Answer the following questions:

A) What are the fields related to semantics? Give a definition for each of them.

Solution:-

Research fields

➢ Linguistics

* Meaning in natural languages is mainly studied by linguist

➢ Computer Science

* Theorical computer scientists and logicians think about artificial languages
* e.g. in machine translation computer scientists have to design artificial languages for representing meanings

➢ Psychology

* Semantic memory is memory for meaning
* in other words, the aspect of memory that preserves only the general significance, of
* remembered experience

B) Give an example of polysemous word different from those in the lecture slides.

Solution: Project – to Project , Project work

C) What is the difference between lemma and lexeme? Explain with examples.

Solution:-lexeme – is a unit of meaning. All the forms of a root word which carry the same meaning is called a lexeme. Ex: write, writes, wrote, writing.

Lemma is the headword which defines the meaning for all the canonical forms in the lexeme.

D) What is Information Retrieval?

Solution:- Information retrieval is the field concerned with the structure, analysis,

organization, storage, searching, and retrieval of information

E) What are the IR tasks?

Solution:-

* Build a system that retrieves documents that are most likely relevant to the user
* Classify documents
* Filtering these documents
* Route the documents
* Clustering based on similarity

Other tasks close to IR

❖ Automated document categorization

❖ Automated document clustering

❖ Automated text summarization

❖ Question answering

❖ Information filtering (spam filtering)

❖ Information extraction

❖ Information integration

❖ Recommendation information or products

❖ Searching and ranking in Web 2.0

F) Explain the differences between Data Retrieval (Databases) and Information Retrieval.

Data Retrieval (DR)  Information Retrieval (IR)

Matching Exact match Partial match, best match

Inference Deduction Induction

Model Deterministic Probabilistic

Classification Monothetic Polythetic

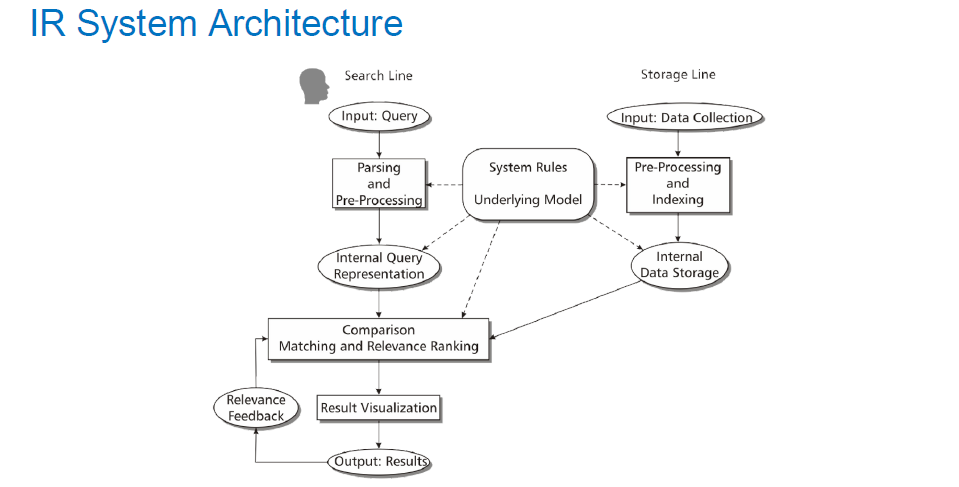
Query language Artificial Natural

Query specification Complete Incomplete

Items wanted Matching Relevant

Error response Sensitive Insensitive

G) How does an IR system works? Sketch the architecture and explain shortly each component.



**Underlying Model :-**

Framework for representation of Queries and Data Items and their relationships

Ex: Probabilistic Model.

**Internal Data Representation:-**

Representation preserves the meaning of the data /context, efficient query processing, less memory usage.

Ex:- Inverted Index structure

**Pre-Processing and Indexing:-**

Tokenizing, Lemmatizing, Recognizing entities, Stopword elimination, Syntactic and Statistical analysis

**Querying:-**

May be NL, Boolean, Stylized NL and Form Based GUI

**Matching & Relevance:-**

Retrieving document matching the queries. Relevance metrics used for ordering.

Ranking can be based on other metrics-Popularity, trust and no. of query terms found, chronology.

**Result Visualization:-**

Ranking, summary of docs listed , displaying relevance score

Interaction :- feedback, query refinement and filtering.

Relevance Feedback:- Improve retrieval – re ranking through users marking relevance and non relevance. Expansion , reformulating and refinement of query.

H) What is a crawler?

Solution:-

Crawls the web to collect documents for the IR system. Updates collection because of the non static nature of web pages.