

# TECHNOLOGY EVOLUTION ANALYSIS BASED ON SPO USING PATENT DOCUMENTS: A CASE STUDY OF INDUCED PLURIPOTENT STEM CELLS

Hu Z.Y., Wen Y., Liu C.J.

Chengdu Library and Information Center, Chinese Academy of Sciences (CLAS, CAS)

Wei L.

School of Information and management, Shanxi University of Finance and Economics (SIM, SUFE)

Qin X.C.

Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences (GIBH, CAS)

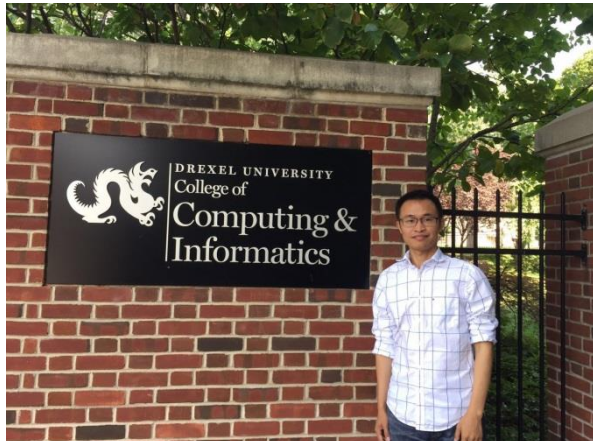
Atlanta, GA, USA

**Sep. 11, 2018**

The background of the slide is a photograph of a bright blue sky filled with fluffy white clouds. At the bottom of the image, there is a horizontal yellow banner. The word "Introduction" is written in a large, bold, red font, centered on the sky background.

# Introduction

*Chengdu Library and Information Center, Chinese Academy of Sciences*



## Liu Chunjiang

Associate Professor

Techmining and Knowledge Discovery Team

Chengdu Library and Information Center, Chinese Academy of Sciences(CAS)

### Ongoing projects:

1. Informationization Special Project of Chinese Academy of Sciences “EScienceApplication for Knowledge Discovery in Stem Cells” (Grant No: XXH13506-203)
2. National Social Science Foundation of China “Theory and Applications Research of Subject-Informatics for Domain Knowledge Discovery” (Grant No: 17ATQ008)



### Stem Cell Innovation Engine



### Biodiversity Knowledge Resource Center



### Graphene Knowledge Resource Center



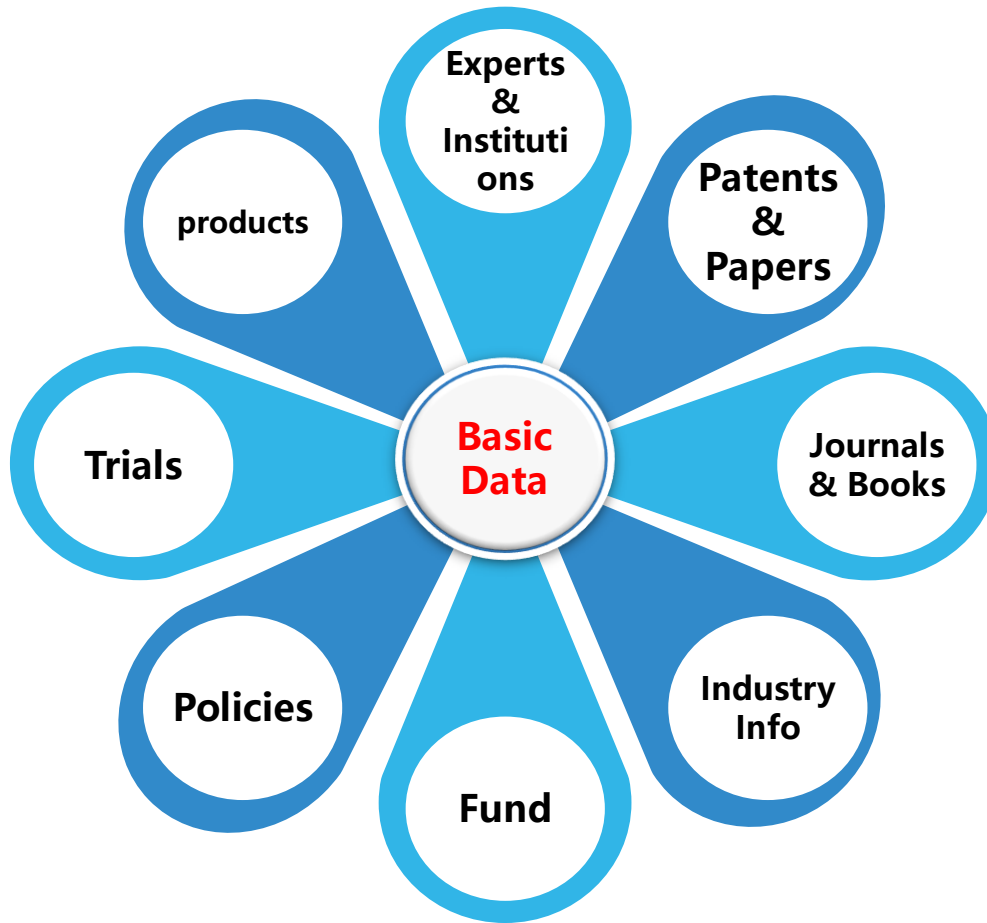
### Neuroscience Knowledge Resource Center



### Medical data analysis platform

Chengdu Library and Information Center, Chinese Academy of Sciences

# stem cell knowledge map



## ■ S&T Information Data

- More than **120,000** records, From **12** kinds of scientific and technological information such as patents, fund, etc.

## ■ Knowledge Mining

- From the perspectives of scientific instruments, animal models, experimental techniques, cell organs, disease genes, **20,000+** knowledge entities

## ■ Knowledge Links

- Based on citation, acknowledgement, knowledge entity.

## ■ Knowledge organization

- Based on R&D projects of international, China and CAS.

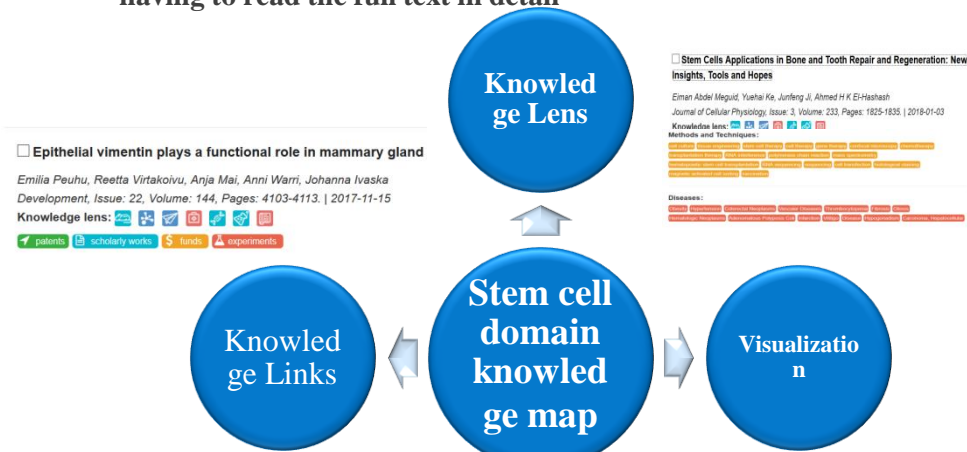


# Some Products



Cooperate with Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences (GIBH, CAS)

1. Quick and comprehensive knowledge of scientific literature without having to read the full text in detail



Stem cell Letters

广东省科技厅信息

广东省科技厅办公室 2018年4月18日

广东多措并举支持基础研究，  
“世界首例亨廷顿舞蹈症基因敲入猪”重大科技成果诞生

Policy advices



Research reports



Knowledge service

2. Realize knowledge links between papers, patents, etc. based on knowledge maps

3. Visual display of stem cell knowledge

4. Multi-view, fine-grained disclosure of stem cell technology information, providing accurate retrieval and knowledge navigation services

Chengdu Library and Information Center, Chinese Academy of Sciences

The background of the slide is a photograph of a clear blue sky filled with fluffy white clouds. At the bottom of the image, there is a horizontal strip showing the tops of trees with bright yellow leaves, suggesting an autumn setting.

# **Research background**

*Chengdu Library and Information Center, Chinese Academy of Sciences*

# Knowledge Discovery in Biomedical Literature (KDiBL)

## ■ KDiBL

- Knowledge Discovery in Biomedical Literature is the process of identifying and extracting the new, useful, potential and understandable patterns from the biomedical literatures in a credible way.

## ■ Information extraction

- Information extraction plays an important part in KDiBL , which automatically extracts the specific terms, the corresponding characteristics and the semantic relations among them from the texts as the basic knowledge unit of knowledge discovery. SPO predication is a typical form of extraction.

## ■ SPO predications

- SPO predications consist of a Subject argument (noun phrase), an object argument (noun phrase), and the relation that binds them (verb phrase), which can represent science and technology (S&T) information with more details in a simple manner and have been widely applied in Knowledge Discovery in Biomedical Literature (KDiBL)

# Technology Evolution Analysis

## ■ Object

- By means of bibliometric analysis, natural language processing, text mining and other technical means to outline the development trend of technological innovation in the field. In the process, the technical connotation of multi-source heterogeneous data in the field will be extracted, identified, analyzed, reconstructed, screened and calculated.

## ■ Effect

- It can trace the main technological development paths in the field from the perspective of birth, development and downfall of technology, provide comprehensive decision-making information from the complete vision of the technology chain, help decision-makers to clarify the mainstream of technology development, identify key technologies and emerging technologies, and grasp the future of technology direction of development.



The background of the slide is a photograph of a bright blue sky filled with fluffy white clouds. At the bottom of the image, there is a horizontal yellow banner. The title 'Literature Review' is centered in the upper half of the slide, overlaid on the sky. The text is in a bold, red, sans-serif font.

# Literature Review

*Chengdu Library and Information Center, Chinese Academy of Sciences*

# Present State of Research



## Scientific Literatures

- Mainly focus on the method that can be used to draw technology evolution map of keywords by calculating the distributions of keywords over the documents cluster groups.



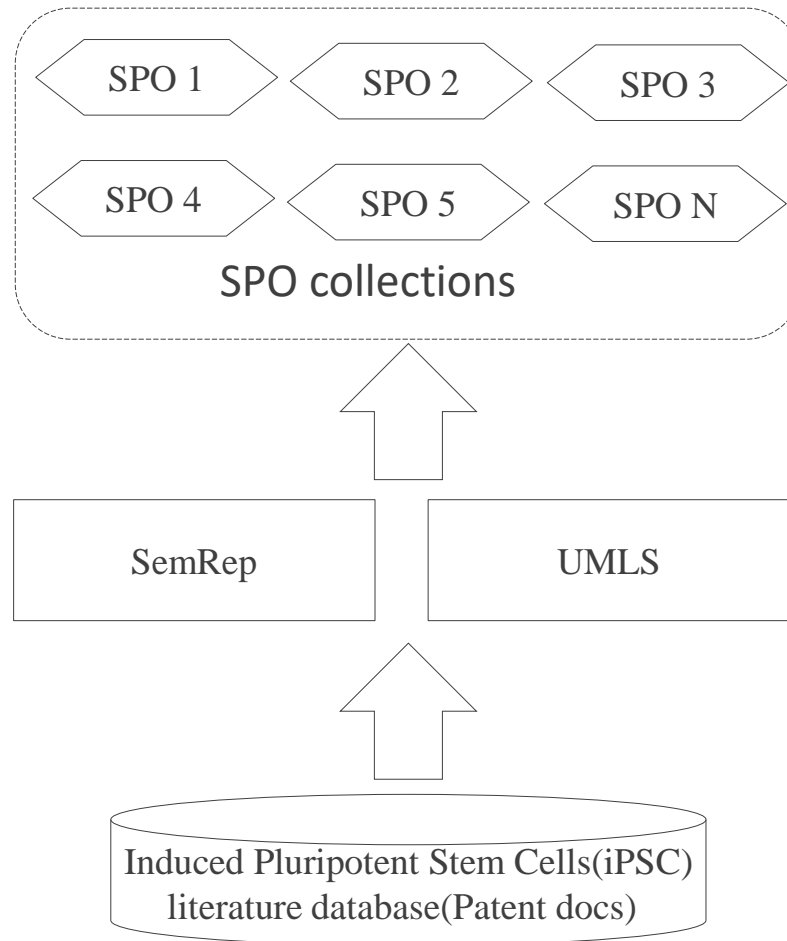
## Limitations

- Lack of semantic information: keywords lack semantic information, knowledge discovery needs automatically extracts information such as entities, entity attributes and semantic relations between entities as the basic knowledge unit.
- Lack of repeatable: the performance of various extraction tools is different, and the applicability is different.
- Technology Evolution Analysis turns into Topic Evolution Analysis: Technologies Evolution Analysis is composed of the evolution analysis of technology's problems, solutions, functions or effects and their combination, Topic Evolution Analysis is only part of Technology Evolution Analysis.

# **Metholodogy**

*Chengdu Library and Information Center, Chinese Academy of Sciences*

# Step 1: Extracting SPO Structures



# Tools in KDiBL

## ■ Unified Medical Language System, UMLS<sup>1</sup>

- UMLS is a biomedical integrated super-thesaurus system researched. It integrates multiple vocabularies in the fields of biomedicine, health and health, and uses a combination of string-term-concept to standardize the terminology in the biomedical field and provides an interoperable interface for computer processing.

## ■ MetaMap<sup>2</sup>

- MetaMap is a tool for mapping free words to UMLS concepts. It can mark the UMLS concepts contained in texts. It is widely used in various fields of KDiBL as a basic text processing tool.

## ■ SemRep<sup>3</sup>

- SemRep is one of the important achievements of the NLM semantic knowledge representation project. It is a semantic knowledge extraction and presentation tool for biomedical literature based on UMLS and MetaMap.

1. NCBI.UMLS® Reference Manual[EB/OL].<http://www.ncbi.nlm.nih.gov/books/NBK9676/>.

2. ARONSON A R, LANG F. An overview of MetaMap: historical perspective and recent advances[J]. Journal of the American Medical Informatics Association, 2010, 17(3): 229-236.

3. ARNOLD P, RAHM E. Semrep: a repository for semantic mapping[EB/OL]. [https://dbs.uni-leipzig.de/en/publication/title/semrep\\_a\\_repository\\_for\\_semantic\\_mapping](https://dbs.uni-leipzig.de/en/publication/title/semrep_a_repository_for_semantic_mapping).

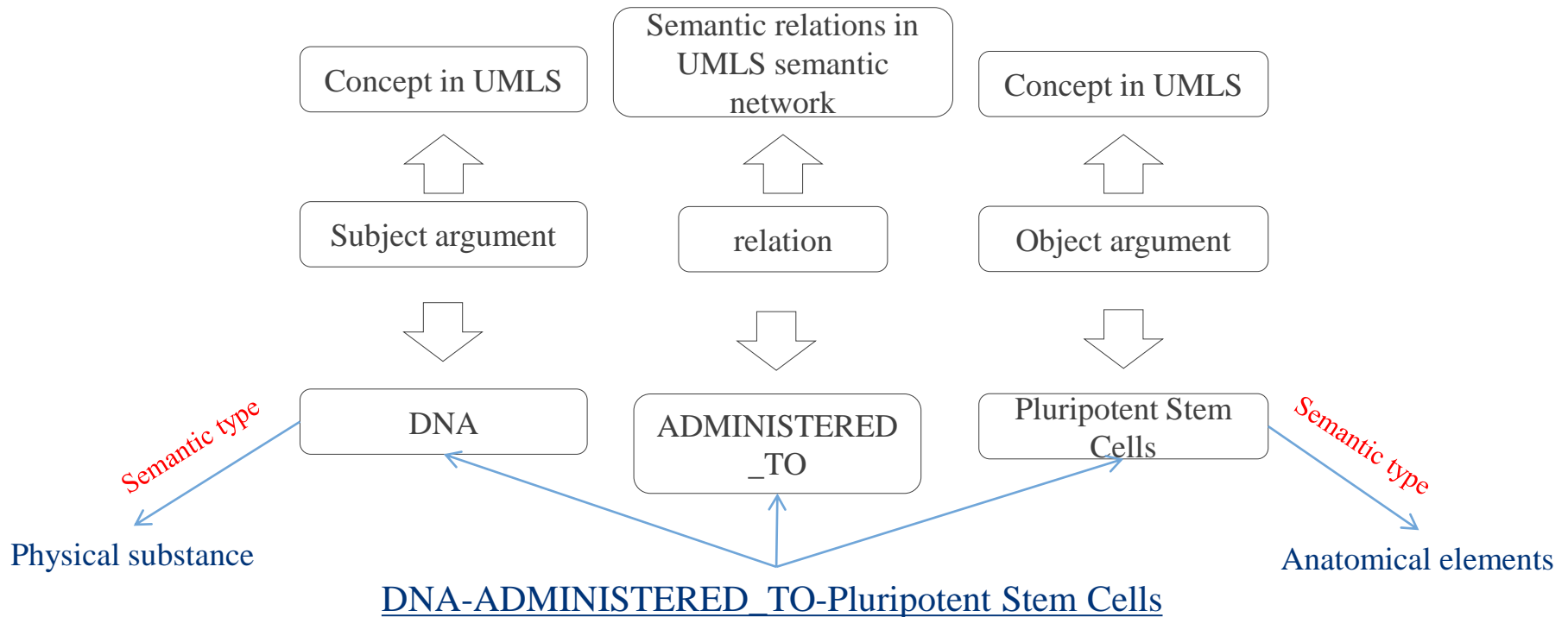


# Tools in KDiBL

## ■ . The UMLS Semantic

- 133 semantic types and 54 semantic relationships

## ■ Example



## Step 2: Clustering the patent documents

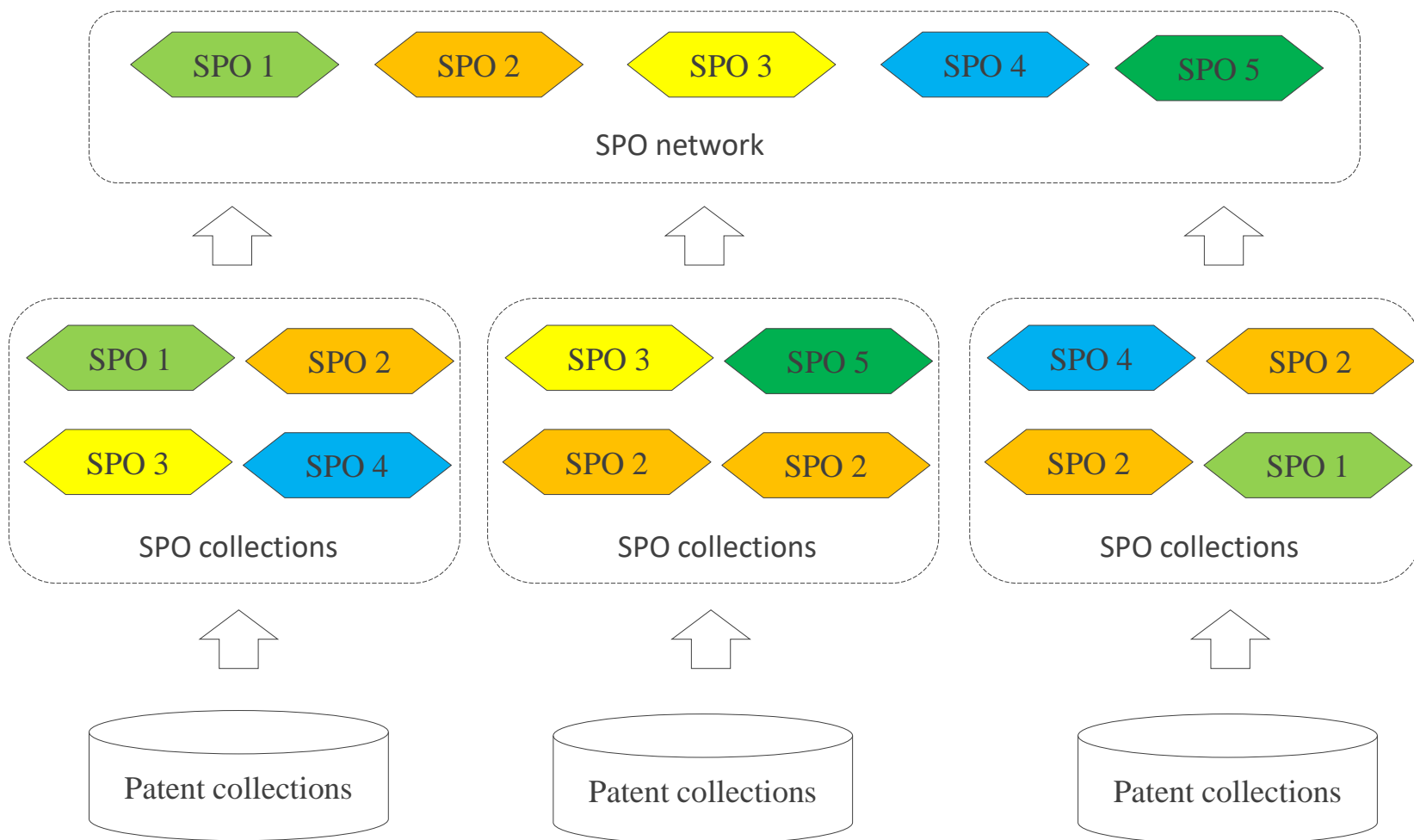
### ■ Fuzzy clustering analysis

- Clustering analysis using fuzzy mathematics, it shows better than other clustering analysis in terms of runtime and accuracy.

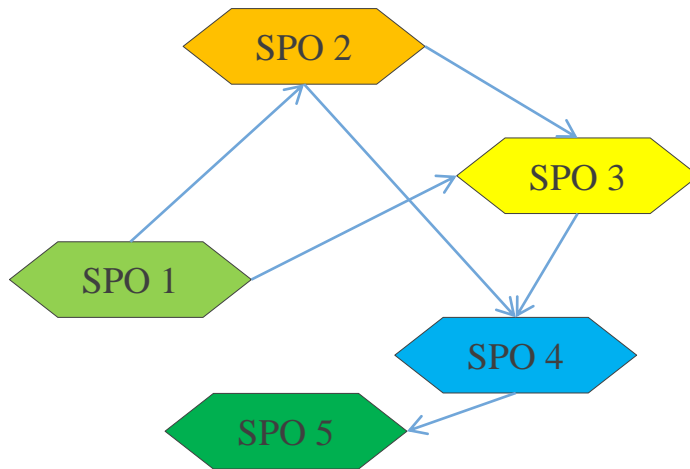
### ■ FCM Algorithm:

- Standardized data matrix;
- Establish a fuzzy similarity matrix and initialize the membership matrix;
- The algorithm starts iterating until the objective function converges to a minimum value;
- According to the iterative result, the class to which the data belongs is determined by the last membership matrix, and the final clustering result is displayed.
- FCM is sensitive to the initial clustering center, and needs to manually determine the number of clusters, which is easy to fall into the local optimal solution.

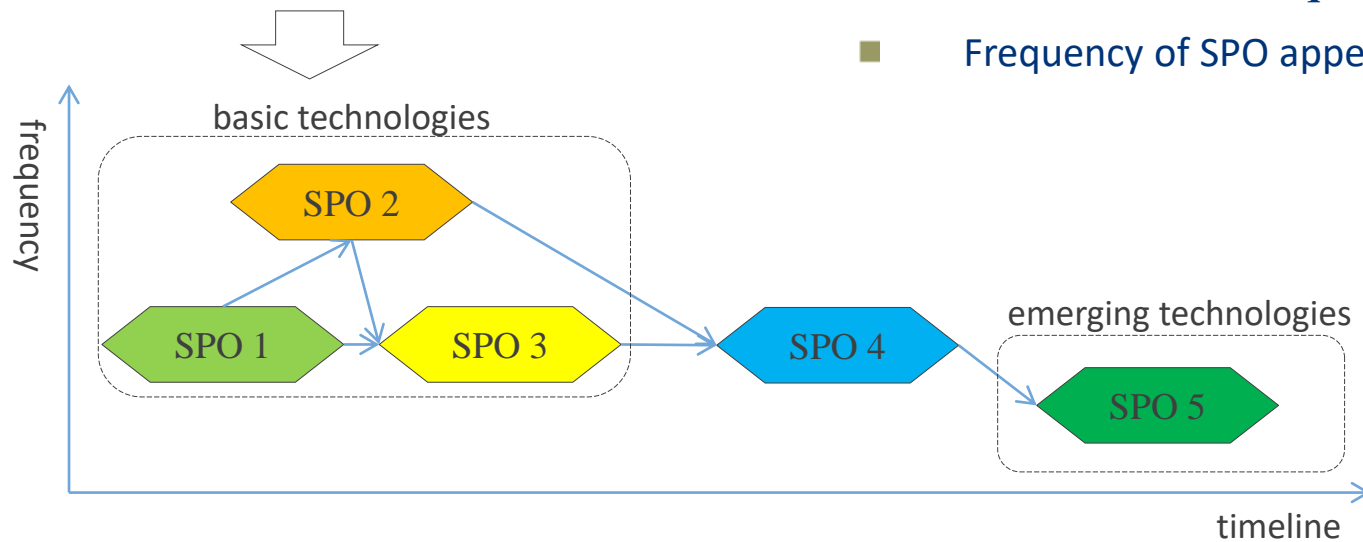
## Step 3: Forming semantic network of SPO Predications



## Step 4: Drawing technology evolution map



- horizontal axis of timeline:
  - The earliest filling date is the earliest priority date or application date of patent documents in which the SPO appears.
- vertical axis of frequency:
  - Frequency of SPO appears in each group.





# **Case Study & Conclusion**

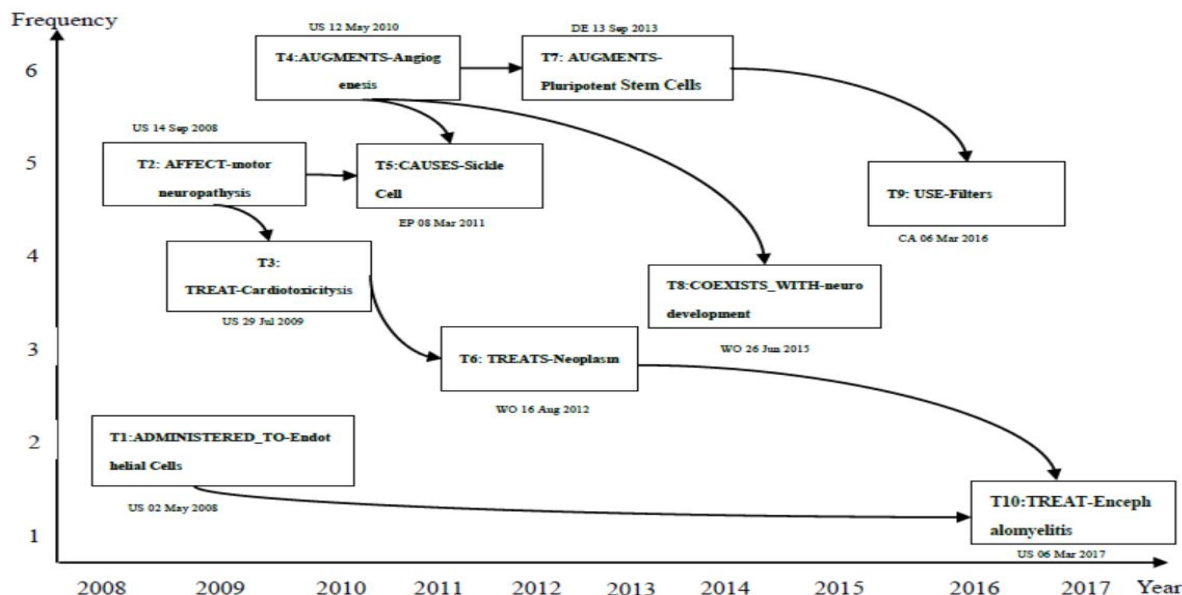
*Chengdu Library and Information Center, Chinese Academy of Sciences*





## Case Study(Ongoing)

- IPSC patents were selected as a case study, Derwent Innovations Index (DII) as data source and 1,282 patent documents are obtained from 2008 to 2017. Following the above methodology, the technology evolution map of IPSC patents is drawn. A part of the map is shown in Figure 1.



## Conclusion

- The result indicates that SPO predications which contain more semantic information are more suitable for technology evolution analysis than keywords. But the nodes need to be attaching understandable labels. SPO predications can be used to generate topics and draw more comprehensive technology evolution map.



中国科学院成都文献情报中心

Chengdu Library and Information Center,  
Chinese Academy of Sciences



中国科学院广州生物医药与健康研究院

GUANGZHOU INSTITUTES OF BIOMEDICINE AND HEALTH, CHINESE ACADEMY OF SCIENCES



# Thank You!

Acknowledgement: The work was supported by the Informationization Special Project of Chinese Academy of Sciences “EScienceApplication for Knowledge Discovery in Stem Cells” (Grant No: XXH13506-203).

*Chengdu Library and Information Center, Chinese Academy of Sciences*