

## K3S: Knowledge-Driven Solution Support System

Yu Zhang, Morteza Saberi, Min Wang, Elizabeth Chang

University of New South Wales, Canberra

Northcott Dr, Campbell, ACT 2612, Australia

{yu.zhang, min.wang}@student.adfa.edu.au, m.saberi@unsw.edu.au, e.chang@adfa.edu.au

### Abstract

As the volume of scientific papers grows rapidly in size, knowledge management for scientific publications is greatly needed. Information extraction and knowledge fusion techniques have been proposed to obtain information from scholarly publications and build knowledge repositories. However, retrieving the knowledge of problem/solution from academic papers to support users on solving specific research problems is rarely seen in the state of the art. Therefore, to remedy this gap, a knowledge-driven solution support system (K3S) is proposed in this paper to extract the information of research problems and proposed solutions from academic papers, and integrate them into knowledge maps. With the bibliometric information of the papers, K3S is capable of providing recommended solutions for any extracted problems. The subject of intrusion detection is chosen for demonstration in which required information is extracted with high accuracy, a knowledge map is constructed properly, and solutions to address intrusion problems are recommended.

### Introduction

With the dramatic expansion of the volume of academic publications, there is an urgent need for novel knowledge management techniques that can retrieve, store and manage the massive amount of scientific information effectively and efficiently. Current knowledge management applications have been used to collect desired information from various sources, such as academic articles, news papers, web pages, advertisements and email correspondence. Once obtained, the data can be added to existing datasets or integrated to build knowledge repositories. The constructed knowledge repositories are then used as the brain of various types of applications for sharing domain knowledge or providing technical support (Mayfield et al. 2014).

Although developing applications to manipulate knowledge has been lasting for decades in both academia and industries, automatically extracting knowledge from scientific papers and reconstructing knowledge maps to provide intelligent support are still at the beginning stage. Knowledge maps are two-dimensional spatial arrays that represent information in the form of a node-link-node diagram (Wiegmann et al. 1992). Intelligent systems assisting to build so-

lution support knowledge maps demand three main components. Firstly, Natural Language Processing (NLP) techniques are required for computers to understand texts and extract requested information (Poon et al. 2010). Secondly, predefined knowledge fusion rules are needed to support information integration, probabilistic reasoning and semantic analysis (Sánchez, Moreno, and Del Vasto-Terrientes 2012). Lastly, a good understanding of knowledge maps is of great help for sorting the collected knowledge accordingly in the maps (Dong et al. 2014). Many works have been done to explore the above topics individually, but little research has combined them to process scholarly publications for providing scientific solutions. Therefore, to remedy this gap, a knowledge-driven solution support system (K3S) is proposed in this paper.

The superiority of K3S is that it is able to not only extract research problems and solutions proposed in academic papers, but also rank the solutions based on the bibliometric information of the papers. To achieve these advantages, K3S introduces a new type of customized information network, named ‘solution-oriented information network (SIN)’. As the knowledge base of K3S, the *SIN* contains nodes representing research problems (*RPs*) and proposed techniques (*PTs*), and one relation ‘tackled’ for connecting *RPs* and *PTs*. The *RPs* and *PTs* are obtained from solution-oriented academic papers in which authors propose techniques or solutions to address specific research issues. Additionally, K3S includes an academic paper ranking module to sort the knowledge contained in the *SIN*. In this manner, driven by the developed knowledge base and ranking module, K3S is able to provide recommended solutions to each extracted research problem in specific subjects.

### K3S Conceptual Framework

K3S relies on four main modules, namely, *Document Filter Module*, *Knowledge Fusion Module*, *Uncertainty Investigation Module* and *Knowledge Ranking Module*. Fig. 1 shows the conceptual framework of the K3S and illustrates how the four modules are connected.

In Fig. 1, keywords of a desired subject are fed into the system as input, and the *Document Filter Module* firstly deploys Scopus as a searching tool to access academic papers based on the keywords. The module assists in collecting solution-oriented academic papers of the desired subject

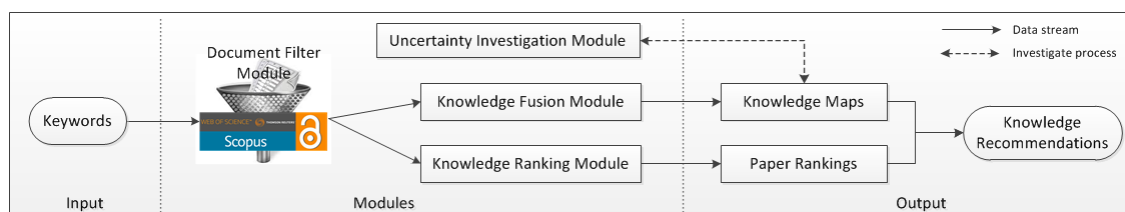


Figure 1: Conceptual Framework of K3S

by leveraging three stages, namely, keyword searching, topic filtering and quality screening. Secondly, the *Knowledge Fusion Module* receives the collected papers and extracts *RPs* and *PTs* from them by adopting a designed NLP program. Thirdly, several knowledge fusion rules are defined based on academic writing convention to connect the extracted information. A knowledge map is now constructed after integrating the extracted information in the Knowledge Fusion Module. Furthermore, the constructed map is inspected by the *Uncertainty Investigation Module* using Crowdsourcing techniques to double check the information extracted and connected by the previous modules. Meanwhile, the *Knowledge Ranking Module* ranks the papers in collection based on their bibliometric information extracted by the Document Filter Module. With the constructed knowledge map and the paper rankings, K3S is able to provide valuable suggestions and solutions to users.

Experiments have been conducted in which ‘intrusion detection in Cyber Security’ was chosen as the subject-specific field. 1358 topic-related papers were collected and processed in our proposed system, and the system achieved almost 98.75% completion in terms of processing data for building the information network. Detailed information can be found in our forthcoming publication (Zhang, Saberi, and Chang 2018). For demonstration, only 140 papers are randomly collected due to the privacy issue of using Scopus data.

### Demonstration Tools and Dataset

The interface of K3S is designed to build a user friendly faceplate for the purpose of demonstrating and providing subject-specific suggestions to users. The interface displays three main panes which are *Knowledge Map Pane*, *Description and Comment Pane* and *Solution Recommendation Pane*. The panes and functions are demonstrated based on a WEB page which is developed mainly by using HTML, CSS and JavaScript languages for the front end, and back end by Python, JSON and Django database.

Since NLP program only takes plain texts as input, the downloaded PDF files (papers) are converted to texts beforehand and saved in a dataset. The plain texts are parsed into a JSON array by a Python script, then the system takes data from the JSON document and loads it into a Django database which is rendered into the faceplate as a web page using HTML, CSS and JavaScript. In addition, the bibliometric information of the papers is extracted using Python and Scopus API, and the information is stored separately in CSV files for processing.

### Conclusion and Future Work

A knowledge-driven solution support system (K3S) is proposed in this paper. K3S is able to extract problem/solution information from academic publications, integrate it to solution-oriented information networks, and provide recommended solutions to users. The subject of intrusion detection is chosen for demonstration in this work, and the system performs well in terms of implementing the designed functions. In future extension of this work, our proposed solution support system, the K3S, will be tested in diverse research fields for a more comprehensive evaluation.

### References

- Dong, X.; Gabrilovich, E.; Heitz, G.; Horn, W.; Lao, N.; Murphy, K.; Strohmman, T.; Sun, S.; and Zhang, W. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 601–610. ACM.
- Mayfield, J.; McNamee, P.; Harmon, C.; Finin, T.; and Lawrie, D. 2014. Kelvin: Extracting knowledge from large text collections. In *AAAI Fall Symposium on Natural Language Access to Big Data*. AAAI Press, November.
- Poon, H.; Christensen, J.; Domingos, P.; Etzioni, O.; Hoffmann, R.; Kiddon, C.; Lin, T.; Ling, X.; Ritter, A.; Schoenmackers, S.; et al. 2010. Machine reading at the university of washington. In *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, 87–95. Association for Computational Linguistics.
- Sánchez, D.; Moreno, A.; and Del Vasto-Terrientes, L. 2012. Learning relation axioms from text: An automatic web-based approach. *Expert Systems with Applications* 39(5):5792–5805.
- Wiegmann, D. A.; Dansereau, D. F.; McCagg, E. C.; Rewey, K. L.; and Pitre, U. 1992. Effects of knowledge map characteristics on information processing. *Contemporary educational psychology* 17(2):136–155.
- Zhang, Y.; Saberi, M.; and Chang, E. 2018. A semantic-based knowledge fusion model for solution-oriented information network development: a case study in intrusion detection field. *Scientometrics*. Forthcoming. DOI: 10.1007/s11192-018-2904-6.