INJNTU.COM

4-1 1 mid of **Information Eetrieval Systems**

Q)Dissemination of information is applicable for --> **Text**
Q)The rank-frequency law of Ziph is--> **Frequency * rank = constant**
Q)The rank-frequency law of Ziph is--> **Frequency * rank = constant**
Q)IRS uses the following character code --> **Unicode**
Q)The process of parse the item into logical sub-divisions is called --> **Zoning**
Q)Example for inter-ward symbol --> **Alphabet**
Q)An item in the IRS is --> **Document**
Q)Measure in IRS is --> **Precision**
Q)Recall defined on --> **Number-retrieved-relevant /number-total-retrieved.**
Q)In the following which is not IRS --> **Webseek**
Q)First step in the IRS --> **Item normalization**
Q)ntegration of DBMS and Information Systems is --> **Very important**
Q)Which one of the following commercial databases is first to integrate IRS and DBMS. --> **INQUIRE DBMS**
Q)_____ provides the capability to dynamically compare newly recurred items in the information system against standing statements of users and deliver the item to those users whose statement of interrupt matches the contents of the item. --> **Selective dissemination**
Q)_____ Data is represented by tables. --> **Structured data**
Q)Search in DBMS is done --> **According to iterative search**
Q)_____ Assist in disambiguation of a particular word --> **Characterization**
Q)The object stop/list algorithm is to eliminate the set of searchable processing tokens those have little value to the system. Justify? --> **True**
Q)_____ reduces system overhead in expanding a search term to similar token representation --> **Stemming**
Q)_____ eliminates the set of searchable processing tokens those that have little value. --> **Stopping**
Q) When the search statement is satisfied the item is placed in the _____ associated with the profile. --> **Mail files**
Q)_____ is used to restrict the distance allowed within an item between two search terms --> **Proximity**
Q)The Query "Find any item containing any two of the terms: "AA","BB, "CC" can be expanded into Boolean searches follows --> **((AA AND BB) OR (AA AND CC) OR (BB AND CC))**
Q)The algorithms used for searching in IRS are called --> **Natural language processing**
Q)Which of the following systems allow for natural languages --> **Topic and retrieval wave and InQuery**
Q)The Weighting of search terms in retrieval wave system is in between --> **0.0 and 1.0**
Q)Data mart is a --> **Subsets of a data warehouse or a small data warehouse**
Q)Which one is more focused on structured data and decision support technologies --> **Data warehouse**
Q)Search and retrieval of data with concern for establishing standards on Contents of the system is done in which of the following --> **Digital library**
Q)The process of search in Data warehouse is called --> **Data mining**
Q)Acronym of KDD related to Data warehouse --> **Knowledge Discovery in databases**
Q)A thesaurus is typically _____expansion of a term to other terms --> **one-level or two level**
Q)Which of the following is tree structure.
(i) Thesaurus (ii) concept class
--> **(ii) only**
Q)The typical format of proximity is TERM within " m" "units" of TERM2 Here m is meant for_____ --> **Dolitance**

INJNTU.COM INJNTU.COM INJNTU.COM

INJNTU.COM

Q)Which one of the following is not a contiguous word phrase (CWB) --> **America Manufacturing**
Q)Which one of the following is "N" ary Operator --> **CWP (Contiguous Word Phrase)**
Q)_____ provide the capability to locate spelling of words that are similar to the entered search term. --> **Proximity and CWP**
Q) A Fuzzy search on the term computer would not include the following word --> **common**
Q)Typically relevance scores are normalized to a value between --> **0.0 and 1.0**
Q)_____allows the users to rank the items to order the output for other user queries that are similar --> **collaborating filtering**
Q)The following query belongs to "Find where Bill Clinton is discussing Cuban refugees and there is a picture of a boot". --> **multimedia search**
Q)Two types of Browsing capabilities are____ --> **line item status and data visualization**
Q)_____is an estimate of the search system on how closely the item satisfies the search statement. --> **relevance score**
Q)In which thesaurus it is very difficult to name (or) understand thesaurus class by viewing it what caused its creation.
i. Semantic thesaurus
ii. Statistical thesaurus
--> **(ii)only**
Q)Which of the following Query allow a uses to enter a prose statement that describes the information that the user wants to find (i) Boolean query (ii) natural language query. --> **(ii) only**
Q)To accommodate negation concept natural language systems, it can use Boolean logic capability. --> **True for all systems or true for only some of the commercial systems**
Q)The correlation between different parts of a query against different modalities is usually based upon _____ --> **Time or location**
Q)Modality is the factor associated with which of the following --> **Multimedia queries**
Q)_____ lets the user quickly focus on the potentially relevant parts of the text to scan for item relevance. --> **highlighting**
Q)The DCARS system acts as a rser frontend to_____ system allows the user to browse an item in the order of paragraphs or individual words. --> **Retrieval ware search**
Q)In graphical visualization techniques, browsing of items could be done using 2 or 3 dimensional graphs. In this, location of points on the graph represents _____ --> **relative relationship.**
Q)____ browsing capability can be used in displaying individual items and the terms that contributed to the items selection. --> **information visualization**
Q)_____is the idea of locality and passage based search and retrieval from a hit item that has to be reviewed. --> **zoning**
Q)_____ provide the user with the capability to determine which items are of interest and select those to be displayed. --> **browse capabilities**
Q)Which one of the following is not browse capability? --> **canned query**
Q)An item with default relevance value 0.0 is_____ --> **non-relevant**
Q)The default relevance value of an item is used to _____ in search --> **start processing items**
Q)In graphical visualization techniques, browsing of items could be done using 2 or 3 dimensional graphs. In this points on the graph represent _____ --> **item**
Q)The feature of allowing variables to be inserted into the auery and bound to specific values at execution time. --> **stored query**
Q)The standard organization most involved in information systems standards is in _____ --> **USA**
Q) _____ is the process of refining the results of a previous search to focus on relevant items. --> **iterative**

INJNTU.COM INJNTU.COM INJNTU.COM

INJNTU.COM

search

Q)The search history log contains _____ **--> query, status and no. of hits of previous search.**

Q)_____ capability is used to name a query and store it to be retrieved and executed during a later user session. **--> canned or stored query**

Q)Highlighting has always been useful in Boolean system because of the _____ between the terms in search and the terms in the item. **--> indirect Mapping**

Q)Information visualization appears to be better than highlights in _____**--> helping users to formulate his query**

Q)_____ provides the capability to display in alphabetical sorted order words from the document database. **--> vocabulary browse**

Q)In vocabulary browsing sorted list of words along with the count of _____ will be displayed to the user. **--> number of items in which the word is found.**

Q)In iterative search hit file contains _____ **--> more iltems than the user wants to review**

Q)Extended services of Z39.50 standard does not include saving one of the following. **--> Resource-report**

Q)In Z39.50 standard client and server identified as **--> origin and target**

Q)An International version of Z39.50 called _____ **--> Search and Retrieval standard(SR)**

Q)Which one of the following is not belongs to eight operation types of Z39.50 standard. **--> Organize**

Q)Which one of the following is not belongs to five types of queies of Z39.50 standard.**--> type 103**

Q)NISO acronym is _____**--> National information standard organization**

Q)The standard organization most involved in information systems standards in United states is _____ **--> ANSI/NISO**

Q)_____ is the standard of ANSI/NISO, which provides informational retrieval applications service definition and protocol specification. **--> Z39.50**

Q)The objective of Z39.50 standard is _____ **--> To overcome incompatibilities with multiple database searching**

Q)The first version of z39.50 approved in _____ **--> 1992**

Q)The achronym of RAP, which is the communication protocol of CNRI. **--> Repository Archive Protocol**

Q)The acronym of IETF is . **--> Internet Engineering Task Force**

Q)Achronym of CNRI is _____ **--> Center for National Research Initiatives**

Q)Handle server Architecture belongs to _____ **--> CNRI**

Q)The communication protocol to retrieve items from existing systems is being provided by CNRI that is called as _____ **--> RAP**

Q)_____ is the de facto standard for many search environments on the INTERNET. **--> WAIS**

Q)Acronym of WAIS **--> Wide Area Information service**

Q)Which of the following standard had nonmarkovian process. **--> WAIS**

Q)Which of the following standard focused on a bibliographic MARC record structure against structured files. **--> Z39.50**

Q)Which of the following is true. **--> WAIS overcomes the deficiencies of Z39.50**

Q)The full text searchable data structure for items in the document file provides a view class of indexing called _____ **--> .total document indexing**

Q)_____ is a finite set of index terms from which all index terms must be selected. **--> controlled vocabulary**

Q)The earliest cataloging system DIALOG contains hop many no. of indexing databases by 1988. **--> 320 Index databases**

Q)Throughout the history of the libraries indexing has been done by _____ **--> Professional Indexers**

INJNTU.COM INJNTU.COM INJNTU.COM

INJNTU.COM

associated with libraries
Q)Which one of the following cataloging system allows the sharing of indexes between libraries. --> **MARC**
Q)Which one of the following is the earliest cataloging system. --> **DIALOG**
Q)In late 1800s subject Indexing became hierarchical. The example for this is _____ --> **Dewey Decimel system**
Q)Acronym of MARC is --> **Machine Readable Cataloging**
Q)DIALOG is the earliest cataloging system is developed by _____--> **Lock heed corporation in NASA**
Q) The earliest cataloging system was developed in _____--> **1965**
Q)Low exhaustivity has _____ --> **an adverse effect on both precision and recall**
Q)Low specificity has _____ --> **an adverse effect on precision but no effect on recall**
Q)_____ indexing uses ranking of items and clustering of items similar to a physical library. --> **electronic indexes**
Q)_____ is the extent to which the different concepts in the item are indexed. --> **exhastivity**
Q)_____ relates to the preciseness of the index terms used in linking . --> **specificity**
Q)A controlled vocabulary, in manual indexing environment makes _____ --> **indexing process slower, simplifies search process**
Q)An uncontrolled vocabulary in manual indexing environment makes ____ --> **indexing faster, search process difficult**
Q)_____ Concerns with the information needs of all users of the library system. --> **public file indexes**
Q)_____ Concerns with the information needs of individual users --> **private file indexes**
Q)_____ Concerns with the information needs of all users of the library system. --> **public file indexes**
Q)_____ Concerns with the information needs of individual users. --> **private file indexes**
Q)_____ saves the indexes from entering index terms that are identical to words in the document. --> **full document file indexes**
Q)Which of the following is not true about weighted automatic indexing --> **The weight of a term cannot be used to represent the importance of the term in the item.**
Q)Which one of the following is not the advantage of automatic indexing over human indexing --> **concurrency**
Q)Which of the following is not true about automatic indexing --> **It is slower than human indexing**
Q)Which one of the following is not true about unweighted automatic indexing --> **The last item presented in the file is the last item relevant to the user's information used.**
Q)If only title and abstract zones are indexed. Then it leads to _____ --> **loss of both precision and recall**
Q)Which one of the following is not true about weighting of index terms. --> **it is very each to add weights to index terms**
Q)The process of creating term linkages at index creation time is called _____ --> **precoordination**
Q)The process of creating term linkages at search time is called _____ --> **post coordination**
Q)_____ is the capability for the system to automatically determine the index terms to be assigned to an item. --> **Automatic indexing**
Q)Acronym of DR-LINK --> **Document Retrieval through linguistic knowledge**
Q)Which one of the following level is not exist in DR-LINK --> **stautural**
Q)A Bayesian network is a _____ --> **directed acyclic graph**
Q)What are the two types of probabilities that are needed to calculated in weighting approach for index terms using Bayesian network --> **prior probability and conditional probability**
Q)If c1 and c2 are nodes representing "The item contains concept ci" and F is the node representing "The

INJNTU.COM INJNTU.COM INJNTU.COM

INJNTU.COM

item has words Fij" then formula for calculating posterior probability --> **P(ci/Fi1,&..Fim)=P(ci) p(Fi1,&,Fim/ci)/P(Fi1,&,Fim)**
Q)In vectorized information system, vector represents --> **document**
Q)In vectorized information system, vector contains --> **weights**
Q)In vectorized information system, a location in the vector denotes --> **processing token**
Q)In vectorized information system, a value zer the word was not in the document --> **the word was not in the document**
Q)In vectorized Information system, search is accomplished by calculating _____ --> **Distance between query vector & document vector**
Q)Multimedias Indexing of vedio is divided into --> **frames**
Q)Which is the unit of search in multimedia indexing --> **vector of frames**
Q)The high dimensional vectors that are used to represent word items, items and queries in concept by indexing method of match plus are called _____ --> **Content vectors**
Q)Which one of the following is incorrect interpretation if for any word item is, its content vector vk is an n-dimensional vector with each component j in concept by index --> **vk is infinite if word k contradicts feature j**
Q)What are the three levels that are associated with the multimedia indexing --> **Raw data level, feature level and semantic level**
Q)_____ determines a canonical set of concepts based upon a test set of terms and uses them as a basuis for indexing all terms --> **Indexing by concept or latent semantic indexing**
Q)The example system for concept indexing --> **Match plus**
Q)How many neural networks are used in concept indexing of match-plus system --> **two**
Q)The goal of match plus system that uses the concept indexing is _____ --> **to be able to determine from the corpus of items, word relation ships and the strength of these relation ships and use that information in generating content vectors**
Q)In match-plus system that uses concept indexing, the following are represented by high dimensional vector --> **Word items,items& queries**
Q)Which one of the following metric of information extraction refers to the amount of irrelevant information that is extracted --> **Over generation**
Q)Which one of the following metric of information extraction refers to how much a system assigns incorrect slot filters as the number of potential incorrect slot fillers --> **Fall out**
Q)Which one of the following is not metric of the information extraction --> **Fall in**
Q)Which one of the following metric of information extraction refers to how much information was extracted versus how much should have been extracted from the item --> **Recall**
Q)Which one of the following metric of information extraction refers to how much information was extracted versus the total information extracted --> **Precision**
Q)In multimedia indexing of India, video presentations are made using SMIL, acronym of SMIL --> **Synchronized multimedia Integration language**
Q)What are the two mechanisms that are use to correlate the different modalities indexing search in multimedia indexing --> **Positional & temporal**
Q)The following mechanism is used when modalities are interspersed in a linear sequential composition of multimedia indexing --> **positional**
Q)The following mechanism to correlate different modalities during search of multimedia --> **Positional & permanent**
Q)What is the term that is used to define particular category of information to be extracted --> **slot**
Q)The stem "compu" could not associate with one of the following word --> **"communicate"**

INJNTU.COM INJNTU.COM INJNTU.COM

INJNTU.COM

Q)The stem "calculat" could not associate with one of the following word --> **"calcution"**
Q)Which among the following data structures allows the creator of an item to manually or automatically create imbedded links within one item to a related item --> **hypertext**
Q)Mapping of multiple morphological variants to a single representation(stem) is called_____ --> **conflation or stemming**
Q)The goal of stemming algorithm is to improve _____ --> **performance and recall**
Q)Which one of the following is not belongs to IRS data structures --> **Document file manager**
Q)_____is the process of reducing diversity of representations of a concept (word) to a canonical morphological statement. --> **Stemming algorithm**
Q)Which among the following data structures minimizes secondary storage access when multiple search items are applied across the total database --> **Inverted file**
Q)Which among the following data structures that breaks processing tokens into smaller string units and uses the token fragments for search --> **N-gram**
Q)Which among the following data structures view the text of an item as a single long stream versus a juxtaposition of words --> **PAT-trees and Arrays**
Q)The first condition of Port stemming algorithm is $C(CV)^mV$ here 'V' represents --> **vowels**
Q)The first condition of Port stemming algorithm is $C(CV)^mV$ here 'm' represents --> **no of VC repeats**
Q)Which of the following stemming algorithm applies stemming to both user's query and incoming text. --> **Porter algorithm**
Q)How many stem conditions are there in Port stemming algorithm --> **5**
Q)The first condition of Port stemming algorithm is $C(CV)^mV$ here 'C' represents --> **consonents**
Q)NLP means --> **Natural Language Processing**
Q)Which of the following stemming technique removes suffixes and prefixes recursively or iteratively. --> **Affix removal technique**
Q)Which of the following stemming algorithm leads to loss of precision and introduces some anomalies related to integrity of the system --> **Porter algorithm**
Q)Which of the following stemming technique determines prefix overlap as the length of a stem is increased. --> **Successor stemmer**
Q)In INQUERY system _____stemming algorithm is used --> **Kstem algorithm**
Q)After the application of rule 4 and rule 1b1 of Porter stemming algorithm , the word "delectable" is converted to _____ --> **duplicate**
Q)After the application of rule 4 , rule 1b1 and rule3 of Porter stemming algorithm , the word "duplicatable" is converted to _____--> **duplic**
Q)The meaning of pattern *d of Porter stemming algorithm is__ --> **stem ends in double consonants**
Q)The meaning of pattern *o of Porter stemming algorithm is _____ --> **stem ends with consonant-vowel-consonant sequence**
Q)After the application of rule 4 of Porter stemming algorithm , the word duplicatable is converted to _____ --> **duplicat**
Q)This is the example string when m=0 in Port stemming algorithm --> **"free"**
Q)This is the example string when m=1 in Port stemming algorithm --> **"frees"**
Q)This is the example string when m=2 in Port stemming algorithm --> **"prologue" or "compute"**
Q)The meaning of pattern *v* of Porter stemming algorithm is _____ --> **stem contains a vowel**
Q)if n=length of N-grams and λ=no of process able symbol from the alphabet Then maximum no of unique N-grams that can be generated --> **$MaSeg_n = \lambda n$**
Q)The name PAT is hort for --> **Patricia Trees**
Q)Which of the following is not true about inversion lists --> **Inversion list are not well suited to store**

INJNTU.COM INJNTU.COM INJNTU.COM

INJNTU.COM

concepts and their relationships

Q)Which of the following I not true about N-gram --> **N-grams are not at all relate to stemming**

Q)Which of the following is the first use of N-grams --> **used by cryptographers in world War II**

Q)How many methods are there for segmenting a word using successor varieties --> **4**

Q)._____of a segment of a word in a set of words is the number of distinct letters that occupy the segment length plus one character --> **successor variety**

Q)Which of the following is not a method of successor variety --> **complete sentence method**

Q)Paice has introduced a stemming performance measure ERRT to compare stemming algorithms --> **Error rate relative to truncation**

Q)Which of the following basic files are used by inverted file structure --> **Document file, inversion list and dictionary**

Q)HTTP stands for --> **Hyper Text Transfer Protocol**

Q)URL full form --> **Uniform Resource Locator**

Q)HTML defined by which model --> **DOM**

Q)XML is defined by which model --> **DTD**

Q)Which of the following data structure is used extensively in the internet environment --> **HTML and XML**

Q)PAT tree is a --> **Unbalance Binary digital tree**

Q)Search of signature matrix requires_____search time --> **O(n)**

Q)XML is the short form for --> **extensible markup language**

Q)Which of the following is web browser --> **Netscape**

Q)What is the full form of the model HMM --> **Hidden markov Model**

INJNTU.COM INJNTU.COM INJNTU.COM