Documentation

1. Design of database scheme
   1. Forward index
      1. Key
         1. Page id
      2. Value
         1. Word id followed by word positions
         2. e.g.
         3. say 2231 is apple’s wid
         4. <wid> <pos> <pos> <pos>;<wid> ….
         5. 2231 1 2 3;8787 5 6 7
      3. Function
         1. allows rapid searching on words with their page positions contain inside the specific page
         2. allows easy update by changing only record related to the specific page
   2. Inverted index
      1. Key
         1. word id
      2. Value
         1. Page id
      3. Function
         1. Allows rapid searching of target page ID containing the specific word
   3. Mapping table from word ID to word
      1. Key
         1. Word id
      2. Value
         1. Word
      3. Function
         1. Allow use of word ID instead of the whole word which save memory space
   4. Mapping table from Page ID to page informations
      1. key
         1. PageID (int)
      2. value
         1. Page information includes
            1. Child links (string array)
            2. Size (int)
            3. URL (string)
            4. Modified (string)
            5. Title (string)
      3. Function
         1. Allow use of shorter word ID instead of word to save memory space
         2. Store page information such as last modified date to allow checking on update status of page easily