Automatic Construction of Technology Function Matrix

Xiang Shi(presenter), Zikun Feng, Jiawei Liu,
Qikai Cheng and Wei Lu

Wuhan University, China

3rd Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents
EEKE2022 @ JCDL2022
Background

- **Era of Big Data**: the number of patent documents is increasing explosively. It is becoming more and more difficult to accurately grasp the development trend of science and technology.

- **Technology Function Matrix (TFM)**: It is an important basis for patent analysis, such as high-value technology discovery and potential technology function prediction.
Contribution

- Construction process of Technology Function Matrix

Contribution

- Technology Framework
- Semi-Supervised Method
- TFM Construction System
**Methodology**

**Technology Phrase**
- 电极材料（electrode material）
- 碳纳米管薄膜（carbon nanotube film）
- 卷绕式超级电容器（wound supercapacitor）

**Function Phrase**
- 降低等效电阻（reduce equivalent resistance）
- 提高电导率（increase conductivity）
- 提高功率密度（increase power density）

**Technology phrase extraction** based on dependency analysis and pre-trained language model

**Function phrase extraction** based on dependency analysis and template

本发明提供了一种卷绕式超级电容器制备方法，包括...极大地提高了卷绕式超级电容器电极的电导率...(The invention provides a preparation method of a winding supercapacitor, which comprises... The conductivity of the wound supercapacitor electrode is greatly improved...)

**Result**
Dataset

Espacenet (Open Source)

New Energy Vehicles

Keywords

IPC, CPC, and Chinese abstracts

Web Crawler

https://worldwide.espacenet.com/
**Function Phrase Extraction**

**STEP1**
Function sentence recognition

**STEP2**
Function phrase extraction

1. select seed words
   such as improve, increase, reduce, etc

2. semantic dependency parser (spaCy)

3. construct template

4. calculate F1 score and repeat above steps

---

- **Function sentence**: 0.9  
- **Non function sentence**: 0.1

Bert Model

本发明能很好地提高汽车的加速性能和爬山性能...(The invention can well improve the acceleration performance and mountain climbing performance......)
Technology Phrase Extraction

- Semantic Dependency Analysis

  core word: domain vocabulary

  context: left and right five words

  ancestor: syntactic parent of the core word

  sub-word: core word that removes modifier.

- Training set generation

  core word + context + ancestor + sub-word
Technology&Function Phrase Merging

- Directory tree crawling.
- Abbreviation recognition.
  - Maximum Entropy Model
  
  \[ p(y|x) = \arg \max_p \sum_{x,y} -\log p(x)p(y|x) \log p(y|x) \]
- Domain triplet recognition.

  "Support Vector Machine" and "SVM"

- Suffix tree pattern recognition.
  - suffix tree string matching algorithm
  
  "author LDA" and "LDA"
Experiment

- Evaluation Data
  - 1,000 function sentences;
  - 532 function phrases;
  - 907 technology phrases

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAO</td>
<td>20.14</td>
<td>26.16</td>
<td>22.76</td>
</tr>
<tr>
<td>SDP + Template</td>
<td>56.83</td>
<td>48.59</td>
<td>52.39</td>
</tr>
</tbody>
</table>

Table 1: Comparative experimental results of function sentence recognition.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>65.86</td>
</tr>
<tr>
<td>Word2Vec+MLP</td>
<td>65.67</td>
</tr>
<tr>
<td>Bert</td>
<td>89.13</td>
</tr>
</tbody>
</table>

Table 2: Experimental results of function phrase extraction.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Span-BERT</td>
<td>35.55</td>
<td>81.25</td>
<td>49.46</td>
</tr>
<tr>
<td>+ Ancestor</td>
<td>53.81</td>
<td>52.14</td>
<td>51.61</td>
</tr>
<tr>
<td>+ Sub</td>
<td>46.94</td>
<td>63.50</td>
<td>52.96</td>
</tr>
<tr>
<td>+ Ancestor + Sub</td>
<td>47.90</td>
<td>60.52</td>
<td>53.48</td>
</tr>
</tbody>
</table>

Table 3: Technology Phrase Extraction Measurement
**System Overview**

**STEP 1**
**Project Creation**

**STEP 2**
**Patent Retrieval**

**STEP 3**
**Technology Selection**

**STEP 4**
**Function Selection**

- **Project Type**
  - Literature
  - Patent

- **User Name**

- **Title**

- **Description**

**Search**

**Keywords**
- Intelligent Vehicle
  - 2733

**Description**

- Intelligent shared small electric bus is an AI smart shared electric bus. Replace motorcycles, tricycles, bicycles and other non motor vehicles for people to travel on urban auxiliary roads.

**Selection**

- **Selection**
  - Storage battery
  - Control system
  - Clutch
  - Sensor
  - Electric generator
  - Motor controller

- **Selected 6 Technology Words**
  - Storage battery: 98
  - Control system: 80
  - Clutch: 73

- **Selected 5 Function Words**
  - Energy consumption: 583
  - Cost: 2187
  - Safety: 2070
System Overview

STEP5
Visualization
Thanks