Welcome to the documentation of how to implement the **Lects Search** System on the new machine. LectSearch system consist of 3 sub modules: 1) Indexing module 2) Web application modules and 3) Concept ontology analysis module (*obsolete*). This documentation will cover all the necessary steps required to deploy above 3 sub modules in order for the LectSearch to be fully operational. It will also guide you on how to populate or remove documents/collection into and from the system.

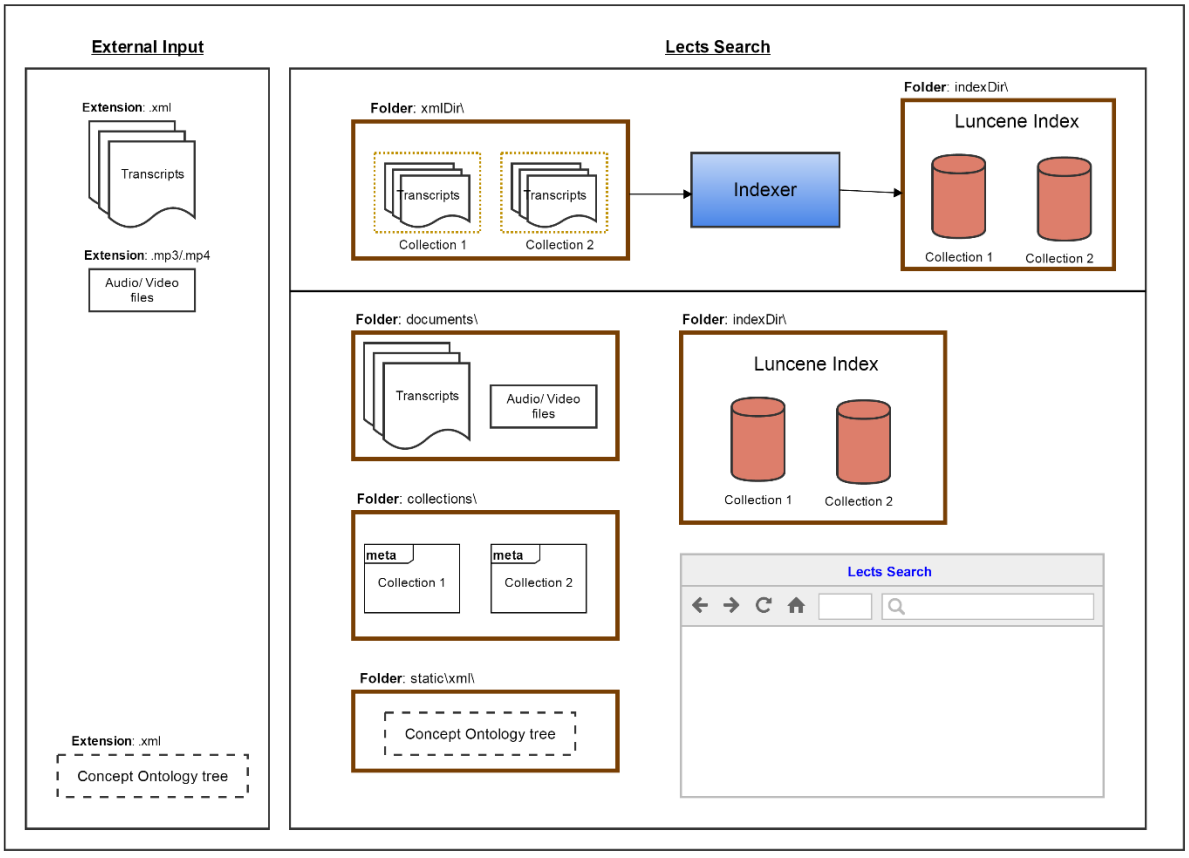

Figure 1: Lects Search Files Organization

Table of Contents

Terminology

Document – Lecture video transcription in .xml (e.g. Lecture1MITAerospace.xml)

Collection – Domain or Database (e.g. aerospace, signal processing)

1. Setting up Module 1: Indexing Module

Prerequisite:

a. Java or Latest Java Runtime Environment (JRE)

b. Executable JAR files , indexXML.jar (indexXML)

1. Get latest version of JAVA JDK or JRE from www.oracle.com
2. Obtain **indexXML.jar** from www.gitHub.com/kzintun/fyp-test.

2. Setting up Module 2: Web Application

Prerequisite:

a. Apache web server and PHP (XAMPP is recommended)

In this section, you will be instructed to set up local Apache server with PHP to run LectSearch web application. We recommend to use XAMPP, open source package, as it comes with both Apache serve and PHP on top of other useful applications and available on multiple platforms.

1. Go to https://www.apachefriends.org/index.html to get latest XAMPP and install it.
2. Obtain '**lectssearch'** folder from www.gitHub.com/kzintun/fyp-test
3. Once installation is complete, copy '**lectssearch** web application folder to the htdocs directory under XAMPP. The default path on Window is C:\xampp\htdocs .
4. Run XAMPP control panel and press Start button for Apache module.
5. Finally, go to your web browser and enter this URL (localhost/lectssearch/index.php) in the address bar.

You should now be able to see Lects Search homepage if all steps are taken successfully!

3. Populate 1 or more Collections

Section 3.1: Indexing video transcript

Prerequisite:

a. Lecture transcription in .xml format (Contents must follow format, Refer to Appendix X)

b. Executable JAR files , indexXML.jar (indexXML)

Our goal in this section is to perform reverse indexing on the transcript files into Lucene index. Lucene index, a group of segment files, is stored under a folder in the computer file directory system. Module 2 will use those segment files to facilitate user's search query from the LectSearch web application.

1. Create a new folder of your desired name (*Folder A*).
2. Create a new folder called xmlDir under '*Folder A*'. Copy all transcription XML files that you want to index under xmlDir folder and categorized them by Collection names.
3. Copy indexXML into *Folder A*. If you have followed steps correctly, the contents of *Folder A* should look as follow:

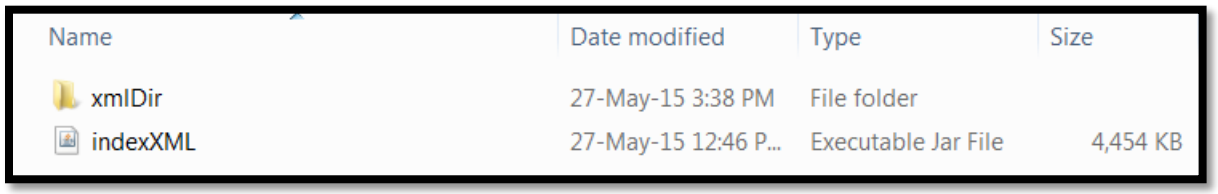

Figure 2: Folder A contents before indexing

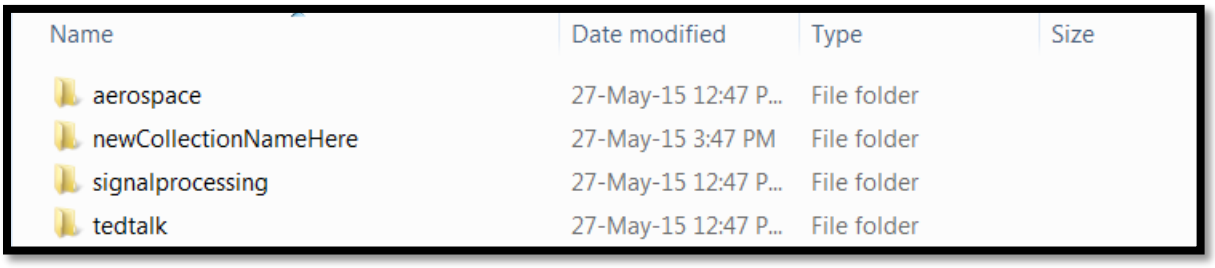

Figure 3: xmlDir Contents

4. In this step, we will be indexing all the new Collections folders (i.e. aerospace, signalprocessing, tedtalk). Now run (double click) on indexXML in *Folder A*. Alternatively, you can run it from command line by typing `java –jar indexXML.jar`. Running from command line allows you to observe the progress of indexing.
5. The indexing process will take some time to finish. The larger the Collection (folder), the longer the indexing time. The following message should be displayed on terminal/command line when the indexing process is completed.

```
Successfully indexed all documents under the folders: [aerospace, newCollectionN
ameHere, signalprocessing, tedtalk]
```

Figure 4: Indexing is complete

6. The indexing procedure is now completed when you see the message in Figure 3. You should be able to see the new Lucene Index folder called indexDir under *Folder A*. indexDir folder, categorized by Collections, contains the index segments files and will be used by search engine. At the end of this Section, your file directory should be looked as follow:

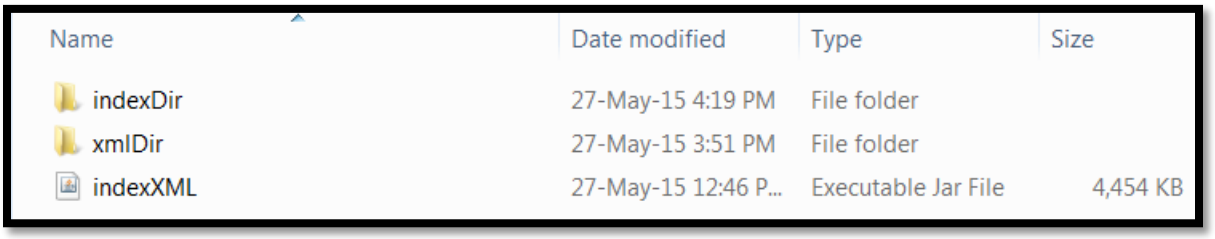

Figure 5: Folder A contents after indexing

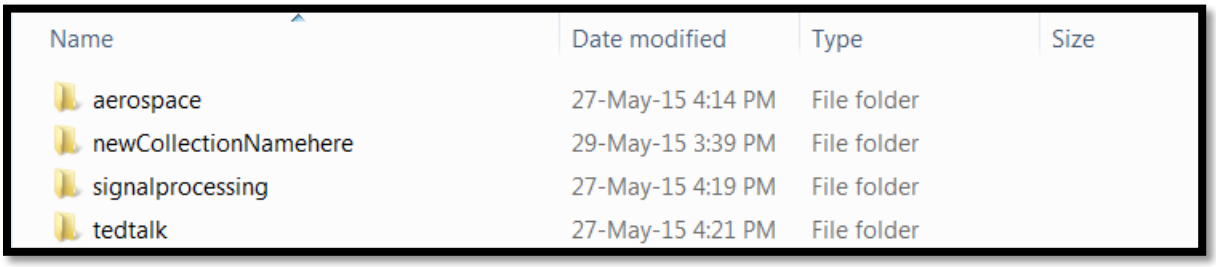

Figure 6: indexDir contents

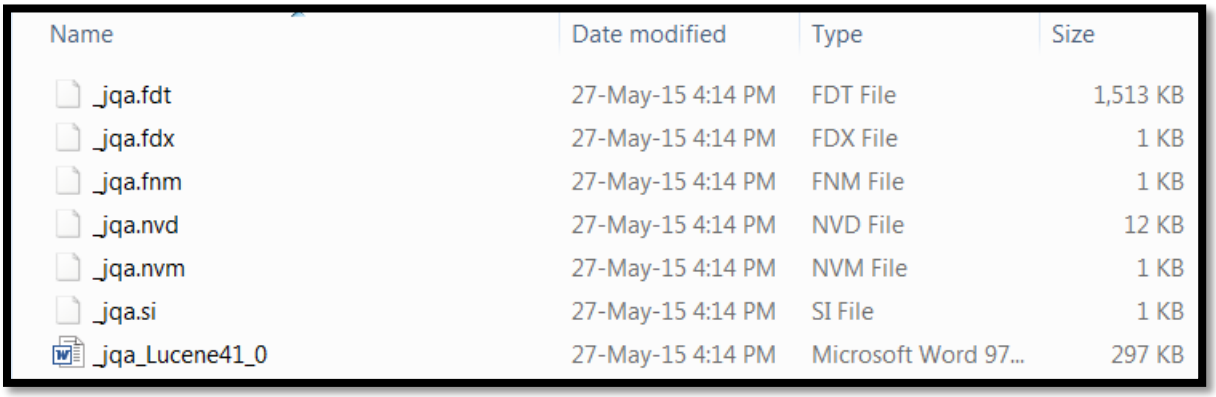

Figure 7: Sample contents of index segments (indexDir/aerospace/)

Section 3.2: Preparing files for web application

Prerequisite:

a. Lecture transcription in .xml format (Contents must follow format, Refer to Appendix X)

b. Lecture audio or video files

c. A thumbnail for each collection

d. An index direcotry indexDir created from previous Section 3.1.

To populate a new Collection(s) entry on the homepage, the above mentioned files need to be placed in the correct directory and a few scripts needs to be run to generate metadata file for each collection and word-Map thumbnail for each document.

1. Copy all lecture transcription and multimedia files of all collections into 'documents' folder of LectSearch.
2. Copy the contents of index directory indexDir into the existing index directory of LectSearch [.\htdocs\lectssearch\indexDir].
3. Copy the thumbnail image of each collection to C:\xampp\htdocs\lectssearch\img\collections .
4. Run the scripts in C:\xampp\htdocs\lectssearch\static\js to generate metadata file and word-Map thumbnail.

4. Populate 1 or more new Documents to existing Collection

Section 4.1: Indexing video transcript

Prerequisite:

a. Lecture transcriptions in .xml format (Contents must follow format, Refer to Appendix X)

b. Executable JAR files , indexXML.jar (indexXML)

Our goal in this section is to add new documents to the current index related to its Collection. The existing index folder of the Collection is required since we are going to *append* new index entries.

1. Create a new folder of your desired name (*Folder A*).
2. Create a new folder called xmlDir under '*Folder A*'. Copy all new transcription XML files that you want to index under xmlDir folder and **folderised** by collection names.
3. Copy indexXML into *Folder A*.
4. Copy the **existing index directory** indexDir of the Collection that you want to add documents into to *Folder A*. If you have followed steps correctly, it should look as follow:

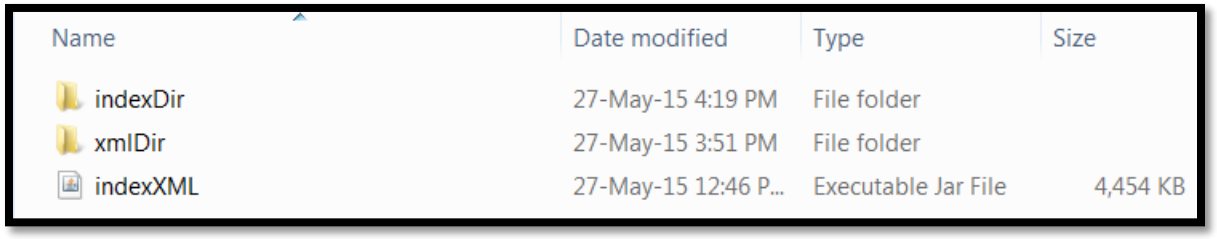

Figure 8: Folder A contents

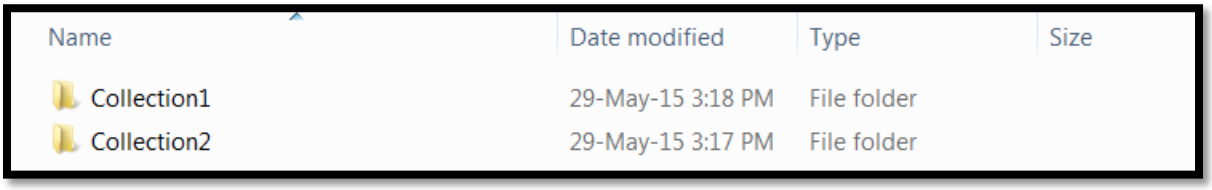

Figure 9: xmlDir contents

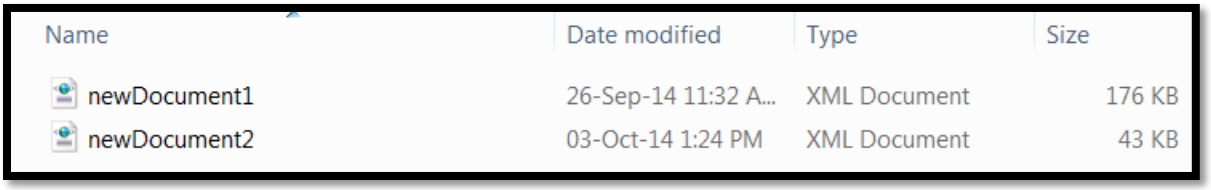

Figure 10: Collection1 contents

5. In this step, we will be indexing all new documents into existing Collection (i.e. Collection1, Collection2). Now run (double click) on indexXML in *Folder A*. Alternatively, you can run it from command line by typing `java –jar indexXML.jar`. Running from command line allows you to observe the progress of indexing.
6. The indexing process will take some time to finish. The larger the collection (folder), the longer the indexing time. The following message should be displayed on terminal/command line when the indexing process is completed.

```
Successfully indexed all documents under the folders: [Collection1, Collection2]
```

Figure 11: Indexing is complete

7. The indexing procedure is now completed when you see the message in Figure 11. The new documents are indexed and appended onto the current index directory indexDir . At the end of this Section, your file directory should be looked the same as before with increased in file size of indexDir as follow:

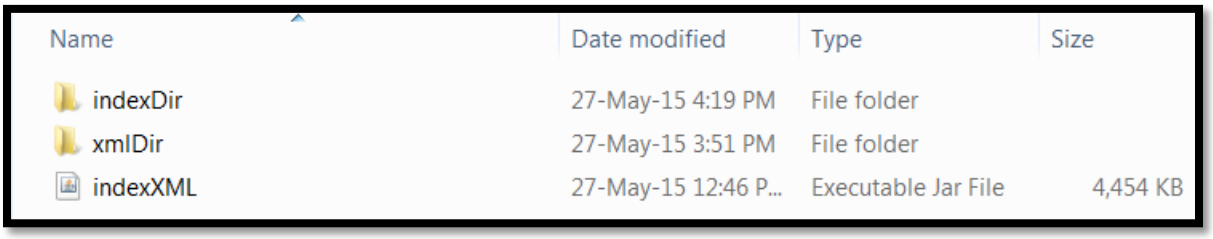

Figure 12: Final look of Folder A

Section 4.2: Preparing files for web application

Prerequisite:

a. New lecture transcription in .xml format (Contents must follow format, Refer to Appendix X)

b. New Lecture audio or video files

c. An index direcotry indexDir created from previous Section 4.1.

To add new documents to the LectSearch, simply place the above mentioned files in the correct directory and run the scripts to update metadata file of the collection and word-Map thumbnail for each new document.

1. Copy all new lecture transcription and multimedia files of all collections into 'documents' folder of LectSearch.
2. Copy and replace the **contents** of index directory indexDir , re-indexed Collection(s), into the existing index directory of LectSearch [.\htdocs\lectssearch\indexDir]. {Make sure to only delete the old Collection index folder(s) that has the same name as the new Collection}.
3. Run the scripts in C:\xampp\htdocs\lectssearch\static\js to update and generate metadata file and word-Map thumbnail respectively.

5. Remove 1 or more Collections

Section 5.1: indexing

None.

Section 5.2: Web application

1. Manually select and delete **all** the transcription and multimedia files of that particular collection(s) in 'documents' folder of Lects Search.
2. Delete the thumbnails of the collection and word-Map of the documents.
3. Delete the metadata file of the collection (collectionName.xml) in 'lectssearch\collections\'
4. Delete the index directory of the collection in 'lectssearch\indexDir\'.

6. Remove 1 more Documents from existing Collection

Section 6.1: Web application

1. Manually select and delete the particular transcription and multimedia file(s) of the collection in 'documents' folder of Lects Search.
2. Delete the word-Map thumbnails of the document(s).
3. Delete the whole index directory of the collection (e.g. lectssearch\indexDir\collectionThatcontainsTheDocumentsYouWantToDelete)

Section 6.2: Indexing

The whole collection is required to be re-indexed with remaining documents. Please follow **Section 3. Populate 1 or more Collections** to indexing process.