

## Appendix K: XML content Removal:

```
try {
File TranFile = new File ("C://Users/u180384/Dropbox/Final Year/Final Year
Project/test_xml/English_With_Tags.txt");
File fXmlFile = new File("C://Users//u180384//Dropbox//Final Year//Final Year
Project//test_xml//clef2015.test.queries-EN.xml");
    DocumentBuilderFactory dbFactory = DocumentBuilderFactory.newInstance();
    DocumentBuilder dBuilder = dbFactory.newDocumentBuilder();
    Document doc = dBuilder.parse(fXmlFile);

    doc.getDocumentElement().normalize();
    // Normalize to stop things being separated by the new line
    // character before tag.

System.out.println("Root element :" +doc.getDocumentElement().getNodeName());
    // root element should be topics for eClef
    NodeList nList = doc.getElementsByTagName("top");

    for (int temp = 0; temp < nList.getLength(); temp++) {
        Node nNode = nList.item(temp);
        if (nNode.getNodeType() == Node.ELEMENT_NODE) {
            Element eElement = (Element) nNode;
            String Num = "<num>" +
            eElement.getElementsByTagName("num").item(0).getTextContent()
            + "</num>";

            String Query = "<query>" +
            eElement.getElementsByTagName("query").item(0).getTextContent(
            )+ "</query>";

            System.out.println(Num + "\r\n" + Query);
            BufferedWriter f = null;
            try {
                f = new BufferedWriter(new FileWriter(TranFile, true));
                f.append(Num + "\r\n");
                f.append(Query + "\r\n");
            }
            // use /r/n as a way to delimit the input onto different lines going into db for
            later translation.
        }
    }
}
```

Synopsis of what is going on:

- TranFile and XMLFile are files that are being read from (XML) and to (TRANS).
- Document builder is used to build the new file.
- If statement is used to remove the right XML content, as found by Node.ELEMENT\_NODE.
- If matches the tag we are looking at, it gets the content between the two defined tags.
- BufferedWriter writes these contents to Trans file with the use of a filewriter, separated by the new line character for later use in the translation process.