Visualization and Analysis of Multidisciplinary Research Networks

A Case Study of UNIST's College of Information & Biotechnology

Sumin Park
20211113

2024.06.07

IE45001 Project Lab
2024 Spring

Motivation

increase in interdisciplinary studies -> topics are not confined to the department -> difficult to find **all labs** in fields of interest

UNIST freshman student interested in
HCI and Human factor

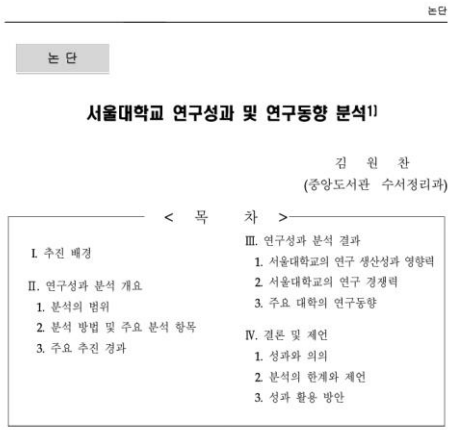
BME	BCI lab DNCE lab Color lab …
DES	Emotion lab DECS lab IPD lab …
CSE	iVHCI lab TACT lab …

Goal

- Graphical representation and analysis of each lab's research interest and overall research trend at the UNIST College of Information & Biotechnology (IB).
- Presentation of analysis methods in the field of study

Related works

1. How to analyze research trends



논 단

서울대학교 연구성과 및 연구동향 분석[1]

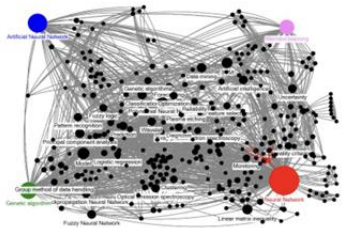
김 원 찬
(중앙도서관 수서정리과)

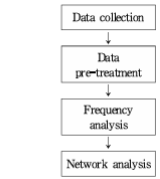
< 목 차 >

2. Using graph method

Analysis of major research trends in artificial intelligence through analysis of thesis data

Myoung-Sug Chung, Seong-Hyeon Park, Byeong-Hoon Chae, Joo-Yeoun Lee
New industry convergence technology R&D center, Ajou university





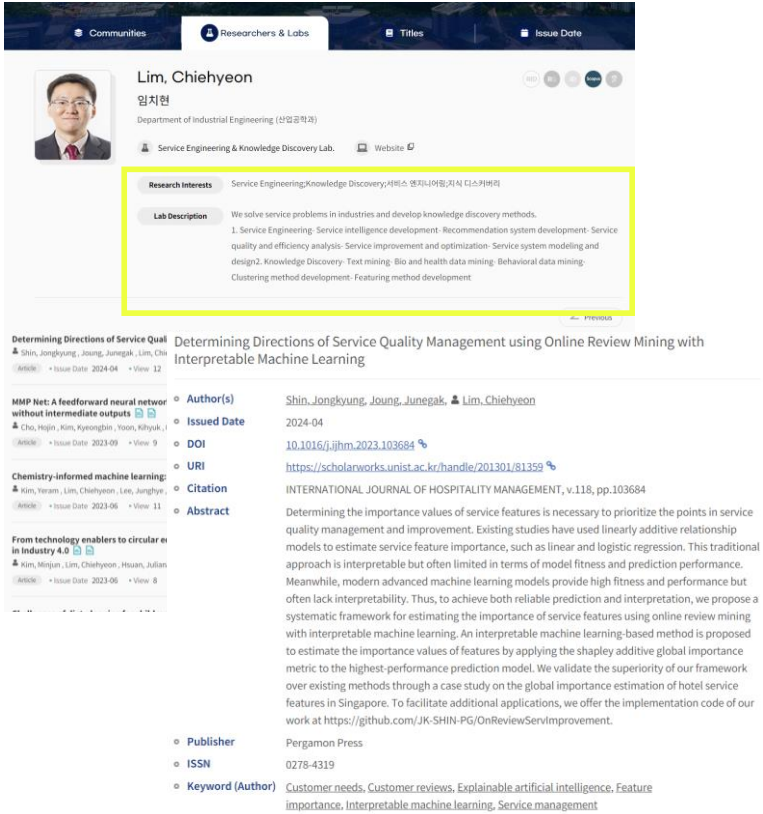
Data analysis process

[Fig. 4] An Analysis of the Keyword Network for 20 Years of Global Catalog

3. Research Trends Analysis Site

https://alinlab.kaist.ac.kr/KCSS

Data



92 laboratory (8 departments) **->Node**

1. Lab URL
2. Professor name
3. Lab name
4. Field [IE, BME, EE,,]
5. research interest
6. Lab description
7. Title and Keywords of Recent 20 Articles (1388 papers)

->Feature

[Web Crawling]

https://scholarworks.unist.ac.kr/researcher/f6235f85-e746-4387-9fbd-f72e507dacd7

Process

Preprocessing	Vectorizer	Similarity	Representation	Clustering
	TF-IDF **BERT Model**		2d / 3d	**K-means** DBSCAN

Noise
(special characters,
mixed case usage)

+ searching part

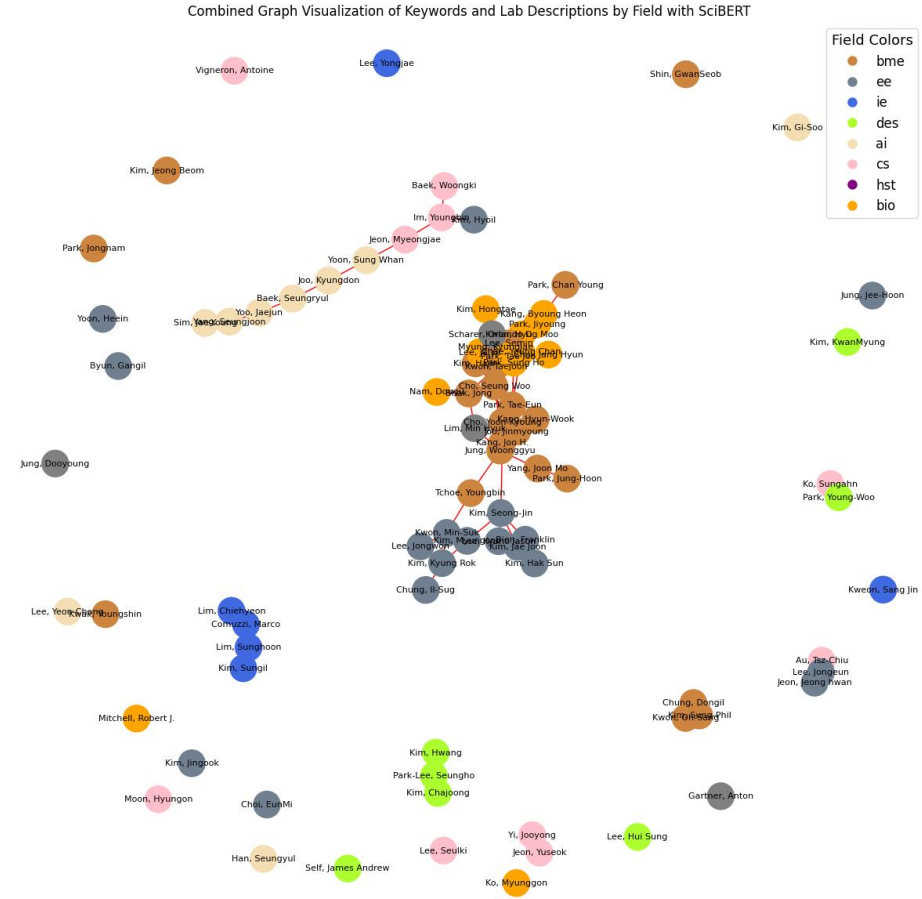
Process

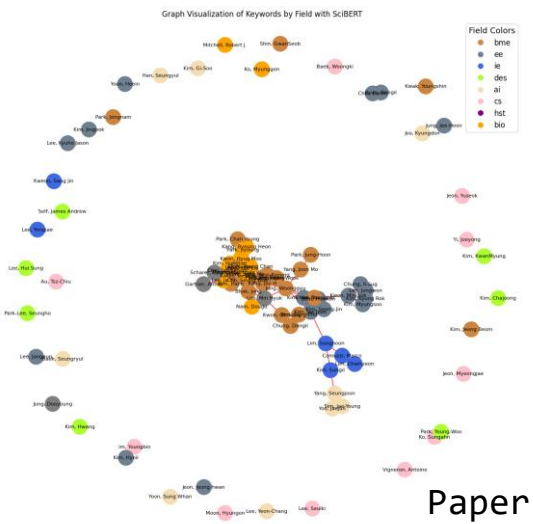
Model Selection

	TF-IDF	BERT model
Language model	Term Frequency - Inverse Document Frequency	BERT : textual embedding (sci_BERT) Strengths in **understanding the context** of word use

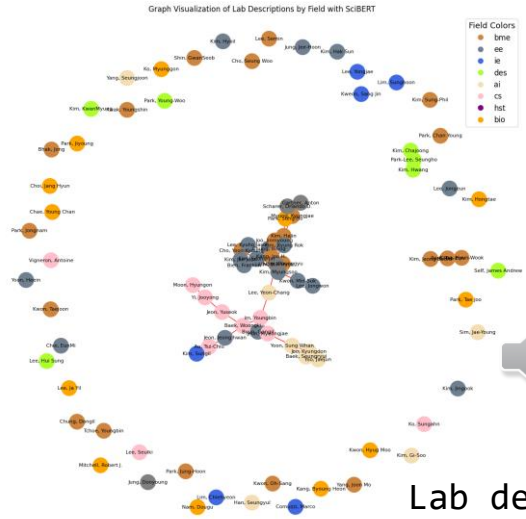
	K-means	DBSCAN
Clustering model	When data is well separated into spherical cluster. Based on **distance between cluster centers** and data point	when shape of cluster is arbitrary May not work when **difference between clusters are not large**

Result





Paper keywords

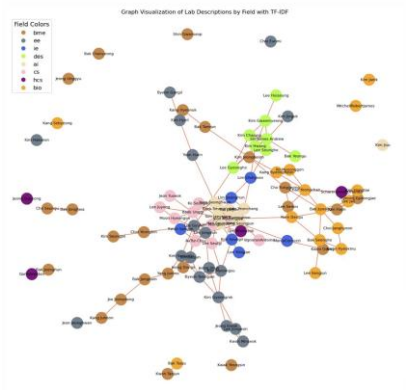


Lab description

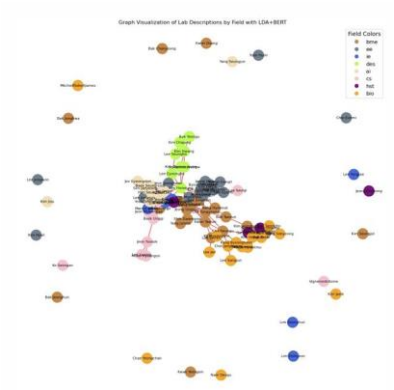
Results

Sci_BERT

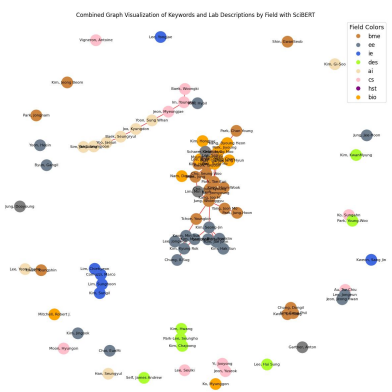
Bidirectional Encoder Representations from Transformers



TF-IDF
Silhouette Score:
0.0148



BERT
Silhouette Score:
0.1094



SciBERT
Silhouette Score:
0.0597

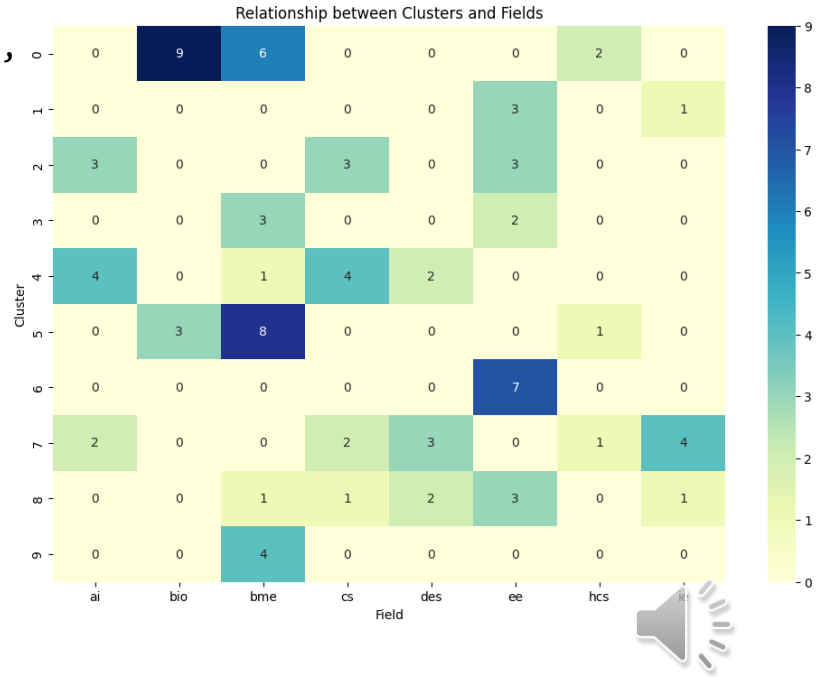
SciBERT (Scientific BERT):
Pretrained model (data of Biomedical & Computer science paper)

Result

```
search_keyword = 'hci'
>>
Most similar cluster for 'hci':
Cluster 8 with similarity 0.7826
```

UNIST freshman student interested in
HCI and Human factor

```
Cluster 8 keywords : 'interaction,
mobility, future, electronic,
humanrobot, driving, theory,
robots, robotics, self',
```



Researcher Name	Keywords	Field
Park, Young-Woo	MENTALHEALTHCARE,INTERVENTIONS,METHODOLOGY,DEP...	des
Kim, KwanMyung	MOVEMENT,SURFACE ELECTROMYOGRAPHY, CHIN TUCK,EL...	des
Kim, Myungsoo	MEMRISTOR,2D material,hexagonal boron nitride,...	ee
Moon, Hyungon	, , , ,domainspecific architecture,instruc...	cs
Byun, Gangil	FRONT,DESIGN,PLANAR,BROADBAND,WAVE PLATE PHASE	ee

Conclusion

Represent the researchers' keywords and lab descriptions to identify research trends

Method: SciBERT embeddings and K-means clustering

Key Findings

SciBERT Embeddings:Specialized in scientific texts
Accurately reflect lab characteristics

K-means Clustering:

Forms clear and interpretable clusters
More effective than DBSCAN in forming meaningful clusters

Conclusion

1. **Help to Choose Department & Laboratory for UNIST Students**

2. **Suggest Analysis Method for Multidisciplinary Research**:

3. **Provide Motif for Multidisciplinary Research**:

4. **Reference in Department Classification Process**:

Discussion

limitation

1. Data limitation
2. Limitation of clustering algorithms
3. Limitation of embedding model
4. Subjectivity in cluster interpretation

Future work

1. Comparison with other university or Global research trend
2. Dynamic Graph Representation with this data (change of research area and lab – detect trend)
3. Experiment with Additional Embedding Models:
4. Automatic Cluster Labeling: