2021

Intention to Use Abstract Sentence Classification Technology

Connor Stead
Macquarie University, connor.stead@hdr.mq.edu.au

Stephen Smith
Macquarie University, stephen.smith@mq.edu.au

Peter Busch
Macquarie University, peter.busch@mq.edu.au

Savanid Vatanasakdakul
Carnegie Mellon University, savanid@cmu.edu

Intention to Use Abstract Sentence Classification Technology

Research-in-progress

Connor Stead
Department of Computing, Faculty of Science and Engineering
Macquarie University
New South Wales, Australia
Email: connor.stead@hdr.mq.edu.au

Stephen Smith
Department of Computing, Faculty of Science and Engineering
Macquarie University
New South Wales, Australia
Email: stephen.smith@mq.edu.au

Peter Busch
Department of Computing, Faculty of Science and Engineering
Macquarie University
New South Wales, Australia
Email: peter.busch@mq.edu.au

Savanid Vatanasakdakul
Carnegie Mellon University
Doha, Qatar
Email: savanid@cmu.edu

Abstract

This paper introduces research in progress to study the intention of researchers to use academic abstract sentence classification technology when undertaking literature acquisition activities. We introduce an enhanced prototypical academic abstract sentence classification system capable of performing on demand sentence classification for metadata results from several academic literature indices. We also outline a preliminary theoretical information systems model developed to explore the intention of researchers to use the system when searching for literature via digital means. Additionally, we provide the survey instrument to be used for review. The overarching body of work this paper introduces will benefit the research community as it is the first-time primary research has been conducted to examine the utility of this technology to improve the way researchers interact more efficiently with the large body of literature digitally available.

Keywords abstract sentence classification, machine learning, natural language processing, intention to use, information systems success

1 Introduction

Academic abstract sentence classification (AASC) is a concept emanating from the computer science artificial intelligence and deep learning research domain referring to the classification of sentences from academic literature abstracts using natural language processing methods. Sentences are classified into literature characteristic classes, with common examples including: 'purpose', 'methodology', 'findings', 'contributions', 'limitations' and 'implications'. To achieve AASC, researchers have developed classification models using artificial intelligence and deep learning approaches. State of the art models can now achieve very high degrees of performance, in some cases exceeding 94% accuracy rates. To learn more about these models please refer to the taxonomy provided in Stead et al. (2021).

Much work has been done in the field on developing increasingly higher performing classification models, however, no research exists – bar our preliminary research in progress work – to posit an information system deploying this capability into a real-world information system to observe and permit human computer interaction in a context where this capability could be beneficial. The context within scope for this paper is the activity of literature discovery, whereby a researcher attempts to retrieve literature relevant to their research pursuits from literature indices – such as Google Scholar and Microsoft Academic. The significance of the research this paper introduces is that it will define a method by which AASC can be utilised to progress academic research, reduce the information burden on researchers and evolve into a real-world application from exclusively an artefact of the computer science domain.

2 Academic Abstract Sentence Classification

AASC refers to "the development of machine/deep learning prediction models capable of classifying academic abstract sentences into classifications representative of key literature characteristics, such as the explicit section headings observed in structured abstracts (Purpose, Method, Findings etc.)." (Stead et al., 2020b, p. 2). Essentially, this technology applies machine/deep learning and natural language processing techniques to classify sentences from academic abstracts as being indicative of 'Purpose', 'Method', 'Findings' or 'Contribution' sentences, which we refer to as 'literature characteristics'. These characteristics relate to the adoption of structured abstracts.

The 'structuring' of abstracts refers to the use of explicit headings within academic abstracts to create a form of 'structure', as opposed to abstracts containing purely free text. An abstract is 'structured' if it features sentences categorised within explicitly defined literature characteristic headings such as 'Purpose', 'Method', 'Findings' and 'Contributions'. The value of structured abstracts has been well documented in the literature. One example is the suggestion by multiple studies that the use of structured abstracts increases researcher's literature discovery capability (Eldredge, 2006, Budgen et al., 2008, Hartley et al., 1996). Eldredge (2006), for example, notes that "the structured abstract format enables busy librarians to assess quickly whether a professional communication, either published or unpublished, contains needed evidence and then to extract that evidence." (p. 345). Hartley et al. (1996) found that researchers can extract information more quickly and accurately from structured abstracts when compared to traditional non-structured variants. Budgen et al. (2008) also conducted quantitative research on the utility of structured abstracts through a survey of 64 researchers and students. They determined that non-structured free text abstracts "are likely to omit substantial amounts of relevant information" (p. 457) and that structured variants "are significantly more complete and clearer than unstructured abstracts" (p. 457).

Table 1: Summary of state-of-the-art academic abstract sentence classification models

Paper	Algorithm	Performance
Dernoncourt et al. (2016)	Neural Network	PubMed 200k (Dernoncourt and Lee, 2017): F1-score: 89.9%
Gonçalves et al. (2019)	Neural Network	PubMed 20k: 90.9% precision, 90.8% recall /F1-score
Jiang et al. (2019)	Text convolutional neural network (CNN) + bidirectional recurrent neural network (bi-RNN).	PubMed 200k: 94.4% accuracy (p. 8)

Advances in artificial intelligence, deep learning approaches and natural language processing techniques have led to development of classification models capable of classifying sentences sourced from academic abstracts as indicative of key literature characteristics, such as a study's 'Purpose', 'Method', 'Findings' or 'Contribution'. This modelling capability can be used to transform unstructured abstracts and reap

the observed benefits of structured abstracts observed in the literature without requiring a restructure of an abstract. These models and the literature contributions associated with their development are products of the computer science and language technology disciplines. Table 1 provides a summary of the highest performing classification models, identifying their algorithm and performance characteristics. These are a sample from ongoing taxonomy efforts we are undertaking to capture the development of classification models over time.

3 Purpose and Motivation

The focus of the research community concerned with academic abstract sentence classification modelling has been to continue to develop increasingly higher performing classification models, however, no research has been conducted to examine the utility and perceived success of this technology should it be adopted by users within a defined context. Accordingly, the purpose of this branch of our research is to explore what intention researchers have to use AASC capability should it be made available to them in a information system within the context of literature acquisition. We are motivated to conduct this study primarily due to the absence of any research to date that has explored what utility and value this technology has for real world application, particularly for the research community.

4 Academic Abstract Sentence Classification Information System

In prior work (Stead et al., 2020a) we have outlined the ongoing development of a prototypical AASC information system, one which deploys AASC and leverages it to transform unstructured abstracts. The system is a web-based platform consisting of a backend framework for processing inbound literature search queries/subsequent sentence classification with a front-end user interface for querying and result observation. The backend framework is currently by setup to interact with Scopus, the Directory of Open Access Journals (DOAJ) and AIS eLibrary to retrieve literature metadata. The system is capable of interacting with any literature metadata source via either an API call or flat file import. The backend framework is hosted on a Flask (https://flask.palletsprojects.com/en/2.0.x/), a Python based web framework and was used as it enabled a straightforward method to deploy the literature index outbound query, metadata transformation, sentence extraction/tokenization, classification and JSON conversion logic – all of which was developed independently in Python – into a single solution. Flask allows Python code to be callable on receipt of a web query to the Flask instance. The system by default is setup to listen to a HTTP GET request of the following format:

http://127.0.0.1:8000/classifier?query=[**x**]&article_n=[**y**]&id=[**z**]

The 'query' parameter is for a Boolean search query to be queried against the literature index or metadata flat file resource. There are varying methods for interacting with each of the three academic literature indices deployed to date. For DOAJ the DOAJ article search API (https://doaj.org/api/v2/docs) is used – in the form of another GET request. For Scopus queries their API is also used (https://dev.elsevier.com/api_docs.html), again in the form of authenticated GET requests. AIS eLibrary leverages a bespoke query bot we developed to automate interaction with AIS eLibrary's front end search platform (https://aisel.aisnet.org/). Post retrieval of metadata (particularly abstracts) from the academic literature indices, the metadata is then processed by a sentence tokenizer which splits the metadata dataframe, containing one observation per resulting article, into a dataframe with one observation per sentence. The other corresponding metadata is maintained with this transformation and each article's sentences are maintained together within the sentence level dataframe using a unique article identifier. The sentence level observations are then processed by a deployed classification model. At this stage we are using a classification model developed by ourselves for the purpose of researching the utility of AASC adoption in the context of literature discovery activities, such as through the development of this system. The model was developed using the XLNet generalized autoregressive pretraining modelling method (Yang et al., 2019) and trained using the Emerald 20k AASC dataset (Stead et al., 2019), which contains 201,452 author/publisher pre-classified sentences from 20,000 multidisciplinary abstracts.

After the classification is complete the system converts the resulting classified and scored dataframe to a JSON string which is returned to the origin request. The JSON is structured as an array with elements at an article level. Key article metadata is included in each element including: the DOAJ article identifier, the title, authors, publication (source) and author(s). In addition to this metadata, four arrays are provided per article element. The first array contains an incremental counter for each sentence extracted from the article's corresponding abstract. The second array contains an element for each extracted sentence. The third array contains the classification assigned to each sentence, such as 'purpose' or

'findings'. The final array contains the corresponding score which is indicative for the strength of the sentence's assigned classification. The resulting JSON string is demonstrated with the below example:


```
[{"id": "0a9c4fdbf93e4453bb814c5df05cd6b0","title": "Hematologic profile of Amazon
river dolphins Inia geoffrensis and its variation during ……","authors": "Daniela M
D de Mello; Vera M F da Silva; ","source": "PLoS ONE","year": "2019","sentence_n":
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14],"sentence": ["Hematological values
are of primary importance when investigating the.", "The objectives of this study are
to ……],"classification": ["purpose", "purpose", "design_ methodology_approach",
"design_methodology_approach" ……, "originality_value"],"classification_score":
[0.43291670083999634, 0.9791497588157654, 0.9478005170822144, 0.66685664 65377808
……]}]
```


The front end is a web interface coded in HTML and JavaScript. The JavaScript interacts with the backend and processes the resulting JSON creating a literature index like interface with article metadata shown along with resulting abstracts. The JavaScript also applies classification dependent formatting to the sentences when displayed. Each sentence features a background colour depending on its classification. This allows the end user to easily distinguish the separate components of the classified abstracts, similar to how a structured abstract would be presented. A 'tooltip' is provided on the page, which shows a small black box upon the cursor hovering on a sentence. The 'tooltip' contains the classification and the classification score for the sentence. The frontend also contains a search bar to query for articles, similar to, ordinary literature indices. There is also a classification bar which allows the user to turn certain literature characteristic classes on and off. A CSV download button is also present, which exports the classified abstracts shown on the page to a .CSV file and forces the download to the user's internet browser session. Figure 1 shows a screenshot of the system. The 'method/approach' characteristic class is turned off. The sentence with no highlighting shown is the corresponding 'method/approach' sentence. The 'tooltip' is also shown as the user is hovering over the first 'findings' sentence. There is also a 'Download CSV' button which downloads to the user's web browser a sentence level classified .CSV file showing each result'

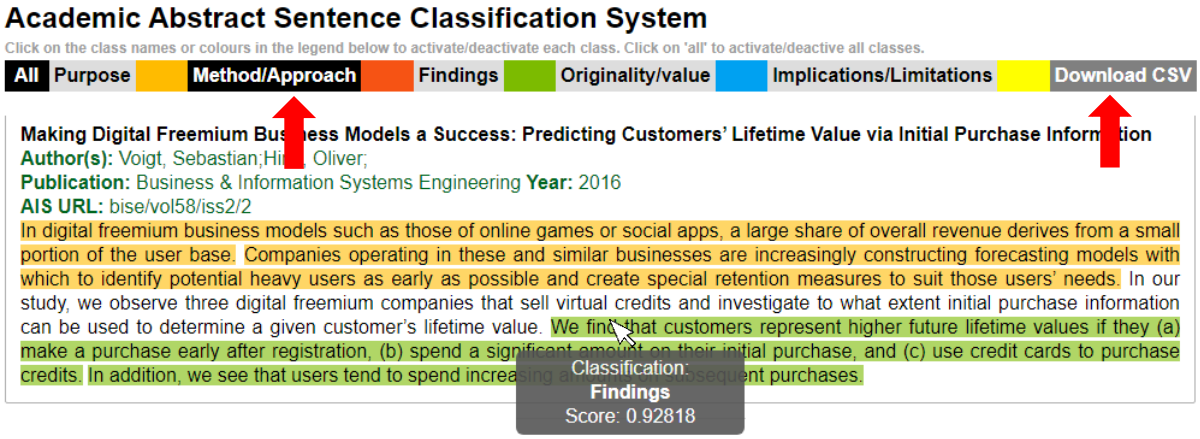

Figure 1: A screenshot of the abstract sentence classification system

5 Intention to Use Academic Abstract Sentence Classification

We are at a stage now in our study on AASC to begin primary research on the perceived success of the technology being adopted by researchers in their efforts to increase efficiencies in their literature acquisition activities. To do this, we have developed a preliminary theoretical model leveraging a component of DeLone and McLeans's (1992, 2003) information systems success model. The examination of this preliminary model will be then used to assist in the development of a more comprehensive theoretical model to be examined in the future through further primary research. Our preliminary model sets forward to examine whether the quality of the prototypical AASC system and the information output produced will lead to an intention of surveyed users to use the system in the context of literature acquisition. If it is determined that users do have an intention to use the system then further research is warranted on AASC utility and adoption, if not, then further research is to be conducted determining why this is the case asnd how best AASC can be positioned to be leveraged by researchers.

The intention to use construct is defined as the degree and manner in which the AASC it intended to be used by researchers to assist in conducting literature acquisition. The system quality construct reflects

the desirable characteristics of the AASC system, with a focus on the way in which the system permits interaction with the user. The information quality construct reflects the desirable characteristics of the AASC system's output, with a focus on the transformation of the corresponding article abstracts using AASC model capability. To examine the system quality and information quality constructs we adopt first and second order constructs. These are used when constructs are complex in nature and difficult to study directly (Hair et al., 2014). For the system quality construct we used the reliability and integration first order constructs. Reliability refers to the ability for the researcher to rely on the system in accurately classifying the abstract sentences to a corresponding literature class. Integration refers to the ability of the system to integrate into their approach to literature acquisition. For the information quality construct we used the first order constructs of accuracy, completeness and format. Accuracy refers to the reliability of the transformed classified output of the AASC system – for example, the ability for it to present 'finding' sentence accurately. This also applies to the export function, in that the system displays results classified correctly in the output file. Completeness refers to the classified output communicated to the user. For completeness to occur all sentences should be classified appropriately and the user should not be provided incomplete classification output. Format refers to the presentation of the classified output, in both the sentence highlighting on the literature index type webpage and the .CSV export. All of the first order constructs adopted to study both DeLone and McLeans's (1992, 2003) system quality and information quality constructs were adapted from Nelson et al. (2005). Figure 3 shows our preliminary theoretical model.

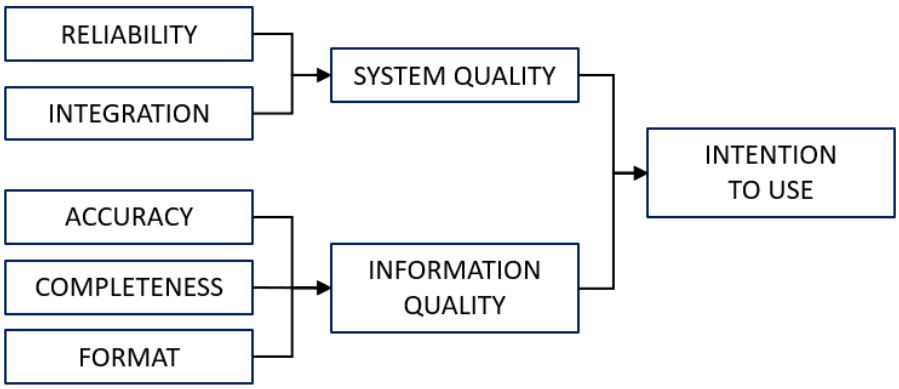

Figure 3: A preliminary abstract sentence classification intention to use model

We are now conducting primary research to explore the intention of researchers to use AASC technology to improve efficiencies in literature acquisition. To do this, we have formed a series of survey questions shown in appendix 1. The questions capture both participant demographics and use a 7-point Likert scale to capture participants perspectives on the constructs identified in the theoretical model (Figure 3). The participant demographics section will allow us to understand how the responses vary depending on the characteristics of researchers, such as their experience, location and frequency of academic literature index (Google Scholar, Scopus, AIS eLibrary or Pubmed) use. The remaining questions primarily seek to assess the first and second order constructs in the theoretical model. The bold text within square brackets shown in most of the questions in the Likert scale section identifies which first or second order construct is captured by the question. The are 4 additional questions at the end of this section which seek to capture participant perspectives on the utility of AASC within the context of researcher literature acquisition. These questions are a pre-cursor for further research into benefits of AASC, but do not directly relate to our preliminary theoretical model which specifically targets intention to use.

We have recently obtained human research ethics approval from Macquarie University to begin surveying participants. We intend to select a sample of authors who have published research indexed in the information systems literature index AIS eLibrary (https://aisel.aisnet.org/). AIS eLibrary is a "... central repository for research papers and journal articles from the information systems academic community. Created and maintained for AIS Members and AIS eLibrary Subscribers, you can browse and download papers from ICIS, AMCIS, ECIS, PACIS and from all AIS conferences, affiliated conferences and content from AIS SIGs, Chapters and the most prominent academic IS journals including JAIS, CAIS, TRR, THCI, MISQE, PAJAIS and AIS affiliated journals" (https://aisel.aisnet.org/). We assume that authors of literature indexed in AIS eLibrary would be likely users of AIS eLibrary to search for and retrieve literature.

From the sample of participants, we will send at a maximum two emails inviting them to participate in a survey on the utility of academic abstract sentence classification technology. If they accept the invitation, the user will be provided with a link to the system, which is hosted on a secure web portal. The prototype system is available at https://www.abstractsentenceclassification.com. The user will not be provided search functionality with literature indices, rather, they will be provided a prepared result containing 100 pre-processed abstracts. The reason we are not permitting search functionality is to control the sample of abstracts reviewed by participants – to ensure that the variety in results does not impact their perspective on the technology. We are using Qualtrics (https://www.qualtrics.com) to collect survey responses and will use structural equational modelling – partial least squares to statistically analyse the results.

6 Ongoing Research

We are actively collecting survey responses and it is hoped that a sufficient response size will be collected in time for ACIS 2021 so that indicative findings can be discussed. We are also working on improving the prototypical system and developing a more comprehensive theoretical model to examine the role this technology can play in improving the way researchers explore the extensive body of academic literature. This study is the primary research component of a PhD research thesis and is the first-time participants have been involved in this long term research project. It is exciting to study how the research community reacts to the technology and how our project will evolve as a result.

7 References

Budgen, D., Kitchenham, B. A., Charters, S. M., Turner, M., Brereton, P., & Linkman, S. G. (2008). Presenting software engineering results using structured abstracts: a randomised experiment. Empirical Software Engineering, 13(4), 435-468.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. Information systems research, 3(1), 60-95.

DeLone, W. H., & McLean, E. R. J. J. o. m. i. s. (2003). The DeLone and McLean model of information systems success: a ten-year update. 19(4), 9-30.

Dernoncourt, F., Lee, J. Y., & Szolovits, P. (2016). Neural Networks for Joint Sentence Classification in Medical Paper Abstracts. arXiv preprint arXiv:1612.05251.

Eldredge, J. (2006). Evidence-based librarianship: the EBL process. Library hi tech, 24(3), 341-354.

Emerald Group Publishing. (2019). A 6 step guide to writing an abstract. Retrieved from http://www.emeraldgrouppublishing.com/authors/guides/write/abstracts.htm

Gonçalves, S., Cortez, P., Moro, S. J. N. C., & Applications. (2019). A deep learning classifier for sentence classification in biomedical and computer science abstracts. 1-15.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). A Primer on Partial Least Squares Structural Equation Modeling (SEM-PLS). Thousand Oaks, California: SAGE Publications, Inc.

Hartley, J., Sydes, M., & Blurton, A. (1996). Obtaining information accurately and quickly: are structured abstracts more efficient? Journal of Information Science, 22(5), 349-356.

Jiang, X., Zhang, B., Ye, Y., & Liu, Z. (2019). A Hierarchical Model with Recurrent Convolutional Neural Networks for Sequential Sentence Classification. Paper presented at the CCF International Conference on Natural Language Processing and Chinese Computing.

Mulrow, C. D. (1987). The medical review article: state of the science. Annals of internal medicine, 106(3), 485-488.

Nakayama, T., Hirai, N., Yamazaki, S., & Naito, M. (2005). Adoption of structured abstracts by general medical journals and format for a structured abstract. Journal of the Medical Library Association, 93(2), 237.

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: an empirical examination within the context of data warehousing. Journal of management information systems, 21(4), 199-235.

Stead, C., Smith, S., Busch, P., & Vatanasakdakul, S. (2019). Emerald 110k: A Multidisciplinary Dataset for Abstract Sentence Classification. Proceedings of the 17th Annual Workshop of the Australasian Language Technology Association (ALTA)

Stead, C., Smith, S., Busch, P., and Vatanasakdakul, S. (2020a). "Towards an Academic Abstract Sentence Classification System,". Research Challenges in Information Science 2020 Proceedings

Stead, C., Smith, S., Busch, P., and Vatanasakdakul, S. (2020b). "An Activity Theory Investigation of Academic Abstract Sentence Classification,". Twenty-Third Pacific Asia Conference on Information Systems, Dubai, UAE

Stead, C., Smith, S., Busch, P., and Vatanasakdakul, S. (2021). "A Taxonomy of Academic Abstract Sentence Classification Modelling,". 18th conference of the Italian Chapter of AIS

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., and Le, Q. V. (2019). "Xlnet: Generalized autoregressive pretraining for language understanding," Advances in Neural Information Processing Systems, pp. 5754-5764.

Appendix 1 – Survey Questions

(1) Please state your primary research field. **(2)** If employed, which sector do you work in: Educational institution (e.g., university), Government, Private industry, Other **(3)** In which country do you reside? **(4)** How many years have you been conducting academic research? 0 – 5 years, 6 – 10 years, 11 – 15 years, 15+ years **(5)** On average, how many times a week do you use a literature index - such as Google Scholar, Scopus, AIS eLibrary or Pubmed?

7-point Likert scale: 1. Strongly Disagree, 2. Disagree, 3. Somewhat Disagree, 4. Neutral, 5. Somewhat Agree, 6. Agree, 7. Strongly Agree

[RELIABILITY] The abstract sentence classification system **(1)** operates reliably, **(2)** performs consistently

(3) [RELIABILITY] The operation of the abstract sentence classification system is dependable.

(4) [ACCESSIBILITY] The abstract sentence classification system is **(4)** accessible, **(5)** responsive

(6) [ACCESSIBILITY] Information can be obtained from the abstract sentence classification system when needed.

(7) [SYSTEM QUALITY] Overall, the abstract sentence classification system is of high quality.

(8) [SYSTEM QUALITY] Overall, I would give the abstract sentence classification system a high rating.

(9) [SYSTEM QUALITY] Overall, the functions provided by the system are of high quality.

(10) [ACCURACY] The sentence classification output is reliable.

(11) [ACCURACY] The sentence classification output contains minimal errors.

(12) [ACCURACY] There is a low frequency of error in the sentence classification output.

(13) [COMPLETENESS] The breadth, depth, and scope of information contained in the output is sufficient.

(14) [COMPLETENESS] The information communicated by the sentence classification output is comprehensive.

(15) [COMPLETENESS] The sentence classification output is appropriate for the research discovery task at hand.

(16) [FORMAT] The sentence classification output is well presented.

(17) [FORMAT] The sentence classification output is presented in a useful format.

(18) [FORMAT] The sentence classification output is beneficial and provides advantages as a result.

(19) [INFORMATION QUALITY] Overall, the sentence classification output is of high quality.

(20) [INFORMATION QUALITY] Overall, I would give the sentence classification output a high rating.

(21) [INFORMATION QUALITY] Overall, the sentence classification system provides high quality information.

(22) [INTENTION TO USE] If it were made available, I intend to use abstract sentence classification: **(23)** consistently, **(24)** frequently, **(25)** constantly, **(26)** regularly.

(27) If it were made available, abstract sentence classification would help my research.

(28) If it were made available, abstract sentence classification would help me find relevant research easier.

(29) If it were made available, abstract sentence classification would introduce efficiencies in my work.

(30) If it were made available, abstract sentence classification would be beneficial for researchers.

(31) Please provide any comments on your experience using academic abstract sentence classification technology.

Acknowledgements

The authors wish to acknowledge the Australian Government Research Training Program Scholarship which enabled this research to take place.

Copyright