Motivation

Observations from a survey of 205 evaluations of plagiarism detectors:
(101 on text, 104 on code)

<table>
<thead>
<tr>
<th>Evaluation Aspect</th>
<th>Text</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>precision, recall</td>
<td>43%</td>
<td>18%</td>
</tr>
<tr>
<td>manual, other</td>
<td>57%</td>
<td>82%</td>
</tr>
<tr>
<td><strong>Corpus Acquisition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>existing corpus</td>
<td>20%</td>
<td>18%</td>
</tr>
<tr>
<td>homemade corpora</td>
<td>80%</td>
<td>82%</td>
</tr>
<tr>
<td><strong>Comparative Evaluation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>46%</td>
<td>51%</td>
</tr>
<tr>
<td>yes</td>
<td>54%</td>
<td>49%</td>
</tr>
</tbody>
</table>

→ Measuring detection performance is not well-understood.
→ There is no standardized evaluation corpus.
→ Half of the evaluations don’t compare different approaches.

We introduce the first standardized evaluation framework for plagiarism detection.

Performance Measures

Let $s \in S$ denote plagiarism cases.

Let $r \in R$ denote plagiarism detections.

The formulas below measure the detection performance of $R$ with regard to $S$.

The well-known precision and recall:

\[
\text{prec}(S, R) = \frac{1}{|R|} \sum_{r \in R} \frac{|\bigcup_{s \in S} (s \cap r)|}{|r|}
\]

\[
\text{rec}(S, R) = \frac{1}{|S|} \sum_{s \in S} \frac{|\bigcup_{r \in R} (s \cap r)|}{|s|}
\]

The granularity measures the average number of detections of all detected cases:

\[
\text{gran}(S, R) = \frac{1}{|S_R|} \sum_{s \in S} |R_s|
\]

The domain of the granularity is $[1, |R|]$.

Combining the three concepts:

\[
\text{plagdet}(S, R) = \frac{F_1}{\log_2(1 + \text{gran}(S, R))}
\]

Evaluation Corpus

The PAN plagiarism corpus 2010 (PAN-PC-10):

27 073 documents in which
68 558 plagiarism cases have been inserted.

4 000 cases were created manually using Amazon's Mechanical Turk, the rest artificially.

A high diversity of cases was achieved by varying 7 different parameters.

See the framework in action at http://pan.webis.de