sanskrit was the medium of serious communication for all discourses and scholarly communications until recent times. We also have a very rich language and grammatical tradition in India. Without knowing the structure of the Sanskrit language and grammar, it is often difficult to work for the development of Sanskrit in this technological era. Also, without knowing how to do programming in Sanskrit through a computer and how to teach Sanskrit to a computer, this will be a bit difficult to work in the field of Sanskrit Computational Linguistics. In the same way, until we do not know the process by which the human mind understands the language and works according to it, we will not be able to teach the computer either. Therefore, each unit of this course is very relevant.

5. Outcome (फलितांश)

Through this course, we can keep our traditional and Indian knowledge tradition alive through technology for the next generation. This course will also open new job opportunities for Sanskrit

students in various educational, industrial and IT sectors.

6. Evaluation (मूल्याङ्कन प्रक्रिया)

Continuous assessment will be done through assignments, group discussions, written tests, practical sessions and PowerPoint presentations.

7. Prescribed books (निर्धारित पाठ्यग्रन्थ)

Sem I - PGDCSN01 & Sem II - PGDCSN06

व्याकरण, भाषाविज्ञान एवं शाब्दबोध - **1 [PGDCSN01] &** व्याकरण, भाषाविज्ञान एवं शाब्दबोध – **2 [PGDCSN06]**

NOTE: Only those portion which is necessary will be taught not the full text.

1. अष्टाध्यायी
2. सिद्धान्तकौमुदी
3. Phonetics in Ancient India, W S Allen 1971
4. Sandhi, W S Wallen
5. Morphology
6. Syntax
7. Karaka Analysis
8. शाब्दतरंगिणी
9. शाब्दबोधमीमांसा
10. Indian Theories of Meaning
11. Philosophy of Word and Meaning
12. Sanskrit Philosophy of Language
13. Logic, Language, Reality
14. The Sanskrit Language: An Overview - History and Structure, Linguistic and Philosophical Representations
15. Annotation schemes for sandhi, morph, compound, karaka
16. शब्दशक्तिप्रकाशिका
17. नव्यन्यायभाषाप्रदीप, महेश न्यायरत्न
18. An introduction to sanskrit language and linguistics

Sem I - PGDCSN02 & Sem II – PGDCSN07

Natural Language Processing – 1 [PGDCSN02] & Natural Language Processing – 2 [PGDCSN07]

1. NLP: A Paninian perspective by Akshar Bharati, Vineet Chaitanya, Rajeev Sangal, prentice hall of India, 1995
2. Speech and Language Processing by Daniel Jurafsky and James H Martin
3. Annotation Guidelines developed by Sanskrit Consortium
4. Relevant research papers in the field of Machine Translation, NLP, Computational Linguistics, Sanskrit Computational Linguistics etc
5. Available Sanskrit Computational Linguistics Tools (Demo)
6. Related Papers and presentations

Sem I - PGDCSN03 & Sem II – PGDCSN08

Computer Programming – 1 [PGDCSN03] & Computer Programming – 2 [PGDCSN08]

1. Fundamentals of computer
2. UNIX power tools by Jerry Peek, Shelley Powers, Tim O'Reilly, Mike Loukides
3. Programming with Python
4. HTML and XML

Sem I – PGDCSN04 & Sem II - PGDCSN09

Machine Learning – 1 [PGDCSN04] & Machine Learning – 2 [PGDCSN09]

1. Explorations in Artificial Intelligence and Machine Learning, A CRC Press FreeBook, Taylor and Francis Group
2. Neural Networks for Natural Language Processing, Yoav Goldberg, Synthesis Lectures on Human Language Technologies #37, Morgan and Claypool Publishers, 2017
3. K Nearest Neighbor by O'Reilly
4. Related articles and presentations

8. Reference book (सन्दर्भ ग्रन्थ)

1. Ashtadhyayi of Panini, Dr. Naresh Jha, Chaukhamba Surbharati Prakashan, 2014
2. Dimensions of Panini Grammar, Kapil Kapoor, D K Print World, 2020
3. वैयाकरणसिद्धान्तकौमुदी with the commentaries बालमनोरमा – तत्वबोधनी, Motilal Banarassidas, 2010
4. कारकम् – महाबलेश्वर भट्ट,संस्कृत भारती, बेंगलूरु, 1999
5. तर्कसङ्ग्रहः शाब्दबोधसहितः-गिर्याचार्यव्याख्या, श्रीराघवेन्द्रस्वामिमठः, मन्त्रालयः, १९८०
6. Language and Logic, Navya Nyaya Perspective, Dr. Tirumala Kulkarni and Jaideep Joshi, Manipal University, 2013
7. Phonetics in Ancient India, W S Allen 1971
8. शब्दशक्तिप्रकाशिका, जगदीश तर्कालंकार, कलिकता, १९०४
9. Natural Language Processing with Python, Steven Bird, Edwan Lein and Edward Loper, O'Reilly
10. NLP: A Paninian perspective by Akshar Bharati, Vineet Chaitanya, Rajeev Sangal, prentice hall of India, 1995
11. Speech and Language Processing, Danial Jurafsky, James H Martin, 2009
12. समासपारिजात, Dr. C Poornananda Sastri, २००३
13. भाषाशास्त्रप्रवेशिनी, r s Venkatarama Shastri, Kuppuswami RESEARCH INSTITUTE, Madras, 1996
14. Language, Bloomfield, Motilal Banarsidass
15. Theories of Language: Oriental and Occidental, Prof. Korada Subramaniyam
16. शाब्दबोधमीमांसा, एन् अस् आर् ताताचार्य, Institute Franciasde Pondicherry/ RS Vidyapeeth 2006
17. An introduction to Sanskrit Linguistics, M Sriman Narayan Murti, DK Publications, Delhi
18. Programming Python: Powerful Object-Oriented Programming, 4th Edition, Mark Lutz, O'Reilly, 2011
19. Designing Web Interface, Bill Scott, O'Reilly, 2022
20. Python Cookbook: Recipes for Mastering Python 3, 3rd Edition, David Beazley, O'Reilly, 2013

21. Sabda, A Study of Bhartrhari's Philosophy of Language, Tandra Patnaik, D K Print World, 2017
22. Roots, Verb-Forms and Primary Derivatives of the Sanskrit Language, William Dwight Whitney, D K Print World, 2008
23. Paninian Tradition of Grammar and Linguistics, Rama Nath Sharma, D K Print World, 2017
24. Philosophy of Language in Classical Indian Tradition, K S Prasad, 2002
25. Linguistic Representations, Ramesh Chandra Pradhan, D K Print World, 2012
26. Artha: Meaning, Jonardon Ganeri, Oxford University Press, 2006
27. शाब्दतरंगिणी, सुब्रह्मण्यशास्त्री, KTP edition, 2006
28. The Word and The World, B K Matilal, 1992
29. Indian Theories of Meaning, K Kunjuni Raja, 1963
30. Philosophy of Word and Meaning, Gaurinath Shastri, 1959
31. Sanskrit Philosophy of Language, J F Stall, 1969
32. Explorations in Artificial Intelligence and Machine Learning, A CRC Press FreeBook, Taylor and Francis Group
33. Neural Networks for Natural Language Processing, Yoav Goldberg, Synthesis Lectures on Human Language Technologies #37, Morgan and Claypool Publishers, 2017
34. Speech and Language Processing (3rd ed. draft) Dan Jurafsky and James H. Martin (https://web.stanford.edu/~jurafsky/slp3/)
35. Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999.

9. Model Question paper (प्रारूप प्रश्न पत्र)

Multiple choice, Objective type questions, short questions, long essay, practical (lab session)

10. Name, address, telephone and e-mail of experts (विशेषज्ञों के नाम, पता, दूरभाष एवं ई-मेल)

1. Prof. Amba Kulkarni, 9440893578, ambapradeep@gmail.com, Hyderabad Central University

2. Prof. Malhar Kulkarni, 8369752472, malhar@iitb.ac.in, IIT Mumbai

3. Prof. Soma Paul, 9912910148, soma@@iiit.ac.in, IIIT Hyderabad

4. Prof. Srinivas Varkhedi, 9483501353, srivara@gmail.com, Central Sanskrit University

5. Prof. Lalit Kumar Tripathi, 9389428935, prof.lalit.tripathi@csu.co.in, CSU, GNJHA Campus

6. Prof. Girish Nath Jha, girishjha@jnu.ac.in, Jawaharlal Nehru University

7. Prof. Shivani V, 9497680167, Karnataka Sanskrit University

8. Prof. Pawan Goyal, 9434202818, pawang.iitk@gmail.com, IIT Khragpur

9. Dr. Sridhar Subanna

10. Dr. Sukhada Sharma, 9291504271, sukhada.hss@iitbhu.ac.in, IIT BHU

11. Dr. Anil Gupta, 9581181590, anil.lalit@gmail.com

12. Dr. Pavan Kumar Satuluri, 9493949376, pavankumarsatuluri@gmail.com, Chinmay University

13. Dr. Tanuja Ajotikar, 09420356462, gtanu30@gmail.com, MaharshiVedic University, Vlodrop, The Netherlands

14. Prof. Kavinarayan Murthy, 9573242653, knmuh@yahoo.com, University of Hyderabad

Objective and Outcome of each Paper

Semester I

PGDCSN01 - Unit I - व्याकरण, भाषाविज्ञान एवं शाब्दबोध – 1

Objective :

- to introduce the rich Indian Grammatical Tradition and concepts of Vyakarana with reference to various issues in NLP and familiarise the students with the parallel linguistic terminology and concepts to enable them to read and understand the latest works in the area of computational linguistics

- general introduction to the study of languages and Sanskrit language from a linguistics perspective

- to introduce the prominent concepts of Shabdabodha.

- to introduce the declensions, verbal roots and word formation

- to introduce the rich morphological tradition

- to introduce the concept of case-endings and the various meanings they carry

Outcome :

- know the language construction

- know the various phonetic changes occurring in Sanskrit

- know the importance of language and Paninian grammar

- know how to solve the cases of multi-sense of a single word

- have basic knowledge on the process of shabdabodha

- chart out the map of a sentence

PGDCSN02 - Unit II - Natural Language Processing – 1

Objective :

- to provide an introduction to NLP

- to understand the definition of morph, word and sentence for processing

- to introduce the available tagging schemes

- to introduce the difference between rule-based and machine learning-based machine translation

systems

Outcome :

- the students should be able to understand the system of language understanding

- can access our traditional linguistic resources vis-a-vis the modern linguistics resources

- can assess the relevance of fundamental principles and concepts in Indian theories to modern languages.

- also understand how to analyse language components.

PGDCSN03 - Unit III - Computer Programming – 1

Objective :
- to get exposure to the fundamentals of computers, and applications of computers in linguistics and language studies.
- to introduce the students to various Unix tools and scripting languages so that students can develop small interfaces on top of existing tools, process corpus, and do preliminary linguistic and statistical analysis of the corpus.
- to introduce the basic concepts of programming.
- to introduce the reader to various kinds of expressions and operators.

Outcome :
- know the basics like operators and parameters of computer programming
- know the key regular expressions

PGDCSN04 - Unit IV - Machine Learning – 1

Objective:

- to introduce the concepts of artificial intelligence and machine learning

- to introduce the concept of unsupervised and supervised learning, the Naive Bayes model, HMM etc. with NLP applications

- to introduce the use of classial Machine Learning methods for NLP tasks

Outcome:

- know the basics of Artificial Intelligence and Machine Learning

- know how to apply basic Machine Learning coonepts for NLP problems

PGDCSN05 - Unit V - Project 1

Semester II

PGDCSN06 - Unit I - व्याकरण, भाषाविज्ञान एवं शाब्दबोध – 2

Objective :

- to introduce the concepts of Vyakarana with reference to various issues in NLP and familiarise the students with the parallel linguistic terminology and concepts.

- to provide in-depth knowledge about suffixes, compounds

- to decide the meaning of a word according to context

- to give in-depth knowledge of the Shabdabodha process of sentences

Outcome :

- know the use of suffixes

- know the word paradigm (how to generate and how to analyse)

- know the Sanskrit compound-related rules, their generation and analysis

- know the deep Paninian concepts

- chart out the map of more sentences

PGDCSN07 - Unit II - Natural Language Processing – 2

Objective :

- to introduce various Indian language machine translation tools

- to introduce machine translation systems developed for Sanskrit and their components

Outcome :

- know how to and where to use Sanskrit computational linguistics tools

- know how they can develop the available tools

PGDCSN08 - Unit III - Computer Programming – 2

Objective :

- to introduce the concepts of programming and data structure, programming with python for text processing

- to introduce the students to scripting languages like HTML and XML so that students can

develop small interfaces on top of existing tools, process corpus, and do preliminary linguistic and statistical analysis of the corpus.

Outcome :

- know the fundamentals of text processing using python

- know how to develop small modules using HTML and XML

PGDCSN09 - Unit IV - Machine Learning – 2

Objective :

- to introduce the state-of-the-art methods from Neural Networks for NLP

- to introduce deep learning paradigms and standard tools for NLP in low-resource languages

Outcome:

- Know the basic premise of neural networks and deep learning

- Know how to apply advanced deep learning concepts for NLP problems

PGDCSN10 - Unit V - Project 2