AMITY UNIVERSITY

----------Uttar Pradesh----------

Term Paper on

NATURAL LANGUAGE PROCESSING, APPLICATIONS & IMPLEMENTATION

Submitted to

Amity Institute of Information Technology

In partial fulfilment of the requirements for the award of the degree of

Bachelor of Computer Applications

Submitted To

DR. SAPNA SINHA
ASSISSTANT DIRECTOR (academics)
AMITY INSTITUTE OF INFORMATION TECHNOLOGY
AUUP, NOIDA

Submitted By

ADNAN QUDSI
ENROLMENT NO: A1004819171
BATCH: 2019-22

TABLE OF CONTENTS

DECLARATION BY STUDENT

I **ADNAN QUDSI** student of BCA 3C hereby declare that the report titled "**NATURAL LANGUAGE PROCESSING, APPLICATIONS AND IMPLEMENTATION** " which is submitted by me to **DR. SAPNA SINHA**, Amity Institute Of Information Technology, Amity University Uttar Pradesh, Noida, in partial fulfilment of requirement for the award of degree of **BACHELORS OF COMPUTER APPLICATIONS**, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

The author attests that permission has been obtained for the use of any copy righted material appearing in the dissertation/ project report other than brief excerpts requiring only proper acknowledgement in scholarly writing and all such use is acknowledged.

Place:- Noida

(student signature)

ADNAN QUDSI
Enrolment no: A1004819171
Semester: 3

GUIDE CERTIFICATE

I hereby certify that the project report by **ADNAN QUDSI**, student of **BCA 3C**, (enrolment no: **A1004819171**) with title "**NATURAL LANGUAGE PROCESSING, APPLICATIONS & IMPLEMENTATION**" which is submitted to **Amity Institute Of Information Technology**, **Amity University Uttar Pradesh**, Noida in partial fulfilment of requirement for the award of the degree of **BACHELORS OF COMPUTER APPLICATIONS** is an original contribution with existing knowledge and faithful record of work carries out by him/her under my guidance and supervision and to the best of my knowledge this work has not been submitted in part or full for any degree or diploma to this university or elsewhere.

Noida

Date:

(signature)
Dr. Sapna Sinha
Assistant Director (Academics)
Amity Institute Of Information Technology
AUUP, Noida

ACKNOWLEDEGEMENT

It is privilege for me to express my deep sense of gratitude to those entire faculty Members who helped me in the completion of the project, specially my internal guide **Dr. Sapna Sinha** who was always there at hour of need.
My special thanks to all other faculty members, Batch mate & Seniors of Amity Institute of Information Technology, Amity University, Uttar Pradesh for helping me in the completion of project work and its report submission.

ADNAN QUDSI

A1004819171

OVERVIEW TO NLP

Anything and everything we express either verbally or it could be in writing carries a huge amount of information . the topic we select, our pitch & tone, selection of words, everything adds up to some type of information that can be interpreted and meaningful value extracted out from it. In theory, we can understand and even predict human behaviour using this information .

But there is a problem : one person may generate hundreds or thousands of words in a declaration, each sentence with its corresponding complexity . if you want to scale out and analyse several hundreds, thousands or even millions of people or declarations in a geographical area, then the situation is unmanageable.

Data generated from conversations , declarations, posts, WhatsApp or even tweets are examples of unstructured data. Unstructured data don't go well with already existing row and column structures of manageable relational databases, and nevertheless represent the vast majority of bulk data available in the actual world .

So coming to the 21st century according to the industry estimate only 21% of the available data is in structured form while the unstructured occupying a whopping 79% of all (adrian bridgwater, 2010). it is definitely hard and messy to manipulate such data and definitely not at all easy to teach machine how to understand it.

Nevertheless, thanks to advancements in technological disciplines like machine learning, a big revolution is going on a very large scale regarding this topic, nowadays it is no longer about elucidating a text or speech based on its paraphrased words , but about understanding the hidden meaning behind those words (the cognitive way). This way we could easily speculate figures of speech like irony, or it's even possible to perform sentimental analysis.

so , a question comes up;

(Q. what is the use of structured data

Ans. Structured data helps search engines in understanding the searched material better and more efficiently , it helps them analyse accurately what the user wants.

there are further more uses of structured data.)

now, as we know majority of the data is in text format; process of deriving meaningful and relevant insights out of such unstructured data that too which exist in natural language text is known as TEXT MINING.

TEXT MINING is done through ordering/organising data in a structured format, deriving possible patterns out of it and finally evaluating and interpreting it to gather meaningful data.

And hence come up questions like how to achieve our goal and how to derive meaning out of our data which we just collected through above mentioned technique . or even the biggest one remains ; how come computers understand human language , have you ever wondered how your computer get what you want or what you need. that is the natural language, etc. etc.

Hence comes up our topic NATURAL LANGUAGE PROCESSING; although these two terms- NLP and text mining goes hand in hand and many other related topic could come up , my main focus would remain at NATURAL LANGUAGE PROCESSING although I'd try my best to cover up all the related stuff .

WHAT IS NLP?

NLP IS A FIELD OF COMPUTER SCIENCE & ARTIFICIAL INTELLIGENCE WHICH CATERS THE METHOD OF COMMUNICATING WITH AN INTELLIGENT SYSTEM USING NATURAL LANGUAGE; *HENCE GIVING MACHNES ABILITY TO READ, UNDERSTAND AND DERIVE MEANINGS FROM HUMAN LANGUAGE* (margarte rouse, n.d.).

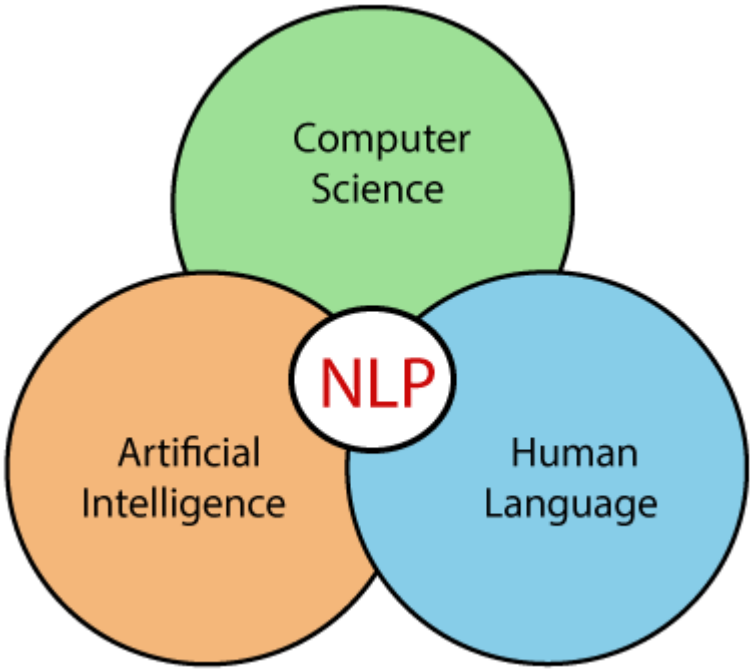


Figure 1: NATURAL LANGUAGE PROCESSING

well, you might have seen modern chatbots like SIRI, or GOOGLE ASSISTANT. They pretty much understand everything we say right?

They know you want to set an alarm or play music when you say so, they can set a reminder for you and will remind you when it's time. They can call anyone on you contacts when you say so, can search for you on google and find out location and can talk to you!! Well that's quite fascinating. A computer system being so smart that it can process and understand you via natural language!! Well that's what NATURAL LANGUAGE PROCESSING is. Making them understand the natural language.

APPLICATIONS OF NLP AND ITS NECESSITY

As we talk of applications of natural language processing it is used widely to get meaningful information from sources which could produce very less or even no meaning at all and is very essential so as to get that information into the computer and let the computer understand what it is.

Natural language processing techniques have been wide implemented in numerous work areas and fields or expertise some of them which are very popularly used in day to day life by common people are:

1. Sentimental analysis
   Also known as opinion mining or emotion AI is an interesting type of data mining technique its quite subjective , it just derives the conclusion out in a to the point manner; as in movie ratings , restaurant ratings or even posts on Facebook have many lucrative symbols to represent ones sentiments regarding something in a quite simpler way.
2. Speech recognition
   Is one of the most significant and prestigious achievement in natural language processing which has allowed modern systems to recognise the human speech directly ; unlike before humans can now just simply speak up and the computer would listen to them , understand them , and take its as an input respectively and accordingly , it is being used in google search or even games hence replacing keyboard . and many applications like Cortona, Siri and google assistant is using it to not only take inputs but to give appropriate output as in form of speech only.
3. Machine translation
   Unlike human we can use modern systems to translate content from one language to another , making it far more easy , time saving and sustainable. Basically what's done is computer understands what the content is then converts it into another language.

4. Spell checking
   It is a very minimal but extremely game changing application of NLP , it checks out for errors being done while writing a document and points out where ever an error occurs or smarter ones even auto correct the spellings .
5. Keyword search
   It is very commonly used application where human search something through search engines and computer just simply not search for what's written but try to understand what's the hidden meaning behind it and tries to make the search more efficient just to match the needs of the person searching and displaying appropriate content.
6. Information extraction
   This means extracting meaningful and structured data in reference to computer systems out of unstructured data present in form of either human language or machine code.
7. Advertisement matching
   In this language processing is used to drive out areas in which a particular person is interested in using search history of a person what he/she has been searching lately or studying lately and matching respectively advertisements that suits him/her.

8. Plagiarism check
   Plagiarism check is also a used case of NLP, what it does is simply traverse through the given text fie and check for the similar content on the web browsers, some smart software can even detect similar but no same words that are being copied.

There are bunch of many such cool application of NLP

But whilst we move forward , question may arise that why do we need NLP so, as just mentioned modern systems need to understand what human language means as in increasingly competent world NLP is used to cater to the need of industrialists and big business firms and many more area of work.

It makes computers more efficient and intelligent saves time and money and human interference .

And definitely as human nowadays generate and uses gallons and gallons of such huge amount of data , it creates an aura for computer systems to be intelligent enough to deal with it , hence there is a lot of advancement going on in this regard.

As in today NLP is widely being used in health care to digitally generate and receive data that is both perceivable by humans directly and to the machine also , many reports regarding underlying health diseases are directly being generated by the system using NLP and even big business groups use them to speculate their products sale , popularity and acceptance of their product in society using different techniques of NLP , big firms like yahoo and Gmail provide their email services, they filter and classify emails using NLP to analyse texts that flow through their servers and restricting spams to enter your inbox.

Even news firms use NLP techniques to technically determine whether the written- to be printed text is politically or regionally biased, so as to make things more reliable , transparent and trustworthy. And this is all advancement in computer systems and artificial intelligence.

And many big organisations have been trying to develop a sophisticated chatbot that could be really very exciting. Many have already been developed and many are in transition e.g. SWELLY, EBAY and many more.

Science is all about moving forward and advancing serving the needs of human and make the pace of competent life more easy.

And NLP plays a vital role in technological development ,yet remaining covert, NLP alongside advancement in HUMAN COMPUTER INTERACTION (of which NLP is a part itself) has played a an essential role in connecting the technology with humans and if it were not so, things could be pretty different today.

HOW IS NLP IMPLEMENTED

As we move on to how NLP is implemented and what steps does it follow to finally let computer understand the natural human language i.e. how the processing of natural human language takes place ; we should know that the language could get very complicated having multiple meaning for a single sentence and to process the real meaning out of it could get very disturbing even for humans , henceforth is very complicated for computer systems too.

E.g. "Environmental regulators grill business owners over illegal coal fires"

Now technically this line could be meant in 2 different ways. Although logically only one meaning is possible for a given statement, unless otherwise it's made purposely.

taking this in reference to the computer systems they are referred as AMBIGUITIES which can possibly makes things tougher for computers to understand and process for sure, so we need to sought them out.

AMBIGUITY are defined as the ability of sentences to have more than one meaning or being understood in more than one way

---

So dealing with ambiguities would be our very first step to implement NLP, there are 3 most common and sought of well formulated ambiguities in existence that comes up in English language

First being ;

1. LEXICAL AMIGUITY
   Lexical ambiguity is ambiguity of a single word, it is ambiguous with respect to its syntactic class.
   - e.g. she is looking for a match.
     (as in for cricket match or a groom?)
   - e.g. the fishermen went to the bank
     (as in the river bank or the bank dealing with finances?)

2. SYNTACTICAL AMBIGUITY
   syntactical ambiguity is also referred as structural ambiguity this in reference to the word formulation in such a way that it carries with it multiple meanings and could be interpreted in multiple ways whilst each being correct.
   - e.g. I saw a women with binoculars.
     (could be interpreted in more than one way!)
   - e.g. visiting relatives can be boring.
   - e.g. the chicken is ready to eat

now all these sentences carries ability to be interpreted in different ways and both could be correct for a human it is easy to know which one is real but for a computer it could get challenging,

3. REFERENTIAL AMBIGUITY
   Referential ambiguity is the ambiguity of a word which possess the quality of referencing , and when it get ambiguous – the word which refers to someone it gets difficult to speculate about whom is the sentence oriented for.
   - e.g. the boy told his father about the theft, he was very upset.
     (in this the word he is in reference to someone but as we don't know for whom it gets ambiguous referentially)

So , these are basic ambiguities that exist in the English language , there are few more but they all can be considered as a cumulated sum of these only. So for knowing how NLP is implemented this was a basic prerequisite moving on to how it is really planned and executed.

Moving forth as we know dealing with anything difficult and complex in AI or Machines usually means building up a pipeline(that is to formulate specific steps in accordance to which we'll proceed and would finally result in the final implementation of NLP. (they must be in synchronisation with the basic set of rules of the computational techniques we'd be using in each step.)

As we talk about the implementation of natural language processing; it's done in two phases, as in NLP is divided into two sub parts and then each part is executed in their own respective ways,

The subparts are:

1. **<u>natural language understanding;</u>**
   this section usually deals with the system processes taking the data items as an input , as the word suggest we deal with input data here , apply proper programming, make them run through are software and play them around our algorithms which in turn result in producing a summarised structured review of wholesome data that is perceivable to computer and thus , the system can understand it.
2. **<u>natural language generation</u>**
   while the natural language generation is the technology that would simply turn your data into a plain English language form making it perceivable to humans without interference of them. Just like any other human analyst would have done. Organisations generally look onto considering it, because its more secure and does not disclose information to others(as it's a computer.) (sciforce, 2019)

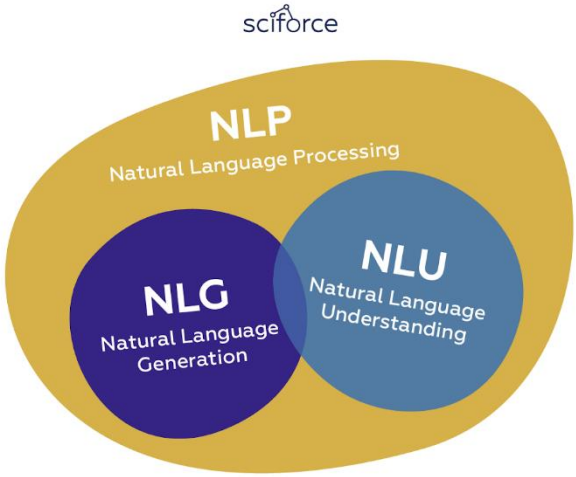


Figure 2:SUB-PARTS OF NLP

SKILLS/REQUISITES THAT ARE NEEDED TO IMPLEMENT NATURAL LANGUAGE PROCESSING IN MODERN SYSTEMS AND SOFTWARES.

- Deep understanding of text representation techniques.
- Knowledge of machine translations and compilers.
- Knowledge of python, java and R programming languages.
- Analytical mind with problem solving ability
- Skills to classify text and clustering.
- Knowledge of areas like designing algorithms and statistics.
- Machine learning framework and libraries; experience.
- Knowledge of CI/ CD pipelines.

As we talk of implementation of NLU that is a subpart of NLP; which would let computer understand the human language, there are multiple rules- statistical , logical and even theoretical. This field has proven to be a game changing technology and a heavy amount of researches are going on regarding optimising and simplifying the methods of NLP.

As mentioned earlier-natural language processing is an umbrella like term that explains holistically the procedure of turning unstructured data into structured form using multiple techniques.

As described previously, NLP is just a technique- so it used in many software programs that deals with data and intend to convert unstructured data input into structured one so that it could be used efficiently, but it’s not a full-fledged program in itself, as there is no software that could be named as NLP.

Although there are software that are specifically used for structuring data alongside providing statistical aids and analytics that are used in big business firms and trading companies.

Big it companies like GOOGLE, FACEBOOK, TWITTER, INSTAGRAM, LINKEDIN, etc. side by side many big websites and digital platforms through which a huge amount of data flow on regular basis uses this technique in their own respective software and servers in order to organise the data in structured format. So it could be used for analysing, storing information, recommending adds, or even answer your questions properly using the data they collected of you.

There are numerous techniques used as per conditions and cases, even many new one have come up and many more are in line, and to discuss them all would take a whole lot of experience, knowledge and time.

Hence, the ones I'll be discussing here are the most widely used and commonly collides with ideal cases; which are easily traceable and could be implemented in wide variety of computer programming languages.

TECHNIQUES USED TO IMPLEMENT NLP:

(Implemented step by step/ NLP PIPELINE)

1.SENTENCE SEGMENTATION

Refers to the technique of dismantling sentence from a paragraph giving them individual identity. So as for any given paragraph- after sentence segmentation, will have respective no of single sentences.

2.TOKENIZATION

Tokenisation is a well-defined and widely used technique that leads to the implementation of NLP, what it does is, it breaks down the sentences into small parcels/ pieces comprising of a single word. It's simply breaking down of text into its basic form WORDS. So simply it results into a list of words used in the text.

e.g. WE ARE IN TROUBLE WE MUST LEAVE!
So this techniques when applied would break this sentence into single word entities like:

Note: depending on precision & demands punctuations may or may not be omitted from getting tokenized.
Now tokenization has its own pros & cons but is widely used.
And is used after sentence segmentation is achieved.

3.STEMMING

stemming is a technique of slicing either the beginning or the ending of a word (referred to as token here*) with the intention of removing suffixes or prefixes so as to match it to its respective root word. So that it be identified and dealt with accordingly.
Affixes that are attached at a beginning of a word are called PREFIXES and the ones added at the last are called SUFFIXS
E.g. FUL in helpful is a suffix, while Astro in astrobiology is a prefix.

So , 

play

new 

the stemming would something be like this, mapping the words correctly to their root word whilst dealing carefully with the suffixes or prefixes that would change the meaning of entire word hence could result in processing the language wrongly.
Such tricky suffixes or prefixes are called as: inflectional affixes
So words like affecting, affects, affected, affection, affective all have a same root word and hence after stemming is applied to them all would be processed into data bases as AFFECT.
So for dealing with this stage of NLP there are multiple techniques and approaches and multiple algorithm, it depends upon programmer to programmer on how to do it(python and R

programming language have specific libraries which deals with such issues.)

Although there are drawbacks for this stemming process also as it supports algorithmic approach it may or may not deal with different strings in respective manners for each but would end up processing all of it similarly, which may end up in a completely different word.

Then Why to use them? Because they are fast, effective and easy to handle and dealing with a simple model of NLP, speed is must, hence stemming could be a way to go.

4. LEMMATISATION

Lemmatization is another very important key technique which aids NLP,

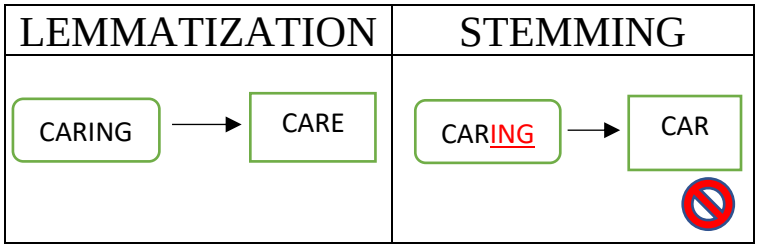
It is used to reduce a word to its base form, or in simple words it combines together different forms of the same word.

e.g. as in synonyms (best and good; would be both treated as 'good') & in (went and go; both be treated as go.)

so like this each word is standardized and referred to computers in its very basic form, using the pre-existing rules of English and dictionary.

Although it appears to be too similar to stemming, it isn't and uses a totally different approach.

Lemmatization dissolves the word into its dictionary unit (called lemma) for which systems requires detailed dictionaries in which they can look into and algorithmically link words to their lemmas (towardsdatascience.com, 2019).



Lemmatization can also look into the word's context in order to remove problems of disambiguation. But only with the help of a tag that is part of speech, definitely while applying lemmatization if a word is not ambiguous and we know whether it is a noun, verb, adverb or something else then it becomes comparatively easy to link it to the words lemma correctly, hence we can differ between the cricket's bat and the bat & between the different types of banks. Making computer systems intelligent.

5. POS TAGS

(PARTS OF SPEECH TAGS)

As we just discussed the lemmatization works efficiently if a word is tagged with its corresponding part of speech. This technique of POS tagging deals with the same.

In this technique each word is compared, with its surrounding words and is mapped with multiple sets of rules and pre-defined grammar which finally tag the part of speech.

We should know that computer is just a machine and what it does is processing each word just in a manner it is told using statistics and each step and procedures are defined by humans, so computers just process it but can't understand it in a manner we humans do.

After the POS are tagged properly and lemmatization has been done, computer can somehow starts to understand what a sentence is oriented for or talking about.

As in, the system now knows that LONDON & CAPITAL are nouns in the sentence and scanning through statistical cases, it can tell that sentence is talking about London.

POS TAGGING is used as a statistical NLP task it helps distinguishing the sense of the word which is very helpful in text realisation.

LONDON	IS	THE	CAPITAL	AND	MOST	POPULOUS	CITY...
(proper noun)	(verb)	(determiner)	(noun)	(conjunction)	(adverb)	(adjective)	

6. DEPENDENCY PARSING

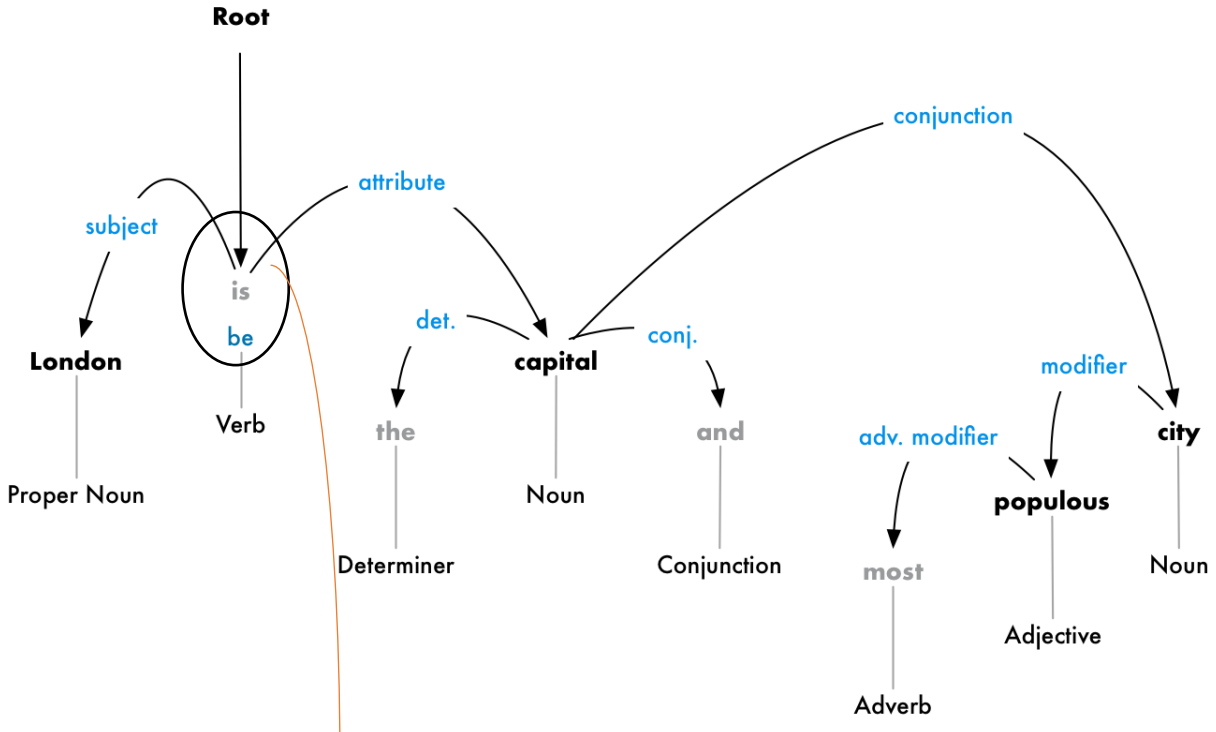
Is the technique to determine how each word of a sentence is related to each other.

It is achieved by assigning a root word to each word of a sentence in a tree structure and the root word of the tree would be the main verb of the sentence; which is again attained thorough logics and processes developed by humans using statistics and the rules of language's grammar.

We could go even a step further be not only making a parse tree but deriving the relationship between two words

e.g. so in the figure below parse tree shows that our subject is noun, "London" and had got a relationship 'be' with the 'capital'. Hence a productive information is extracted through an unstructured text data. That is- LONDON IS A CAPITAL!

We can go on deriving relationships like this further in this sentence, and this how meanings are extracted by computers. Not all together but in phases.



(this is achieved through lemmatization)

This technique has widely been recognised and many organisations and people have already developed many algorithms and libraries for programming in this area.
Like Google's PARSEY MCPARSFACE and many more.
A lot of research is already going on in regard to this topic for optimising the dependency parsing technique.

7. NAME ENTITY RECOGNITION

as we have done all the school level grammar exercise, we can now go beyond that to extracting ideas.
In name entity we determine/ recognise through names that is for sure the noun, what the noun is. That is we determine whether the noun is name of human, entity, organisation, place or anything, it's just recognising the noun.
We adopt statistical approach and algorithmically scan the rules of English grammar and the data previously used alongside dictionary and context clues to recognise the nouns present in the string and implant name entity recognition.

e.g. SUNDAR PICHAI INTRODUCED PIXEL 3 IN NYC AT GOOGLE'S HEADQUARTERS.
So after named entity recognition is achieved SUNDAR PICHAI would be tagged as PERSON, PIXEL 3 as an ENTITY/PRODUCT, NYC as a CITY and GOOGLE as an ORGANISATION.

8. COREFERENCE RESOLUTION

By this time we have already analysed our sentence, we know relationship between words, parts of speech , & we know names and entities and places.
However we still are left with a problem and that is our language is full of pronouns and what we have achieved up till now through scrutinising the sentence, surfing through data implementing algorithms and applying grammar rules would fail

because the words for which we did this all, may or may not be referred as something else in following sentences of the paragraph.
Although humans can keep track of for what word we are using what pronoun. Our NLP models are meant to examine only one sentence at a time.
So continuing our sentence which we used to explain dependency parsing i.e. "LONDON IS THE CAPITAL AND MOST POPULOUS CITY OF ENGLAND. IT WAS FOUNDED BY ROMANS, WHO NAMED IT LONDINIUM"

If we used are dependency parsing technique here. We could et our results as "it was founded by romans" but what we wanted is to let our system know that it was LONDON that was founded by romans and that, it is LONDON which is referred as it here.

Coreference resolution does the same for computers, it is responsible for keeping track and linking the pronouns to its former noun token. It figures out all the words that are referring to the same entity.

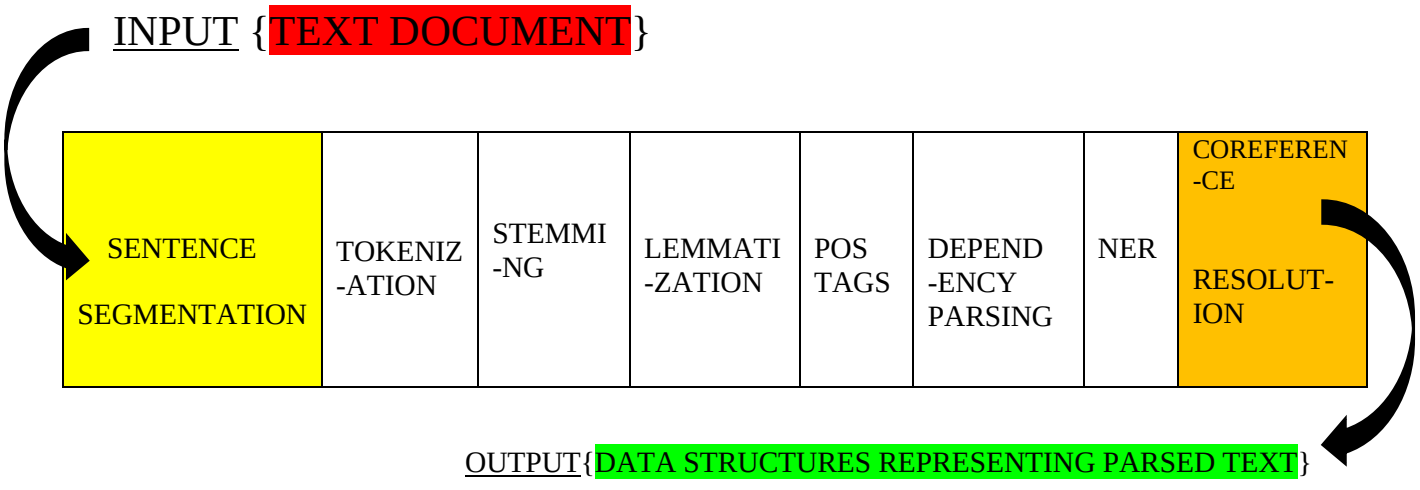
Coreference resolution is one the most difficult techniques to implement but is worthy enough. It requires a lot of expertise and wide knowledge of stats and big data and data science stuff to implement this.

A lot of work is already going on in this field and many advances od deep learning often collides with this particular technique. But none of it is perfect yet.

NLP PIPELINE

Now are least step remains to finally make up a pipeline which would guide our way to NLP implementation using these techniques step by step.

Our NLP pipeline would look something like this:



Now, that's all we have got to do, CODE IT ALL! these are the steps of a typical NLP pipeline. Now if you are an expert or your program or don't need to have it all, you can alter some steps out or even add some in according to your needs.

As we move forward we now are left with other subpart of NLP that is NLG (NATURAL LANGUAGE GENERATION)

As we know idea behind NLG is just reverse to that of NLP, as the latter one used to deal with unstructured data providing us with structured information. NLG will deal with the that structured data that is perceivable to computers and use techniques to frame it in natural language which is understandable to humans.

Some of the **MAJOR APPLICATIONS OF NLG** are;

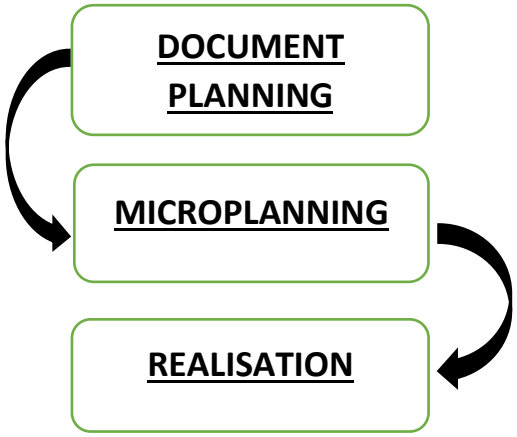
writing of data-driven:

1. financial reports.

2. Product description
3. Meeting memos
4. Major application remains writing of analysis and communicating necessary information to customers digitally. **These applications of NLG serves practically in wide area of professions;** such as
    - Analysis for business intelligence dashboards.
    - Reporting on business data/ data analysis.
    - IOT device status and maintenance reporting.
    - Individual client financial portfolio summaries and updates.
    - Personalised customer communication.

There are three stages of NLG process that leads towards its implementation.

TYPICAL NLG PIPELINE:



DOCUMENT PLANNING deals with content planning and document outline. While MICRO PLANNING deals with referring expressions, word choice and aggregation and REALISATION part takes into account converting specifications to real text.

There are multiple approaches to implement this pipeline, again you can construct one of your own and there are many others in process already.

Some of the practical and widely used algorithmic techniques which have their libraries available in several programming languages are:

1. Markov chain
   It is one of the first algorithms to achieve language generation. This algorithm predicts the possibility of next word in a sentence using current word and speculating unique relationship between each words and thereby predicting next word using probability.
   This technique is used in smartphones to predicts next word in the keyboards.
2. Recurrent neural network
   Is a type of algorithm that tries t mimic human brain, it basic idea is that the output of the previous process is fed back again as input to the current process. It is helpful in generating/ predicting the next word in the sentence as it remembers what was said last.
   It is one of the coolest yet hard to implement algorithms.
3. LSTM
   LSTM stands for long short term memory, it is an alternative to RNNs to address the problem of long-range dependencies. It consists of four parts- unit, input door, output door and the forgotten door. These allow the systems to forget remember or to forget words at any time by adjusting information flow of the unit door. They all adjust themselves according to the encounters to the periods and may or ay not drop current unit state information considering the idea might change, so on and so forth. These processes may get much complicated and they definitely are driven by principle of applied statistics and the language syntactics and semantics.
   Memory of LSTM is limited to few hundred words only due to their complex sequential paths

4. Transformer

   This type of algorithm uses encoders and decoders to generate natural language, stack of encoders is used to process inputs of varying lengths and another set of decoders to generate sentences as outputs. In contrast to previously discussed algorithms transformer only performs small but recurrent constant no. of steps, whilst a different module working alongside that simulates the relation among all the words called- self attention mechanism. Some examples of transformer are: OPEN AI, GPT 2 language model, etc.

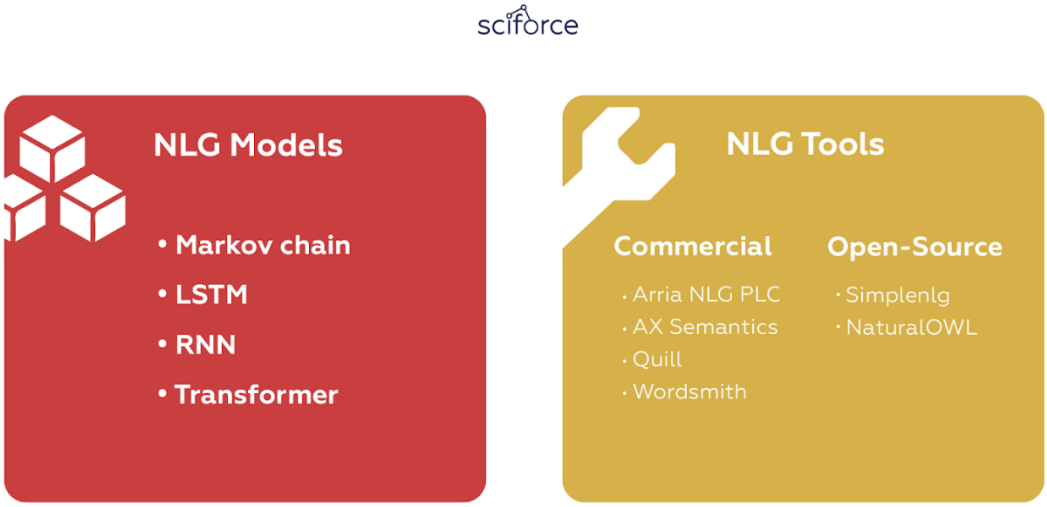


Figure 3

SOFTWARES/ APPLICATION USED TO IMPLEMENT NLP

Well talking about implementation of natural language processing specifically, we need to deal with a lot of data and analytics and deep learning and related stuff. So we need to select a programming language that is good in all of these parameters and should be fast enough and as we deal with natural language it would be great if we deal with a high level language. So the languages that come up ideally fulfilling these parameters are PYTHON, JAVA, MATLAB, OCATAVE, R ,etc.

These are some of the most used languages for implementation of natural language processing although we can use any language even C++ to achieve our goal.

Languages are chosen on the basis of expertise, interests or needs and that may differ from person to person and from programmer to programmer. Python and Java had been the most popular choice among the audience as these languages are very easy to handle with and these two languages even provide specific libraries that supports multiple techniques of natural language processing.

Some more tools/software that supports NLP implementation are:

Stanford NLP soft., apache open NLP, Scala NLP, snowball, etc.

Some very famous libraries of PYTHON & JAVA that enables above mentioned NLP(NLG/NLU) techniques are:

PYTHON	JAVA
NATURAL LANGUAGE TOOL KIT (NLTK)	CORE NLP
TEXT BLOB	OPEN NLP
CORE NLP	LUCENE
SPACY	GATE
POLYGLOT	MALLET
SCIKIT LEARN	REVERB
PATTERN	NLP4J

HOW DOES THESE LIBRARIES HELP IN IMPLEMENTATION OF NLP?

They delivers a helping hand to the programmer as they have some predefined keywords and some identifiers along with some fixed operations that are commonly used, already defined to themselves.

So a NLP programmer can easily apply techniques to their program because he now just need to call on functions like stemming or parsing, or coreference resolution instead of coding it all up.

Libraries may have some pathways already defines to surf through internet of data files so you don't need to program in that regard, that's already taken care of. Hence making programming way easier.

Imagine programming in c language where you even need to define what in int is or that when typing for a loop has to start and then defining loop again!

Using these techniques a company may get its NLP software ready through a company or can get it done personally so that they can cope with their business demands or BIG FIRMS can install NLP programs in their respective software to update their services.

HOW TO DIFFERENTIATE BETWEEN NLP AND RELATED FIELDS THAT COULD OFTEN BE CONFUSED WITH EACH OTHER.

Now after learning about all this people may question regarding the difference between BIG DATA and NLP so its simple as we know big data is a huge collection of data – structured or unstructured. And simply driving required data that is our target is TEXT MINING. Manipulating the mined data into structured format if required, that exists in natural human language or is to be presented in natural language is NATURAL LANGUAGE PROCESSING, and when we use all of these to feed information into computers so that they could gain experience through it and can respond to future problems with respect to similarity of the data already present is learning, called deep learning or MACHINE LEARNING. All of these makes

computers advance and intelligent as they are now capable of doing smart work and that's ARTIFICIAL INTELLIGENCE.

As far as data science is considered as it is also a method of extracting and deriving data, we must know that NLP is a sub field of COMPUTER SCIENCE & ARTIFICIAL INTELLIGENCE and that it used to achieve things like machine translation or keyword search anything that is related to language processing, now even data science or at some point of its implementation we might need processing of natural language and that's where NLP comes in. similarly talking of machine learning, it too have some used cases of NLP and may require it at various instances of its execution. Hence we must not confuse these terms with each other. Even machine learning often collides with data science as it may provide it with meaningful structures insights that may help in processing future data with reference to the information already present.

All these big terms like DATA SCIENCE, MACHINE LEARNING, ARTIFICIAL INTELLIGENCE, ETC. often run into each other for implementation of something bigger and anywhere at any stage if there is a need of natural language processing, it is just implemented using the above mentioned techniques and respective algorithms. It works as an aid to the computer systems.

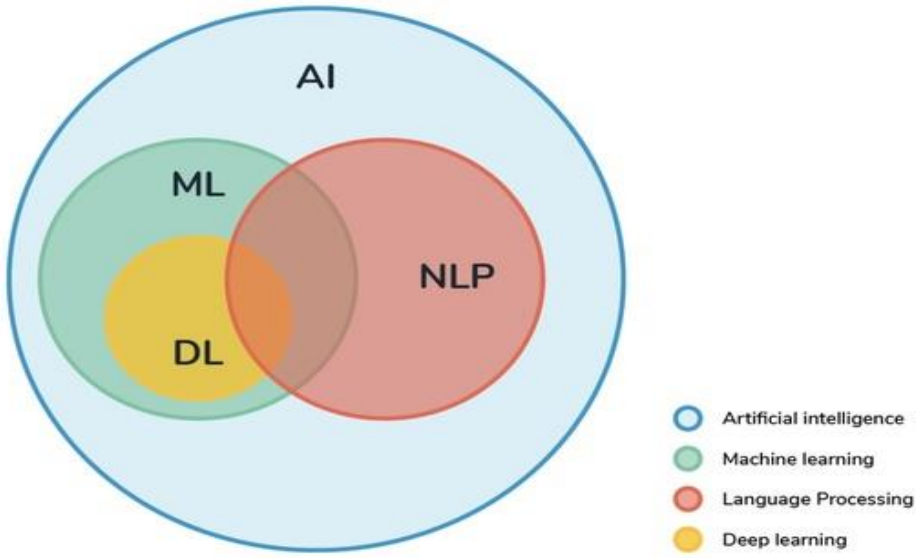


Figure 4: NLP & RELATED FIELDS

TABLE SHOWING BRIEF HISTORY OF EVOLUTION OF NLP

SOFTWARE/ TECHNOLOGY	YEAR	CREATOR	DESCRIPTION
GEROGETOWN EXPERIMENT	1954	GEROGETOWN UNIVERSITY	FULLY AUTOMATIC TRANSLATION OF 60 RUSSIAN SENTENCE TO ENG.
STUDENT	1964	DANIEL BOBROW	COULD SOLVE HIGH SCHOOL ALGEBRA WORD PROBLEMS
ELIZA	1964	JOSEPH WEIZENBAUM	A SIMULATION OF REPHRASING RESPONSES WITH GRAMMER RULES
SGRDLU	1970	TERRY WINNOGARD	NATURAL LANGUAGE SYSTEM WORKINGWITH RESTRICTED VOCABULARY.
PARRY	1972	KENNETH COLBY	A CHATTERBOT
KL-ONE	1974	SONDHEMIER ET AL.	KNOWLEDGE REPRESENTATION SYSTEM IN A TRADITION OF SEMANTIC NETWORKS AND FRAMES.
MARGIE	1975	ROGER SCHANK	--
TALESPIN SOFT.	1976	MEEHAN	--
QUALM	-	LEHNERT	
LIFER/ LADDER	1978	HENDRIX	A NATURAL LANGUAGE INTERFACE TO A DATABASE OF INFORMATION ABOUT US NAVY
SAM SOFTWARE	1978	CULLINGFORD	--
PAM SOFTWARE	1978	ROBERT WILENSKY	--
POLITICS SOFT.	1979	CARBONELL	--
PLOT UNITS	1981	LEHNERT	--

JABBERWACKY	1982	ROLLO CARPENTER	CHATTERBOT TO SIMULATE NATURAL LANGUAGE IN A FUNNY MANNER
MUMBLE SOFT.	1982	MCDONALD	--
RACTER	1983	WILLIAM CHAMBERLAIN & THOMAS ETTER	CHATTERBOT THAT GENERATED ENGLISH LANGUAGE PROSE AT RANDOM
MOPTRANS	1984	LYTINEN	--
KODIAK SOFT.	1986	WILENSKY	--
ABSITY SOFT.	1987	HIRST	--
WATSON (AI SOFTWARE)	2006	IBM	A QUESTION ANSWERING SYSTEM THAT WON THE JEOPARDY CONTEST, DEFEATING HUMAN PLAYERS.
SIRI	2011	APPLE	A VIRTUAL ASSISTANT DEVELOPED BY APPLE
AMAZON ALEXA	2014	AMAZON	VIRTUAL ASSISTANT
GOOGLE ASSISTANT	2016	GOOGLE	VIRTUAL ASSISTANT

From the very start humans have been trying to make computers smart enough to process the natural language and what they have achieved till now, they have achieved in phases. Someone proposed the idea another one provided the pipeline someone designed the algorithm and someone implemented it, from understanding to translations NLP has grown a lot and now modern systems are capable enough to interact like any other human beings and that too way smartly than humans as they are aided with internet and searching facilities.

Above mentioned table depicts the timeline of how NLP has been growing and at what stage does it stand today.

HOW IS NLP GROWING AND ITS FUTURE PROSPECTS

As we have learnt a lot about NLP, from what it is? to what are its applications and where and how they are used!, how its finally implemented and achieved , its techniques and algorithms. We have seen NLP's history and what all work has been done. We might think of what is the future of NLP? How is it growing and whether it's worth it to get into it or not.

As we know NLP enables applications to be intuitive and intelligent and almost all of use these every day and has been transforming the ways in which humans communicate to computers. They are driven by semantic and cognitive technologies. Future of NLP will be able to capitalize on its potential of understanding of text of speech similar to that of humans via numerous applications.

As we have seen that NLP is already playing such a crucial role in defining a computer as a smart entity, NLP has got lot more to offer from making machine learning more efficient to definitely making artificial intelligence more agile.

NLP will not only cater to business needs in future by providing companies a lot more effective and consumer oriented product details and requirements but also make systems extremely comfortable with human language so as to process even ironies, sarcasm , etc.

Around 50 years ago humans began the earliest attempts to analyse human language using computational approaches, but the idea has become more relevant in recent times due to the increase of human interaction with computers and digital systems hence producing huge amount of data for producing alongside the enhancement in computational power of modern systems so to deal with such enormous amount of data.

NLP is the voice behind famous SIRI, ALEXA similarly customer service chatbots harness the force of NLP to drive customised responses for e-commerce, healthcare and businesses. Some of the

universal applications of NLP sentimental analysis, customer service, machine translation and keyword search.

As the technology evolves, the future applications of NLP are going to be more user friendly and close to the need of users. As in NLP systems will be able to answer lot more difficult questions than before and would imply implications also.

NLP systems are not only limited to e-commerce and health care assistance but are also used to provide technical aid as if a system is smart enough it can respond to not only your question ”what is the problem with my network?” accurately but moreover can also provide you with ways to resolve the issue. Similarly in the fields of science and research modern systems can be trained with past research work and scientific information alongside multiple approaches and different theories making them an expert in science, so during a research instead of a human making points and paraphrasing everything, computers can come to their help, by framing sentences and generating reports in required format furthermore by memorising the research work and suggesting and raising possible questions and ways that could lead the research work, NLP is the future of robotics as this field moves forward to make robots more intelligent just like human they need to process language efficiently as in case of famous human robot SOPHIA.

In the coming time NLP will be integrated with other technologies like facial recognition and could deal with gestures making systems more agile.

Keeping all this in mind and the amount of data that is being generated industry leaders have predicted that the demand of NLP experts have been and will increase a lot in coming future.

Alongside the often discussed most popular and growing fields of computer science like artificial intelligence, machine learning, deep learning & data science, NLP provides aid to all when it comes to deal with natural language. Hence future belongs to NLP!

CHALLENGES OF NLP AND RESEARCH WORKS

as far as researches are concerned there are multiple things that are being dealt with there's a lot of work going on in optimising the pre-existing techniques of implementation of natural language processing and its algorithms, new compact, easy and optimal algorithms are being designed to implement different techniques like dependency parsing & named entity recognition so as to give computers a strong idea of what they are dealing with.

A lot of emphases is put on neural networks and disambiguation so as to simplify the sentences and speculate concrete results. Other research areas include hamming problems and plagiarism detection, and information retrieval.

As we talk about hot topics of research and challenges that NLP faces, most popular and sorted out areas remain MACHINE LEARNING, DEEP LEARNING & TEXT MINING, whatever we do in whatever area we like being it of NLP it is directly or indirectly related to AI AND ML. whether it be parsing or entity recognition we are here to make computer systems smart and that could not only be achieved by simply implementation but what we need is to make these systems learn from previous information and past experiences so to automatically deal with such processes intelligently.

As we have learned NLP is a subpart of machine learning and artificial intelligence and that both of these go hand in hand with each other. So we need to make our NLP modules such that they are compatible enough with the advances and needs of enhanced AI & ML modules. as the machine learning and artificial intelligence is getting advanced so do the need of NLP. We need to make our systems fast enough and smart enough to just simply know what the user has typed in, so as to make our keyword searches accurate and our machine translation perfect and text mining be so accurate that it boosts up the business and everything should synchronise in such a way that it is not only confined into books or researches anymore but be so simple that it benefits all the sectors of life from science to

healthcare, from business, journalism, e-governance, politics, marketing and robotics.

Some famous work done in the field of NLP are, Stanford universities' chatbot named Woebot, which claimed to help people suffering with depression and anxiety acting as a therapist, moreover Microsoft and google both have tried their attempts to diagnose about the underlying diseases like pancreatic cancer and influenza respectively which failed though. On march 2016 twitter released a chatbot on its website with the idea that interaction with more and more people will enhance the bots performance, but instead bot got trained badly and was removed due to its unacceptable behaviour. Microsoft attempted it's another experiment with the launch of another chatbot named ZO, it converse intelligently and is capable of remembering information related to one single conversation (academia, n.d.).

Although the future looks very challenging and demanding but the work is going on at a very large pace and will change a lot and bring in lot more technologies with it.

And with such constant advances let the humanity change its future, it is science only that never stops and so the needs of human with one step forward there comes more demand and needs. The coming era is wholly dependent on computer science and NLP is going to play an extremely vital role in it. It is the future of ours. Future of robotics, future of marketing, future of healthcare, and future of scientific researches, future of arts and commerce.

IT IS THE FUTURE OF THE WORLD.

SKILL SETS

Moving forth when we discuss about skills and knowledge one should possess to implement NLP or to be an expert, following are the most crucial and demanding skill one should have:

- First of all as we discussed one must have good mathematical, analytical and statistical (in particular) knowledge and expertise. As we deal with a lot of data that has to be statistically organised.

- Also one must be good at language he/she is dealing with so as to derive and design algorithms if needed be, so that he can formulate and comprehend ways out of text according to the needs.

- Moreover, one should have a good background or have at least studied, about the machine language, translations and compilers.

- Should possess good programming skills in at least one programming language preferably python, java or even R programming language.

- Should have a deep understanding of text representation techniques, and how sentences are designed that is grammar.

- And should have an analytical mind set with problem solving ability. And skills to classify text and clustering.

- Again one should have knowledge of designing algorithms and manipulating them according to needs and statistics definitely.

- One should be very good at machine learning frameworks and libraries, have experience in them. And should have knowledge about pipelines- CI/CD.
- Should have good software engineering skills and be able to quickly build prototypes if needed.

- Should have a good reading skills so that one could go through text way quickly if required and be able to scan and paraphrase and draw important figures out of it.

- Alongside statistics one should also have good knowledge about probability and possibly be able to derive permutation and combinations.

As we now know what it is required out of an NLP expert. And we have already discussed about what NLP is, where it is used and why is it necessary. Moreover we learned about how it is implanted finally and discussed about its subparts- NLG & NLU separately.

Furthermore we discussed popular techniques and algorithms used for implementation and saw their structural pipelines (approach), we had a brief look upon software tools and popular libraries and we dealt with differences with fields related to NLP, we had a quick look over NLP's timeline and history and talked about the future prospects of NLP and how it is growing.

At last we covered up topics like researches and work going on in regard to NLP that will shape its future and also the skills and knowledge under which NLP's implementation and the entire genre falls. With that all being set we now move onto concluding our topic.

CONCLUSION

NATURAL LANGUAGE PROCESSING is a method of processing human language so to make it understandable to computer systems, it is a subfield of computer science and artificial intelligence with provide optimised ways through which a human interacts to a computer using their natural language.

NLP is used to derive meanings out of unstructured text data item, which may or may not be collected through various techniques, one such is data mining, using this technique usable insights are collected from a pile of unstructured data & then through NLP techniques meaning are driven out if they exist in human language. NLP serves as a backbone to a lot of modern applications which deals with human language and organise them in a structured format providing a common platform which helps in both computers and humans in understanding each other. Some of these cool applications are: MACHINE TRANSLATION, KEYWORD SEARCH, SENTIMENTAL ANALYSIS, SPEECH RECOGNITION, SPELL CHECK, ETC.

Using all these applications and techniques NLP is serving a lot of famous applications like SIRI, GOOGLE ASSISTANT, etc. and many business firms use NLP tools to generate useful insights, it is used in health care and scientific research as well, it is serving in areas like journalism as well and provide help for advances in HUMAN COMPUTER INTERACTION.

NLP is divided into two sub parts- NLU & NLG that is natural language understanding and natural language generation respectively. Implementation of both of these collectively is referred to as implementation of NLP. They are implemented using statistical, analytical and probabilistically driven approaches, computers are trained with rules of language and possible orders and grammar, and a multiple of techniques are used according to needs and at last computers understands what the sentence means.

Both NLU and NLG have got there typical pipelines for implementation. Some of the popular sorted out techniques which are widely used are: sentence segmentation, tokenisation, stemming, lemmatisation, pos tags, dependency parsing, name entity recognition. And for NLG – markov chain, LSTM, RNN, transformers.

Programming languages that suits to the persona of NLP are R, java, python, matlab, etc. as they all are data and stats driven but most widely used ones are java and python. And some of the most famous libraries are; NLTK, SPACY, CORE NLP, NLP4J.

NLP has been into human's consideration form very long and human has been deriving ways out to achieve it, throughout many years, many advances has been noticed, earlier from simple translation of a single paragraph to now digitally not only convert the language but to play it out loud so that human can listen, NLP has grown a lot and has yet been growing.

NLP has a lot to offer and huge potent for business and scientific research sectors, it will not only make things faster, but easier, and more accurate and reliable, it increases security, decreases human utilisation and make things more audience targeted. NLP's usage has been growing and many big organisations are already investing in it.

NLP has become an integral part of artificial intelligence and machine learning as both of them faces human language and if they are more structured and machine understandable, it will add up to the advantages more over it is helping more and more people to connect to the technology.

A lot of research work is already going on in NLP, so to derive optimised ways to achieve the goal. A lot of advances are being done in already existing NLP model and used techniques and a lot of new ideas are coming up and being welcomed. It is the growing age of NLP and hence it does have a lot to offer to the industry experts.

At last we have the skills and knowledge that one needs to have to be an NLP expert and the areas of studies under which NLP would fall and they are: mathematics- statistics and probability, analytical thinking, basics of algorithms, expertise in at least one programming languages, good knowledge of the human language one would deal with, good software engineering skills, experience in machine language, translations and compilers, knowledge of CI/CD pipelines and text clustering. That is what NLP is with that all being set, you are ready to make your own chatbot!

References & Bibliography

adrian bridgwater. (2010, october 26). *IBM: 80 percent of our global data is unstructured (so what do we do?)*. Retrieved from computerweekly.com: https://www.computerweekly.com/blog/CW-Developer-Network/IBM-80-percent-of-our-global-data-is-unstructured-so-what-do-we-do

dataversity. (2018, december 6). *dataversity.net*. Retrieved from The Future of NLP in Data Science: https://www.dataversity.net/future-nlp-data-science/

Department of Information Technology, Kannur University, Kerala, India. (2014, october 5). *http://www.ijircce.com/.* Retrieved from http://www.ijircce.com/upload/2014/sacaim/59_Paper%2027.pdf

edureka! (2018). natural language processing & text mining [Recorded by edureka!]. india.

expertsystem.com. (2016, october 18). *The future of natural language processing*. Retrieved from expertsystem.com: https://expertsystem.com/future-natural-language-processing/

freelancermap. (2020, january 10). *www.freelancermap.com.* Retrieved from what-does-nlp-engineer-do: https://www.freelancermap.com/blog/what-does-nlp-engineer-do/

freelancermap.com. (2020, january 10). *what does nlp engineer do* . Retrieved from freelancermap.com: https://www.freelancermap.com/blog/what-does-nlp-engineer-do/

Madhavan, M. (n.d.). *Natural_Language_Generation_Scope_Applications_and_Approaches.* Retrieved from www.academia.edu: https://www.academia.edu/3879518/Natural_Language_Generation_Scope_Applications_and_Approaches

margarte rouse. (n.d.). *searchbusinessanalytics.techtarget.com*. Retrieved from natural-language-processing-NLP: https://searchbusinessanalytics.techtarget.com/definition/natural-language-processing-NLP

medium. (2017, march 24). *Ambiguity in Natural Language Processing*. Retrieved from medium.com: https://medium.com/@jrodthoughts/ambiguity-in-natural-language-processing-15f3e4706a9a

medium.com. (2018, july 10). *Natural Language Processing is Fun!* Retrieved from medium.com: https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e

medium.com. (2019, january 16). *What are some of the challenges we face in NLP today?* Retrieved from medium.com: https://medium.com/datadriveninvestor/what-are-some-of-the-challenges-we-face-in-nlp-today-2e9d94da1f63

Oleksii Kharkovyna. (2019, december 19). *Natural Language Processing (NLP): Top 10 Applications to Know*. Retrieved from towardsdatascience.com: https://towardsdatascience.com/natural-language-processing-nlp-top-10-applications-to-know-b2c80bd428cb

opensource.com. (2019, march 22). *12 open source tools for natural language processing*. Retrieved from opensource.com: https://opensource.com/article/19/3/natural-language-processing-tools

sciforce. (2019, july 4). *medium.com*. Retrieved from A Comprehensive Guide to Natural Language Generation: https://medium.com/sciforce/a-comprehensive-guide-to-natural-language-generation-dd63a4b6e548

som, k. (2018, october 5). *WHY THE FUTURE IS BRIGHT FOR NATURAL LANGUAGE PROCESSING*. Retrieved from analyticsinsight.net: https://www.analyticsinsight.net/why-the-future-is-bright-for-natural-language-processing/

towardsdatascience.com. (2018, january 30). *5 Heroic Tools for Natural Language Processing*. Retrieved from towardsdatascience.com: https://towardsdatascience.com/5-heroic-tools-for-natural-language-processing-7f3c1f8fc9f0

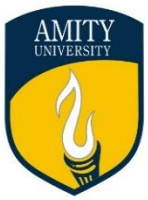
towardsdatascience.com. (2019, jan 16). *your-guide-to-natural-language-processing-nlp*. Retrieved from www.towardsdatascience.com: https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1

wikipedia. (2013, september 30). *wikipedia.org*. Retrieved from history of natural language processing: http://www.basic-english.org/projects/333/History%20of%20natural%20language%20processing%20-%20Wikipedia,%20the%20free%20encyclopedia.htm

wikipedia. (2020, may 11). *wikipedia.org*. Retrieved from Named-entity_recognition: https://en.wikipedia.org/wiki/Named-entity_recognition

www.kdnuggets.com. (n.d.). *named-entity-recognition-practitioners-guide-nlp*. Retrieved from www.kdnuggets.com: https://www.kdnuggets.com/2018/08/named-entity-recognition-practitioners-guide-nlp-4.html

www.linuxlinks.com. (2019, september 27). *8 Excellent Java Natural Language Processing Tools*. Retrieved from www.linuxlinks.com: https://www.linuxlinks.com/excellent-java-natural-language-processing-tools/



AMITY UNIVERSITY

UTTAR PRADESH

NTCC WEEKLY PROGRESS REPORT

TERM PAPER [ETTP100]

NATURAL LANGUAGE PROCESSING,APPLICATIONS AND IMPLEMENTATION

ADNAN QUDSI

A1004819171

(BCA-3C)

AMITY UNIVERSITY
UTTAR PRADESH

TERM PAPER; WEEKLY PROGRESS REPORT –1st

(FOR WEEK:18 APRIL- 25 APRIL, 2020)

ENROLLMENT NUMBER: A1004819171

PROGRAM: BCA

BATCH: 2019-22

STUDENT NAME: ADNAN QUDSI

FACULTY GUIDE'S NAME: DR. SAPNA SINHA

PROJECT TITLE:

NATURAL LANGUAGE PROCESSING; APPLICATIONS AND TECHNIQUES.

TARGET SET FOR THE WEEK

Learning about what NATURAL LANGUAGE PROCESSING is and why is it essential for modern computer systems and in what areas of life it is used. How is it beneficial

PROGRESS/ ACHIVEMENTS FOR THE WEEK :

- NLP is an applied technique of computer science and AI that enables modern computers to understand and process natural human language directly; it helps human communicate to computer systems using natural language.
- It is used in business, health care, journalism- as in keyword search, sentimental analysis, machine translation, speech recognition, spell checking, information extraction, etc.
- In this competent era computers need to be smart enough to process natural language efficiently and NLP caters to their demands using above mentioned techniques , and many more is in progress
- NLP helps in processing unstructured data alongside- TEXT MINING

FUTURE WORK PLANS

- I'll work on how Computer systems understand the NATURAL LANGUAGE using NLP techniques, and what all techniques are used for the same.
- what is the layout of the whole procedure, and what all things are done for the processing
- I'll figure out what all skills are taken into account for the implementation of NLP.

AMITY UNIVERSITY
UTTAR PRADESH

TERM PAPER; WEEKLY PROGRESS REPORT –2nd

(FOR WEEK:26 APRIL- 2 may, 2020)

ENROLLMENT NUMBER: A1004819171

PROGRAM: BCA

BATCH: 2019-22

STUDENT NAME: ADNAN QUDSI

FACULTY GUIDE'S NAME: DR. SAPNA SINHA

PROJECT TITLE:

NATURAL LANGUAGE PROCESSING; APPLICATIONS AND TECHNIQUES.

TARGET SET FOR THE WEEK

- I'll work on how Computer systems understand the NATURAL LANGUAGE using NLP techniques, and what all techniques are used for the same.
- what is the layout of the whole procedure, and what all things are done for the processing
- I'll figure out what all skills are taken into account for the implementation of NLP.

PROGRESS/ ACHIVEMENTS FOR THE WEEK :

I learned about and drafted my report accordingly for the following given topics.

- NLP PIPELINE, TECHNIQUES AND ALGORITHMS to implement nlp
- IMPLEMENTATION OF BOTH NLU AND NLG; THEY ARE SUBPARTS OF NLP
- PROGRAMMING LANGUAGES THAT SUPPORT IMPLEMENTATION OF NLP
- POPULAR LIBRARIES THAT DEAL WITH IMPLEMENTATION OF NLP
- SKILLS NEEDED TO BECOME AN NLP EXPERT

FUTURE WORK PLANS

- I'll work on brief history of NLP
- Future prospects of NLP
- And fields related to NLP that are often tangible to NLP, like what is the relationship between NLP and machine learning/data science/ AI/ human computer interaction. And how to differentiate between them.

AMITY UNIVERSITY
UTTAR PRADESH

TERM PAPER; WEEKLY PROGRESS REPORT –3rd

(FOR WEEK:3 may- 9 may, 2020)

ENROLLMENT NUMBER: A1004819171

PROGRAM: BCA

BATCH: 2019-22

STUDENT NAME: ADNAN QUDSI

FACULTY GUIDE'S NAME: DR. SAPNA SINHA

PROJECT TITLE:

NATURAL LANGUAGE PROCESSING; APPLICATIONS AND TECHNIQUES.

TARGET SET FOR THE WEEK

- I'll work on brief history of NLP
- Future prospects of NLP
- And fields related to NLP that are often tangible to NLP, like what is the relationship between NLP and machine learning/data science/ AI/ human computer interaction. And how to differentiate between them.

PROGRESS/ ACHIVEMENTS FOR THE WEEK :

I've successfully read and collected information regarding following topics

- history timeline of NLP
- fields that are related to NLP and are often confused with each other & their differences
- I have covered future prospects of NLP in regard to machine learning and AI.

FUTURE WORK PLANS

- Ill work on challenges and research works going on NLP
- Ill cover the skills and knowledge one should possess for implementation of NLP
- Ill see and speculate through my learnings and provide suggestions and ideas related to NLP (what should be there?/ what changes if any! ,etc, etc,)

AMITY UNIVERSITY
UTTAR PRADESH

TERM PAPER; WEEKLY PROGRESS REPORT –4th

(FOR WEEK:10 may- 16 may, 2020)

ENROLLMENT NUMBER: A1004819171

PROGRAM: BCA

BATCH: 2019-22

STUDENT NAME: ADNAN QUDSI

FACULTY GUIDE'S NAME: DR. SAPNA SINHA

PROJECT TITLE:

NATURAL LANGUAGE PROCESSING; APPLICATIONS AND TECHNIQUES.

TARGET SET FOR THE WEEK

- Ill work on challenges and research works going on NLP
- Ill cover the skills and knowledge one should possess for implementation of NLP
- Ill see and speculate through my learnings and provide suggestions and ideas related to NLP
- (what should be there?/ what changes if any! ,etc, etc,)

PROGRESS/ ACHIVEMENTS FOR THE WEEK :

I've successfully read and collected information regarding following topics

- Challenges and research going on NLP
- Future prospect of nlp
- Skills and knowledge needed for its implementation

FUTURE WORK PLANS

Complete- ill just draft, check & proofread my work.

NTCC TERM PAPER[ETTP100]

DAILY DIARY

NATURAL LANGUAGE PROCESSING

ADNAN QUDSI
(BCA) A1004819171

AMITY UNIVERSITY
UTTAR PRADESH

AMITY INSTITUTE OF INFORMATION TECHNOLOGY

DAILY DIARY

NAME: **ADNAN QUDSI** WPR OF WEEK: **1**
ENROLLMENT NUMBER:**A1004819171** WPR REMAINING: **3**
TOPIC: **NLP, APPLICATIONS & IMPLENTATION** PROGRAM: **BCA**

WEEK'S SUMMARY

DAYS	SUMMARY
Monday	Finding of the topic and got it confirmed from the guide Got it registered.
Tuesday	Making a layout of what topics to cover
Wednesday	Short listing areas to be covered
Thursday	Learning about what natural language processing is
Friday	Learning about its applications and necessity and where and how it is used
Saturday	Summarising and understanding of the concepts
Sunday	Finally writes down the report on covered topics

AMITY UNIVERSITY
UTTAR PRADESH

AMITY INSTITUTE OF INFORMATION TECHNOLOGY

DAILY DIARY

NAME: **ADNAN QUDSI** WPR OF WEEK: **2**
ENROLLMENT NUMBER:**A1004819171** WPR REMAINING: **2**
TOPIC: **NLP, APPLICATIONS & IMPLENTATION** PROGRAM: **BCA**

WEEK'S SUMMARY

DAYS	SUMMARY
Monday	Learning more about NLP, its subfield and related areas
Tuesday	Learned about the approaches and techniques used in implementation.
Wednesday	Learning about ambiguities in language and sub fields of NLP: NLG & NLU
Thursday	Learning about pre requisites for NLP implementation.
Friday	Learning about techniques used for implementation of NLP
Saturday	Learning about implementation of NLP and drafting the report.
Sunday	Writing down about the topics covered and preparing a draft.

AMITY UNIVERSITY
UTTAR PRADESH

AMITY INSTITUTE OF INFORMATION TECHNOLOGY

DAILY DIARY

NAME: **ADNAN QUDSI** WPR OF WEEK: **3**
ENROLLMENT NUMBER:**A1004819171** WPR REMAINING: **1**
TOPIC: **NLP, APPLICATIONS & IMPLENTATION** PROGRAM: **BCA**

WEEK'S SUMMARY

DAYS	SUMMARY
Monday	Learning about what a pipeline is and about typical NLP pipeline.
Tuesday	Drafting of a typical NLP pipeline based on learning
Wednesday	Describing and differentiating NLG and NLU. Studying about it and drafting of report.
Thursday	Learning about typical pipeline of NLG , learning about work done through NLG and NLU specifically and how they make up to be called NLP
Friday	Learning about the techniques used and approaches to adopt for implementing NLG.
Saturday	Learning about all these stuff and relating them to each other. Learning about how it all goes hand in hand
Sunday	Finally preparing a draft for topics covered up till now.

AMITY UNIVERSITY
UTTAR PRADESH

AMITY INSTITUTE OF INFORMATION TECHNOLOGY

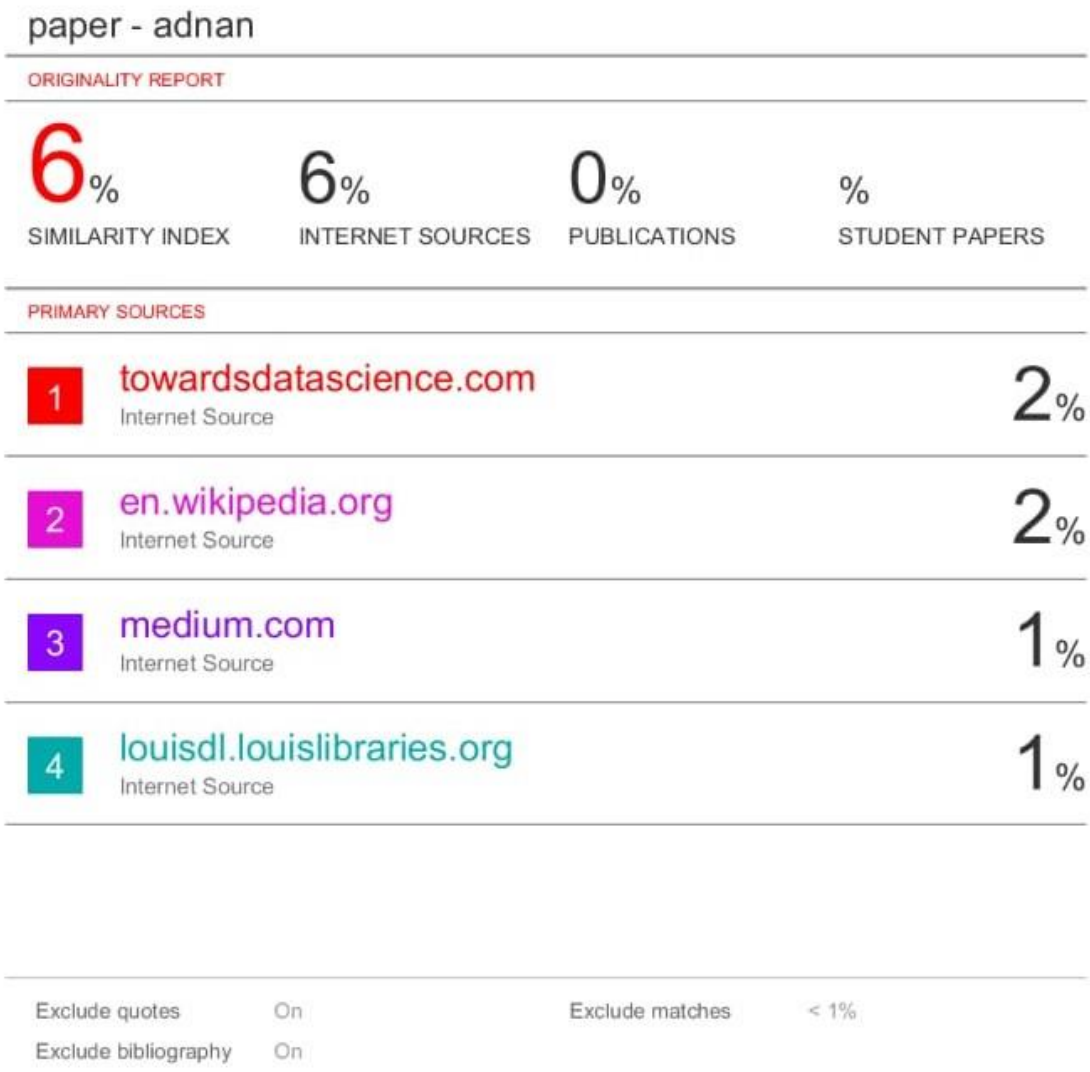
DAILY DIARY

NAME: **ADNAN QUDSI** WPR OF WEEK: **4**
ENROLLMENT NUMBER:**A1004819171** WPR REMAINING: **0**
TOPIC: **NLP, APPLICATIONS & IMPLENTATION** PROGRAM: **BCA**

WEEK'S SUMMARY

DAYS	SUMMARY
Monday	Learning about software and applications used to implement NLP, and all aid used in for the same.
Tuesday	Learning about NLP and related fields like AI, ML , data science ,etc that are often confused with each other.
Wednesday	Learning about importance of libraries in a programming languages and how they serve a program. Specifically NLP.
Thursday	Learning about history of NLP , and how its has been evolving and how it has changed over the years.
Friday	Learning about future prospects and challenges faced by NLP and how its growing. And skill sets and knowledge needed .
Saturday	Preparing a final draft of what all has been covered. And concluding everything.
Sunday	Editing of final draft and giving up final touch.

PLAGIARISM REPORT OF THE TERM PAPER



paper - adnan

ORIGINALITY REPORT

6%	6%	0%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	towardsdatascience.com Internet Source	2%
2	en.wikipedia.org Internet Source	2%
3	medium.com Internet Source	1%
4	louisdl.louislibraries.org Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		