# INTRODUCTION TO NATURAL LANGUAGE PROCESSING

What is NLP?

➢ Natural Language Processing (NLP) is the sub-field of Computer Science especially Artificial Intelligence (AI) that is concerned about enabling computers to understand and process human language.

What is NLP?

➤ Natural Language Processing (NLP) is the sub-field of Linguistics ,Computer Science and Artificial Intelligence (AI) concerned with the interaction between computers and human languages

➤ In particular how to enable computers to understand, process, and analyse large amounts of natural language data.

SAXON
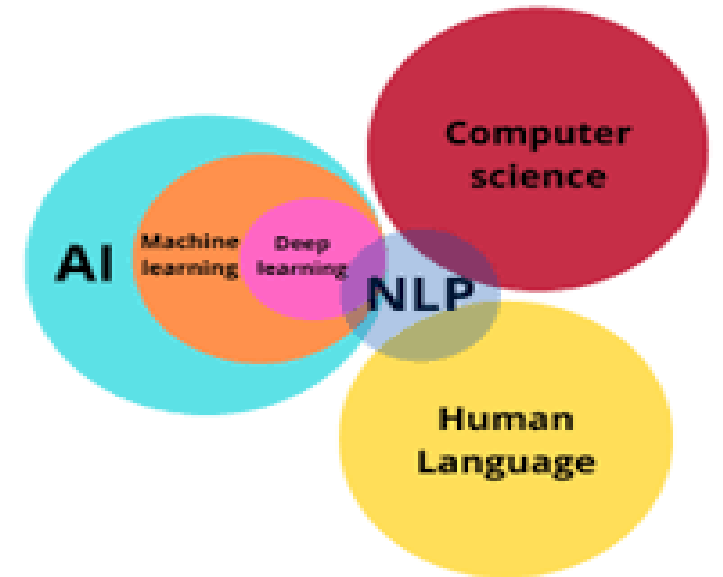
Natural Language

➢ In neuropsychology, linguistics, and philosophy of language, a natural language or ordinary language is any language that occurs naturally in a human community by a process of use, repetition, and change without conscious planning or premeditation.

➢ It can take different forms, namely either a spoken language or a sign language.

➢ Natural languages are distinguished from constructed and formal languages such as those used to program computers or to study logic
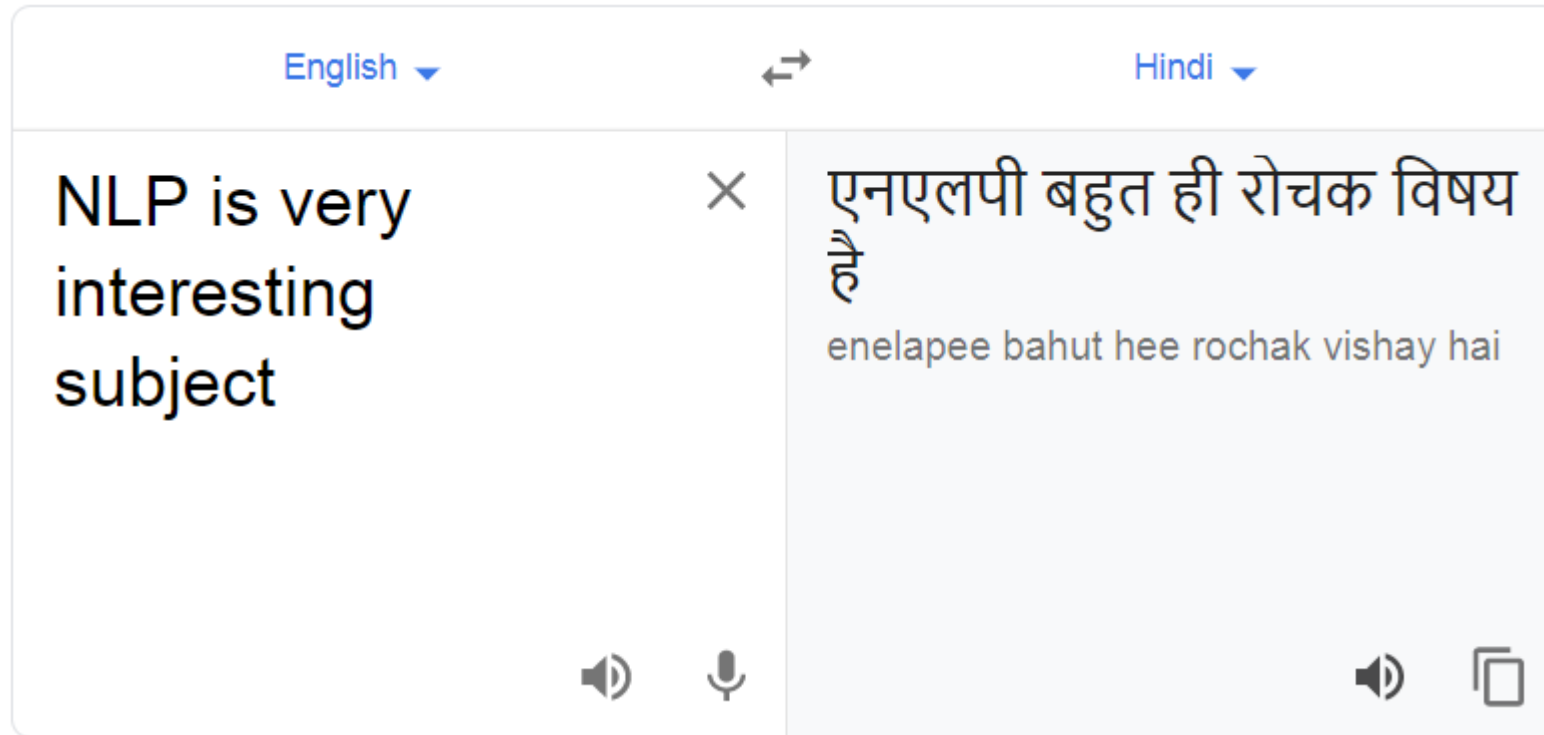
Goal of NLP

Fundamental and Scientific Goal
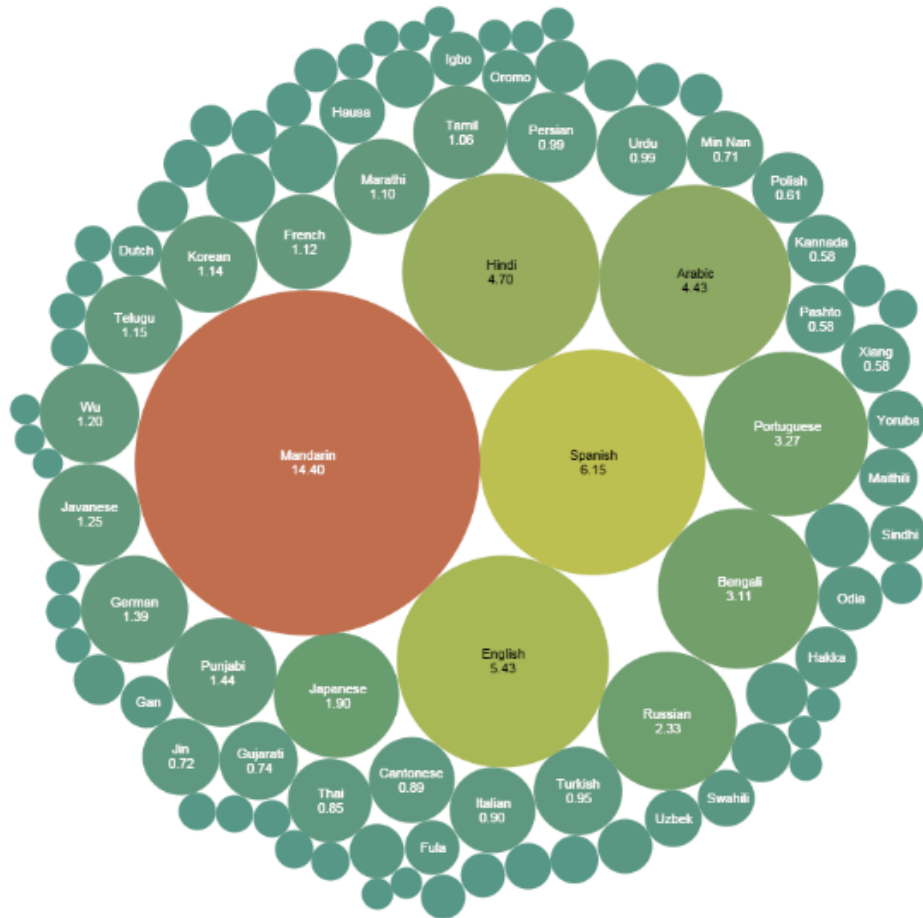
○ Deep understanding of broad language

Engineering Goal

◦ Design, implement, and test systems that process natural languages for practical applications

Goals can be very ambitious: Good quality translation

Why study NLP?

➢ Text is the largest repository of human knowledge.

➢ NLP is all about processing data.

➢ Sources:News articles, web pages, scientific articles, patents, emails, government documents ,Tweets, Facebook posts, comments, Quora

Why is Language Processing hard?

Consider trying to build a system that would answer email sent by customers to a retailer selling laptops and accessories via the Internet.

• Has my order number 4291 been shipped yet?

• Is FD5 compatible with a 505G?

• What is the speed of the 505G?

Assume the query is to be evaluated against a database containing product and order information, with relations such as the following:

ORDER

Order number	Date ordered	Date shipped
4290	2/2/02	2/2/02
4291	2/2/02	2/2/02
4292	2/2/02	

USER: Has my order number 4291 been shipped yet?

DB QUERY: order(number=4291,date shipped=?)

RESPONSE TO USER: Order number 4291 was shipped on 2/2/02

➢ Very similar strings can mean very different things, while very different strings can mean much the same thing

◦ 1. How fast is the 505G?

◦ 2. How fast will my 505G arrive?

◦ 3. Please tell me when I can expect the 505G I ordered.

➢ Ambiguity

➢Do you sell Sony laptops and disk drives?

➢Do you sell (Sony laptops) and (disk drives)? Do you sell (Sony (laptops and disk drives))?

➢Hand-coding such knowledge in NLP applications has turned out to be impossibly hard to do

History of NLP

➢ First Phase (Machine Translation Phase) – Late 1940s to late 1960s

➢ Second Phase (AI Influenced Phase) – Late 1960s to late 1970s

➢ Third Phase (Grammatico-logical Phase) – Late 1970s to late 1980s

➢ Fourth Phase (Lexical & Corpus Phase) – The 1990s

Study of Human Languages

Two forms – written and spoken

Discipline	Problems	Tools
Linguists	How phrases and sentences can be formed with words? What curbs the possible meaning for a sentence?	Intuitions about well-formedness and meaning. Mathematical model of structure. For example, model theoretic semantics, formal language theory.
Psycholinguists	How human beings can identify the structure of sentences? How the meaning of words can be identified? When does understanding take place?	Experimental techniques mainly for measuring the performance of human beings. Statistical analysis of observations.
Philosophers	How do words and sentences acquire the meaning? How the objects are identified by the words? What is meaning?	Natural language argumentation by using intuition. Mathematical models like logic and model theory.
Computational Linguists	How can we identify the structure of a sentence How knowledge and reasoning can be modeled? How we can use language to accomplish specific tasks?	Algorithms Data structures Formal models of representation and reasoning. AI techniques like search & representation methods.

Ambiguity and Uncertainty in Language

➤ Ambiguity is the capability of being understood in more than one way.

➤ Natural language is very ambiguous.

➤ NLP has the following types of ambiguities:

  ➤ Lexical Ambiguity

  ➤ Syntactic Ambiguity

  ➤ Semantic Ambiguity

  ➤ Anaphoric Ambiguity

  ➤ Pragmatic ambiguity

1. **Lexical Ambiguity**

   • The ambiguity of a single word is called lexical ambiguity.

   • Treating the word silver as a noun, an adjective, or a verb

2. **Syntactic Ambiguity**

   • Ambiguity occurs when a sentence is parsed in different ways.

   • "The man saw the girl with the telescope".

   • It is ambiguous whether the man saw the girl carrying a telescope or he saw her through his telescope.

3. Semantic Ambiguity

◦ semantic ambiguity happens when a sentence contains an ambiguous word or phrase.

◦ "The car hit the pole while it was moving" is having semantic ambiguity

◦ because the interpretations can be

◦ "The car, while moving, hit the pole"

◦ "The car hit the pole while the pole was moving".

4. Anaphoric Ambiguity

- This kind of ambiguity arises due to the use of anaphora entities in discourse.
- the use of a word referring back to a word used earlier in a text or conversation, to avoid repetition- pronouns
- The horse ran up the hill. It was very steep. It soon got tired. Here, the anaphoric reference of "it" in two situations cause ambiguity.

5. Pragmatic ambiguity

- Such kind of ambiguity refers to the situation where the context of a phrase gives it multiple interpretations.
- The sentence "I like you too" can have multiple interpretations like I like you (just like you like me), I like you (just like someone else does).
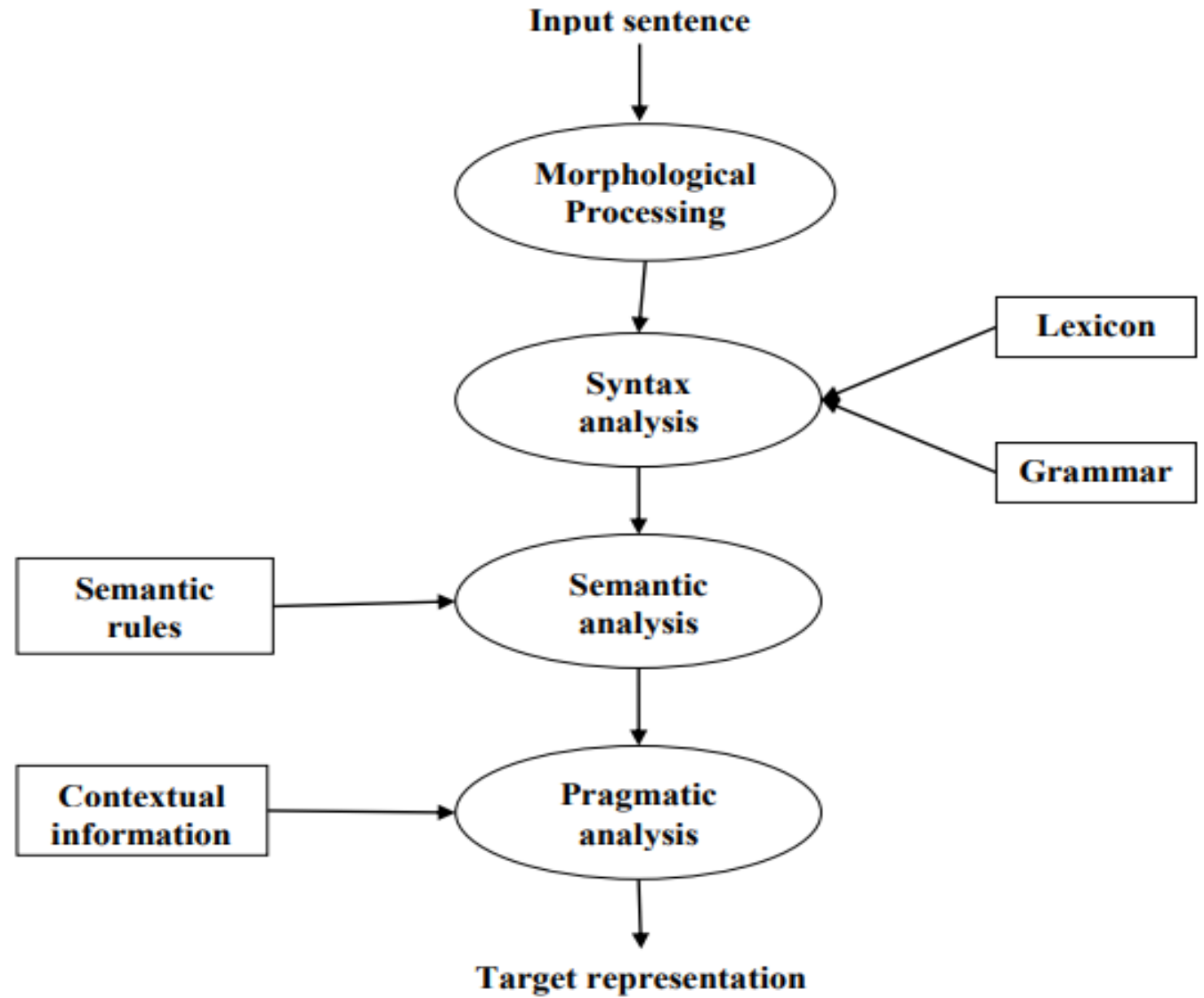
Other challenges

Contextual words

Slangs, colloquialism idioms,phrases

Synonyms

Irony, sarcasm, tone difference, spelling errors

NLP Phases

➢ **Morphological Processing**

 ➢ First phase of NLP.

 ➢ The purpose of this phase is to break chunks of language input into sets of tokens corresponding to paragraphs, sentences and words.

 ➢ Word like "uneasy" can be broken into two sub-word tokens as "un-easy".

➢ **Syntax Analysis**

 ➢ Second phase of NLP.

 ➢ The purpose of this phase is two folds:

  ➢ To check that a sentence is well formed or not.

  ➢ To break it up into a structure that shows the syntactic relationships between the different words

 ➢ The sentence like "The school goes to the boy" would be rejected by syntax analyzer or parser.

➢**Semantic Analysis**

   ➢Third phase of NLP.

   ➢The purpose of this phase is to draw exact meaning.

   ➢The text is checked for meaningfulness.

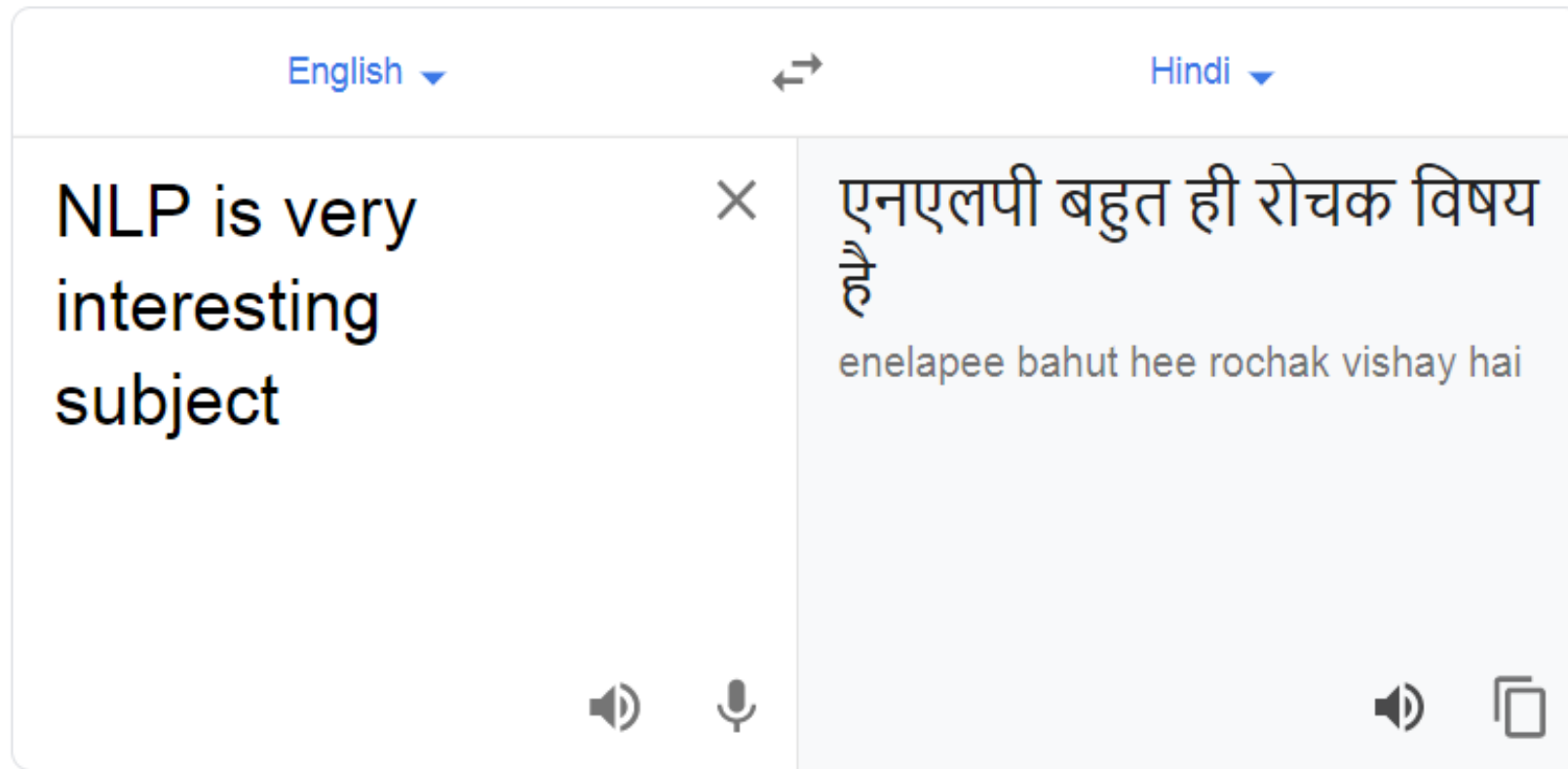   ➢Semantic analyzer would reject a sentence like "Hot ice-cream".

➢**Pragmatic Analysis**

   ➢Fourth phase of NLP.

   ➢Pragmatic analysis simply fits the actual objects/events, which exist in a given context with object references obtained during the last phase (semantic analysis).

   ➢The sentence "Put the banana in the basket on the shelf" can have two semantic interpretations and pragmatic analyzer will choose between these two possibilities.

Popular applications of NLP:

➢ **Machine Translation.**

➢ **Speech Recognition. …**

➢ **Sentiment Analysis. …**

➢ **Question Answering. …**

➢ **Automatic Summarization. …**

➢ **Chatbots. …**

➢ **Market Intelligence. …**

➢ **Text Classification.**

Machine Translation.

English ⇄ Hindi

NLP is very interesting subject

एनएलपी बहुत ही रोचक विषय है

enelapee bahut hee rochak vishay hai

Sentiment Analysis

Sentiment analysis is the interpretation and classification of emotions (positive, negative and neutral) within text data using text **analysis** techniques.

Sentiment analysis tools allow businesses to identify customer **sentiment** toward products, brands or services in online feedback.
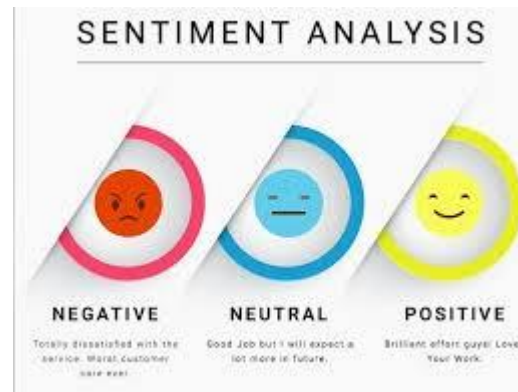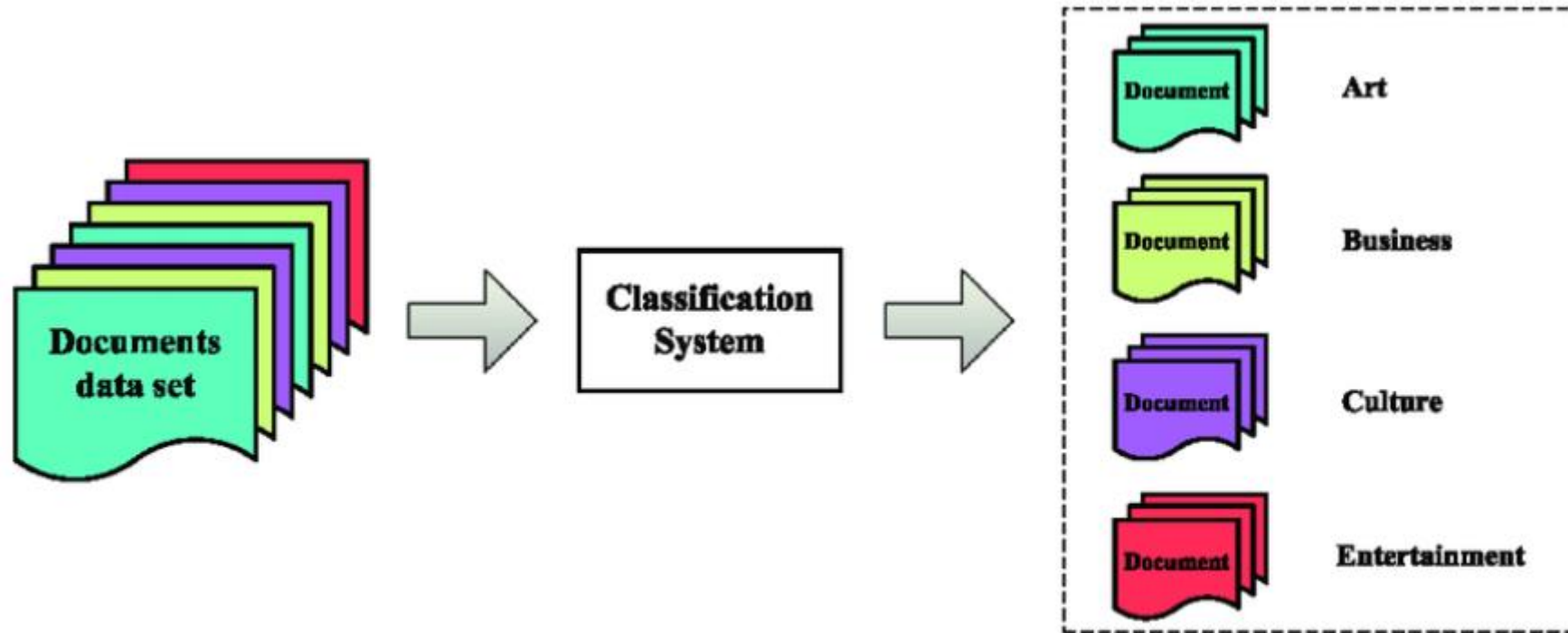
Question Answering

Question Answering is concerned with building systems that automatically answer questions posed by humans in a natural language.

Text Classification

A simple example algorithm framework for text categorization.

Named-entity recognition

(NER) locate and classify **named entities** in text into pre-defined categories such as the **names** of persons, organizations, locations, expressions of times, quantities.

"There was nothing about this storm that was as expected," said Jeff Masters, a meteorologist and founder of Weather Underground. "Irma could have been so much worse. If it had traveled 20 miles north of the coast of Cuba, you'd have been looking at a (Category) 5 instead of a (Category) 3."

| Person | Organization | Location |

Chatbot

A **chatbot** is a computer program that's designed to simulate human conversation.

Users communicate with these tools using a chat interface or via voice, just like they would converse with another person.

Examples Chatbot

Weather bot. Get the weather whenever you ask.

Grocery bot. Help me pick out and order groceries for the week.

News bot. Ask it to tell you when ever something interesting happens.

Life advice bot. I'll tell it my problems and it helps me think of solutions.

Personal finance bot. It helps me manage my money better.

Scheduling bot. Get me a meeting with someone on the Messenger team at Facebook.

A bot that's your friend. In China there is a bot called Xiaoice, built by Microsoft, that over 20 million people talk to.

Other Popular applications of NLP:

➢Speech Recognition. ...

➢Automatic Summarization. ...

➢Market Intelligence. ...

➢**Spam Detection**

➢**Topic Modelling**

➢**Spell checking /Grammar**