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 to index and search documents, and they often provide users with a variety of ways to interact with the system.  
  
2. \*\*Contrast 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 designed for searching and retrieving information, while a DBMS is designed for storing and managing data.  
  
3. \*\*Illustrate the Inverted File structure and PAT Data structure ? Pg. No. [82]\*\*  
  
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 Pg. No. [20]\*\*  
  
Multimedia database search is a type of information retrieval that focuses on finding information from multimedia documents. Multimedia documents can include text, images, audio, and video. Multimedia database search is often used to find information from websites, e-books, and other online resources.  
  
5. \*\*Explain the terms i) Recall ii) Precision Pg. No. [5]\*\*  
  
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? Pg. No. [10-17]\*\*  
  
A total information retrieval system (TIRS) is a system that attempts to retrieve all relevant documents from a collection. TIRSs typically use a variety of techniques to index and search documents, and they often provide users with a variety of ways to interact with the system.  
  
7. \*\*Explain Digital Libraries and Dataware House? Pg. No. [21-22]\*\*  
  
A digital library is a collection of digital objects, such as text, images, audio, and video. Digital libraries are often used to store and manage information that is not easily accessible in traditional formats, such as paper documents. A data warehouse is a repository of data that is collected from a variety of sources. Data warehouses are often used to analyze data and to make decisions.  
  
8. \*\*Explain the conditions of Porter Stemming Algorithm —\_? Pg. No. [75-77]\*\*  
  
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 they are applied in a specific order.  
  
9. \*\*Explain the PAT Data structure? — Pg. No. [88-93]\*\*  
  
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 to find documents that contain patterns.  
  
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 being modeled, and the second level of nodes represents the factors that influence the 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 to retrieve information from a database that is frequently requested by users.  
  
(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 to represent the context of a word.  
  
(c) N-Gram [85]  
  
An n-gram is a sequence of n words. N-grams are often used to represent the context of a word.  
  
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 uses statistical methods to index documents. Rule