UE21CS342BA2: Algorithms for Information Retrieval and Intelligence Web (4-0-0-4-4)

|  |  |
| --- | --- |
| **Project Work (A group consisting of two members)** | **8 (It will be conducted for 20 marks but scaled down to 8 marks)** |

1. Implementation of a Basic Boolean Retrieval System
2. Developing an Extended Boolean Model for Information Retrieval
3. Construction of an Inverted Index and Dictionary for Document Retrieval
4. Building a Positional Indexing System for Phrase Queries
5. Implementing Tolerant Retrieval Techniques in Information Retrieval
6. Index Compression Techniques and Evaluation in Information Retrieval Systems
7. Implementation of Vector Space Model for Document Scoring
8. Utilizing tf-idf and Its Variants for Document Ranking
9. Comparative Analysis of Retrieval Performance Using Different IR Models
10. Designing and Evaluating an Information Retrieval System for a Specific Domain
11. Implementation of a Simple Scoring and Ranking System for Document Retrieval
12. Designing and Evaluating Parametric and Zone Indexes for Efficient Information Retrieval
13. Construction of Tiered Indexes for Improved Scalability in Web Search
14. Development of a Query Parser for Parsing User Queries in a Search Engine
15. Investigating Query Term Proximity Techniques for Enhanced Relevance in Search Results
16. Aggregating Scores from Multiple Ranking Factors for Document Ranking
17. Performance Measurement and Evaluation of Information Retrieval Systems
18. Building a Web Application with Search Algorithms for Document Retrieval
19. Experimenting with Relevance Feedback Techniques for Iterative Search Improvement
20. Implementing Query Expansion Methods to Enhance Search Results in Information Retrieval Systems
21. Implementation of PageRank Algorithm for Link Analysis in Web Search
22. Building a Web Crawler and Indexing System for Information Retrieval
23. Utilizing Lucene as a Search Engine for Document Retrieval
24. Comparative Analysis of Search Engines: Lucene, Solr, and Google
25. Developing a Multimodal Information Retrieval System with Content-Based Visual Information Retrieval
26. Implementing Meta Data Searching in Multimodal Information Retrieval Systems
27. Designing Query by Example Feature for Visual Information Retrieval
28. Semantic Retrieval Techniques in Multimodal Information Retrieval
29. Application of Machine Learning Approaches in Multimodal Information Retrieval
30. Evaluation of Content Comparison using Image Distance Measures in Multimedia Retrieval
31. Implementation of a Factoid Question Answering Model using Neural Networks
32. Building an Entity Linking Model for Question Answering Systems
33. Designing a Knowledge-based Question Answering System using Neural Models
34. Utilizing Pretrained Language Models for Question Answering in Information Retrieval
35. Comparative Analysis of Factoid QA Models for Information Retrieval Tasks
36. Development of a Neural Ranking Model for Document Retrieval
37. Implementation of Neural Models for Passage Ranking in Question Answering
38. Experimenting with Neural Retrieval Models for Information Retrieval Tasks
39. Fine-tuning Pretrained Language Models for Question Answering on Specific Domains
40. Evaluation of Neural Models for Information Retrieval using Benchmark Datasets

There are numerous sources where you can find machine learning datasets for various purposes. Here are some popular ones:

1. **UCI Machine Learning Repository**: One of the oldest and most widely used repositories for machine learning datasets. It covers a wide range of domains and includes both real-world and synthetic datasets.
2. **Kaggle**: Kaggle hosts a large collection of datasets contributed by the community. It covers diverse topics ranging from image classification to natural language processing.
3. **Google Dataset Search**: Google's search engine dedicated to finding datasets. It aggregates datasets from various sources across the web, including government websites, academic institutions, and other data repositories.
4. **OpenML**: OpenML is an open platform where researchers can share datasets, algorithms, and experiments. It hosts a vast collection of datasets and provides tools for collaborative research and experimentation.
5. **GitHub**: Many researchers and organizations share datasets on GitHub repositories. You can search for datasets using GitHub's search feature or explore curated lists of machine learning datasets.
6. **Government Websites**: Government agencies often publish datasets related to public health, transportation, economics, and other domains. Websites like data.gov (for the United States) and data.gov.uk (for the United Kingdom) are good starting points.
7. **Academic Repositories**: Many universities and research institutions maintain repositories of datasets used in published research papers. You can find datasets related to specific research areas by exploring the repositories of relevant institutions.
8. **ImageNet**: ImageNet is a large-scale dataset for visual object recognition. It contains millions of labeled images across thousands of categories and has been widely used for training deep learning models.
9. **Natural Language Processing (NLP) Datasets**: Websites like the Natural Language Processing Archive (NLP Arch) and the Stanford NLP Group provide datasets specifically curated for NLP research.
10. **Social Media APIs**: Social media platforms like Twitter and Facebook provide APIs for accessing public data. You can collect datasets of social media posts, user interactions, and other relevant information through these APIs.