**PSG COLLEGE OF TECHNOLOGY**

**Department of Applied Mathematics and Computational Sciences**

**IX MSc SS**

**Information Retrieval Lab**

**Problem Sheet 3**

In this assignment, implement three different ranking functions and four different evaluation functions on TIME dataset. (<https://ir.dcs.gla.ac.uk/resources/test_collections/time/>).

It contains binary relevance judgements for 83 queries on a collection of 423 documents.

Build an inverted index using Lucene (Java) - <https://lucene.apache.org/> or PyLucene(Python) -<https://lucene.apache.org/pylucene/install.html> or Elasticsearch(Python) - <https://pypi.python.org/pypi/elasticsearch>.

Compare the following ranking functions:

1. **Cosine similarity:** Convert each query and document to tf-idf and sort the documents according to the cosine similarity between query and document.
2. **BIM:** Calculatetrk score to calculate probability of documents to rank
3. **BM25**:
4. For each system, compute the following evaluation metrics:

* Precision
* Recall
* F1
* Mean Average Precision (MAP)