Matching Ukrainian Red Links with English Wikipedia’s Articles

https://github.com/Katerali/redlinks_linking
https://doi.org/10.6084/m9.figshare.11550774

Kateryna Liubonko, UCU (aloshkina@ucu.edu.ua)
Diego Sáez-Trumper, Wikimedia Foundation (diego@wikimedia.org)
Problem & Approach

Not to be confused with A.C. Ancona.

Unione Sportiva Ancona 1905 [ITA], commonly referred to as Ancona, was an Italian football club based in Ancona, Marche.

The club changed its name to U.S. Ancona 1905 from S.S. Piano San Lazzaro in 2010, due to the bankruptcy of the main football team of the city, A.C. Ancona. However, after a relegation in 2017, U.S. Ancona 1905 withdrew from the football pyramid.

817 892 articles*
5 719 743 articles*

* Wikipedia dumps from the 20th of September, 2018

Results & Contribution

<table>
<thead>
<tr>
<th>Similarity metrics</th>
<th>fp</th>
<th>tn</th>
<th>tp</th>
<th>fn</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>BabelNet</td>
<td>4</td>
<td>65</td>
<td>108</td>
<td>456</td>
<td>0.964</td>
<td>0.191</td>
<td>0.32</td>
</tr>
<tr>
<td>Incoming links (top1, th=0.26)</td>
<td>385</td>
<td>149</td>
<td>64</td>
<td>37</td>
<td>0.143</td>
<td>0.634</td>
<td>0.233</td>
</tr>
<tr>
<td>Concurrent links (top1, th=0.1)</td>
<td>263</td>
<td>360</td>
<td>2</td>
<td>10</td>
<td>0.008</td>
<td>0.167</td>
<td>0.014</td>
</tr>
<tr>
<td>Levenshtein (top1, th=0.39)</td>
<td>295</td>
<td>24</td>
<td>41</td>
<td>275</td>
<td>0.122</td>
<td>0.13</td>
<td>0.126</td>
</tr>
<tr>
<td>Multi-factored model (top1)</td>
<td>31</td>
<td>1445</td>
<td>339</td>
<td>90</td>
<td>0.92</td>
<td>0.79</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Table 2: Evaluation results on similarity metrics as independent models and on a multi-factored model based on the results of the independent ones. The best _F₁_ score is in bold.

Dataset of 2 957 927 pairs of red links and the candidate articles in English Wikipedia for the most frequent 3 171 red links from Ukrainian Wikipedia. Publicly released at https://doi.org/10.6084/m9.figshare.11550774