1. **Base Crawler** implements web crawler that crawls Wikipedia starting at given seed in BFS/DFS manner to collect web pages.
2. **LinkAnalysis&PageRank** calculates page rank scores for web pages by applying Link Analysis and PageRank Algorithm as random surfer.
3. **Inverted Indexer** generates the corpus by parsing web pages and creates inverted indexes. Analyzes Zipfian curve on vocabulary.
4. **Lucene** ranks documents using Lucene Analyzer as well as Vectors Space Cosine Similarity ranking to return top documents in response of a query**.**

**Inverted Indexer** is the folder, which provides inverted index “file.out” and document length “tokenPerDoc.out” to the **Lucene**.

* Both works separately, Lucene has those 2 data structure files in its own folder.
* Want to build structures again? : Inverted Indexer project -> Menu Option 1
* Otherwise, use directly in Lucene, already provided.