

NLP Assignment-2.

- Q Explain with suitable example following between the meaning Homonymy, Polysemy, Antonymy, Hyperonymy
- * Homonymy - It may be defined as the word having same spelling or same form but having different and unrelated meaning

For example. The word 'Bat' is a homonymy word because it can be implemented to hit a ball or bat is a flying object

* Polysemy - This means 'many signs'. It is a word or phrase with different but related sense. Polysemy has the same spelling but offer different

For example + 'Bank'

following Meaning -

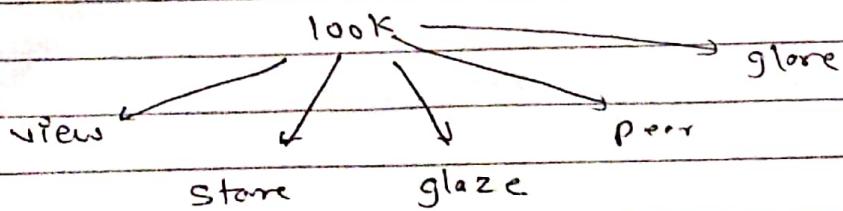
1) financial Institution

2) also synonym for 'to reply on'

* Antonymy :- It is the relation b/w 2 lexical items having symmetry between their components relative to an axis

for example - rich / poor , life / death , father / son etc

* Hyperonymy : Hyperonymy is the condition of being inclusive of all other synergy of a given of synonym



Teacher's Sign.: _____

* **Hyponymy** :- It may be defined as the relationship between generic term and instance of that generic term. Here generic term is called as hypernymic instances are called as hyponym.

For example : The word color is hypernym and color blue are hyponyms.

* **Meronymy** - It may be defined as a semantic relation specific to linguistics. A meronym is a part or contained part of a member of something.

Framework : finger is ~~also~~ meronym of hand.

Similarly, wheels is a meronym of automobile.

a) Why is word sense disambiguation a challenging problem in NLP?

→ In natural language processing, word sense disambiguation is the problem of determining which sense of word is activated by the use of the word in a particular context, a process which appears to be largely unconscious in people.

WSD is a natural classification problem : Given a word and its possible senses, as defined by a dictionary, classify an occurrence of the word into

into one or more of its sense classes. The feature of the context provide the evidence for classification

for example - little John was looking for history box finally he found it. The box was in the pen . John was very happy.

wordnet lists 5 senses for the word pen:-

- i) Pen - a writing implements with a pt from which ink flows.
- 2) Pin - an enclosure for confining livestocks
- 3) Pen - a portable enclosure in which babies may be left to play
- 4) Pen - a correctional institution for those connected Major , crime
- 5) Pen - female

words are typically to have a finite & discrete sentence , a gross simplification of the complexity of word meaning . There are different algorithm for different application . In machine translation, the problem takes the form of target word selection . Here the senses are words in the target language which often correspond to significant meaning distinctions in the source language

for example - Bank could translate into financial bank or edge of River

a) Draw and explain shift-reduce parser in Natural language processing.

→ The Shift-Reduce parser by maintaining a state of the current passed tree, with the words of the sentence on a queue and partially completed trees on a stack and applying transitions to the state into the queue is empty and the current stack may ~~not~~ contain a finished tree.

The initial state is to have all of the words in order on the queue with an empty stack. The transitions which can be applied are:-

* Shift :- A word moves from the queue onto the stack

* Unary Parse : A label of the first constituent on the stack changes. There is a different unary transition for every possible unary node in the treebank used for training

* Binary Reduce : The first two nodes on the stack are combined into a new label.

They are either right sided or left sided indicating which child is treated as the head.

* Finalize :- A tree is not considered finished until the parser chooses the finalize transition

Part of speech tags are not assigned by this parser and are in fact used as features. This is accomplished by pre-tagging the text, meaning the POS annotation needs to be used.

for example "the dog saw man in the park"

1. Initial state

stack

Removing text

the dog saw a man in the park

2. After a shift

stack

Remaining text

the

dog saw a man in the park

3. After reduce shift-reduce

stack.

Remaining text

Det

N

I

I

saw a man in the park

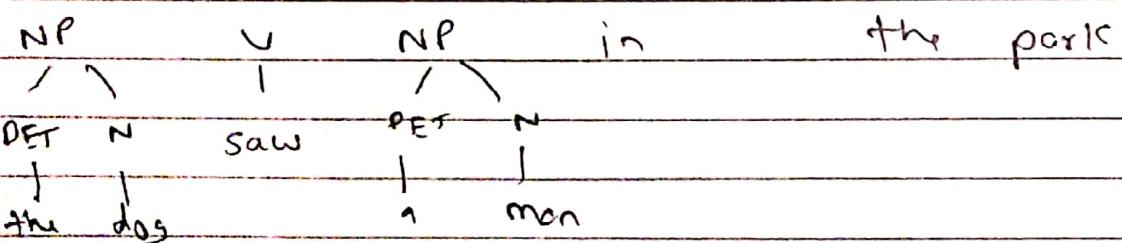
the

dog

4. After recognizing the second NP.

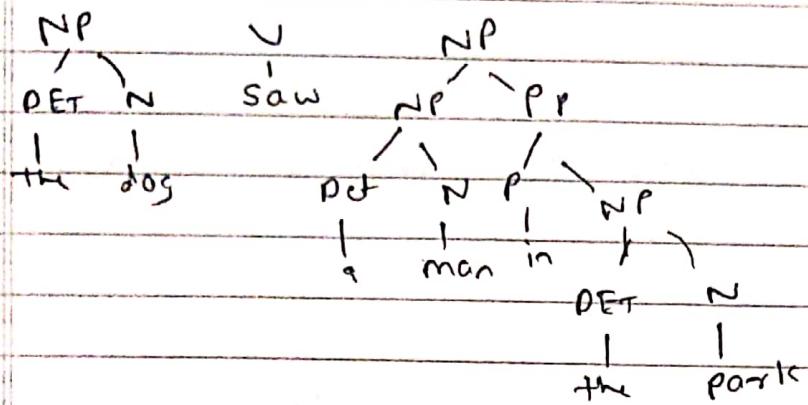
stack.

Remaining text

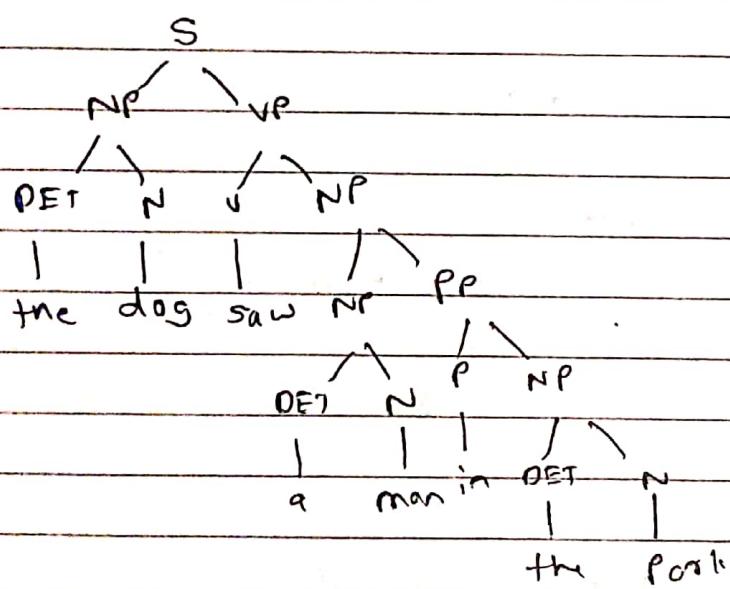


Teacher's Sign.: _____

5) After building a complex NP
Stack

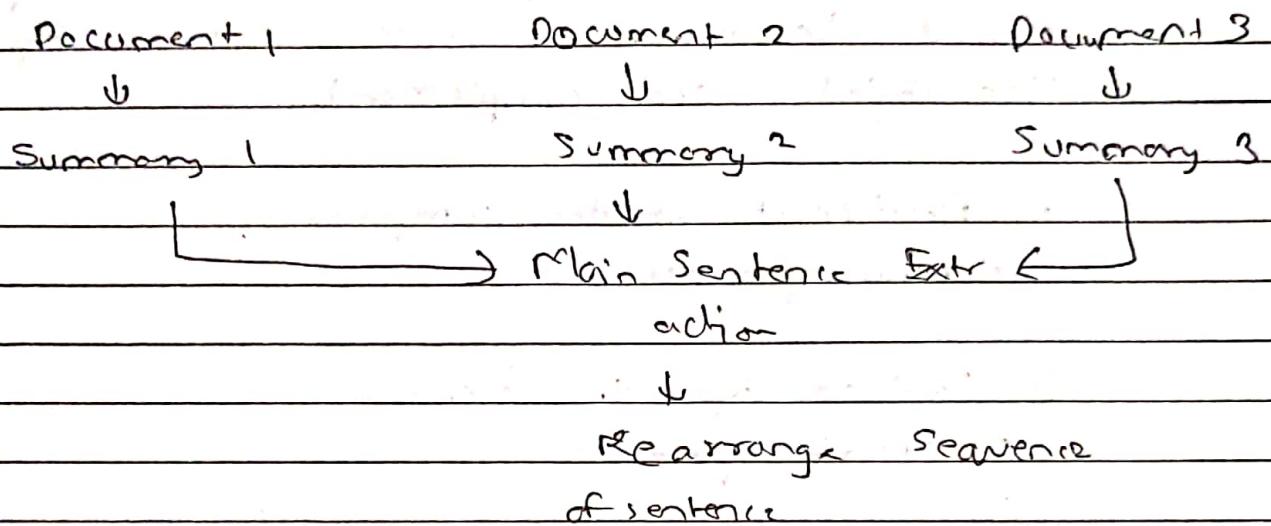


6. Built a complete parse tree
Stack



f Explain the process of multi - document summarization

An Multi-document summary is the process of dealing with a large amount of documents only the essential element or main idea in document inter space . Multi-document summarization by sentence extraction is the same as single document.



a Explain noun phrase and verb phrase in language
Identify the head morphological type

- i) president of the company
- ii) look up the chimney
- iii) Angry or hippo
- iv) rapidly like a bat

→ & Noun phrase +

- A noun phrase is phrase that has a noun as its head or perform the same grammatical function as a noun

⇒

→ verb phrase

- A verb is a syntactic unit composed of at least one verb and its dependents like object and complement

i) President of the company

→ Noun phrase

ii) Looked up the chimney

→ verb phrase

iii) Angry as hippo

→ Adjective phrase

iv) Rapidly like a bat

→ Adverbial phrase

a write a short note on QA system

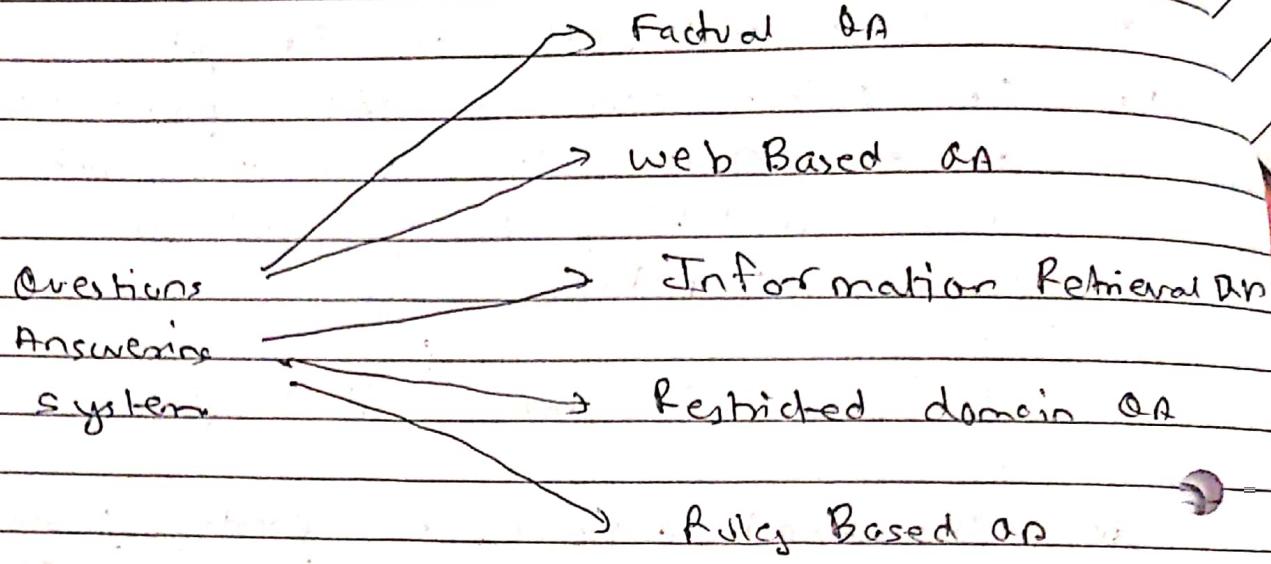
Ans QA system aim to retrieve point to point answers rather than flooding with documents by ever matching passage and most of information Retrieval System.

ex: "Who is the ~~first~~ first prime minister of India?"

The exact answers is Pandit Jawaharlal Nehru but not intended to read through the passage or document that match with the word like first, prime minister etc. It also aims to recognise the errors linguistic question which ~~allow~~ allow the user to ask questions & obtain the answer in their native language.

QA research attempts to deal with wide range of questions types including facts, list fiction how, why and 2 type ; closed domain QA system open domain QA system

Open domain QA system deal with question under a specific domain & can be seen a easier task

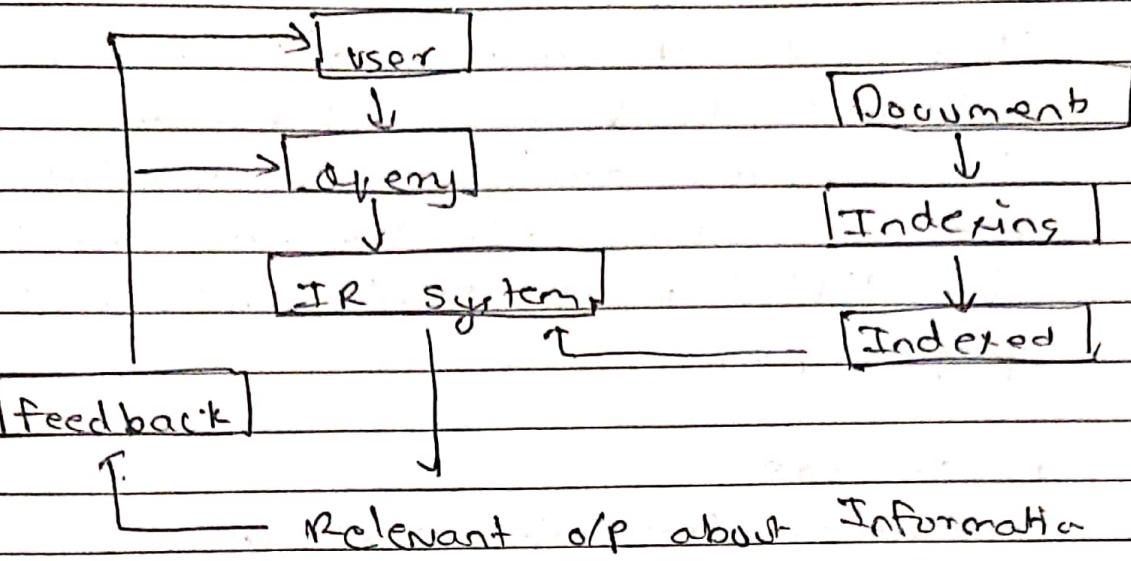


a) write a short note on IER.

a) Information Retrieval system -

IR may be defined as a software program that deals with organization, storage, retrieval and evaluation of information from document repositories particularly textual information.

The system assist user in finding information they require but it does not explicitly return the answer of the question.



Mathematically, a retrieval Model consist of

D - Representation of documents

R - Representation of queries

T = The modelling framework for D, R along with relationship b/w them

b) Machine translation -

→ Machine translation is a sub field of computational linguistics that investigates the use of software to translate text or speech from one language to another. On a basic level, Machine translation performs simple substitution of words in one language for words in another, but that alone usually cannot produce a good translation of text.

Improved O/P quality can also be achieved by human intervention. Some systems are able to translate more accurately. If the user has unambiguously identified which words, in the text are proper names.

Major issues

- * Machine translation could produce some non-understandable phrase
- * Disambiguation
- * Non-standard speech
- + Named Entities

c) Semantic analysis

⇒ The purpose of semantic analyser is to draw exact meaning or dictionary meaning from the text. The work of semantic analyzer is to check them for all meanings. Semantic Analysis can be divided into 2 parts:
i) Studying meaning of individual word.
ii) Studying the combination of individual word.

i) Individual word -

It is the first part of semantic analysis. It is also called lexical semantics.

ii) Combination of individual words -

A more contextual analysis of the words.
Ex - "Ram is good"

In this sentence, Ram can either be person or god.

a) Wordnet -

Wordnet is a lexical database of semantic relations between words in more than 200 languages. Wordnet groups words into relation including synonymy, meronymy etc. The synonyms are grouped into synsets with short definitions and usage examples. Wordnet can be thus be seen as combination of dictionary and thesaurus.

While it is accessible to human users via a web browser, its primary use is in automatic text analysis and NLP.

Limitations

- It is easy to ~~easy~~ create hyponymy relationships to capture that one is a type or tree but is difficult to classify emotions like fear or happiness.

- wordnet does not include info about etymology and pronunciation of word, and limited info about usage
- wordnet does not include much domain-specific terminology

* Application -

- wordnet has been used from a number of purposes in info system. Including word sense disambiguation, Info Retrieval, automatic text classification, automatic text summarization etc.
- A common use of wordnet is to determine the similarity b/w words. It can also be used as link btw vocabulary