**Basic Outline:**

Frontend GUI: JavaServer Pages (JSP)

Server: Apache Tomcat

JRE – 8

JDK – 7

IDE: Eclipse Juno

Knowledge Base: Wikipedia

NLP techniques incorporated: Parts of Speech Tagging, Named Entity Recognition (Stanford Parser, Stanford CoreNLP used), jsoup (HTML parser)

**QA System Implementation steps:**

1. Download Apache Tomcat

<https://www.youtube.com/watch?v=bP66y108xAc>

<http://www.herongyang.com/JSP/Tomcat-Launch-Tomcat-7-Server.html>

1. Download Eclipse Juno and integrate with Apache TomCat

<http://www.coreservlets.com/Apache-Tomcat-Tutorial/tomcat-7-with-eclipse.html>

1. Create first servlet and integrate with tomcat

<http://www.vogella.com/tutorials/EclipseWTP/article.html>

1. First HelloWorld jsp

<http://www.srccodes.com/p/article/2/JSP-Hello-World-Program-using-Eclipse-IDE-and-Tomcat-web-server>

1. Add first Jar – jsoup added (*this is for real-time extraction of information*)

<http://www.wikihow.com/Add-JARs-to-Project-Build-Paths-in-Eclipse-(Java)>

6) Added Stanford ner jar to the MediHelp project in Eclipse. Tried out an example code.

Example output:

Hello I am <PERSON>Samrudhi</PERSON>, I stay in <LOCATION>Mumbai</LOCATION> and I am 22 years old.

<https://ganeshpachpind.wordpress.com/2014/02/28/stanford-named-entity-recognizer/>

<http://www.informit.com/articles/article.aspx?p=2265404>

1. Creating the GUI jsp page – MediHelp.jsp (*GUI created*)
2. Adding jsoup code for extracting web content (*Real Time Extraction*)

<http://stackoverflow.com/questions/9151075/jsoup-extract-text-from-wikipedia-article>

<http://www.joe0.com/2014/09/21/how-to-parse-wikipedia-infobox-html-using-java-and-jsoup/>

<http://www.mkyong.com/java/jsoup-html-parser-hello-world-examples/>

1. Adding jquery functionality for basic query checking!
2. Add question mappings txt file to properties folder. It is used to train Stanford NER.

<http://nlp.stanford.edu/software/crf-faq.shtml>

1. Stanford parser. Jar added

<http://nlp.stanford.edu/software/lex-parser.shtml>

1. Started with the Query Analyzer phase. The main task involves detecting the topics, attributes, and question type.

The class names for those programs are: TopicDetect, AttributeDetect and QuestionTypeDetect. For this we shall be using the Standford Parts of Speech tagger. For these programs to detect the different parts of speech it is absolutely necessary to add the Stanford jar and Stanford CoreNLP jars to the eclipse projects. [All the jars under Stanford CoreNLP]

The different POS labels are:

1. CC Coordinating conjunction
2. CD Cardinal number
3. DT Determiner
4. EX Existential there
5. FW Foreign word
6. IN Preposition or subordinating conjunction
7. JJ Adjective
8. JJR Adjective, comparative
9. JJS Adjective, superlative
10. LS List item marker
11. MD Modal
12. NN Noun, singular or mass
13. NNS Noun, plural
14. NNP Proper noun, singular
15. NNPS Proper noun, plural
16. PDT Predeterminer
17. POS Possessive ending
18. PRP Personal pronoun
19. PRP$ Possessive pronoun
20. RB Adverb
21. RBR Adverb, comparative
22. RBS Adverb, superlative
23. RP Particle
24. SYM Symbol
25. TO to
26. UH Interjection
27. VB Verb, base form
28. VBD Verb, past tense
29. VBG Verb, gerund or present participle
30. VBN Verb, past participle
31. VBP Verb, non­3rd person singular present
32. VBZ Verb, 3rd person singular present
33. WDT Wh­determiner
34. WP Wh­pronoun
35. WP$ Possessive wh­pronoun
36. WRB Wh­adverb

<http://stackoverflow.com/questions/1833252/java-stanford-nlp-part-of-speech-labels>

<http://stackoverflow.com/questions/16523067/how-to-use-stanford-parser/>

<http://stackoverflow.com/questions/9492707/how-can-i-split-a-text-into-sentences-using-the-stanford-parser?rq=1>

<http://stackoverflow.com/questions/15431139/java-program-to-get-parse-score-of-a-sentence-using-stanford-parser?rq=1>

<http://stackoverflow.com/questions/27912176/stanford-parser-java-error?rq=1>

While detecting the topics and attributes we remove all the stopwords from the query. The class name is QueryCleanAndParse. Stopwords removed are : "is","was","the","a","and","has","have","had","in","into","an","as","at" 🡪 The list has to be expanded further

<http://stackoverflow.com/questions/1218335/stop-words-list-for-english>

Stanford Parser Tutorials:

<http://stackoverflow.com/questions/19429106/how-can-i-integrate-stanford-parser-software-in-my-java-program>

<http://www.programcreek.com/2012/07/java-example-for-using-stanford-postagger/>

<http://mendicantbug.com/2007/10/25/the-noobs-guide-to-parsing/>

<http://www.programcreek.com/java-api-examples/index.php?api=edu.stanford.nlp.parser.lexparser.LexicalizedParser>

<http://stackoverflow.com/questions/22318532/lexicalized-parser-vs-dependency-parser>

<http://stackoverflow.com/questions/14598250/perfomance-issue-while-using-stanford-lexicalized-parser-in-java>

<http://www.quora.com/What-tools-are-available-for-extracting-lexicalized-PCFGs-from-a-parsed-corpus>

<https://rajvardhan.wordpress.com/2012/11/11/stanford-pos-tagger-in-eclipse/>

<http://utkarshu.in/PosTagger.java>

<http://stackoverflow.com/questions/3733587/how-to-get-pos-tagging-using-stanford-parser>

(*Query’s Topic, Question Type, and attributes have been parsed at this stage*)