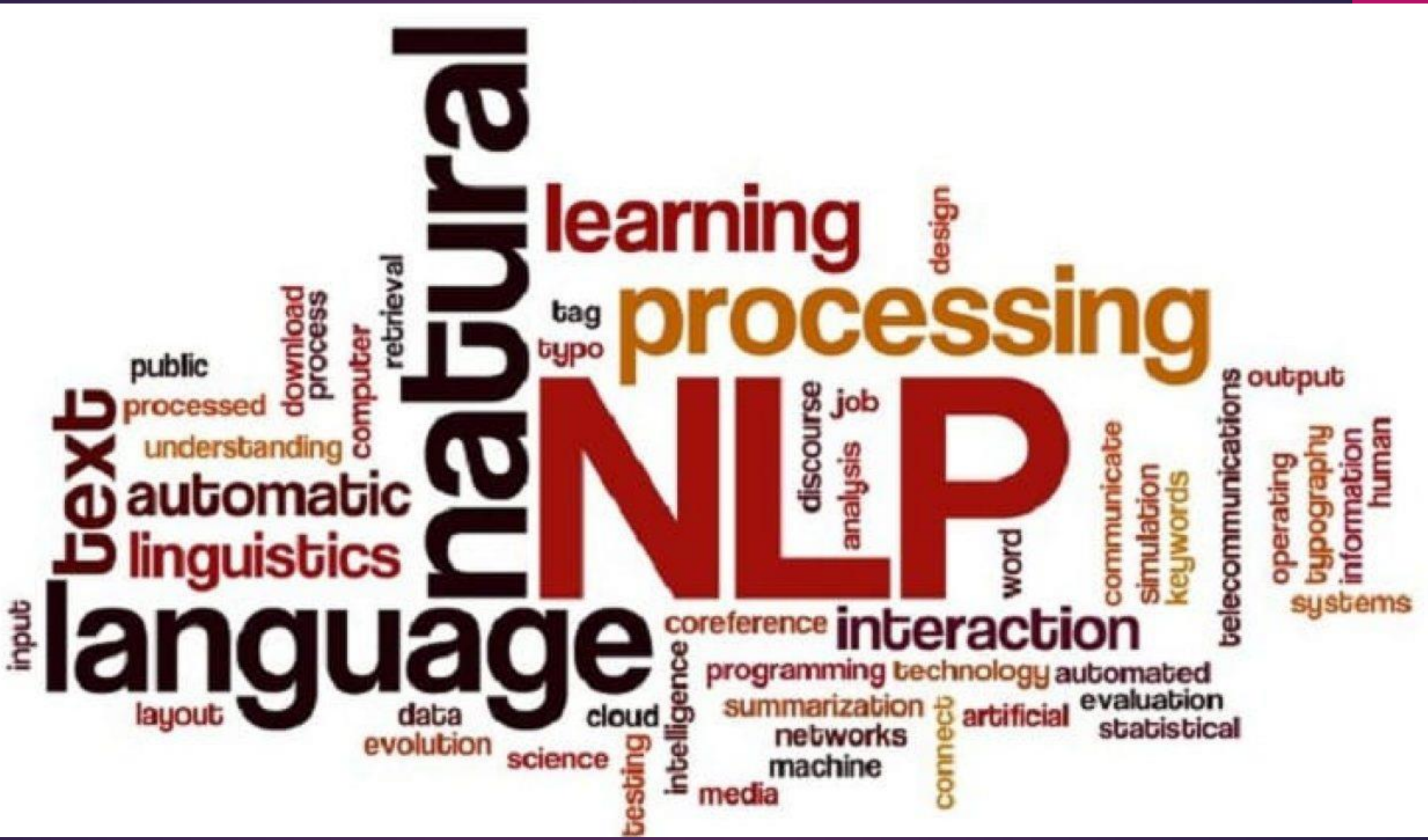
Data Science


natural
learning
processing
NLP
language
text
automatic
linguistics
interaction
public
processed
understanding
download
process
computer
retrieval
tag
typo
design
discourse
job
analysis
word
communicate
simulation
keywords
telecommunications
output
operating
typography
information
human
systems
input
coreference
programming
technology
automated
evaluation
statistical
artificial
summarization
networks
machine
media
connect
intelligence
testing
cloud
data
evolution
science
layout

Natural Language Processing

NLP

معالجة اللغات الطبيعية

Ahmad Shhadeh
5/31/2020

Agenda

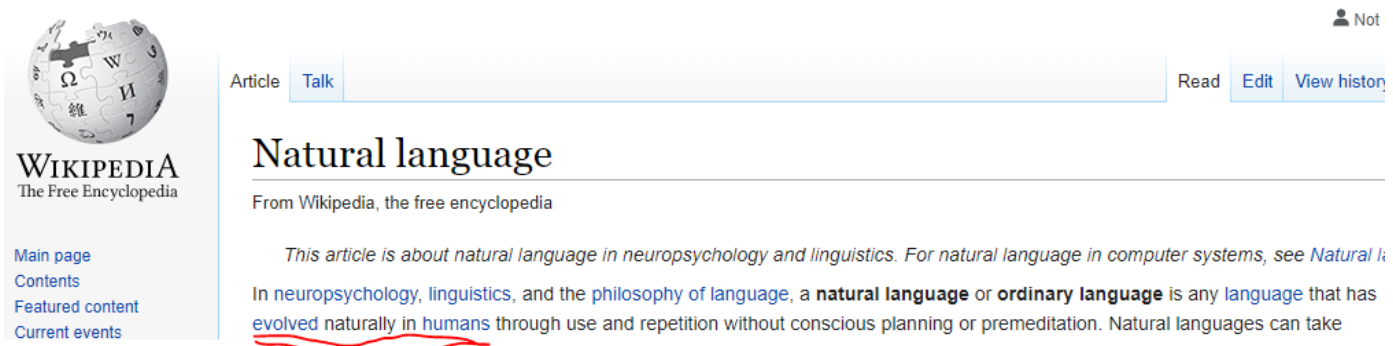
1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

What is Natural Language Processing(NLP)?

- **Natural Language (Human Language)**
  - Arabic language
  - English language
  - French Language
  - …..etc



- Human get **the edge** due to the communication skills he has
- Roughly **6,500 languages** are spoken in the world today
- **Programming language** (Python language, C++,Java..etc) is different

- **Processing**
  - How computers carries out instructions .How to **deal with Text data**

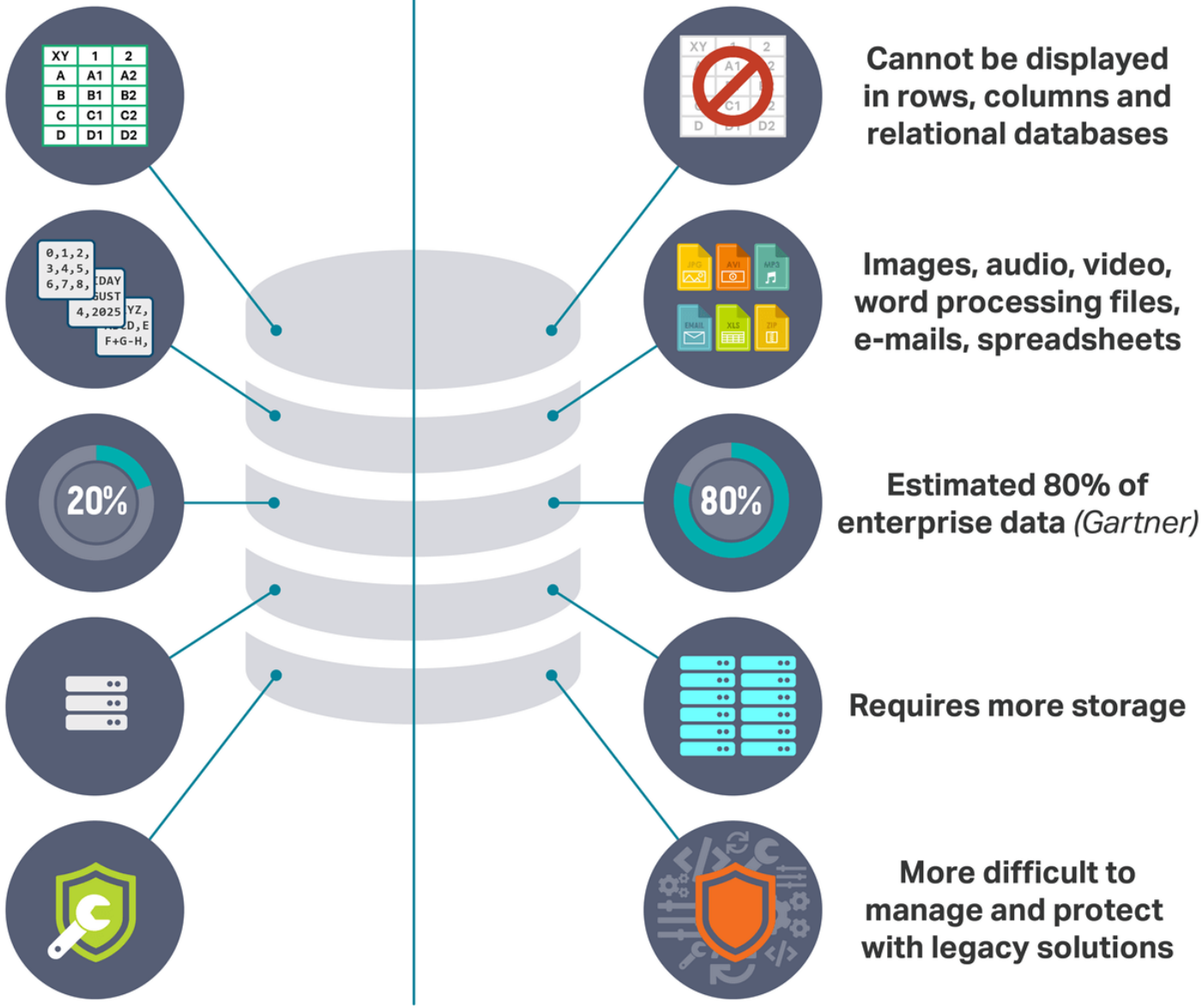
What is NLP?

- **Natural Language Processing (NLP)** is defined as the automatic manipulation of natural languages, such as speech and text, by using software or any programming language.
- The ultimate objective of NLP is to **read**, **decipher**, **understand**, **and make sense** of the human languages in a manner that is valuable.
- **Transforming free-form text into structured data and back**
- Most NLP techniques rely on **machine learning** to derive meaning from human languages..
- As a business tool, NLP helps to drive better decision-making by applying computer intelligence. It also identifies hot discussion topics and consumers' interest charts. For instance, marketers use sentiment analysis for consumer insights regarding brand preference.
- It's not an easy task teaching machines to understand **how we communicate**.☺☺

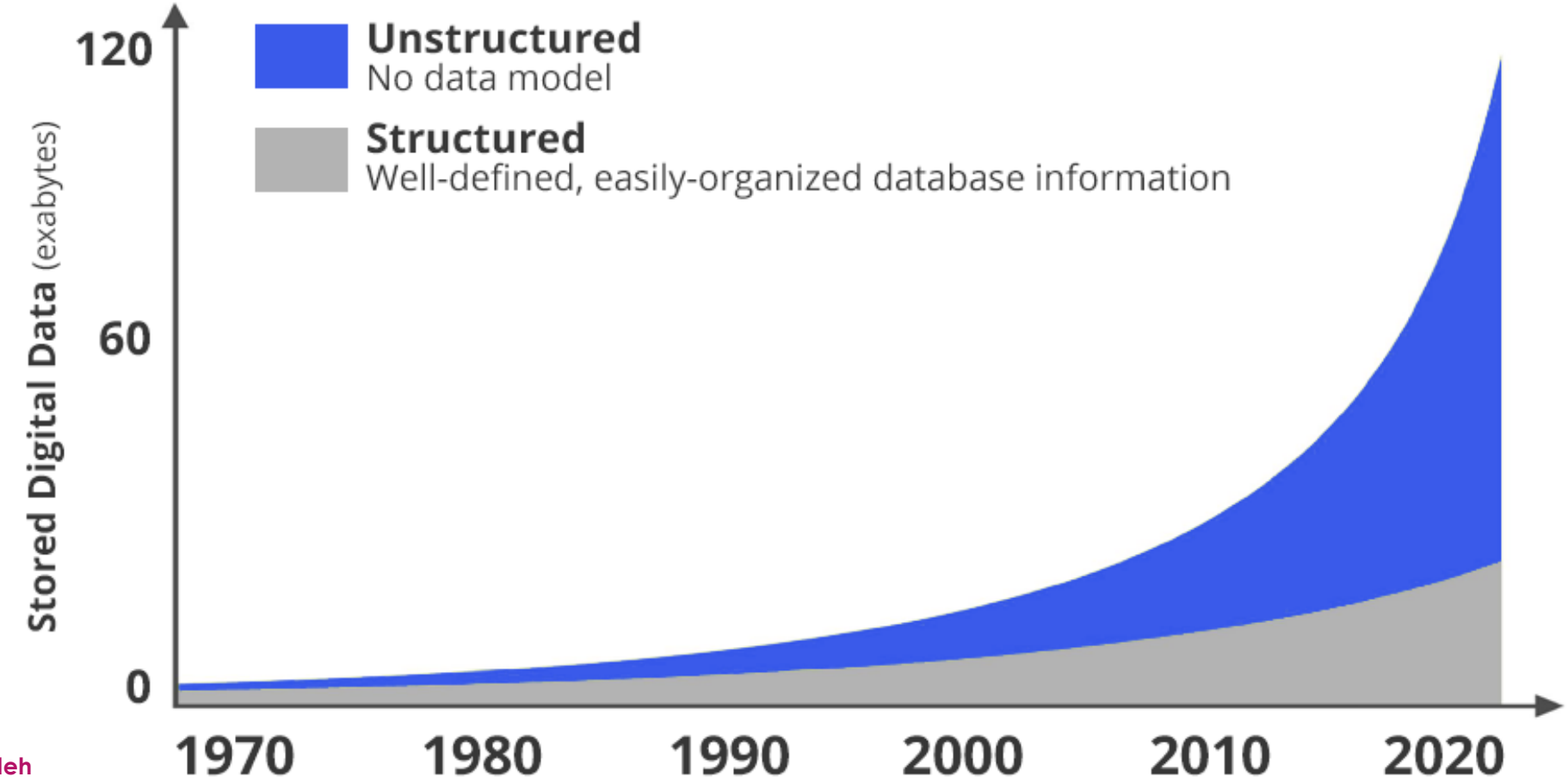
Agenda

1. What is NLP?
2. **Structured Data vs. Unstructured Data**
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

Structured Data vs Unstructured Data



Structured Data vs. Unstructured Data



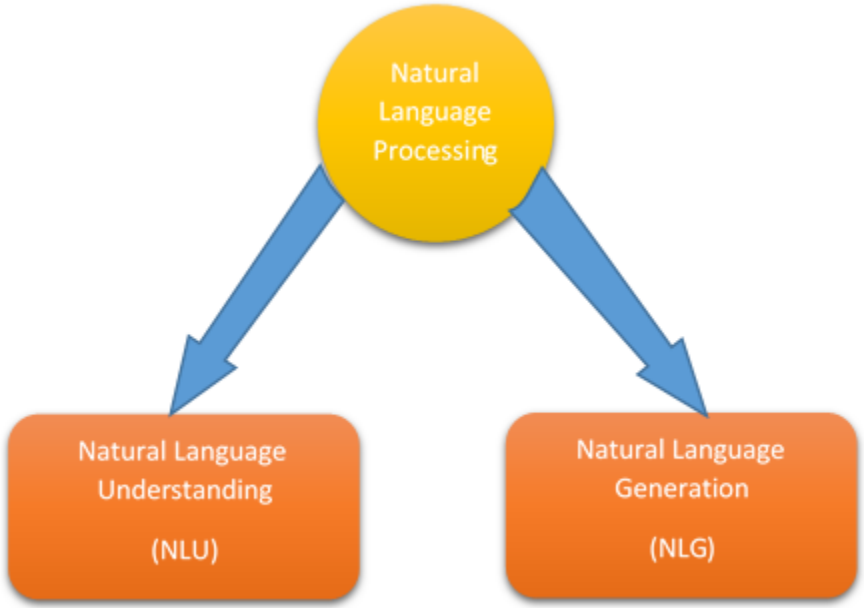
Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. **NLP Components**
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

NLP Components

- The term NLP can be divided into two major components
- **Natural language understanding (NLU)**
- **Natural language generation (NLG)**

- Or, in simple terms **:NLP consist of :**
  - Turning text into data, Then
  - Turning data into text.



Natural Language Processing(NLP)

- In fact, a typical interaction between humans and machines using Natural Language Processing could go as follows:

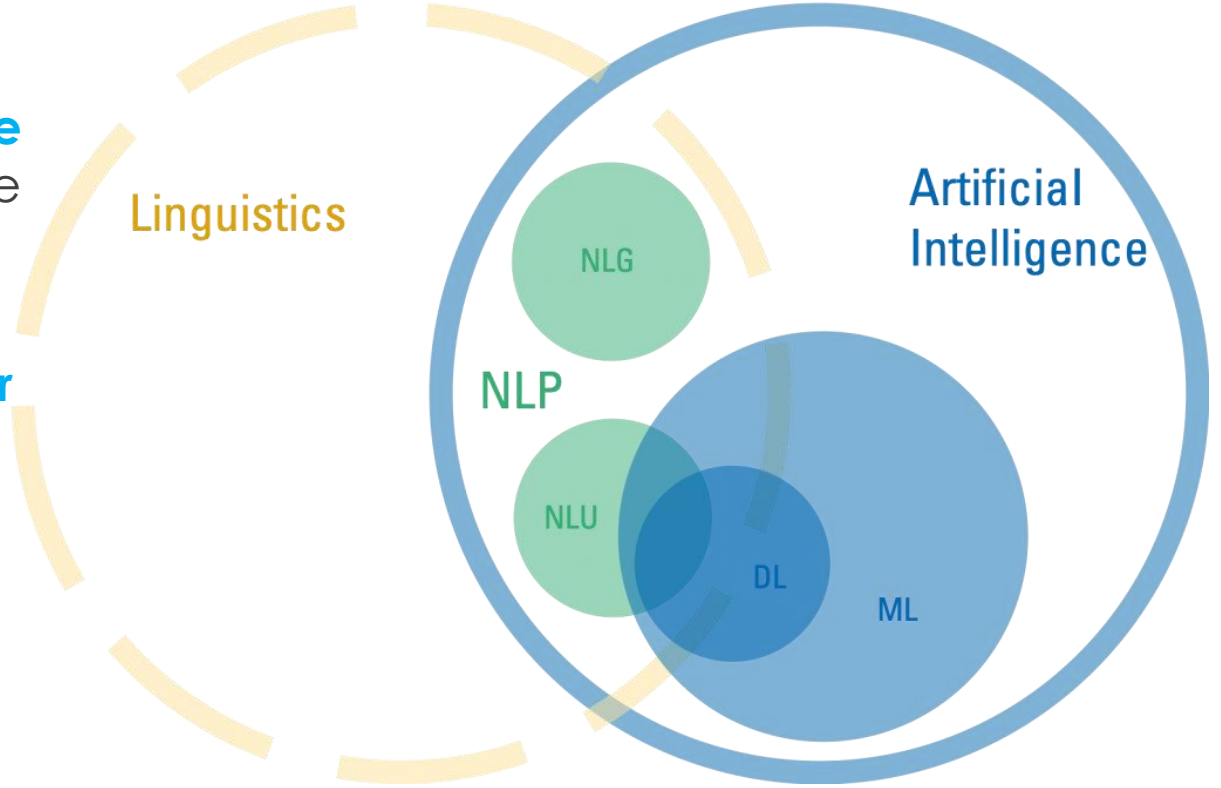
1. A human talks to the machine
2. The machine captures the audio
3. Audio to text conversion takes place
4. Processing of the text's data
5. Data to audio conversion takes place
6. The machine responds to the human by playing the audio file

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. **NLP, Artificial intelligence & Machine Learning**
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

NLP, Artificial intelligence & Machine Learning

- **NLP**: is a branch of **artificial intelligence .**
- Human readable natural language processing **is the biggest AI- problem.** It is all most same as solving the central artificial intelligence problem and making computers as intelligent as people.
- NLP usually using **common ML –algorithms with other domains ,and has it is own special algorithms .**
- **Open the door for ML specialization**



Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. **Main approaches in NLP(TimeLine)**
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

Main approaches in NLP(TimeLine)

1. **Rule-based methods**
   - Regular expressions
   - Context-free grammars
   - …
2. **Machine Learning or 'Traditional' Approach**
   - Likelihood maximization
   - Linear classifiers
   - …
3. **Deep Learning (State of the Art)**
   - Recurrent Neural Networks
   - Convolutional Neural Networks
   - check the time line

Timeline

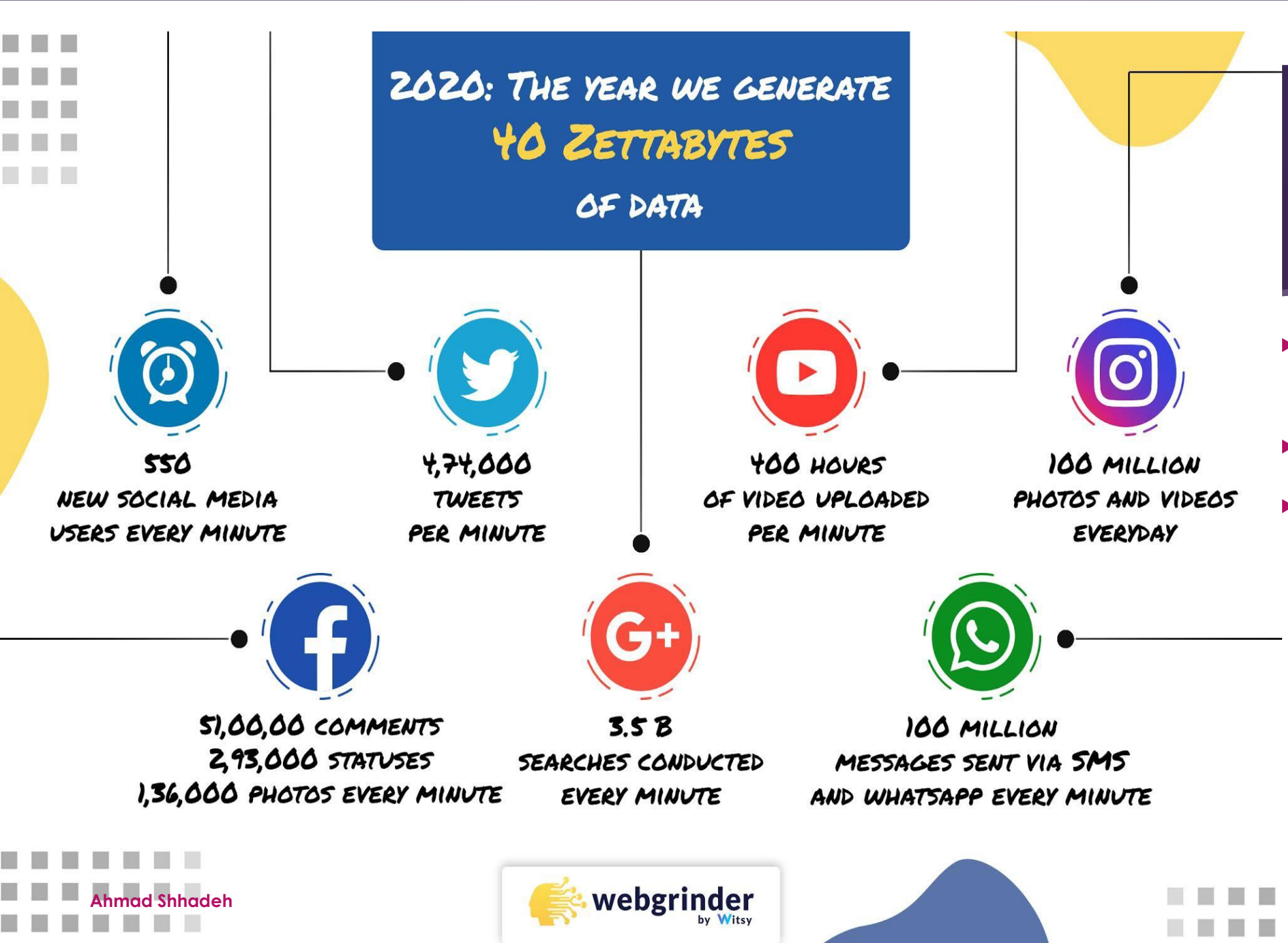
2001	Neural language models
2008	Multi-task learning
2013	Word embeddings
2013	Neural networks for NLP
2014	Sequence-to-sequence models
2015	Attention
2015	Memory-based networks
2018	Pretrained language models

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. **Why NLP is very important?**
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

Why NLP is important?

- **NLP is everywhere even if we don't realize it**.
- **The majority of activities performed by humans are done through language.**
- There are millions of gigabytes of data generated by **Social media** (Facebook, Instagram, Twitter, YouTube etc.),**Apps messages** (Whatsapp, WeChat, Telegram etc.),**Forums** (Quora, Reddit etc.),**Blogs**, news publishing platforms, **google searches** and many other channels.
- All these channels are constantly generating large amount of **text data every second**.
- And because of the large volumes of text data as well as the highly unstructured data source, we can no longer use the common approach to understand the text and this is where NLP comes in.
- NLP produces **new and exciting results** on a daily basis, and is a very large field.



- 20 % of data are in structured form
- 80% unstructured
- And then majority of data come on text form

Why NLP is important?

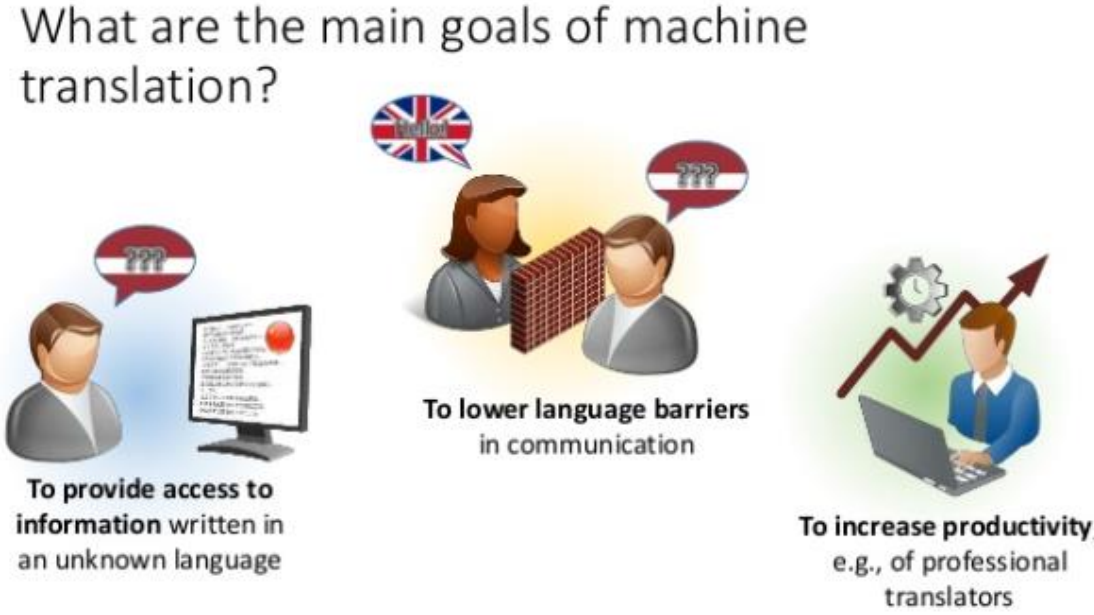
- NLP allows companies to **track, manage, and analyze billions of ever-changing data points**. This way, companies make sense of all this information and use it to make decisions about their businesses
- NLP helps systems **analyze data faster** By combining **the power of artificial intelligence**, **computational linguistics** and **computer science**.
- **NLP helps bring semantic understanding to languages:** NLP systems help resolve confusing, ambiguous language by adding structure to the data they receive.
- With NLP, there are several successful implementations with search engine like **Google; social websites like Facebook's news feeds; speech engines like Apple Siri**; and spam filters.

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. **Natural Language Processing Applications**
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

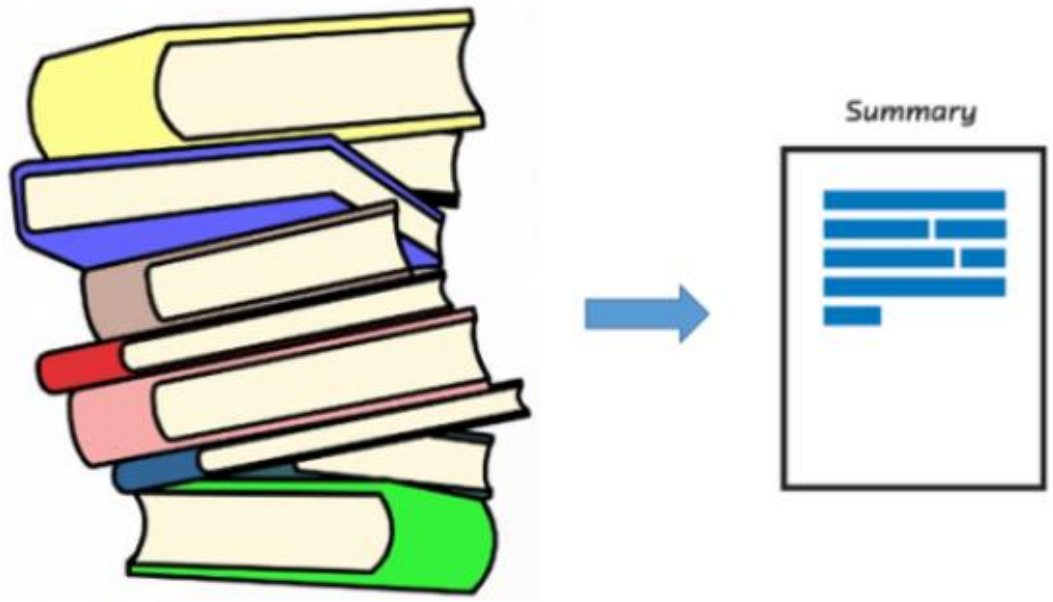
NLP Applications :Machine Translation

- As the **amount of information available online is growing**, the need to access it becomes increasingly important and the value of natural language processing applications becomes clear.
- The challenge with **machine translation technologies is not in translating words, but in understanding the meaning** of sentences to provide a true translation.



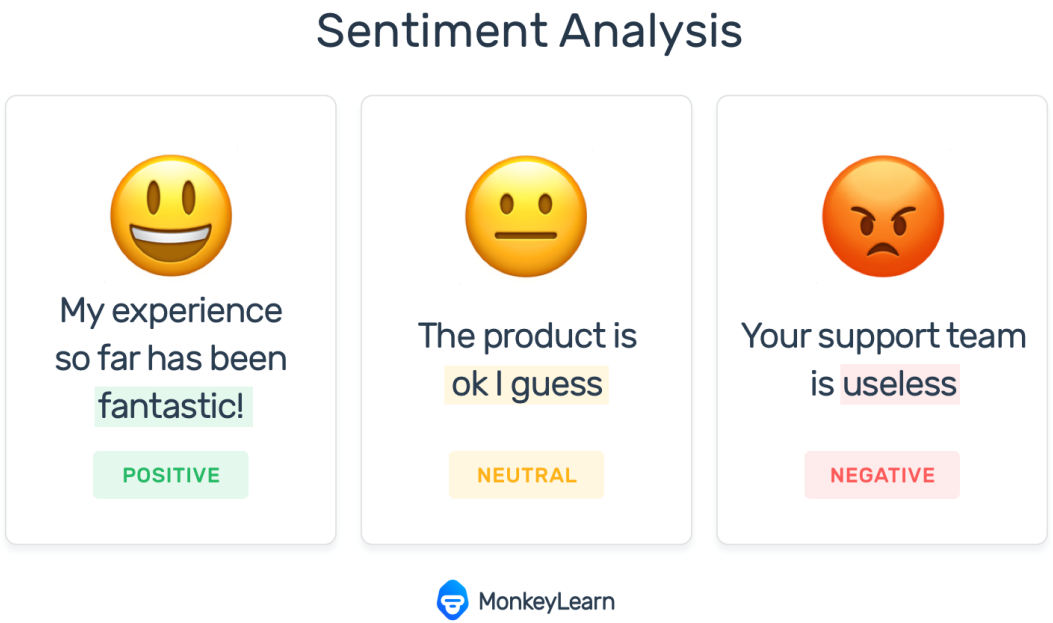
NLP Applications :Automatic summarization

- **Information overload is a real problem** when we need to access a specific, important piece of information from a huge knowledge base.
- Automatic summarization is relevant not only for **summarizing the meaning of documents and information**, but also for **understand the emotional meanings inside the information**, such as in collecting data from social media.
- **Automatic summarization** is especially relevant when used to provide an overview of a news item or blog posts, while avoiding redundancy from multiple sources and maximizing the diversity of content obtained.



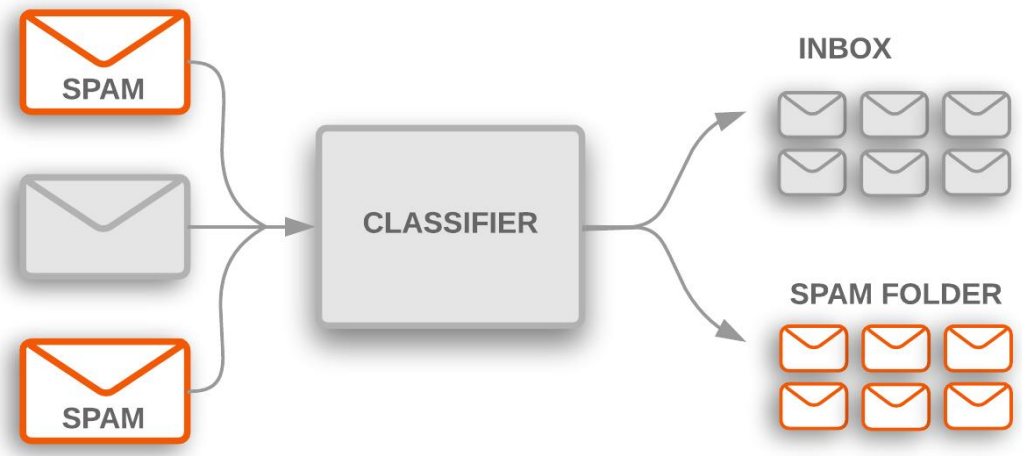
NLP Applications :Sentiment analysis

- **The goal of sentiment analysis is to identify sentiment among several posts** or even in the same post **where emotion is not always explicitly expressed**.
- Companies use sentiment analysis, to identify opinions and sentiment online to help them **understand what customers think about their products and services**
- Beyond determining simple polarity, sentiment analysis understands sentiment in context to help you better understand what's behind an expressed opinion, which can be extremely relevant in **understanding and driving purchasing decisions**.



NLP Applications :Text classification

- **Text classification** makes it possible to assign predefined categories to a document and **organize it to help you find the information** you need or simplify some activities. For example, an application of text categorization is **spam filtering in email**.



NLP Applications :Question Answering

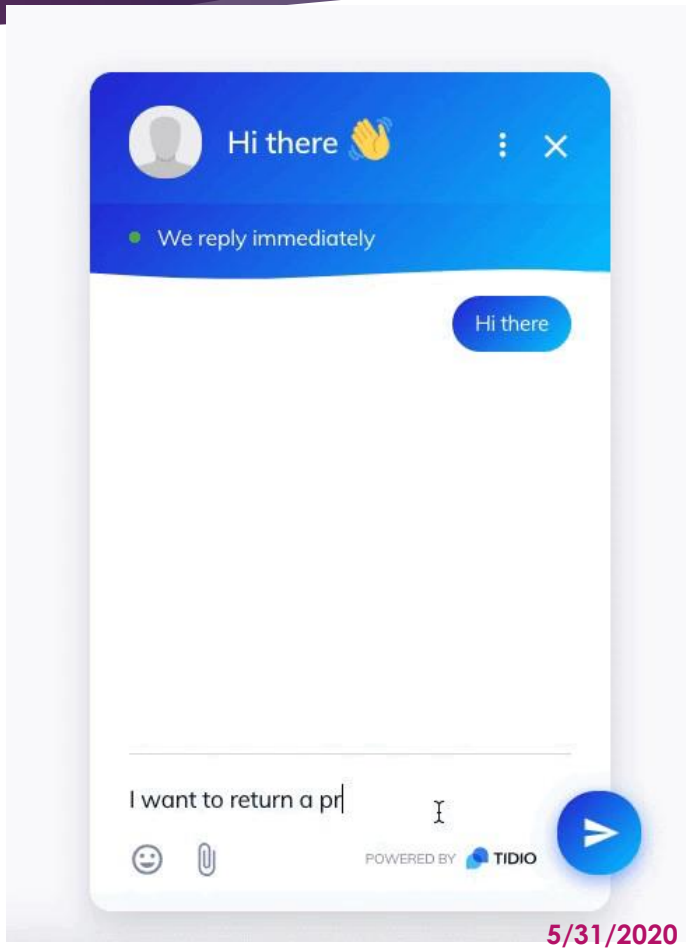
- As speech-understanding technology and voice-input applications improve, **the need for NLP will only increase.**
- **Question-Answering (QA)** is becoming more and more popular thanks to applications such as **Siri, OK Google, chat boxes and virtual assistants**.
- A QA application is a system capable of coherently answering a human request.

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. **Areas That Leverage NLP Technology**
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

Areas That Leverage NLP Technology: Chatbots

- The use of chatbots in **maintaining business workflow** is considerable and beneficial for every Industry.
- It also enables bots to respond to customer **queries faster than a human being**. The faster responses help in building **customer trust** and more business.
- NLP, when paired with **voice recognition** technology, can make chatbots smarter.
- Chatbot interactions nowadays can be easily confused with human interactions because they are intelligent and also can **recognize human emotions**.
- NLP helps chatbots **analyze, understand, and prioritize complex questions.**
- **Gartner** has predicted that chatbots will account for 85% of customer interactions in 2020.



Areas That Leverage NLP Technology: E-commerce

- With the exponential growth of multi-channel data like social or mobile data, businesses need solid technologies in place to assess and evaluate customer sentiments. So far, businesses have been happy analyzing customer actions, but in the current competitive climate, **that type of customer analytics is outdated.**
- Now businesses need to **analyze and understand customer attitudes, preferences, and even moods** – all of which come under the purview of sentiment analytics. Without NLP, business owners would be seriously handicapped in conducting even the most basic sentiment analytics.
- With the help of NLP, machines can easily pick out **what phrases and words** are generally used by humans while searching on **a particular product** on any ecommerce website.
- NLP helps in **customizing** the searches for users using search engines. The system finds what the user is **exactly searching** for by using its understanding of language and sentence structure. It also detects patterns and creates links between messages to discover the meaning of unstructured text.
- Smart Product Recommendations

Areas That Leverage NLP Technology..Cont

Sentiment Analysis

- **A classic example of NLP**, sentiment analysis can help estimate how customers feel about the brand when it comes to adjusting sales and marketing strategy.
- This technology is also known as opinion mining and is capable of analyzing news and blogs and assigning a value to the text (**positive, negative, or neutral**).
- NLP algorithms enable you to identify emotions such **as happy, annoyed, angry, and sad**. In addition, a sentiment analysis tool increases customer loyalty, drives business changes, and achieves an appropriate return on sales and marketing investments.

Hiring & Recruitment

- By utilizing NLP, HR professionals can significantly **speed up candidate searches**, filtering out relevant resumes and creating bias-proof and gender-neutral job descriptions. By using semantic analysis,
- NLP-based software helps recruiters to detect candidates that meet a job's requirements.
- **Textio** is a real example of using semantic categorization to tweak job descriptions in a way to maximize the number of job applicants.

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. **Why is NLP so difficult?**
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

Why is NLP difficult?

- It's the **nature of the human language** that makes NLP difficult.
- Human gets the edge due to the communication skills he has.
- There are **hundreds of natural languages**, each of which has different syntax rules. Words can be ambiguous where their meaning is dependent on their context.
- The rules that dictate the passing of information using natural languages are **not easy** for computers to understand.
- Some of these rules can be high-leveled and abstract; for example, when someone uses a **sarcastic** remark to pass information.
- Comprehensively understanding the human language requires understanding both **the words** and **how the concepts are connected** to deliver the intended message.
- While humans can easily master a language, the ambiguity and imprecise characteristics of the natural languages are what make NLP difficult for machines to implement.

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. **What are the techniques used in NLP?**
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

What are the techniques used in NLP?

I. Syntax analysis :

Syntax refers to the **arrangement of words** in a sentence such that they make grammatical sense. In NLP, syntactic analysis is used to assess how the natural language aligns with the grammatical rules. Here are some syntax techniques that can be used:

- **Lemmatization** : It entails reducing the various inflected forms of a word into a single form for easy analysis.
- **Stemming**: It involves cutting the inflected words to their root form.
- **Morphological segmentation**: It involves dividing words into individual units called morphemes.
- **Word segmentation**: It involves dividing a large piece of continuous text into distinct units.
- **Part-of-speech tagging**: It involves identifying the part of speech for every word.
- **Parsing**: It involves undertaking grammatical analysis for the provided sentence.
- **Sentence breaking**: It involves placing sentence boundaries on a large piece of text.

What are the techniques used in NLP?

2. Semantics Analytics

- ✓ **Semantics** refers to the meaning that is conveyed by a text. Semantic analysis is one of the difficult aspects of Natural Language Processing that has not been fully resolved yet.
- ✓ It involves applying computer algorithms to understand the **meaning** and interpretation of words and how sentences are structured.

Here are some techniques in semantic analysis:

- **Named entity recognition (NER):** It involves determining the parts of a text that can be identified and categorized into preset groups. Examples of such groups include names of people and names of places.
- **Word sense disambiguation:** It involves giving meaning to a **word based on the context**.
- **Natural language generation**: It involves using databases to derive semantic intentions and convert them into human language.

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. **Libraries and tools**
12. The Future of NLP
13. Arabic Natural Language Processing

Libraries and tools

- **NLTK**
  - Small but useful datasets with markup
  - Preprocessing tools: tokenization, normalization…
  - Pre-trained models for POS-tagging, parsing…
- **Stanford parser**
- **spaCy:**
  - python and cython library for NLP
- **Gensim**
  - python library for text analysis, e.g. for word
  - embeddings and topic modeling
- **MALLET**
  - Java-based library, e.g. for classification,
  - sequence tagging, and topic modeling
-

Libraries and tools-NLTK

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. **Arabic Natural Language Processing**
13. The Future of NLP

Arabic Natural Language Processing

Content languages for websites

Rank	Language	Percentage
1	English	59.3%
2	Russian	8.4%
3	Spanish	4.2%
4	German	2.9%
5	Turkish	2.9%
6	Persian	2.8%
7	French	2.8%
8	Japanese	2.4%
9	Portuguese	2.2%
10	Chinese	1.3%
11	Vietnamese	1.3%
12	Italian	1.0%
13	Arabic	0.9%
14	Polish	0.9%
15	Greek	0.7%
16	Dutch	0.7%

Internet users by language

Rank	Language	Internet users	Percentage
1	English	1,186,451,052	25.9%
2	Chinese	888,453,068	19.4%
3	Spanish	363,684,593	7.9%
4	Arabic	237,418,349	5.2%
5	Indonesian / Malaysian	198,029,815	4.3%
6	Portuguese	171,750,818	3.7%
7	French	151,733,611	3.3%
8	Japanese	118,626,672	2.6%
9	Russian	116,353,942	2.5%
10	German	92,525,427	2.0%

Arabic Natural Language Processing

- Arabic language is recognized as the **4th most used language of the Internet**.
- Despite its cultural, religious, and political significance, Arabic has received **comparatively little attention in modern computational linguistics.**
- Complexities of the Arabic language: Beyond the traditional challenges of natural language processing for English, there are unique complexities for the Arabic language.
  - Lack of diacritical marks
  - Free word order
  - The right-to-left direction of the text.
  - Cursive writing,
  - presence of extra non-significant characters, and more.
  - …..etc

Agenda

1. What is NLP?
2. Structured Data vs. Unstructured Data
3. NLP Components
4. NLP, Artificial intelligence & Machine Learning
5. Main approaches in NLP(TimeLine)
6. Why NLP is very important?
7. Natural Language Processing Applications
8. Areas That Leverage NLP Technology
9. Why is NLP so difficult?
10. What are the techniques used in NLP?
11. Libraries and tools
12. Arabic Natural Language Processing
13. The Future of NLP

The Future of NLP

- Natural language processing (NLP), **one of the most exciting components of AI**
- NLP is the voice behind **Siri and Alexa, likewise, customer service chatbots harness the power of NLP to drive customized responses in e-commerce, healthcare and business utilities. Some of the more omnipresent applications of NLP today include virtual assistants, sentiment analysis, customer service, and translation.**
- According to many market statistics, data volume is doubling every two years, but in future this time span may get further reduced. **The vast portion of this data (about 75 percent) is text data**.
- NLP is the sub-branch of Data Science that attempts to extract insights from "text." Thus, NLP is assuming an **important role in Data Science**. Industry experts have predicted that the demand for NLP experts **will grow exponentially in the near future.**
- **Using natural language processing for creating a seamless and interactive interface** between humans with machines will continue to be a top priority for today's and tomorrow's increasingly cognitive applications.
- NLP everywhere , **there is a potential ,Opportunities, Jobs and Money**.

The Future of NLP

- https://en.wikipedia.org/wiki/Languages_used_on_the_Internet
- https://www.datasciencecentral.com/profiles/blogs/comparison-of-top-6-python-nlp-libraries
- https://lawtomated.com/structured-data-vs-unstructured-data-what-are-they-and-why-care/

THANK YOU

Ahmad Shhadeh
5/31/2020