JPlag Web Report Requirements

Evgeni Cholakov

Abstract

The following document examines an old web report of JPlag[2] and identifies the elements which were displayed 1. In second section 2 the JPlag API[1] is observed. The result metrics which can be obtained from a report result through the API are identified. Section 3 describes the functional requirements of the new web report.

Contents

1 Old Report Elements

The old report is structured as follows:

- Main Page 1.1
- Pages for each match 1.2
- Help Pages 1.3

1.1 Main Page

On the main page we can find the following information:

1. Path to input submissions/directory of submissions
2. Ids of the analyzed submissions
3. Language and language version of programs
4. Total number of submissions with indication and number of failed submission parses
5. Id of failed submissions
6. Number of found matches for 2 criteria (average and maximum similarity) as well as threshold (in percentage) for detecting vital match
7. Date of run
8. Sensitivity of matches (minimum matched tokens to declare a match)
9. File extensions
10. Distribution of matches from 0 to 100 percent
11. Table showing average similarity between tuples of submissions (vital matches only)
12. Table showing maximum similarity between tuples of submissions (vital matches only)

The page has links to:

1. Help pages (average similarity, maximum similarity).
2. Links to pages for each vital match

1.2 Page for a match

On the page for a match we can find the following information:

1. Ids of the submissions being matched
2. Table showing the matched lines, described by:
    - corresponding lines of both submissions
    - length of the match in tokens
3. Two text fields, one for each submission source code, with same text coloring for matching code parts

The page has the following links

1. Link to the main page
2. Help page link (how to read results)
3. Links to matching code parts (on pressed scrolls both text field to the lines which are matching)

1.3 Help pages

On the help pages can be found help about a certain topic (How is a result read, what is max similarity, etc.).
On the pages there are links to:

1. Main page
2. Email of Guido Malpohl jplag@ira.uka.de

2 JPlag API

In the following section we observe what information can be extracted from a result of submissions analysis.

2.1 Obtained information

The analysis returns an object of type *JPlagResult*. With this object we have access to its internal attributes, the options of the analysis, instance of *JPlagOptions* and comparisons done in the analysis, instances of *JPlagComparison*. From these classes we can obtain the following information:

1. Total amount of submissions.
2. Duration of the analysis (in ms assume, no docs).
3. Name of the base code file, if present otherwise null.
4. Description of comparison mode.
5. Exclusion file name, if present otherwise null.
6. Name of the submissions programming language
7. Tokens of the language.
8. Minimum tokens match of the language.
9. Analysis max number of matches.
10. Root directory of the submissions.
11. Subdirectory name if present otherwise null.
12. Name of the similarity metric.
13. Similarity threshhold.
14. Suffixes of the files.
15. Distribution of similarity in percentage.
16. List of comparisons.

From each comparison we can obtain:

1. Name of the submissions compared.
2. Maximum similarity.
3. Minimum similarity.
4. Average similarity.
5. Similarity of first submission.
6. Similarity of second submission.
7. Similarity of first submission to base code (if present).
8. Similarity of second submission to base code (if present)
9. List of matches *Match* object.

In each match we have:

1. Length of the match.
2. Start of the match in first submission (line of code)
3. Start of the match in second submission (line of code)

3 Functional Requirements

The new JPlag Web Report (WR) should have the following features:

FN 1 The WR should have a main page, displaying the overall information.

FN 2 The WR should provide a single page for each match.

FN 3 The WR should provide help information about reading the matches page and explanations on the metrics.

FN 4 The main page should contain the following overall information:

- **FN 4.1** Path to the folder containging the submissions.
- **FN 4.2** Base code path and name, if present.
- **FN 4.3** Programming language of the submissions.
- **FN 4.4** File extensions of the submissions. (do we need it, if we have language info?)
- **FN 4.5** Number of total submissions.
- **FN 4.6** Number of failed to be parsed submissions.
- **FN 4.7** List of files excluded from the comparison.
- **FN 4.8** The comparison metrics with their threshold.
- **FN 4.9** Minimum match length to determine a match.
- **FN 4.10** Date of execution.
- **FN 4.11** Execution duration.

FN 5 The main page should provide the distribution of matches in a diagram.

FN 6 The main page should provide a list or table containing the top N matches with their match percentage for the maximum similarity metric.

FN 7 The main page should provide a list or table containing the top N matches with their match percentage for the average similarity metric.

FN 8 A match page should provide names of the matched submissions.

FN 9 A match page should provide percentage of the matched submission.

FN 10 A match page should provide name of the metric according to which it is matched.

FN 11 A match page should provide tabs with code snippets of all matched code, including names of files matched, start and end lines of matched code.

FN 12 A match page should provide the option ot change view to side by side full code preview with matched code lines identically colored.

References

[1] *JPlag API Documentation*. URL: https://github.com/jplag/JPlag/wiki/1.-How-to-Use-JPlag#using-jplag-programmatically.

[2] *Old Web Report Example*. URL: https://github.com/jplag/ExampleReport.