1. **Course Number and Course Title:**
   ML 513 – Advanced Natural Language Processing

2. **Credit Hours:** 3 – 0 – 3

3. **Prerequisites and/or Co-Requisites:**
   Prerequisite: Approval of the CSE Head of Department
   Co-requisites: None
   Competencies: Undergraduate-level knowledge of programming and language grammars.

4. **Name and Contact Information of Instructor:**
   Name: Dr. Michel Pasquier
   Email: mpasquier@aus.edu
   Office: ESB 2057
   Phone: 06 515 2883
   Office Hours: Posted on office door and iLearn; also by appointment

5. **Course Description (Catalog Description):**
   Introduces the fundamental concepts and algorithms for Natural Language Processing (NLP). Covers text classification, language modeling, vector semantics, word embedding, sequence labeling, language grammars, constituency parsing, dependency parsing. Examines advanced applications such as information extraction, sentiment analysis, conversational agents, machine translation, and text generation.

6. **Textbook and other Supplemental Material:**
   Textbook:
   - D. Jurafsky, J. H. Martin, Speech and Language Processing, 3rd ed., Prentice Hall 2022.

   Supplemental material:
   - J. Eisenstein, Natural Language Processing, Lecture notes, 2018.
   - Selected articles and journal papers, NLP software tools and data.

7. **Course Learning Outcomes:**
   Upon completion of the course, students will be able to:
   1. Appraise the growing importance of Natural Language Processing in many domains.
   2. Categorize the fundamental concepts and techniques of Natural Language Processing.
   3. Design a text classifier using various text features for any task.
   4. Evaluate different language characteristics and models.
   5. Assess various algorithms for sampling, smoothing, and sequence labeling.
   6. Compare different parsing algorithms and their applications.
   7. Formulate advanced NLP tasks such as for dialogue, translation, speech recognition.
   8. Select state-of-the-art software tools to realize Natural Language Processing tasks.

8. Teaching and Learning Methodologies:

Methods include lectures, problem and project-based learning methods (assignments, exams, research project, presentation), and class discussions.

9. Course Topics and Schedule:

Topic/Activity	Weeks
Introduction to Natural Language Processing	Week #1
Text classification, bag of words, tokens, Naive Bayes classifier	Week #2
Logistic regression, Neural networks for text classification	Week #3
Convolutional neural networks, sentiment analysis	Week #4
Language models, N-grams, sampling, smoothing	Week #5
Lexical semantics, vector semantics, similarity measures	Week #6
Word embedding, models, biases, evaluation	Week #7
Sequence labeling, part-of-speech tagging, Viterbi algorithm	Week #8
Midterm exam	Week #9
Hidden Markov Models, neural sequence labeling	Week #10
Context-free grammars, constituency parsing	Week #11
Dependency relations, dependency parsing	Week #12
Conversational agents, dialogue, questions and answers	Week #13
Information extraction, speech recognition, text-to-speech	Week #14
Machine translation, statistical and neural approaches	Week #15
Final Exam	Week #16

10. Schedule of Laboratory and other Non-Lecture Sessions:

The semester-long project will involve one to three students and will focus on Natural Language Processing, either focusing on core NLP methods or using NLP in support of an empirical research question. The project will comprise four components: (1) Project proposal: Students will propose the research question to be examined, motivate its rationale as an interesting question worth asking, and assess its potential to contribute new knowledge by situating it within related literature in the scientific community. (2) Intermediate report: By the middle of the course, students should present initial experimental results and establish a validation strategy to be performed at the end of experimentation. (3) Presentation: At the end of the semester, teams will present their work in a video/demo. (4) Final report: It will include a complete description of the work undertaken for the project, including data collection, development of methods, experimental details (complete enough for replication), comparison with past work, and a thorough analysis. Projects will be evaluated according to standards including clarity, originality, soundness, substance, evaluation, meaningful comparison, and impact (of ideas, software, and/or datasets). Students will be evaluated individually based on their respective contributions.

11. Out-of-Class Assignments with Due Dates:

Assignment	Due Date (tentative)
Homework 1 – Text classification	Week #4
Homework 2 – Language modeling	Week #7
Homework 3 – Labeling and parsing	Week #9

Homework 4 – Applications of NLP	Week #11
Research project – 1 Proposal	Week #5
Research project – 2 Intermediate report	Week #8
Research project – 3 Presentation/video	Week #13
Research project – 4 Final report	Week #15

12. Student Evaluation:

Assessment	Weight	Due Date (tentative)
Homework	15 %	cf. section 11
Research Project	30 %	cf. section 10
Midterm Exam	25 %	Week #9
Final Exam	30 %	Week #16

13. Assessment Instruments:

Assessment	Course Learning Outcomes
Homework	O2–O6
Research Project	O3–O8
Midterm Exam	O1–O5
Final Exam	O2–O7

14. Contribution of Course to Program Outcomes:

MSML Program Outcomes	Emphasis in this course	Course Learning Outcomes
1. Perform research emphasizing creativity, independent learning, and scientific methods in the field of Machine Learning.	◑	O3–O8
2. Apply advanced mathematics, computer science knowledge, and software tools in identifying, formulating, and solving real world problems.	●	O2–O7
3. Demonstrate an in-depth understanding of modern Machine Learning approaches, algorithms, and tools.	○	O2–O3, O7
4. Select and use techniques, skills, and modern tools necessary for research or professional practice.	◑	O3–O8
5. Communicate effectively through technical presentations and reports.	○	O1–O7
6. Recognize the need for, and engage in, lifelong learning in professional areas.		
7. Attend to professional and ethical responsibilities within global and societal contexts.	○	O1, O4, O7

Emphasis: ● High; ◑ Medium; ○ Low; Blank – Nothing Specific Expected

15. Letter Grade Policy:

Total (T)	Letter Grade
$90 \leq T$	A
$85 \leq T < 90$	A-
$80 \leq T < 85$	B+
$75 \leq T < 80$	B
$70 \leq T < 75$	B-
$65 \leq T < 70$	C+

$60 \leq T < 65$	C
$T < 60$	F