FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

DEPARTMENT OF INFORMATION AND MEDIA TECHNOLOGY

SECOND SEMESTER 2017/2018 EXAMINATION

COURSE CODE:	IMT524
COURSE TITLE:	INFORMATION RETRIEVAL SYSTEMS
CREDIT UNITS:	3 UNITS
TIME ALLOWED:	2hrs 30min
COURSE LECTURER(S):	Dr. I. O. Oyefolahan and Mrs Stella Etuk
NUMBER OF QUESTIONS:	4
NUMBER OF PAGES:	2 (INCLUDING THIS PAGE)

INSTRUCTIONS

- Answer all questions
- Do not use red pen
- Please use a clear handwriting
- This exam is closed book. closed notes. closed laptop and closed cell phone
- Please use non-programmable calculators only

Question 1.

(a) Inverted indexes are considered the most efficient and flexible index structure and thus, all modern search engine indexes are based on inverted indexes. Explain the concept of an inverted index.

(b) Why is an inverted index with counts better than ordinary document with posting for the document ranking model?

(c) When a phrase is being queried such as "Information System", what type of inverted index is the best predictor of a relevant document and why?

Question 2.

(a) Why is it important to evaluate a search engine?

(b) One of the primary distinctions made in the evaluation of search engines is between effectiveness and efficiency. Differentiate both as related to search engines

(c) The two most common effectiveness measures for evaluating a search engine are recall and precision. Explain

Question 3.

(a) Explain the processes involved in retrieving web pages from the web servers by web crawlers.

'The task of the document statistics component is simply to gather and record statistical information about words, features, and documents. This information is used by the ranking component to compute scores for documents. The types of data generally required are the counts of index term occurrences in individual documents, the positions in the documents where the index terms occurred, the counts of occurrences over groups of documents and the lengths of documents in terms of the number of tokens. The actual data required is determined by the retrieval model and associated ranking algorithm'

(b) Considering the word collection given above as a document, show in a tabular form using Zipf's law the first 6 most common words in the document together with their frequency (f), rank (r), probability of occurrence (Pr) and its (r. Pr) values.

(c) From the table generated in 3b above, calculate the number of words with the same frequency given that a word that occur n times has a rank r_n

Question 4

Explain in detail the following important functions of a search engine architectural components:

(a) Indexing process

(b) Query process of Information Retrieval.