Data Search using Deep Learning

The Web is a rich source for tables:
- Many tables describe the same real world entities
- Most tables contain only partial information
- Tables are scattered across different websites

Table Augmentation
- Augment an input table with information from a table corpus
- Task not trivial due to heterogeneity and size of table corpus

<table>
<thead>
<tr>
<th>Movie</th>
<th>Premiere</th>
<th>Director</th>
<th>Actors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biutiful</td>
<td>05/17/2010</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>True Grit</td>
<td>12/22/2010</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>The Social Network</td>
<td>10/01/2010</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>The King’s Speech</td>
<td>1/07/2010</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>127 Hours</td>
<td>09/04/2010</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>The Fighter</td>
<td>12/10/2010</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

How can we use transformer-based models for Table Augmentation?
Project Goal
Experiment with State-of-the-Art NLP Transformer Models and use them to search for Tabular Data

Involves
- Data Profiling, Data Preprocessing
- Model Training, Evaluation, Selection

Learning Targets
- Gain technical experience with State-of-the-Art Data Search Technologies
- Gain work experience as Data Scientist

Requirements
- Data Science & Engineering Skills, Programming Experience (Python)
- Relevant Courses: Web Data Integration, Data Mining I & II, Information Retrieval & Web Search

Organization: 4-6 people, 6 months, work as a complete team and in subgroups
Instructors: Alexander Brinkmann, Christian Bizer