The Hong Kong University of Science and Technology (Guangzhou)

Syllabus for DSAA 3051: Introduction to Natural Language Processing and Knowledge Engineering

Course Title: Introduction to Natural Language Processing and Knowledge Engineering

Course Code: DSAA 3051

No. of Credits: 3

Any pre-/co-requisites: DSAA 2012 Deep Learning

Instructor Name: Jiaheng WEI

Email: jiahengwei@hkust-gz.edu.cn

Office Hours: TuTh 7:30 PM -9:00 PM (E3-305)

TA's Name: Yuting ZHANG, Zhen SUN

Office Hours: Flexible (Remotely via Email)

Course Description

This course provides a comprehensive overview of modern approaches for (1) natural language processing and (2) knowledge graphs. Students will explore text processing, linguistic preprocessing, and foundational language models, progressing from n-grams to word embeddings and neural language models. The curriculum covers semantic and discourse analysis, information extraction, and the integration of knowledge bases. Key topics include knowledge-intensive tasks, and the construction and application of knowledge graphs, along with ontologies and semantic web technologies like RDF, OWL, and SPARQL. The course also delves into large language models, emphasizing transformer architectures and ethical considerations of NLP and knowledge systems such as bias, fairness, and privacy. Through lectures, discussions, and practical exercises, students will gain critical insights into the development, maintenance, and application of NLP and knowledge systems, preparing them to address real-world challenges in these fields.

Intended Learning Outcomes (ILOs)

Upon completion of this course, students are expected to be able to do the following:

	Course ILOs	**Nature of the learning outcomes** *(A - Knowledge/Content Related; B - Academic Skills/Competencies; C - Others)*
1	Explain the fundamental concepts of natural language processing.	A
2	Analyze fundamental tasks in NLP.	A
3	Use some hands-on skills for solving NLP problems.	A
4	Differentiate some classic and state-of-the-art techniques.	A
5	Communicate effectively to convey scientific knowledge.	A, B

6	Develop a broad interest in NLP and Knowledge Engineering.	A, C

Weekly schedule & Weekly ILOs

A weekly schedule and corresponding ILOs for clear, week-by-week guidance is attached below.

Week	Topics	Weekly ILOs
1	Introduction to NLP and Knowledge Engineering	Course structure and expectations; Historical context and evolution of NLP and Knowledge Engineering.
2	Text Processing and Linguistic Preprocessing	Tokenization, stemming, and lemmatization; Introduction to semantics: word senses and meanings.
3	N-Gram Models and Basics of Language Modeling	N-Gram language models and their applications; Limitations of statistical language models.
4	Word Embeddings and Semantic Representation	Word2Vec, GloVe, and fastText; Semantic similarity and word sense disambiguation.
5	Neural Language Modeling	From word embeddings to contextual embeddings; Neural sequence labeling.
6	Information Extraction and Knowledge Bases	Relation extraction and entity linking; Introduction to knowledge bases and their integration with NLP.
7	Midterm Exam	
8	Ontologies and Semantic Web Technologies	Basics of ontologies and their role in knowledge engineering; Introduction to RDF, OWL, and SPARQL
9	Knowledge Graphs Construction, Maintenance and Applications	Techniques for constructing and maintaining knowledge graphs; Knowledge graphs for downstream applications.
10	Applications of Knowledge Engineering in NLP	Knowledge-intensive tasks: entity linking and semantic search; Integration of KG and KB with NLP systems.
11	Large Language Models	Transformer architecture to the generative paradigm; Pre-training, supervised fine-tuning and RLHF.
12	Safety and Security: Bias, Fairness and Privacy	Additional Considerations of NLP systems; Safeguard the NLP system.
13	Summary and Discussions	
14	Final Exam	

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment will be provided on the released date, outlining the criteria used for evaluation.

Assessments:

[List specific assessed tasks, exams, quizzes, their weightage, and due dates; perhaps, add a summary table as below, to precede the details for each assessment.]

Assessment Task	Contribution to Overall Course grade (%)	Due date
Mid-Term Test	20%	Oct 16th, 2025 (6:00PM—8:00PM)
Course Participation	10%	N/A
Written Assignment	30%	3-4 times
Final examination	40%	Dec 4th, 2025 (6:00PM—8:00PM)

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs
Course Participation [10%]	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5, CILO-6
Written Assignments [30%]	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5, CILO-6
Midterm test [20%]	CILO-1, CILO-2, CILO-3
Final exam [40%]	CILO-4, CILO-5, CILO-6

Grading Rubrics

Detailed rubrics for each assignment will be provided. These rubrics clearly outline the criteria used for evaluation. Students can refer to these rubrics to understand how their work will be assessed.

Final Grade Descriptors:

As appropriate to the course and aligned with university standards

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a comprehensive grasp of subject matter, expertise in problem-solving. Exhibits a high capacity for scholarship and collaboration, going beyond core requirements to achieve learning goals.
B	Good Performance	Shows good knowledge and understanding of the main subject matter, competence in problem-solving, and the ability to analyze and evaluate issues. Displays high motivation to learn and the ability to work independently.
C	Satisfactory Performance	Possesses adequate knowledge of core subject matter, competence in dealing with familiar problems, and some capacity for analysis. Shows persistence and effort to achieve broadly defined learning goals.
D	Marginal Pass	Has threshold knowledge of core subject matter, potential to achieve key professional skills, and the ability to make basic judgments. Benefits from the course and has the potential to develop in the discipline.
F	Fail	Demonstrates insufficient understanding of the subject matter and lacks the necessary problem-solving skills. Shows limited ability to think critically or analytically and exhibits minimal effort towards achieving learning goals. Does not meet the threshold requirements for professional practice or development in the discipline.

Course AI Policy

Homework assignments will be done individually. All help from others (from the web, books other than text, LLMs, or people other than the TA or instructor) must be clearly acknowledged. Presenting other's work as your own is dishonest and is called plagiarism. Code similarity will be checked. The teaching staffs reserve the right to reach out to individual students to check on their understanding of the submitted homework solutions.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include [specific details, e.g., strengths, areas for improvement]. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission & Late Day Policy

Each student has a total of **7 late days** for use on assignments. They extend the due date by **24 hours** and a maximum of **3 late days** can be used towards any individual assignment.

Grading Written and Programming assignments will be hand graded. If you have used up your 7 late days, you will be penalized 25% per day, up to two days max, with no credit after two days. In cases of medical or other emergencies which interfere with your work, or resubmission needs, contact the instructor for approval is required.

Required Texts and Materials

Reading materials will be available in each week's lecture slides.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST(GZ)'s Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Regulations for Academic Integrity and Student Conduct for the University's definition of plagiarism and ways to avoid cheating and plagiarism.