

UNIVERSITY OF MUMBAI

MCQ QUESTION BANK

Course Code and Name: ITDLO8042 Information Retrieval System	
Class:	BE
Semester:	VIII
Name of Subject Incharge	Prof. Shaila K. Pawar(Deore) Prof. V. M. Kharche Prof. Shivshankar Bhutekar

Module No/Contact Hours	Weightage in Hrs	Q NO	QUESTION (2 marks per question)	OPTIONS				Correct Answer	Level of question (E/M/H)
				A	B	C	D		
1	5	1	_____ of a retrieval system has to translate his information need into a query in the language provided by the system.	The manager	The user	The designer	The administrator	B	E
		2	In the language of the World Wide Web, retrieval and browsing are called as _____	accessing actions.	modifying actions.	pulling actions.	pushing actions.	C	E
		3	_____ is usually provided by most modern information retrieval systems.	Information and knowledge retrieval	Information or knowledge retrieval	Information and data retrieval	Information or data retrieval	C	E
		4	Overhead from a user's perspective is the time required to find file information needed,	excluding the time for actually reading the relevant data.	including the time for actually reading the relevant data.	excluding the time for actually reading the non-relevant data.	including the time for actually reading the non-relevant data.	A	H
		5	search composition, search execution, and _____ are all aspects of information retrieval overhead.	reading relevant items	reading non-relevant items	writing relevant items	writing non-relevant items	B	M
		6	Once the logical view of the documents is defined, the database manager (using the DB Manager Module) builds _____ of the text.	an Appendix	the Bibliography	the Glossary	an index	D	E
		7	The full text is clearly the most _____ of a document but its usage usually implies higher	complete logical view	incomplete logical view	complete physical view	incomplete physical view	A	E
		8	The task whose main objectives are not clearly defined in the beginning and whose purpose might change during the interaction	retrieval	serching	browsing	sorting	C	E

		9	Which operations transform the original documents and genera'te a logical view of them?	The text operations	The query operations	Indexing operations	Searching Operations	A	M
		10	Which of the following is correct search statement for "Select all items that discuss software and not storage or hardware in the item"?	SOFTWARE OR (STORAGE NOT HARDWARE)	SOFTWARE OR STORAGE NOT HARDWARE	SOFTWARE AND NOT STORAGE OR HARDWARE	SOFTWARE OR NOT STORAGE AND HARDWARE	C	H
				OPTIONS					
Module No	Weigh atge in Hrs	Q NO	QUESTION (2 marks per question)	A	B	C	D	Correct Answer	Level of question (E/M/H)
2	9	1	In the information model quadraple [D,Q,F,R(qi,dj)], what D represents...	is a set composed of logical review for the user information need	is a set composed of logical review for the document in the collection	Framework for representing documet	Ranking function	B	M
		2	Major drawback suffered by Boolean model is due to the fact	Model based on set theort and bollean algebra	Queries are specified as boolean expresion	Retrieval strategy is based on binary decision criteria	Model predict that each document is either relevant or non-relevant	C	E
		3	Which of the following is not a classical model of information retrieval?	Boolean	Cluster	Probabilistic	Vector space	B	E
		4	Which of the following is False with respect to boolean model?	It does not perform query spell checking	It does not capture information regarding term position in the document	It does not consider the document structure	It uses term frequency information to rank results	D	H
		5	The resultant effect retrieved by the _____ model is document answer set	Boolean	Vector	Probabilistic	Set theoretic	A	E
		6	The probabilistic model was introduced by...	John Butler	Robertson and Sparck Jone	B. Fourozon	Jones Smith	B	E
		7	In which model, the query term defines a fuzzy	Boolean	Vector	Probabilistic	Set theoretic	D	E
		8	Document surrogates are the _____ of full document	Limited representation	Predefined structure	Classification	Detail representation	A	M
		9	What is the reason for having information retrieval model ?	As it helps to store the data securely	as it is giving entire information about information	as the models can serve as a blueprint to implement an actual retrieval system	as it used for information security	C	E

		10	In Boolean retrieval, each item in the list which records that the term is appeared in the document is called as	Ranking	Posting	Indexing	Grepping	B	H
		11	An instance of a sequence of characters in the particular document that are grouped together as a useful semantic unit for processing is	Type	Term	Token	Index	C	M
		12	Relevance is determined by	structure of document	structure of query	the similarity matching function	structure of query and document both	C	M
		13	_____provide a formal definition or mathematical framework for querying semi-structured textual databases	Set Theorotic model	Strucured text retrieval model	Latent semantic index model	extended boolean model	B	E
		14	Given a document containing the sentence "I left my left bag at my home" the number of tokens in the sentence is	2	4	6	8	D	H
		15	Which of the following items is not a component of a complete search system?	Document cache	Indexes	Spell correction	Horizontal index	D	M
		16	Weighted zone scoring is referred to as:	ranked Boolean retrieval	Zipf retrieval	Ad Hoc query retrieval	Jaccard retrieval	A	E
		17	A group of related documents against which information retrieval is employed is called:	Corpus	Text Database	Index Collection	Repository	A	E
		18	The standard approach to information retrieval system evaluation involves around the notion of:	Quantity of documents in the collection	Relevant and non relevant documents.	Accuracy	user happiness	B	E
		19	What is the advantage of boolean model?	retrieval criteria is based on binary decision criteria	In reality, It is a more data retrival model	Provide a framework which is easy to grasp by common users	Boolean expression has precise semantics	C	E
		20	Which of the following statement is false with regards to boolean retrieval model?	It answers query based on boolean expression	It views document as a set of terms	It is very precise as it meets a very specific condition	It cannot combine two operators "AND NOT" and "OR-NOT"	D	H
				OPTIONS					
Module No	Weightage in Hrs	Q NO	QUESTION (2 marks per question)	A	B	C	D	Correct Answer	Level of question (E/M/H)
3	10	1	Context queries uses search for the word _____	Which appear close to each other	Which appear far away from each other	Which appears Continuos	Whch appear only once	A	M

2	_____ is a sequence of single-word queries	Phrase	Pattern	Sentence	Keywords	A	M
3	With proximity query, a query "enhance retrieval" is searched such that two word should appear with the four word. Correct match will be	Enhance power of retrieval	Enhance the power of retrieval	Enhance retrieve process	Enhance power of all the retrieval	B	H
4	In boolean query syntax tree leaves corresponds to	A operands	result of operation	a operator	a basic query	D	M
5	Which of the boolean query selects all documents that satisfy e1 and not e2 ?	e1 OR e2	e1 AND e2	e1 BUT e2	e1 NOT e2	C	E
6	With boolean querying system _____ of document is not provided	Highlights occurrence of the word	Ranking	Sorting	Feedback	B	M
7	If a pattern "ter" matches with the word is matches document "tester" then it is called as	Prefix matching	String matching	Parse matching	Suffix matching	D	H
8	If a pattern "any flow" matches with the word is matches document "many flowers" then it is called as	Prefix matching	Sub-String matching	Parse matching	Suffix matching	B	
9	Which of the following operator can be used to build the statement, "if e1 and e2 are regular expressions, then (e1 e2)matches what e1 or e2 matches " ?	Union	Intersection	Concatenation	Repetition	A	E
10	_____ is a query protocol is based on the classical boolean model .	WAIS	CCL	CD-RDx	Z39.50	B	M
11	The user is presented with a list of the retrieved documents and,after examining them, marks those which are relevant. This strategy is called as	User relevance feedback	System relevant feedback	Initiation strategy	Retrieve and examine strategy	A	E
12	The application of relevance feedback to the _____ considers that the term-weight vectors of the documents identified as relevant (to a given query)have similarities among themselves	Vector model	Boolean model	Probabilistic model	Fuzzy set model	A	M
13	Simplicity of query expansion and term reweighting of vector model is because of:	modified term weights are computed directly from the set of retrieved documents.	optimality criterion is adopted:	modified query vector does reflect a a portion of the intended query semantics	a given query q is known in advance	A	M

14	In user relevance feedback which of the following is a disadvantage of probabilistic model ?	the feedback process is directly related to the derivation of new weights for query terms	document term weights are taken in to account during the feedback loop	weights of terms in the previous query formulations are regarded	No query expansion is used	D	H		
15	The classic probabilistic model introduced by _____ is composed of classes which group correlated terms in the context of the whole collection.	Robertson and Sparck Jones	Croft	Claude Shannon	Donald Kluth and George	A	E		
16		Similarity Thesaurus	Global clustering	Local clustering	Statistical Thesaurus	D	M		
17	The similarity between two clusters is defined as	the maximum of the similarities between all pairs of outer-cluster documents the minimum of the similarities between all pairs of inter-cluster documents	The minimum of the similarities between all pairs of inter-cluster documents the minimum of the similarities between all pairs of inter-cluster documents	the maximum of the similarities between all pairs of inter-cluster documents the minimum of the similarities between all pairs of outer-cluster documents	the minimum of the similarities between all pairs of inter-cluster documents the minimum of the similarities between all pairs of outer-cluster documents	B	H		
18	A pattern is set of	Systematic and meaning full ranking of documents	Combination of boolean feature and ranking	Symmetric feature	Syntactic feature	D	E		
19	In pattern matching, more powerful the set of pattern is allowed then	Implementation of search function becomes more complex	Search function can be implemented easily	Large no of document will be retrieved	Probability of pattern matching decreases	A	H		
20	This is the most basic pattern in pattern matching	Prefix	Suffix	Word	Substring	C	E		
Module No	Weightage	Q NO	QUESTION (2 marks per question)				Options	Correct	Level of question
			A	B	C	D			
4	10	1	Which of the following is NOT a benefit of index compression?	Simplified algorithm design	Reduction of disk space	Faster transfer of data from disk to memory	Increased Use of caching	A	E
		2	A web server communicates with a client (browser) using which protocol:	HTML	HTTP	FTP	Telnet	B	E

3	A model of information retrieval in which we can pose any query in which search terms are combined with the operators AND, OR, and NOT:	Ad Hoc Retrieval	Ranked Retrieval Model	Boolean Information Model	Proximity Query Model	C	M
4	An approach to computing scores in an IR system that orders documents in the posting list of a term by decreasing order of term frequency is called:	Champion list	Impact ordering	Cluster pruning	Tiered indexes	B	H
5	Which of the following item is not a component of a complete search system?	Document cache	Indexers	Spell correction	Horizontal index	D	E
6	Which of the following is a technique for context sensitive spelling correction?	Jaccard Coefficient	Soundex algorithms	k-gram overlap	Levenshtein distance	B	E
7	Which of the following software text search algorithm is fastest?	Brute Force Approach	Knuth-Morris-Pratt	Boyer-Moore	Robin-Karp	C	H
8	The bitmap Image format proposed for the Internet is	JBIG	WSQ	PNG	TGA	C	E
9	_____ provides a set of evolving features and architecture extensions to HTML and Web browsers that include cascading style sheets and document object model	DOM	VRML	SGML	SDML	B	M
10	A Thesaurus stands for...	Objective of removing affixes and allowing the retrieval of document containing syntactical variations of query term	Objective of filtering out words with very low discrimination value for retrieval purpose	Objective of treating digits, hyphens, punctuation marks, and the case of letters	Method of allowing the expansion of original query with related term	D	E
11	A metric derived by taking the log of N divided by the document frequency where N is the total number of documents in a collection is called:	Document frequency	tf-idf weight	collection frequency	inverse document frequency	D	M
12	_____ is designed to transform and style highly structured, data rich documents written in XML.	XLL	XSL	MathML	SMIL	B	M
13	_____ was designed to share document without losing control over the content, structure, and layout of that document	ODA	SDML	VML	STEP	A	E
14	A good compression algorithm is able to reduce the text to _____ of its original	20-25%	30-35%	40-45%	25-30%	B	E

15	The motivation for building thesaurus is based on the functional idea of using a _____ for the indexing and searching	Controlled Vocabulary	Posting List	Dictionary	Tokenisation	A	H
16	Lemmatization is a technique for:	Ranking documents	Case folding	Normalization	Tokenization	D	E
17	The gain obtained from compressing text is that	it requires less computation time	it requires less access time	it requires less storage space	it requires less formatting	C	E
18	One important consequence of using _____ is the possibility of performing direct searching on compressed text.	byte Huffman coding	Huffman coding	Adaptive dictionary methods p	Arithmetic coding	A	M
19	In which compression method, it is first necessary to establish the parameter b for each term?	Arithmetic	Ziv-Lempel	Golomb	Word Huffman	C	H
20	Which compression technique has slow compression speed?	Arithmetic	Character Huffman	Word Huffman	Ziv-Lempel	A	E
21	Compression models can be adaptive, static, or _____	semi-static	static	semi-dynamic	dynamic	A	E

Module No	Weightage	Q NO	QUESTION (2 marks per question)	OPTIONS				Correct	Level of question
				A	B	C	D		
5	11	1	Pick up the correct statement	Inverted files are quite amenable to compression	Inverted files are quite intractable to compression	Inverted files are quite uncooperative to	Inverted files are quite noncompliant to compression	A	E
		2	The third step of the GEMINI methodology is to____	Finds one or more numerical feature-extraction functions, to provide a 'quick-and-dirty' test.	Shows that the distance in feature space lower-bounds the actual distance.	Determines the distance measure between two time series.	Uses a SAM (e.g., an R-tree), to store and retrieve the <i>f</i> -D feature vectors.	B	E
		3	_____generation of Web query languages were aimed at combining content with structure	the third generation	the second generation	the first generation	the fourth generation	C	E
		4	_____ search in the Web is equivalent to sequential text searching	Dynamic search	Semi-dynamic search	Static search	Semi-Static search	A	E
		5	The _____ can also be compressed independently of the index.	image	audio	text	video	C	E

6	All ways of compressing signature files are based on the fact that _____	only last 5 bits are set in the whole file.	only a first 5 bits are set in the whole file.	only a few bits are set in the whole file.	only alternate bits are set in the whole file.	C	E
7	Broad query uses _____ to find good starting points?	search engines	web directories	encyclopedia	web dictionaries	B	M
8	The problem of having more than one meaning for a word is called _____	polisemy	pollysemy	polysemy	pollisemy	C	M
9	_____ search exploits the intuition that relevant documents often have neighbors that are relevant.	Whale Search	Shark search	Sea Lion Search	Fish search	D	M
10	In general, maintaining an inverted index can not be done by _____	Incremental updates	Rebuild	Intermittent merge	Intermittent copy	D	H
11	Which is a free Web navigation service that can be attached as a tool bar at the bottom of any browser and accompanies the user in his surfing?	Alexa	Mapuccino	SurfSerf,	WebTaxi	A	H
12	Searching and compression were traditionally regarded as _____ operations.	inclusive	exclusive	dependent	independent	B	E
13	Direct indexing and searching on compressed text provides better (sometimes much better) _____ at the same time.	time performance and less space overhead	time performance and less complexity	time performance and less computational overhead	time performance and less management	A	E
14	Which generation of languages, called Web data manipulation languages?	The first generation	The second generation	The third generation	The fourth generation	B	E
15	A _____ for a text of n characters can be built in $O(n)$ time.	suffix tree	B-tree	R-tree	prefix trees	A	E
16	The _____ represents a spatial object by its minimum bounding rectangle	R-tree	prefix trees	B-tree	suffix tree	A	E
17	Which technique involves taking advantage of the intrinsic parallelism of the bit operations inside a computer word (of w bits)?	Shift-And	Shift-ExOr	Shift-Not	Shift-Or	D	E
18	The second generation of web languages maintain the emphasis on semi-structured data	semi-structured data	structured data	unstructured data	fully-structured data	A	M
19	Vague queries use _____ and improve the query formulation based on relevant answers.	web search engines	web directories	web dictionaries	hyperlink	A	M

		20	Shark search search does a better relevance assessment of	odd pages	even pages	neighboring pages	non neighboring pages	C	M
		21	Knuth-Marris-Pratt algorithm moves	a window over the text and a pointer inside the window	a window over the text and a pointer outside the window	a window below the text and a pointer inside the window	a window below the text and a pointer outside the window	A	H
		22	The automaton accepts a text position as the end of a match with k errors whenever the (k +1)-th leftmost state is active.	leftmost state is active	Rightmost state is active.	leftmost state is inactive	Rightmost state is inactive.	A	H
Module No	Weigh atge	Q NO	QUESTION (2 marks per question)	OPTIONS				Correc t	Level of question
				A	B	C	D		
6	5	1	If the user has control of how and when feedback is provided, then the system provides an	internal locus of control	External locus of control	high sense of self-efficacy	Low sense of self-efficacy	A	E
		2	Which design principal provide a mechanisms for keeping track of choices made during the search process?	Offer informative feedback	Provide alternative interfaces for novice users.	Reduce working memory load	Provide alternative interfaces for expert users.	C	M
		3	Humans are highly attuned to _____ for visualization and understanding	Animation	Image	Text	inherently abstract information	A	M
		4	In graphical query specification which one of the following is not a property described by shneiderman	continuous representation of the object of interest	physical actions or button presses instead of complex syntax	Faceted groups are created for efficient result setd	rapid incremental reversible operations	C	E
		5	How many maximum query terms can be ANDed together in a standard Venn diagram.	Six	Two	Three	Four	C	E
		6	Graphical depictions of Venn diagrams Hertzum and Frokjaer found that	Information is represented as lists or icons within a 2D space.	improvements over standard Boolean syntax by providing users with a direct manipulation filter-flow model.	a simple Venn diagram representation produced faster and more accurate results than a Boolean query syntax.	providing a list of entry labels, as suggested above for specifying facets	C	H
		7	Graphical depictions of Venn diagrams _____ found improvements over standard Boolean syntax by providing users with a direct manipulation filter-	Robertson and Sparck Jones	Croft	Claude shannon	Young and Shneiderman	D	M

	8	List of information about documents in order of their computed relevance to the query is called as	Document surrogates	Query term hits within the document	Mapuccino	keyword-in-context	A	E
	9	_____ represents text in a manner resembling columns of newspaper text, with one `line' of text on each horizontal line of the strip.	VIBE	InfoCrystal	Mapuccino	Seesoft	D	E
	10	_____ allows the user to issue a query on a particular Web site.	VIBE	InfoCrystal	Mapuccino	Seesoft	C	E

