Advanced Natural Language Processing and Information Retrieval

Course Description

Alessandro Moschitti
Department of Computer Science and Information Engineering
University of Trento
Email: moschitti@disi.unitn.it
Teachers

- Prof. Alessandro Moschitti, PhD
- Dr. Olga Uryupina, PhD
- Dr. Antonio Uva
- Dr. Massimo Nicosia
- Dr. Daniele Bonadiman
- Dr. Kateryna Tymoshenko, PhD
Teachers

- Dr. Gianni Barlacchi
- Dr. Lingzhen Chen (Liah)
- Dr. Irina Haponchychk
Student List

- https://goo.gl/0nWz5V
Course Schedule

- Lectures
  - Thursday, 11:30 – 13:30 (Theory)
    - Room A213
  - 16:30 – 18:30 (Laboratory)
    - Room PC A201
  - In the first month some theory lectures in lab

- Consulting hours:
  - Sending email is recommended
Syllabus

Introduction to Information Retrieval (IR)

Statistical Machine Learning:
- Kernel Methods, Classification, Clustering, Ranking, Re-Ranking and Regression and hints to practical machine learning, Neural Networks: CNNs, LSTM
Syllabus

Performance Evaluation:
- Performance Measures, Performance Estimation, Cross validation, Held Out and n-Fold Cross validation

Statistical Natural Language Processing:
- Syntactic Parsing: shallow and deep Constituency Parsing, Dependency Syntactic Parsing.
- Social Media: sentiment analysis and event extraction from Twitter
Syllabus

- Statistical Natural Language Processing:
  - Discourse Parsing: Coreference Resolution and discourse connective classification
Syllabus

Joint NLP and IR applications:

- Deep Linguistic Analysis for Question Answering: QA tasks (open, restricted, factoid, non-factoid), NLP Representation, Question Answering Workflow, QA Pipeline, Question Classification and QA reranking.

- Fine-Grained Opinion Mining: automatic review classification, deep opinion analysis, automatic product extraction and review, reputation/social media analysis
Lab 1

- Search Engines
- Kernel Methods and SVMs
  - Automated Text Categorization
  - Question Classification
  - Answer Reranking
- Syntactic Parsing and Named Entity Recognition
- Sentiment Analysis
  - Neural Networks
Lab 2

- Our UIMA pipeline implementing a pseudo Watson (4-5 lectures)
  - all NLP processors seen before
  - Question Answering full pipeline
  - Community Question Answering full pipeline
PART I: Essential Notions of Information Retrieval and Machine Learning

- Feb 22: Alessandro
  - Introduction to the course and IR, performance measures, machine learning, text categorization

- Mar 1: Alessandro (live video lecture)
  - Perceptron, SVMs (theory)
  - Kernel Methods, Question Classification (theory)
  - Practical examples on the above
PART I: Essential Notions of Information Retrieval and Machine Learning

- Mar 8
  - Alessandro (live video lecture): Classification, Multi-classification, Ranking, Regression and Structured Output Models (theory)
  - Irina: Ranking, Multi-classification, Regression, Structured Perceptron (Lab)
PART II: Basics of Natural Language Processing

- Mar 15:
  - Olga: sequence labeling: POS-tagging and Named Entity Recognition
  - Antonio: sequence labeling: POS-tagging and Named Entity Recognition (Lab)

- Mar 22
  - Olga Coreference Resolution (theory)
  - Mar 22: Irina: Coreference Resolution (Lab)

- Mar 29: Easter
PART II: Basics of Natural Language Processing

Apr 5
- Olga: Syntactic Parsing (theory)
- Antonio: Syntactic Parsing (Lab)

Apr 12
- Gianni: Pandas for text data analysis

Apr 19:
- Kateryna: Question Answering with a UIMA pipeline
- Antonio: Community Question Answering with a UIMA Pipeline
PART III: Neural Networks for NLP and IR

- Apr 26: Alessandro
  - Introduction to Neural Networks (theory)
  - Neural Models for NLP (theory)

- May 3: Daniele
  - Neural Networks models and implementations:
    - Pytorch development environment: examples on Sentiment Analysis

- May 10: Daniele
  - Neural networks for Question Answering
  - Convolutional Networks, Long Short Term Memory
PART III: Neural Networks for NLP and IR

- May 10: Daniele
  - Neural networks for Question Answering
  - Convolutional Networks, Long Short Term Memory

- May 17:
  - Liah: Networks for NER and sequence to sequence models.
  - Massimo: Neural Networks for end-to-end systems
Where to study?

- Course Slides at
  http://disi.unitn.it/moschitti/teaching.html
  - ANLP-IR section
    (you can watch the old NLP-IR section)

- Book - IR:
Where to study?

- **Book – NLP:**
  - SPEECH and LANGUAGE PROCESSING. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Second Edition by Daniel Jurafsky and James H. Martin
Where to study?

- Course Slides at
  http://disi.unitn.it/moschitti/teaching.html

- NLP-IR section:
  - Slides of IIR available at: http://informationretrieval.org
Natural Language Processing and Information Retrieval

Information Retrieval Lectures

- Motivation and presentation of the course + inverted index
- In depth on tokenization, normalization and optimization (optional ppt)
- Preprocessing, data structures, n-grams and wildcards
- Vector Space Model and weighting schemes
- Efficient methods for document retrieval
- Performance Measures and Query Expansion

The above presentations are heavily if not totally based on the IR courses of my friends Chris and Hinrich, who with Prabhakar Raghavan have built an excellent didactic tool. I would like to express my sincere thanks and appreciation for their nice work: their ppts are available at:


whereas the book (also adopted in my course) is available at:
