Subject Code	COMP6402
Subject Title	Natural Language Processing Techniques
Credit Value	3
Level	600 level
Pre-requisite / Co-requisite/ Exclusion	For DALS students, Taking FH6051 (Computational Linguistics) first is preferred, but not absolutely necessary
Objectives	This subject aims to achieve the following goals: • To introduce students the challenges of empirical methods for natural language processing (NLP) applications. • To introduce basic mathematical models and methods used in NLP applications to formulate computational solutions. • To provide students with the knowledge on designing procedures for natural language resource annotation and the use of related tools for text analysis and hands-on experience of using such tools. • To introduce students research and development work in information retrieval, information extraction, and knowledge discovery using different natural language resources. • To give an overview of the major technologies in speech recognition and synthesis including tools for acoustic analysis and hands-on experience of using such tools • To give students opportunities to sharpen their programming skills for computational linguistics applications
Intended Learning Outcomes	Upon completion of the subject, students will be able to: (a) Understanding of the fundamental mathematical models and algorithms in the field of NLP. (b) Apply these mathematical models and algorithms in applications in software design and implementation for NLP. (c) Understand the principles of language resource annotation and its use in machine learning applications and apply the above principles in analysis of data and acquire intended information through the use of available tools. (d) Understand the design and implementation issues in various NLP applications such as information retrieval and information extraction. (e) Understand the complexity of speech and the challenges facing speech engineers. (f) Understand the principles of automatic speech recognition and

<table>
<tr><td></td><td>synthesis.
(g) Problem solving using systematic ways and learning independently.</td></tr>
<tr><td>Subject Synopsis/ Indicative Syllabus</td><td>• Introduction and overview
• Statistical models, information theory, and vector space models of data representation
• Formal language, automaton, and their applications in NLP
• Lexical resources, corpora, and annotations
• Statistical language models and machine learning algorithms
• Text classification and indexing
• Advanced applications such as Information retrieval, information extraction and knowledge discovery
• Introduction to speech technology including phonetics, text-to-speech synthesis, and automatic speech recognition</td></tr>
<tr><td>Teaching/Learning Methodology</td><td>The course will be taught in a combined form of seminars and lab sessions. Readings will be assigned every week and students are expected to participate in discussion during the seminars. Students are also expected to participate in lab sessions and complete lab exercises on computers.</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Learning Outcomes</td><td>
<table>
<tr><td rowspan="2">Specific assessment methods/tasks</td><td rowspan="2">% weighting</td><td colspan="7">Intended subject learning outcomes to be assessed (Please tick as appropriate)</td></tr>
<tr><td>a</td><td>b</td><td>c</td><td>d</td><td>e</td><td>f</td><td>g</td></tr>
<tr><td>1. Homework assignments</td><td>60%</td><td>✓</td><td></td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>2. Lab exercises</td><td>10%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>3.Tests</td><td>30%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Total</td><td>100 %</td><td colspan="7"></td></tr>
</table>

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:
All the above assessment methods are appropriate for evaluating students' understanding of course materials and their problem solving skills. Individual assignments provide assessment on a regular basis, which also</td></tr>
</table>

<table>
<tr><td></td><td colspan="2">serve as a means of self-monitoring for students. Lab exercises will emphasize the ability to apply knowledge to real-world problems and have hands-on experience of using certain tools. Tests will assess students' overall understanding of the concepts and algorithms learnt in class.</td></tr>
<tr><td rowspan="8">Student Study Effort Required</td><td>Class contact:</td><td></td></tr>
<tr><td>▪ Seminars</td><td>36 Hrs.</td></tr>
<tr><td>▪ Tutorial/Lab</td><td>6 Hrs.</td></tr>
<tr><td>Other student study effort:</td><td></td></tr>
<tr><td>▪ Reading</td><td>35 Hrs.</td></tr>
<tr><td>▪ Homework assignments and preparation for tests</td><td>40Hrs.</td></tr>
<tr><td>Total student study effort</td><td>117Hrs.</td></tr>
<tr><td>Reading List and References</td><td colspan="2">Text Book:
There is no single text book for this classes. Selected readings from text of the references will be suggested, along with relevant research papers.

References:
Dan Jurafsky and James H. Martin, Speech and Language processing, 2nd Addition, Prentice Hall, 2008
Christopher Manning and Hinrich Schuetze, Foundations of Statistical Natural Language Processing, the MIT Press, 1999
Ruslan Mitkov, The Oxford Handbook of Computational Linguistics, Oxford University Press, 2005
Christopher Manning, Prabhakar Raghavan, and Hinrich Schuetze, Introduction to Information Retrieval, Cambridge University Press, 2008
Charu C. Aggarwal and ChengXiang Zhai, Mining Text Data, Springer, 2012
Steven Bird, ewan Klein, and Edward Loper, Natural Language Processing with Python, O'Reilly Media, 2009
Hopcroft, J.E. and Ullman, J.D., Introduction to Automata, Theory and Languages, Addison-Wesley, 1979
Mark Lutz, Learning Python, O'Reilly Media; Third Edition edition (October 29, 2007)</td></tr>
</table>