# Federal University of Technology Minna
## School of Information and Communication Technology
## Department of Information and Media Technology
## 2013/2014 Second Semester Examination

**Course Title:** Information Retrieval Systems – 2units **Course Code:** IMT 524
**Instruction:** Answer only **Four** (4) questions **Time Allowed:** 2 hours

**Question 1**

a. Describe the key components of an Information Retrieval Systems (IRS)
b. Elucidate on the concept that requires each document to have one or more descriptive attribute.
c. Use the table below to answer questions i & ii:

| P/R | Query 1 | Query 2 | Query 3 | Query 4 | Query 5 |
|---|---|---|---|---|---|
| System A | 0.9/0.1 | 0.7/0.4 | 0.45/0.5 | 0.3/0.6 | 0.1/0.8 |
| System B | 0.8/0.2 | 0.5/0.3 | 0.4/0.5 | 0.3/0.7 | 0.2/0.8 |
| System C | 0.9/0.4 | 0.7/0.6 | 0.5/0.7 | 0.3/0.8 | 0.2/0.9 |

i. Plot a Precision – Recall diagram for the three (3) systems.
ii. Using a fixed interval value of 0.25, 0.5 and 075 for recall what will be the values of precision for system 1, 2 and 3

**Question 2**

a. Mention any five (5) challenges of precision and recall in evaluating an IRS
b. Describe the major differences between computerised and inverted index
c. i. Briefly explain what you understand by the term vector and vector representation of document space?
ii. What is the mathematical representation of the two (2) terms respectively?
d. Discuss in detail the concept and relationship between recall and precision

**Question 3**

a. Discuss the concept of Abstraction Principle as it relates to IRS
b. As IRS specialist, how will you identify a functionaldocument model?
c. There are some features that are unpublished by search engines, identify any three (3) of these features that differentiate google, altavista and yahoo search engines.
d. A database contains 80 records on a particular topic, a search was conducted on that topic and 60 records were retrieved; of the 60 records retrieved, and 45 were relevant. Calculate the precision and recall scores for the search.

**Question 4**

a. Distinguish between data, information and knowledge in the context of IRS
b. What is quadruple in a formal IRSmodels and what are the parameters
c. Discuss the merit and demerit of a probabilistic model of an IRS
d. Discuss any four (4) advance IRS that is adopted by search engines

**Question 5**

a. Outline any five difference between data retrieval and information retrieval
b. Use the keys and table below to answer questionsi-iii:

**TN / True Negative:** case was negative and predicted negative,
**TP / True Positive:** case was positive and predicted positive,
**FN / False Negative:** case was positive but predicted negative,
**FP / False Positive:** case was negative but predicted positive

| | Predicted Negative | Predicted Positive |
|---|---|---|
| Negative Cases | TN: 9,760 | FP: 140 |
| Positive Cases | FN: 40 | TP: 60 |

i. What percent of your predictions were correct?
ii. What percent of the positive cases did you catch?
iii. What percent of positive predictions were correct?

c. An IRS was deployed for your institution, outline the techniques you will use to assess the system
d. What are the likely problems you will encounter when implementing boolean search for you IRS

PTO

**Question 6**

a. Discuss in detail why we need information retrieval systems
b. Elucidate on any five (5) issues affecting information retrieval systems
c. Briefly discuss in any three (3) examples of advance information retrieval system