**Wikipedia Introduction Extractor**

**Format to enter the input:**

1. Through command line -

After compilation,

Java UserInput *The topic with words separated by spaces*

1. Through GUI -

Enter the text in the text field and click the button.

**Explanation of the code:**

Two classes are used for this project. “*UserInput.java*” and “*Connection.java*”.

1. The *UserInput* class is used to accept the input from the user in two ways:
2. From the command line arguments entered by the user.
3. In this case, the command line arguments are used to form the input string.
4. An object of the *Connection* class is created which is used to access the findResult method.
5. The input string which is passed as an argument to the findResult method is converted to the format accepted by the url with the help of the WordUtils.capitalizeFully(str).
6. The extra tags in the resulting JSON file are removed using regular expressions.

The "(?<=\"extract\":\").\*" regex is used to get the content after the occurrence of “extract”: where the content starts.

The resulting string is further split using the regex "\"}" to remove the “}” occurring at the end.

1. The final output string obtained is passed to the *UserInput* class where it is printed in the text area.
2. From the graphical interface using Swing from which the input query is accepted from the user.
3. When the user clicks on the button, an object of the second class i.e. *Connection* class is created and the method findResult is called by passing the user input string as argument.
4. In the findResult method, an input stream from the following URL is stored in a JSON file.
5. The rest of the steps are the same as steps c to e mentioned above.

**External JARS used:**

org-apache-commons-lang.jar

It is used for converting the string entered by the user to the format accepted by the URL.

The WordUtils.capitalizeFully(str) method in this library helps in capitalizing the first alphabet of each word of the input string and also converting the rest of the characters of the words to a lower case.