

Short Title:	Natural Language Processing APPROVED
Full Title:	Natural Language Processing
Module Code:	MHLT H6011
ECTS credits:	10
NFQ Level:	9
Module Delivered in	1 programme(s)
Module Contributor:	Arnold Hensman
Module Description:	The aims of this module are: To give practical exposure to the student to the main research areas in that part of computer science concerned with the processing of written and spoken human language within software- Natural Language Processing. Concentrate on the practical application of advanced structures and algorithmic techniques in solving language problems in software. Provide the student with the ability to undertake parsing and generation of human language in software. Provide students with the ability to apply Natural Language Processing techniques, software solutions and computational intelligence in the areas of internet development, search engines, machine translation, speech processing and human computer interfaces.
Learning Outcomes:	
<i>On successful completion of this module the learner will be able to</i>	
<ol style="list-style-type: none"> 1. Select the appropriate techniques, software algorithms and structures for use in solving practical language related software-engineering problems. 2. Understand how human languages are constructed, apply appropriate techniques to tokenise, parse and generate human language in software, and to build grammars in software. 3. Understand how internet based solutions that process language are constructed, i.e. search engine software. 4. Apply computational approaches to language and speech processing in software. 5. Critically analyse technical research articles and papers related to natural language processing. 	

Module Content & Assessment

Indicative Content
Applications of Natural Language Processing (10%) Internet Search Engines, Human speech interfaces to a computer, Human language processing, Machine translation and localisation, Speech processing by computers.
What is a language (10%) Human Languages, Chomsky hierarchy of languages, Understanding of human language in software systems, Model of a human language grammar, The role of the lexicon.
Structures and Algorithms for Natural Language Processing (20%) The nature of the problems of NLP, String and text processing, The structure of the lexicon, Structures and algorithms: Finite state automata, Graphs structures, Lists, queues, stacks, Trees, binary trees and string tries.
Natural Language Processing (20%) How natural languages are processed in software and the problems that can arise (e.g. polysemy, reference), Defining a grammar of a language, The interaction of syntax and semantics, Definite clause grammars, Unification grammars, Context free grammars, Generative grammars, Semantics: Macro and thematic roles, Aktionsarten, computational semantics.
Elements of a Parser and NLP Development Paradigms (30%) Tokens and tokenising algorithms, Composition of a parser, Parser types (Terminal, shallow and composite parsers), HPSG, Java, XML, Prolog, Considerations, benefits, advantages, How to design a parser and generator, The role of a grammar, Different types of grammars, Unification approaches to grammar, Encoding a grammar in software, Regular expressions, Testing a parser and generator, Machine translation.
Current Issues in Natural Language Processing (10%) Current issues in NLP in the following areas: Social Media and Web Data, QA (Question-Answering) Systems, Information Extraction (IE), General Text Analytics, NLP/TA for Big Data, Semantic Web and Linked Data, Natural Language in Conceptual/Ontological Modelling.

Indicative Assessment Breakdown	%
Course Work Assessment %	100.00%

Course Work Assessment %				
<i>Assessment Type</i>	<i>Assessment Description</i>	<i>Outcome addressed</i>	<i>% of total</i>	<i>Assessment Date</i>
Practical/Skills Evaluation	Practical exercise based on lecture material	1,2,3,4,5	20.00	Every Week
Project	This Practical Assessment will typically consist of a choice of the following: • Defining elements of the semantics and syntax of a language • Building a parser in software • Building and applying text and string processing algorithms • Building natural language interfaces to software applications for use on the Internet • Building and implementing a computational online lexicon to support a grammar • Implementing voice capabilities in software for naturalistic speech synthesis and recognition	1,2,3,4,5	40.00	n/a
Project	This research Assessment Research work will typically involve case studies and papers on topics such as: • Natural Language processing and natural language applications for the Internet • Text and string processing issues with Internet search engines. • Current issues in natural language processing in software • Current issues in machine translation. • Architectures for machine translation engines. • Design and practical issues relating to Text-to-Speech Processing • The use of XML/ Speech Mark-up Languages/ Speech API with speech enabled applications	1,2,3,4,5	40.00	n/a

No Final Exam Assessment %

Indicative Reassessment Requirement
Coursework Only <i>This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.</i>

ITB reserves the right to alter the nature and timings of assessment

Indicative Module Workload & Resources

Indicative Workload: Full Time	
Frequency	Indicative Average Weekly Learner Workload
Every Week	28.00
Every Week	28.00
Every Week	56.00

Indicative Workload: Part Time	
Frequency	Indicative Average Weekly Learner Workload
Every Week	28.00
Every Week	28.00
Every Week	56.00

Resources
<i>Recommended Book Resources</i>
<p>Daniel Jurafsky, James H. Martin 2009, <i>Speech and language processing</i>, Pearson Prentice Hall Upper Saddle River, N.J. [ISBN: 0131873210]</p> <p>Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze 2008, <i>Introduction to information retrieval</i>, Cambridge University Press New York [ISBN: 0521865719]</p> <p>Grant S. Ingersoll, Thomas S. Morton, Andrew L. Farris, <i>Taming Text</i>, Manning Publications [ISBN: 193398838X]</p>
<i>Supplementary Book Resources</i>
<p>I van A. Sag, Thomas Wasow, Emily M. Bender 2003, <i>Syntactic theory</i>, Center for the Study of Language and Information Stanford, Calif. [ISBN: 1575864002]</p> <p>Christopher D. Manning, Hinrich Schütze, <i>Foundations of statistical natural language processing</i>, Cambridge, Mass. ; MIT Press, c1999. [ISBN: 0262133601]</p>
<i>This module does not have any article/paper resources</i>
<i>This module does not have any other resources</i>

Module Delivered in

Programme Code	Programme	Semester	Delivery
BN_KMHLT_R	Master of Science in Computing in Multimodal Human Language Technology	1	Mandatory