

Text Mining Project/Lab

Behrang QasemiZadeh
behrangatoffice@gmail.com

Objectives of Today's session

- To give us an understanding of:
 - Who you are and what are your expectations?
 - Who is teaching the course and what are his expectations?

Objectives of Today's session

- To give us an understanding of:
 - Who you are and what are your expectations?
 - Who is teaching the course and what are his expectations?

About the instructor

<http://atmykitchen.info>

Expectation

- The major objective is:
 - To Enable you to translate theoretical lessons into cool practical applications:

Expectation

- The major objective is:
 - To Enable you to translate theoretical lessons into cool practical applications:
 - Cool in this context means:
 - To provide you with an exciting place where you can investigate, analyse and apply theories and make abstract concepts concrete.

Expectation

- The major objective is:
 - To Enable you to translate theoretical lessons into cool practical applications:
 - Cool in this context means:
 - To provide you with an exciting place where you can **investigate, analyse and apply** theories and make abstract concepts concrete.

Your role is to learn and reflect (this may not a smooth path).

My role is to make sure this is going to be an smooth procedure.

Expectation

- **Your role is to learn and reflect:**
 - **That is, you need to have a growth mind-set;**
 - **That is, you need to be self-motivated;**
 - **The natural product of this process is a project.**

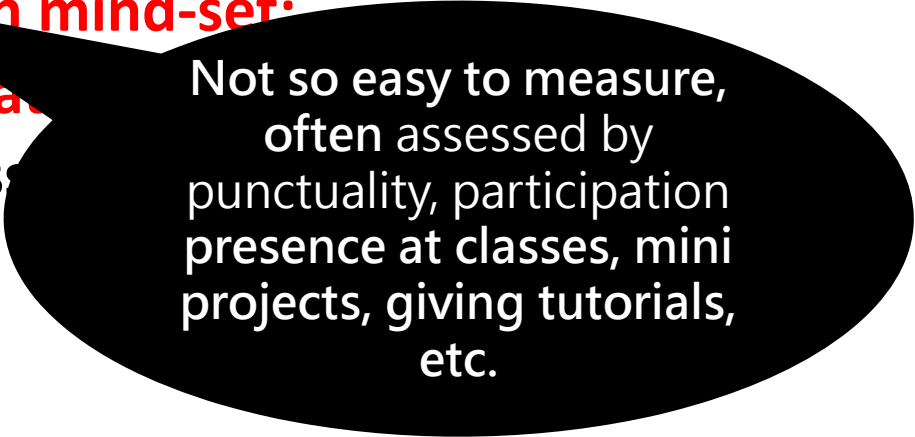
Assessment

- **Your role is to learn and reflect:**
 - **That is, you need to have a growth mind-set;**
 - **That is, you need to be self-motivated;**
 - **The natural product of this process is a project.**

Assessment

- **Your role is to learn and reflect:**

- **That is, you need to have a growth mind-set:**
- **That is, you need to be self-motivated:**
- **The natural product of this process is:**



Not so easy to measure,
often assessed by
punctuality, participation
presence at classes, mini
projects, giving tutorials,
etc.

Assessment

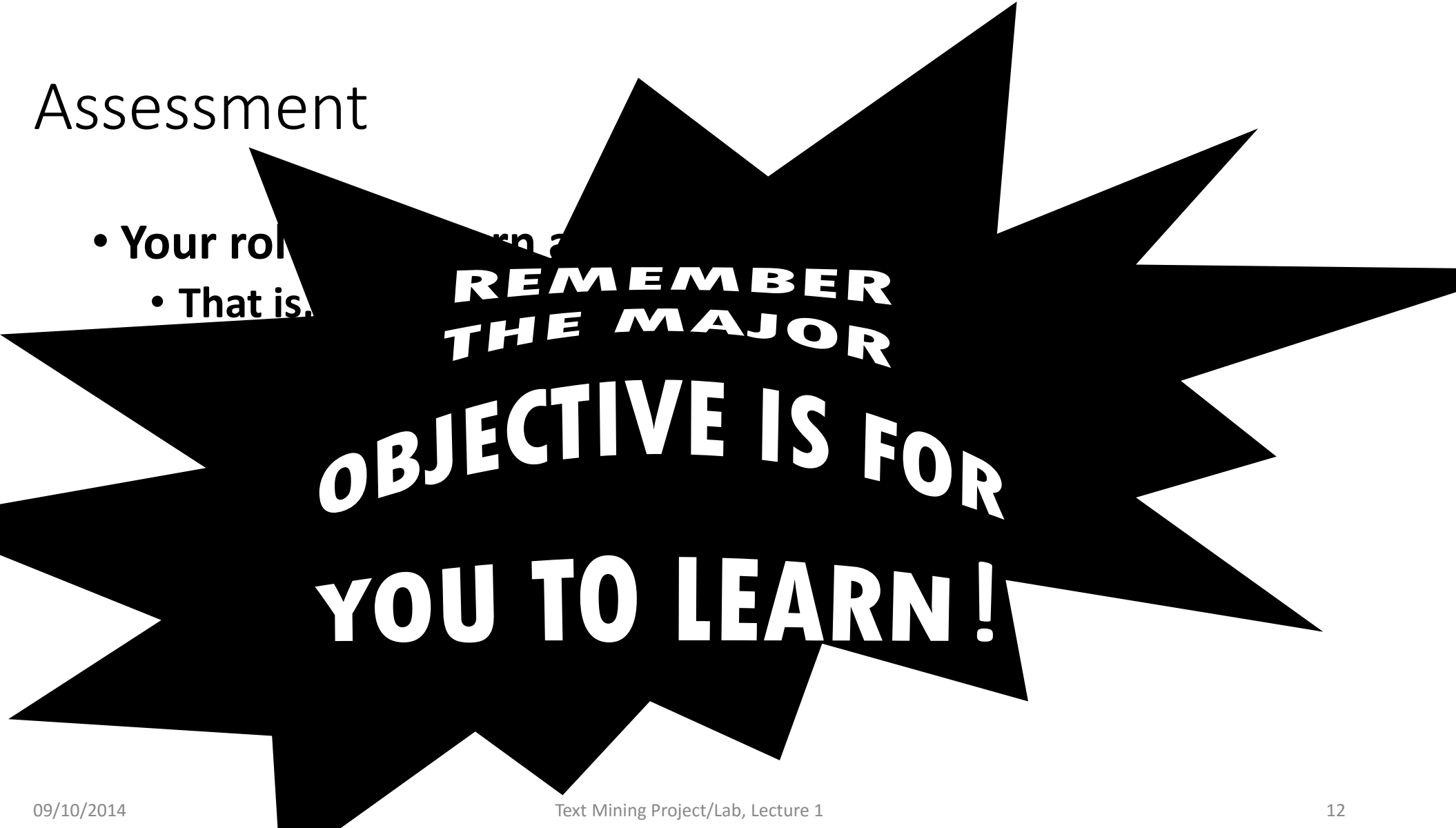
- **Your role is to learn and reflect:**
 - That is, you need to have a growth mind-set;
 - That is, you need to be self-motivated;
 - The natural product of this process is a **project**.



Relatively easy
to assess

Assessment

- Your role in the assessment process
- That is, the assessment process



**REMEMBER
THE MAJOR
OBJECTIVE IS FOR
YOU TO LEARN!**

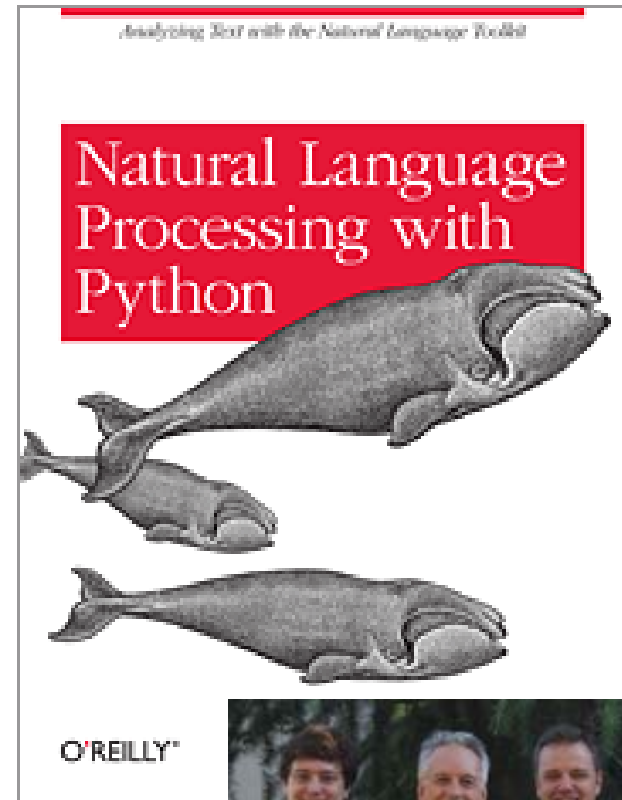
Standard Tools to Achieve Our Goal

- Natural Language Toolkit of Python

WWW.NLTK.ORG

***If you can, please buy the
book for a good cause!***

***Otherwise get a copy from library and READ IT!
No one can do that on your be half!***



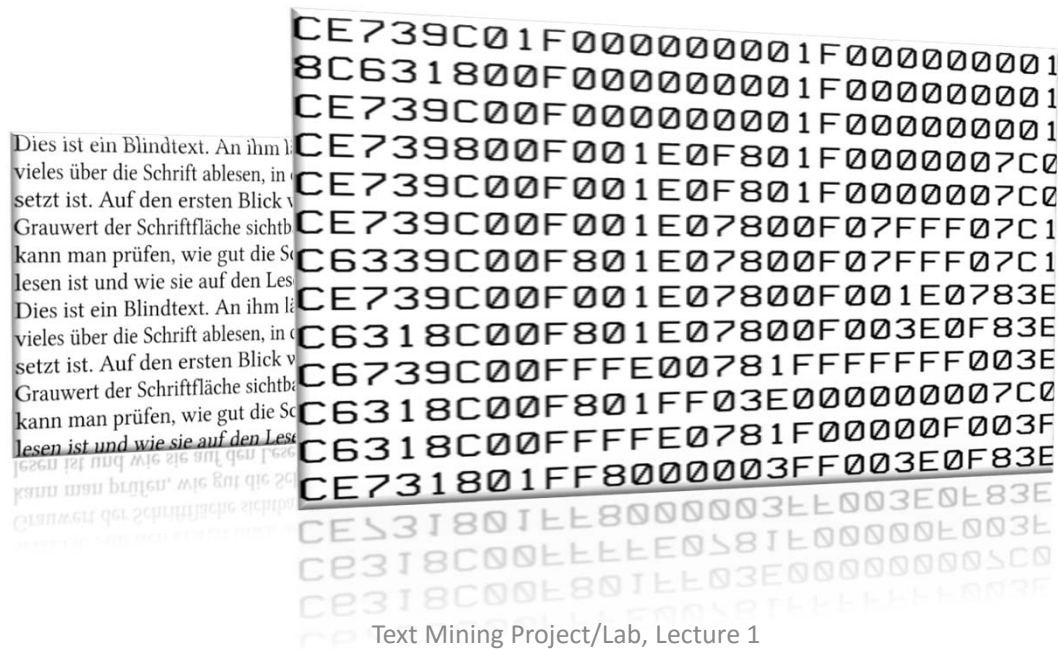
Intended Plan and Structure

- To enable you to choose or define a project as soon as possible.
 - That is, to enable you to go after your own text analysis idea independently.
 - This will be achieved through 7 weeks of intensive reading and coding.

An overview to Techniques Covered

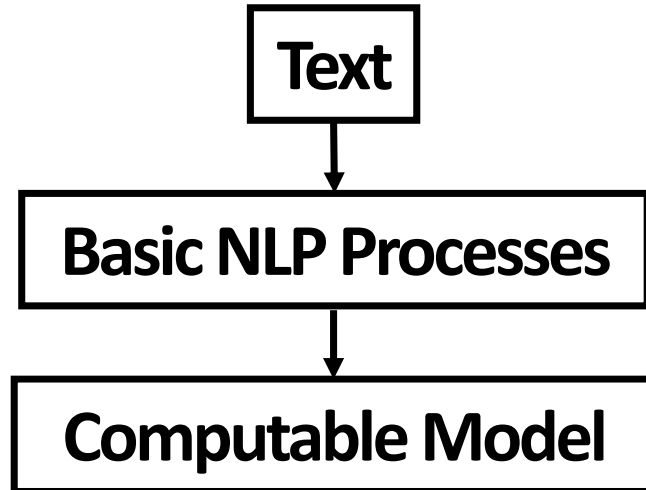
Natural Language Text Processing

- For Computers, natural language text is simply a sequence of bits and bytes.

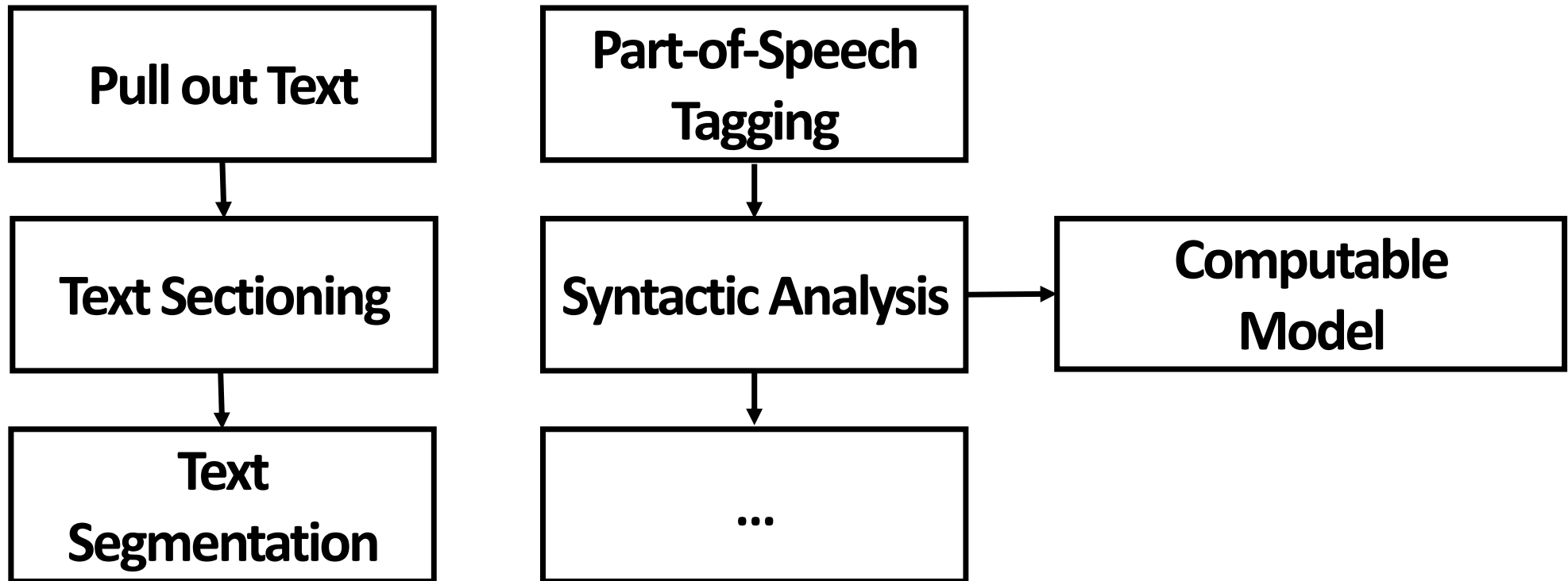


Natural Language Text Processing

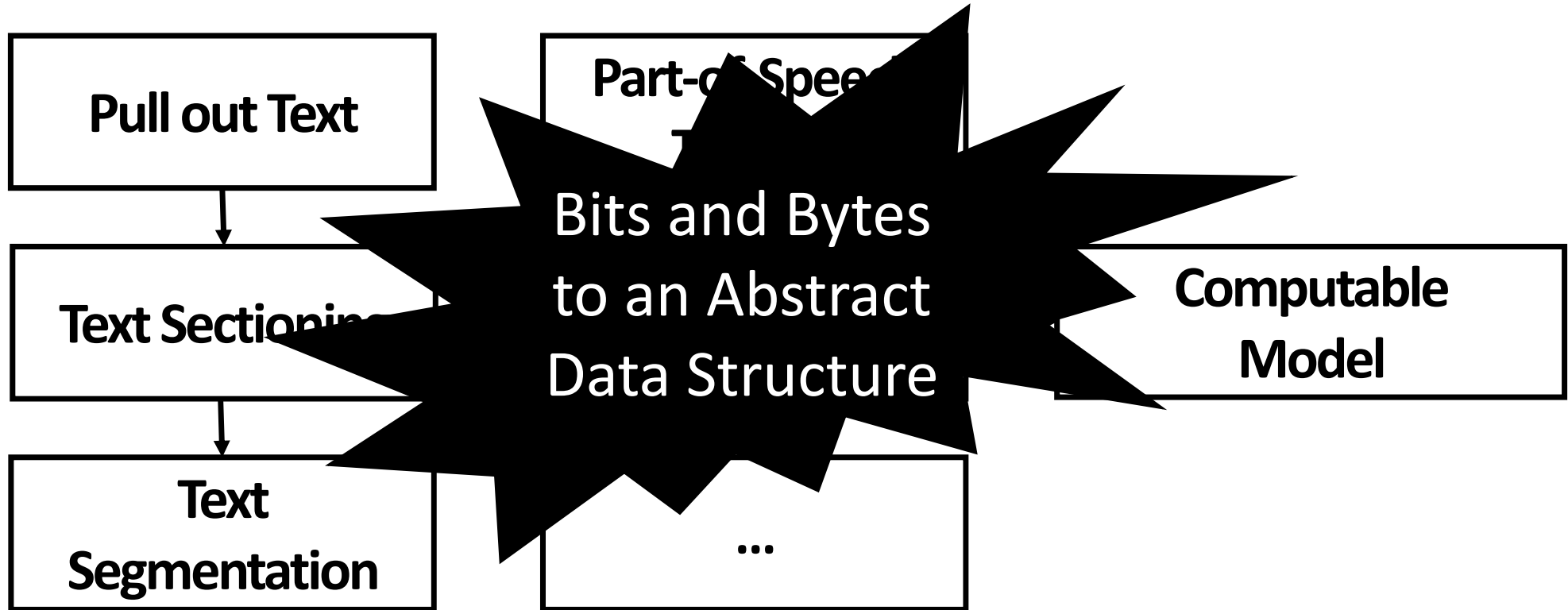
- For Computers, natural language text is simply a sequence of bits and bytes.



Natural Language Text Processing



Natural Language Text Processing



Intended Plan

Intended Plan and Structure

1. Accessing Text Corpora and Lexical Resources
2. Processing Raw Text
3. Categorizing and Tagging Words
4. Syntactic Analysis and Parsing
5. Learning to Classify Text (Application Example 1)
6. Extracting Information from Text (Application Example 2)

Expectation

- The major objective is:
 - To Enable you to translate theoretical lessons into cool practical applications:
 - Cool in this context means:
 - To provide you with an exciting place where you can **investigate, analyse and apply** theories and make abstract concepts concrete.

Your role is to learn and reflect (this may not be run smoothly).

My role is to make sure this is going to be an smooth procedure.

Expectation

- **Availability**

I am available Wednesday and Thursday for Teaching and meeting (requires booking).
You can also email me and I get back to you as soon as possible.

- **Flexibility**

Lab and project sessions goes hand in hand with the course.
The content/pace will be adopted inasmuch as possible.

PLEASE TELL ME HOW CAN I EXTEND THIS LIST

- Your role is to learn, analyse and reflect (this may not be run smoothly).
- My role is to make sure this is going to be an smooth procedure.

**By the way, DO YOU
KNOW PYTHON?**

If NO, then no worries:

NLTK book is designed to teach you NLP and Python in one go!

And, Python is EASY!

a short introduction
to Python

What is Python?

- Python is a high-level programming language.
- The emphasis is on:
 - Readability;
 - Expressivity of the code.
- Python can be viewed as either an *object-oriented* or a *procedural programming* language.

What is Python?

- Python employs a dynamic type system.
- Python programmers implement managed codes:
 - That is, Python takes care after memory management automatically.

Python Data Types

- We can identify four data types in Python:
 - Variables
 - Lists
 - Sets
 - Dictionary
- See Python's documentations!

Built-in variable types

- Basic *VARIABLE* Types

- Integer;
- Long integer;
- Float;
- String.

- Boolean (truth value):

- Values NONE, FALSE, 0.0, 0, empty set, dictionary, etc. is considered false (see Python's documentations).

STRING

- In python strings are seen as lists of characters.
- Basic String operations are:
 - Concatenation;
 - Comparison;
 - Indexing;
 - Extracting substrings;
 - Finding substrings;
 - Splitting;
 - Stripping;
 - Changing case.

Built-in Functions in Python

- Built-in functions for basic operations
 - functions for data structure manipulation;
 - arithmetic operations;
 - input/output.
- Python Standard Modules (built-in functions) are loaded automatically at the beginning of each session:

```
>>> dir(__builtins__)
```

- Content of this directory is automatically loaded when you start!

Extra Resources to Learn Python

- First lesson: mastering a programming language comes with practice.
- Look at the very nice teaching project by:
- I myself learned Python from **Damir Cavar**:
 - His slides are available online;
 - Check .

IT IS YOUR TURN NOW!

- ANY QUESTION BEFORE THAT?
- Please tell us
 - Your name;
 - Year of study;
 - Languages that you speak (natural and programming);
 - And, a hobby of yours.

Assignment

Write a Python program that:

1. Reads a large text file (i.e., a ***corpus***), e.g. a book or any kind of text; ***tokenize*** it using Python String built-in functions;
2. Make a ***dictionary*** of tokens and their frequencies;
3. And, write the dictionary into another text file (each line of the output file is a token followed by the frequency of that token in tab separated format).

