**PageInfo**  
Description: Contains information about each page.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| PageID | Long | Unique key ID for pages |
| PageTitle | String | Title of the page |
| URL | String | URL of the page |
| LastModificationDate | String | Date of the last modification |
| SizeofPage | Long | Size of the page |

Data:

|  |  |
| --- | --- |
| 1 | COMP4321 Project Website|[www.testing.net](http://www.testing.net)|11/03/2024|96 |

**PageURLMapping**  
Description: Maps URL to their respective PageID.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| URL | String | URL of the page |
| PageID | Long | Unique key ID for pages |

Data:

|  |  |
| --- | --- |
| [www.testing.net](http://www.testing.net) | 1 |

**PageChildMapping**  
Description: Maps child links to their respective parent pages.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| PageID | Long | ID of the parent page |
| ChildLink | String | URL for the child page |

Data:

|  |  |
| --- | --- |
| 1 | [www.testing2.net](http://www.testing2.net), [www.testing3.net](http://www.testing3.net) |

**InvertedBodyWord**  
Description: Stores inverted index for words found in the body of pages.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| WordID | Long | Unique identifier for body words |
| PageID | Long | ID of the page containing the word |
| Frequency | Long | Frequency of the word in the body |
| TFIDF | Float | TF-IDF value of the word |

Data:

|  |  |
| --- | --- |
| 1 | 1|5|2.1, 2|4|1.8 |

**BodyWordMapping**  
Description: Maps body words to the pages they appear in.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| PageID | Long | ID of the page |
| WordID | Long | ID of the body word |
| Frequncy | Integer | Frequency of the word |

Data:

|  |  |
| --- | --- |
| 1 | 1|3, 3|4, 6|5, |

**InvertedTitleWord**  
Description: Stores inverted index for words found in the title of pages.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| TitleWordID | Long | Unique identifier for title words |
| PageID | Long | ID of the page containing the word |
| Frequency | Long | Frequency of the word in the title |
| TFIDF | Float | TF-IDF value of the word |

Data:

|  |  |
| --- | --- |
| 3 | 1|5|2.1, 2|4|1.8 |

**TitleWordMapping**  
Description: Maps title words to the pages they appear in.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| PageID | Long | ID of the page |
| TitleWordID | Long | ID of the title word |

Data:

|  |  |
| --- | --- |
| 1 | 1, 3, 6 |

**WordMapping**  
Description: Maps body word IDs to their respective words.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| WordID | Long | ID of the body word |
| Word | String | Body word |

Data:

|  |  |
| --- | --- |
| 1 | Apple |