

Registration No:

--	--	--	--	--	--	--	--	--	--

Total Number of Pages: 02

B.Tech.  
PIT5H004

**5<sup>th</sup> Semester Regular Examination 2017-18**

**Information Retrieval**

**BRANCH: IT**

**Time: 3 Hours**

**Max Marks: 100**

**Q.CODE: B479**

**Answer Question No.1 and 2 which are compulsory and any four from the rest.  
The figures in the right hand margin indicate marks.**

**Q1 Answer the following questions: *multiple type or dash fill up type* (2 x 10)**

- a) Which of the following is necessary for effective information retrieval?
  - (i) The database is well designed.
  - (ii) The query designer adequately understands the desired output.
  - (iii) The query designer has a thorough knowledge of the database table structures and the nature of the data in the tables.
  - (iv) All of the above
- b) Consider the following sentence "I left my left food in San Francisco". How many 2-grams does the sentence have?
  - (i) 2 (ii) 4 (iii) 6 (iv) 7
- c) TF-IDF is \_\_\_\_\_ when t occurs in virtually all documents.
  - (i) Low (ii) Medium (iii) High (iv) Not applicable
- d) For o'neill, which of the following is the desired tokenization?
  - (i) o'neill (ii) o neill (iii) o' neill (iv) All of the above
- e) Which of the following is odd?
  - (i) be (ii) with (iii) us (iv) has
- f) Which process may be thought of as the inverse of information retrieval?
  - (i) Selective dissemination of information
  - (ii) Store dissemination of information
  - (iii) Selective dissemination of index
  - (iv) Store dissemination of index
- g) TF-IDF weight of a term is the \_\_\_\_\_ of its tf weight and its idf weight.
  - (i) sum (ii) difference (iii) product (iv) ratio
- h) Accuracy = \_\_\_\_\_.
  - (i)  $(FP + FN) / (TP + TN + FP + FN)$
  - (ii)  $(TP + TN) / (TP + TN + FP + FN)$
  - (iii)  $(FP + TN) / (TP + TN + FP + FN)$
  - (iv)  $(FN + TP) / (TP + TN + FP + FN)$
- i) Automated information retrieval systems are used to reduce \_\_\_\_\_.
  - (i) Attention economy (ii) Information pollution (iii) Continuous partial attention
  - (iv) Information overload
- j) How to calculate the Jaccard index?
  - (i)  $((\text{the number in both sets}) / (\text{the number in either set})) \times 100$
  - (ii)  $((\text{the number in either set}) / (\text{the number in both sets})) \times 100$
  - (iii)  $(\text{the number in either set}) \times (\text{the number in both sets}) \times 100$
  - (iv) None of the above

**Q2 Answer the following questions: *Short answer type* (2 x 10)**

- a) List out the components of information retrieval block diagram and Explain.
- b) Define unigram language.
- c) What are the disadvantages of Boolean model?
- d) List the differences between data retrieval and information retrieval.
- e) What are the desirable properties of a clustering algorithm?
- f) What is web crawler? How it is different from focused crawler?

- g)** What are the characteristics of relevance feedback?
  - h)** Define precision. Give a numerical example.
  - i)** Define term frequency.
  - j)** Explain latent semantic indexing.
  
- Q3**
  - a)** Define information retrieval. Explain the objectives of the information retrieval system. **(10)**
  - b)** Explain the functional overview of the information retrieval system. **(5)**
  
- Q4**
  - a)** What is stemming algorithm? Explain how stemming algorithm helps to improve the recall. **(10)**
  - b)** Describe the similarities and differences between term stemming algorithm and n-grams. **(5)**
  
- Q5**
  - a)** What is automatic indexing? Briefly describe data flow in information processing system. **(10)**
  - b)** Explain concept indexing and hypertext linkages. **(5)**
  
- Q6**
  - a)** Write about information visualization technologies. **(10)**
  - b)** What is TREC results? How it can be used in information system evaluation? **(5)**
  
- Q7**
  - a)** Explain various software text search algorithms. **(10)**
  - b)** Compare and contrast hardware and software text search algorithms. **(5)**
  
- Q8**
  - a)** Discuss various methods for generation of a Thesaurus. **(10)**
  - b)** Write short notes on similarity measures and ranking. **(5)**
  
- Q9**
  - a)** Describe various hypertext and XML data structures. **(10)**
  - b)** Compare and contrast term clustering and item clustering. **(5)**