

Database folder that store all the following table: [main.db](#)

Database tables related to Indexer (5 tables):

Table name: **body_inverted_index**

Columns:

- wordId (format: uuid4)
- value (format: uuid4;frequency;position1,position2,...
uuid4;frequency;position1,position2,...)

Example:

<u>wordId</u>	<u>value</u>
11	34;1;456 275;3;75,80,221
23	97;2;45,992

Description: Given a word id, return the url id, word frequency, positions

Table name: **title_inverted_index**

Columns:

- wordId (format: uuid4)
- value (format: uuid4;frequency;position1,position2,...
uuid4;frequency;position1,position2,...)

Example:

<u>wordId</u>	<u>value</u>
11	34;1;456 275;3;75,80,221
23	97;2;45,992

Description: Given a word id, return the url id, word frequency, positions

Table name: **forward_index**

Columns:

- urlId (format: uuid4)
- value (format: uuid4 uuid4 ...)

Example:

<u>urlId</u>	<u>value</u>
101	4 70 91 301
200	2 67 9

Description: Given a url id, return the word id(s)

Table name: **word_to_id**

Columns:

- word (format: string)
- wordId (format: uuid4)

Example:

<u>word</u>	<u>wordId</u>
fruit	7
read	99

Description: Given a word, return the word id

Table name: **id_to_word**

Columns:

- wordId (format: uuid4)
- word (format: string)

Example:

<u>wordId</u>	<u>word</u>
fruit	7
read	99

Description: Given a word id, return the word

Database tables related to Spider (8 tables):

Table name: **url_to_id**

Columns:

- url (format: string)
- urlId (format: uuid4)

Example:

<u>url</u>	<u>urlId</u>
https://www.cse.ust.hk/~kwtleung/COMP4321/testpage.htm	75
https://www.cse.ust.hk/~kwtleung/COMP4321/ust_cse.htm	1001

Description: Given a url, return the url id

Table name: **id_to_url**

Columns:

- urlId (format: uuid4)
- url (format: string)
- Example:

<u>url</u>	<u>urlId</u>
https://www.cse.ust.hk/~kwtleung/COMP4321/testpage.htm	75
https://www.cse.ust.hk/~kwtleung/COMP4321/ust_cse.htm	1001

- Description: Given a url id, return the url
-

Table name: **crawled_page_to_id**

Columns:

- url (format: uuid4)
- urlId (format: uuid4)

Example:

<u>url</u>	<u>urlId</u>
https://www.cse.ust.hk/~kwtleung/COMP4321/testpage.htm	75
https://www.cse.ust.hk/~kwtleung/COMP4321/ust_cse.htm	1001

Description: Given a url, return the url id

Table name: **id_to_page_title**

Columns:

- urlId (format: uuid4)
- pageTitle (format: string)

Example:

<u>urlId</u>	<u>pageTitle</u>
75	Test page
1001	CSE department of HKUST

Description: Given a url id, return the page title

Table name: **id_to_last_modification_date**

Columns:

- urlId (format: uuid4)
- lastModificationDate (format: date)

Example:

<u>urlId</u>	<u>lastModificationDate</u>
75	17/12/2024 21:09:37
1001	3/3/2025 08:02:58

Description: Given a url id, return the last modification date

Table name: **id_to_page_size**

Columns:

- urlId (format: uuid4)
- pageSize (format: string)

Example:

<u>urlId</u>	<u>pageSize</u>
75	603
1001	108

Description: Given a url id, return the page size

Table name: **id_to_children_url_id**

Columns:

- urlId (format: uuid64)
- childrenUrlId (format: uuid64 uuid64 ...)
- Example:

<u>urlId</u>	<u>childrenUrlId</u>
75	7 71 546 7839
1001	938

- Description: Given a url id, return all the children url id. Note that if the given url has no more children (excluding cyclic that go back to any parents), they should not be shown in the table.
-

Table name: **id_to_parents_url_id**

Columns:

- urlId (format: uuid64)
- parentsUrlId (format: uuid64 uuid64 ...)
- Example:

<u>urlId</u>	<u>parentsUrlId</u>
57	1 551
1003	2

- Description: Given a url id, return all the parents url id. Since the TA may modify the testing webpage (i.e., hidden test cases), we CANNOT assume there is only ONE parent for each urlId. Note that if the given url has no more parent (excluding cyclic that go back to any children), they should not be shown in the table.