Automatic Construction of Technology Function Matrix

Xiang Shi(presenter), Zikun Feng, Jiawei Liu,

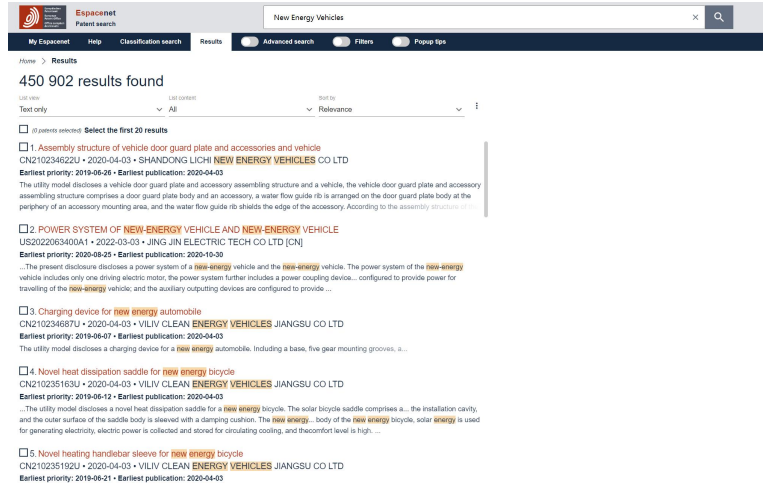
Qikai Cheng and Wei Lu

Wuhan University, China

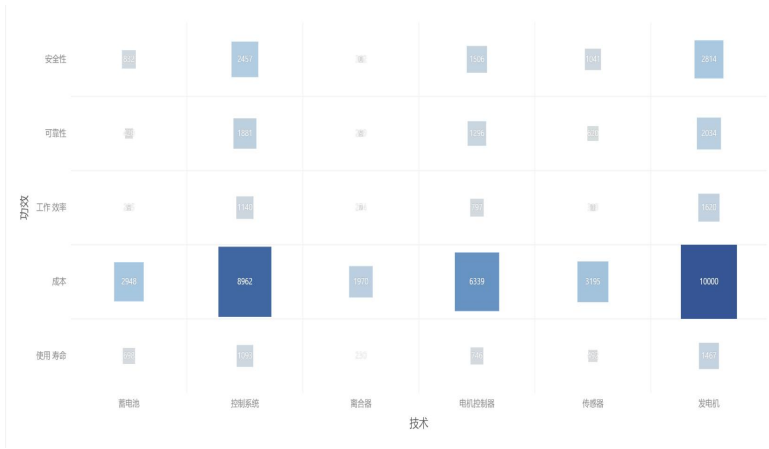
3rd Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents
EEKE2022 @ JCDL2022

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.

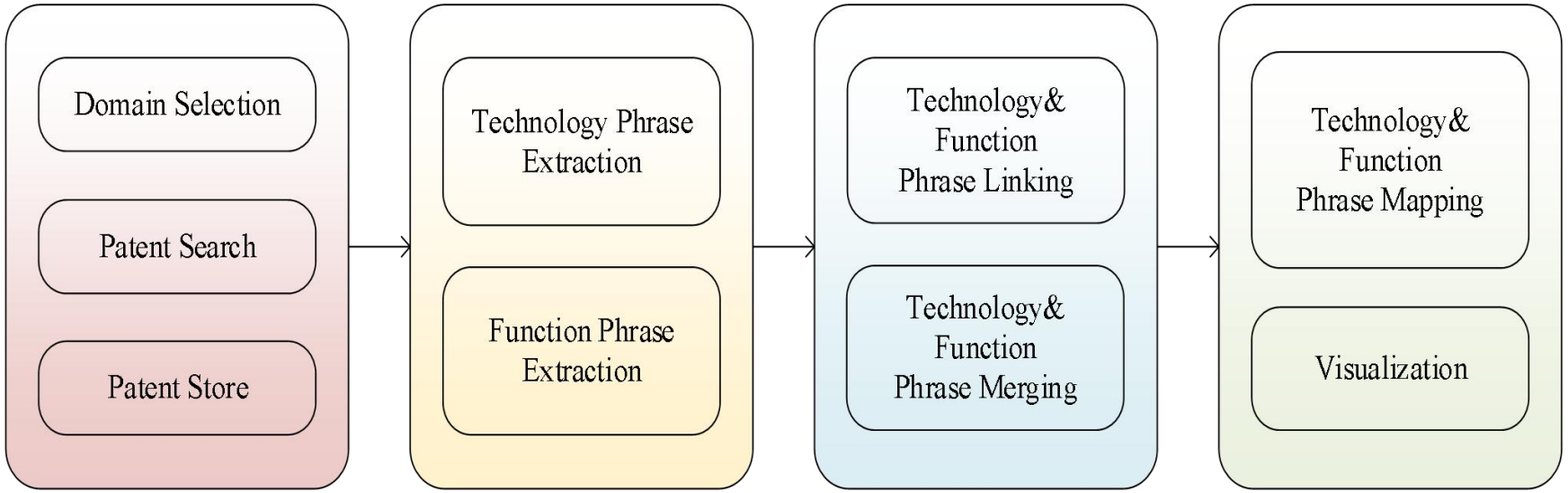


- **Technology Function Matrix**(TFM)： It is an important basis for patent analysis, such as high-value technology discovery and potential technology function prediction.



Contribution

- Construction process of Technology Function Matrix



- Contribution

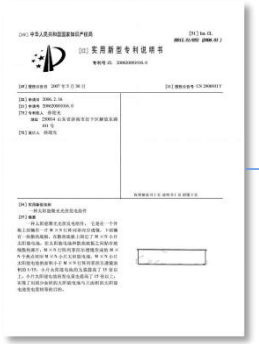
Technology Framework

Semi-Supervised Method

TFM Construction System

Methodology

Patent



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

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...)

Function phrase extraction based on dependency analysis and template

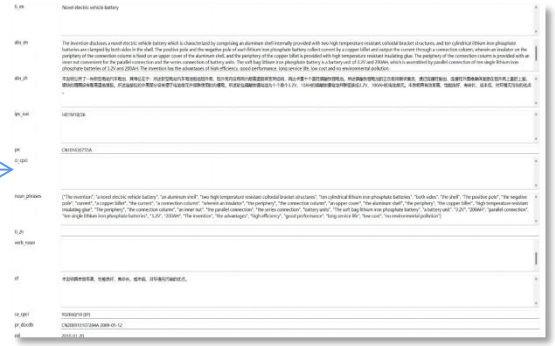
Technology Phrase

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）

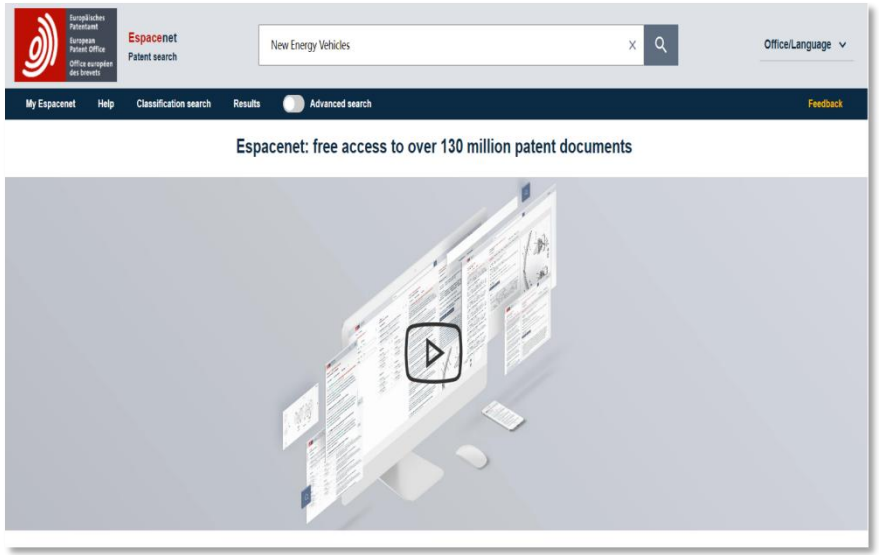
降低等效电阻（reduce equivalent resistance）
提高电导率（increase conductivity）
提高功率密度（increase power density）

Function Phrase

Result



Dataset

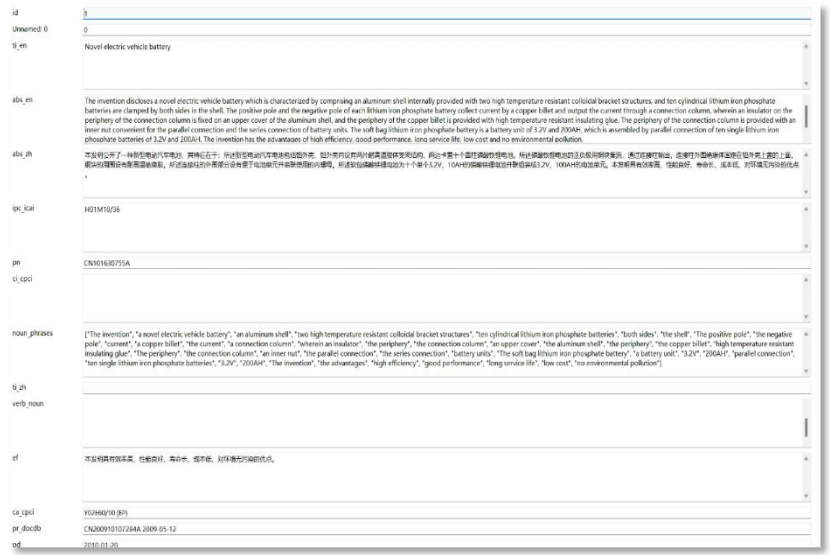


Espacenet(Open Source)

New Energy Vehicles

Keywords

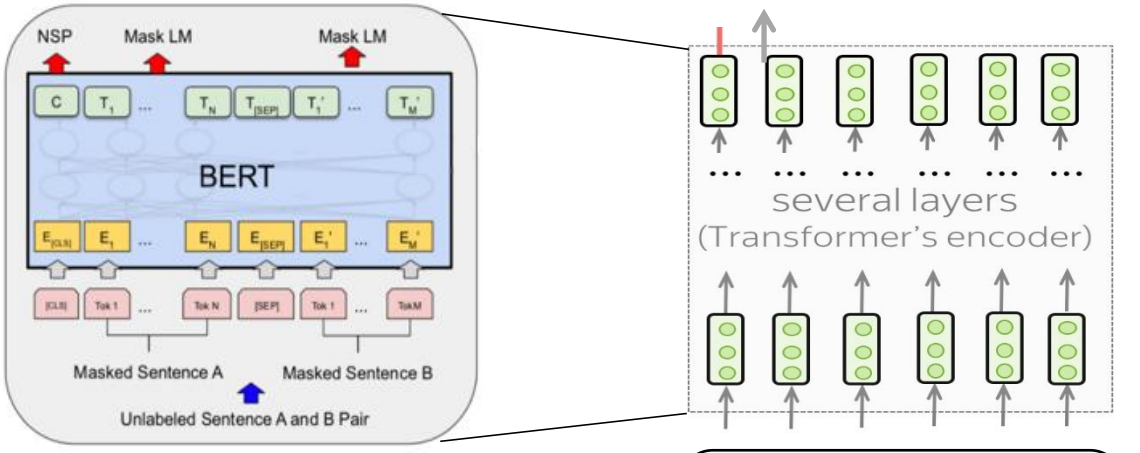
Web Crawler



IPC, CPC, and Chinese abstracts

https://worldwide.espacenet.com/

Function Phrase Extraction

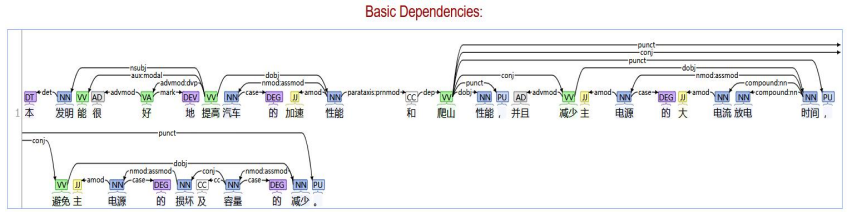


Bert Model

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

STEP1
Function sentence recognition

STEP2
Function phrase extraction

Technology Phrase Extraction

- Semantic Dependency Analysis

***core word*: domain vocabulary**

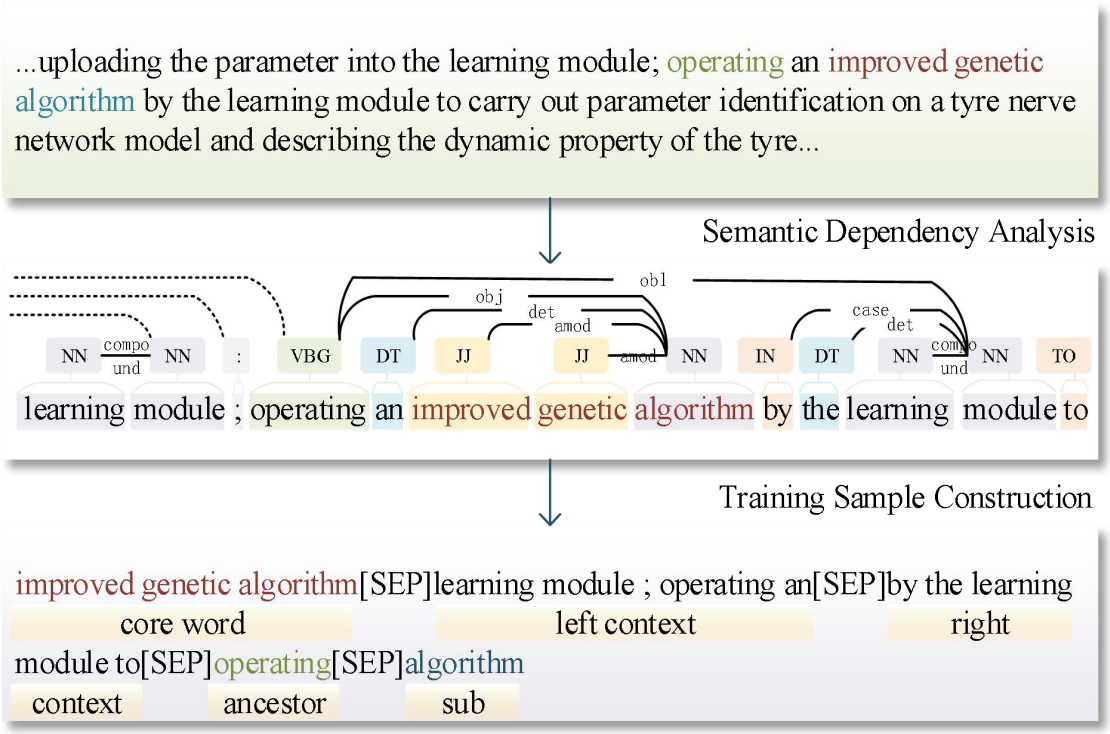
***context*: left and right five words**

***ancestor*: syntactic parent of the core word**

***sub-word*: core word that removes modifier.**

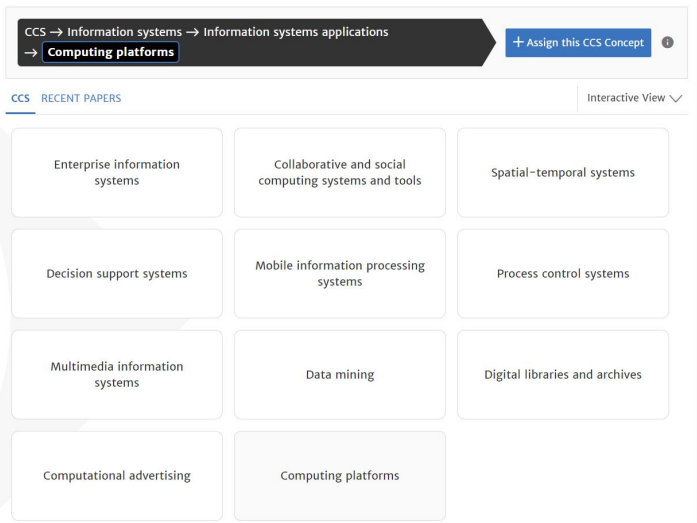
- Training set generation

core word + context + ancestor + sub-word



Technology&Function Phrase Merging

- Directory tree crawling.



- Suffix tree pattern recognition.

—*suffix tree string matching algorithm*

"author LDA" and "LDA"

- Abbreviation recognition.

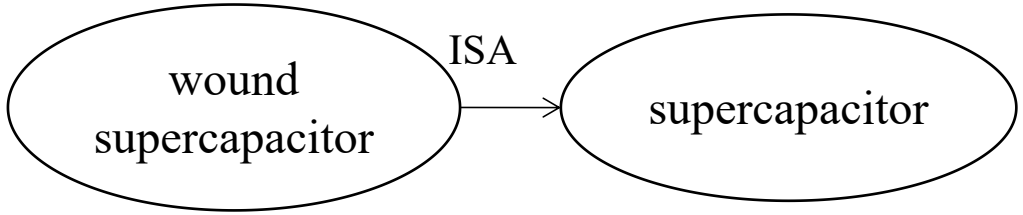
—*Maximum Entropy Model*

$$p(y|x) = \underset{p}{argmax} \sum_{x,y} -\tilde{p}(x)p(y|x)\,log\,p(y|x)$$

—*Example*

"Support Vector Machine" and "SVM"

- Domain triplet recognition.



Experiment

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

Table 1: Comparative experimental results of function sentence recognition.

Algorithm	Accuracy
Naive Bayes	65.86
Word2Vec+MLP	65.67
Bert	89.13

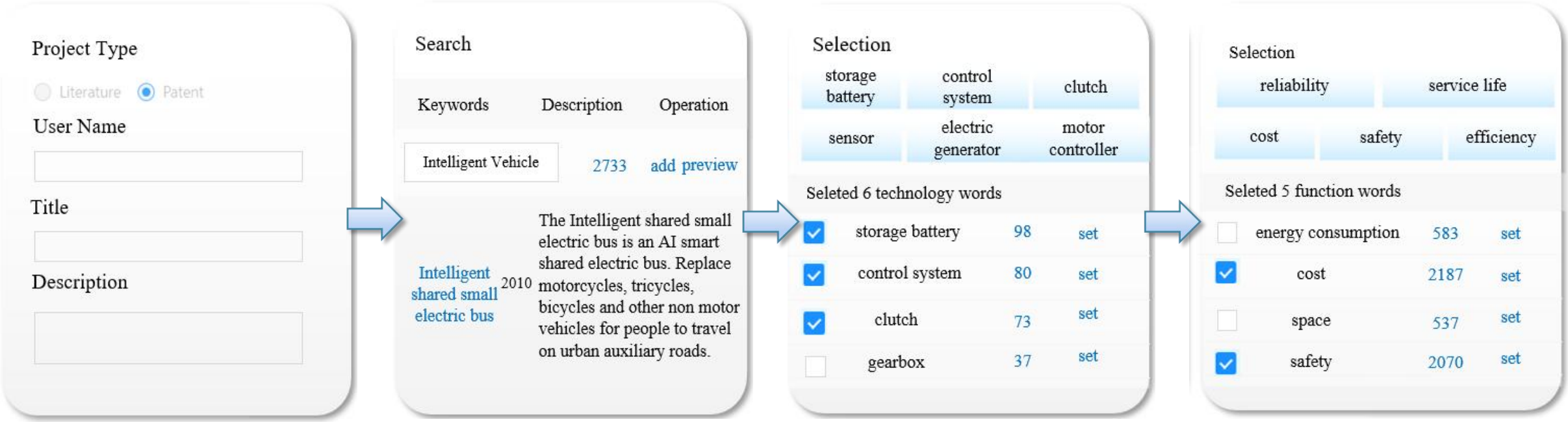
Table 2: Experimental results of function phrase extraction.

Algorithm	Precision	Recall	F1 score
SAO	20.14	26.16	22.76
SDP + Template	56.83	48.59	52.39

Table 3: Technology Phrase Extraction Measurement

Algorithm	Precision	Recall	F1 score
Span-BERT	35.55	81.25	49.46
+ Ancestor	53.81	52.14	51.61
+ Sub	46.94	63.50	52.96
+ Ancestor + Sub	47.90	60.52	53.48

System Overview



STEP1
Project Creation

STEP2
Patent Retrieval

STEP3
Technology Selection

STEP4
Function Selection

System Overview

STEP5
Visualization

Thanks