**Documentation**

* Team members :-

**Dharm Dutt Mishra**

**Suman Singh**

**Deliverable​ ​1:-**

We have chosen the five relevant data source from these websites that are mentioned below -

1. Fileinfo.com (https://fileinfo.com/)
2. Filext.com (filext.com/file-extension)
3. Pc.net (https://pc.net/)
4. Reviversoft.com (https://www.reviversoft.com/file-extensions/net)
5. Extension.net (extensionfile.net)

We have found that, the exact file type from these above resources that contains almost all the extension files and details about those extension files. We used three among all five resources in our code implementation to find the all type of file extension and merge them in output.

These resource will provide the following specific details about the input file type.

Filext.com (filext.com/file-extension)

1. **File extension:**PDF
2. **File type:**Portable Document Format

**Detailed information for file extension PDF:**

1. **Primary association:**[Acrobat](http://www.adobe.com/)
2. **Company:**[Adobe Systems Incorporated](http://www.adobe.com/)
3. **Mime type:**application/pdf, application/x-pdf, application/acrobat, applications/vnd.pdf, text/pdf, text/x-pdf
4. **Identifying characters**Hex: 25 50 44 46 2D 31 2E, ASCII: %PDF-1.
5. **Related links**

Reviversoft.com (<https://www.reviversoft.com/file-extensions/net>)

* What is pdf file?
* How to open a .pdf file
* Application that open a .pdf file :-
* Acrobat Reader
* Adobe Reader
* PDF Colony PDF Index Generator
* Google Chrome
* Acrobat Reader
* PDF Colony PDF Index Generator
* Google Chrome

Extension.net (extensionfile.net)

* This will provide all details about the file type

**Deliverable​ 2​:-**

**Tools for Implementation of Project:-**

* Eclipse

**Implementation of Project:-**

* We are using the Web Crawling technique to collect the data from resources for the input file type.
* At first we created a maven in eclipse with package name com.blueOptima.
* For Web Crawling implementation we added the Dependency of Jsoup.jar file in our maven project.
* We use the FileReader for the input stream of file type with file name text23.txt that is situated at specified location “H://text23.txt” in my PC.
* We use the FileWriter for output stream with the file name “testout.txt” which is situated in our specified location “H://text23.txt” in my PC.
* I have attached all the input and output files with this project.