ESRA: Explainable Scientific Research Assistant

Pollawat Hongwimol[1*], Peeranuth Kehasukcharoen[1*], Pasit Laohawarutchai[1*],
Piyawat Lertvittayakumjorn[2], Aik Beng Ng[3], Zhangsheng Lai[3], Timothy Liu[3], Peerapon Vateekul[1†]

[1] Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Thailand
[2] Department of Computing, Imperial College London, United Kingdom
[3] NVIDIA

✉ peerapon.v@chula.ac.th

[*] Equal contributions, [†] Corresponding author

Abstract

- We introduce Explainable Scientific Research Assistant (ESRA), a literature discovery platform that augments search results with relevant details and explanations.
- ESRA provides three main features: explanation (for why a paper is returned to the user), list of facts (that are relevant to the query), and graph visualization (drawing connections between the query and each paper with surrounding related entities).
- The experimental results with humans involved show that ESRA can accelerate the users' search process with paper explanations and helps them better explore the landscape of the topics of interest by exploiting the underlying knowledge graph.
- We provide the ESRA web application at http://esra.cp.eng.chula.ac.th/

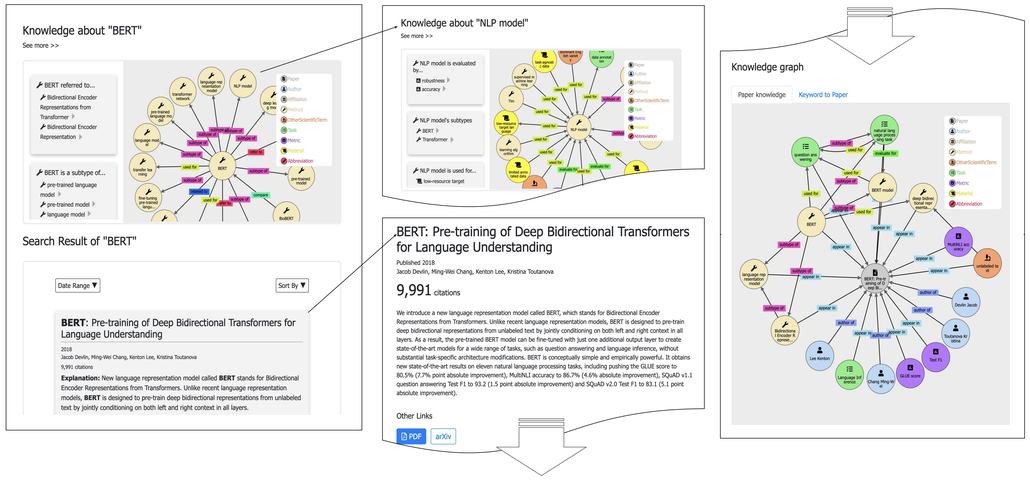
Introduction

Existing literature search platforms:

- Mostly present only metadata of papers
- Requires users to read the entire abstracts to understand the brief contents
- Cannot suggest next search term because users need to reflect on the knowledge of the papers themselves

It is time consuming to gradually expand their understanding of the field using existing platforms

Web Application



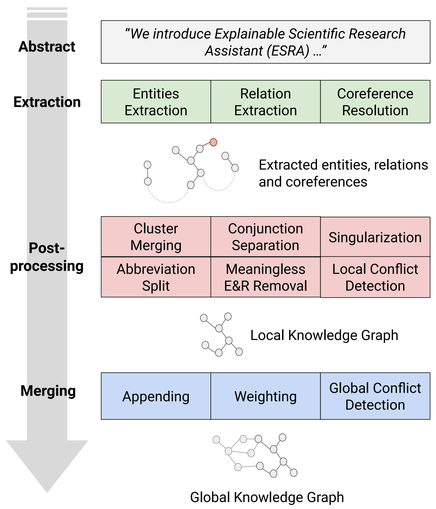
ESRA has three main features:

- "Explanation" explains how the query and each returned paper are related.
- "Fact list" suggests top-related keywords with their relationships to the query supporting exploration of related scientific concepts.
- "Graph visualization" provides a subgraph illustrating related knowledge around the query and the returned papers.

These features aim to help researchers quickly discover and understand a collection of literature they are looking for.

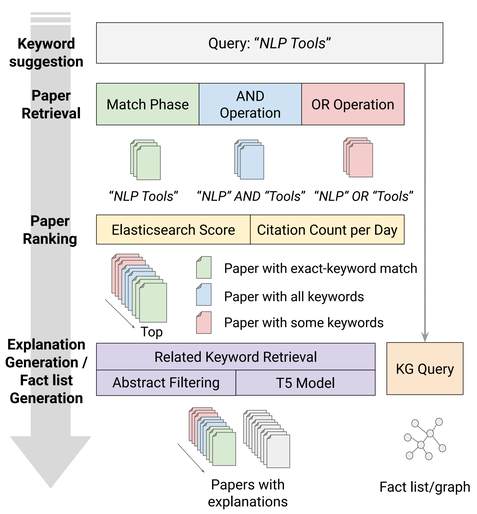
Pipelines

There are two important pipelines in ESRA which are uses for extracting a knowledge graph and searching.



Knowledge Graph Construction Pipeline

This pipeline is for extracting knowledge from the abstracts of papers to create a scientific knowledge graph.



Paper Searching Pipeline

This pipeline is for retrieving ranked papers, explanations, fact list, and graphs to show on the web application.

Results and Evaluation

Knowledge graph construction

- We compare four models including SciIE [1], DyGIE++ [2], SpERT [3], and our model.
- We later evaluate the models on three different tasks, including NER, RE, and CR.
- Our model is the combination of two models – SciIE for the CR task and SpERT for the best performance on NER and RE tasks.

Conclusion: Our model can retain the performance of SpERT on NER and RE tasks, while it slightly sacrifices the performance of SciIE on CR task.

Model	F1 Score (evaluated on SciERC)		
	Named-Entity Recognition (NER)	Relation Extraction (RE)	Coreference Resolution (CR)
SciIE	64.20	39.30	48.20
DyGIE++	67.50	48.40	-
SpERT	70.33	50.84	-
SpERT+SciIE (Ours)	70.33	50.84	45.87

Human Evaluation

- We recruited 32 participants from Computer Science and Engineering areas.
- Three dimensions are focused—understandability, usefulness, and visual appeal.
- The score ranges from 1 to 5—strongly disappointed, disappointed, neutral, satisfied, and strongly satisfied, respectively.

Conclusion: The average score from all participants on each dimension falls within the range between 3.6 and 4.2, meaning that our system could reasonably satisfy users with some room for further improvement.

Features	Average satisfaction score (from 1 to 5)		
	Explanation	Fact list	Graph visualization
Understandability	4.25	3.69	3.94
Usefulness	3.94	3.91	3.88
Visual appeal	3.81	3.81	3.81

Feature Comparison

Table below compares prominent features of existing graph-based literature platforms to our ESRA system.

Feature	Platform				
	Microsoft Academic	Semantic Scholar	ORKG	AceMap	ESRA (ours)
Scientific KG	✘	✔	✘	✘	✔
Metadata Graph	✔	✔	✔	✔	✔
Explanation	✘	✘	✘	✘	✔
Fact List	✘	✘	✘	✘	✔
Graph Visualization	✘	✘	✔	✔	✔

Conclusion

Our literature discovery platform, ESRA, uses a scientific knowledge graph to enhance user's experience. Based on the human evaluation, ESRA can help users screen through papers faster using the generated explanations and capture important facts about the query and the papers using the fact list and the graph visualization. In the future, we aim to expand the coverage of our knowledge graph by extracting facts from the full documents to enhance the quality of ESRA results.

References

[1] Yi Luan et al., "Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction". EMNLP 2018

[2] David Wadden et al., 2019. "Entity, relation, and event extraction with contextualized span representations". EMNLP-IJCNLP 2019

[3] Markus Eberts and Adrian Ulges. 2019. "Span-based joint entity and relation extraction with transformer pre-training". ECAI 2020