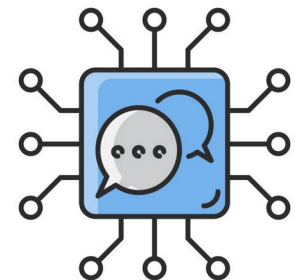


NLP Applications

Tushar B. Kute,
<http://tusharkute.com>



Lexicon

- A lexicon is the vocabulary of a language or branch of knowledge (such as nautical or medical). In linguistics, a lexicon is a language's inventory of lexemes.
- The word lexicon derives from Greek word λεξικόν (lexikon), neuter of λεξικός (lexikos) meaning 'of or for words'.
- Linguistic theories generally regard human languages as consisting of two parts: a lexicon, essentially a catalogue of a language's words (its wordstock); and a grammar, a system of rules which allow for the combination of those words into meaningful sentences.

Dictionary

- A dictionary is a listing of lexemes from the lexicon of one or more specific languages, often arranged alphabetically (or by radical and stroke for ideographic languages), which may include information on definitions, usage, etymologies, pronunciations, translation, etc.
- It is a lexicographical reference that shows inter-relationships among the data

Thesaurus

- A thesaurus (plural thesauri or thesauruses) or synonym dictionary is a reference work for finding synonyms and sometimes antonyms of words.
- They are often used by writers to help find the best word to express an idea:
...to find the word, or words, by which [an] idea may be most fitly and aptly expressed

Transliteration

- Transliteration refers to the method of mapping from one system of writing to another based on phonetic similarity.
- With this tool, you type in Latin letters (e.g. a, b, c etc.), which are converted to characters that have similar pronunciation in the target language.
- For example, in Hindi transliteration, you can type in "namaste" to get "नमस्ते", which sounds like "namaste". A list of candidate transliterations may show for you to select.
- Note that “transliteration” is different from “translation”: the conversion is based on the pronunciation, not the meaning.

Spell Checker and Auto Correct

- Spell Checker is a sequence-to-sequence pipeline that detects and corrects spelling errors in your input text.
- It's based on Levenshtein Automaton for generating candidate corrections and a Neural Language Model for ranking corrections.
- You can download the pretrained pipeline that comes ready to use.

Connected Applications

- Grammar Checker
- Domain Classification
 - Machine Learning Techniques
- Language Identification
 - ML / DL Methods
- Auto suggest/ Auto complete
 - RNN Sequence Prediction

Machine Translation

- Machine translation, sometimes referred to by the abbreviation MT (not to be confused with computer-aided translation, machine-aided human translation or interactive translation), is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one language to another.

Neural Machine Translation

- Neural machine translation (NMT) is an approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.
- They require only a fraction of the memory needed by traditional statistical machine translation (SMT) models.
- Unlike conventional translation systems, all parts of the neural translation model are trained jointly (end-to-end) to maximize the translation performance.
- First large-scale NMT system was launched by Baidu in 2015. Next year Google launched an NMT system too, followed by others.

Information Extraction

- The process of **sifting through unstructured data and extracting vital information** into more editable and structured data forms is known as information extraction.
- Working with a large volume of text data is usually stressful and time-consuming.
- As a result, many businesses and organizations rely on Information Extraction techniques to use clever NLP algorithms to automate manual tasks.
- Information extraction can save time and money by reducing human effort and making the process less error-prone and efficient.

Information Extraction

- Deep Learning and NLP techniques like Named Entity Recognition may be used to extract information from text input.
- If we're starting from scratch, though, we should evaluate the sort of data we'll be dealing with, such as bills or medical records.
- Information Extraction System is used in a variety of NLP-based applications.
- For example, extracting summaries from vast collections of text like Wikipedia, conversational AI systems like chatbots, extracting stock market announcements from financial news, and so on.

Web Scrapping

- Web scraping is the process of **collecting and parsing raw data from the Web**, and the Python community has come up with some pretty powerful web scraping tools.
- The Internet hosts perhaps the greatest source of information—and misinformation—on the planet.
- Many disciplines, such as data science, business intelligence, and investigative reporting, can benefit enormously from collecting and analyzing data from websites.

Question Answering

- Conversational question answering (QA) requires the ability to correctly interpret a question in the context of previous conversation turns.
- We address the conversational QA task by decomposing it into question rewriting and question answering subtasks.

Speech Technologies

- Speech technology is a type of computing technology that enables an electronic device to recognize, analyze and understand spoken word or audio.
- This is typically done by inputting digital sound signals and matching its pattern against a library of stored patterns.
- Speech technology relies on the concepts of signal processing and machine learning.
- Signal processing is used to extract relevant information from speech, such as speaker characteristics, background noise and frequency

OCR / Handwriting Recognition

- Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example: from a television broadcast)

Chatbots

- A chatbot or chatterbot is a software application used to conduct an on-line chat conversation via text or text-to-speech, in lieu of providing direct contact with a live human agent.
- Designed to convincingly simulate the way a human would behave as a conversational partner, chatbot systems typically require continuous tuning and testing, and many in production remain unable to adequately converse, while none of them can pass the standard Turing test

Thank you

This presentation is created using LibreOffice Impress 7.4.1.2, can be used freely as per GNU General Public License



@mitu_skillologies



@mITuSkillologies



@mitu_group



@mitu-skillologies



@MITUSkillologies

kaggle

@mituskillologies

Web Resources

<https://mitu.co.in>

<http://tusharkute.com>



@mituskillologies

contact@mitu.co.in
tushar@tusharkute.com