Question and Answers Document

1. \*\*What is information retrieval?\*\*  
  
Information retrieval (IR) is the process of finding documents that are relevant to a user's information need. IR systems typically use a combination of techniques, such as keyword searching, full-text indexing, and machine learning, to identify and retrieve relevant documents.  
  
2. \*\*What are the differences between IR and DBMS?\*\*  
  
IR and DBMS are both concerned with storing and retrieving data, but they have different goals. IR systems are designed to retrieve documents that are relevant to a user's information need, while DBMS are designed to store and retrieve data that is accurate and consistent.  
  
3. \*\*What is the inverted file structure?\*\*  
  
The inverted file structure is a data structure that is used to store the terms that occur in a document and the documents that contain those terms. The inverted file structure is often used in IR systems to improve the efficiency of keyword searching.  
  
4. \*\*What is the PAT data structure?\*\*  
  
The PAT data structure is a data structure that is used to store the patterns that occur in a document. The PAT data structure is often used in IR systems to improve the efficiency of pattern matching.  
  
5. \*\*What is multimedia database search?\*\*  
  
Multimedia database search is the process of searching for multimedia data, such as images, videos, and audio files. Multimedia database search is often used in applications such as image retrieval, video retrieval, and audio retrieval.  
  
6. \*\*What are recall and precision?\*\*  
  
Recall is the fraction of relevant documents that are retrieved by an IR system. Precision is the fraction of retrieved documents that are relevant.  
  
7. \*\*What is the total information retrieval system?\*\*  
  
The total information retrieval system is a model of an IR system that includes all of the components that are necessary for effective information retrieval. The total information retrieval system includes the following components:  
  
\* User interface: The user interface allows users to interact with the IR system.  
\* Query processor: The query processor parses and interprets user queries.  
\* Indexer: The indexer creates an index of the documents in the IR system.  
\* Search engine: The search engine retrieves documents that are relevant to user queries.  
\* Ranking algorithm: The ranking algorithm ranks the retrieved documents in order of relevance.  
\* Presentation component: The presentation component presents the retrieved documents to the user.  
  
8. \*\*What are digital libraries and data warehouses?\*\*  
  
Digital libraries are collections of digital objects, such as documents, images, videos, and audio files. Data warehouses are collections of data that are used for decision-making.  
  
9. \*\*What is Porter stemming algorithm?\*\*  
  
Porter stemming algorithm is a stemming algorithm that is used to reduce words to their root form. Stemming is the process of removing suffixes from words to reduce them to their base form.  
  
10. \*\*What is PAT data structure?\*\*  
  
The PAT data structure is a data structure that is used to store the patterns that occur in a document. The PAT data structure is often used in IR systems to improve the efficiency of pattern matching.  
  
11. \*\*What are two-level Bayesian networks?\*\*  
  
Two-level Bayesian networks are a type of 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 hidden.  
  
12. \*\*What is canned query?\*\*  
  
A canned query is a pre-defined query that is used to retrieve documents from an IR system. Canned queries are often used in applications where users do not have the time or expertise to formulate their own queries.  
  
13. \*\*What are bigrams, trigrams, and pentagrams?\*\*  
  
Bigrams, trigrams, and pentagrams are n-grams, which are sequences of n words. Bigrams are sequences of two words, trigrams are sequences of three words, and pentagrams are sequences of five words. N-grams are often used in IR systems to improve the accuracy of keyword searching.  
  
14. \*\*What are the classes of automatic indexing?\*\*  
  
The classes of automatic indexing are:  
  
\* Full-text indexing: Full-text indexing is the process of indexing the entire text of a document.  
\* Keyword indexing: Keyword indexing is the process of indexing the keywords that occur in a document.  
\* Phrase indexing: Phrase indexing is the process of indexing the phrases that occur in a document.  
\* Statistical indexing: Statistical indexing is the process of indexing the words and phrases that occur in a document based on their frequency and location in the document.