

## Week 1 Overview

[Help Center](#)

# Week 1: NLP, Text Representation, and Word Association Mining

On this page:





[Instructional Activities](#)   [Time](#)   [Goals and Objectives](#)   [Key Phrases/Concepts](#)

[Guiding Questions](#)   [Readings and Resources](#)   [Video Lectures](#)   [Tips for Success](#)

[Getting and Giving Help](#)

## Instructional Activities

Below is a list of the activities and assignments available to you this week. See the [How to Pass the Class](#) page to know which assignments pertain to the badge or badges you are pursuing. Click on the name of each activity for more detailed instructions.

Relevant Badges	Activity	Due Date*	Estimated Time Required
	<a href="#">Week 1 Video Lectures</a>	Sunday, July 5	4 hours
 	<a href="#">Programming Assignments Overview</a>	Sunday, July 5	30 minutes
 	<a href="#">Week 1 Quiz</a>	Sunday, July 5	1 hour

\* All deadlines are at 11:55 PM Central Time ([time zone conversion](#)) unless otherwise noted.

## Time

This module will last **7 days** and should take **approximately 6 hours** of dedicated time to complete, with its lectures and assignments.

## 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 and why it is hard for computers to

precisely understand natural language text.

- Explain what are paradigmatic and syntagmatic relations between words
- Explain how statistical approaches can be used to mine paradigmatic and syntagmatic relations from text with no human effort
- Explain some basic concepts in information theory, including entropy, conditional entropy, and mutual information

## Key Phrases/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.

- Part of Speech tagging; syntactic analysis; semantic analysis; ambiguity
- Text representation, especially “Bag of words” representation
- Vector space model
- Context of a word; context similarity
- TF (Term Frequency) Transformation, BM25, IDF (Inverse Document Frequency)
- Entropy
- Conditional entropy
- Mutual information

## Guiding Questions






















Develop your answers to the following guiding questions while completing the readings and working on assignments 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?
- How does BM25 transformation work?
- Why do we want to do IDF (Inverse Document Frequency) weighting?
- What is entropy? For what kind of random variables does the entropy function reach its minimum and maximum, respectively?
- What is conditional entropy?
- What is the relation between conditional entropy  $H(X|Y)$  and entropy  $H(X)$ ? Which is larger?
- How can conditional entropy be used for discovering syntagmatic relations?
- What is mutual information  $I(X;Y)$ ? How is it related to entropy  $H(X)$  and conditional entropy  $H(X|Y)$ ?
- What's the minimum value of  $I(X;Y)$ ? Is it symmetric?
- For what kind of  $X$  and  $Y$ , does mutual information  $I(X;Y)$  reach its minimum? For a given  $X$ , for what  $Y$  does  $I(X;Y)$  reach its maximum?
- Why is mutual information sometimes more useful for discovering syntagmatic relations than conditional entropy?

# Readings & Resources

- 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.

## Video Lectures

Video Lecture	Lecture Notes	Transcript	Video Download	SRT Caption File	Forum
 <b>1.1 Overview Text Mining and Analytics: Part 1</b> (00:11:43)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.2 Overview Text Mining and Analytics: Part 2</b> (00:11:44)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.3 Natural Language Content Analysis: Part 1</b> (00:12:48)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.4 Natural Language Content Analysis: Part 2</b> (00:04:25)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.5 Text Representation: Part 1</b> (00:10:46)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.6 Text Representation: Part 2</b> (00:09:29)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.7 Word Association Mining and Analysis</b>		Forthcoming...	Forthcoming...	Forthcoming...	

(00:15:39)					
 <b>1.8</b> <b>Paradigmatic</b> <b>Relation</b> <b>Discovery Part</b> <b>1</b> (00:14:31)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.9</b> <b>Paradigmatic</b> <b>Relation</b> <b>Discovery Part</b> <b>2</b> (00:17:53)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.10</b> <b>Syntagmatic</b> <b>Relation</b> <b>Discovery:</b> <b>Entropy</b> (00:11:00)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.11</b> <b>Syntagmatic</b> <b>Relation</b> <b>Discovery:</b> <b>Conditional</b> <b>Entropy</b> (00:11:57)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.12</b> <b>Syntagmatic</b> <b>Relation</b> <b>Discovery:</b> <b>Mutual</b> <b>Information:</b> <b>Part 1</b> (00:13:55)		Forthcoming...	Forthcoming...	Forthcoming...	
 <b>1.13</b> <b>Syntagmatic</b> <b>Relation</b> <b>Discovery:</b> <b>Mutual</b> <b>Information:</b> <b>Part 2</b> (00:09:42)		Forthcoming...	Forthcoming...	Forthcoming...	

## 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. You can access the Help Center by clicking on the **Help** link at the top right of any course page. If you can not find an answer in the documentation, you can also report your problem to the Coursera staff by clicking on the **Contact Us!** link available on each topic's page within the Learner Help Center.
- Use the [Content Issues](#) forum to report errors in lecture video content, assignment questions and answers, assignment grading, text and links on course pages, or the content of other course materials. University of Illinois staff and Community TAs will monitor this forum and respond to issues.

As a reminder, the instructor is not able to answer emails sent directly to his account. Rather, all questions should be reported as described above.

---

Created Tue 11 Dec 2012 10:59 AM PST

Last Modified Mon 8 Jun 2015 7:00 AM PDT