**Phase 1 RocksDB Database Schema**

Group 19

|  |  |
| --- | --- |
| **RocksDB Name** | **Mapping** |
| idToPageInfo | pageID -> page<page title, URL, last modified date, size>   * pageID: String (directly converted by URL) * page: PageInfo’s basic information   + PageInfo is a class which contains:   + title: String   + url: String   + modDate: Date   + size: int * Given a pageID, the corresponding PageInfo’s basic information is returned.   Explanation/Notes:   * The PageInfo class is used to store the information of a web page. Accessing these information directly from pageID is easy by using this database. |
| idToLinks | pageID -> pageID…, pageID…   * pageID: String (directly converted by URL) * Given a specific pageID, the corresponding parent pageID and children pageID is returned. Different parent pageID and children pageID separated by common and parent pageID and children pageID separated by space |
| InvertedIndex  (separated by titleIndex and bodyIndex) | wordID -> {pageID, <word position>}   * wordID: String (directly converted by the Word) * idAndPosition: the pageID(URL) and corresponding word position   + <word position> is the positions of the wordID in this specific pageID. <word position> is represented using a String containing the positions separated by comma. * Given a specific wordID, the corresponding pageID and of pages and their positions contained in the page is returned. |
| forwardIndex | pageID -> {wordID, <word position>}   * pageID: String (directly converted by the URL) * wordAndPosition: the wordID(word) and corresponding word position   + <word position> is the positions of the specific wordID in this pageID. <word position> is represented using a String containing the positions separated by comma. * Given a specific pageID, the corresponding wordID of words and their positions contained in the page is returned.   Explanation/Notes:  Store the words appearing in a page along with their positions so that searching for phrases will be easier. |