Abstract Sentence Classification Acceptance

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

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

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

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

Abstract Sentence Classification Acceptance

Full research paper

Connor Stead
School of Computing
Macquarie University
Email: connor.stead@hdr.mq.edu.au

Stephen Smith
School of Computing
Macquarie University
Email: stephen.smith@mq.edu.au

Peter Busch
School of Computing
Macquarie University
Email: peter.busch@mq.edu.au

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

Abstract

Background: Abstract sentence classification modelling is a natural language processing approach to the classification of sentences from academic literature abstracts as indicative of a study's key literature characteristics commonly used as structured abstract headings, such as background, aims, method and results. Much work has been done to advance abstract sentence classification modelling capability in the computer science domain, however, there has been no examination on the acceptance of this technology by researchers in the context of conducting academic literature discovery.

Aims: This study aims to explore acceptance of abstract sentence classification modelling capability by information systems researchers using the technology acceptance model theoretical framework, which was extended to include the indicators originating from the information retrieval and perceived web quality research domains.

Method: A prototype system deploying abstract sentence classification modelling capability into a pseudo academic literature index interface was developed. A conceptual model grounded in the technology acceptance model was then produced after consideration of existing relevant literature, before a survey instrument was prepared using measurement items identified and extended from existing literature. Information systems researchers who had published in leading information systems conference proceedings between 2019 and 2021 were invited to participate in the survey, with their survey responses used to evaluate the measurement and structural models using partial least squares structural equational modelling.

Results: Results indicated that the capability of the prototype abstract sentence classification modelling system to perform sentence classification, the appearance of the prototype system and supplementary system features all positively and significantly effected perceived usefulness and ease of using abstract sentence classification modelling. Perceived usefulness and ease of using abstract sentence classification modelling had a positive and significant effect on participant attitude towards abstract sentence classification modelling and intention to use abstract sentence classification modelling. Participant attitudes towards the concept of abstract sentence classification modelling also had a positive and significant effect on intention to use abstract sentence classification modelling.

Keywords Structured Abstract, Abstract Sentence Classification, Technology Acceptance, TAM.

1 Introduction

The volume of academic literature available on the internet is unprecedented and is problematic for researchers, as they seek to identify papers relevant to their research. Researchers use online 'academic search engines and bibliographic databases' (ASEBDs) (Gusenbauer, 2019) to conduct literature reviews (Haines et al. 2010, Gusenbauer, 2019). Examples of ASEBDs include Google Scholar; Microsoft Academic (now defunct), ProQuest, Scopus (Gusenbauer, 2019) and the Association for Information Systems (AIS) eLibrary. By serving as the portal to academic literature, ASEBDs provide a means to gauge the volume of academic literature available. Google Scholar is commonly known as the most popular ASEBD (van Noorden, 2014) and, accordingly, has been the subject of substantial research seeking to examine the size of its scholarly research collection (Mayr and Walter, 2007; Aguillo, 2012, Khabsa and Giles, 2014, Orduña-Malea et al., 2015, Orduña-Malea et al., 2014). Gusenbauer (2019), extending the work of others, explored the volume of material indexed on ASEBDs using a method known as 'query hit count' (QHC), previously employed by Vaughan and Thelwall (2004) and Orduña-Malea et al. (2015). The QHC method involves submitting a series predefined queries to an ASEBD and observing the resulting response size. Using the size of the resulting queries a minimum indication of the size of the ASEBD is determined. Gusenbauer's (2019) results are remarkable. In 2018, Google Scholar returned a QHC of 389 million records. Other ASEBDs with QHCs in the 100s of millions include: WorldWideScience; a collection of ProQuest databases, Bielefeld Academic Search Engine (BASE), a collection of Web of Science databases and a collection of EbscoHost databases (Gusenbauer, 2019). Another interesting observation from Gusenbauer (2019) is the extent of average monthly ASEBD growth rates, with suggestions that Google Scholar adds an additional 1.6 million literary material per month (Gusenbauer, 2019, p. 194).

The term 'literature overload' is referred to and exists as a motivational practical problem across a range of studies concerned with finding innovative ways for researchers to access relevant research (Galli et al., 2020, Frisoni et al. 2022, Hobert and von Wolff, 2019). This overload impedes a researcher's ability to conduct efficient literature reviews and comprehensively assess a research domain's literature in a timely and sane manner. The phenomenon has another impact on the evolution of research in a domain – the increasing likelihood that highly cited literature will continue to be retrieved and cited, instead of more recent studies with potentially greater relevance (Farrell et al., 2022), which "… can result in slowing of scientific progress as transformative ideas are less likely to permeate and make substantive impact" (Farrell et al., 2022, p. 1). Westgate et al. (2018) argue that due to the volume of academic literature available, a 'synthesis gap' has formed. In describing the synthesis gap, Westgate et al. (2003) cite McKinnon et al. (2015) who explore the policy implications of only a subset of scientific research used as evidence for science-policy interactions. McKinnon et al. (2015) provide examples of the evidence, citing Fisher et al. (2013), suggesting that "…many of the tens of thousands of documents that are produced every year to assess the impacts of sustainability policies and programmes are never read" (McKinnon et al., 2015 p. 185, citing Fisher et al., 2013). McKinnon et al. (2015) provide another example of the variance between the evidence subset of overall scientific resources, citing Doemeland and Trevino's (2014) finding, that 32% of the World Bank's policy reports published between 2008 and 2012 were never downloaded (p. 13).

One approach to improving researcher literature discovery and access is the adoption of structured abstracts. The structuring of abstracts refers to the inclusion of explicit headings reflecting key characteristics of a study - such as its purpose, method, findings, and contributions. Structured abstracts provide greater access to key findings of studies (Mosteller et al., 2004); contain information of higher perceived relevance (Budgen et al., 2008), are easier to read and more accessible (Kitchenham et al., 2008 and Budgen et al., 2008), as well as in general being of higher quality and of greater value to researchers than unstructured abstracts (Sharma and Harrison, 2006, Guimarães, 2006). Academic abstract sentence classification (AASC) refers to the leveraging of artificial intelligence (AI) natural language processing (NLP) methods which classify sentences academic abstract sentences into structured abstract subheadings. In practice, AASC involves the development of NLP models using datasets constructed from pre-classified structured abstract sentences. The resulting models are referred to as abstract sentence classification models (ASCM), and are achieved using a range of AI NLP methods, including conditional random fields (Hirohata et al., 2008, Chung, 2009, Kim et al., 2011 and Hassanzadeh et al., 2014); support vector machines (Shimbo et al., 2003, McKnight and Srinivasan, 2003, Ito et al., 2004, Yamamoto and Takagi, 2005, Liu et al., 2012 and Nam et al., 2016) and neural networks (Dernoncourt et al., 2016, Jin and Szolovits, 2018 and Gonçalves et al., 2018). For a comprehensive analysis of known ASCM refer to Stead et al. (2021a). We have also published an online taxonomy of known ASCM development at www.abstractsentenceclassification.com/taxonomy.html.

Whilst ASCM has matured, there are no known implementations of the technology into public production information systems intended to aid researchers in searching for and retrieving academic literature, such as ASEBDs. There exists a mismatch between the extent of ASCM technological capability and the evaluation of its utility for researchers conducting literature discovery activities. Researchers have managed to develop high performing ASCM, however, with no known deployment of ASCM into production information systems, there is no understanding of the utility of AASC to researchers conducting literature discovery activities. Without this understanding, further investment in ASCM development and evaluation is questionable, as its value to the broader body of knowledge and research community is unknown. It seems the only motivating factor at this time in relation to further ASCM development is a desire to improve on the performance characteristics of prior ASCM development. With the problems facing researchers with respect to the volume of literature and the growing burden of conducting comprehensive literature reviews, innovative approaches are required to help them conduct literature discovery activities more efficiently and effectively. It's believed that deploying the capability of ASCM into a researcher facing information system would be an innovative and novel implementation of ASCM, which is theoretically possible given the known capability of ASCM. The next section will introduce a prototype information system deploying ASCM into a pseudo ASEBD for examination of ASCM utility.

2 Prototype Development

We have developed a prototype system known as the 'Academic Abstract Sentence Classification System', available online (www.abstractsentenceclassification.com). The prototype system architecture consists of a web application connected to a Python instance running Flask, a Python web framework enabling REST request processing. Via a REST query, the Python instance forwards a Boolean search query originating from the web application to either the DOAJ, Elsevier Scopus or AIS eLibrary ASEBDs of the user's choice and awaits a response from the ASEBD. Upon receipt of the abstracts and other literature metadata from the ASEBD, the Python instance pre-processes the resulting dataset and prepares it for classification. The core pre-processing activity is the deconstruction of abstracts into sentences, as well as sub tasks such as the removal of additional text such as copyright statements and author contact information. The Python instance then submits the resulting ASEBD dataset containing abstracts and other literature metadata to a bespoke XLNet (Yang et al., 2019) ASCM we have developed, which assigns each sentence a literature characteristics, such as 'Purpose'; 'Method/Approach' and 'Findings'.

Upon completion of the sentence classification process, the Python instance converts the resulting dataframe into a JSON response for the web application. Each element in the JSON response contains the following data for each publication received from the ASEBD: (1) the paper's title; (2) the paper's authors; (3) the origin publication name (i.e., journal or conference); (4) the origin publication's year; and (5) an array of sentences for the paper's abstract. For each sentence in the array the sentence's literature characteristic, or structured abstract heading class (i.e., 'Purpose', 'Method/Approach', 'Findings'…) is identified, as well as a confidence score ranging from 0 to 100 indicating the reliability of the classification. The web application displays literature resulting from the search query alongside metadata and abstracts with sentences graphically categorised into literature characteristics using text highlighting and a popup box following the user's cursor identifying a sentence's corresponding literature characteristic as the user moves their cursor around the page. Further explanation of the prototype can be found in Stead et al., 2020a; Stead et al., 2020b; Stead et al., 2021a and Stead et al., 2021b.

3 Theoretical Model

With the prototype system deploying ASCM into a production public facing environment for researchers developed, its utility and value for researchers was examined. To do so, the IS technology acceptance theoretical framework, specifically the technology acceptance model (TAM) (Davis, 1985), was adopted. The acceptance of new and evolving technology by users and organisations is a common line of inquiry in the IS discipline. A popular and accepted theoretical framework is known as technology acceptance (Yousafzai et al., 2007; Marangunic and Granic, 2015), an adaptation of the theory of reasoned action (Davis, 1989; Davis et al., 1989). TAM was first proposed in a PhD thesis by Davis (1985) at the Massachusetts Institute of Technology. TAM has two objectives; firstly, it should improve the understanding of user acceptance by providing insights into the design and implementation of information systems (Davis, 1985, p. 7) and it "… should provide the theoretical basis for a practical 'user acceptance testing' methodology that would enable system designers and implementors to evaluate proposed systems…" (Davis, 1985, p. 7). Davis (1985, citing Bewley et al., 1983; Card et al., 1978 and

Gould et al., 1983) notes that the IS literature suggests "… testing system prototypes with actual prospective users is an effective way of evaluating and refining proposed designs" (Davis, 1985, p. 9). Davis (1985) positions TAM to understand the relationship between the characteristics of an information system and use of such system. Specifically, TAM "… is intended to describe the motivational processes that mediate between system characteristics and user behaviour" (Davis, 1985, p. 10).

The core constructs of TAM are 'Perceived Usefulness'; 'Perceived Ease of Use'; 'Attitude Toward Using' and 'Actual System Use'. Davis (1985) describes the relationship between these constructs as follows. Attitude towards using a system is a major determinant of whether the user uses the system (Davis, 1985, p. 24). Attitude towards using a system is a function of both the perceived usefulness of the system and the perceived ease of using the system (Davis, 1985, p. 24). Perceived ease of use has a causal effect on perceived usefulness (Davis, 1985, p. 24). The relationships between the constructs are linear (Davis, 1985, p. 25) and both perceived usefulness and perceived ease of use are posited as of "primary relevance for computer acceptance behaviour" (Davis et al., 1989, p. 985). Davis (1985) defines perceived usefulness as "the degree to which an individual believes that using a particular system would enhance his or her job performance" (Davis, 1985, p. 26; Davis, 1989, p. 320). Davis (1989) invokes the definition of useful when further explaining the construct, being "capable of being used advantageously" (p. 320), and when a user believes a 'positive use-performance relationship' exists then a system features high perceived usefulness (p. 320). Perceived ease of use is defined as "the degree to which an individual believes that using a particular system would be free of physical and mental effort" (Davis, 1985, p. 26). Davis (1989) suggests that "an application perceived to be easier to use than another is more likely to be accepted by users" (p. 320). TAM is adopted in this research study to examine the acceptance of a prototypical system deploying ASCM capability by researchers in the social situation of the conduct of literature discovery activities. TAM will guide the theoretical framework to permit an assessment of the perceived usefulness; ease of use, attitude towards using and intention to use AASC.

Instead of directly applying the typical TAM conceptual model to our area of study we included indicators addressing the characteristics of the literature discovery context under examination. 'Literature Index Experience (LIE)' was proposed as a new construct seeking to capture the experience the participant has in using ASEBDS to search for literature. LIE is believed to have a causal effect on both perceived ease of use and perceived usefulness, as it is posited from our understanding of the technology and development of the prototype, that those who have used ASEBDS before and use ASEBDS frequently would find the prototype easy to use and useful due to its enhancement of typical ASEBD functionality. An additional construct capturing unique characteristics of the literature discovery context was posited, that being 'Researcher Experience (RE)', which intended to capture participant's overall experience in conducting academic research. It was posited that those who have less experience in conducting academic research would have a perception that ASCM is more useful and of greater ease of use compared to participants who have greater experience in conducting academic research. Examples of participants with a less experience would include undergraduate students, Master of Research students or perhaps those undertaking PhD research. It is thought that such persons would appreciate the capabilities afforded by ASCM due to the technology's capability to introduce the benefits associated with structured abstract adoption (Mosteller et al., 2004; Budgen et al., 2008; Kitchenham et al., 2008; Budgen et al., 2008; Sharma and Harrison, 2006; and Guimarães, 2006), making it easier to find relevant literature in new research domains. Those with greater experience in conducting academic research may not require the additional benefits obtained from the introduction of ASCM into an ASEBD, as they may already have a solid grasp of their research domain and have established practices of discovering relevant literature, perhaps without the use of ASEBDS.

Our conceptual model also features additional indicators with causal effects on both perceived usefulness and perceived ease of use, measuring various characteristics of the prototype system and ASCM functionality. The constructs of 'Content Quality'; 'Technical Adequacy', 'System Appearance' and 'Specific Content' were adopted from online health information retrieval (HIR) literature (Bliemel and Hassanein, 2006). Bliemel and Hassanein (2006) developed a theoretical model featuring these constructs to examine consumer satisfaction with HIR systems, which refers to the acquisition of health-related information on information systems, such as the internet (Bliemel and Hassanein, 2006) or social media, academic journals and hospital records (Goeuriot et al., 2016). 'Content Quality' refers to the "usefulness, clarity, completeness, currency, conciseness, and accuracy" (Bliemel and Hassanein, 2006, p. 88) of the information provided by a HIR. In the context of this study and our prototype system, 'Content Quality' is renamed to 'Classification Quality' and relates to the aforementioned characteristics main functionality of the system - sentence classification, particularly, the ability to retrieve key literature characteristics (purpose, method etc.) from abstracts. 'Technical Adequacy' refers to "the technical adequacy of a Web site comprises aspects such as the speed with which pages load, searching capabilities, personalization and customization features, and the ease of accessing the site" (Bliemel and Hassanein, 2006, p. 13, citing Aladwani & Palvia, 2002). 'System Appearance' refers to "the appearance dimension of system quality includes overall attractiveness, organization, proper use of fonts, proper use of colors, and proper use of multimedia" (Bliemel and Hassanein, 2006, p. 13, citing Aladwani & Palvia, 2002) and 'Specific Content' reflects "information about the Web site and authors such as contact information, privacy policies, and support information" (Bliemel and Hassanein, 2006, p. 12, citing Aladwani & Palvia, 2002).

TAM was selected as an appropriate theoretical model to apply to this due to its demonstrated use in exploring the acceptance of novel technologies in individual settings particularly in the IS discipline, among others. TAM is used as a common theoretical model in IS, with a number of studies in recent AIS conference proceedings featuring TAM (Jabagi et al., 2021; Zamir and Kim, 2021; Zhang et al., 2022; Joshi and Sondhi, 2023; David and Zu, 2023; David and Zu, 2023; Amirabdolahian et al., 2022; Dagliyan et al., 2022; Chawla and Joshi, 2020; Baah-Peprah et al., 2022; Weidhaas et al., 2021). An alternative to TAM is the Unified Theory of Acceptance and Use of Technology (UTUAT) (Venkatesh et al., 2003). TAM is one of the 8 IS acceptance models 'unified' in UTUAT, which makes UTUAT a comprehensive instrument to examine technology acceptance. TAM, however, was deemed more applicable to the current phenomenon under examination due to the novelty of both the prototype system and the concept of using ASCM in the context of researcher literature discovery. Should acceptance be demonstrated using TAM in this study, and if advancements can be made in further development of the prototype system, UTUAT may be more applicable in future research of ASCM acceptance, particularly if such acceptance is studied at an organisational or institutional level.

4 Hypotheses Development

The focus of this study is to determine the researcher's perceived intention to use AASC capability when conducting literature discovery activities. A review of the literature determined there is substantial research in the following domains, all of which relate to AASC and the social situational context of seeking to improve researcher capacity to conduct literature discovery activities:

- AASC/ASCM: for over two decades researchers in the computer science discipline developed NLP models to classify sentences from academic abstracts into literature characteristic or structured abstract heading classes.
- Structured Abstracts: researchers from several disciplines explored the adoption of structured abstracts and the benefits of such adoption in the context of literature discovery.
- Technology Acceptance: the technology acceptance theoretical framework and the TAM have been leveraged extensively in IS literature to explore the acceptance of innovative IS.

ASCM development has matured significantly in the computer science NLP research field, with state-of-the-art models achieving high degrees of performance. The capability of these models, however, has not been deployed into a system to address real world practical problems - such as difficulties encountered when conducting literature discovery, with a specific example being the process of undertaking a literature review of a research area using ASEBDs.

The focus here is to explore the utility and value of the capabilities offered by AASC in the social situational context of researcher literature discovery activities. Technology acceptance will allow the determination of the perceived usefulness, ease of using, attitude towards and intention to use AASC. Noting (a) the performance capabilities of AASC in the computer science domain, (b) the absence of IS deploying AASC technology to address real-world use cases, and (c) the absence of literature examining the utility and value for researchers in the social situational context of conducting literature review discovery activities, the aim here is to expand the body of knowledge through the use of TAM as a theoretical framework, leveraging the prototype information system deploying AASC capability, as a proxy to examine the perceived utility and value of AASC in the social situational context of conducting literature discovery activities. Research questions and corresponding hypotheses are outlined below and are categorised into key components of the theoretical model (figure 1).

Use of Academic Literature Indices

RQ1: Does the experience researchers have in using academic literature indices positively and significantly affect their perceived usefulness of, and ease of using, abstract sentence classification?

H1: A person's experience in using academic literature indices has a positive and significant effect on the perceived usefulness of abstract sentence classification.

H2: A person's experience in using academic literature indices has a positive and significant effect on the perceived ease of using abstract sentence classification.

Research Experience

RQ2: Does the experience researchers have in conducting academic research negatively and significantly affect their perceived usefulness of, and ease of using, abstract sentence classification?

H3: A person's overall experience in conducting academic research has a negative and significant effect on the perceived usefulness of abstract sentence classification.

H4: A person's overall experience in conducting academic research has a negative and significant effect on the perceived ease of using abstract sentence classification.

Classification (Content) Quality

RQ3: Does the quality of sentence classification capability afforded by ASCM positively and significantly affect the perceived usefulness of, and ease of using, abstract sentence classification?

H5: A higher level of perceived content quality has a positive and significant effect on the perceived usefulness of abstract sentence classification.

H6: A higher level of perceived content quality has a positive and significant effect on the perceived ease of using abstract sentence classification.

System Quality

RQ4: Does the technical adequacy of the prototype system serving as a proxy for ASCM positively and significantly affect the perceived usefulness of, and ease of using, abstract sentence classification?

H7: A higher level of perceived technical adequacy has a positive and significant effect on the perceived usefulness of abstract sentence classification.

H8: A higher level of perceived technical adequacy has a positive and significant effect on the perceived ease of using abstract sentence classification.

RQ5: Does the system appearance of the prototype system serving as a proxy for ASCM positively and significantly affect the perceived usefulness of, and ease of using, abstract sentence classification?

H9: A higher level of perceived system appearance has a positive and significant effect on the perceived usefulness of abstract sentence classification.

H10: A higher level of perceived system appearance has a positive and significant effect on the perceived ease of using abstract sentence classification.

RQ6: Do the ancillary functions of the prototype system serving as a proxy for ASCM positively and significantly affect the perceived usefulness of, and ease of using, abstract sentence classification?

H11: A higher level of perceived specific content has a positive and significant effect on the perceived usefulness of abstract sentence classification.

H12: A higher level of perceived specific content has a positive and significant effect on the perceived ease of using abstract sentence classification.

Perceived Usefulness

RQ7: Does the perceived usefulness of abstract sentence classification have a positive and significant effect on the intention to use, and attitude towards, abstract sentence classification?

H13: Perceived usefulness of abstract sentence classification has a positive and significant effect on the attitude towards abstract sentence classification.

H14: Perceived usefulness of abstract sentence classification has a positive and significant effect on the intention to use abstract sentence classification.

Perceived Ease of Use

RQ8: Does the perceived ease of using of abstract sentence classification have a positive and significant effect on the intention to use, and attitude towards, abstract sentence classification?

H15: Perceived ease of using abstract sentence classification has a positive and significant effect on the attitude towards abstract sentence classification.

H16: Perceived ease of using abstract sentence classification has a positive and significant effect on the intention to use abstract sentence classification.

Attitude Towards Academic Abstract Sentence Classification

RQ9: Does the attitude towards abstract sentence classification have a positive and significant effect on the intention to use abstract sentence classification in future literature discovery activities?

H17: The attitude towards abstract sentence classification has a positive and significant effect on the intention to use abstract sentence classification.

5 Approach

The target population of this study is researchers in the IS discipline. A representative sample of the target population was deemed to be authors of IS literature in international IS conferences. This resulted in the desired participants being authors of ICIS, ECIS, PACIS, AMCIS, ACIS and HICSS AIS conference proceedings (research-in-progress or full papers) published between 2019 and 2021. This time frame was specified to ensure active IS researchers were invited to participate. AIS eLibrary was used as the source of literature from these conferences, as it is the AIS's principal literature database, containing literature published across the association's conferences, journals, and other literature sources. To acquire the contact information for authors of literature published at the aforementioned conferences, the following steps were undertaken: (1) download AIS eLibrary paper metadata export .TXT file for each conference; (2) develop a Python script to download each paper's .PDF file from AIS eLibrary; (3) develop a Python script to extract from each paper's .PDF file author email addresses; and (4) export author email addresses for each conference, compatible with a Microsoft Outlook Mail Merge function.

Metadata was downloaded from the AIS eLibrary for a total of 6,172 papers. Most papers originated from HICSS (2,106), followed by AMCIS (1,512) and ICIS (1,152). A smaller number was acquired for PACIS (729), ECIS (474) and ACIS (199). The number of paper metadata was relatively distributed across each year, with approximately 2,000 papers published in each of the acquired years: 2019, 2020 and 2021. A Python script was developed to automate the paper PDF download process. This Python script then scraped author email addresses identified on the title page of each paper. A total of 11,560 unique email addresses were obtained. A survey instrument was developed in Qualtrics using the measurement items identified in Appendix 1. An invitation to participate in the research was then emailed to the 11,560 email addresses. The invitation instructed users, should they wish to participate, to use the prototype prior to completing the survey. The survey was configured to identify the origin of the user, either directly from the email or from the website hosting the prototype. If a user originated directly from the email invitation, they were encouraged to use the prototype prior to completing the survey. To ensure all participants experienced the same functionality of the prototype, ASEBD searching was disabled, and all participants were shown the same sample of 260 AIS conference proceedings which had been pre-processed by the prototype system's ASCM. All ethical aspects of this study were approved by the Macquarie University (Australia) Human Research Ethics Committee.

6 Results

6.1 Prototype Interaction and Survey Completion

Of the 254 respondents who began the survey, 211 (83.07%) provided satisfactory responses. A satisfactory response was the completion of all questions, with some leniency for missing responses which could be resolved using missing value methods, such as imputation and likelihood analysis. Only 43 respondents (16.93%) did not provide satisfactory survey responses. Use of the prototype system was monitored using Mixpanel (https://mixpanel.com), a cloud-based product analytics solution. Mixpanel event logs captured 5,354 events from 745 distinct users. Per user, there was a minimum event count of 1; a maximum count of 184 and an average (mean) of 7. Of the 745 distinct users, 723 (97.05% of all distinct users) only used the system on one day, leaving 22 users (2.95% of all distinct users) who accessed the system more than once - of which 15 (2.01% of all distinct users) visited on two days.

There were 47 distinct countries of origin identified from analysis of the Mixpanel event logs. The most popular country of origin for users was the USA, with 952 events (17.78% of all event occurrences), by 173 distinct users (23.22% of all distinct users). Australia closely followed, identified for 861 events (16.08% of all event occurrences), with 84 distinct users (11.28% of all distinct users). Combined, the USA and Australia were identified as origin countries for over a third of all events (33.86%) and distinct users (38.93%). There were 3 other countries identified with greater than 4% of total events: Germany, identified in 409 events (7.64% of all event occurrences), by 34 distinct users (4.56% of all distinct users); Switzerland, identified in 224 events (4.18% of all event occurrences), by 22 distinct users (2.95% of all distinct users) and Canada, identified in 221 events (4.13% of all event occurrences), by 44 distinct users (5.91% of all distinct users). The remaining 42 countries were identified for 2,077 events (38.79% of all event occurrences), by 290 distinct users (38.93% of all distinct users. Of the 254 respondents who commenced the survey, 211 (83.07%) provided satisfactory responses, meaning the completion of all survey questions, with some leniency for missing question responses which could be resolved using missing value methods, such as imputation and likelihood analysis. Only 43 respondents (16.93%) did not provide satisfactory responses to the survey. At the time of the survey, 65 of the 211 respondents (30.81% of all respondents) had completed between 0 and 5 years of academic research; 44 (20.85% of all respondents) completed between 6 and 10 years of research, 29 (13.74% of all respondents) had completed between 11 and 15 years of research and 72 (34.12% of all respondents) had completed more than 15 years of academic research.

These results indicate respondents had a variety of years of research experience. Interestingly, both the 15+ years (34.12% of all respondents) and 0-5 years (30.81% of all respondents) categories, represented almost 65% of all responses, suggesting the sample represents persons at both the initial and later stages of long research careers. Of the 211 respondents, 179 respondents (84.83%) were employed by an educational institution; 14 respondents (6.64%) were not employed, 8 respondents (3.79%) were employed in private industry and 5 respondents (2.37%) were employed in government. There were 2 respondents (0.95%) who selected 'Other' and 3 respondents (1.42%) who did not answer the question. These results indicate a lack of variety in respondent's field of employment; however, the results are unsurprising considering the participants were authors of literature published in the proceedings of IS conferences, who anecdotally tend to be affiliated with educational institutions.

6.2 Model Evaluation

Structural and measurement model evaluation was performed using partial least squares structural equational modelling (PLS-SEM), with Smart PLS 4 software (Ringle et al., 2022) used to conduct PLS-SEM analysis. PLS-SEM is a commonly used measurement model evaluation in IS research, as is the use of Smart PLS software to conduct PLS-SEM analysis (e.g., Jabagi et al., 2021; Kim and Lee, 2021; Zamir and Kim, 2021; Mo and Yang, 2021; Andijani and Kang, 2022; Zhang et al., 2022; Jha and Kathuria, 2022). A Smart PLS 4 PLS algorithm was executed using the following configuration parameters: mean replacement missing data algorithm; 3000 maximum iterations; a stop criteria of 10^{-7}; standardized result type output; and path weighting scheme. Initial assessment of outer loadings and collinearity statistics variance inflation factor (VIF) was performed to determine measurement item suitability. Of the original 48 measurement items, 8 were withdrawn due to VIF values above the threshold of 5 (Ringle et al., 2022). Measurement items withdrawn are identified in appendix 1. The PLS algorithm was then re-executed using the same parameters identified.

Constructs were then examined for internal consistency reliability. Composite reliability (rho_c) was evaluated, with all construct rho_c values exceeding the minimum acceptable threshold of 0.70 (Jöreskog, 1971; Hair et al., 2022). Cronbach's alpha was also evaluated, again with all construct Cronbach's alpha values exceeding the acceptable threshold of 0.70 (Hair et al., 2022). Convergent validity was then assessed through examination of the average variance extracted (AVE), which represents the communality of a construct (Hair et al., 2022, p. 78). All constructs featured AVE values exceeding the minimum acceptable threshold of 0.50 (Hair et al., 2022). Internal consistency reliability and convergent validity statistics are shown in table 2. Discriminant validity was evaluated using the Fornell-Larcker criterion (Fornell and Larcker, 1981), whereby discriminant validity between latent variables is demonstrated if the square root of each construct's AVE exceeds any other inter-construct correlation co-efficients; this was achieved, as shown in the Fornell-Larcker criterion matrix in table 2.

With minimum thresholds met for consistency reliability, convergent validity and discriminant validity of the measurement model, the structural model was then evaluated. Firstly, collinearity issues were ruled out through examination of the inner model VIF values, with all construct VIF values below the maximum threshold of 5 (Hair et al., 2022). The bootstrapping algorithm in Smart PLS 4 was then executed with the following configuration parameters: complete complexity; percentile bootstrap confidence interval method; 10,000 samples; fixed seed; two tailed test type with significance level of 0.05. Through examination of the bootstrapping results structural (inner) model path relationship coefficients and their significance were evaluated, as well as R^2 statistics as measures of both the model explanatory and predictive powers. Of the 17 construct path relationships and corresponding hypotheses, 10 were accepted – featuring significant positive path coefficients. Table 1 shows the structural model path coefficients, significance statistics and hypotheses results in relation to the construct relationship paths. The bootstrapping results are graphically shown in figure 1.

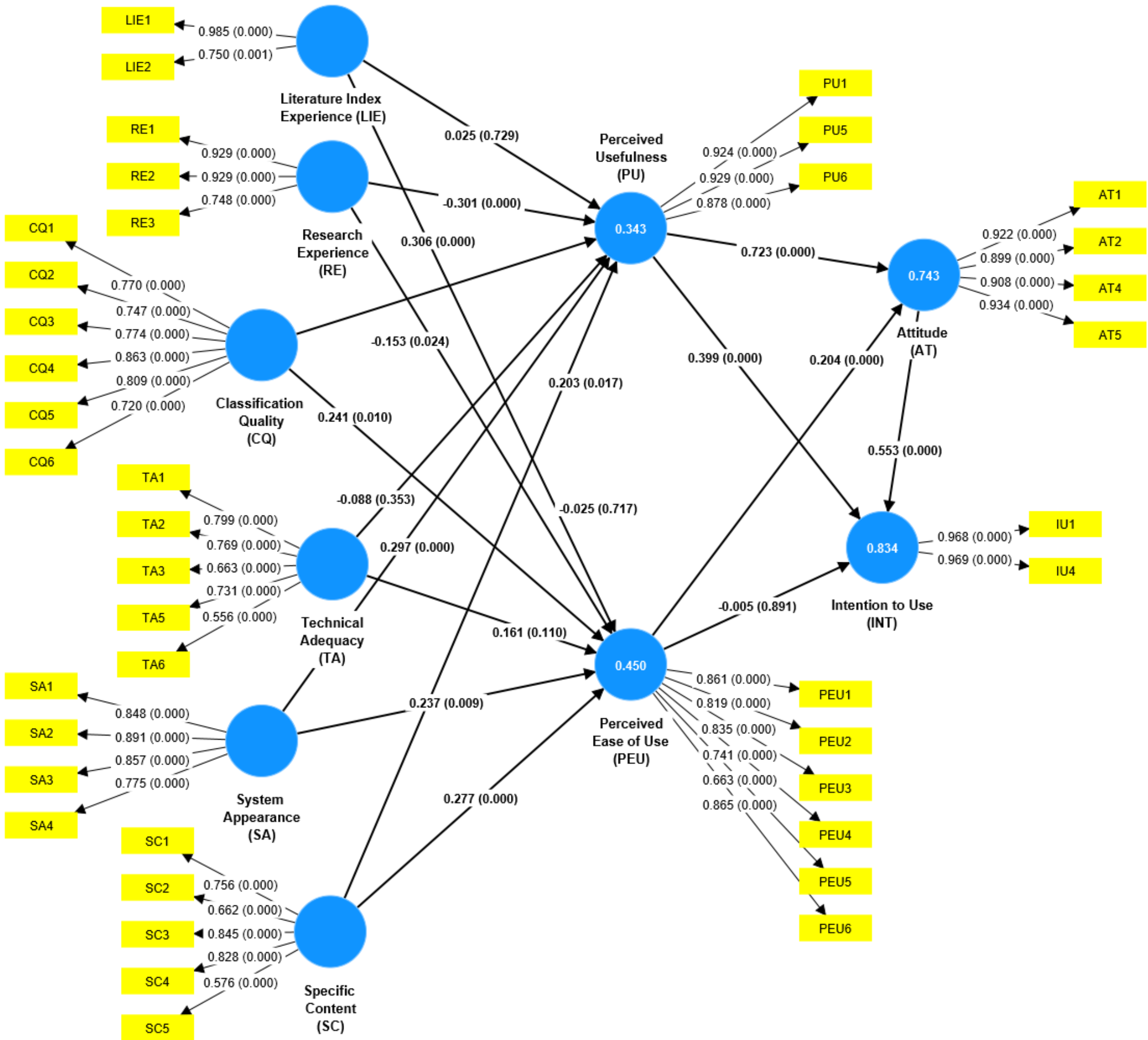

Figure 1: Measurement model PLS-SEM bootstrapping graphical output

Table 1: Path coefficients β, T statistics, P values and hypotheses results (<= 0.05, ** <= 0.01, *** <= 0.001)*

Hypothesis	Path	Path coefficient β	T statistic	P value	Result
H1	LIE -> PU	0.025	0.346	0.729	Reject
H2	LIE -> PEU	-0.025	0.363	0.717	Reject
H3	RE -> PU	-0.301***	4.16	0.000	**Accept**
H4	RE -> PEU	-0.153*	2.257	0.024	**Accept**
H5	CQ -> PU	0.306***	3.583	0.000	**Accept**
H6	CQ -> PEU	0.241**	2.578	0.010	**Accept**
H7	TA -> PU	-0.088	0.928	0.353	Reject
H8	TA -> PEU	0.161	1.597	0.110	Reject
H9	SA -> PU	0.297***	4.395	0.000	**Accept**
H10	SA -> PEU	0.237**	2.599	0.009	**Accept**
H11	SC -> PU	0.203*	2.379	0.017	**Accept**
H12	SC -> PEU	0.277***	3.655	0.000	**Accept**
H13	PU -> AT	0.723***	18.451	0.000	**Accept**
H14	PU -> INT	0.399***	6.723	0.000	**Accept**
H15	PEU -> AT	0.204***	4.06	0.000	**Accept**
H16	PEU -> INT	-0.005	0.137	0.891	Reject
H17	AT -> INT	0.553***	9.151	0.000	**Accept**

Table 2: Construct Level (inner model latent variables) Cronbach's Alpha, Composite Reliability (rho_c), AVE and Discriminant validity (Fornell-Larcker criterion)

Construct	*CA*	*rho_c*	*AVE*	AT	CQ	INT	LIE	PEU	PU	RE	SC	SA	TA
Attitude (AT)	0.936	0.954	0.839	**0.916**									
Classification Quality (CQ)	0.876	0.904	0.611	0.440	**0.782**								
Intention to Use (INT)	0.933	0.968	0.938	0.888	0.382	**0.968**							
Literature Index Experience (LIE)	0.769	0.866	0.766	-0.014	0.169	-0.045	**0.875**						
Perceived Ease of Use (PEU)	0.887	0.914	0.641	0.641	0.495	0.591	0.042	**0.801**					
Perceived Usefulness (PU)	0.897	0.936	0.830	0.846	0.393	0.865	-0.082	0.604	**0.911**				
Research Experience (RE)	0.842	0.905	0.762	-0.101	0.197	-0.135	0.619	-0.037	-0.190	**0.873**			
Specific Content (SC)	0.791	0.856	0.548	0.390	0.490	0.333	0.270	0.526	0.344	0.164	**0.740**		
System Appearance (SA)	0.866	0.908	0.712	0.421	0.296	0.367	-0.004	0.430	0.403	0.054	0.289	**0.844**	
Technical Adequacy (TA)	0.752	0.833	0.503	0.262	0.518	0.203	0.291	0.490	0.242	0.155	0.586	0.310	**0.709**

7 Discussion and Future Research

This study aimed to explore the acceptance of ASCM technology by IS researchers in the context of the conduct of academic literature reviews. Through the development of a novel prototype deploying ASCM capability into a pseudo ASEBD interface, the prototype system emulated the experience a researcher would have using an ASEBD with ASCM deployed to enhance user experience, in practice through the classification of abstract sentences according to literature characteristics. TAM was used as a theoretical framework and a model was developed to examine the perceived usefulness and ease of using ASCM, attitude towards using ASCM and intention to use ASCM. Survey responses from 211 international IS researchers were used to evaluate the measurement and structural models using PLS-SEM statistical analysis, which resulted in 12 of the 17 proposed hypotheses being accepted.

In relation to RQ 1, researcher experience in using academic literature indices was not demonstrated to have a significant effect on both perceived usefulness or perceived ease of using abstract sentence classification (P-values of 0.729 and 0.717 for LIE to PU and PEU respectively). Literature index experience had a non-significant weak positive path coefficient (PC) to PU (PC: 0.025) and a significant negative path coefficient to PEU (PC: -0.025). Both hypotheses H1 and H2 were rejected accordingly. This result suggests that the extent of researchers' prior total experience and regularity of using academic literature indices is not an indicator of the perceived usefulness or ease of using ASCM. A different finding was made for RQ 2, concerned with the degree of participant experience in conducting academic research. H3 and H4 posited that the greater experience a participant has in conducting academic research the higher likelihood for a negative and significant relationship to be observed with PU and PEU. This was demonstrated to be the case with research experience featuring a moderate significant negative relationship with PU (PC: -0.301***) and a weak significant negative relationship with PEU (PC: -0.153*). Measurement items for both literature index experience and research experience indicated participants in general had high levels of experience in literature index usage and the conduct of academic research in general, which suggests that participants with both low and high levels of relevant experience had varying responses concerning PU and PEU.

RQ 3 assessed if the performance of the ASCM technology deployed in the prototype in performing classification of sentences as indicative of literature characteristics had a positive and significant relationship on PU (H5) and PEU (H6). Evaluation of the structural model indicates that sentence classification quality does have a positive and significant relationship on both PU (PC: 0.306***) and PEU (PC: 0.241**). These results indicate that the core functionality of the prototype system, the classification of sentences as indicative of literature characteristics, positively and significantly effects the participant's perception that ASCM is useful and easy to use, in the context of them performing academic literature discovery. This is a key finding, suggesting that ASCM positively effects IS researcher's perception of the technology's usefulness in improving literature discovery capability and ease of using academic literature indices featuring ASCM as a user experience enhancement feature.

RQ 4 explored if the technical adequacy of the prototype system positively and significantly effected PU (H7) and PEU (H8) ASCM. Structural model evaluation did not support RQ 4 and resulted in both H7 and H8 being rejected. Technical adequacy of the prototype system had a particularly weak negative relationship with PU and a weak positive effect on PEU. Both relationships, however, were insignificant. Measurement items used to examine technical adequacy assessed the functionality of the prototype system separate to the core ASCM functionality, such as navigation through different sections of the website, the domain name, and the operation of buttons. The rejection of H7 and H8 implies that non-core functionality of the website had no bearing on IS researcher perceptions of the usefulness and ease of using ASCM, however, suggests that further improvements may be required should another version of the prototype by developed.

RQs 5 and 6 addressed additional non-core functionality aspects of the prototype system including its visual appearance and additional supplementary information provided on the website such as instructions on how to use the system, ability to navigate throughout the system, the system's terms of use and contact information of the webmaster. Hypotheses H9 and H10 posited that a higher level of user perceived quality of the system's appearance would have a positive and significant effect on PU and PEU respectively. This was demonstrated to be the case with both H9 and H10 being accepted. System appearance featured a significant moderate positive effect on PU (PC: 0.297***) and PEU (PC: 0.237 **). Specific content had a weak positive effect on both perceived usefulness (PC: 0.203*) and a moderate positive effect on ease of use (PC: 0.277***). This finding suggests that the way in which ASCM is presented in the prototype through a user enhancement to a typical academic literature interface is an appropriate medium to deploy ASCM.

The effect of PU and PEU on both attitude towards ASCM and intention to use ASCM was the focus of RQs 7 and 8. H13 posited that perceived usefulness would have a positive and significant effect on attitude towards ASCM. This hypothesis was accepted as perceived usefulness was shown to have a significant substantial positive effect on attitude towards ASCM (PC: 0.723***). H14 was concerned with the effect of perceived usefulness on intention to use ASCM, positing that a high degree of PU would result in greater intention to use ASCM. H14 was accepted as perceived usefulness was demonstrated to feature a significant positive effect on intention to use ASCM (PC: 0.399***). PEU was demonstrated to have a positive and significant effect on attitude towards ASCM in H15 (PC: 0.204***), however, the posited significant positive effect of perceived ease of use on intention to use ASCM in H16 was rejected as it was instead shown to be a weak negative and non-significant (PC: 0.005). The mixed findings for RQs 7 and 8 suggests that whilst researchers found the ASCM capability demonstrated by the prototype system easy to use (with indicator measurement mean values ranging between 4.8 and 5.9) it isn't the ease of using the prototype system that is a primary driver behind participant attitude towards ASCM or their intention to use ASCM, it is instead their perception of the usefulness of ASCM in the context of the literature acquisition activity. This is a core finding of the study, demonstrating that ASCM is perceived to be a useful technology and its usefulness effects an intention by IS researchers to use ASCM in the future.

Finally, RQ 9 investigated the relationship between attitude towards ASCM and intention to use ASCM. Hypothesis 17 posited that a more positive attitude towards ASCM would have a positive and significant effect on intention to use ASCM. H17 was accepted with attitude towards ASCM was demonstrated to have a strong positive and significant effect on intention to use ASCM (PC: 0.553***). This outcome again underscores the core finding of this study that IS researchers have a positive attitude towards the role ASCM can play in introducing efficiencies in the literature discovery workflow, and this positive attitude influences an intention by IS researchers to utilise ASCM in the future should it be made available. The findings of this study, particularly those demonstrated through an assessment of RQs 7, 8 and 9, motivates ongoing research we are conducting into how to ASCM can be further incorporated into researcher activities and introduce efficiencies in literature discovery.

By demonstrating that ASCM is perceived as useful by IS researchers, that there is a positive attitude towards the technology and they intend to use the technology should it be available in the future, this study warrants further research into ASCM technology and how it can help alleviate literature overload issues faced by researchers. The results may indicate acceptance of ASCM by cross disciplinary researchers in general, however, due to the unique characteristics of literature discovery methods across disciplines, further research into cross disciplinary acceptance of ASCM is encouraged.

8 References

AGUILLO, I. F. 2012. Is Google Scholar useful for bibliometrics? A webometric analysis. Scientometrics, 91, 343-351.

ALADWANI, A. M. & PALVIA, P. C. 2002. Developing and validating an instrument for measuring user-perceived web quality. Information Management, 39, 467-476.

ANDIJANI, A. & KANG, K. 2022. The determinants of the use of social commerce in Saudi Arabia across different age groups. PACIS 2022.

BAAH-PEPRAH, P. 2023. Explaining reward crowdfunding backers' intentions and behavior. Baltic Journal of Management, 18, 262-281.

BEWLEY, W. L., ROBERTS, T. L., SCHROIT, D. & VERPLANK, W. L. 1983. Human factors testing in the design of Xerox's 8010 “Star” office workstation. Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 1983. 72-77.

BLIEMEL, M. & HASSANEIN, K. 2006. An empirical study of consumer satisfaction with online health information retrieval. Special Interest Group on Human-Computer Interaction 2006.

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, 435-468.

CARD, S. K., ENGLISH, W. K. & BURR, B. J. 1978. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. Ergonomics, 21, 601-613.

CHAWLA, D. & JOSHI, H. 2020. Segmenting and Profiling Mobile Wallet Users in India. ICIS 2020.

CHUNG, G. Y. 2009. Sentence retrieval for abstracts of randomized controlled trials. BMC Medical Informatics Decision Making, 9, 10.

DAGLIYAN, G., GRANADOS, N. F. & RUSSELL, C. A. 2022. Adopting AI-Enabled Technology: Taking the Bad with the Good. AMCIS 2022.

DAVID, D. & ZU, X. 2023. Examining Underrepresented Communities' Intention Towards Digital Entrepreneurship: A Dual-Theory Framework. Americas Conference on Information Systems: AMCIS 2023.

DAVIS, F. D. 1985. A technology acceptance model for empirically testing new end-user information systems: Theory and results. Massachusetts Institute of Technology.

DAVIS, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 319-340.

DAVIS, F. D., BAGOZZI, R. P. & WARSHAW, P. R. 1989. User acceptance of computer technology: A comparison of two theoretical models. Management science, 35, 982-1003.

DERNONCOURT, F., LEE, J. Y. & SZOLOVITS, P. 2016. Neural Networks for Joint Sentence Classification in Medical Paper Abstracts. arXiv preprint arXiv:1612.05251.

DOEMELAND, D. & TREVINO, J. 2014. Which World Bank reports are widely read? World Bank Policy Research Working Paper.

FARRELL, M. J., BRIERLEY, L., WILLOUGHBY, A., YATES, A. & MIDEO, N. 2022. Past and future uses of text mining in ecology and evolution. Proceedings of the Royal Society B, 289, 20212721.

FISHER, B., BALMFORD, A., FERRARO, P. J., GLEW, L., MASCIA, M., NAIDOO, R. & RICKETTS, T. H. 2014. Moving Rio forward and avoiding 10 more years with little evidence for effective conservation policy. Conservation Biology, 28, 880-882.

FRISONI, G., MIZUTANI, M., MORO, G. & VALGIMIGLI, L. 2022. BioReader: a Retrieval-Enhanced Text-to-Text Transformer for Biomedical Literature. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022. 5770-5793.

GALLI, C., COLANGELO, M. T. & GUIZZARDI, S. 2020. Striving for modernity: Layout and abstracts in the biomedical literature. Publications, 8, 38.

GOEURIOT, L., JONES, G. J., KELLY, L., MÜLLER, H. & ZOBEL, J. 2016. Medical information retrieval: introduction to the special issue. Information Retrieval Journal, 19, 1-5.

GONÇALVES, S., CORTEZ, P. & MORO, S. 2018. A Deep Learning Approach for Sentence Classification of Scientific Abstracts. International Conference on Artificial Neural Networks, 2018. Springer, 479-488.

GOULD, J. D., CONTI, J. & HOVANYECZ, T. 1983. Composing letters with a simulated listening typewriter. Communications of the ACM, 26, 295-308.

GUIMARÃES, C. A. 2006. Structured abstracts: narrative review. Acta cirurgica brasileira, 21, 263-268.

GUSENBAUER, M. 2019. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. Scientometrics, 118, 177-214.

HAINES, L. L., LIGHT, J., O'MALLEY, D. & DELWICHE, F. A. 2010. Information-seeking behavior of basic science researchers: implications for library services. Journal of the Medical Library Association: JMLA, 98, 73.

HAIR JR, J. F., HULT, G. T. M., RINGLE, C. M., SARSTEDT, M., DANKS, N. P. & RAY, S. 2021. Partial least squares structural equation modeling (PLS-SEM) using R: A workbook, Springer Nature.

HARTLEY, J. 2008. Academic Writing and Publishing, London, Routledge.

HASSANZADEH, H., GROZA, T. & HUNTER, J. 2014. Identifying scientific artefacts in biomedical literature: The Evidence Based Medicine use case. Journal of Biomedical Informatics, 49, 159-170.

HIROHATA, K., OKAZAKI, N., ANANIADOU, S. & ISHIZUKA, M. 2008. Identifying sections in scientific abstracts using conditional random fields. Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I, 2008.

HOBERT, S. & MEYER VON WOLFF, R. 2019. Say hello to your new automated tutor–a structured literature review on pedagogical conversational agents. WIRTSCHAFTSINFORMATIK 2019.

ITO, T., SHIMBO, M., YAMASAKI, T. & MATSUMOTO, Y. 2004. Semi-supervised sentence classification for medline documents. Methods, 138, 141-146.

JABAGI, N., CROTEAU, A.-M., AUDEBRAND, L. & MARSAN, J. 2021. Who's the Boss? Measuring Gig-Workers' Perceived Algorithmic Autonomy-Support. ICIS 2021.

JHA, S. & KATHURIA, A. 2022. Size Matters for Cloud Capability and Performance. AMCIS 2022.

JIN, D. & SZOLOVITS, P. 2018. Hierarchical Neural Networks for Sequential Sentence Classification in Medical Scientific Abstracts. arXiv preprint arXiv:.06161.

JÖRESKOG, K. G. 1971. Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

JOSHI, H. & SONDHI, N. 2023. Integrating Technology Readiness, Learning Goal Orientation with TAM to explain E-learning adoption. Americas Conference on Information Systems: AMCIS 2023.

KHABSA, M. & GILES, C. L. 2014. The number of scholarly documents on the public web. PloS one, 9, e93949.

KIM, S. N., MARTINEZ, D., CAVEDON, L. & YENCKEN, L. 2011. Automatic classification of sentences to support evidence based medicine. BMC bioinformatics, 2011. BioMed Central, S5.

KIM, S.-S. & LEE, H.-Y. 2021. Research on the Influential Indicators and Measurements of Mediating and Moderating Roles on the Performance of SMEs by Industrial Sectors. Proceedings of the 54th Hawaii International Conference on System Sciences: HICSS 2021

KITCHENHAM, B. A., BRERETON, O. P., OWEN, S., BUTCHER, J. & JEFFERIES, C. 2008. Length and readability of structured software engineering abstracts. IET Software, 2, 37-45.

LIU, Y., WU, F., LIU, M. & LIU, B. 2013. Abstract sentence classification for scientific papers based on transductive SVM. Computer Information Science, 6, 125.

MARANGUNIĆ, N. & GRANIĆ, A. 2015. Technology acceptance model: a literature review from 1986 to 2013. Universal access in the information society, 14, 81-95.

MAYR, P. & WALTER, A. K. 2007. An exploratory study of Google Scholar. Online information review, 31, 814-830.

MCKINNON, M. C., CHENG, S. H., GARSIDE, R., MASUDA, Y. J. & MILLER, D. C. 2015. Sustainability: Map the evidence. Nature, 528, 185-187.

MCKNIGHT, L. & SRINIVASAN, P. Categorization of sentence types in medical abstracts. AMIA Annual Symposium Proceedings, 2003. American Medical Informatics Association, 440.

MO, Y. & WANG, Q. 2021. Exploring the Influence of Live Streaming in Social Commerce on Impulse Buying from a Affordance Perspective. Wuhan International Conference on E-Business (WHICEB 2021).

MOSTELLER, F., NAVE, B. & MIECH, E. J. 2004. Why we need a structured abstract in education research. Educational Researcher, 33, 29-34.

NAM, S., KIM, S.-K., KIM, H.-G., NGO, V. & ZONG, N. 2016. Structuralizing biomedical abstracts with discriminative linguistic features. Computers in Biology Medicine, 79, 276-285.

ORDUÑA-MALEA, E., AYLLÓN, J. M., MARTÍN-MARTÍN, A. & DELGADO LÓPEZ-CÓZAR, E. 2015. Methods for estimating the size of Google Scholar. Scientometrics, 104, 931-949.

ORDUÑA-MALEA, E., AYLLÓN, J. M., MARTÍN-MARTÍN, A. & LÓPEZ-CÓZAR, E. D. 2014. About the size of Google Scholar: playing the numbers. arXiv preprint arXiv:.05251.

RINGLE, CHRISTIAN M., WENDE, SVEN, & BECKER, JAN-MICHAEL. 2022. SmartPLS 4. Oststeinbek: SmartPLS. Retrieved from https://www.smartpls.com

SCHULZ, K. F., ALTMAN, D. G. & MOHER, D. 2011. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. Annals of Internal Medicine, 154, 291-292.

SHARMA, S. & HARRISON, J. E. 2006. Structured abstracts: do they improve the quality of information in abstracts? American Journal of Orthodontics Dentofacial Orthopedics, 130, 523-530.

SHIMBO, M., YAMASAKI, T. & MATSUMOTO, Y. 2003. Using sectioning information for text retrieval: a case study with the medline abstracts. Proceedings of Second International Workshop on Active Mining (AM'03).

STEAD, C., SMITH, S., BUSCH, P., VATANASAKDAKUL, S. & PANG, V. 2020a. An Activity Theory Investigation of Academic Abstract Sentence Classification. 24th Pacific Asia Conference on Information Systems: PACIS 2020.

STEAD, C., SMITH, S., BUSCH, P. & VATANASAKDAKUL, S. 2020b. Towards an Academic Abstract Sentence Classification System. The 14th International Conference on Research Challenges in Information Science: RCIS 2020.

STEAD, C., SMITH, S., BUSCH, P. & VATANASAKDAKUL, S. 2021a. A Taxonomy of Abstract Sentence Classification Modelling. 18th conference of the Italian Chapter of AIS (Association for Information Systems) 2021: itAIS 2021.

STEAD, C., SMITH, S., BUSCH, P. & VATANASAKDAKUL, S. 2021b. Intention to Use Abstract Sentence Classification Technology. Australasian Conference on Information Systems 2021: ACIS 2021.

VAN NOORDEN, R. 2014. Online collaboration: Scientists and the social network. Nature news, 512, 126.

VAUGHAN, L. & THELWALL, M. 2004. Search engine coverage bias: evidence and possible causes. Information Processing, 40, 693-707.

VENKATESH, V., MORRIS, M. G., DAVIS, G. B. & DAVIS, F. D. 2003. User acceptance of information technology: Toward a unified view. MIS quarterly, 425-478.

WEIDHAAS, R., SCHLÖGL, S., HALTTUNEN, V. & SPIEß, T. 2021. Watch this! The influence of recommender systems and social factors on the content choices of streaming video on demand consumers. Innovation Through Information Systems: Volume II: A Collection of Latest Research on Technology Issues, 2021. Springer, 738-753.

WESTGATE, M. J., HADDAWAY, N. R., CHENG, S. H., MCINTOSH, E. J., MARSHALL, C. & LINDENMAYER, D. B. 2018. Software support for environmental evidence synthesis. Nature Ecology & Evolution, 2, 588-590.

YAMAMOTO, Y. & TAKAGI, T. 2005. A sentence classification system for multi biomedical literature summarization. 21st International Conference on Data Engineering Workshops (ICDEW'05), 2005. IEEE, 1163-1163.

YANG, Z., DAI, Z., YANG, Y., CARBONELL, J., SALAKHUTDINOV, R. & LE, Q. V. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding. arXiv preprint arXiv:.08237.

YOUSAFZAI, S. Y., FOXALL, G. R. & PALLISTER, J. G. 2007. Technology acceptance: a meta-analysis of the TAM: Part 1. Journal of Modelling in Management, 2, 251-280.

ZAMIR, Z. B. & KIM, D. D. 2021. The Impact of Quality Dimensions on Knowledge Sharing and User Satisfaction. Americas Conference on Information Systems: AMCIS 2021.

ZHANG, F., PAN, Z., LU, Y. & HU, Q. 2022. The Dark Side of Artificial Autonomy: Formation and Mitigation of AI Technostress in Smart Voice Assistants (SVAs). Pacific Asia Conference on Information Systems: PACIS 2022.

ZHANG, L., TULU, B., DJAMASBI, S., SANKAR, G. & MUEHLSCHLEGEL, S. 2023. Investigating User Satisfaction: An Adaptation of IS Success Model for Short-term Use. Hawaii International Conference on System Sciences: HICSS 2023.

Appendix 1 - Measurement Items

ID	Construct	Measurement Item	Mean	VIF
LIE1	Literature Index Experience	I have a high level of experience in using academic literature indices - such as Google Scholar, Microsoft Academic and Scopus.	6.019	1.732
LIE2	Literature Index Experience	I regularly use academic literature indices - such as Google Scholar, Microsoft Academic and Scopus.	6.109	2.079
LIE3	Literature Index Experience	Academic literature indices - such as Google Scholar, Microsoft Academic and Scopus - are key tools in my research.	Withdrawn	
RE1	Research Experience	I have a high level of experience in conducting academic research.	5.791	3.407
RE2	Research Experience	I have been conducting academic research for a long time.	5.412	3.234
RE3	Research Experience	A significant part of my career is to conduct academic research.	5.957	1.493
CQ1	Classification Quality	The classified literature characteristics (purpose, method, findings…) within the abstract sentence classification website are useful.	5.581	2.129
CQ2	Classification Quality	The classified literature characteristics (purpose, method, findings…) within the abstract sentence classification website are complete.	5.09	2.181
CQ3	Classification Quality	The classified literature characteristics (purpose, method, findings…) within the abstract sentence classification website are clear.	5.448	2.459
CQ4	Classification Quality	The classified literature characteristics (purpose, method, findings…) within the abstract sentence classification website are informative.	5.436	3.000
CQ5	Classification Quality	The classified literature characteristics (purpose, method, findings…) within the abstract sentence classification website are concise.	5.36	2.284
CQ6	Classification Quality	The classified literature characteristics (purpose, method, findings…) within the abstract sentence classification website are accurate.	5.118	1.775
TA1	Technical Adequacy	The academic abstract sentence classification website links and buttons work as expected.	5.839	2.257
TA2	Technical Adequacy	The academic abstract sentence classification website is easy to navigate through.	5.787	1.961

ID	Construct	Measurement Item	Mean	VIF
TA3	Technical Adequacy	The academic abstract sentence classification website pages load quickly.	5.967	1.937
TA4	Technical Adequacy	The domain: www.abstractsentenceclassification.com is a suitable domain name and is easy to access online.	Withdrawn	
TA5	Technical Adequacy	The academic abstract sentence classification website is interactive.	5.257	1.476
TA6	Technical Adequacy	The academic abstract sentence classification website can be customised to suit my needs.	4.687	1.232
SC1	Specific Content	In the academic abstract sentence classification website, I can find information about its purpose.	5.573	1.404
SC2	Specific Content	In the academic abstract sentence classification website, I can find instructions and guidance on how to use the website.	5.379	1.320
SC3	Specific Content	In the academic abstract sentence classification website, I can find the terms and conditions.	5.673	4.356
SC4	Specific Content	In the academic abstract sentence classification website, I can find the privacy information.	5.571	4.150
SC5	Specific Content	In the academic abstract sentence classification website, I can find contact details for the webmaster.	4.882	1.332
SA1	System Appearance	The abstract sentence classification website looks attractive.	4.519	2.345
SA2	System Appearance	The abstract sentence classification website looks organised.	5.185	2.583
SA3	System Appearance	The abstract sentence classification website uses fonts properly.	5.412	2.385
SA4	System Appearance	The abstract sentence classification website uses colours properly.	5.019	1.914
PU1	Perceived Usefulness	The functionality of abstract sentence classification is useful to me.	5.166	3.127
PU2	Perceived Usefulness	I consider the functions of abstract sentence classification useful.	Withdrawn	
PU3	Perceived Usefulness	Overall, abstract sentence classification is useful.	Withdrawn	
PU4	Perceived Usefulness	Using abstract sentence classification would help me to conduct my research more productively.	Withdrawn	
PU5	Perceived Usefulness	Using abstract sentence classification would introduce efficiencies to my research.	4.948	3.206
PU6	Perceived Usefulness	Using abstract sentence classification would improve my ability to discover literature relevant to my research.	4.852	2.303
PEU1	Perceived Ease of Use	Learning to use abstract sentence classification is easy.	5.81	4.047
PEU2	Perceived Ease of Use	It is easy to review abstracts using abstract sentence classification.	5.545	2.181
PEU3	Perceived Ease of Use	Remembering how to use abstract sentence classification is not hard.	5.919	3.728
PEU4	Perceived Ease of Use	Interacting with abstract sentence classification does not require a lot of mental effort.	5.706	1.831
PEU5	Perceived Ease of Use	Using abstract sentence classification is a convenient way to find relevant literature.	4.815	1.474
PEU6	Perceived Ease of Use	Overall, I find abstract sentence classification easy to use.	5.701	3.112

ID	Construct	Measurement Item	Mean	VIF
AT1	Attitude	Using abstract sentence classification is a good idea.	5.355	3.841
AT2	Attitude	I enjoy using abstract sentence classification.	4.896	3.032
AT3	Attitude	I am positive towards abstract sentence classification.	Withdrawn	
AT4	Attitude	Other researchers should use abstract sentence classification.	4.953	3.592
AT5	Attitude	Using abstract sentence classification makes sense.	5.284	4.608
IU1	Intention to Use	Assuming I had access to abstract sentence classification, I intend to use it.	5.009	4.27
IU2	Intention to Use	If I had access to abstract sentence classification, I intend to use it regularly.	Withdrawn	
IU3	Intention to Use	Given that I have access to abstract sentence classification, I would use it as part of my research.	Withdrawn	
IU4	Intention to Use	I would recommend other researchers to use abstract sentence classification.	4.943	4.27

Acknowledgements

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