**Instruction**

This instruction will guide you through every step for lauching a submission run on the validation tool.

1. **To specify the Directories of the CJK Wikipedia Collection**.

The directory or the home of the CJK Wikipeida corpora will be the directory holding the ***zh*** , or ***ja*** , or ***ko*** folder, or all above folders unzipped from the downloaded Wikipedia collection. For example,

|  |
| --- |
| F001.png |
| Figure 1. the collections home |

So, /data/corpus/wikipedia/CJK/xml-v2 is the home directory of the CJK Wikipeida corpora.

1. **A dump of the Wikipedia collections**:

* For the first time using this tool, you must specify the directories of the Wikipedia collections by using the menu bar function ***Corpus*** under ***Utility*** (see Figure 2 to 6).
* Once the directories of the collections have been set up correctly, the tool will show the current directories for both collections (see Figure 7). You do not need to specify them again, except the directories have been changed.

1. Please place the Wikipedia DTD file (i.e. ***article.dtd***) just outside the directory of the package.

* As we can see each Wikipedia document has been specified the DTD, *<!DOCTYPE article SYSTEM "../article.dtd">*, the DTD file must be placed in the right location for the Wikipedia document to refer (see Figure 12 and 13).

1. **To Load a submission**:
2. **Sample Wikipedia submission runs**:

* Then you should select the function ***Load*** under ***Utility*** to load a submission. Once the loading is success, the tool will display all content you need to check the submission run.
* One sample submission XML files, ***SAMPLE\_A2F\_RUN\_01.xml***, have been prepared in the directory, ***Sample\_Submission***, for this instruction. You can use ***Load*** function to browse one file at a time.
* If the loading was not success or the tool did not display the content correctly, it means there are something going wrong with the submission XML format.

1. **Topic Pane, Link Pane, RHS (i.e. Table) Pane interaction**:

* Please refer to Figure 8 to 11 regarding to the area of each pane and the interaction scheme.
* You can click on an Anchor Text or Best Entry Point icon on the Topic Pane to just to the 1st respective outgoing or incoming target link.
* On the Link Pane, you can simply either right click to go to the *Next* link or left click to go back the *Previous* link.
* The RHS pane offers a convenient means for the user to navigate to any of the topic and the target link by just clicking on a Table Cell. This Table also show the Offset and Length for each Anchor and BEP.

1. **Outgoing and Incoming Links**:

* There are 2 menu item functions, ***Outgoing*** and ***Incoming***, are located under the menu function, ***Linking***.
* Incoming mode has been disabled because only outgoing links need to be specified.

1. **The detail interaction scheme is described below**.

**The Detailed Description**

|  |  |  |
| --- | --- | --- |
| 1. For the first time, the tool will ask you to specify the directories of your Wikipedia collection (see Figure 1). 2. Click **OK** to let the tool launch entirely 3. From the Menu bar, there is a menu item called **Utility**. Click on it and select the option, **Corpus**. 4. As you can see, the Link-the-Wiki Collections Dialog is open and you can browse and select the directories where the Wikipedia and TeAra collections were located.    * **Note01**: If you are only interested in either the Wikipedia tasks or the TeAra tasks, you still need to specify both of them, just make sure the one you are interested is specified correctly. Otherwise the tool will not be running correctly.    * **Note02**: For the Wikipedia collection, you must place/combine all relative XML files into a single folder according to its naming convention (see Figure 4). Therefore, there will be only one directory (e.g. *Wikipedia\_2009*) containing all these sub-folders (e.g. *001*, *002* …, *999*). For example, the file, ***1978308.xml***, must be placed into the folder, ***308***.    * **Note03**: For the TeAra collection, you must extract the ZIP file downloaded from the INEX site into a single folder (see Figure 5). Therefore, there is only one directory (e.g. *TeAra\_2009*) containing all these sub-folders.    * **Note**: What you need to select is that “only one directory” (e.g. *Wikipedia\_2009* and *TeAra\_2009*), then you will see the dialog show the absolute directory path for both collections (e.g. *E:\\INEX\_Collections\_Dump\TeAra\_2009*). Then click **OK** to go Next. 5. ***Once you have specified the directories of the collections, the tool will remember it and show it to you every time you launch the tool (see Figure 7). The directories only need to be changed when the collections have been moved.*** 6. The 2nd step you need to do is to load a submission XML file. From the Menu bar, there is a menu item called **Utility**. Click on it and select the option, **Load**.   A Window file browser will open for you to select a submission XML file one at a time. It means you can only load a submission file and display it on the tool each time.   1. As you can see from the Figure 8 to 11, they show the Wikipedia and TeAra submission for outgoing and incoming links respectively. | |  | | --- | | Figure 2. Specification of the collection directories |     Figure 2. Empty directories of the collections    Figure 6. Link-the-Wiki collections selection dialog    Figure 7. The CJK Wikipeida collections directory message    **Table Pane**  **Link Pane**  **Topic Pane**  Figure 8. The crosslink submission |
| 1. You can click on either an *Anchor Text* or *BEP Icon* on the Topic pane to display its relative links showing on the Link Pane. 2. *Right-Click* on the Link Pane will move to the Next link while *Left-Click* on the Link Pane will be one link backward. 3. You can also click on a Table Cell on the RHS Pane to jump into that particular link.   According to the **Row** you click, the tool will display the corresponding topic and link on the respective pane.   1. The general information (e.g. *topic title*, *anchor name* and *target document title*) will be showed on the top banner just below the menu bar.   From the RHS Table Pane, you can see the *Offset* and *Length* of each anchor and BEP. According to this information, you can adjust your system to correct the position of each text anchor and best entry point. A general idea is that althought an exact “hit” of position (i.e. offset and length) is perfect, these anchors and BEPs are NOT expected to match exactly their respective position on the document but nearby or overlapping. | |

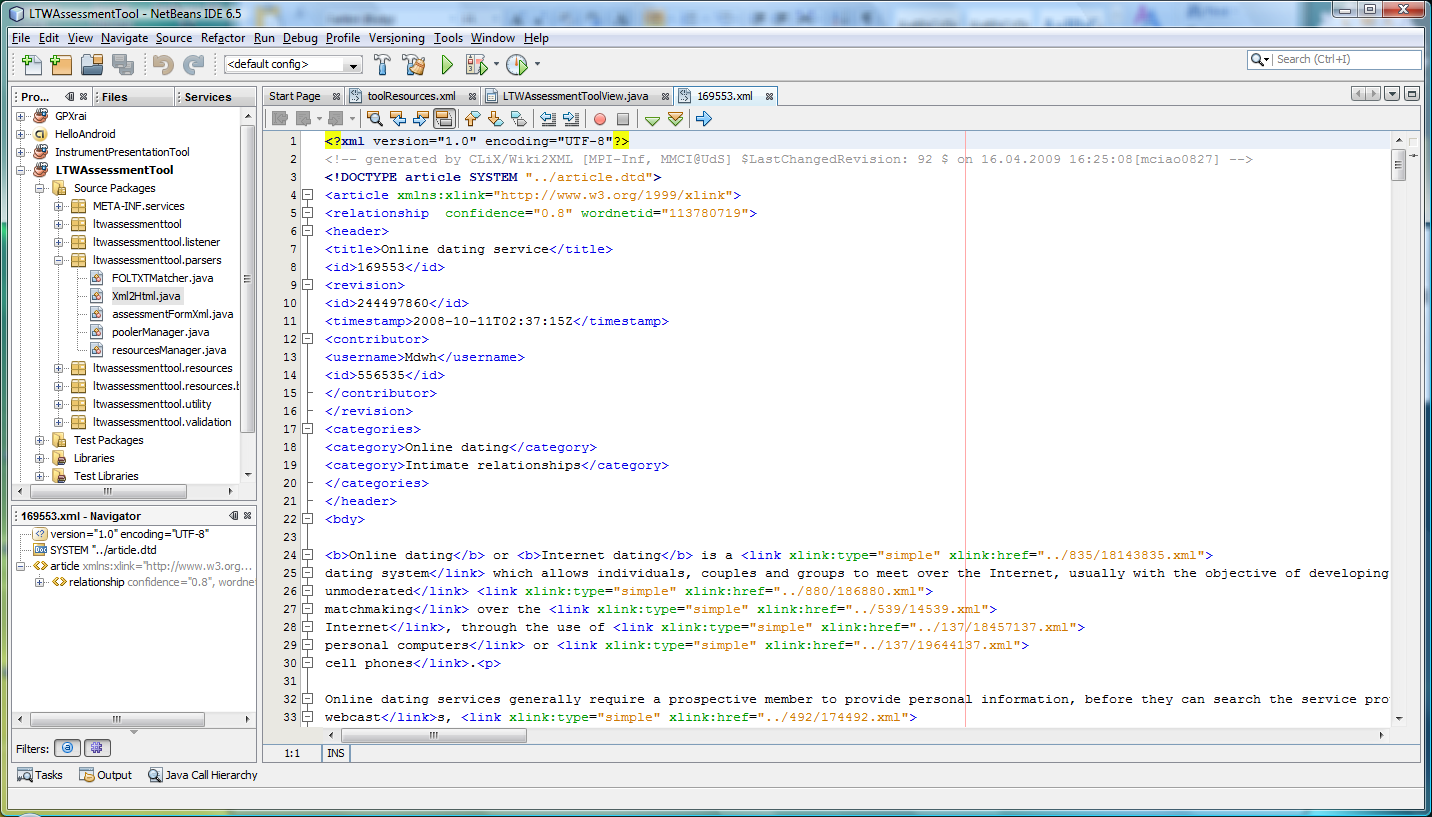


Figure 12. The Wikipedia DTD Specification

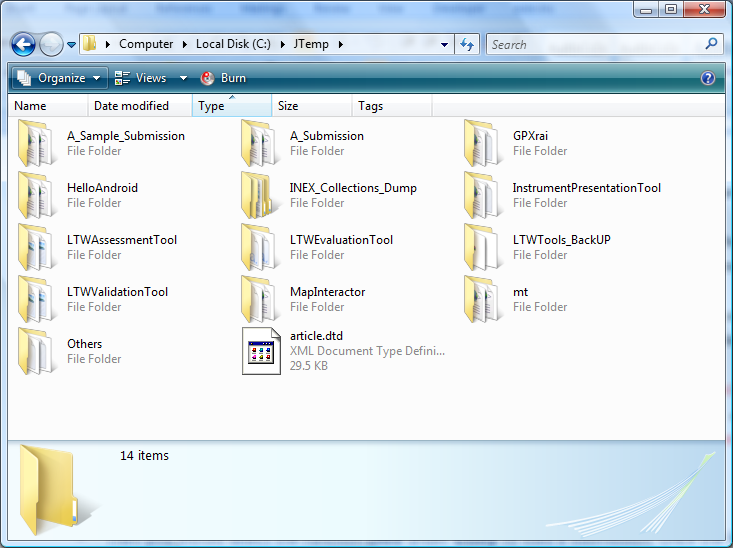


Figure 13. The location of DTD and the tool Package