The innovation driven E-School

**Format No. QSP/7.1/01.F01 (B)** **Issue No.05 Rev. No 5 Dated: Jan 1, 2017**

---

**UNIVERSITY OF PETROLEUM & ENERGY STUDIES**

**College of Engineering Studies**

**Dehradun**

**COURSE PLAN**

Programme : B. Tech – (BAO)

Course : Information Retrieval and Search Engines

Course Code : CSEG-393

No. of credits : 3

Semester : V

Session : 2017-18

Batch : 2015-19

Prepared by : Dr. Durgansh Sharma

Email : durgansh.sharma@ddn.upes.ac.in

**Approved By**

HOD/ Prog. Head

UPES Campus
"Energy Acres"
P.O. Bidholi, , Dehradun

Tel : +91-135-2770137
Fax : +91 135- 27760904

# COURSE PLAN

A. **PREREQUISITE:**

a. Basic Knowledge of DBMS, Data Structures, NLP.
b. Basic Knowledge of Set Theory and Statistics.

**B. PROGRAM OUTCOMES (POs) and PROGRAM SPECIFIC OUTCOMES (PSOs) for Business Analytics and Optimization (BAO):**

**B1. PROGRAM OUTCOMES (POs)**

**B2. Program Specific Outcomes (PSOs)**

**C. COURSE OUTCOMES FOR AUTOMOTIVE TRANSMISSION SYSTEMS: At the end of this course student should be able to**

**Table: Correlation of POs and PSOs v/s COs**

| PO/CO | PO 1 | PO 2 | PO 3 | PO 4 | PO 5 | PO 6 | PO 7 | PO 8 | PO 9 | PSO 1 | PSO 2 | PSO 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CO1 | | | | | | | | | | | | |
| CO2 | | | | | | | | | | | | |
| CO3 | | | | | | | | | | | | |
| CO4 | | | | | | | | | | | | |
| CO5 | | | | | | | | | | | | |
| CO6 | | | | | | | | | | | | |

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

**D. PEDAGOGY**

- **Presentation,**
- **flipped classroom session,**
- **think-pair and share,**
- **youtube videos as a startup**

**E. COURSE COMPLETION PLAN**

| | |
|---|---|
| **Total Class room** | 36 |
| **Total Quizzes** | 02 |
| **Total Test** | 02 |
| **Total Assignment** | 02 |

One Session =60 minutes

## F. EVALUATION & GRADING

Students will be evaluated based on the following 3 stages.

5.1 Internal Assessment - 30%
5.2 Mid-term Examination - 20%
5.2 End term Examination - 50%

**F1. INTERNAL ASSESSMENT: WEIGHTAGE – 30%**

Internal Assessment shall be done based on the following:

| Sl. No. | Description | % of Weightage out of 30% |
|---|---|---|
| 1 | Class Tests | 30% |
| 2 | Quizzes | 20% |
| 3 | Assignments (Problems/Presentations) | 30% |
| 4 | Attendance and Class Participation | 20% |

**F2.** ***Internal Assessment Record Sheet (including Mid Term Examination marks)*** *will be displayed online at the end of semester i.e. last week of regular classroom teaching.*

**F3. CLASS TESTS/QUIZZES:** Two Class Tests based on descriptive type theoretical & numerical questions and Two Quizzes based on objective type questions will be held; one class test and one quiz at least ten days before the Mid Term Examination and second class test and second quiz at least ten days before the End Term Examination. Those who do not appear in quiz examinations shall lose their marks.

*The marks obtained by the students will be displayed on LMS a week before the start of Mid Term and End Term Examinations respectively.*

**F4. ASSIGNMENTS:** After completion of each unit or in the mid of the unit, there will be home assignments based on theory and numerical problems. Those who fail to submit the assignments by the due date shall lose their marks.

**F5. GENERAL DISCIPLINE:** Based on student's regularity, punctuality, sincerity and participation in the interactions.

*The marks obtained by the students will be displayed on LMS at the end of semester.*

**F6. MID TERM EXAMINATION: WEIGHTAGE – 20%**

Mid Term examination shall be Two Hours duration and shall be a combination of Short and Long theory Questions.

***Date of showing Mid Term Examination Answer Sheets: Within a week after completion of mid Sem examination.***

**F7. END TERM EXAMINATION: WEIGHTAGE – 50%**

End Term Examination shall be Three Hours duration and shall be a combination of Short and Long theory/numerical Questions.

**F8. GRADING:**

The overall marks obtained at the end of the semester comprising all the above three mentioned shall be converted to a grade.

## G. COURSE DELIVERY PLAN

| TOPICS/SUB TOPICS | NO. OF SESSION | Course Outcomes Addressed | Assignment(s)/Quizzes/ Tests |
|---|---|---|---|
| **UNIT 1: Introduction to Information Retrieval** | | CO1 | |
| Basic Concepts | 6 | | Assignment – 1 |
| Retrieval Process, Modeling | | | |
| Classic Information Retrieval | | | |
| Set Theoretic, Algebraic and Probabilistic Models | | | |
| Structured Text Retrieval Models | | | |
| Retrieval Evaluation, Word Sense Disambiguation | | | |
| **UNIT 2: Querying** | | | |
| Languages, Keyword based querying | 6 | CO2 | |
| Pattern Matching, Structural Queries | | | Quiz -1 |
| Query Operations | | | |

| | | | |
|---|---|---|---|
| User Relevance Feedback, Local and Global analysis | | | |
| Text and Multimedia Languages | | | |
| **UNIT 3: Text Operations and User Interface** | | | |
| Document Preprocessing, Clustering, Text compression | 7 | CO3 | |
| Indexing and Searching, Inverted files, Boolean Queries | | | |
| Sequential Searching, Pattern Matching, | | | |
| User Interface and Visualization, Human Computer Interaction, Access Processes | | | **Test-1** |
| Starting Points, Query Specifications, Context-User Relevance Judgement, Interface for search | | | |
| **Mid TERM** | | | |
| **UNIT-4: Multimedia Information Retrieval** | | | |
| Data Models, Query Languages | 6 | | **Assignment-2** |
| Spatial Access Models, Generic Approach | | | |
| One Dimensional Time Series | | | |
| Two Dimensional Color Images, Feature Extraction | | | |

| | | | |
|---|---|---|---|
| **Unit 5: Search Engine** | | | |
| Overview of Web search, Web crawler | 7 | CO4 | |
| Map reduce and Web indexing | | | |
| Link analysis (Pagerank, HITS), Learning to Rank | | | **Quiz-2** |
| Content based recommendations, collaborative filtering | | | |
| **Unit 6: Applications** | | | |
| Searching the web – challenges, Characterizing the Web | 6 | CO4 | |
| Search Engines, Browsing, Meta Searchers | | | |
| Online IR Systems, Online Public Access Catalogs, | | | Test-2 |
| Digital Libraries, Architectural Issues, Document Models | | | |
| Representations and Access, Prototypes and Standards | | | |
| **Revision** | 03 | | |
| **43** | | | |

## H. SUGGESTED READINGS:

### H1. TEXT BOOK:

**Books:** "Modern Information Retrieval" by Ricardo Baeza-Yate, Pearson Education Asia.

"Information Retrieval: Algorithms and Heuristics" by David A. Grossman, Ophir Frieder, Academic Press.

### H2. REFERENCE BOOKS:

**E-Book:** "An Introduction to Information Retrieval" by Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze

## H3. OTHER RESOURCES

Web Link :

1. https://nlp.stanford.edu/IR-book/

## H4. VIDEO RESOURCES:

1. NPTEL Lectures –will be available - \\10.2.1.33 (intranet)
2. **VIDEO RESOURCES: - PPT & VIDEO - Info during Course Curriculum**

## I. GUIDELINES

***Cell Phones and other Electronic Communication Devices*:** Cell phones and other electronic communication devices (such as Blackberries/Laptops) are not permitted in classes during Tests or the Mid/Final Examination. Such devices MUST be turned off in the class room.

***E-Mail and online learning tool:*** Each student in the class should have an e-mail id and a pass word to access the LMS system regularly. Regularly, important information – Date of conducting class tests, guest lectures, via online learning tool. The best way to arrange meetings with us or ask specific questions is by email and prior appointment. All the assignments preferably should be uploaded on online learning tool. Various research papers/reference material will be mailed/uploaded on online learning platform time to time.

***Attendance:*** Students are required to have **minimum attendance of 75%** in each subject. Students with less than said percentage shall **NOT** be allowed to appear in the end semester examination.

***Passing criterion:*** Student has to secure minimum 30%/40% marks of the "highest marks in the class scored by a student in that subject (in that class/group class)" individually in both the 'End-Semester examination' and 'Total Marks' in order to pass in that paper.

- Passing Criterion for B. Tech: Minimum 30% and 40% of the highest marks in the class applicable to the students admitted before July 2015 and onwards July 2015 respectively.
- Passing Criterion for M. Tech: minimum 40% of the highest marks in the class

## J. Course outcome assessment

To assess the fulfilment of course outcomes two different approaches have been decided. Degree of fulfillment of course outcomes will be assessed in different ways through direct assessment and indirect assessment. In Direct Assessment, it is measured through quizzes, tests, assignment, Mid-term and/or End-term examinations. It is suggested that each examination is designed in such a way that it can address one or two outcomes (depending upon the course completion). Indirect assessment is done through the student survey which needs to be designed by the faculty (sample format is given below) and it shall be conducted towards the end of course completion. The evaluation of the achievement of the Course Outcomes shall be done by analyzing the inputs received through Direct and Indirect Assessments and then corrective actions suggested for further improvement.

**Sample format for Indirect Assessment of Course outcomes**

| |
|---|
| NAME: |
| *ENROLLMENT NO*: |
| SAP ID: |
| COURSE: |
| PROGRAM: |

Please rate the following aspects of course outcomes of Information Retrieval and Search Engines.
Use the scale 1-4*

| Sl. No. | | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 1 | CO1. Classify and exploit mathematical information retrieval models | | | | |
| 2 | CO2. Importance of Query and Query Structures in Information Retrieval | | | | |
| 3 | CO3. Understanding the importance of Preprocessing in Text Operations and User Interface | | | | |
| 4 | CO4. Apply Information Retrieval Models in the domain of Multimedia | | | | |
| 5 | CO5. Understand and formulate the searching and ranking methods | | | | |
| 6 | CO6. Explore and Understand the Application, Challenges and Future Scope in Information Retrieval Systems | | | | |

* 1 Below Average 3 Good

2 Average 4 Very Good