**Database scheme design:**

1. *pageID* (URL, PageID)

* It is used for assigning a pageID for each retrieved website, which maps the page ID with the URL.

1. *pageUrl* (PageID, URL)

* This table uses the page ID to find out the corresponding URL. It’s useful when we want to find out the URL of a page. For example, when we need to find out the child link of a page, what we get from another jdbm is a pageID of the child page. We need to use this table to find out the URL and print out the result.

1. *pageInfo* (url, title, pageID, lastModification, pageSize)

* It stores the page details, such as page title, size and the last modification time with respect to the page ID.

1. *childLinks* (parentPageID, List<(childPageIDs)>)

* It stores a list of child page IDs of a page. We use it to find out what child pages do a page have, using the page ID of the parent page as the key.

1. *wordID* (Word, WordID)

* It maps the word with a specific wordID.

1. *idWord* (WordID, Word)

* Usage: To get the actual word given the word ID

1. *wordInfo* (wordID, {pageID, tf})

* This one will store the posting list (pageID, tf) with respect to the wordID, so that we can find out the pages which contain that word and the term frequency.

1. *pageWord* (pageID, {keywords})

* A forward index that can efficiently retrieve the term frequency of a word in a page from the *wordInfo* index.

The following JDBM is not implemented in Phase 1, but will be used in final submission:

1. *parentLinks* (PageID, List<(parentPageIDs)>)

* This one will be used when we need to print the parent page given a child page ID. A list of parent page ID will be stored in list type. The key is the page ID of the child page.