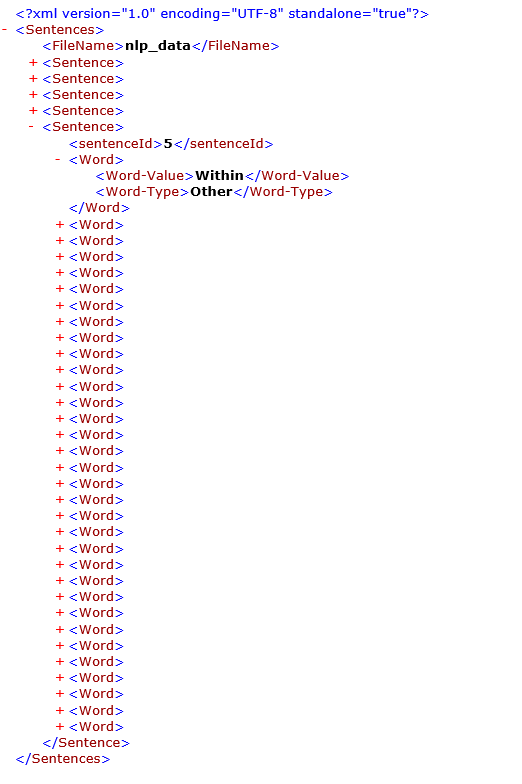
This Application searches a file to identify the each sentence boundaries and tokenize each word of sentence. Its generates the output in XML representation as following :

1. Under <Sentences> all sentences of a named file are listed.
2. Under <FileName> file name is mention where all these <sentences> are present in .txt file.
3. <sentenceId> attribute gives the sentence number of the sentence in txt file.
4. <Word> attribute tell information about each word of the sentence.
5. <Word-Value> attribute is the tokenized word of sentence.
6. <Word-Type> attribute tells if a word is NER(“named identity”) or not(“Other”)



**How to run :**

To run the module of the assignment, use following files respectively and run as java application

/NLP/src/com/digitalreasoning/module1/ApplicationMain1.java

/NLP/src/com/digitalreasoning/module1/ApplicationMain1.java

/NLP/src/com/digitalreasoning/module1/ApplicationMain1.java

**Results :**

Results of module 1 are in \NLP\_test\module1 output

Results of module 2 are in \NLP\_test\module2 output

Results of module 3 are in \NLP\_test\module3 output