COURSE SYLLABUS  
**CSC15008 – Natural Language Processing Applications**

# GENERAL INFORMATION

|  |  |
| --- | --- |
| Course name: | Natural Language Processing Applications |
| Course name (in Vietnamese): | Xử lý ngôn ngữ tự nhiên ứng dụng |
| Course ID: | CSC15008 |
| Knowledge block: | Elective - Computer Science |
| Number of credits: | 4 |
| Credit hours for theory: | 45 |
| Credit hours for practice: | 30 |
| Credit hours for self-study: | 90 |
| Prerequisite: | Introduction to Natural Language Processing |
| Prior-course: | None |
| Instructors: | Nguyen Hong Buu Long |

# COURSE DESCRIPTION

This course offers a comprehensive introduction to the essential applications in the realm of Natural Language Processing (NLP). It focuses on providing students with a foundational understanding of key NLP tasks including text classification, text similarity, text summarization, and machine translation, among others. Throughout the course, students will engage in practical exercises that allow them to experience firsthand the development of basic NLP applications. These exercises are designed to be systematic and effective, ensuring a thorough understanding of the process. A notable aspect of the course is its incorporation of Large Language Models (LLMs), which have garnered significant interest in recent times. This inclusion enriches the course content and equips students with current and relevant knowledge in the field of NLP.

# COURSE GOALS

At the end of the course, students are able to

|  |  |  |
| --- | --- | --- |
| **ID** | **Description** | **Program LOs** |
| G1 | Work on an individual level and collaborate with a team to build reports on NLP applications | 2.2, 2.3.1 |
| G2 | Understand and explain English terminologies in NLP | 1.7.4, 2.4.3, 2.4.5 |
| G3 | Develop skills in describing, analyzing and modeling a real-world NLP application | 1.7.1, 1.7.4 |
| G4 | Know and understand metrics to evaluate NLP applications | 1.7.4 |
| G5 | Construct some simple NLP applications | 1.7.4, 5.1.1, 5.1.3, 5.3.2, 6.1.1 |
| G6 | Understand NLP toolkits and libraries | 1.3.6 |

# COURSE OUTCOMES

|  |  |  |
| --- | --- | --- |
| **CO** | **Description** | **I/T/U** |
| G1.1 | Establish, organize, operate and manage groups | I, T |
| G1.2 | Participate in group discussions on subject topics | U |
| G1.3 | Analyze, synthesize and write technical documents according to a given template individually or in group collaboration | I, T |
| G2.1 | Know and understand specialized English terminologies in NLP | I |
| G2.2 | Read and understand English documents/references related to class lectures | I |
| G2.3 | Understand similarities and differences between English and Vietnamese | I |
| G3.1 | Understand the concepts of corpora, types of corpora and how to collect a corpus for a specific NLP application | I,T |
| G3.2 | Analyse and establish a plan for building a specific NLP application | I,T,U |
| G3.3 | Understand and evaluate the role of NLP applications in practice | I,T,U |
| G4.1 | Ability to comment, evaluate the quality of NLP applications | I,T |
| G5.1 | Describe the main features of a specific NLP application | I,T,U |
| G5.2 | Outline processes for analysis, design of NLP applications | I,T,U |
| G5.3 | Systematically and methodically build basic components at a simple level | I,T,U |
| G6.1 | Use some tools (Label Studio…) and support libraries (NLTK, CoreNLP…) | I,T,U |

# TEACHING PLAN

**THEORY**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Topic** | **Course outcomes** | **Teaching/Learning Activities (samples)** | **Assessments** |
| 1 | Introduction & Problems in NLP | G1.1, G1.2, G1.3, G2.1, G2.2, G2.3 | **At home**:   * Read slides   **In class**:   * Discussion * Q&A * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A |
| 2 | NLP applications: An introduction | G1.2, G1.3, G2.1, G2.2, G2.3, G3.1, G3.2 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A |
| 3 | Large Language Models (LLMs) | G3.1, G3.2, G3.3, G4.1 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A |
| 4 | Text classification | G3.1, G3.2, G3.3, G4.1 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A * QZ1 * HW1 |
| 5 | Text similarity | G3.1, G3.2, G3.3, G4.1 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A |
| 6 | Text summarization | G3.1, G3.2, G3.3, G4.1 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A |
| 7 | Machine translation | G3.1, G3.2, G3.3, G4.1 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A |
| 8 | Question-Answering | G3.1, G3.2, G3.3, G4.1 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A * QZ2 * HW2 |
| 9 | Chatbot | G3.1, G3.2, G3.3, G4.1 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A |
| 10 | NLP applications with LLMs: Deployments and MLOps | G5.1, G5.2, G5.3, G6.1 | **At home**:   * Read slides   **In class**:   * Lecturing | **At home**:   * Discussion   **In class**:   * Q&A |
| 11 | Seminar | G1.1, G1.3,  G3.1, G3.2, G3.3, G4.1, G5.1, G5.2, G5.3, G6.1 | **Student**: the group presents their projects.  **Teacher**: organize discussions, help students to ask questions and answer groups. Teacher comments, evaluates, explains and gives further instructions. | Project1  Project2 |

**LABORATORY**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Topic** | **Course outcomes** | **Teaching/Learning Activities (samples)** | **Assessments** |
| 1 | Language models | G5.1, G5.3 | Provide instructions  Demonstration, discussion | LW1 |
| 2 | Text classification | G5.1, G5.3 | Provide instructions  Demonstration, discussion | LW2 |
| 3 | Text clustering | G5.1, G5.3 | Provide instructions  Demonstration, discussion | LW3 |
| 4 | Text similarity | G5.1, G5.3 | Provide instructions  Demonstration, discussion | LW4 |
| 5 | Text summarization | G5.1, G5.3 | Provide instructions  Demonstration, discussion | LW5 |

# ASSESSMENTS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Topic** | **Description** | **Course outcomes** | **Ratio (%)** |
| **A1** | **Assignments** |  |  | **30%** |
| A11 | Quizzes: QZ1 and QZ2. | Small quizzes in class for each topic | G5.1, G5.2 | 5% |
| A12 | Homework: HW1 and HW2 | HW1, HW2: practicing based on knowledge taught in class | G5.1, G5.2, G5.3 | 5% |
| A13 | Weekly lab work: LW1-LW5 |  | G5.1, G5.2, G5.3 | 20% |
| **A2** | **Projects** |  |  | **70%** |
| A21 | Project1: Seminar | Seminar about a specific NLP application | G1.1, G1.2, G1.3, G2.1, G2.2, G2.3, G3.1, G3.2, G3.3, G4.1 | 30% |
| A22 | Project2: Application | Implement and write report about a specific NLP application | G1.1, G1.2, G1.3, G2.1, G2.2, G2.3, G3.1, G3.2, G3.3, G4.1, G5.1, G5.2, G5.3, G6.1 | 40% |

# RESOURCES

# Textbooks

* Xử lý ngôn ngữ tự nhiên , Đinh Điền, 2006, Tp HCM: NXB. ĐHQG-HCM
* Speech and Language Processing: An Introduction to Natural Language Processing, Dan Jurafsky and James H. Martin, Computational Linguistics, and Speech Recognition. Third Edition draft, 2023
* Measurement of Text Similarity: A Survey, Jiapeng Wang and Yihong Dong, Information 2020
* Neural Machine Translation, Philipp Koehn, Cambridge University Press 2020
* Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems, Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana. O’Reilly 2020.
* Mark Treveil et. al., 2020, Introducing MLOps, O'Reilly Media, Inc.

# Papers

* Shervin Minaee et. al., Large Language Models: A Survey, 2024, url: https://arxiv.org/abs/2402.06196
* Chien Van Nguyen et. al., A Survey of Small Language Models, 2024, url: https://arxiv.org/abs/2410.20011
* Zishan Guo et. al., Evaluating Large Language Models: A Comprehensive Survey, 2023, url: https://arxiv.org/abs/2310.19736
* Ehsan Kamalloo et. al., 2023. Evaluating Open-Domain Question Answering in the Era of Large Language Models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 5591–5606, Toronto, Canada. Association for Computational Linguistics.
* Xiaofei Sun et. al., Text Classification via Large Language Models, 2023, url: https://arxiv.org/abs/2305.08377
* Lochan Basyal and Mihir Sanghvi, Text Summarization Using Large Language Models: A Comparative Study of MPT-7b-instruct, Falcon-7b-instruct, and OpenAI Chat-GPT Models, 2023, url: https://arxiv.org/abs/2310.10449
* Yunfan Gao et. al., Retrieval-Augmented Generation for Large Language Models: A Survey, 2023, url: <https://arxiv.org/abs/2312.10997>

# GENERAL REGULATIONS & POLICIES

* All students are responsible for reading and following strictly the regulations and policies of the school and university.
* Students who are absent for more than 3 theory sessions are not allowed to take the exams.
* For any kind of cheating and plagiarism, students will be graded 0 for the course. The incident is then submitted to the school and university for further review.
* Students are encouraged to form study groups to discuss the topics. However, individual work must be done and submitted on your own.