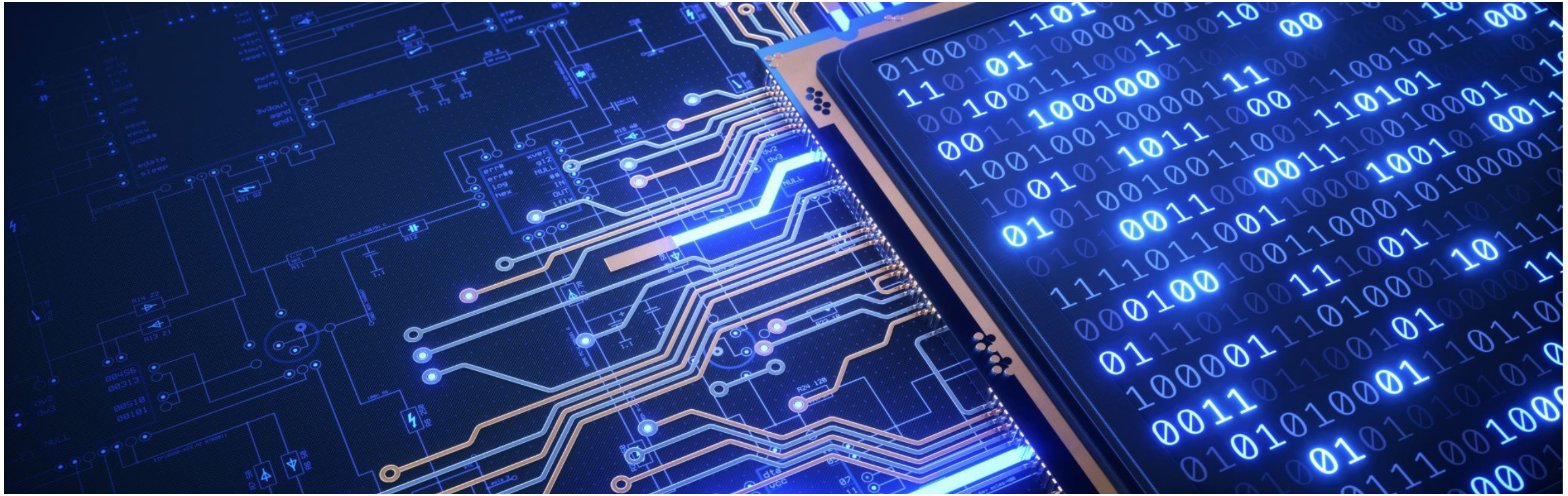


NATURAL LANGUAGE PROCESSING
AIMS 2022-2023

INTRODUCTION

Elvis Ndah, PhD

Contact

- elvis.ndah@gmail.com

Consultant – Machine Learning Engineer

- Anju Software
- European Commission

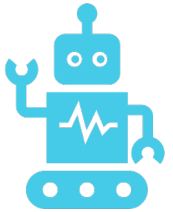
PREREQUISITES

Programming language: Python

Recommended: intermediary

Acceptable: Basic

Probability and statistics



Machine learning (deep learning)

THE FOCUS OF THIS COURSE

01

Introduction to human language understanding.

02

Why natural language processing is difficult?

03

Understanding of the modern techniques for NLP

04

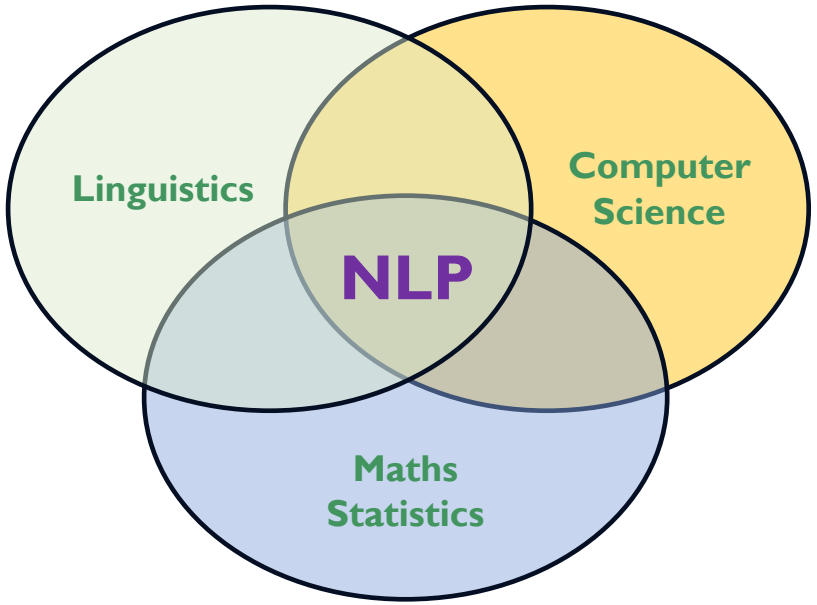
Learn to build systems for major problems in NLP

LECTURE 1: INTRODUCTION TO NLP

- Course Overview
  - What is NLP? Why it is important?
  - Applications of NLP
  - What are the challenges?
  - What types of ML methods used in NLP?

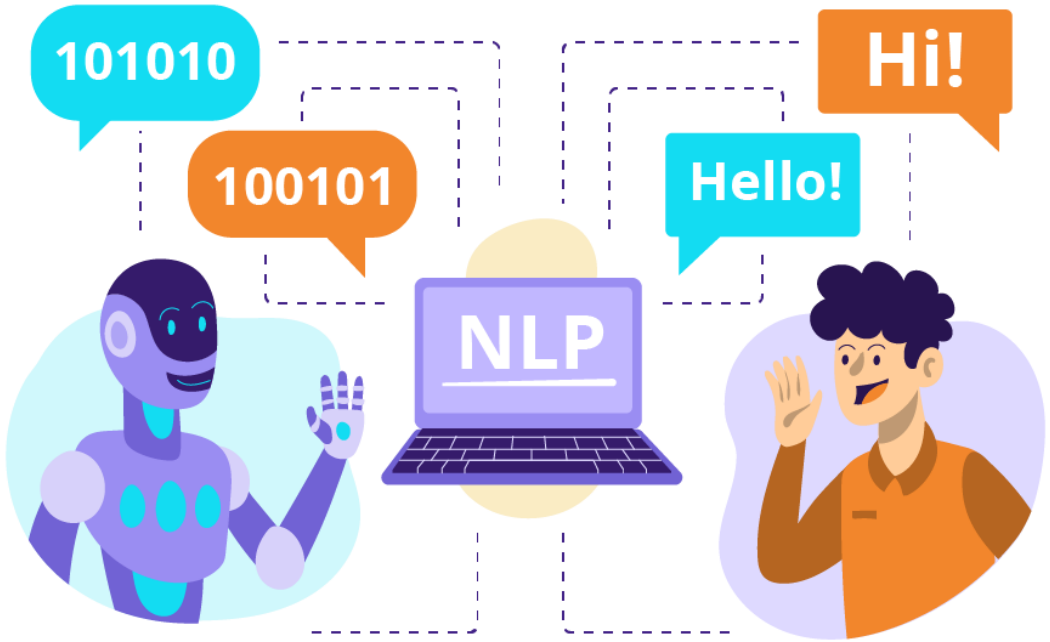
WHAT IS NLP?

Wiki: Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages.



WHAT IS NLP?

It concerns with the interaction between computing devices and human natural languages.



WHAT IS NLP?

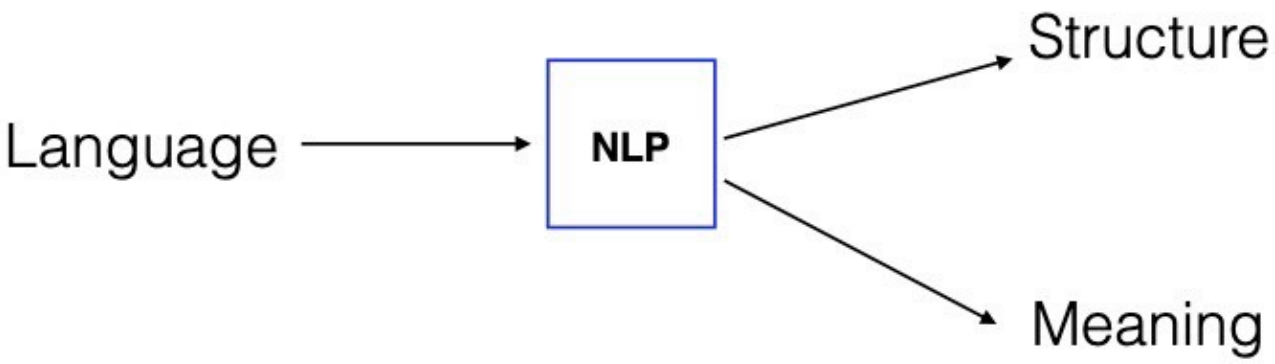
It concerns with the interaction between computing devices and human natural languages.

https://slideplayer.com/slide/17527407/

WHAT IS NLP?

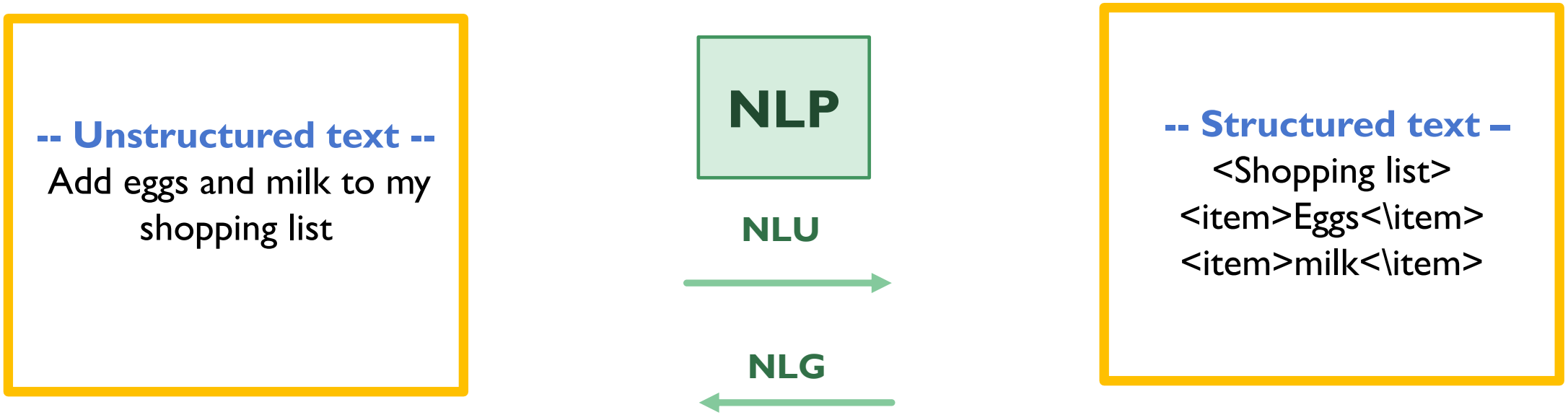
For NLP systems to be be useful they need to:

- Identify and understand structure and meaning of words, sentences, texts and conversations.
- Deep understanding of broad language.



WHAT IS NLP?

- We can break down NLP into two main classes



WHY WORK ON NLP?

Build systems that help humans communicate

Help humans interact with each other and/or devices.

Useful systems

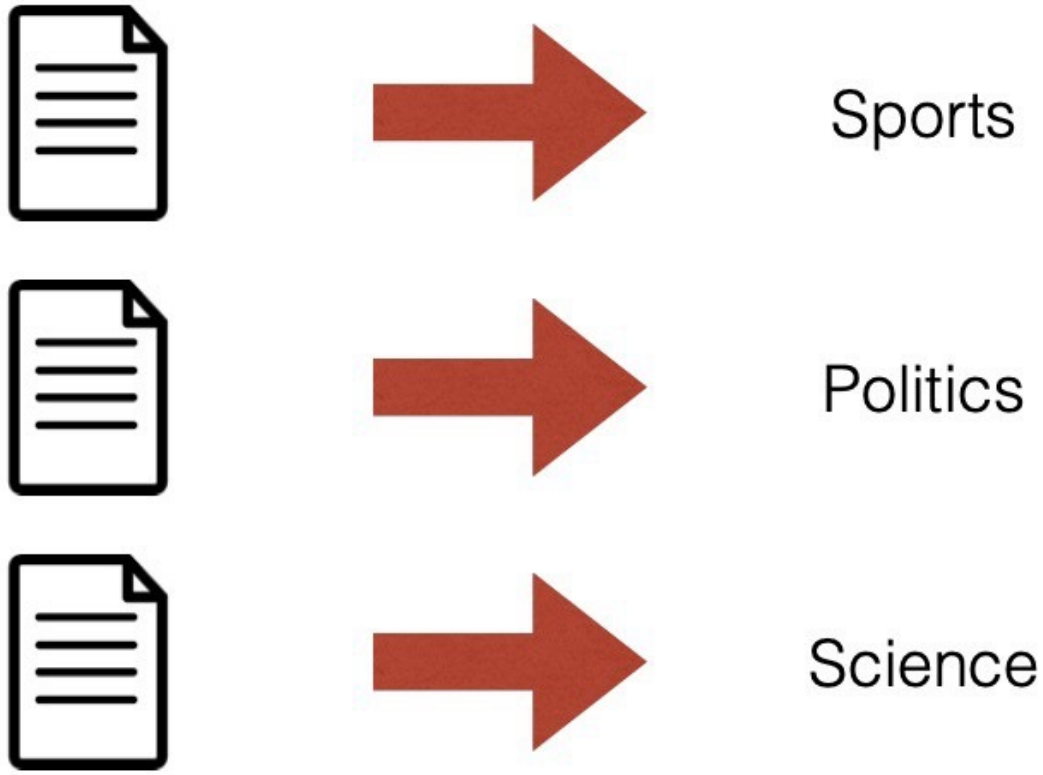
Text classification
Automatic text summarization
Communicate without language barrier
Model and analyse properties of language
Speech recognition

NLP PROGRESS

1950s: Rule based systems → **Late 1980s:** Statistical techniques → **Early 2010s:** Neural networks – deep learning

Rule-based and statistical methods ares still relevant today.

NLP APPLICATIONS – TEXT OR DOCUMENT CATEGORIZATION



NLP APPLICATIONS – INFORMATION EXTRACTION

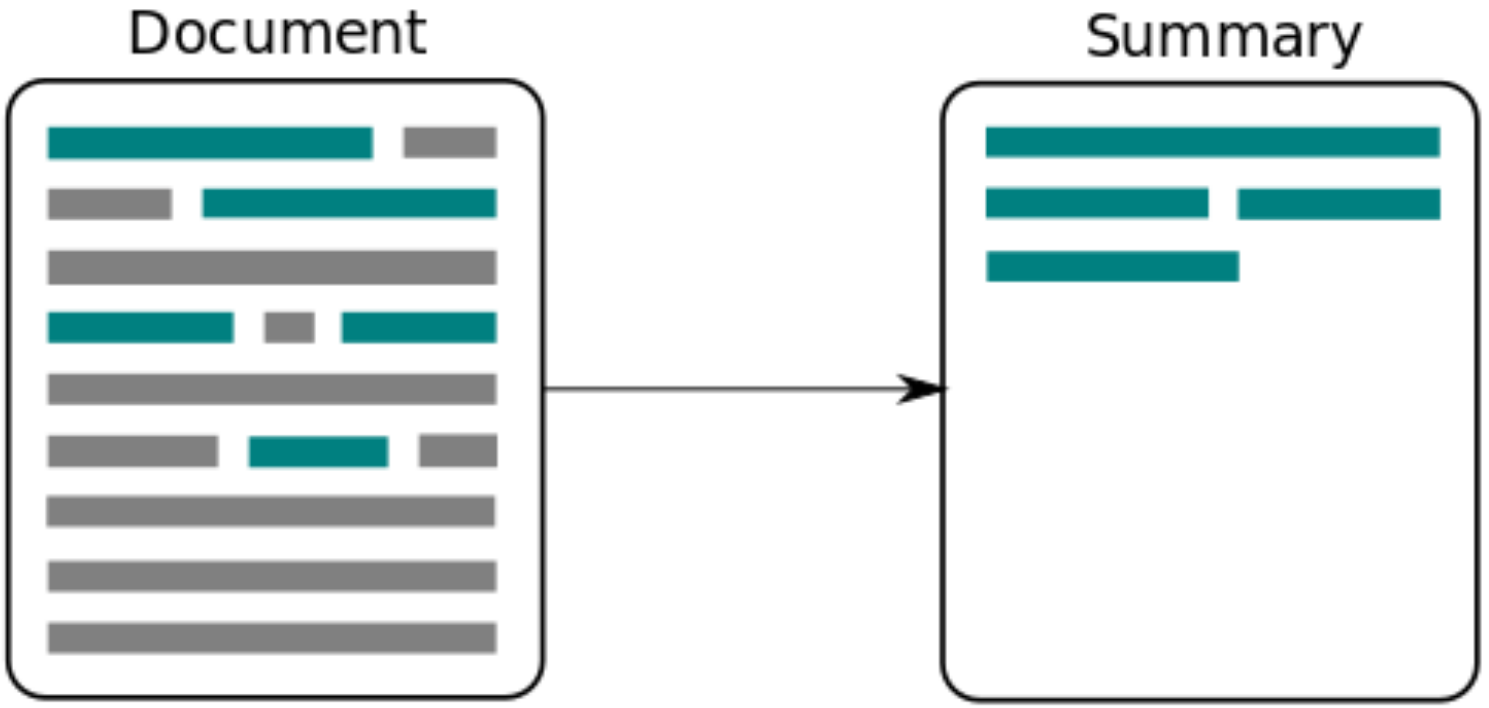
The task of **Information Extraction** involves extracting meaningful information from unstructured text

New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

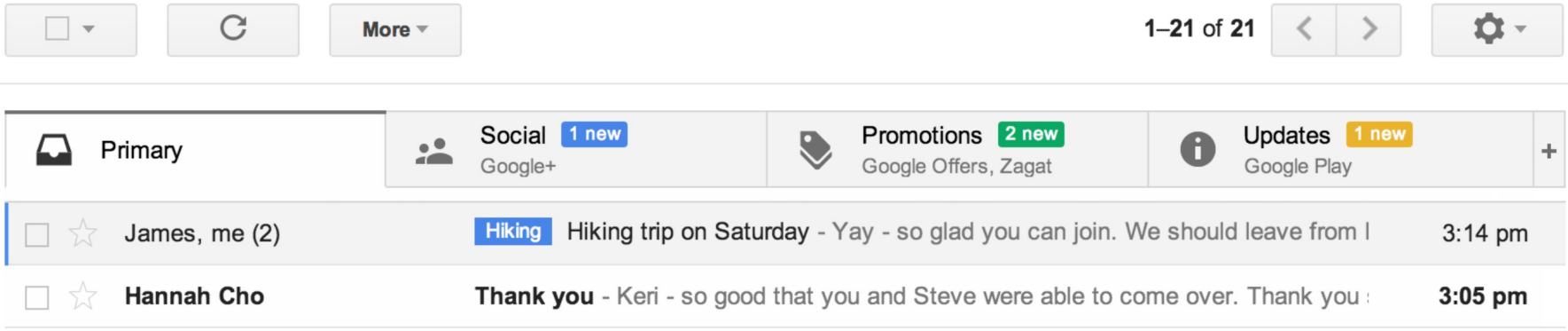
https://slideplayer.com/slide/17527407/

NLP APPLICATIONS – SUMMARIZATION



https://hackernoon.com/summarization-with-wine-reviews-using-spacy-b49f18399577

NLP APPLICATIONS – TEXT CLASSIFICATION

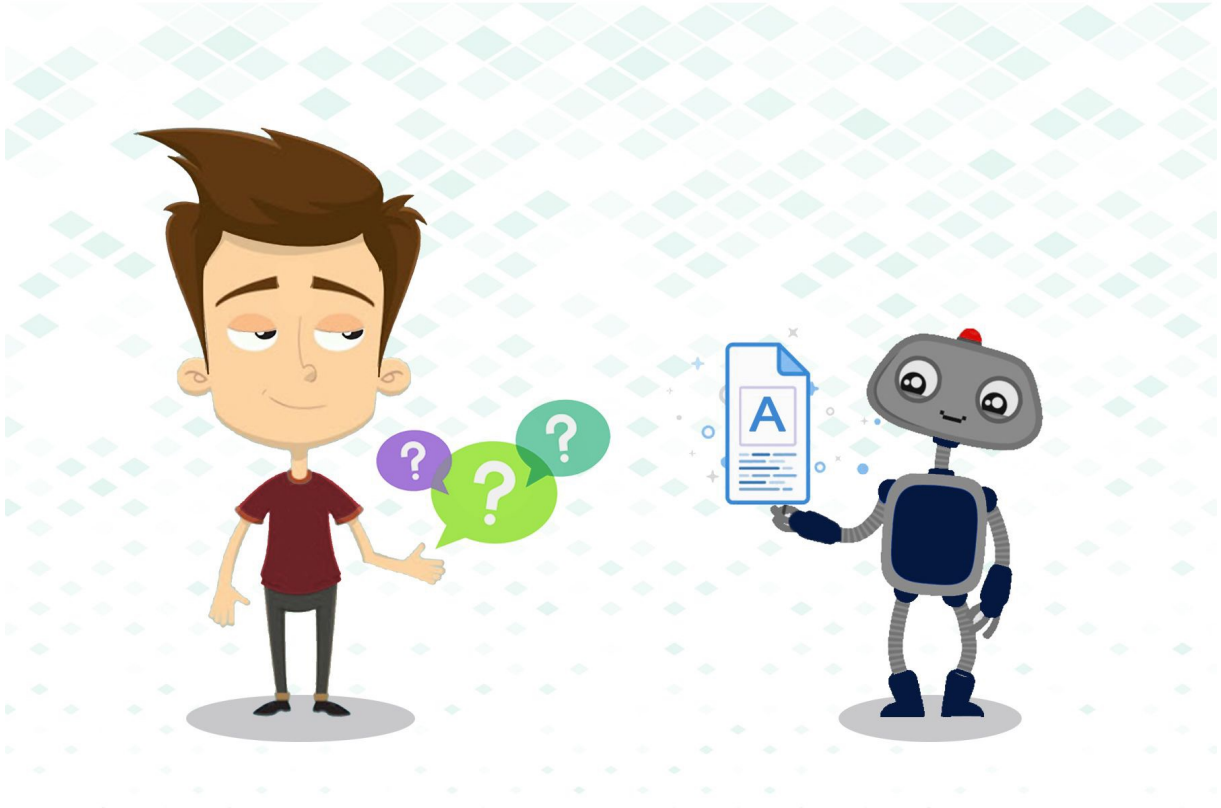


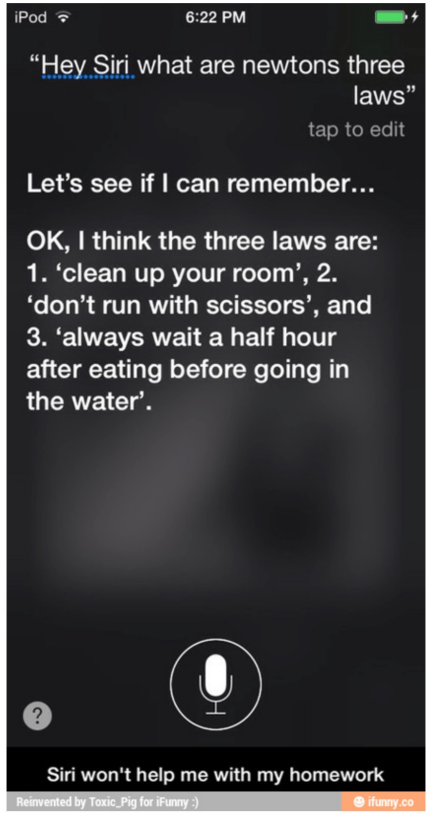
NLP APPLICATIONS – DIGITAL PERSONAL ASSISTANTS

Semantic parsing – ubderstand the task

https://www.bestaiassistant.com/siri/do-not-tell-siri-112/

NLP APPLICATIONS – QUESTION ANSWERING

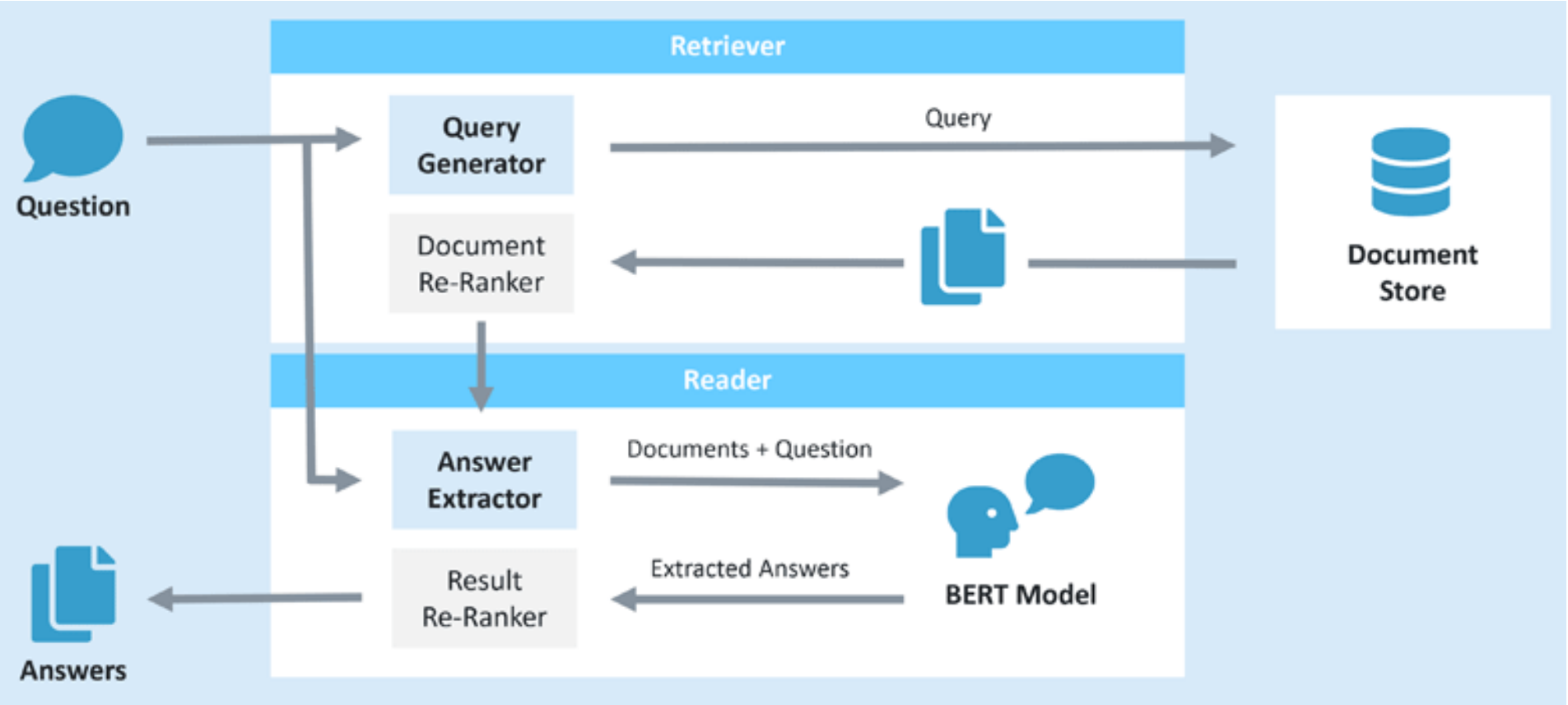




credit: ifunny.com

https://blog.paperspace.com/question-answering-models-a-comparison/

NLP APPLICATIONS – QUESTION ANSWERING



https://blog.seeburger.com/natural-language-question-answering-systems-get-quick-answers-to-concrete-questions/

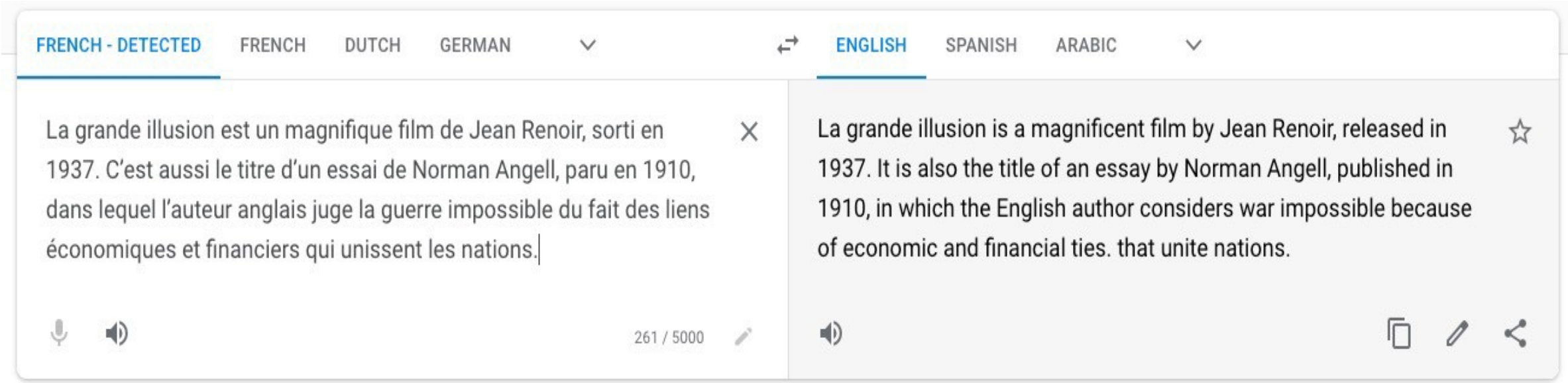
NLP APPLICATIONS – READING COMPREHENSION

More than a decade ago, Carl Lewis stood on the threshold of what was to become the greatest athletics career in history. He had just broken two of the legendary Jesse Owens' college records, but never believed he would become a corporate icon, the focus of hundreds of millions of dollars in advertising. His sport was still nominally amateur.

Eighteen Olympic and World Championship gold medals and 21 world records later, Lewis has become the richest man in the history of track and field – a multi-millionaire.

- Who is Carl Lewis?
- Did Carl Lewis break any world records?
- Is Carl Lewis wealthy? What about Jesse Owens?

NLP APPLICATIONS – MACHINE TRANSLATION



WHAT IS SPECIAL ABOUT NATURAL LANGUAGE?

Lingustic analysis

- Phonology - sounds that make up language.
- Morphology – study of words and how they are formed.
- Syntax - structure of phrases, how words modify one another.
- Semantics - meaning of language in the world.
- Discourse: relations between clauses and sentences

WHAT ARE THE CHALLENGES OF NLP?

AMBIGUITY - LANGUAGES ARE AMBIGUOUS

VARIABILITY - LANGUAGES ARE COMPLEX

UNDERSTANDING REQUIRES VAST KNOWLEDGE AND EXPERIENCE

WHY IS NLP HARD – SYNTACTIC AMBIGUITY

Semantic ambiguity: *occurs when a word, phrase or sentence, taken out of context, has more than one interpretation.*

By a tree, a stolen painting was found

A tree found a stolen painting

WHY IS NLP HARD – SYNTACTIC AMBIGUITY

Syntactic ambiguity: *two or more possible meanings within a single sentence.*

"Finally, a computer that understands you like your mother" (Ad , 1985)

- The computer understands you as well as your mother understands you.
- The computer understands that you like your mother.
- The computer understands you as well as it understands your mother.

WHY IS NLP HARD – LEXICAL AMBIGUITY

Lexical ambiguity: *two or more possible meanings of a single word*

Finally, a computer that understands your lie cured mother"

- The word *lie* can have multiple meanings in sentence the and will not change the context of the sentence.
- The ambiguity is on what cured mother
  - *lie:* an intentionally false statement
  - *lie:* spice or home-made remedy

WHY IS NLP HARD? – VARIABILITY

There are many ways to express the same meaning in language.

PWD ends up with 6 points.

PWD climbs by 6 points in the table.

PWD gains 6 points

Key computational challenge in NLP is to compute similarity of the above phrases.

WHY IS NLP HARD? – LANGUAGE IS NOT STATIC

New words added to dictionary

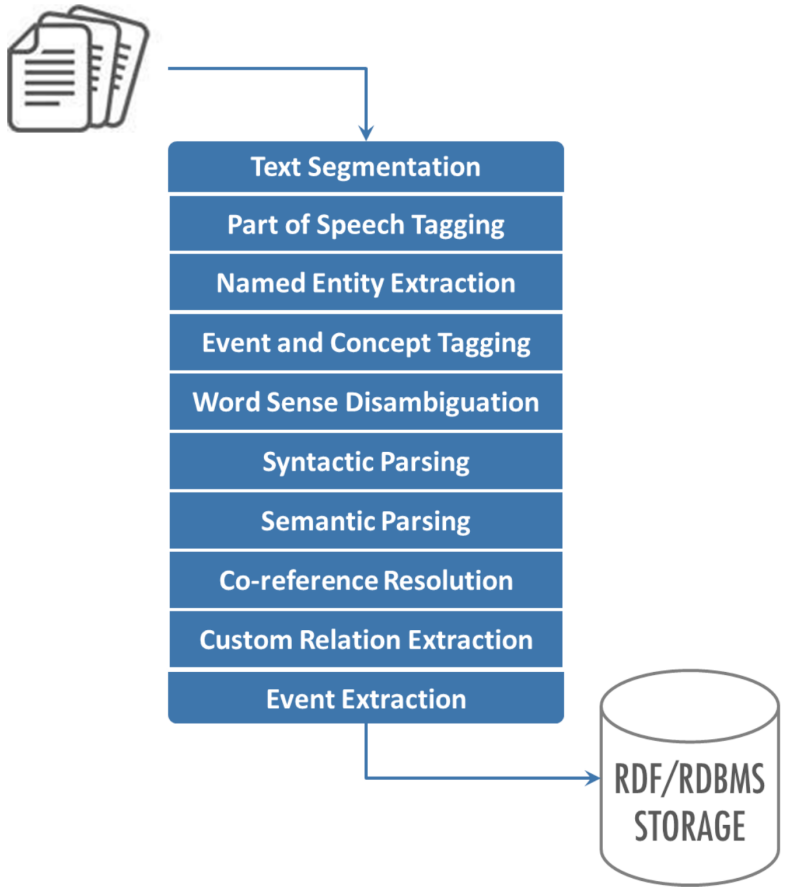
google, googling
laggy
Greenwash

cyber lingo

#TBT => throwback Thursday
DM => direct message
LOL => laugh out loud
AMA => ask me anything
Troll => online troll
Epic fall => when some one fails

WHAT YOU WILL LEARN

- Theory:
  - understanding of popular methods and techniques
- The NLP Pipeline
  - Key components for understanding text
- NLP Systems and application
  - Current limitations & techniques
- Build realistic NLP tools



TEXT BOOKS

1. Speech and Language Processing 3rd ed, Jurafsky and Martin. https://web.stanford.edu/~jurafsky/slp3/
2. Natural Language Processing, Jacob Eisenstein. https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf