Eclipse JDT/Core

• Plot of the distribution of all the terms in the software lexicon. You can use Excel or anything like it.

• List of the 10 most frequent terms in the corpus (report their tf).

• List of the 10 terms with the highest document frequency (report their df).

• List of the 3 documents with the highest number of unique terms.

• List of the 3 documents with the lowest number of unique terms.

Weka

• Plot of the distribution of all the terms in the software lexicon. You can use Excel or anything like it.

• List of the 10 most frequent terms in the corpus (report their tf).

• List of the 10 terms with the highest document frequency (report their df).

• List of the 3 documents with the highest number of unique terms.

• List of the 3 documents with the lowest number of unique terms.