Question and Answers Document

1. \*\*What is an information retrieval system?\*\*  
  
An information retrieval system (IRS) is a system that helps users find information from a collection of documents. IRSs typically use a variety of techniques, such as keyword searching, full-text search, and natural language processing, to help users find the information they are looking for.  
  
2. \*\*Contrast the differences between IRS and DBMS?\*\*  
  
An information retrieval system (IRS) is a system that helps users find information from a collection of documents. A database management system (DBMS) is a system that helps users store and manage data. The main difference between an IRS and a DBMS is that an IRS is focused on finding information, while a DBMS is focused on storing and managing data.  
  
3. \*\*Illustrate the Inverted File structure and PAT Data structure?\*\*  
  
An inverted file is a data structure that stores information about the terms that appear in a collection of documents. Each term is associated with a list of the documents that contain the term. The PAT data structure is a data structure that stores information about the patterns that appear in a collection of documents. Each pattern is associated with a list of the documents that contain the pattern.  
  
4. \*\*Multimedia Database Search\*\*  
  
Multimedia database search is a type of information retrieval that focuses on finding information from multimedia documents, such as images, videos, and audio files. Multimedia database search is more challenging than traditional information retrieval because multimedia documents are often more complex and difficult to index.  
  
5. \*\*Explain the terms i) Recall ii) Precision\*\*  
  
Recall is the fraction of relevant documents that are retrieved by a search system. Precision is the fraction of retrieved documents that are relevant.  
  
6. \*\*Explain the Total Information Retrieval System?\*\*  
  
The total information retrieval system (TIRS) is a model of information retrieval that consists of three components: a user, a document collection, and a retrieval system. The user submits a query to the retrieval system, which searches the document collection and returns a set of documents that are relevant to the query. The user then selects the documents that are most relevant to their needs.  
  
7. \*\*Explain Digital Libraries and Dataware House?\*\*  
  
A digital library is a collection of digital resources, such as text, images, videos, and audio files. A data warehouse is a collection of data that is used for analysis. Digital libraries and data warehouses are both important tools for information retrieval.  
  
8. \*\*Explain the conditions of Porter Stemming Algorithm —\_?\*\*  
  
The Porter stemming algorithm is a stemming algorithm that is used to reduce words to their root form. The algorithm is based on a set of rules that are applied to words. The rules are designed to remove suffixes from words and to return the root form of the word.  
  
9. \*\*Explain the PAT Data structure? —\*\*  
  
The PAT data structure is a data structure that stores information about the patterns that appear in a collection of documents. Each pattern is associated with a list of the documents that contain the pattern. The PAT data structure is used in information retrieval to find documents that contain patterns that are similar to a query pattern.  
  
10. \*\*Two-level Bayesian networks with neat diagram pg. No. [61]\*\*  
  
A two-level Bayesian network is a Bayesian network that has two levels of nodes. The first level of nodes represents the variables that are observed, and the second level of nodes represents the variables that are unobserved. The two-level Bayesian network is used to model the relationships between the observed and unobserved variables.  
  
11. \*\*Explain\*\*  
  
(a) Canned Query pg. No. [43]  
  
A canned query is a query that is pre-defined and stored in a database. Canned queries are often used in information retrieval to find documents that are relevant to a specific topic.  
  
(b) Bigrams, Trigrams and Pentagrams [85]  
  
A bigram is a sequence of two words. A trigram is a sequence of three words. A pentagram is a sequence of five words. Bigrams, trigrams, and pentagrams are often used in information retrieval to find documents that are relevant to a specific topic.  
  
(c) N-Gram [85]  
  
An n-gram is a sequence of n words. N-grams are often used in information retrieval to find documents that are relevant to a specific topic.  
  
12. \*\*What are Classes of Automatic Indexing pg No. [106]\*\*  
  
There are three main classes of automatic indexing: statistical indexing, rule-based indexing, and hybrid indexing. Statistical indexing is based on the frequency of words in a document. Rule-based indexing is based on a set of rules that are applied to words. Hybrid indexing is a combination of statistical indexing and rule-based indexing.