Matching product names from different retail businesses

Dependable Systems - Dirk Nowotka

Project Description

The data and questions of this thesis are provided by the Department of Market Theory at the Institute of Agricultural Economics, CAU which also provides the data and content-related support.

Background

For the analysis of price series for food in online grocery retailing, data was collected by web scraping from the Department of Market Theory. The data contains product names and prices. However, the product names at the different suppliers (Amazon, Rewe, etc.) are not always or even predominantly the same for identical products. Sometimes there are also changes over time.

Task of Thesis

Identical products are now to be matched with the help of an AI algorithm and provided with a uniform product name in the data set. The following table shows a few examples:

<table>
<thead>
<tr>
<th>Applicable For</th>
<th>Bachelorstudents</th>
<th>Masterstudents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skillset</td>
<td>Programming</td>
<td>Machine Learning</td>
</tr>
<tr>
<td></td>
<td>Natural Language Processing</td>
<td></td>
</tr>
</tbody>
</table>

Keywords

Machine Learning
Natural Language Processing
Similarity analysis

Contact

Karolina Ochs
@ kao@informatik.uni-kiel.de