



18th ISKO2024

ISKO International Conference 2024
Wuhan, China March 20-22, 2024



中国科学院成都文献情报中心
NATIONAL SCIENCE LIBRARY(CHENGDU), CHINESE ACADEMY OF SCIENCES

The Method and Evaluation of Use of LLM to Build Task Ontology

—Taking the Hull-less Barley as an Example

Chai Miaoling^{1,2} Zhang Xian^{1,2} * Zeng Yi³ Dawazhuoma⁴

1. National Science Library (Chengdu), Chinese Academy of Sciences;

2. Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, China

3. Information System, College of Business, City University of Hong Kong

4. The Institution of Science and Technology Information of Xizang

Wuhan China

Mar. 21, 2024



中国科学院大学
University of Chinese Academy of Sciences

FUNDED BY THE CHINESE NATIONAL SOCIAL SCIENCE FOUNDATION PROJECT (NO. 18BTQ067).

*ZHANG XIAN, CORRESPONDING AUTHOR.



18th ISKO 2024

ISKO International Conference 2024
Wuhan, China March 20-22, 2024

Who We Are ?



Chai Miaoling

Department of
Information Resources
Management, School of
Economics and
Management, University
of Chinese Academy of
Sciences



National Science
Library (Chengdu),
Chinese Academy of
Sciences

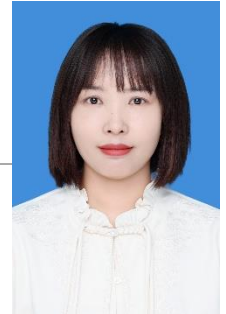


Zhang Xian

The Institution of
Science and
Technology
Information of Xizang



Information System,
College of Business,
City University of Hong
Kong, China



Dawazhuoma



Zeng Yi



中国科学院
CHINESE ACADEMY OF SCIENCES



中国科学院大学
University of Chinese Academy of Sciences



香港城市大學
City University of Hong Kong
專業 創新 觸摸全球
Professional-Creative
For The World



中国科学院成都文献情报中心
NATIONAL SCIENCE LIBRARY(CHENGDU), CHINESE ACADEMY OF SCIENCES



18th ISKO2024

ISKO International Conference 2024
Wuhan, China March 20-22, 2024

contents

01

• Background & Why

02

• Building Knowledge Fusion Model & Method

03

• LLM Evaluation & Selection

04

• Knowledge Fusion: Classification

05

• Results & Discussion





中国科学院成都文献情报中心
NATIONAL SCIENCE LIBRARY(CHENGDU), CHINESE ACADEMY OF SCIENCES



18th ISKO2024

ISKO International Conference 2024
Wuhan, China March 20-22, 2024

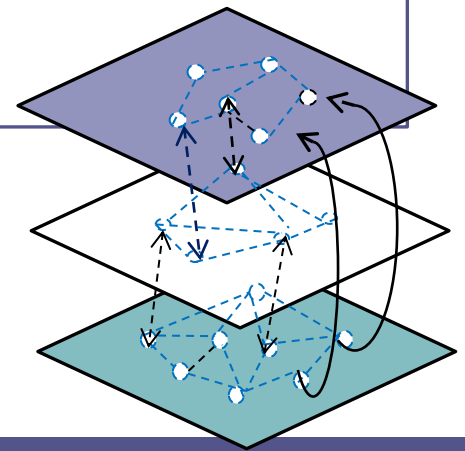
1 Background & Why



中国科学院大学
University of Chinese Academy of Sciences

Focus on

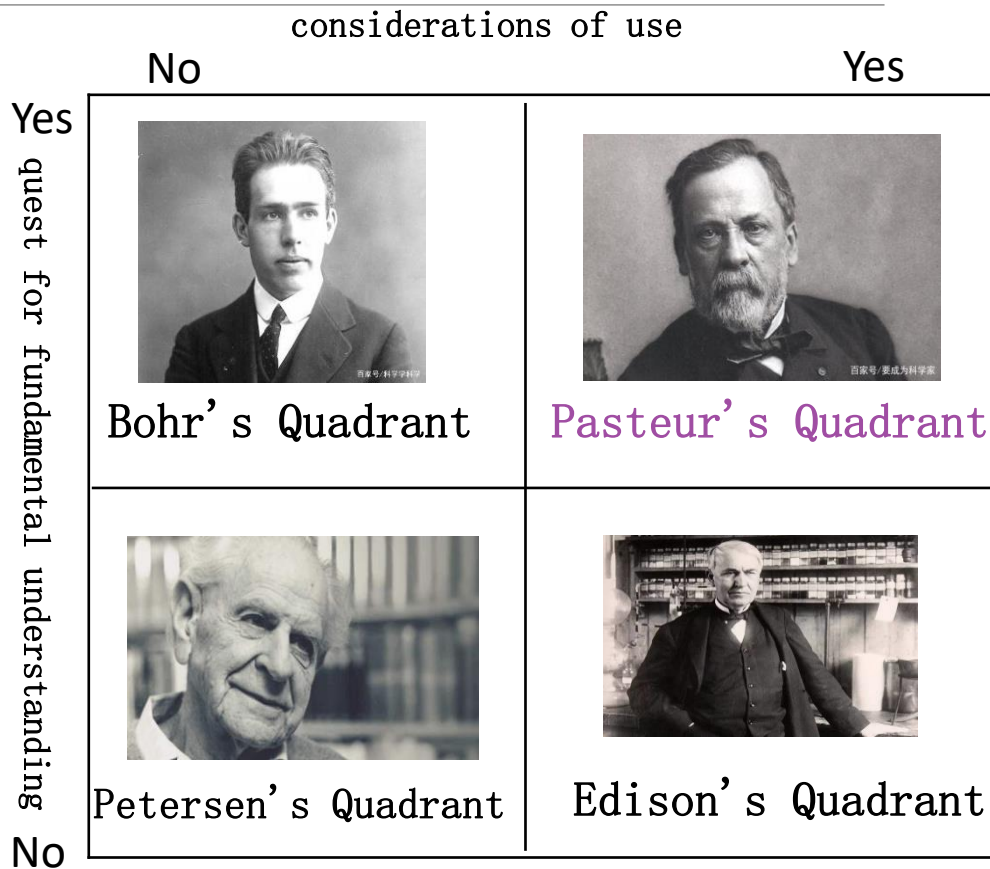
- Identified Scientific Questions (refer to **Pasteur's Quadrant**) from Multimodal data



1.1 Background: What is Pasteur's Quadrant

1997, D. E. Stokes propose the 4 quadrants of Scientific Research.

Pasteur's Quadrant means the research that pursues both fundamental understanding and practical applications, known as use-inspired basic research.



F1. Stokes' 2D-model

1.2 Why Ontology: Multimodal data

T1. The Current Situation and Existing Problems of Data Usage

Data selection	Problem
<i>Journal papers = Science, Patents = Technology</i>	<i>Journal papers ≠ Science, and Patents ≠ Technology</i>
Add conference paper, thesis, reports, news, monographs, funds, funds guides, etc.	Lack of fusion logical semantic granularity Why not multimodal data? Just like Image etc.
How about Multimodal data?	How to extract the semantic. How to fusion the knowledge?


Ontology

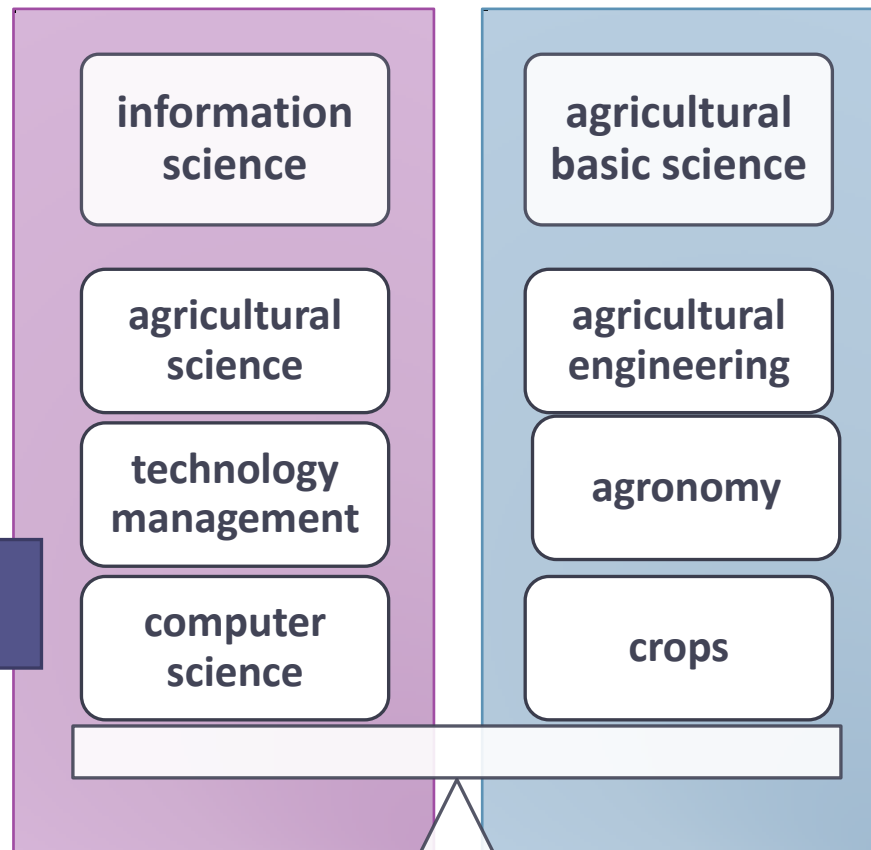


1.2 Why LLM: Understanding Deviation in discipline research

Understanding bias between

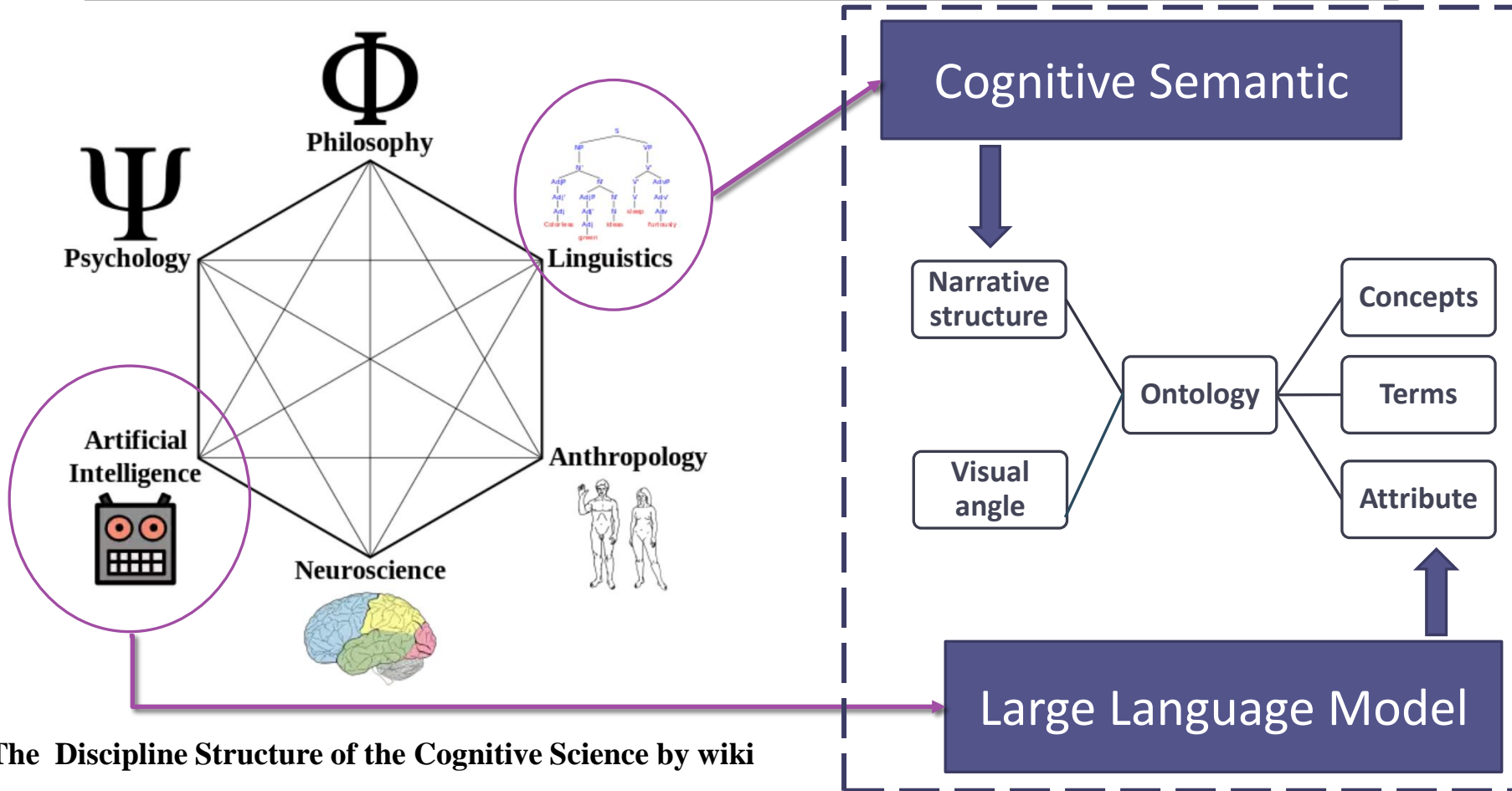


es:
ts



- Why not LLM?
- How to control the uncertainty of LLM?

1.3 How to Combine Them?





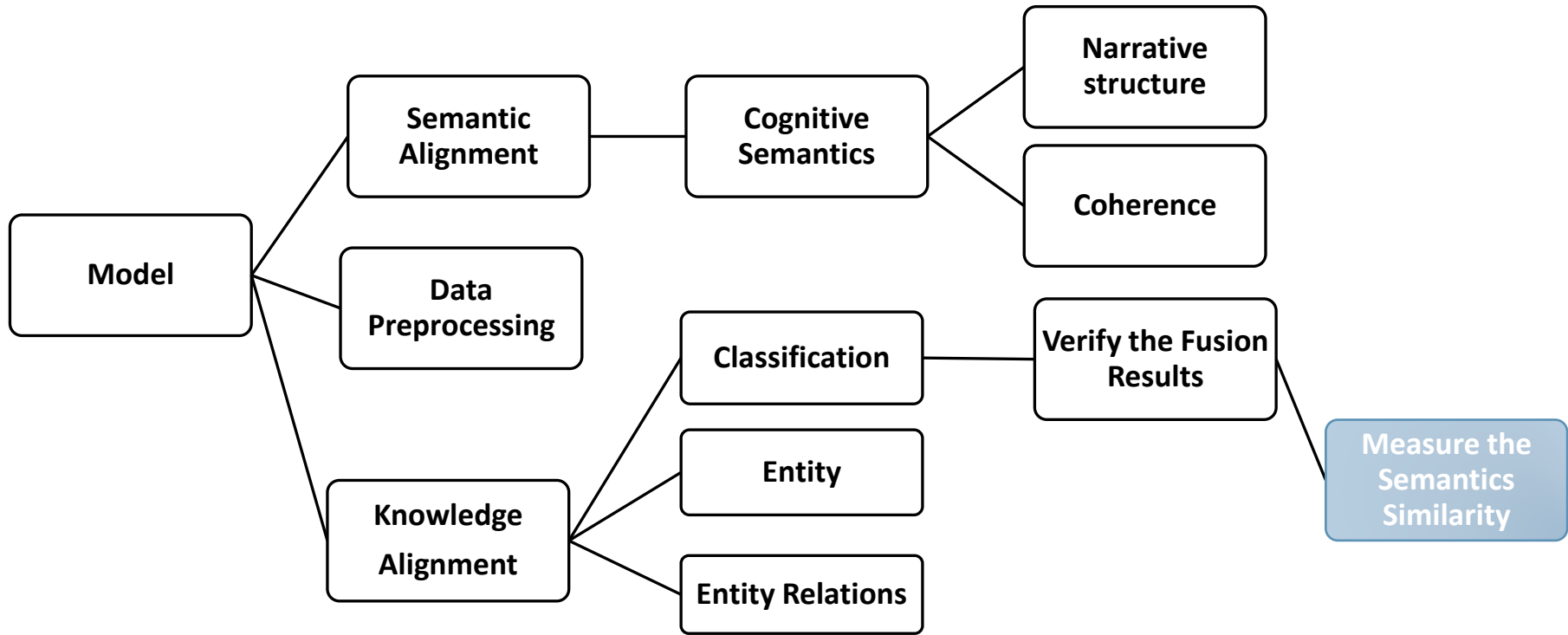
中国科学院成都文献情报中心

NATIONAL SCIENCE LIBRARY(CHENGDU), CHINESE ACADEMY OF SCIENCES

2 Building Knowledge Fusion Model & Method

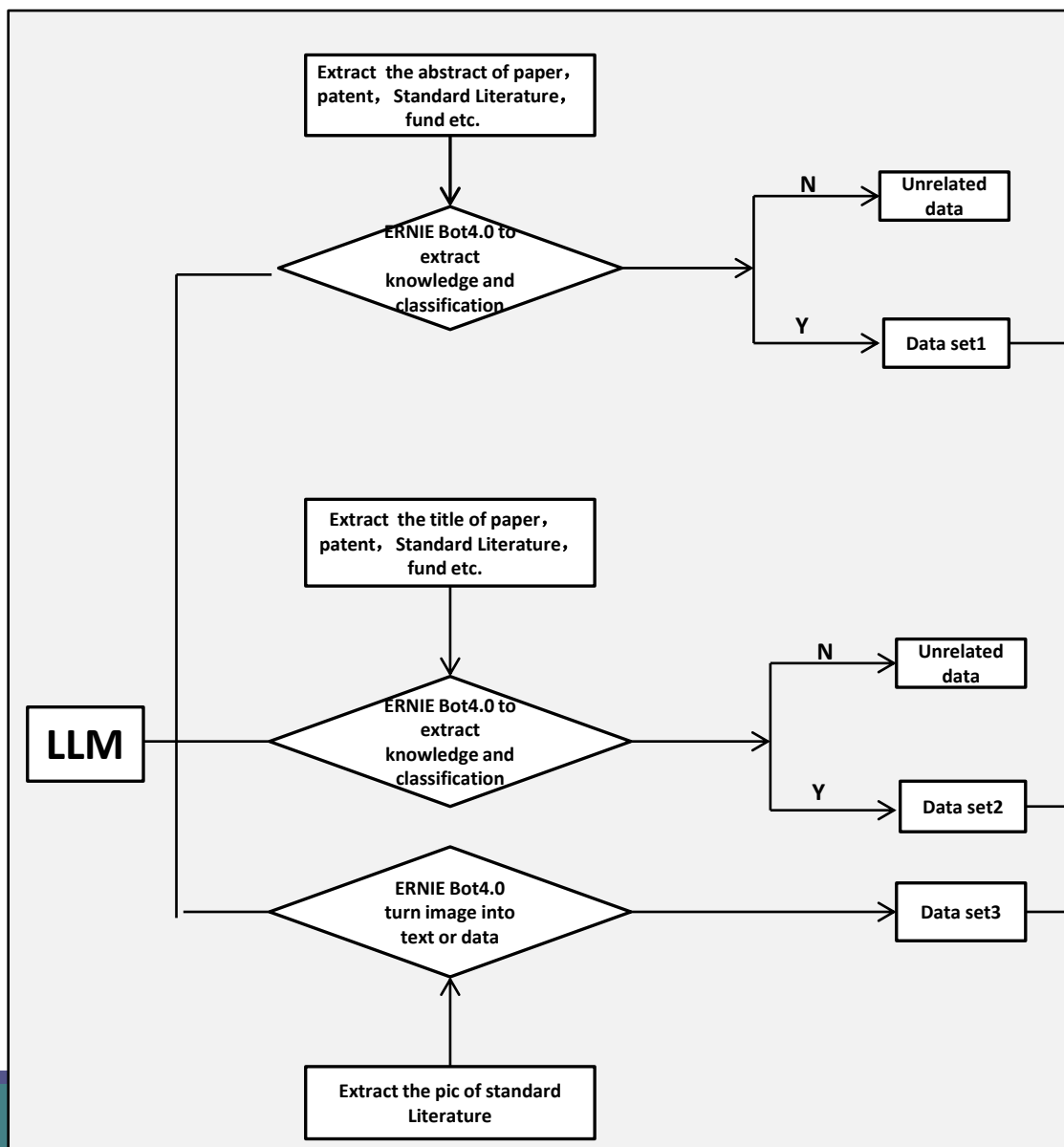


中国科学院大学
University of Chinese Academy of Sciences

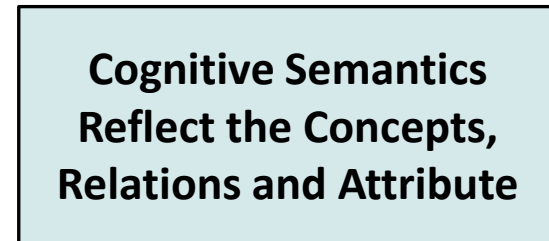


F3. The knowledge fusion Model

Method

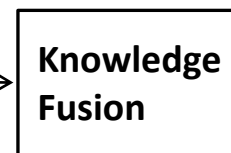


3.1

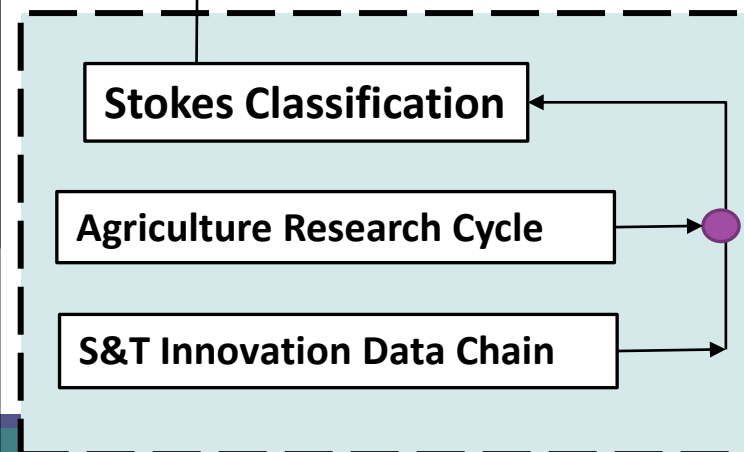
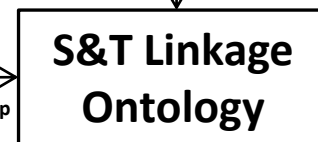


Top-down

3.2



Bottom-up

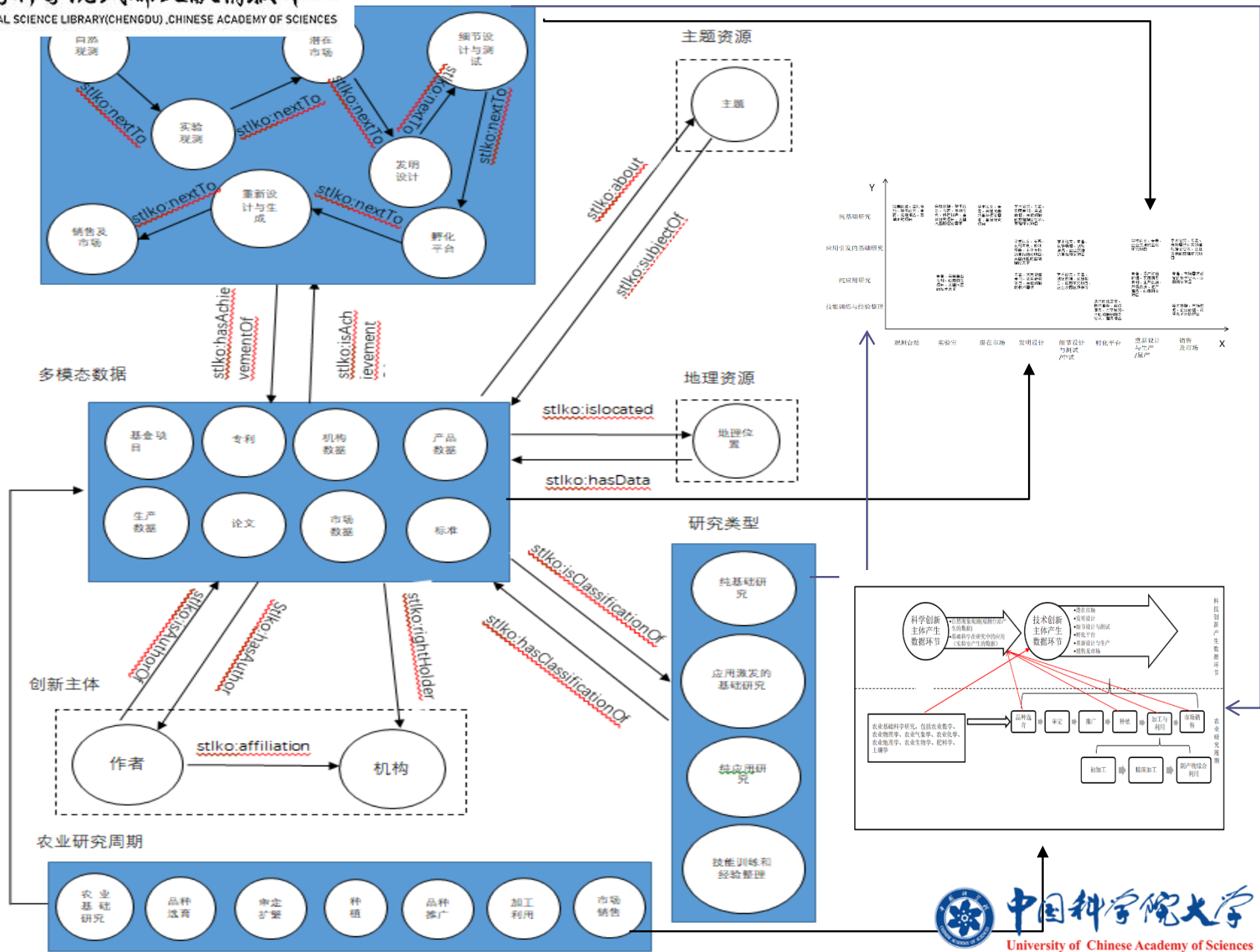


F4. The knowledge Fusion Method

3.1 Cognitive Sematic & Ontology Structure

T2. Cognitive Semantics Reflect the Concepts, Relations and Attribute

Cognitive System	Schema & Related Concept			Mapping to S&T Linkage Ontology
Non-language Cognitive System	Pattern-forming cognitive system	Narrative structure	Narrative Structure	<ul style="list-style-type: none"> S&T Innovation Agriculture Research Cycle
			Cognitive structure	Pasteur's Quadrant
			Factor	Ontology Concepts
			Factor relationship	Ontology Relations
Language Cognitive System	The Schematic system of Perspective State	Coherence	Coherent scene	S&T Innovation Scene
			Coherence timeliness	The Multimodal Data Produced in the Coherent Scene



F5. The Concepts of the Ontology



中国科学院成都文献情报中心
NATIONAL SCIENCE LIBRARY(CHENGDU), CHINESE ACADEMY OF SCIENCES



18th ISKO2024

ISKO International Conference 2024
Wuhan, China March 20-22, 2024

4 LLMs Evaluation & Selection



中国科学院大学
University of Chinese Academy of Sciences

4.1 Multimodal data selection

T3. Indicator System of Reliable Data

- **Hull-less barley**

is a unique high-altitude crop in China.

- Based on the FAIR principle build **Indicator System of Reliable Data**, and select 12 types data.

- Finally, select 10 multimodal data.

Primary indicator	Importance Degree	Secondary indicator	Importance Degree
Findable	4	White	1
		Gray	2
		Black	3
Accessible	3	Open	1
		Internal	2
		NO	3
Interoperable	2	Direct use	1
		Need intermediary	2
		NO	3
Reusable	1	Primary data	1
		Secondary data	2
		tertiary data	3

4.2 LLMs Service Evaluation

T4. Evaluation Index System for LLM Service Maturity in Ontology Construction

Primary indicators	data handling complexity (1 is the top)	Secondary indicator	Difficulty level (1 is the top)
Image extraction	1	Line break	1
		Cross-page	2
		Title, header, notes	3
Knowledge extraction	2	Classification	1
		RE	2
		NER	3
Supplementation	3	Characterization	1
		Breeding unit	2
		Ref.	3
English \rightleftharpoons Chinese	4	Corpus	1
		Inference	2
		Interoperable	3

4.2 LLMs Service Evaluation

LLMs: ERNIE Bot4.0, ChatGPT3.5, Claude3, Gemini

Target: Data cleaning, Classification, NER, RE

Materials : 50 Pieces paper abstract (English)

Prompt: in Chinese

T5. The LLMs Comparison in Classification, NER and RE

LLMs	Classification	NER	RE
ERNIE Bot4.0	100%	717	386
ChatGPT3.5	70%	685	506
Claude3	/	/	/
Gimini	72%	806	681

- Gemini has a higher number of extracts in NER and RE recognition.
- Google's rich corpus and the ability of English recognition is beneficial for identifying entities and entity relationships?



1. Gemini' answers looked like more logical, and natural
2. ERNIE Bot4.0, speak too much, looked like more reliable but not make sense.
3. Chat GPT3.5, match the answer one by one, like a real e.g.

Abstract: The freeze-thaw of early spring in China's Qinghai-Tibet Plateau is often accompanied by severe droughts. *Artemisia annua*, widely distributed in the region, releases allelopathic substances, mainly artemisinin, to the environment and exerts a wide range of effects on crops. This paper studied the physiological effects of barley under freeze-thaw, drought, and artemisinin stress through indoor simulation experiments. The physiological response characteristics of superoxide dismutase activity, catalase (POD) activity, net photosynthetic rate, relative water content (RWC), relative electrical conductivity, malondialdehyde (MDA) content, and soluble protein content in highland barley were analyzed. The results showed that artemisinin and drought contributed to the increase of SOD activity and the decrease of POD activity at the freeze-thaw stress, the SOD and POD activities both decreased firstly and then increased, but the effect of compound stress on POD was more complicated. Artemisinin, drought, or low temperature could reduce the net photosynthetic rate of highland barley. Low temperature had more significant impacts on photosynthesis than compound stress would show a single stress superimposed effect. Artemisinin, drought, and low temperature could reduce the RWC of highland barley, and increase relative electrical conductivity and the concentration of soluble protein (except for low temperature stress above zero, which reduces the concentration of soluble protein). However, the effect of compound stress on soluble protein is more complex. The single stress of artemisinin and drought had no obvious effect on MDA content, but MDA content was increased significantly under the freeze-thaw stress and the compound stress of artemisinin and drought, and the MDA content reached its peak. These results are helpful to explore the effects of freeze-thaw, drought and artemisinin stress on the growth of highland barley under the background of the aridification of the Qinghai-Tibet Plateau.

Gemini

◆ 科学技术创新管理专家:

根据您提供的摘要描述,可以将该过程分为以下几个阶段:

1. 基础研究:

- 自然观测: 记录客观自然现象
- 实验观测: 收集基础科学应用产生的实验数据

2. 应用研究:

- 潜在市场分析: 分析市场前景、存在问题、提出潜在研究需求
- 发明设计: 包括基础研究成果转化和市场需求引导两个方向促成的发明设计活动

3. 技术开发:

- 细节设计和测试: 类似中试阶段的数据,包括细节设计和测试

4. 成果转化:

- 孵化平台: 服务成果转化的平台

5. 产业化:

- 重新设计及生产: 产出最终阶段的数据

在这里输入你的问题



ChatGPT 3.5

● ChatGPT

(1) 根据提供的摘要描述,将研究分类如下:

- 自然观测: 描述了中国青藏高原早春的冻融情况以及高原大麦在冻融、干旱和青蒿素胁迫下的生理效应。
- 实验观测: 通过室内模拟实验研究了高原大麦在冻融、干旱和青蒿素胁迫下的生理响应特征。
- 潜在市场分析: 摘要中未提及相关内容。
- 发明设计: 摘要中未提及相关内容。
- 细节设计和测试: 室内模拟实验的设计和结果分析可视为细节设计和测试的一部分。
- 孵化平台: 研究结果可为合理的农业管理提供思路和方法,可将其视为孵化平台的一部分。
- 重新设计及生产: 摘要中未提及相关内容。
- 销售及市场: 摘要中未提及相关内容。

(2) 根据提供的摘要描述,将研究分类如下:

- 应用激发的基础研究: 研究的目的是探索高原大麦在冻融、干旱和青蒿素胁迫下的生理响应特征,这种基础研究源于对农业实践中出现的问题的深入理解,符合应用激发的基础研究的特点。
- 纯基础研究: 虽然研究目的是为了解决农业实践中的问题,但整个研究过程更侧重于对高原大麦的生理响应进行探索和理解,缺乏直接的应用目的,因此不符合纯应用研究的特征。
- 纯应用研究: 摘要中未提及相关内容。
- 技能训练与经验整理: 摘要中未提及相关内容。

Message ChatGPT...

ChatGPT can make mistakes. Consider checking important information.



中国科学院大学
University of Chinese Academy of Sciences

(1) 根据提供的摘要描述,该研究主要涉及通过文心大模型4.0 大麦在冻融、干旱和黄花蒿素胁迫下的生理效应。这包括对不同胁迫条件下超氧化物歧化酶(SOD)活性、过氧化物酶(POD)活性、净光合速率、相对含水量(RWC)、相对电导率、丙二醛(MDA)含量和可溶性蛋白含量等生理指标的分析。这些研究内容与“自然观测-实验观测-潜在市场分析-发明设计-细节设计和测试-孵化平台-重新设计及生产-销售及市场”过程中的“实验观测”阶段最为相符。因为该阶段主要涉及基础科学应用产生的实验数据,而本研究正是通过模拟实验来探究不同环境胁迫对高原大麦生理指标的影响,属于基础科学研究的范畴。

(2) 根据摘要描述的研究内容,可以将其归类为“应用激发的基础研究”,即巴斯德象限。这是因为该研究是通过室内模拟实验来探究高原大麦在冻融、干旱和黄花蒿素胁迫下的生理效应,旨在解决实际应用中可能遇到的问题。虽然研究起源于实际应用中的挑战,但最终目的是对基础知识的深入理解和发展,符合应用激发的基础研究的定义。同时,该研究并没有明确的应用目标或技术解决方案,因此不属于纯应用研究或技能训练与经验整理。背景信息方面,随着全球气候变化和青藏高原干旱化的加剧,冻融、干旱等胁迫对高原农作物的影响日益显著。因此,通过深入研究这些胁迫对农作物生理生化影响机制,可以为合理农业管理提供科学依据,具有重要的现实意义和应用价值。

(3) 该研究属于农业基础科学研究。依据在于,研究通过室内模拟实验探究了高原大麦在特定环境胁迫下的生理响应,这涉及作物生长的基础生物学过程。原因在于,了解作物如何响应环境胁迫是优化种植管理、提高产量的基础,而这一过程属于农业基础科学研究的范畴。

选择插件

使用“/”创建和收藏指令,可通过shift+回车换行

内容由AI生成,无法确保真实准确,仅供参考,请遵守《文心一言用户协议》、《文心一言个人信息保护规则》,基于文心大模型,版本:V2.5.4

F6. The Classification Command Execution Results with Chinese Prompt in ERNIE Bot4.0, ChatGPT3.5 and Gemini

4.2 LLMs Service Evaluation

LLMs: ERNIE Bot4.0, ChatGPT3.5, Claude3, Gemini Target:4 pieces of images

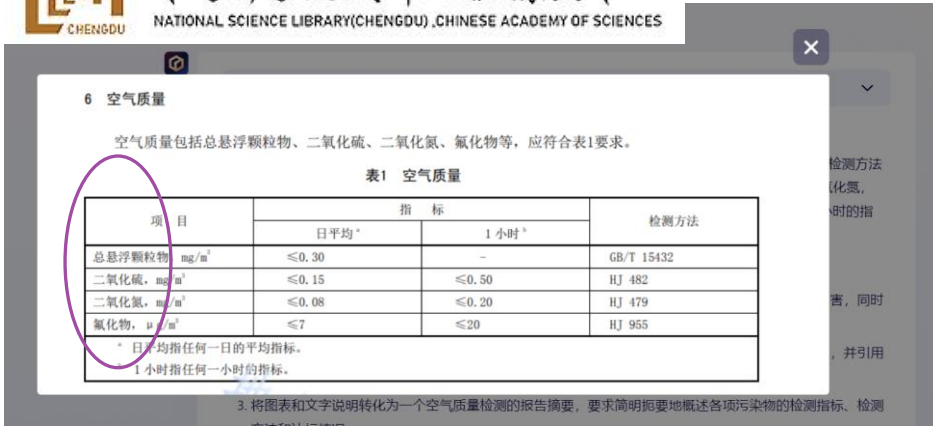
Target: The recognition accuracy of three types of content, text, cross page and line break recognition.

Materials : 4 Pieces of images(standard literatures)

Prompt: in Chinese

T6. The LLMs Comparison in Images

LLMs	Line break	Cross-page	Title, header, notes
ERNIE Bot4.0	75%	100%	100%
ChatGPT3.5	0	0	0
Claude3	/	/	/
Gimini	75%	100%	100%



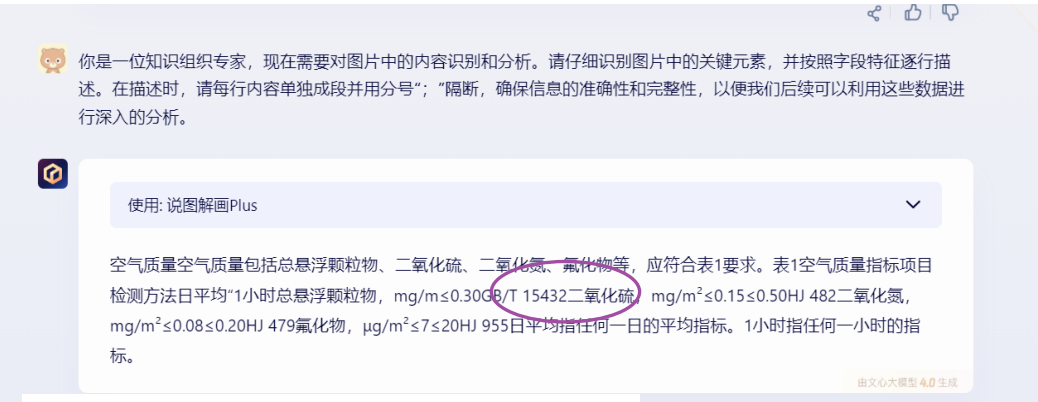
图片信息

- 图片类型: 表格
- 图片主题: 土壤肥力
- 图片来源: <https://baike.baidu.com/item/%E5%9C%9F%E5%A3%A4%E8%82%A5%E5%8A%9E>

关键元素识别

- 表格标题: 土壤肥力
- 表格列名: 项目、指标、检测方法
- 表格内容:

```
| 项目 | 指标 | 检测方法 |
| --- | --- | --- |
| pH值 | ≥6.5 | GB/T 6920 |
| 有机质,g/kg | >11.00 | NY/T 1121.6 |
| 全氮,g/kg | >0.70 | HJ 717 |
| 有效磷,mg/kg | >8.00 | NY/T 1121.7 |
| 速效钾,mg/kg | >70 | NY/T 889 |
```



- 表格标题: 空气质量

- 表格列:
 - 项目
 - 指标
 - 检测方法
 - 日平均
 - 1小时
- 表格内容:
 - 总悬浮颗粒物
 - 二氧化硫
 - 二氧化氮
 - 氟化物

逐行描述:

1. 项目; 指标; 检测方法; 日平均; 1小时
2. 总悬浮颗粒物,mg/m³; ≤0.30; GB/T 15432; -; -
3. 二氧化硫,mg/m³; ≤0.15; ≤0.50; HJ 482; -
4. 二氧化氮,mg/m³; ≤0.08; ≤0.20; HJ 479; -
5. 氟化物,μg/m³; ≤7; <20; HJ 955; -

逐行描述

表名: 水质;
项目: 指标; 检测方法;
pH: 5.5-8.5; GB/T 6920;
总氮: <0.003mg/L; HT 694;
总磷: <0.005mg/L; GB/T 7475;
...
粪大肠杆菌群: <10000个/L; SL 355;

数据分析

- 该水质标准对7项指标进行了规定,包括pH值、重金属、化学需氧量、石油类等。
- 其中,pH值、总汞、总磷、总铅、六价铬、氟化物、化学需氧量、石油类等指标都有明确的限值要求。
- 粪大肠杆菌群指标则要求不超过10000个/L。

结论

该水质标准对饮用水的水质提出了较为严格的要求,旨在确保饮用水的安全。

备注

- 以上分析仅基于图片中的有限信息进行,如有更详细的信息,可进行更深入的分析。

在这里输入你的问题

Gemini 显示的信息 (包括与人类对话内容) 不一定准确, 请注意核实。 但您始终与 Gemini 互动

4.2 LLMs Service Evaluation

LLMs: ERNIE Bot4.0, ChatGPT3.5, Claude3, Gemini

Target: Supplement related knowledge, e.g. Breeding Unit, characterization and reference.

Materials : 11Pieces records(Chinese)

Prompt: in Chinese

T7. The LLMs Comparison in Knowledge Supplementation

LLMs	Breeding unit	Characterization	Reference
ERNIE Bot4.0	11	11	7
ChatGPT3.5	0	1	0
Claude3	/	/	/
Gemini	9	11	6





中国科学院成都文献情报中心
NATIONAL SCIENCE LIBRARY(CHENGDU), CHINESE ACADEMY OF SCIENCES



18th ISKO2024

ISKO International Conference 2024
Wuhan, China March 20-22, 2024

4 Knowledge Fusion: Classification



中国科学院大学
University of Chinese Academy of Sciences

Enhancement of the release of phenolic compounds from white and black Qingke bran by autoclaving and fermentation treatments

od with health-promoting properties, but its by-product, Qingke bran, remains under-exploited. This c compounds (PCs) in Qingke bran and to improve the release of bound PCs by autoclaving and fermentation treatment, based on a comparative study between white and black Qingke bran. In this study, HPLC-ESI-QTOF-MS/MS was used to identify the diversity and composition of PCs in Qingke bran, and the antioxidant capacities of PCs were determined by ABTS, DPPH and FRAP assays. The results showed that both white and black Qingke bran contained a high content of PCs (647.82 and 931.42 mg GAE/100 g) and flavonoids (383.96 and 392.95 mu g RE/100 g), and a wide diversity of PCs consisting of flavonoid glycosides, flavonoid aglycones, phenolic glycosides, phenolic acid aglycones, phenolic acid esters and other PCs. Ferulic acid and vitexin were the major phenolic acids and flavonoids in Qingke bran, and most PCs predominated in the bound form. The autoclaving and fermentation treatment could change the diversity and composition of PCs, enhancing the release of free PCs, especially ferulic acid. The enhancement of the release of PCs contributed to the increase in the antioxidant capacities. In summary, fermentation was proved to be an effective treatment to enhance the release of PCs and to improve the anti-oxidant capacity of Qingke bran. This study provides important insights into the potential health benefits of Qingke bran and suggests a promising approach for the development of functional food products.

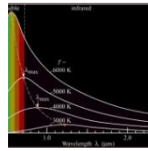
FOOD BIOSCIENCE



A

• Agriculture character

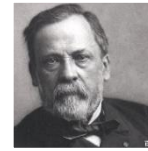
- O
- A1
- A2
- ...
- A6



S

• Scientific & technological character

- O1
- O2
- S1
- ...
- S6



R

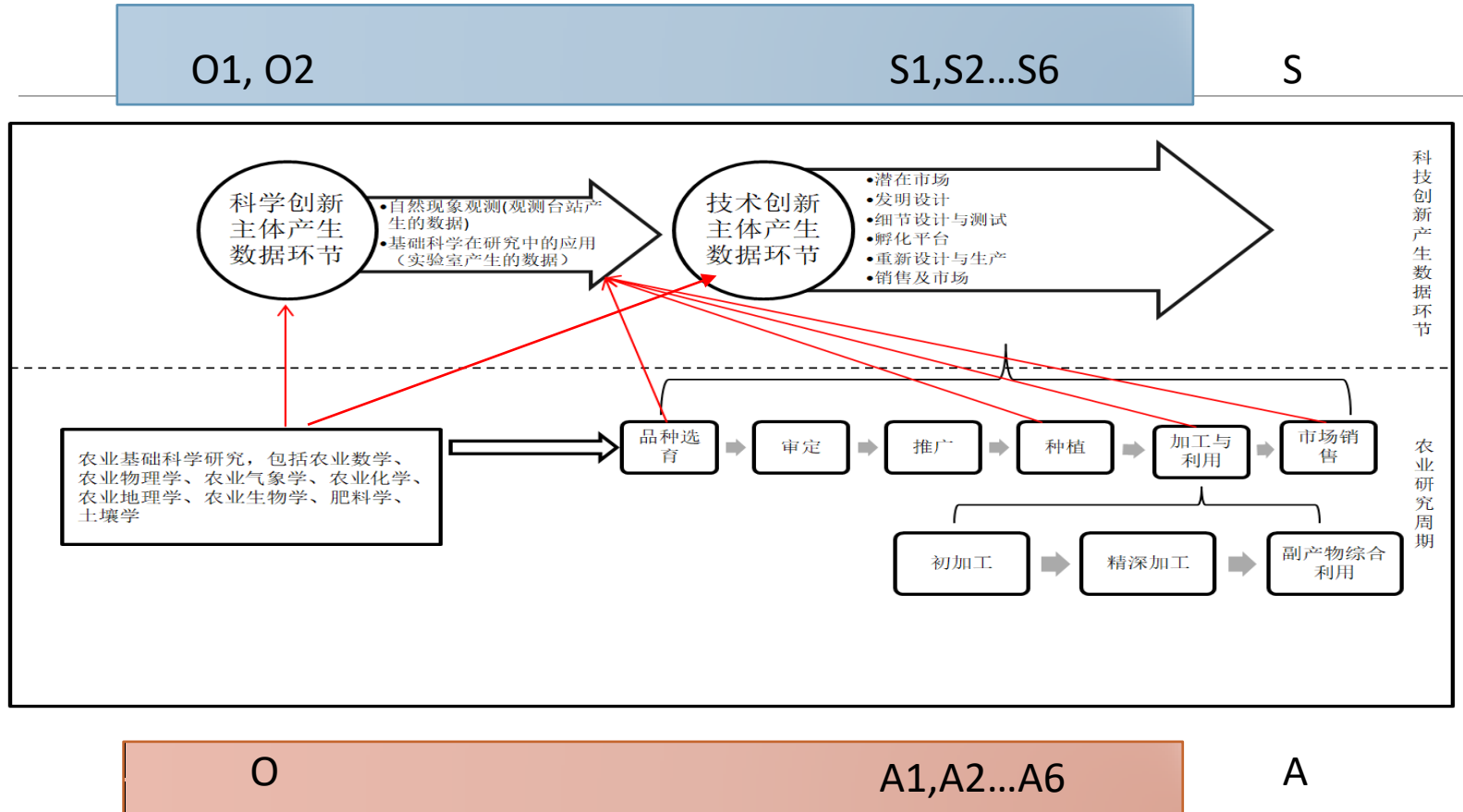
• Stokes science research model

- R1
- R2
- R3
- R4

F8. The 3 Characters in Hull-less Barley Journal Paper



Pasteur's Quadrant



F9. Theoretical Model of Identifying Scientific Problems

The Cross judgment

Cross matching and expert verification methods for further analysis.

1. The basic scientific research in the A is cross matched with the technological innovation link in S,
2. as well as the technological innovation link in the A and the scientific research part in the S.

The mathematical expression is as follows:

Suppose there are sets A and sets S

$$A = \{O, A1, A2 \dots A6\}$$

Where A=Agriculture Research Circle. O represents basic scientific research in agriculture, A1 –A6 present the circle.

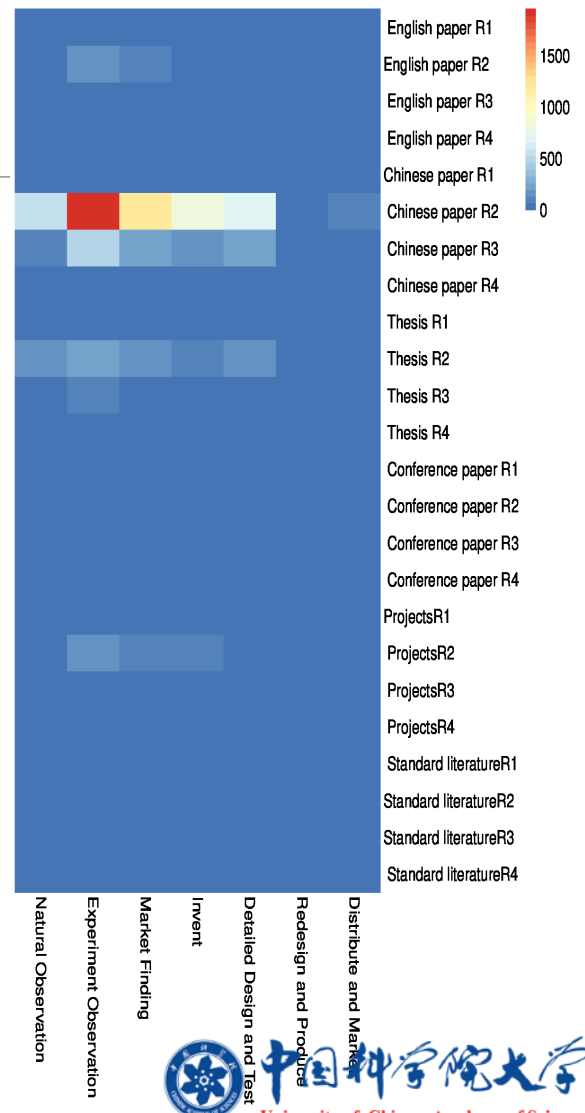
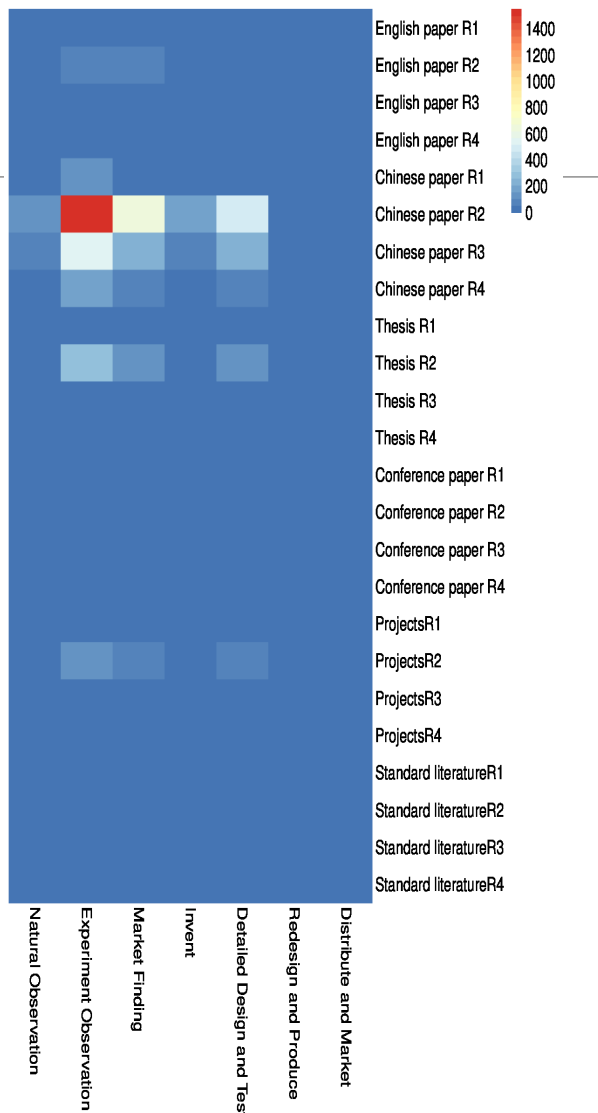
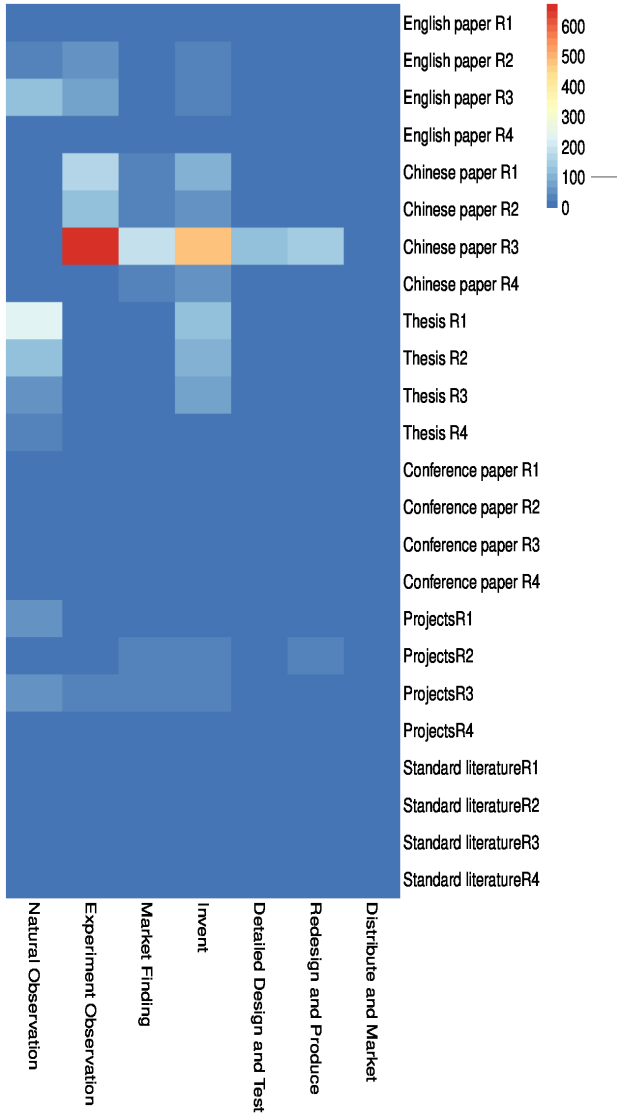
$$S = \{O1, O2, S1, S2 \dots S6\}$$

S represents the Scientific Technology Innovation Chain. O1 represents natural observation, O2 represents experimental observation, S1 to S6 represent innovation chain.

Ax = {A1, A2 ... A6}, and Ax ∈ A, Ox = {O1, O2}, and Ox ∈ S, make R2 represent Past eur' s Quadrant,

if Ax ∩ Ox ≠ ∅, and ∅ is empty set, then R2 ≠ ∅

Heatmap of Stokes Classification



T8. The Result of Cross judgment

Type	Pasteur's Quadrant (%)	Mismatching (%)	Other Quadrants (%)
English paper	76.8	0	23.2
Chinese paper	78.3	0.59	21.1
Conference paper	79.6	0	20.5
Thesis	37.5	5.93	56.6
Funding	79.5	0.54	20
Standard literature	0	0	100

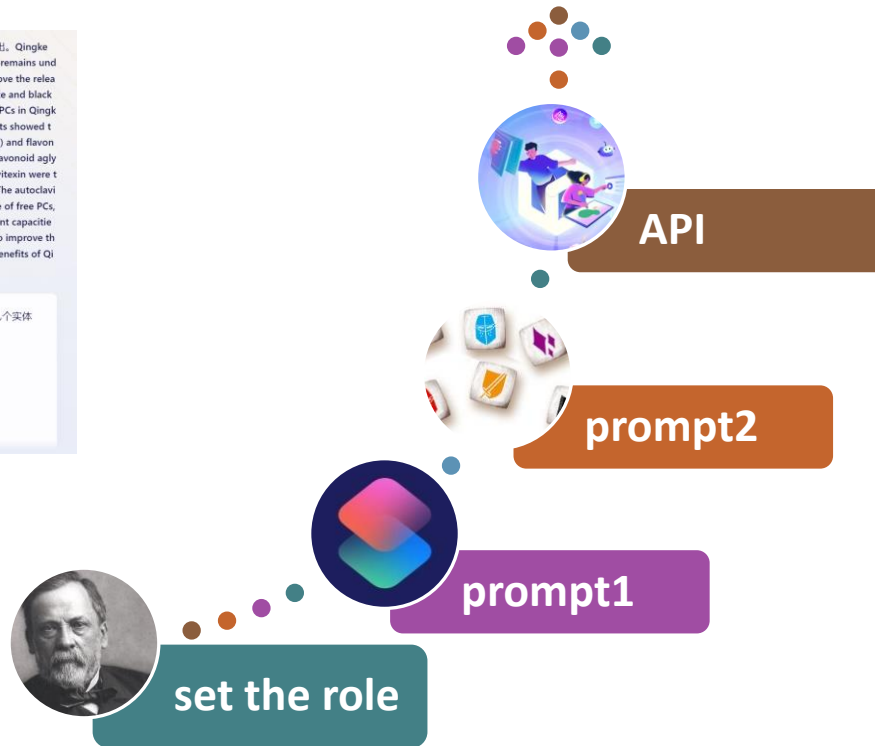
4.3 Knowledge Extraction By LLM

The extraction of entity and relations

你是一位自然语言处理专家，请根据下列数据抽取实体和实体关系，最后以：实体1-关系-实体2，的方式输出。Qingke (highland barley) is a functional food with health-promoting properties, but its by-product, Qingke bran, remains under-exploited. This study aimed to characterize the phenolic compounds (PCs) in Qingke bran and to improve the release of bound PCs by autoclaving and fermentation treatment, based on a comparative study between white and black Qingke bran. In this study, HPLC-ESI-QTOF-MS/MS was used to identify the diversity and composition of PCs in Qingke bran, and the antioxidant capacities of PCs were determined by ABTS, DPPH and FRAP assays. The results showed that both white and black Qingke bran contained a high content of PCs (647.82 and 931.42 mg GAE/100 g) and flavonoids (383.96 and 392.95 μg RE/100 g), and a wide diversity of PCs consisting of flavonoid glycosides, flavonoid aglycones, phenolic glycosides, phenolic acid aglycones, phenolic acid esters and other PCs. Ferulic acid and vitexin were the major phenolic acids and flavonoids in Qingke bran, and most PCs predominated in the bound form. The autoclaving and fermentation treatment could change the diversity and composition of PCs, enhancing the release of free PCs, especially ferulic acid. The enhancement of the release of PCs contributed to the increase in the antioxidant capacities. In summary, fermentation was proved to be an effective treatment to enhance the release of PCs and to improve the anti-oxidant capacity of Qingke bran. This study provides important insights into the potential health benefits of Qingke bran and suggests a promising approach for the development of functional food products.

根据提供内容可知，上文主要讲述了青稞麸皮、酚类化合物和抗氧化能力的相关信息，下面请提取以这几个实体为主体进行信息抽取，并以结构化的格式展示如下：

- 青稞麸皮 (Qingke bran) - 类型 - 功能食品副产品
- 青稞麸皮 (Qingke bran) - 利用状态 - 未充分利用
- 青稞麸皮 (Qingke bran) - 含有成分 - 酚类化合物 (PCs) - 黄酮类
- 酚类化合物 (PCs) - 测定方法 - HPLC-ESI-QTOF-MS/MS
- 酚类化合物 (PCs) - 抗氧化能力测定 - ABTS, DPPH, FRAP分析





中国科学院成都文献情报中心
NATIONAL SCIENCE LIBRARY(CHENGDU), CHINESE ACADEMY OF SCIENCES



18th ISKO2024

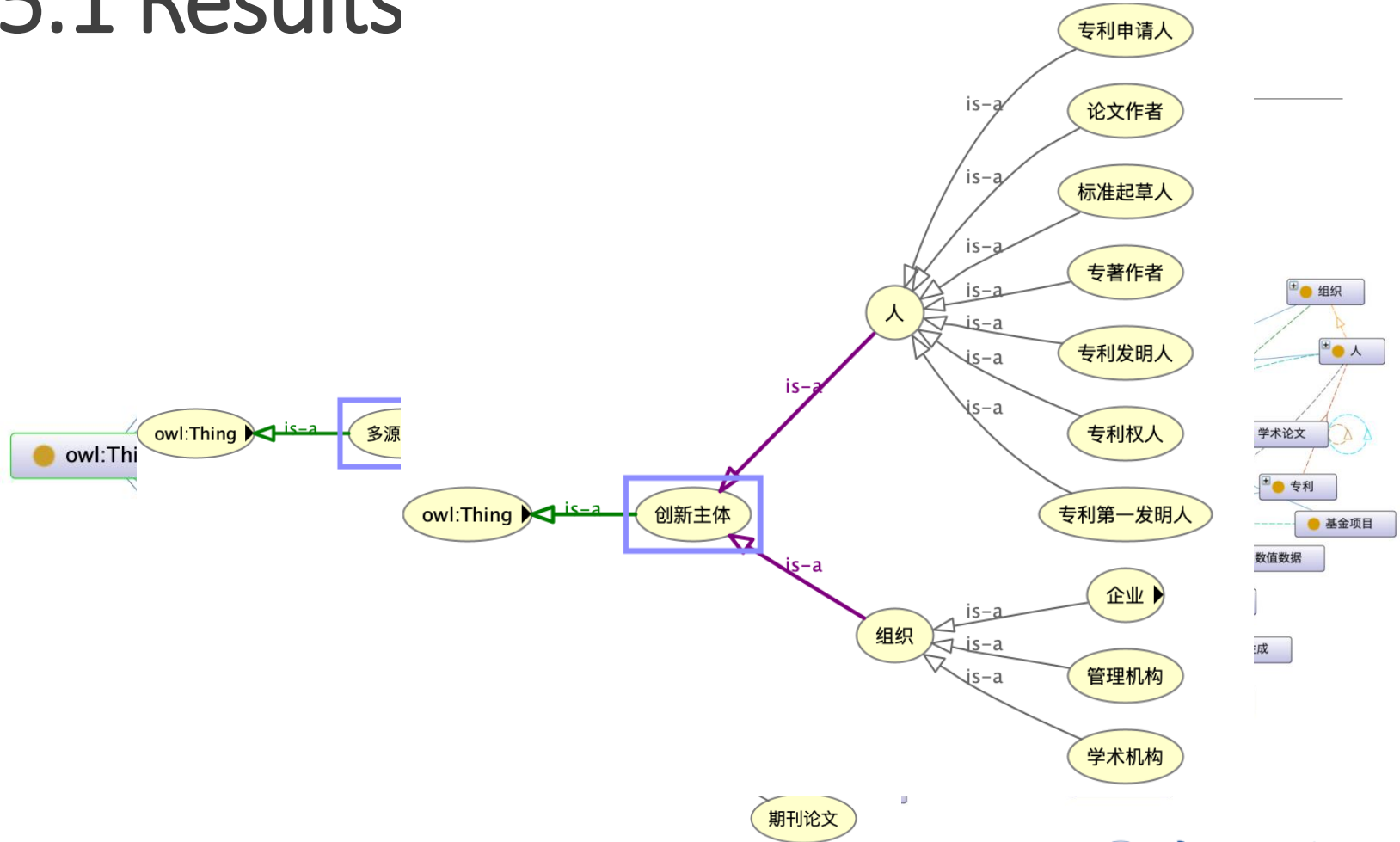
ISKO International Conference 2024
Wuhan, China March 20-22, 2024

5 Result & Discussion



中国科学院大学
University of Chinese Academy of Sciences

5.1 Results



F13. The Concepts Visualization by Protégé

5.1 Results

T9. Application Evaluation of ERNIE Bot4.0 in Building Agricultural Task Ontology

	Image Recognition	Knowledge Supplementation	Knowledge Extraction	Knowledge Classification	Chinese ⇌ English
ERNIE Bot4.0	★★★★☆☆	★★★☆☆☆	★★★★☆☆	★★★★★★	★★★★★★☆
Reason	Once for 1 piece to recognize , and some problems occurred in Line break	With little quantitative data, the supplement are right, but not make sense	In English, the entities and relationships extracted not as much as Gemini and Chat GPT.	Perfect, and meet our needs, the results can be verified	No issues have been identified yet

5.1 Results

- **Prompt:** in Chinese
- **Materials :** 50 Pieces paper abstract in English and 4 images with Chinese text

T10. The Results of Evaluation Index System for LLM Service in Ontology Construction

指标层	权重	指标层	权重	综合权重	指标层	EB4.0	Chat GPT3.5	Claude3	Gemini
图片识别	0.4566	换行	0.6333	0.2892	换行	4	1		4
	0.4566	跨页	0.2605	0.1189	跨页	5	1		5
	0.4566	普通文本	0.1096	0.0485	普通文本	5	1		5
信息补充	0.2212	机构补充	0.5571	0.1232	机构补充	4	1		3
	0.2212	特征补充	0.3202	0.0708	特征补充	4	1		2
	0.2212	参考文献	0.1226	0.0271	参考文献	3	1		2
知识挖掘	0.202	分类	0.5679	0.1147	分类	5	3		3
	0.202	RE	0.3339	0.0674	RE	4	4		5
	0.202	NER	0.0982	0.0198	NER	4	3		4
中英互操作	0.1202	语料	0.6127	0.0736	语料	4	2		3
	0.1202	知识推理	0.2693	0.0324	知识推理	4	3		4
	0.1202	语义理解	0.118	0.0142	语义理解	5	3		4
						EB4.0	Chat GPT	Claude3	Gemini
					分值	4.1981	2.2706	0	3.4588

5.2 Discussion

Advantages

1. Inter-disciplinary

In semantic analysis, concepts understanding, classification

2. Multilingual

Semantic analysis, interoperability.

3. Data Preprocessing & Knowledge Extraction

Quickly and efficiently

4. Image Extraction

In tables

5. Knowledge Inference

It has preference but could be explained

Disadvantages

1. Corpus

Will effect the ability of interfere

2. Supplementation

Not work now.

3. Multimodal data

It will help the user understand the thing form different aspects.

4. Charge

Charge will widen a cognitive gap between poor and rich.

5.2 Discussion

Suggestions

1.Enlarge the Corpus of Disciplinary

Improve the ability of semantic analysis.

2.Elarge the Ability of Multimodal Data Recognition

They will enlarge the natural language knowledge from the Perspective aspect.

3.Improve the Ability of NLP.

For ERNIE Bot4.0, improve the knowledge extraction of English.

And make the corpus more reliable.



- [1] 曹伟, 王得公, 李强, 符子均. 专利与技术创新的关系分析[J]. 情报学报. 2018, 37(06): 642-652.
- [2] 柴苗岭, 邹弈星, 谭荣志等. 面向农业产业知识服务的科学数据与科技文献关联研究与实践[J]. 农业图书情报学报. 2022, 34(03): 37-50.
- [3] Bhattacharya S, Kretschmer H, Meyer M. Characterizing intellectual spaces between science and technology[J]. *Scientometrics*, 2003, 58(2): 369-390.
- [4] Guan J C, He Y. Patent-bibliometric analysis on the Chinese science—technology linkages[J]. *Scientometrics*, 2007, 72(3): 403-425.
- [5] Verbeek A, Debackere K, Luwe M, et al. Linking science to technology: Using bibliographic references in patents to build linkage schemes[J]. *Scientometrics*, 2002, 54(3): 399-420.
- [6] 李睿. 科学-技术关联探测方法研究: 专利引文视角[M]. 成都: 西南交通大学出版社, 2013.
- [7] Gardner P. Representations of the relationship between science and technology in the curriculum[J]. *Studies in Science Education*, 1994, 24(1): 1-28.
- [8] 范内瓦, 布什, 拉什, D. 霍尔特. 科学: 无尽的前沿[M]. 北京: 中信出版集团, 2021.
- [9] Kline, Stephen J. Innovation is not a linear process[J]. *Research Management*. 1985, 28(2): 36 – 45.
- [10] D. E. 司托克斯. 基础科学与技术创新[M]. 北京: 科学出版社, 1999.
- [11] CARPENTER M, COOPER M, NARIN F. Linkage between basic research literature and patents[J]. *Research Management*, 1980, 23(2): 30-35.
- [12] Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy*, 29, 409–434.
- [13] Narin, F., Noma, E., 1985. Is technology becoming science?. *Scientometrics* 7 3–6 , 369–381.
- [14] Narin, F., Hamilton, K.S., Olivastro, D., 1995. Linkage between agency supported research and patented industrial technology. *Research Evaluation* 5 3 , 183–187.
- [15] Kevin W. Boyack, Richard Klavans. Measuring science–technology interaction using rare inventor–author names[J]. *Journal of Informetrics* (2008) 2: 173–182.
- [16] 李睿, 基于专利引文分析的科学-技术关联探测模型改进[D]. 北京: 中国科学院国家科学图书馆博士学位论文, 2011, 143
- [17] 张鹏, 雷家骥. 基于科学的产业发展模式研究: 以心电图和石墨烯产业为例[J]. *科学学与科学技术管理*, 2015, 36(9): 40-53.
- [18] Cui, Xuemiao; Pan, Chen; Lu, Jiping; Han, Yafeng. A Metadata Based Equipment Integrated Logistics Support Data Ontology Modeling Method.[C] 2022 3rd International Conference on Computer Vision, Image and Deep Learning and International Conference on Computer Engineering and Applications, CVIDL and ICCEA 2022, Pages 144-148, 202.
- [19] Casimir H B J. Industries and academic freedom[J]. *Research Policy*, 1971, 1: 3-8.
- [20] Verbeek A, Debackere K, Luwe M, et al. Linking science to technology: Using bibliographic references in patents to build linkage schemes[J]. *Scientometrics*, 2002, 54(3): 399-420.
- [21] Rip A, "Science and Technology as Dancing Partners", *Technological Development and Science in the Industrial Age*, 1992, 231-270.
- [22] Gardner P. Representations of the relationship between science and technology in the curriculum[J]. *Studies in Science Education*, 1994, 24(1): 1-28.
- [23] 范内瓦, 布什, 拉什, D. 霍尔特. 科学: 无尽的前沿[M]. 北京: 中信出版集团, 2021.
- [24] Kline, Stephen J. Innovation is not a linear process[J]. *Research Management*. 1985, 28(2): 36 – 45.
- [25] D. E. 司托克斯. 基础科学与技术创新[M]. 北京: 科学出版社, 1999.
- [26] <https://encyclopedia.thefreedictionary.com/Cognitive+Science> 20240317
- [27] Isac, Daniela; Charles Reiss (2013). *I-language: An Introduction to Linguistics as Cognitive Science*, 2nd edition. Oxford University Press. p. 5. ISBN 978-0199660179.
- [28] <https://guba.eastmoney.com/news,872225,940544701.html> [20240305]
- [29] Adapted from Miller, George A (2003). "The cognitive revolution: a historical perspective". *Trends in Cognitive Sciences* 7.
- [30] 伦纳德·泰尔米(Leonard Talmy). 认知语文学. 北京: 北京大学出版社, 2023.2
- [31] Lan Du, Wray Buntine, Huidong Jin. Sequential Latent Dirichlet Allocation: Discover Underlying Topic Structures within a Document [C], 2010 IEEE International Conference on Data Mining, p. 148-157.
- [32] 陈稳, 陈伟. 科学与技术对比视角下的前沿主题识别与演化分析[J]. *情报杂志*. 2022, 41(01): 67-73+163.
- [33] Yoon B, Park Y. A text-mining-based patent network: analytical tool for hightechnology trend[J]. *The Journal of High Technology Management Research*, 2004, 15(1): 37-50.
- [34] Gangbo Wang, Jiancheng Guan. Measuring science–technology interactions using patent citations and author–inventor links: an exploration analysis from Chinese nanotechnology[J] *Journal of Nanoparticle Research*, 2011(13): 6245–6262.
- [35] Noyons, E. C. M., Buter, R. K., van Raan, A. F. J., Schmoch, U., Heinze, T., Hinze, S., et al. (2003). Mapping excellence in science and technology across Europe: Nanoscience and nanotechnology: Centre for Science and Technology Studies (CWTS). The Netherlands: Leiden University
- [36] Meyer, M. Are patenting scientists the better scholars? An exploratory comparison of inventor–authors with their non-inventing peers in nanoscience and technology. *Research Policy*. 2006, volume 35: 1646–1662.
- [37] Winnink, J. J., Tijssen, Robert J, W. Dynamics and scientific breakthroughs in HIV/AIDS drugs development: the case of integrase inhibitors[J]. *Scientometrics* 2014, 101: 1–16.
- [38] 赵辉, 冷伏海. 从学术型发明人视角看高科技领域科学与技术研究前沿主题的关系: 以石墨烯领域为例[J]. *中国发明与专利*, 2021, 18(01): 3-12.
- [39] H.Y. Xu, J. Winnink, Z. Yue, Z. Liu, G. Yuan. Topic-linked innovation paths in science and technology [J]. *Journal of Informetrics*, 2020, 14 (2), Article 101014
- [40] Jaffe A B . The Real Effects of Academic Research[J]. *American Economic Review*, 1989, 79(5): 957-970.
- [41] 许海云, 武华维, 罗瑞. 基于多元关系融合的科技文本主题识别方法研究[J]. *中国图书馆学报*. 2019, 45(01): 82-94.
- [42] 许海云, 董坤, 隗玲, 等. 科学计量中多源数据融合方法研究述评[J]. *情报学报*, 2018, 37(3): 318-328.
- [43] Martin, C., Cadiou, C., & Jannès-Ober, E. (2017). Data Management: New Tools, New Organization, and New Skills in a French Research Institute



AI 作图



THANKS!

Chaiml@clas.ac.cn