

# Automatic Construction of Technology Function Matrix

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



Wuhan University, China

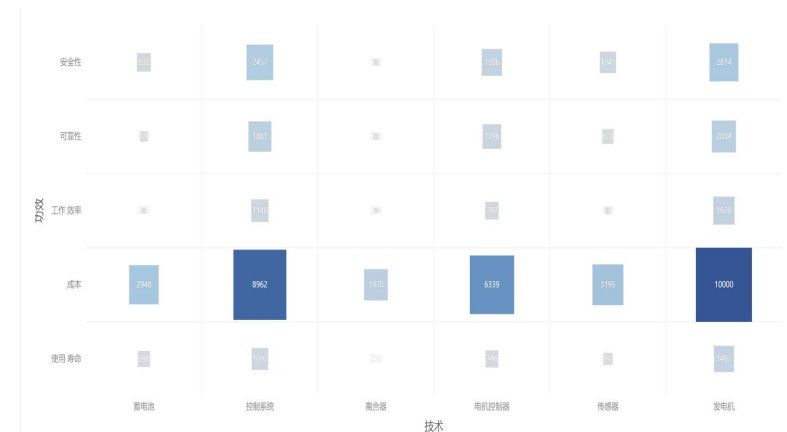
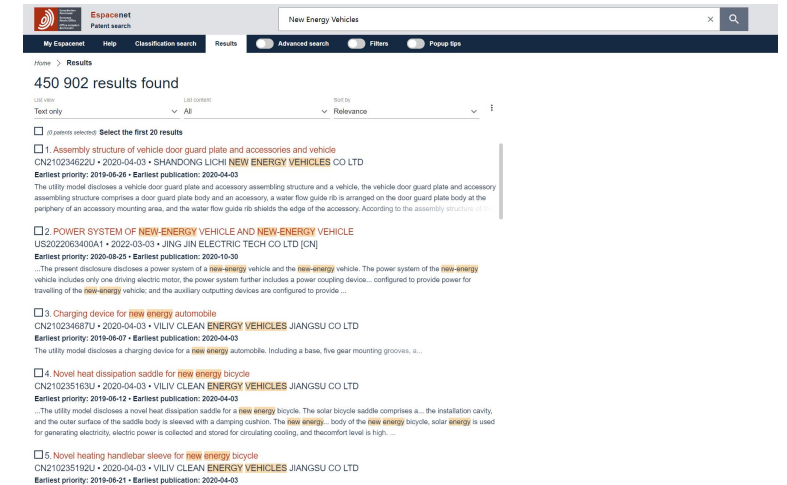
3rd Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents  
EEKE2022 @ JCDL2022





# 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.

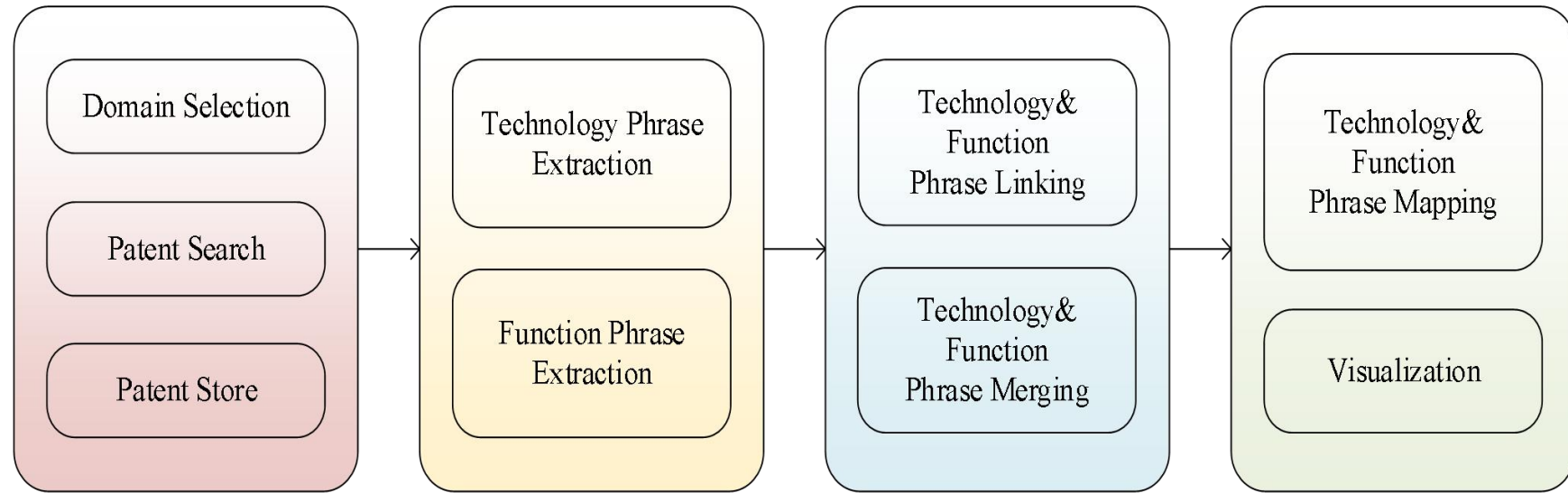




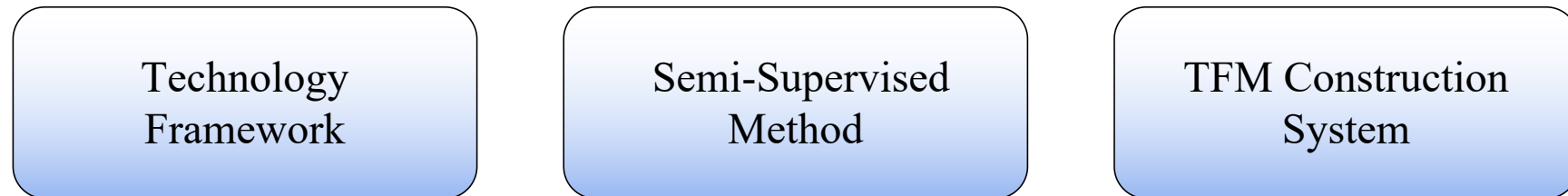


# Contribution

## ➤ Construction process of Technology Function Matrix



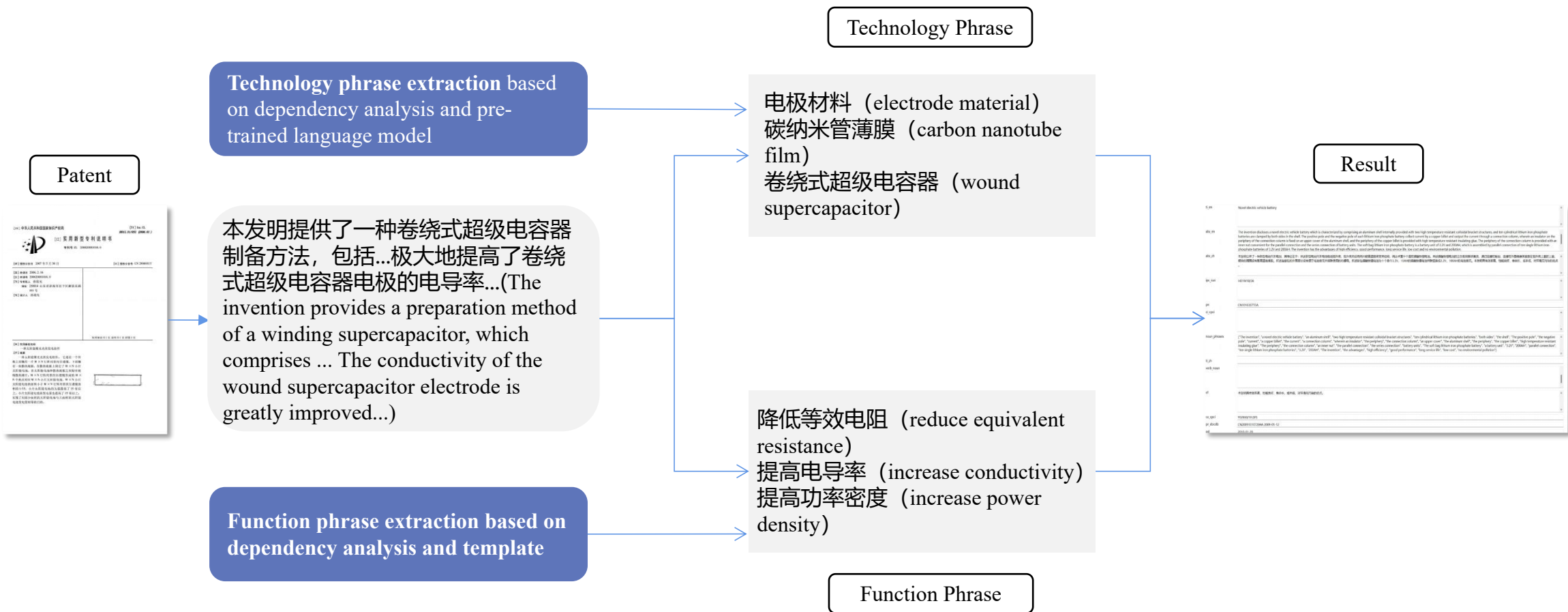
## ➤ Contribution





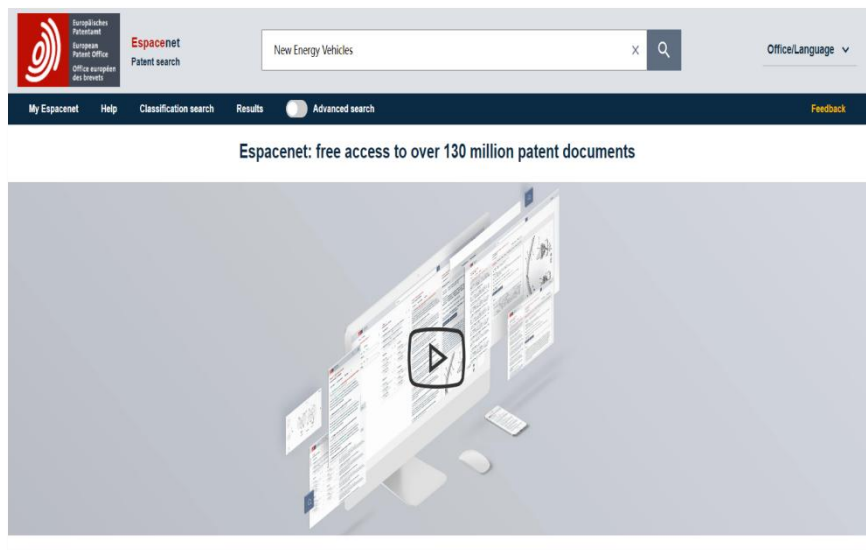


# Methodology





# Dataset



Espacenet(Open Source)

New Energy Vehicles



Keywords



Web Crawler

|             |   |
|-------------|---|
| id          | 0   |
| Unreleased  | 0   |
| is_en       | Novel electric vehicle battery  |
| abs_en      | The invention discloses a novel electric vehicle battery which is characterized by comprising an aluminum shell internally provided with two high temperature resistant colloidal bracket structures, and ten cylindrical lithium iron phosphate batteries are clamped by both sides in the shell. The positive pole and the negative pole of each lithium iron phosphate battery collect current by a copper billet and output the current through a connection column, wherein an insulator on the periphery of the connection column is fixed on an upper cover of the aluminum shell, and the periphery of the copper billet is provided with high temperature resistant insulating glue. The periphery of the connection column is provided with an inner nut convenient for the parallel connection and the series connection of battery units. The soft bag lithium iron phosphate battery is a battery unit of 3.2V and 200Ah, which is assembled by parallel connection of ten single lithium iron phosphate batteries of 3.2V and 200Ah. The invention has the advantages of high efficiency, good performance, long service life, low cost and no environmental pollution. |
| abs_zh      | 本发明公开了一种新型电动汽车电池，其特征在于：所述新型电动汽车电池包括铝壳，铝壳内部设有两个耐高温胶柱支架结构，两个支架十个圆柱形磷酸铁锂电池，锂电池的正极和负极通过铜片夹取电流，铜片夹取电流通过连接柱输出，连接柱夹取电流通过铝壳上的螺母进行上下螺母的连接，螺母的连接方便电池的并联连接和串联连接。所述新型电动汽车电池的铝壳内部设有两个耐高温胶柱支架结构，两个支架十个圆柱形磷酸铁锂电池，锂电池的正极和负极通过铜片夹取电流，铜片夹取电流通过连接柱输出，连接柱夹取电流通过铝壳上的螺母进行上下螺母的连接，螺母的连接方便电池的并联连接和串联连接。所述新型电动汽车电池的铝壳内部设有两个耐高温胶柱支架结构，两个支架十个圆柱形磷酸铁锂电池，锂电池的正极和负极通过铜片夹取电流，铜片夹取电流通过连接柱输出，连接柱夹取电流通过铝壳上的螺母进行上下螺母的连接，螺母的连接方便电池的并联连接和串联连接。   |
| ipc_cls     | H01M10/26   |
| pn          | CN101630755A  |
| ci_cpci     |   |
| non_glnvns  | ["The invention", "a novel electric vehicle battery", "an aluminum shell", "two high temperature resistant colloidal bracket structures", "ten cylindrical lithium iron phosphate batteries", "both sides", "the shell", "The positive pole", "the negative pole", "current", "a copper billet", "the current", "a connection column", "wherein an insulator", "the periphery", "the connection column", "an upper cover", "the aluminum shell", "the periphery", "the copper billet", "high temperature resistant insulating glue", "the periphery", "the connection column", "an inner nut", "the parallel connection", "the series connection", "battery units", "The soft bag lithium iron phosphate battery", "a battery unit", "3.2V", "200Ah", "parallel connection", "ten single lithium iron phosphate batteries", "3.2V", "200Ah", "The invention", "the advantages", "high efficiency", "good performance", "long service life", "low cost", "no environmental pollution"]   |
| is_zh       |   |
| verb_noun   |   |
| df          | 本发明具有效率高、性能设计、寿命长、成本低、无污染等优点。   |
| ci_cpci     | Y02E09/10 (B)   |
| pr_abstract | CN100910102AA 2009-05-12  |
| df          | 2010-01-20  |

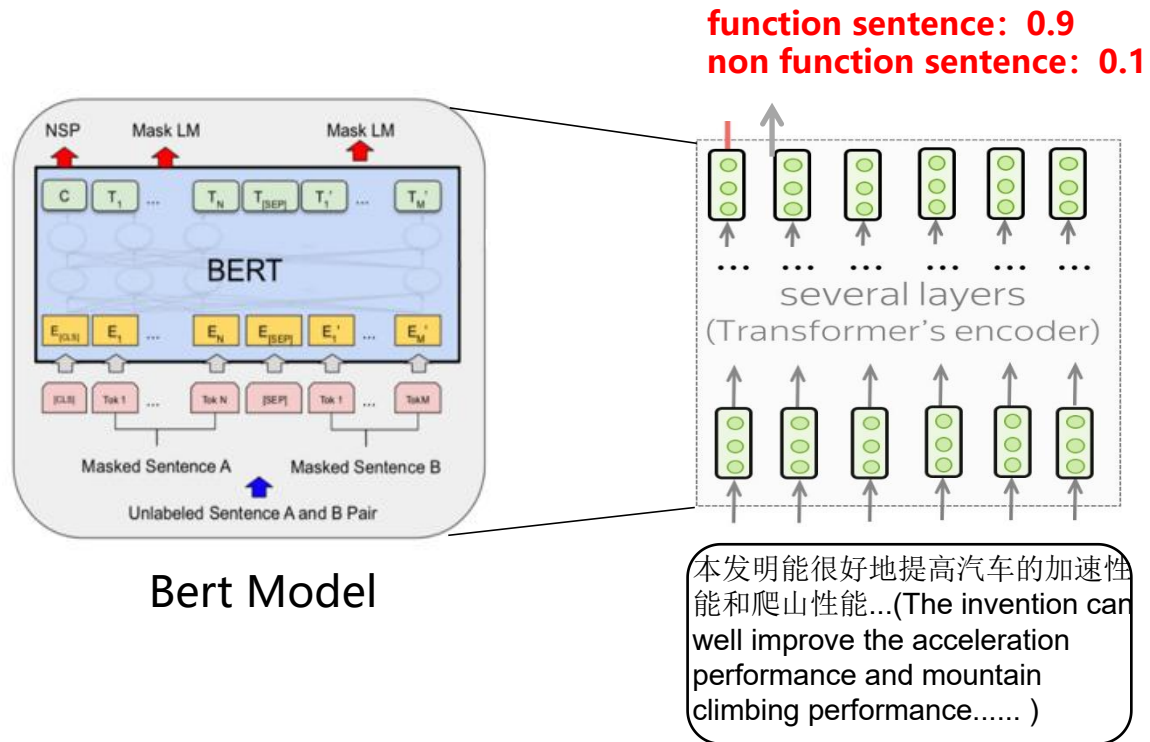
IPC, CPC, and Chinese abstracts

<https://worldwide.espacenet.com/>





## Function Phrase Extraction



## STEP 1

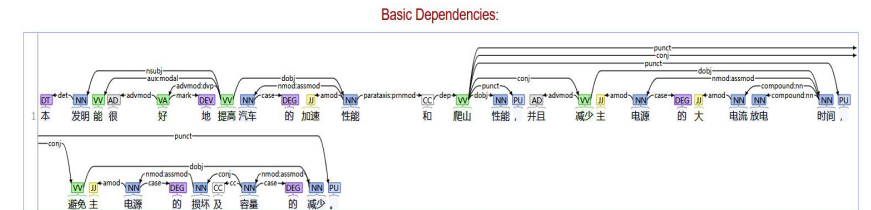
## Function sentence recognition



- 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

## STEP2

## Function phrase extraction





# Technology Phrase Extraction

## ➤ Semantic Dependency Analysis

**core word:** domain vocabulary

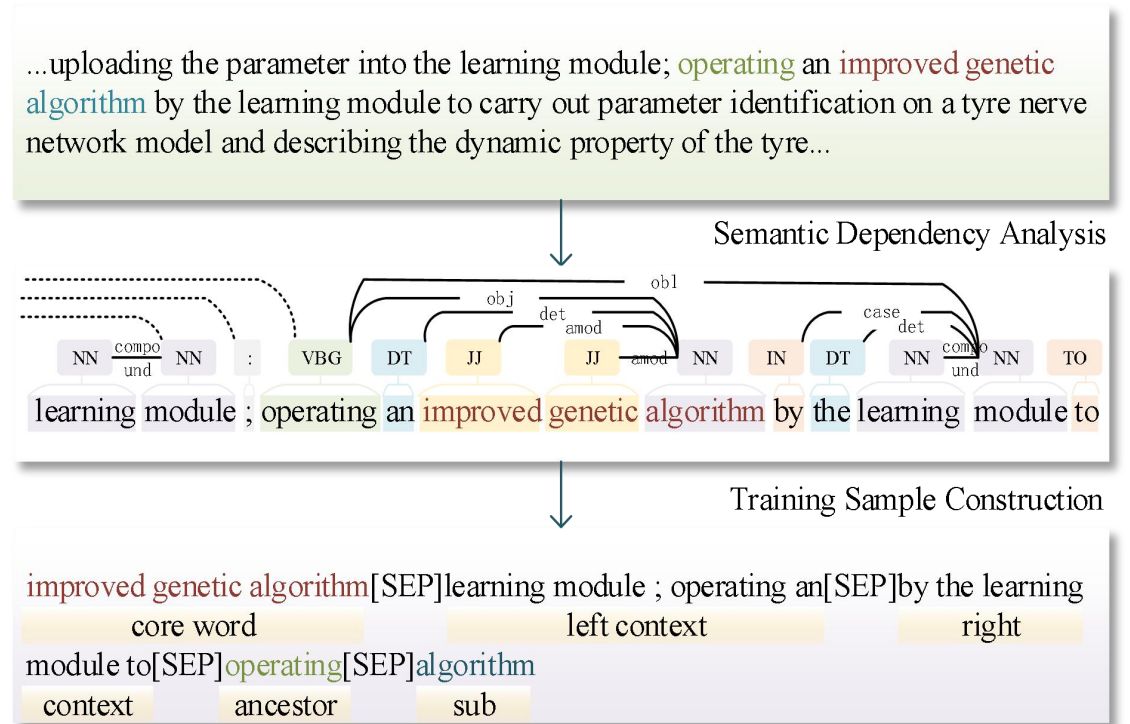
**context:** left and right five words

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

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

## ➤ Training set generation

**core word + context + ancestor + sub-word**

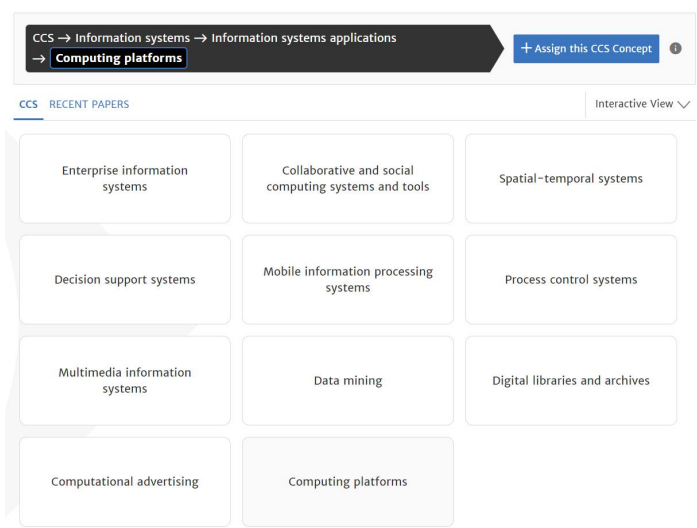






# Technology&Function Phrase Merging

## ➤ Directory tree crawling.



## ➤ Suffix tree pattern recognition.

—*suffix tree string matching algorithm*

"author LDA" and "LDA"

## ➤ Abbreviation recognition.

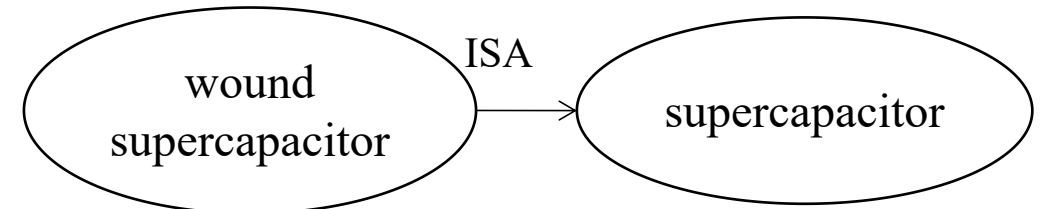
—*Maximum Entropy Model*

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

—*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.**

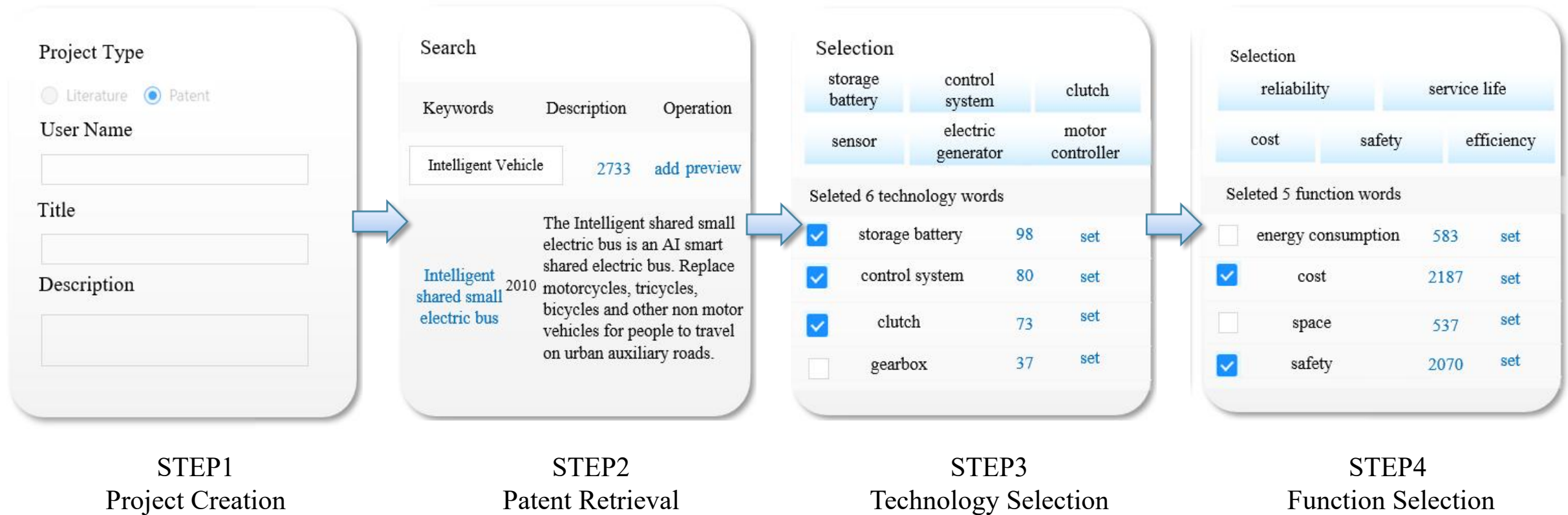
| Algorithm      | Precision | Recall | F1 score |
|----------------|-----------|--------|----------|
| SAO            | 20.14     | 26.16  | 22.76    |
| SDP + Template | 56.83     | 48.59  | 52.39    |

**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





# System Overview



功效矩阵-专利



STEP5  
Visualization





Thanks