Please provide the below content related to the project codebase and report.

1 Project Codebase

Please provide the following regarding the project codebase (**Note**: Python 3.10+/sklearn 1.2+/Scrapy 2.11+/Flask 2.2+):

- A **Scrapy** based **Crawler** for downloading web documents in *html* format - content crawling:

  - Required: Initialize using seed URL/Domain, Max Pages, Max Depth
  - Optional: Concurrent crawling (**AutoThrottle**), Distributed crawling (**scrapyd**),

- A **Scikit-Learn** based **Indexer** for contructing an inverted index in *pickle* format - search indexing:

  - Required: TF-IDF score/weight representation, Cosine similarity
  - Optional: Vector embedding representation (**word2vec**), Neural/Semantic search kNN similarity (**FAISS**)

- A **Flask** based **Processor** for handling free text queries in *json* format - query processing:

  - Required: Query validation/error-checking, Top-K ranked results
  - Optional: Query spelling-correction/suggestion (**NLTK**), query expansion (**WordNet**)

2 Project Report

Please provide the following regarding the project report (**Note**: Can be generated from the *Project Codebase*):

- Abstract - Development summary, objectives, and next steps.

- Overview - Solution outline, relevant literature, proposed system.

- Design - System capabilities, interactions, integration.

- Architecture - Software components, interfaces, implementation.

- Operation - Software commands, inputs, installation.

- Conclusion - Success/Failure results, outputs, caveats/cautions.

- Data Sources - Links, downloads, access information.

- Test Cases - Framework, harness, coverage.

- Source Code - Listings, documentation, dependencies (open-source).

- Bibliography - Reference citations (Chicago style - AMS/AIP or ACM/IEEE).