



Week 1 Overview

During this week's lessons, you will learn the overall course design, an overview of natural language processing techniques and text representation, which are the foundation for all kinds of text-mining applications, and word association mining with a particular focus on mining one of the two basic forms of word associations (i.e., paradigmatic relations).

Time

This module should take **approximately 3 hours** of dedicated time to complete, with its videos and assignments.

Activities

The activities for this module are listed below (with required assignments in bold):

Activity	Estimated Time Required
Week 1 Video Lectures	2 hours
Week 1 Graded Quiz	1 hour

Goals and Objectives

After you actively engage in the learning experiences in this module, you should be able to:

- Explain some basic concepts in natural language processing.
- Explain different ways to represent text data.
- Explain the two basic types of word associations and how to mine paradigmatic relations from text data.

Guiding Questions

Develop your answers to the following guiding questions while watching the video lectures throughout the week.



- What does a computer have to do in order to understand a natural language sentence?
- What is ambiguity?
- Why is natural language processing (NLP) difficult for computers?
- What is bag-of-words representation?
- Why is this word-based representation more robust than representations derived from syntactic and semantic analysis of text?
- What is a paradigmatic relation?
- What is a syntagmatic relation?
- What is the general idea for discovering paradigmatic relations from text?
- What is the general idea for discovering syntagmatic relations from text?
- Why do we want to do Term Frequency Transformation when computing similarity of context?
- How does BM25 Term Frequency transformation work?
- Why do we want to do Inverse Document Frequency (IDF) weighting when computing similarity of context?

Additional Readings and Resources

The following readings are optional:

- C. Zhai and S. Massung, *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. ACM and Morgan & Claypool Publishers, 2016. Chapters 1-4, Chapter 13.
- Chris Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press. Cambridge, MA: May 1999. Chapter 5 on collocations.
- Chengxiang Zhai, *Exploiting context to identify lexical atoms: A statistical view of linguistic context*. Proceedings of the International and Interdisciplinary Conference on Modelling and Using Context (CONTEXT-97), Rio de Janeiro, Brazil, Feb. 4-6, 1997, pp. 119-129.
- Shan Jiang and ChengXiang Zhai, *Random walks on adjacency graphs for mining lexical relations from big text data*. Proceedings of IEEE BigData Conference 2014, pp. 549-554.

Key Phrases and Concepts



Keep your eyes open for the following key terms or phrases as you complete the readings and interact with the lectures. These topics will help you better understand the content in this module.

Coursera



- Part of speech tagging
- Syntactic analysis
- Semantic analysis
- Ambiguity
- Text representation, especially bag-of-words representation
- Context of a word; context similarity
- Paradigmatic relation
- Syntagmatic relation

Tips for Success

To do well this week, I recommend that you do the following:

- Review the video lectures a number of times to gain a solid understanding of the key questions and concepts introduced this week.
- When possible, provide tips and suggestions to your peers in this class. As a learning community, we can help each other learn and grow. One way of doing this is by helping to address the questions that your peers pose. By engaging with each other, we'll all learn better.
- It's always a good idea to refer to the video lectures and chapter readings we've read during this week and reference them in your responses. When appropriate, critique the information presented.
- Take notes while you read the materials and watch the lectures for this week. By taking notes, you are interacting with the material and will find that it is easier to remember and to understand. With your notes, you'll also find that it's easier to complete your assignments. So, go ahead, do yourself a favor; take some notes!

Getting and Giving Help

You can get/give help via the following means:

- Use the **Learner Help Center** to find information regarding specific technical problems. For example, technical problems would include error messages, difficulty submitting assignments, or problems with video playback. If you cannot find an answer in the documentation, you can also report your problem to the Coursera