Applied Text Mining in Python

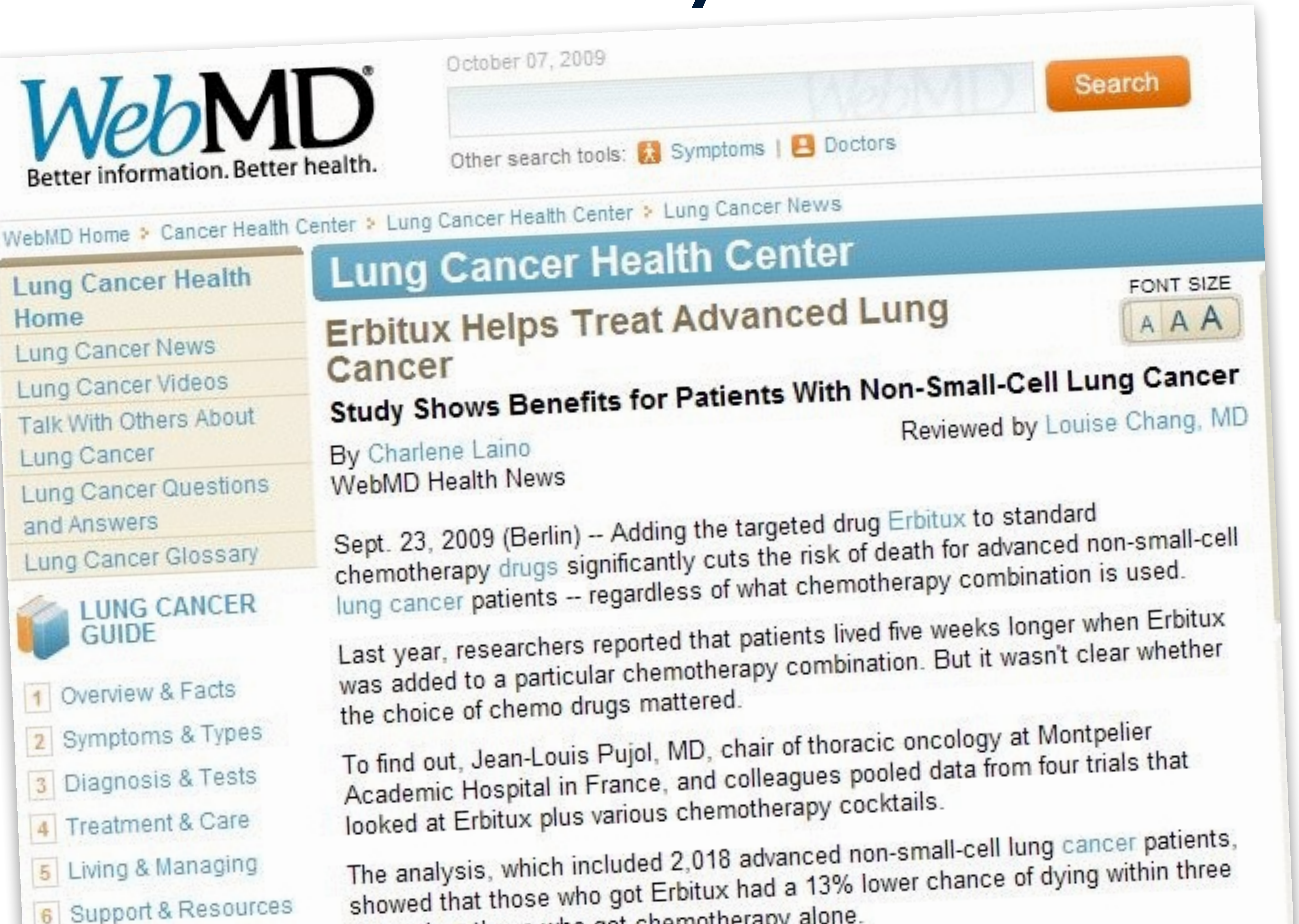
Information Extraction

Information is hidden in free-text

- Most traditional transactional information is structured
- Abundance of unstructured, freeform text
- How to convert unstructured text to structured form?

Information Extraction

- Goal: Identify and extract fields of interest from free text



October 07, 2009

WebMD
Better information. Better health.
Search
Other search tools: Symptoms | Doctors
FONT SIZE
A A A

WebMD Home > Cancer Health Center > Lung Cancer Health Center > Lung Cancer News

Lung Cancer Health Home
Lung Cancer News
Lung Cancer Videos
Talk With Others About Lung Cancer
Lung Cancer Questions and Answers
Lung Cancer Glossary

LUNG CANCER GUIDE

1. Overview & Facts
2. Symptoms & Types
3. Diagnosis & Tests
4. Treatment & Care
5. Living & Managing
6. Support & Resources

Lung Cancer Health Center

Erbitux Helps Treat Advanced Lung Cancer

Study Shows Benefits for Patients With Non-Small-Cell Lung Cancer

By Charlene Laino
WebMD Health News

Reviewed by Louise Chang, MD

Sept. 23, 2009 (Berlin) -- Adding the targeted drug Erbitux to standard chemotherapy drugs significantly cuts the risk of death for advanced non-small-cell lung cancer patients -- regardless of what chemotherapy combination is used.

Last year, researchers reported that patients lived five weeks longer when Erbitux was added to a particular chemotherapy combination. But it wasn't clear whether the choice of chemo drugs mattered.

To find out, Jean-Louis Pujol, MD, chair of thoracic oncology at Montpelier Academic Hospital in France, and colleagues pooled data from four trials that looked at Erbitux plus various chemotherapy cocktails.

The analysis, which included 2,018 advanced non-small-cell lung cancer patients, showed that those who got Erbitux had a 13% lower chance of dying within three chemotherapy alone.

Erbitux helps treat lung cancer

Author: Charlene Laino

Reviewer: Louise Chang, MD

Sept. 23, 2009

Berlin

…

Fields of Interest

- Named entities
  - [NEWS] People, Places, Dates, …
  - [FINANCE] Money, Companies, …
  - [MEDICINE] Diseases, Drugs, Procedures, …
- Relations
  - What happened to who, when, where, …

Named Entity Recognition

- Named entities: Noun phrases that are of specific type and refer to specific individuals, places, organizations, …
- Named Entity Recognition: Technique(s) to identify all mentions of pre-defined named entities in text
  - Identify the mention / phrase: *Boundary detection*
  - Identify the type: *Tagging / classification*

Examples of Named Entity Recognition Tasks

The patient is a 63-year-old female with a three-year history of bilateral hand numbness and occasional weakness.

Within the past year, these symptoms have progressively gotten worse, to encompass also her feet.

She had a workup by her neurologist and an MRI revealed a C5-6 disc herniation with cord compression and a T2 signal change at that level.

Approaches to identify named entities

- Depends on kinds of entities that need to be identified

- For well-formatted fields like date, phone numbers: Regular expressions (Recall Week 1)
- For other fields: Typically a machine learning approach

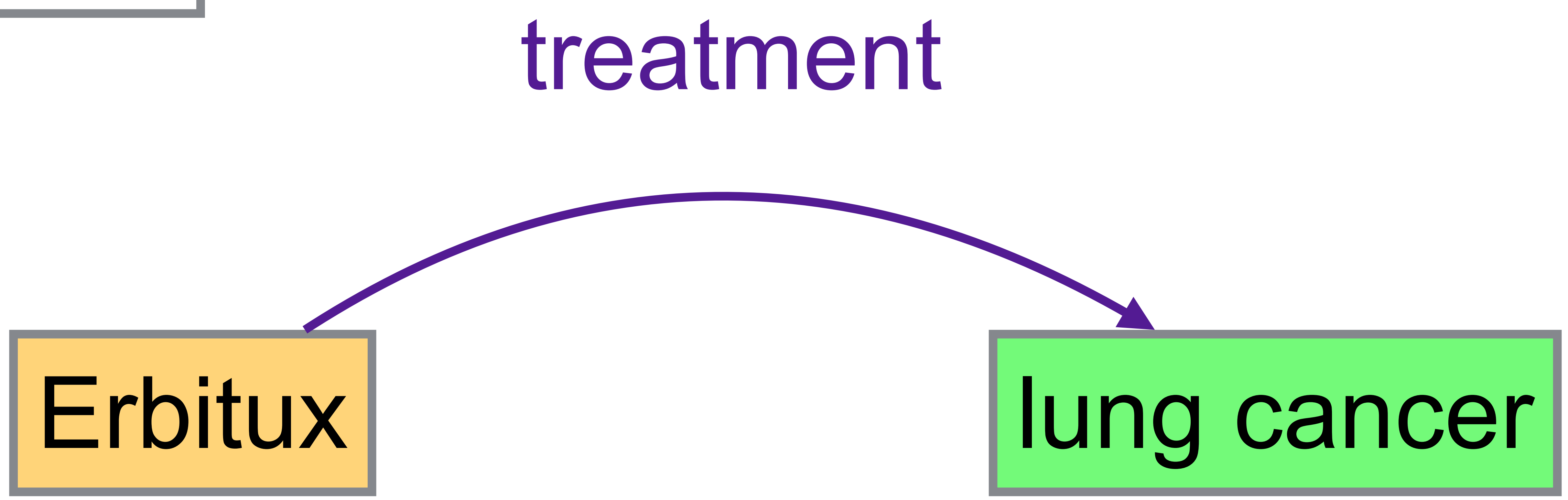
Person, Organization, Location/GPE

- Standard NER task in NLP research community
- Typically a four-class model
  - PER
  - ORG
  - LOC / GPE
  - Other / Outside (any other class)

Relation Extraction

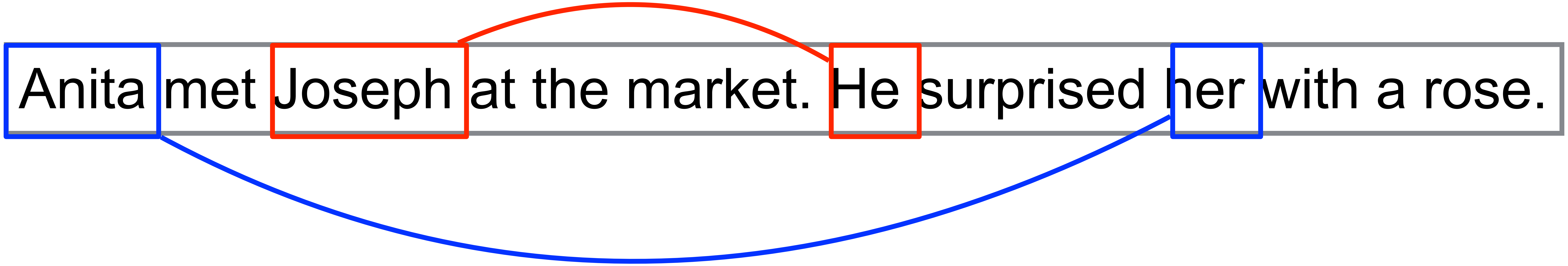
- **Identify relationships between named entities**

Erbitux helps treat lung cancer



Co-reference Resolution

- **Disambiguate mentions and group mentions together**



Anita met Joseph at the market. He surprised her with a rose.

Question Answering

- Given a question, find the most appropriate answer from the text
  - What does Erbitux treat?
  - Who gave Anita the rose?
- Builds on named entity recognition, relation extraction, and co-reference resolution

Take Home Concepts

- Information Extraction is important for natural language understanding and making sense of textual data
- Named Entity Recognition is a key building block to address many advanced NLP tasks
- Named Entity Recognition systems extensively deploy supervised machine learning and text mining techniques discussed in this course