

## UNIT 1

1	<p>A web page whose content doesn't vary from one request to another is called a:</p> <ol style="list-style-type: none"><li>Text Page</li><li>Dynamic Page</li><li>Active Server Page</li><li>Static Page</li></ol>
2	<p>An approach to computing scores in an IR system that pre-computes for each term in the dictionary, the set of documents with the highest weights for the term is:</p> <ol style="list-style-type: none"><li>Champion list</li><li>Impact ordering</li><li>Cluster pruning</li><li>Tiered indexes</li></ol>
3	<p>The basic operation of a web browser is to pass a request to the web server. This request is an address for a web page and is known as the:</p> <ol style="list-style-type: none"><li>UAL: Universal Address Locator</li><li>HTML: Hypertext Markup Language</li><li>URL: Universal Resource Locator</li><li>HTTP: Hypertext transfer protocol</li></ol>
4	<p>A program that captures and indexes content from web pages is known as what insect:</p> <ol style="list-style-type: none"><li>Fly</li><li>Centipede</li><li>Mosquito</li><li>Spider</li></ol>
5	<p>For a moderately large collection of static documents maintained on a single system the most appropriate indexing algorithm would be:</p> <ol style="list-style-type: none"><li>Block sort-based indexing algorithm</li><li>Single-pass in memory indexing algorithm</li><li>Distributed Map-Reduce indexing algorithm</li><li>Dynamic indexing process employing an auxiliary index</li></ol>
6	<p>A measure of similarity between two vectors which is determined by measuring the angle between them is called:</p> <ol style="list-style-type: none"><li>cosine similarity</li><li>sin similarity</li><li>vector similarity</li><li>vector scoring</li></ol>
7	<p>Which of the following is NOT one of the types of queries in a complete search system discussed in our text?</p> <ol style="list-style-type: none"><li>Wildcard Query</li><li>Boolean retrieval</li><li>Phrase Query</li></ol>

	d. Ranked retrieval Query
8	In the context of web search engines the manipulation of web page content for the purpose of appearing high up in search results for selected query terms is called: a. Paid inclusion b. SPAM c. SEO d. Link Analysis
9	The number of times that a word or term occurs in a document is called the: a. Proximity Operator b. Vocabulary Lexicon c. Term Frequency d. Indexing Granularity
10	An approach to computing scores in an IR system that selects a sample of documents randomly from the collection as leaders which are in the index and links similar documents to it (followers) is called: a. Champion list b. Impact ordering c. Cluster pruning d. Tiered indexes

## UNIT 2

1	For a small collection of documents on a personal computer that don't experience any change the most appropriate indexing algorithm would be: a. Block sort-based indexing algorithm b. Single-pass in memory indexing algorithm c. Distributed Map-Reduce indexing algorithm d. Dynamic indexing process employing an auxiliary index
2	An index that includes sequences of words or terms of variable length that have been extracted from a source document is called a: a. Phrase Index b. Biword index c. Positional index d. Inverted Index
3	In order to access a particular web site in the internet, the URL must be converted into an IP address. Which service does this conversion?  a. HTTP b. TNS c. DNS

	d. DHCP
4	The number of times that a word or term occurs in a document is called the: a. Proximity Operator b. Vocabulary Lexicon c. Term Frequency d. Indexing Granularity
5	An approach to compression that takes advantage of the redundancy in the dictionary that results from common prefixes that come from sorted terms is called: a. Front Coding b. Blocked storage c. Prefix Coding d. Variable byte encoding
6	For a moderately large collection of static documents maintained on a single system the most appropriate indexing algorithm would be: a. Block sort-based indexing algorithm b. Single-pass in memory indexing algorithm c. Distributed Map-Reduce indexing algorithm d. Dynamic indexing process employing an auxiliary index
7	Results from a search engine that are based upon the retrieval of items using a method of term weighting such as cosine similarity is a form of: a. Sponsored Search b. Algorithmic Search c. Informational Search d. Navigational Search
8	Given two strings s1 and s2, the edit distance between them is sometimes known as the: a. Levenshtein distance b. isolated-term distance c. k-gram overlap d. Jaccard Coefficient
9	A web link within a web page that references another part of the same page is called a: a. Out link b. Vector c. In link d. Tendril
10	Which of the following is a technique for context sensitive spelling correction: a. the Jaccard Coefficient b. Soundex algorithms c. k-gram indexes d. Levenshtein distance

--	--

### UNIT 3

1	Operational information is needed for a)Day to day operations b)Meet government requirements c)Long range planning d)Short range planning
2	Data by itself is not useful unless a)It is massive b)It is processed to obtain information c)It is collected from diverse sources d)It is properly stated
3	For taking decisions data must be a)Very accurate b)Massive c)Processed correctly d)Collected from diverse sources
4	Measures of Similarity are as Follows : i. The lengths of the Documents. ii. The number of terms in common. iii. Whether the terms are common or unusual. iv. How many times each term appears.  a)i) & ii) b)ii) & iii) c)iii) & iv) d)i),ii),iii) & iv)
5	one of the application of Personalized Search is, a)Google b)Yahoo c)IBM d)Alpha Search Engine
6	Structure of Web has following entities: i. Web Graph ii. Static and Dynamic Pages iii. Hidden web pages iv. Size of web page

	<p>a)i) &amp; ii)  b)i) &amp; ii)  c)iii) &amp; iv)  d)i),ii),iii) &amp; iv)</p>
7	<p>Goal of IR is to find documents _____ to an information need from a large document set</p> <p>a. applicable  b. Relevant  c. knowledgeable  d. none of the above</p>
8	<p>Static collection of documents is known as _____.</p> <p>a. Information  b. corpus  c. data  d. None of the above</p>
9	<p>Information retrieval systems have much in common with_____.</p> <p>a. Filing systems  b. Transaction system  c. Database systems  d. Management systems</p>
10	<p>A data structure that maps terms back to the parts of a document in which they occur is called as_____.</p> <p>a. Postings list  b. Incidence Matrix  c. Dictionary  d. Inverted Index</p>