SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 97

Accredited by NAAC-UGC with 'A' Grade

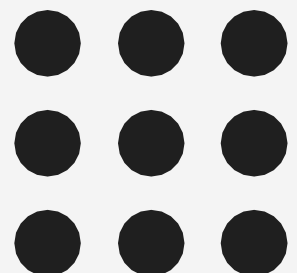**Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai**

Department of Artificial Intelligence and Data Science

Course Name – 19AD601 – Natural Language Processing

III Year / VI Semester

Unit 1 – Introduction

Topic 1- Origins and Challenges of NLP

NLP

Natural Language Processing

- Natural language processing, or NLP, is the field that involves getting systems to understand human languages.

- Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

- NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to 'understand' its full meaning, complete with the speaker or writer's intent and sentiment

Origins and Challenges of NLP

Origin and Brief History of NLP

- The field of natural language processing began in the 1940s, after World War II. At this time, people recognized the importance of translation from one language to another and hoped to create a machine that could do this sort of translation automatically. However, the task was obviously not as easy as people first imagined.

- By 1958, some researchers were identifying significant issues in the development of NLP. One of these researchers was Noam Chomsky, who found it troubling that models of language recognized sentences that were nonsense but grammatically correct as equally irrelevant as sentences that were nonsense and not grammatically correct.

- Chomsky found it problematic that the sentence "Colorless green ideas sleep furiously" was classified as improbable to the same extent that "Furiously sleep ideas green colorless"; any speaker of English can recognize the former as grammatically correct and the latter as incorrect, and Chomsky felt the same should be expected of machine models.

Origins and Challenges of NLP

- From 1983 to 1993, researchers became more united in focusing on empiricism and probabilistic models.

- Researchers were able to test certain arguments by Chomsky and others from the 1950s and 60s, discovering that many arguments that were convincing in text were not empirically accurate.

- Thus, by 1993, probabilistic and statistical methods of handling natural language processing were the most common types of models

Challenges of NLP

Challenges of NLP

NLP is a powerful tool with huge benefits, but there are still a number of Natural Language Processing limitations and problems:

- 1. Contextual words and phrases and homonyms
- 2. Synonyms
- 3. Irony and sarcasm
- 4. Ambiguity
- 5. Errors in text or speech
- 6. Colloquialisms and slang
- 7. Domain-specific language
- 8. Low-resource languages
- 9. Lack of research and development

Challenges of NLP

1.Contextual words and phrases and homonyms
The same words and phrases can have different meanings according the context of a sentence and many words

For example:
I ran to the store because we ran out of milk.
Can I run something past you real quick?

2. Synonyms
Synonyms can lead to issues similar to contextual understanding because we use many different words to express the same idea.

3. Irony and sarcasm
Irony and sarcasm present problems for machine learning models because they generally use words and phrases that, strictly by definition, may be positive or negative, but actually connote the opposite.

Challenges of NLP

4. Ambiguity

Ambiguity in NLP refers to sentences and phrases that potentially have two or more possible interpretations. NLP has the following types of ambiguities

5. Spelling Errors in text

Misspelled or misused words can create problems for text analysis. Spelling mistakes can occur for a variety of reasons, from typing errors to extra spaces between letters or missing letters.

Autocorrect and grammar correction applications can handle common mistakes, but don't always understand the writer's intention

6. Colloquialisms and slang

Informal phrases, expressions, idioms, and culture-specific lingo present a number of problems for NLP – especially for models intended for broad use.

Challenges of NLP

7. Domain-specific language

Different businesses and industries often use very different language. An NLP processing model needed for healthcare, for example, would be very different than one used to process legal documents. These days, however, there are a number of analysis tools trained for specific fields, but extremely niche industries may need to build or train their own models.

8. Low-resource languages

AI machine learning NLP applications have been largely built for the most common, widely used languages. However, many languages, especially those spoken by people with less access to technology often go overlooked and under processed. For example, by some estimations, (depending on language vs. dialect) there are over 3,000 languages in Africa, alone. There simply isn't very much data on many of these languages

9. Lack of research and development

Machine learning requires A LOT of data to function to its outer limits – billions of pieces of training data. The more data NLP models are trained on, the smarter they become.

THANK YOU