



FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA  
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY  
DEPARTMENT OF INFORMATION AND MEDIA TECHNOLOGY

SECOND SEMESTER 2016/2017 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:** 3 (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. Search and Communication is by far the most popular uses of the computer. Explain the following types of search as relation to Information Retrieval (IR).
  - i. Vertical Search
  - ii. Enterprise Search
  - iii. Desktop Search
  - iv. Peer-to-peer Search (2mks)
- b. What are the issues term "The Big Issues" in IR? Explain them (6mks)
- c. How did researchers try to address each of "The Big Issues" listed above? (3mks)
- d. An Architecture is designed to ensure that a System will satisfy its application requirements and goal. Explain the **two (2)** primary goals of a search engine (4mks)

Question 2.

- a. Briefly explain "Query transformation" (3mks)
- b. What do you understand by "Document Feeds" (4mks)
- c. Search Engine architectural components supports 2 major functions: Index processing and query processing. Explain the 'Index Processing'. (8mks)

Question 3.

- a. For crawling individual websites, a single computer is sufficient. However, crawling the entire web requires many computers devoted to crawling. Give **two (2)** reasons for the multiple computers, (4mks)
- b. Briefly explain what the following in relation to web crawling
  - i. Freshness
  - ii. Focused Crawling
  - iii. Deep web (6mks)
- c. Explain how 'Checksumming' technique is used to detect duplicate documents (5mks)

Question 4

- a. **Tokenization** is a technique in Information Retrieval (IR) that involves splitting of words apart and stripping punctuation from words to make them easier to find or retrieve. Explain **five (5)** likely problems associated with tokenization. (5mks)



“The textbook on Information Retrieval is designed to help people understand search engines, evaluate and compare them, and modify them for specific applications. A Search for information on the Web is, for most people, a daily activity. Search and communication are by far the most popular uses of the computer. Not surprisingly, many people in companies and universities are trying to improve search by coming up with easier and faster ways to find the right information. These people are working in the field of Information Retrieval”

- b. Considering the word collection given above as a document, show in a tabular form using Zipf's law
- the first 8 most common words in the document
  - frequency ( $f$ ),
  - rank ( $r$ ),
  - probability of occurrence ( $Pr$ ) and its
  - Calculate the number of words with the same frequency 4

(10mks)