Natural Language Processing: Course Introduction

AMIT DHOMNE

Natural Language Processing

Sub-field of CS concerned with the development of systems that allow computers to interact with human language.

Also known as Computational Linguistics.

Why work on NLP?

- Automatically manage and summarize text
- Natural language computer interaction
- Machine Translation
- Model and analyze properties of language

Why is NLP hard?

- Languages are complex
- Languages are ambiguous
- Understanding requires vast knowledge
- Human input is scarce

Solutions

- Incorporate linguistic knowledge
- Learn from human input, when available
- Automatically learn structure

NLP Tasks

Summarization

Automatically summarize a text

- How to identify people, places, etc.?
- How to identify relations between entities?
- How to recognize events?

Human-Computer Interaction

Allow for human interaction with computers in natural language

- How to model meaning?
- How to infer missing information?
- How to incorporate world knowledge?

Machine Translation

Automatically translate from one human language to another

- How do the words map?
- How does the grammar map?
- Does the output language read fluently?

Course Topics

Working with Corpora

Use a body of text to accomplish NLP tasks

Classification

Classify texts into discrete categories

Classification

Hi, My name is Ivy Irwin. I used to be extremely fat not so long ago and my friends would often mock at me because of this.

It all changed when I found the medication that helped me. I lost 8 kilos of excessive weight in just one month. I didn't do any exercises or go on a diet. All I did was living my life and losing the weight.

It is nothing but the truth, believe me! It is easy to see, here you can find it with a 60% discount.

Spam

Not Spam

Language Modeling

Build a probabilistic model of what
a language looks like

Language Modeling

The University of Texas at _____

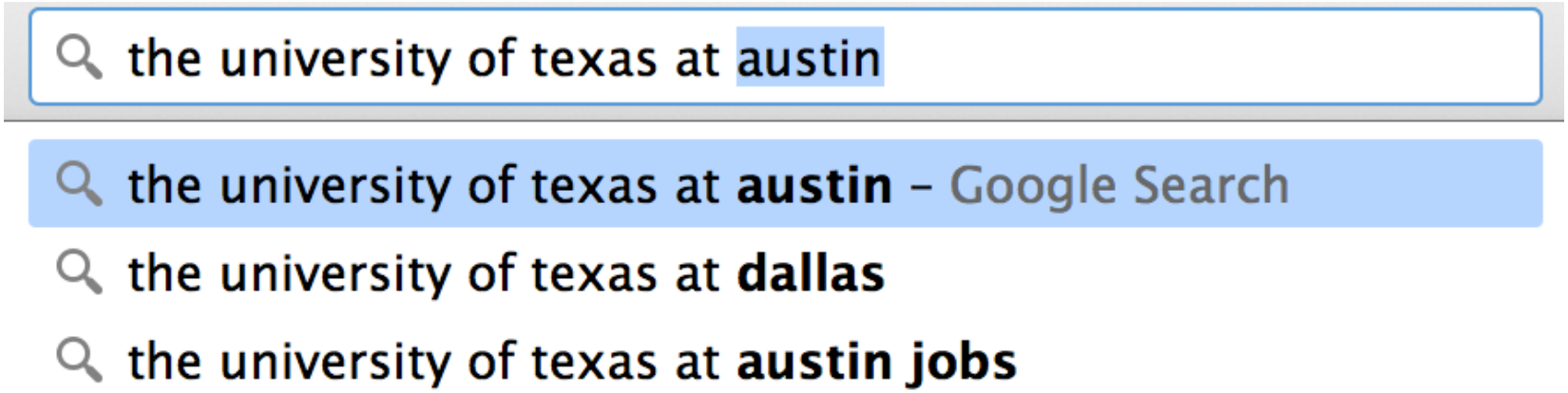
Austin?

Dallas?

Los Angeles?

Giraffe?

Language Modeling



Language Modeling

Can you go be the store on your way home?

Language Modeling

une histoire intéressante 

a story interesting?

a interesting story?

an interesting story?

Syntactic Processing

Tag words with parts of speech

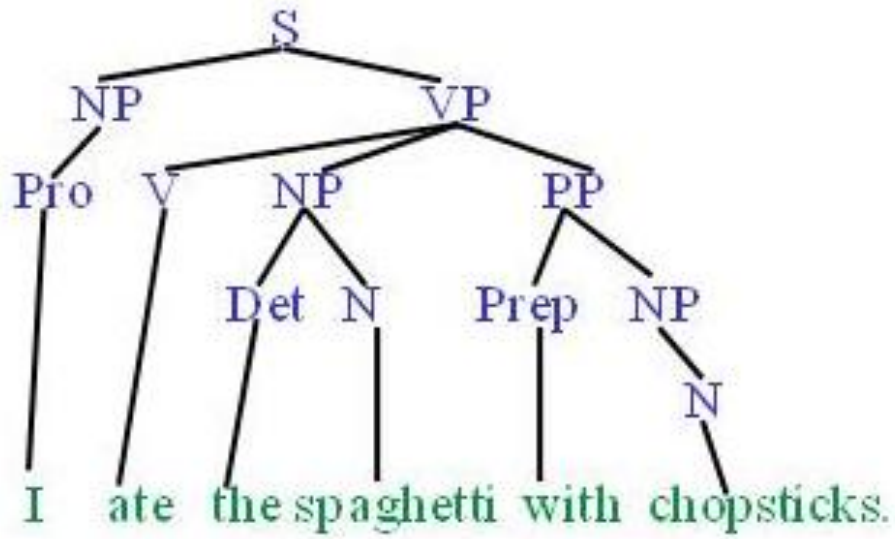
Syntactic Processing

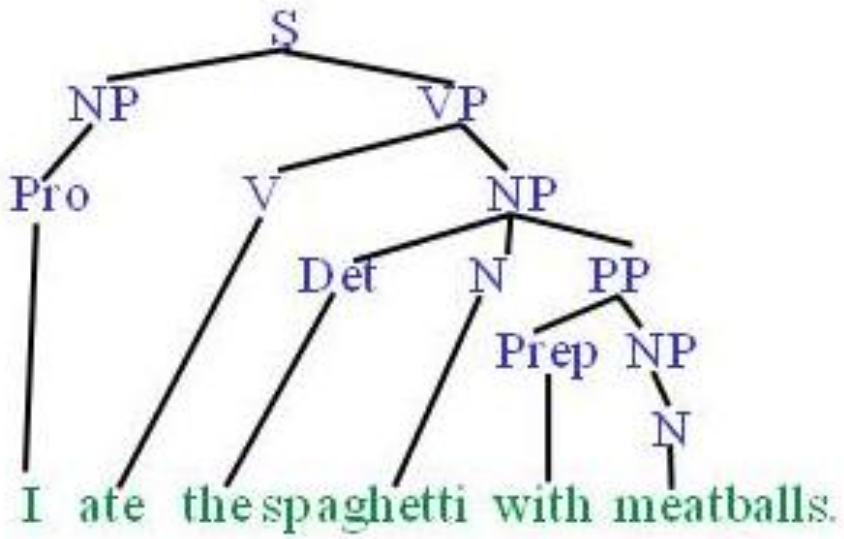
I	saw	the	man	with	the	saw	.
PRP	**VBD**	**DT**	**NN**	**IN**	**DT**	**NN**	**.**

Syntactic Processing

Parse a text

Syntactic Processing





Semantic Processing

Word sense disambiguation

Semantic Processing

I drove by the bank this morning.

The bank forclosed on his home.

The restaurant is on the river bank.

Semantic Processing

Identify entities and their relationships

Semantic Processing

John brought his son to UT, where he has worked for many years.

Entity	Type
John	Person
his	Person
son	Person
UT	Organization
he	Person

Semantic Processing

John brought his son to UT, where he has worked for many years.

Subject	Object	Relation
John	son	Father
son	John	Child
John	UT	Employee
UT	John	Employer

Semantic Processing

Model the meaning of a text

Semantic Processing

John brought his son to UT, where he worked.

$John(x) \& bring(e_1) \& agent(e_1,x) \&$
$son(y) \& patient(e_1,y) \&$
$UT(z) \& to(e_1,z) \&$
$work(e_2) \& agent(e_2,x)$

Course Expectations

Course Website

http://utcompling.github.io/nlpclass-fall2013/

(linked from my website)

TA

Lewis

Course Mailing List

on **Piazza**

(linked from the course website)

Textbook

Speech and Language Processing, 2nd Edition

Dan Jurafsky and Jim Martin

Graded Work

2 exams

7 assignments

Programming

You must be comfortable writing code.

All programming will be done in **Scala**.

Programming

Information about Scala, including instructions for getting started, can be found on the course website.

Programming

A preliminary assignment (#0) will help you:

a) practice Scala
b) determine whether the programming expectations will be a problem for you
c) set things up for future assignments

Assignments

Written homework is due when the **lecture starts**.

Programming homework is due **two hours before class**.

Assignments

Programming takes time. Start early.