Filetype Identification, Version: 1.0 R1

Team Name: BO-QA-Team 001

Problem Statement:

Identify different sources that could be used to identify details of a file type
like following (but not limited to)
   1. Short Description (explaining the usage of the file type)
   2. Language family (Java, Python, Perl, etc).

Sources we are referring to:

   1. file-extensions.org
   2. fileinfo.com
   3. Wikipedia's list of file extensions
   4. dcjtech.info/topic/python-file-extensions/
   5. fileext.com

Justifying our sources:

1. File-extensions.org presented itself as a superset of everything we found on fileinfo.com.
Hence proving to be a broader sources of information. However, the site-structure turns out
to be more complicated.

2. File Info calls itself " The file information database". In our experience we found out that
we were able to get some info even on the most obscure files that we had lying around in
our computers. With a rather simple structure of the page, Scraping it for info was not a
problem.

3. Wikipedia's list of file extensions – As we know, wikipedia houses even the most obscure
details about the most obscure things. And wikipedia can be edited or augmented by various
sources all over the world. This also means, that if there's a relatively newer file type in the
wild, wikipedia would be a great place to start to gather information about it.

4. dcjtech.info/topic/python-file-extensions/ - although just a 3rd party site dedicated to tech
we found this helpful on info for various python related files.

5. filext.com – this website although only mentions a subset of the file extensions found on
the first two resources, goes into details of them unlike the first two , even mentioning the
various ways that we could open up those files.