


Course Outline	
COMP8108 Natural Language Processing (2/1)	
Effective Date 01 February 2016	Study Program Computer Science
	Revision 0

1. Course Description

Natural Language Processing is one field of Artificial Intelligence to perform tasks involving human language, human-machine communication, improving human-human communication, or simply doing useful processing of text or speech. This course provides students with the fundamental techniques of Natural Language Processing such as understanding words and their properties, modeling natural language, doing parsing and getting overview of Natural Language Processing applications. By completing this course, students can explain what Natural Language Processing is and describe how to implement the techniques to build an application. To understand this course appropriately, students need to pass Artificial Intelligence course.

2. Graduate Competency

Each course in the study program contributes to the graduate competencies that are divided into employability and entrepreneurial skills and study program specific outcomes, in which students need to have demonstrated by the time they complete their course.

BINUS University employability and entrepreneurial skills consist of planning and organizing, problem solving and decision making, self management, team work, communication, and initiative and enterprise.

2.1. Employability and Entrepreneurial Skills

Aspect	Key Behaviour

2.2. Study Program Specific Outcomes

Study Program Specific Outcomes
Able to demonstrate knowledge and understanding of algorithm concepts, principles and theories relating to computer science knowledge.
Able to classify problems and to apply design and development principles for specific problems
Able to construct a solution by applying current technologies
Able to depict trend technologies in the future

3. Topics

- Natural Language Processing
- Regular Expressions and Automata
- Words and Transducers
- N-grams
- Part-of-Speech Tagging
- Hidden Markov and Maximum Entropy Models
- Syntactic Parsing
- Statistical Parsing
- Computational Semantics
- Lexical Semantics
- Computational Discourse
- NLP Applications
- Project Presentation

4. Learning Outcomes

On successful completion of this course, student will be able to:

- LO 1: Describe what is Natural Language Processing
- LO 2: Explain fundamental concepts of Natural Language Processing
- LO 3: Apply Natural Language Processing concepts in sentences
- LO 4: Construct Natural Language Processing application

5. Teaching And Learning Strategies

In this course, the lecturers might deploy several teaching learning strategies, including Discussion, Research Papers and presentations, Lecture, Experiment, and Project Research.

6. Textbooks and Other Resources

6.1 Textbooks

1. Daniel Jurafsky & James H. Martin. (2006). *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. 02. Prentice Hall. New Jersey. ISBN: 978-0131873216.
2. Christopher D. Manning. (1999). *Foundations of statistical natural language processing*. 01. Massachusetts Institute of Technology Press. Cambridge, Mass.. ISBN: 0262133601.

The book in the first list is a must to have for each student.

6.2 Other Resources

1. <http://www.youtube.com/watch?v=nfoudtpBV68&list=PL6397E4B26D00A269>

7. Schedule

Theory

Session/ Mode	Related LO	Topics	References
1 F2F	LO 1	Natural Language Processing - What is Natural Language Processing (NLP)? - Knowledge in Speech and Language Processing - Ambiguity - Models and Algorithms - Language, Thought, and Understanding - The State of the Art - Some Brief History	- Natural Language Processing - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing, - Introduction to NLP, http://www.youtube.com/watch?v=nfoudtpBV68&list=PL6397E4B26D00A269
2 F2F	LO 1 LO 2	Regular Expressions and Automata - Regular Expressions - Regular Expressions Patterns - Finite-State Automata - Formal Language - Non-Deterministic Automata - Using an NFSA to Accept Strings - Regular Languages and FSAs	- Regular Expressions and Automata - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
3 GSLC	LO 1 LO 2	Words and Transducers - Orthographic and Morphological Rules - Survey of English Morphology - Combine Morphemes to Create Words - Inflectional Morphology - Derivational Morphology - Cliticization	- Words and Transducers - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition,

		<ul style="list-style-type: none"> - Non-concatenative Morphology - Agreement - Finite-State Morphological Parsing - Building a Finite-State Lexicon - Finite-State Transducers - Lexicon Free FSTs: The Porter Stemmer - Word and Sentence Tokenization 	<ul style="list-style-type: none"> - Foundations of statistical natural language processing,
4 F2F	LO 1 LO 2	<ul style="list-style-type: none"> N-grams - N-Grams - Counting Words in Corpora - Unknown Words: Open versus closed vocabulary tasks - Perplexity - Smoothing 	<ul style="list-style-type: none"> - N-grams - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
5 F2F	LO 2	<ul style="list-style-type: none"> Part-of-Speech Tagging - Parts-of-Speech - English Word Class - Tagsets for English - Part-of-speech Tagging - Rule Based Part-of-speech Tagging - Stochastic Part-of-speech Tagging - Transformation Based Tagging 	<ul style="list-style-type: none"> - Part-of-Speech Tagging - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
6 F2F	LO 2	<ul style="list-style-type: none"> Hidden Markov and Maximum Entropy Models - Sequence Classifier - Markov Chain - The Hidden Markov Models - Computing Likelihood: The Forward Algorithm - Decoding: The Viterbi Algorithm - Training HMMs: The Forward-Backward Algorithm - Maximum Entropy Models - Linear Regression - Logistic Regression 	<ul style="list-style-type: none"> - Hidden Markov and Maximum Entropy Models - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
7 F2F	LO 2 LO 3	<ul style="list-style-type: none"> Syntactic Parsing - Grammar and Lexicon - Syntactic Parsing - Top-Down Parsing - Bottom-Up Parsing - Comparing Top-Down and Bottom-Up Parsing - Ambiguity - Structural Ambiguity - Local Ambiguity - Partial Parsing - Finite-State Rule-Based Chunking - Machine-Learning Approaches to Chunking - Evaluating Chunking System 	<ul style="list-style-type: none"> - Syntactic Parsing - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
8 F2F	LO 2 LO 3	<ul style="list-style-type: none"> Statistical Parsing - Statistical Parsing - Probabilistic Context-Free Grammars - PCFGs for Disambiguation - PCFGs for Language Modeling - Problems with PCFGs - Evaluating Parsers 	<ul style="list-style-type: none"> - Statistical Parsing - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
9 GSLC	LO 3	<ul style="list-style-type: none"> Computational Semantics - Syntax-driven Semantic Analysis 	<ul style="list-style-type: none"> - Computational Semantics - Speech and Language

		<ul style="list-style-type: none"> - Semantic Augmentations to CFG Rules - Quantifier Scope Ambiguity and Underspecification - Unification-Based Approaches to Semantic Analysis - Semantic Attachments for a Fragment of English - Idioms and Compositionality 	Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
10 F2F	LO 3	Lexical Semantics <ul style="list-style-type: none"> - Lexical Semantics - Word Senses - Relations between Senses - Wordnet: A Database of Lexical Relations - Event Participants: Semantic Roles and Selectional Restrictions - Primitive Decomposition - Metaphor 	- Lexical Semantics - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
11 GSLC	LO 3 LO 4	Computational Discourse <ul style="list-style-type: none"> - Coherence - Discourse Segmentation - Text Coherence - Reference Resolution - Reference Phenomena - Coreference Resolution 	- Computational Discourse - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
12 F2F	LO 4	NLP Applications <ul style="list-style-type: none"> - Information Extraction - Question Answering and Summarization - Dialogue and Conversational Agents - Machine Translation 	- NLP Applications - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
13 F2F	LO 2 LO 3 LO 4	Project Presentation <ul style="list-style-type: none"> - Project Presentation - Project Evaluation - Project Submission 	- Project Presentation - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,

Practicum

Session/ Mode	Related LO	Topics	References
1 F2F	LO 1 LO 2	Introduction to Python <ul style="list-style-type: none"> - Input Output - Variable - Selection - Repetition - Creating List and Set - Accessing List and Set - Slicing List and Set - Function - Function Parameters 	- Introduction to Python - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
2 F2F	LO 1 LO 2	Introduction to Python <ul style="list-style-type: none"> - Input Output 	- Introduction to Python - Speech and Language

		<ul style="list-style-type: none"> - Variable - Selection - Repetition - Creating List and Set - Accessing List and Set - Slicing List and Set - Function - Function Parameters 	Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
3 F2F	LO 2 LO 3 LO 4	Natural Language Toolkit <ul style="list-style-type: none"> - Tokenizing - Stop Words - Stemming - Lemmatizing - Part Of Speech Tagging - Named Entity Recognition - Frequency Distribution - Load Corpora from NLTK Data - Load Corpora from Web - WordNet - Extract Text Features - Naïve Bayes - Save Classification with Pickle 	- Natural Language Toolkit - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
4 F2F	LO 2 LO 3 LO 4	Natural Language Toolkit <ul style="list-style-type: none"> - Tokenizing - Stop Words - Stemming - Lemmatizing - Part Of Speech Tagging - Named Entity Recognition - Frequency Distribution - Load Corpora from NLTK Data - Load Corpora from Web - WordNet - Extract Text Features - Naïve Bayes - Save Classification with Pickle 	- Natural Language Toolkit - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,
5 F2F	LO 2 LO 3 LO 4	Natural Language Toolkit <ul style="list-style-type: none"> - Tokenizing - Stop Words - Stemming - Lemmatizing - Part Of Speech Tagging - Named Entity Recognition - Frequency Distribution - Load Corpora from NLTK Data - Load Corpora from Web - WordNet - Extract Text Features - Naïve Bayes - Save Classification with Pickle 	- Natural Language Toolkit - Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, - Foundations of statistical natural language processing,

8. Evaluation

Theory

Assessment Activity	Weight	Learning Outcomes			
		1	2	3	4
Assignment	30%	√	√	√	
Mid Exam	35%		√	√	
Final Exam	35%		√	√	√

Practicum

Assessment Activity	Weight	Learning Outcomes			
		1	2	3	4
Assignment	40%	√	√	√	√
Final Exam	60%	√	√	√	√



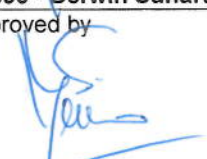

Final Evaluation Score

Aspects	Weight
Theory	80%
Practicum	20%

9. A. Assessment Rubric (Study Program Specific Outcomes)

LO	Indicators	Proficiency Level			
		Excellent (85 – 100)	Good (75 – 84)	Average (65 – 74)	Poor (<= 64)
LO1	1.1. Ability to identify fundamental issues occur in Natural Language Processing	Fundamenta l issues occur in NLP are clearly identified	Fundamenta l issues occur in NLP are identified	Fundamenta l issues occur in NLP are not well identified	Fundamenta l issues occur in NLP are not identified
	1.2. Ability to explain utilization of NLP techniques	Utilization of NLP techniques is totally mastered	Utilization of NLP techniques is affordable	Utilization of NLP techniques is not really realized	Utilization of NLP techniques is unknown
LO 2	2.1. Ability to explain fundamental concepts of NLP	fundamental concepts of NLP are totally mastered	fundamental concepts of NLP are affordable	fundamental concepts of NLP are not really realized	fundamental concepts of NLP are unknown
	3.1. Ability to relate one concept to another	Relationship one concept to another is clearly identified	Relationship one concept to another is identified	Relationship one concept to another is not well defined	Relationship one concept to another is not defined
LO 3	3.1. Ability to build solutions from NLP concepts	Solutions from NLP concepts are very well constructed	Solutions from NLP concepts are constructed	Solutions from NLP concepts are incomple ted	Solutions from NLP concepts are not defined
	3.2. Ability to extract implicit meanings from sentences	Implicit meanings from sentences are extracted	Implicit meanings from sentences are partly extracted	Implicit meanings from sentences are unconvincin g	Implicit meanings from sentences are not extracted
LO 4	4.1. Ability to build NLP applications	NLP applications are well built without any mistakes	NLP applications are well built	NLP applications are built but still have many mistakes	NLP applications are not able to be built

	4.2. Ability to evaluate NLP applications	NLP applications are very well evaluated	NLP applications are appropriately evaluated	NLP applications are little bit evaluated	NLP applications are not able to be evaluated
--	---	--	--	---	---

Prepared by  D3690 - Derwin Suhartono, S.Kom., M.T.I.	Checked by  D3690 - Derwin Suhartono, S.Kom., M.T.I. Acting as Subject Content Specialist
Approved by  D3690 - Derwin Suhartono, S.Kom., M.T.I. Subject Content Coordinator	Acknowledged by  D2923 - Yen Lina Prasetio, S.Kom., M.Comp.Sc. Head of Program - Computer Science

