

Manipal Academy of Higher Education

**Impressions@MAHE**

---

Manipal Institute of Technology, Manipal  
Theses and Dissertations

MAHE Student Work

---

Summer 7-1-2020

## **Data Retrieval, Similarity Check and Analysis using Natural Language Processing Concepts**

Angad Singh

Follow this and additional works at: <https://impressions.manipal.edu/mit>



Part of the [Computer Sciences Commons](#)

---

## Data Retrieval, Similarity Check and Analysis using Natural Language Processing Concepts

Contracts are an essential aspect of any deal. They contain the clauses, the rights of the parties involved as well as the metrics needed. Other documents, such as tenders, ledgers, and license agreements, are no different. Currently, most companies do not have a central repository where they store them. They miss an intelligent system for querying, comparing, and fetching relevant documents and snippets for easy access as required by the companies. Hence, KPMG India decided to address this gap and develop an intelligent contract management system that will enable the user to perform all the tasks mentioned above and more. The project entailed developing an end-to-end solution for ingesting documents, storing them in a hierarchical format, retrieving answers to user queries, and comparing document similarity using Natural Language Processing techniques.

**[Computing Methodologies]:** Artificial Intelligence – Natural language processing, Information extraction

**[Information Systems]:** Information retrieval – Retrieval tasks and goals, Search engine architectures and scalability

**[Applied Computing]:** Document management and text processing – Document searching, Document management, Document capture, Optical character recognition

**Angad Singh**

**160953180**

**B.Tech(CCE)**

**Guide: Dr.Manohara Pai M. M**