**Case Scenario: Process of Web Crawling**

A web crawler (also known as a web spider or web robot) is a program or automated script which browses the World Wide Web (www) in a methodical, automated manner. This process is called Web crawling. Search engines use this concept for fetching the relevant and updated content from the websites. The same can be used to download websites as an offline dump so that it can be referred without using internet afterwards. Parsing of a respective web-page constitute two parallel tasks viz. fetching relevant content from body of webpage and fetching data from the META tag used while scripting. Relevancy relates to the process of checking keyword stuffing, URL redundancy and doorway pages.

Hyperlinks Referred for understanding the above given scenario and for generating its activity diagram using Rational Software Architect (RSA) tool:

1. <http://computer.howstuffworks.com/internet/basics/search-engine1.htm>
2. <http://searchmicroservices.techtarget.com/definition/crawler>
3. <https://en.wikipedia.org/wiki/Web_crawler>
4. <http://www.wpthemesplanet.com/2009/09/how-does-web-crawler-spider-work/>
5. <https://florianhartl.com/nutch-how-it-works.html>