AI

Unit – 12

Natural Language Processing

By :- Dr. Maulik Dhamecha

What is NLP?

- Language is meant for communicating about the world.
- By studying language, we can come to understand more about the world.
- If we can succeed at building computational mode of language, we will have a powerful tool for communicating about the world.
- We look at how we can exploit knowledge about the world, in combination with linguistic facts, to build computational natural language systems.

What is NLP?

- Natural Language Processing (NLP) problem can be divided into two tasks:

1. Processing written text, using lexical, syntactic and semantic knowledge of the language as well as the required real world information.

2. Processing spoken language, using all the information needed above plus additional knowledge about phonology as well as enough added information to handle the further ambiguities that arise in speech.

Steps in NLP

1. Morphological Analysis
2. Syntactic Analysis
3. Semantic Analysis
4. Discourse integration
5. Pragmatic Analysis

Syntactic Processing

- Syntactic Processing is the step in which a flat input sentence is converted into a hierarchical structure that corresponds to the units of meaning in the sentence.

- This process is called parsing.

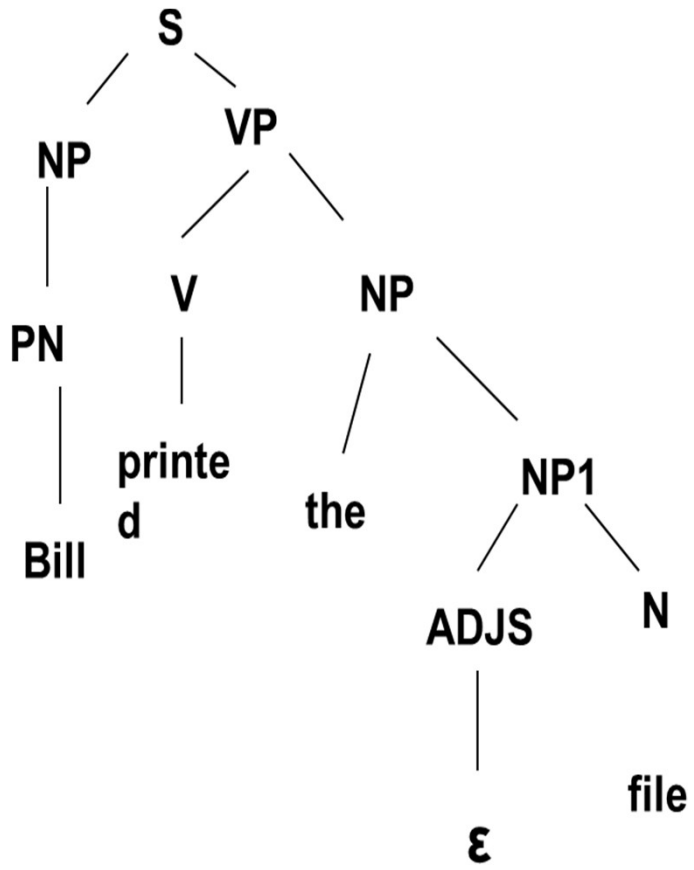
Syntactic Processing

- It plays an important role in natural language understanding systems for two reasons:

1. Semantic processing must operate on sentence constituents. If there is no syntactic parsing step, then the semantics system must decide on its own constituents. If parsing is done, on the other hand, it constrains the number of constituents that semantics can consider.
2. Syntactic parsing is computationally less expensive than is semantic processing. Thus it can play a significant role in reducing overall system complexity.

Example

A Parse tree for a sentence: Bill Printed the file



Augmented transition network (ATN)

- An augmented transition network is a top down parsing procedure that allows various kinds of knowledge to be incorporated into the parsing system so it can operate efficiently.
- ATNs build on the idea of using finite state machines (Markov model) to parse sentences.
- Instead of building an automaton for a particular sentence, a collection of transition graphs are built.
- A grammatically correct sentence is parsed by reaching a final state in any state graph.

Augmented transition network (ATN)

- Transitions between these graphs are simply subroutine calls from one state to any initial state on any graph in the network.
- A sentence is determined to be grammatically correct if a final state is reached by the last word in the sentence.
- The ATN is similar to a finite state machine in which the class of labels that can be attached to the arcs that define transition between states has been augmented.

Semantic Analysis

- Producing a syntactic parse of a sentence is only the first step toward understanding it.
- We must produce a representation of the meaning of the sentence.
- Because understanding is a mapping process, we must first define the language into which we are trying to map.
- There is no single definitive language in which all sentence meaning can be described.
- The choice of a target language for any particular natural language understanding program must depend on what is to be done with the meanings once they are constructed.

Semantic Analysis

- Choice of target language in semantic Analysis :
- There are two broad families of target languages that are used in NL systems, depending on the role that the natural language system is playing in a larger system:
- When natural language is being considered as a phenomenon on its own, as for example when one builds a program whose goal is to read text and then answer questions about it, a target language can be designed specifically to support language processing.
- When natural language is being used as an interface language to another program (such as a db query system or an expert system), then the target language must be legal input to that other program. Thus the design of the target language is driven by the backend program.

Discourse and Pragmatic processing

- To understand a single sentence, it is necessary to consider the discourse and pragmatic context in which the sentence was uttered.
- There are a number of important relationships that may hold between phrases and parts of their discourse contexts

Discourse and Pragmatic processing

- Identical entities. Consider the text:
  - Bill had a red balloon.
  - John wanted it.
  - The word “it” should be identified as referring to red balloon. These types of references are called anaphora.

Discourse and Pragmatic processing

- Parts of entities. Consider the text:
  - Sue opened the book she just bought.
  - The title page was torn.
  - The phrase “title page” should be recognized as part of the book that was just bought.

Discourse and Pragmatic processing

- Parts of actions. Consider the text:
  - John went on a business trip to New York.
  - He left on an early morning flight.
  - Taking a flight should be recognized as part of going on a trip.

Discourse and Pragmatic processing

- Entities involved in actions. Consider the text:
  - My house was broken into last week.
  - They took the TV and the stereo.
  - The pronoun “they” should be recognized as referring to the burglars who broke into the house.

Discourse and Pragmatic processing

- Elements of sets. Consider the text:
  - The decals we have in stock are stars, the moon, item and a flag.
  - I'll take two moons.
  - Moons mean moon decals.

Discourse and Pragmatic processing

- Names of individuals:
  - Dev went to the movies.

Discourse and Pragmatic processing

- Causal chains
  - There was a big snow storm yesterday.
  - The schools were closed today.

Discourse and Pragmatic processing

- Planning sequences:
  - Sally wanted a new car
  - She decided to get a job.

Discourse and Pragmatic processing

- Implicit presuppositions:
  - Did Joe fail CS101?

End of Unit - 12