

# AutoSurvey: Automatic Survey Generation based on a Research Draft

Hen-Hsen Huang

Department of Computer Science, National Chengchi University, Taipei City, Taiwan  
hhuang@nccu.edu.tw

## Abstract

This work presents AutoSurvey, an intelligent system that performs literature survey and generates a summary specific to a research draft. A neural model for information structure analysis is employed for extracting fine-grained information from the abstracts of previous work, and a novel evolutionary multi-source summarization model is proposed for generating the summary of related work. This system is extremely used for both academic and educational purposes.

## 1 Introduction

Summarization of scientific articles is an emerging topic that attracts attention in recent years [Jaidka *et al.*, 2016; Yasunaga *et al.*, 2019; Boni *et al.*, 2020]. Previous work treat the task as multi-document summarization in which the summarization is based on a list of publications, remaining two shortcomings as follows.

1. The users have to provide the related works by themselves. In addition to the inconvenience, junior researchers may not be capable of finding relevant papers from a variety of aspects.
2. Unaware of the target research, the generated summary suffers from the lack of focus.

In this work, we propose an intelligent survey generation system, AutoSurvey,<sup>1</sup> to address these two issues. The input of AutoSurvey is not a list of publications, but a short research draft written in natural language. Our system will automatically do literature survey based on a deep understanding of the research draft and generate a corresponding summary of the related work. An overview of our AutoSurvey system is illustrated as Figure 1, where the three core components are denoted in blue.

AutoSurvey is currently aimed at the natural language processing (NLP) domain. The latest journal articles and conference/workshop proceedings on the study of NLP are collected as the source materials. We employ a neural model for cross-lingual information structure analysis for identifying the fine-grained aspects of every individual paper [Huang

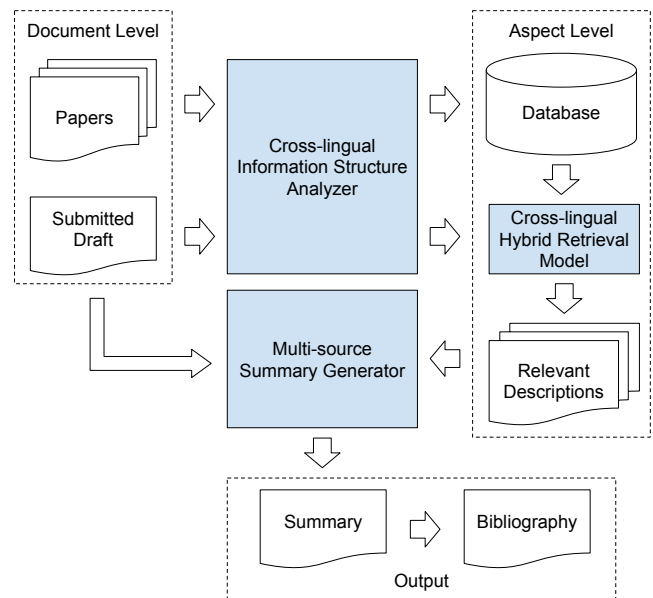


Figure 1: System Overview of AutoSurvey

and Chen, 2017; Huang and Chen, 2018]. The information structure analysis will also perform on the user submitted draft, and a cross-lingual information retrieval model, a hybrid of embedding-based and term-based, will extract related work from the database from different aspects. Finally, a novel evolutionary algorithm for multi-source summarization will generate a summary focus on the research purpose of the draft. The contributions of this work are threefold as follows.

1. We introduce a new direction that extends the scope of scientific article summarization and demonstrate a novel intelligent system, AutoSurvey, that is extremely useful for both academic and educational purposes.
2. We accomplish our goal by exploring cutting-edge NLP methodologies, including cross-lingual information structure analysis and the evolutionary algorithm for multi-source summarization.
3. The cross-lingual ability of our methodologies enables our system handles the research draft written in the language other than English, covering a wider audience.

<sup>1</sup>[http://www.cs.nccu.edu.tw/~hhuang/auto\\_survey/](http://www.cs.nccu.edu.tw/~hhuang/auto_survey/)

## 2 Aspect-based Retrieval of Related Work

We build a crawler that automatically collects abstracts of the latest papers hosted in the ACL Anthology,<sup>2</sup> a repository of top-tier and representative papers in the NLP area. Until now, a total of 11,298 abstracts are crawled.<sup>3</sup> The latest papers from the famous conferences such as ACL, EMNLP, NAACL-CLP, COLING, and LREC are covered from 2016 to 2019. The journal papers in TACL are also included. The papers prior to 2016 are partially included because the BibTex files of earlier papers do not contain the abstract.

For each crawled abstract, our system analyzes its content from five aspects including the purpose, the background, the methodology, the results, and the conclusion of the research by performing information structure analysis. The model for cross-lingual information structure analysis is enhanced from our previous work [Huang and Chen, 2018], achieving an accuracy of 83.29% and a macro F-score of 74.35%. The outcomes of analysis, at the aspect level, are stored in a database.

When a user submits a research draft  $x$ , our system will perform information structure analysis to identify its content from the five aspects and retrieve relevant information from the database. The cross-lingual information retrieval model is a hybrid of embedding-based and term-based approaches. Based on the multi-lingual BERT text-encoder [Devlin *et al.*, 2018; Pires *et al.*, 2019], we train a neural network for measuring the similarity between two descriptions. We also consider the overlapped terminologies between two descriptions since terminologies frequently appear in scientific writing.

The cross-lingual information retrieval will be performed at every aspect. In this way, the resulting survey will be more sensitive to the relationships between the given research draft and every individual related work at different aspects. For example, the draft  $x$  and a paper  $w_i$  share a similar research purpose, then the methodology and the results of  $w_i$  will be highlighted in the summary for suggesting the feasible approaches, while  $x$  and another  $w_j$  share a similar methodology, then the purpose of  $w_j$  will be included in the summary for pointing out the potential applications.

## 3 Multi-source Summarization

Compared to extractive summarization, abstractive summarization attracts much attention in recent years [Liu *et al.*, 2018; Fan *et al.*, 2017]. For multi-document summarization, however, the neural-based abstractive approach is limited in this stage [Bing *et al.*, 2015; Lebanoff *et al.*, 2018; Lin and Ng, 2019]. In this work, we employ the extractive approach to multi-source summarization because of two reasons. Firstly, the large scale dataset for multi-document summarization in scientific writing has yet to construct; it is difficult to train the abstractive summarization model on few data. Secondly and more importantly, the correctness of the generated summary is extremely important for our scenario, while abstractive summarization runs the risk of producing out-of-control results.

<sup>2</sup><https://www.aclweb.org/anthology/>

<sup>3</sup>All materials in the ACL Anthology are available to the general public for non-commercial purposes according to <https://www.aclweb.org/anthology/faq/copyright/>

In general, extractive summarization is aimed at selecting and ordering a subset of sentences that retain the meaning of all sentences as much as possible. Recent work also employs end-to-end neural models [Narayan *et al.*, 2018]. Because of the lack of training data in our domain, we propose a novel evolutionary algorithm for unsupervised multi-source summarization. We formulate the summarization as a constrained optimization problem, where our model has to meet the following four criteria.

1. The length of the summary should not exceed the target length  $l$ .
2. The similarity between the meaning of the summary and the meaning of all retrieved materials should be maximized.
3. The relatedness between the meaning of the summary and the submitted research draft should be maximized.
4. The coherence of the sentence order should be maximized.

Formally, the goal of the summarization model is to select a set of sentences  $s$  from all sentences  $S$  that are produced by the cross-lingual hybrid retrieval model and determine the order of sentences in  $s$ . A fitness function  $f(s)$  is defined to measure how good a solution is, and the solution can be optimized with the genetic algorithm.

For the four criteria, the length is a hard constraint that is straightforward to handle. For the similarity and the relatedness, we train a BERT-based text-encoder for measuring the similarity and the relatedness of two sentence sets. For the coherence, we train another BERT-based text-encoder for modeling sentence order. Note that the first three criteria depend on the combination of  $s$  only, while the fourth criterion depends on the permutation of  $s$ . To reduce the complexity, the summarization is divided into two stages. In the first stage, our system optimizes  $s$  for the first three criteria with the genetic algorithm. In the second stage, our system determines the sentence order by using beam-search.

Some sentences extracted from the original abstracts do not fit the style of a literature survey. Our system contains a post-processing module that fixes the style of the raw summary. For example, the fragment “In this work, We propose ...” will be revised as “*Author Surnames (date)* propose ...”, and the fragment “Our model outperforms ...” will be revised as “The model of (*Author Surnames, date*) outperforms ...”. Figure 2 shows an outcome given the abstract of this paper as input and a limitation of 120 words. Due to the randomized nature of genetic algorithm, our system may generate a slightly different survey given the same input. In future work, sentence compression, a task aimed at pruning long sentences, will be introduced to AutoSurvey for generating more compact summaries.

## 4 Cross-lingual Support

All the three core components shown in Figure 1 are based on cross-lingual text representation. In other words, the user can submit a draft written in a language other than English, providing a friendly interface for users who do not want to write in English. Of course, the cross-lingual processing may not

The work of Giannakopoulos and Pittaras (2019) provides a brief description of the summary evaluation task, the data generation protocol and the resources made available by the MultiLing community, towards improving automatic summarization evaluation. The work of Handler and O'Connor (2018) introduces a new problem, relational summarization, in which the goal is to generate a natural language summary of the relationship between two lexical items in a corpus, without reference to a knowledge base. The work of Goncalo Oliveira (2017) surveys intelligent poetry generators around a set of relevant axis for poetry generation – targeted languages, form and content features, techniques, reutilisation of material, and evaluation – and aims to organise work developed on this topic so far.

Figure 2: A short survey generated by AutoSurvey given the abstract of this paper as input.

be as good as the mono-lingual one. In addition to English, we try to tune the models for handling Chinese input. For other languages, the cross-lingual performance has yet to investigate. Note that our unsupervised summarization model can also generate summaries in other languages. However, the current system can only generate English summaries because all the materials in our database are in English.

## 5 Conclusion and Future Work

This work demonstrates a preliminary idea of automatic survey generation given a research draft. Compared with traditional multi-document summarization, our system can suggest a more dedicated literature survey specific to the user's research direction. Like most generation tasks, performance evaluation is a challenging issue. A comprehensive evaluation and user study will be conducted in the future. The scope of the database will also be extended to other areas in computer science.

## Acknowledgments

This research was partially supported by the Ministry of Science and Technology, Taiwan, under the grants 109-2634-F-002-034- and 109-2634-F-002-041-.

## References

- [Bing *et al.*, 2015] Lidong Bing, Piji Li, Yi Liao, Wai Lam, Weiwei Guo, and Rebecca Passonneau. Abstractive multi-document summarization via phrase selection and merging. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1587–1597, Beijing, China, July 2015.
- [Boni *et al.*, 2020] Odellia Boni, Guy Feigenblat, Doron Cohen, Haggai Roitman, and David Konopnicki. A study of human summaries of scientific articles. *ArXiv*, abs/2002.03604, 2020.
- [Devlin *et al.*, 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.
- [Fan *et al.*, 2017] Angela Fan, David Grangier, and Michael Auli. Controllable abstractive summarization. *CoRR*, abs/1711.05217, 2017.
- [Huang and Chen, 2017] Hen-Hsen Huang and Hsin-Hsi Chen. Disa: A scientific writing advisor with deep information structure analysis. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 5229–5231, 2017.
- [Huang and Chen, 2018] Hen-Hsen Huang and Hsin-Hsi Chen. Cisa: Chinese information structure analysis for scientific writing with cross-lingual adversarial learning. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 5832–5834, 7 2018.
- [Jaidka *et al.*, 2016] Kokil Jaidka, Muthu Kumar Chandrasekaran, Sajal Rustagi, and Min-Yen Kan. Overview of the CL-SciSumm 2016 shared task. In *Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL)*, pages 93–102, June 2016.
- [Lebanoff *et al.*, 2018] Logan Lebanoff, Kaiqiang Song, and Fei Liu. Adapting the neural encoder-decoder framework from single to multi-document summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4131–4141, Brussels, Belgium, October-November 2018.
- [Lin and Ng, 2019] Hui Lin and Vincent Ng. Abstractive summarization: A survey of the state of the art. In *Proceedings of the AAI Conference on Artificial Intelligence*, volume 33, pages 9815–9822, 2019.
- [Liu *et al.*, 2018] Yizhu Liu, Zhiyi Luo, and Kenny Zhu. Controlling length in abstractive summarization using a convolutional neural network. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4110–4119, Brussels, Belgium, October-November 2018.
- [Narayan *et al.*, 2018] Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Ranking sentences for extractive summarization with reinforcement learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1747–1759, New Orleans, Louisiana, June 2018.
- [Pires *et al.*, 2019] Telmo Pires, Eva Schlinger, and Dan Garrette. How multilingual is multilingual bert? *CoRR*, abs/1906.01502, 2019.
- [Yasunaga *et al.*, 2019] Michihiro Yasunaga, Jungo Kasai, Rui Zhang, Alexander R Fabbri, Irene Li, Dan Friedman, and Dragomir R Radev. Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks. In *Proceedings of the AAI Conference on Artificial Intelligence*, volume 33, pages 7386–7393, 2019.