CSIS-503: NATURAL LANGUAGE PROCESSING

Effective Term
Fall 2026

BOT Approval Date
05/15/2025

Discipline
CSIS - Computer Sci/Info Systems
Course Number
503

Course Title
Natural Language Processing

Short Title
Natural Language Processing

Credit Status (CB 04)
D - Credit - Degree applicable

Units

Lecture Units
3.00

Total Units
3.00

Hours

Lecture Contact Hours
48-54
Out of Class Hours Lecture
96-108

Total Contact Hours
48 - 54

Total Out of Class Hours
96 - 108

Total Student Learning Hours
144 - 162

Distance Education
Yes

Distance Education

Distance Education Type
Both Fully Online and Hybrid Online

Fully Online Delivery Requirements:
a. Students must be notified via the college schedule of classes and the syllabus for the class if proctored tests are required for this course
b. Any planned face-to-face meetings, such as an orientation or study session, must be optional

c. The MSJC curriculum committee requires the use of accessible, asynchronous discussion as a component of every fully online course

Are you using publisher content?

No

Regular Substantive Interaction

Instructor-to-student, student-to-student, and student-to-content contact shall be distributed in a manner that will ensure regular substantive interaction (RSI) is maintained not only weekly but also for the duration of the course. Instructor interactions should be equivalent in time, quality and consistency to the engagement students would receive in a face-to-face course—for example, through announcements, discussion participation, personalized feedback, video messages, or other instructional communications. Instructors are expected to be responsive to student inquiries within a 24-72 hour time frame excluding holidays. RSI will be maintained through a variety of methods: • Orientation - Students must be oriented to the online and face-to-face portions at the start of the course. • Announcements - Announcements using the course management system must be posted at least weekly to keep students current on course events, due dates, materials, etc. • Discussion Forums - Content-related discussion forums within the designated CMS will include appropriate instructor participation and will be initiated and maintained, with timely feedback provided. Instructors are required to monitor the discussions and related activities to provide guidance and direction. Open-ended discussion boards may be used to further learning and interaction. • MSJC Communication – Instructor will use MSJC-administered communication tools (such as email or the designated CMS communication tools) to regularly communicate with and provide ease of access to students. • Timely Feedback - Timely feedback on student work will be provided by the instructor, which may include comments and/or rubrics, within the Grades area of the designated CMS. • Scheduled Meetings – Hybrid and synchronous courses will meet at regularly scheduled times. ADA Compliance - All course interactions and content must be designed with accessibility in mind and meet ADA requirements.

Catalog Description

This course explores the fundamental concepts of natural language and text processing and how it enables machines to understand, interpret, and generate human language in a meaningful and contextually relevant way. Students will investigate concepts to manipulate text for language models and topic modeling.

Requisites

Prerequisite(s)

Prerequisite(s) (must be taken before)
CSIS-502 (with a grade of C or better)

Codes

TOP Code (CB 03)

0702.00* - *Computer Information Systems

CIP Code

11.0102 - Artificial Intelligence.

Repeatability Code

No

Grading Option

Letter grade OR P/NP

Minimum Qualifications

Minimum Qualifications	And/Or
Computer Information Systems	End

Learning Objectives

Learning Objectives

Learning Objective
1. Develop an understanding of the foundational concepts and principles of natural language processing.
2. Evaluate text processing techniques such as tokenization, stemming, and lemmatization.

3. Investigate linguistic fundamentals relevant to natural language processing, including part of speech tagging and syntactic analysis.
4. Assess the application of supervised and unsupervised learning techniques for natural language processing tasks.
5. Investigate differing language models, including N-gram models and word embedding.
6. Develop skills in text classification and sentiment analysis.
7. Develop an understanding of semantic analysis, including semantic role labeling and discourse analysis.
8. Investigate advanced natural language processing topics, including neural machine translation and question-answering systems.
9. Evaluate the ethical implications of natural language processing applications, including biases in language models.
10. Evaluate the presence of racial and gender bias in natural language processing models, datasets, and algorithms

Learning Outcome Information

Course Learning Outcomes

Learning Outcome
1. Students will describe natural language processing and define the use of current applications.
2. Students will apply string processing techniques to solve natural language processing problems.
3. Students will employ machine learning algorithms for text classification and other language tasks.
4. Students will use topic modeling to identify clusters or groups of similar words.

Content

Course Lecture Content

1. Introduction to natural language processing
   a. Overview of natural language processing and its applications.
   b. Historical perspectives.
   c. Evolution of natural language processing.
   d. Applications of natural language processing.
   e. Algorithms used in natural language processing.
2. Preprocessing techniques
   a. Tokenization.
   b. Stop-word removal.
   c. Lemmatization.
   d. Tagging.
   e. Syntactic analysis.
3. Statistical and machine learning approaches
   a. Statistical models.
   b. Supervised learning techniques.
   c. Unsupervised learning techniques.
   d. Feature extraction.
4. Language Modes and Word Embedding
   a. N-gram models
   b. Word embedding.
   c. Transformer models.
   d. Attention mechanisms.
5. Information retrieval and extraction
   a. Text classifications.
   b. Sentiment analysis.
   c. Named entity recognition (NER).
   d. Information extraction.
   e. Coreference resolution.
6. Syntax and semantic analysis

a. Dependency parsing.
b. Semantic role labeling.
c. Discourse analysis.

7. Deep learning for natural language processing
   a. Recurrent neural networks.
   b. Long short-term memory.
   c. Sequence-to-sequence models.
   d. Transfer learning.
   e. Fine-tuning for natural language processing.
8. Machine learning algorithms for natural language processing
   a. Neural networks.
   b. Transform words to vectors.
   c. Potential applications of vectors.
   d. Text classification.
   e. Information retrieval.
   f. Markov chains.
   g. Seq2seq.
9. Diversity, equity, and inclusion in Natural Language Processing
   a. Why DEI Matters in Natural Language Processing.
   b. How language technology affects diverse populations
   c. The importance of inclusive design in language models and applications.
   d. How linguistic bias is embedded in training data and models.
   e. Challenges of building representative datasets.
   f. Ethical considerations in data collection

Methods of Instruction

Method

Lecture

Integration

Lectures will be presented with supporting visual materials (slide sets or multimedia). They will introduce conceptual and practical skills such as tokenization, stemming, lemmatization, supervised and unsupervised learning, natural language models, and the ethical implications of natural language processing applications.

DE Adaptations for MOI

Lectures will be delivered using video capture software to embed voice into lecture slides. Larger videos will be broken down into shorter digestible segments. All videos will be closed-captioned with synchronized audio. Transcripts of the lecture will be provided upon request. Any included Images will have alternate text that provides a detailed image description. If external links are used, they will have descriptive text explaining the link's destination. Any resources or documents will comply with accessibility standards. All lectures will be delivered using the course management system. Lectures will encourage students to ask questions through the system's discussion board.

Method

Observation and Demonstration

Integration

Interactive demonstrations will consist of coding sessions to solve natural language problems found in semantic analysis, text classification, N-gram models, word embedding, supervised and unsupervised learning techniques, and linguistic fundamentals.

DE Adaptations for MOI

Interactive demonstrations will be pre-recorded and delivered via video via the course management system. All videos will be closed-captioned with synchronized text. Any external link within the discussion will have descriptive text explaining the link's destination. Documents embedded in any discussion will be provided in accessible file formats such as HTML, accessible PDF, and text. Images included in discussions will have alternate text provided that provides a detailed image description. Any resources or documents will comply with accessibility standards. The discussion board within the course management system will be used for student-to-student

and student-to-instruction interactions. Interactive notebooks (e.g., Jupyter) will supplement hands-on exercises delivered via the course management system.

Method

Discussion

Integration

Weekly discussions will be conducted to explore natural language processing topics such as the ethical implications of natural language processing applications, bias in language models, semantic analysis and role labeling, supervised and unsupervised learning, and foundational concepts.

DE Adaptations for MOI

Weekly discussions will be administered using the course management system's discussion tool. Images included in discussions will have alternate text provided that provides a detailed image description. Any external link within the discussion will have descriptive text explaining the link's destination. Documents embedded in any discussion will be provided in accessible file formats such as HTML, accessible PDF, and text. Any resources or documents will comply with accessibility standards

Methods of Evaluation

Method

Quizzes

Integration

Quizzes for each unit in the class will consist of multiple choice and short answer questions and small coding problems to evaluate the student's knowledge of tokenization, stemming, lemmatization, speech tagging, syntactic analysis, and neural machine translation.

DE Adaptations for MOE

Quizzes will be administered and collected using the quizzes tool within the course management system. All quizzes will have alternate text provided for images that give a detailed image description. If external links are used, descriptive text will explain the link's destination. All videos will be closed-captioned with synchronized audio. Any resources or documents will comply with accessibility standards. Feedback will be provided via the comments section of the quizzes tool.

Method

Exams/Tests

Integration

A midterm and final exam consisting of small coding problems will be given to evaluate the student's understanding of natural language topics such as supervised and unsupervised learning, language models and their biases, semantic analysis, and neural machine translation.

DE Adaptations for MOE

Exams will be administered and collected using the assignments tool within the course management system. Images included in any quiz question will have alternate text that provides a detailed description of the image. Any external link will have descriptive text explaining the link's destination. All videos will be closed-captioned with synchronized audio. Any resources or documents will comply with accessibility standards. Feedback for exams will be provided via the comments section of the quizzes tool within the course management system.

Method

Homework

Integration

Weekly coding assignments will be given to evaluate the student's understanding of tokenization, stemming, lemmatization, N-gram models, text classification, sentiment analysis, and neural machine translation.

DE Adaptations for MOE

Homework assignments will be provided in accessible file formats such as HTML, accessible PDF, and text. Any external link within the assignment will have descriptive text explaining the link's destination. Any images included will have alternate text provided that provides a detailed description of the image. All videos will be closed-captioned with synchronized audio. Any resources or documents will comply with accessibility standards. Assignments will be delivered via the course management system's assignment tool. Feedback will be provided in the comments section of the tool.

Assignments

Assignment Type

Reading

In-class and/or outside of class

Outside of class

Formative or Summative

Formative

Assignment

The goal of this assignment is to develop a foundation of understanding how neural network models can be used in Natural Language Processing, including language modeling, word embeddings, sequence processing, and deep learning applications.
To gain a better understanding, please read the following:
Goldberg, Y. (2016). A Primer on Neural Network Models for NLP. Journal of Artificial Intelligence Research, 57, 345–420.
Once you have finished reading, please answer the following questions in a two- to three-page report.
1. What are the key challenges in processing natural language computationally?
2. How do word embeddings capture semantic relationships?

3. What are the differences between traditional sequence models (e.g., LSTMs) and transformer-based models?

4. What are the differences between traditional sequence models (e.g., LSTMs) and transformer-based models?
This assignment will be based on the following criteria
• Clarity and depth of understanding (40%)
• Critical thinking and reflection (30%)
• Quality of discussion questions (20%)
• Proper citations and academic writing style (10%)

How assignment will be adapted in online format

Assignments will be distributed and collected through the assignments tool of the course management system. Any external link within the assignment will have descriptive text explaining the link's destination. If images are included as part of the assignment, they will have alternate text provided that provides a detailed description of the image. All videos will be closed-captioned with synchronized audio. Any resources or documents will comply with accessibility standards. Feedback for the assignment will be given via the comments section in the grade book.

Assignment Type

Writing

In-class and/or outside of class

Outside of class

Formative or Summative

Summative

Assignment

For this assignment, you will develop a sentiment analysis tool to classify social media posts into positive, negative, or neutral categories using NLP techniques learned during the course.
Overview:
Sentiment analysis is a common application of NLP, classifying text into different sentiments. This project will require you to collect, preprocess, and analyze social media data (like tweets or Facebook posts) to determine the sentiment expressed in each post.
Collect a dataset of social media posts. This can be done using APIs provided by platforms like X and Reddit or using pre-existing datasets. Ensure the dataset is balanced with a mix of positive, negative, and neutral posts.
Perform necessary preprocessing steps on the dataset, including tokenization, normalization, removing stop words, handling emojis, and stemming or lemmatization.
Extract relevant features from the text using techniques such as TF-IDF, word embeddings, or BERT representations.
Implement a machine learning model for sentiment analysis. This could be a traditional model like Naive Bayes, Logistic Regression, or an advanced model using neural networks. Train the model on a portion of the dataset and validate it on a separate set.
Evaluate the model using appropriate metrics, such as accuracy, precision, recall, and F1-score. Conduct error analysis to understand the types of posts where the model performs poorly.
Write a detailed report documenting the entire process, including data collection, preprocessing, model selection, training, evaluation, and error analysis. Include insights and observations from the project, discussing any challenges and how they were addressed.
Please submit your source code for the data collection, preprocessing, model training, evaluation, and a small presentation summarizing the project findings.

Assignments will be graded using the following rubric
20% - All Directions Are Followed
20% - Accuracy of developed tool
20% - If all categories are covered.
30% - If the program runs to completion with proper input and output
10% - If the project is properly documented

How assignment will be adapted in online format

Assignments will be distributed and collected through the assignments tool of the course management system. Any external link within the assignment will have descriptive text explaining the link's destination. If images are included as part of the assignment, they will have alternate text provided that provides a detailed description of the image. All videos will be closed-captioned with synchronized audio. Any resources or documents will comply with accessibility standards. Feedback for the assignment will be given via the comments section in the grade book.

Assignment Type

Writing

In-class and/or outside of class

In-class

Formative or Summative

Formative

Assignment

Load the pre-trained word embedding model using the Python libraries. Explore the relationships between different word pairs by calculating cosine similarities between their embeddings.
Use analogies to explore bias in embeddings. For example, "man is to king as woman is to X" and investigate the output.

Write a reflection on what the results suggest about gender bias in the embedding model. How might these biases affect downstream NLP tasks like sentiment analysis or machine translation?

Submit your written reflection along with the code you have written.

This assignment will be graded based on the following rubric
30% - All Directions Are Followed
30% - Completeness of similarities
35% - If analogies are accurate and complete

How assignment will be adapted in online format

Assignments will be distributed and collected through the assignments tool of the course management system. Any external link within the assignment will have descriptive text explaining the link's destination. If images are included as part of the assignment, they will have alternate text provided that provides a detailed description of the image. All videos will be closed-captioned with synchronized audio. Any resources or documents will comply with accessibility standards. Feedback for the assignment will be given via the comments section in the grade book.

Course Materials

Textbooks

Natural Language Processing: A Textbook with Python Implementation
Raymond S. T. Lee
Springer November 2023 ISBN 978-9819919987

Class Size Information

Class Size

35

Class Size Category

B - Facilitated learning 35

Diversity and Inclusion

Course Design and Materials: Course will incorporate diverse perspectives, authors, and examples that reflect a variety of cultures, identities, and lived experiences. Instructional materials will be reviewed regularly to ensure representation, inclusivity, and alignment with equity-minded practices across disciplines.

Inclusive Learning Environment: Instructors will establish classroom norms and policies that promote respect, civil discourse, and belonging in both online and face-to-face classes. Clear expectations for communication and engagement will support a learning space where all students feel valued and heard.

Accessible Instruction: Course content will be delivered using accessible design. Materials will follow accessibility guidelines within the CMS, including captioned media, readable document formats, and multiple means of engagement to support diverse learning.

Equitable Teaching Practices: Instructors will use transparent and grading criteria, varied assessment methods, and timely feedback to promote fairness and student success. Opportunities for revision, reflection, and multiple demonstrations of learning may be incorporated when appropriate to the discipline.

Culturally Responsive Pedagogies: Instructional strategies will encourage critical thinking about diverse perspectives and systems of power relevant to the course content. Faculty will foster opportunities for students to connect course material to real-world contexts.

Ongoing Reflection and Improvement: Faculty will engage in continuous professional development related to IDEAA principles and reflect on course practices using metacognition to identify opportunities for increased equity, inclusion, and accessibility.

Explain how diversity and inclusion(e.g., gender, culture/race, sexuality, etc.) is infused into the course.

This class's reading material and case studies will include a wide range of perspectives and voices.
This course will discuss bias in natural language processing and data sets. Discussions will also investigate how biases enter natural language processing and the ethical implications of biased natural language processing applications. In addition, real-world cases where natural language processing had biased, or discriminatory outcomes will be examined. Methods will be explored to mitigate such biases.
The course materials in this class will be available in accessible formats to ensure they are available to students with disabilities. Material in this class will be provided that uses inclusive language and is sensitive to students' diverse backgrounds. This includes being mindful of gender-neutral language and cultural nuances.

Describe how assignments, instruction, evaluation, course content, and interactions are contextualized for diverse learners promoting an equitable course.

Assignments will be designed to be accessible to all students, including those with disabilities, and alternative formats will be provided.
The course includes diverse teaching methods (lectures, discussions, interactive demonstrations) catering to different learning styles and preferences.
Feedback that is constructive and sensitive to cultural differences will be provided to support learning and improvement rather than discouraging diverse viewpoints.

Interactions

Explain how the course will incorporate student-to-student interaction.

Collaborative demonstrations and guided practice on statistical machine learning, language models, information extraction, and semantic analysis will facilitate diverse student interactions. The course discussions will encourage students to share experiences and insights on supervised learning techniques, word embedding, information retrieval, and deep learning techniques.

Students will be assigned group projects where they can collaborate on building natural language models, implementing algorithms, and solving real-world problems using natural language processing techniques.

Peer code reviews allow students to exchange code and provide feedback on each other's projects or assignments. This will promote collaboration, help students learn different coding styles and problem-solving techniques, and reinforce key AI concepts.

Explain how the course will incorporate instructor-to-student interaction.

Regular office hours allow students to discuss course materials, seek project guidance, or address concerns on language processing, neural networks, word embedding, and transfer models. Students will be encouraged to schedule one-on-one meetings for personalized feedback and mentoring.
Discussion boards will be provided where the instructor will participate in discussions about language processing, tokenization, lemmatization, supervised learning techniques, and statistical models.

Explain how the course will incorporate student-to-content interaction.

Examples and case studies presented for discussion will be from various domains, such as healthcare, finance, or social media. The discussion of real-world examples will make the content more relatable and interesting.
The lectures will be interactive so that students will see real-time demonstrations of natural language processing concepts such as text parsing, sentiment analysis, or machine learning model training. During demonstrations, students will be encouraged to ask questions and suggest different parameters or methods.
Guided practice and demonstration problems will encourage students to review course materials and posted examples to formulate solutions.

Explain how the course will incorporate student-to-self interaction (metacognitive).

Discussion threads will be created that will encourage students to share their reflections on topics related to natural language processing, such as ethics, potential applications, or biases found in popular data sets. The lecture will include exercises that will prompt students to think about how they learn best. Students will be encouraged to reflect on their study habits, identify effective learning strategies, and suggest potential adjustments.

Key: 3139