



# Classifying the ideational impact of Information Systems review articles: A content-enriched deep learning approach

Julian Prester<sup>a,\*</sup>, Gerit Wagner<sup>b,c</sup>, Guido Schryen<sup>d</sup>, Nik Rushdi Hassan<sup>e</sup>

<sup>a</sup> University of New South Wales, Sydney, Australia

<sup>b</sup> HEC Montreal, Canada

<sup>c</sup> University of Regensburg, Germany

<sup>d</sup> Paderborn University, Germany

<sup>e</sup> University of Minnesota Duluth, Duluth, USA

## ARTICLE INFO

### Keywords:

Ideational impact  
Citation classification  
Academic recommender systems  
Natural language processing  
Deep learning  
Cumulative tradition

## ABSTRACT

Ideational impact refers to the uptake of a paper's ideas and concepts by subsequent research. It is defined in stark contrast to total citation impact, a measure predominantly used in research evaluation that assumes that all citations are equal. Understanding ideational impact is critical for evaluating research impact and understanding how scientific disciplines build a cumulative tradition. Research has only recently developed automated citation classification techniques to distinguish between different types of citations and generally does not emphasize the conceptual content of the citations and its ideational impact. To address this problem, we develop Deep Content-enriched Ideational Impact Classification (Deep-CENIC) as the first automated approach for ideational impact classification to support researchers' literature search practices. We evaluate Deep-CENIC on 1256 papers citing 24 information systems review articles from the IT business value domain. We show that Deep-CENIC significantly outperforms state-of-the-art benchmark models. We contribute to information systems research by operationalizing the concept of ideational impact, designing a recommender system for academic papers based on deep learning techniques, and empirically exploring the ideational impact in the IT business value domain.

## 1. Introduction

Evaluating citing behavior is a complex and multidimensional task in any discipline. It is concerned with evaluating the degree to which researchers build on each other's work and develop a set of shared definitions, concepts, and theories [46]. Because citations serve as symbols for how a citing article (CA) applies research results, ideas, or concepts from other publications, citation counts are often used in research evaluation as an indicator of research impact [5,15]. However, research evaluation based on citation counts alone does not account for the conceptual content of the citations, how relevant that content is to the CA, and how much the cited content is taken up by the field to build its cumulative tradition [4,31,47]. In other words, research evaluation is in need of an academic recommender system that takes into account the conceptual content of what is cited, recommends research that is ideationally relevant, and suggests citations that contribute to knowledge building. Such a dimension of citations is proposed by Truex, Cuellar,

Takeda and Vidgen [61] and Hassan and Loebbecke [22] who refer to it as the ideational dimension and define ideational impact as the uptake of a researcher's ideas and concepts by subsequent research.

Our work is related to a broader line of research surrounding how citation counts are traditionally used to measure the cited article's importance and significance [4,48]. Most of this research is based on the assumption that all citations can be considered of equal importance [25,62]. However, many studies have shown that the bulk of citations are perfunctory to the CA's main contribution [24,47] and therefore do not use or apply the concepts proposed by the cited works [22,56]. The studies that do account for the uptake of ideas and concepts typically apply manual qualitative analyses. Such analyses are not only tedious and error-prone but also difficult to execute reliably at scale [7,21]. Recent studies that apply automated citation classification approaches [26,50,62] rely on traditional machine learning techniques that utilize derivative features inferred from paper-level meta-data, which ignore the ideas and concepts that are being cited.

\* Corresponding author.

E-mail addresses: [j.prester@unsw.edu.au](mailto:j.prester@unsw.edu.au) (J. Prester), [gerit.wagner@hec.ca](mailto:gerit.wagner@hec.ca) (G. Wagner), [guido.schryen@uni-paderborn.de](mailto:guido.schryen@uni-paderborn.de) (G. Schryen), [nhassan@d.umn.edu](mailto:nhassan@d.umn.edu) (N.R. Hassan).

<https://doi.org/10.1016/j.dss.2020.113432>

Received 18 March 2020; Received in revised form 21 October 2020; Accepted 24 October 2020

Available online 4 November 2020

0167-9236/© 2020 Elsevier B.V. All rights reserved.

Our study proposes and evaluates an automated approach that focuses on the ideas and concepts cited and assesses the ideational impact of research papers. We propose a deep learning method called Deep Content-Enriched Ideational Impact Classification (Deep-CENIC) that utilizes the content of citing sentences<sup>1</sup> and its derivative syntactic, semantic, and contextual features. Deep-CENIC extends state-of-the-art citation classification techniques using a word embedding dimension based on the actual content of in-text citations and a novel deep learning architecture. The combination of word embedding and derivative features, which is lacking in current citation classification approaches [25], improves the representation of the conceptual content of citations and identifies impactful citations more effectively.

Our study makes the following contributions. First, we advance the conceptualization of ideational impact of citations and define a model for classifying ideational impact [22,56]. Second, we design a content-enriched deep learning approach (Deep-CENIC) to identify ideational impact. Thus, our methodological contribution to academic recommender systems [52,55,68] advances the design of systems that support researchers in responding to information overload by identifying ideational knowledge flows [3]. Third, our study offers an illustrative ideational impact analysis of the IT business value (ITBV) domain and provides insights into how knowledge grows in the information systems (IS) field vis-à-vis the significant differences between ideational impact and impact based on citation numbers. With this, we contribute to the theoretical discourse in IS, by offering an approach for evaluation of the growth of knowledge and state of theory in the IS discipline [17,66,67].

## 2. Background

### 2.1. Ideational impact and the growth of knowledge

As Keen [31 p. 9] emphasized: “Unless we build on each other’s work, a field can never emerge, however good individual fragments may be.” We focus on this process of building on each other’s works as represented by citations of the works of others. Different approaches such as literature review and meta-analysis have been proposed to analyze this process; however, they focus primarily on synthesizing existing knowledge [20]. Thus, such techniques require empirical evidence and established sets of theoretical constructs, which makes them less effective for tracing ideas and concepts across large literature corpora. Citations, on the other hand, indicate the growth of knowledge and the more those works are cited, the more impact they are assumed to have in building that cumulative tradition [14,42]. However, many studies argue that not all citations are equal [25,62,70]. Each citation can play a different role within the CA and the impact of each citation varies [4,47]. For example, some studies categorize citations into different levels of importance based on the frequency and location [62,70] and context of the citations [25]. What these studies ignore is the conceptual content carried by the citations and their relevance to the citing work. The conceptual content constitutes the ideational dimension of “citations as signs for ideas and concepts offered by and imparted onto the cited text” [22].

The ideational dimension [22] of citing is based on Small’s [56] work on the symbolic perspective of citations and how citations represent ideas as they flow through the citation network. Previous studies of citation importance count the number of times a paper is cited in the CA as an indicator of importance. Specifically, we refer to the uptake of a paper’s ideas and concepts by subsequent research [22 p. 18,56] as “ideational impact.” As Small [57 p. 72] noted, “When scientists agree on what constitutes prior relevant literature, including what is significant in that literature, they are in fact defining the structures of their communities.” Citations that play a significant role in the main arguments of the CA have “ideational impact” and can be expected to contribute to the growth of knowledge of that discipline [22].

The ideational dimension follows from scientometric traditions that foreground qualitative rather than quantitative characteristics of citing

behavior such as that of Merton’s [40] studies of originality and priority. The qualitative characteristics of citations are rarely studied even though scholars of citation analysis had always emphasized the value of qualitative information that could help us better understand the relationship between the citing and cited works [36]. For example, by analyzing the discourse reflected in the context surrounding the citations, it is possible to interpret the function of the citation [24,59]. Others interview authors to qualitatively classify citations [7] and find that when citing authors engage, build or tie using the citations [21] they are more likely to be improving, modifying or extending the cited works.

The qualitative analysis of citations using techniques such as interviewing authors can be very tedious and will not be efficient when analyzing a corpus of studies or citations at the level of a discipline. Several studies have begun investigating this dimension [69], although, not directly addressing the conceptual content. In addition to grammatical and locational features, Teufel et al., [60] applied cue phrases to measure the sentiment of the citing author toward the cited article. Noting these issues, scholars are stressing the need to understand the ideational dimension of citations and to develop automated approaches that can address the above-mentioned challenges at scale [38]. Scientometricians consider such automated approaches to supporting scientific collaboration [52,68], and assessing knowledge transfer [3,9], to be the future of citation analysis [10].

### 2.2. Natural language processing and machine learning in citation classification

Advances in the availability of digital text and computing power have opened up new opportunities for automated analyses of natural language, leading to an increase in the application of natural language processing (NLP) and text classification [29]. Both methods are suitable for assessing links between documents, which are represented by citations. Thus, NLP-based approaches can help to solve challenges such as automated content analyses of research articles, reliable coding of citation patterns, and overcoming subjectivity and variation in qualitative citation analyses [11].

Given the complexity of the human language, NLP approaches typically split language into seven conceptual levels [44]: phonology, morphology, lexicography, syntax, semantics, discourse, and pragmatics. The syntactic and semantic levels are of particular importance for text classification. On the syntactic level, NLP algorithms examine how words are combined and used to form sentences. Major tasks on the syntactic level include word segmentation, stemming, part-of-speech (POS) tagging, and parsing [44]. To identify the meaning of what is written, the semantic level goes beyond the structure of words and sentences. Prominent applications on the semantic level include machine translation, sentiment analysis, named entity recognition, and topic recognition.

These advances in NLP have enabled automated document analyses and text classification approaches including the classification of citations [11,25]. We surveyed the literature in citation classification based on NLP and machine learning to understand the analytical approaches and inform our research design. Table 1 shows the recent studies in automated citation classification together with our own approach for comparison. The novel aspects of our approach are emphasized.

The classified citation impact in most studies distinguishes between important and non-important [25,26,50,62] or influential and non-influential citations [70]. Although this distinction is useful in the sense that it goes beyond simple citation numbers, this classification approach lacks an ideational dimension and does not account for the conceptual content implied by the citations.

This dominant focus on classifying important and non-important, or influential and non-influential citations can largely be attributed to the lack of available datasets for citation classification. Valenzuela et al. [62] provide one prominent dataset, which is often seen as the gold

**Table 1**  
Recent studies on automated citation classification.

Reference	Categories	Dataset	Features	Method	Performance
Valenzuela, Ha and Etzioni [62]	Important Incidental	Computational linguistics <sup>a</sup> (N = 465)	Context Structure Cue-phrases Meta-data	SVM RF	F-score: 75%
Zhu, Turney, Lemire and Vellino [70]	Influential Non-influential	Computer science (N = 3143)	Context Structure Cue-phrases Meta-data	SVM	F-score: 42%
Hassan, Safder, Akram and Kamiran [26]	Important Non-important	Computational linguistics <sup>a</sup> (N = 465)	Context Structure Cue-phrases Meta-data	SVM RF	AUC: 84%
Hassan, Imran, Iqbal, Aljohani and Nawaz [25]	Important Non-important	Computational linguistics <sup>a</sup> (N = 465)	Context Structure Cue-phrases Meta-data Sentiment	SVM RF LSTM	AUC: 89%
Jurgens, Kumar, Hoover, McFarland and Jurafsky [30]	Background Uses Compares Motivation Continuation Future	Computational linguistics <sup>a</sup> (N = 1969)	Structure Cue-phrases Meta-data Argument	RF	F-score: 53%
Qayyum and Afzal [50]	Important Non-important	Computational linguistics <sup>a</sup> (N = 465)	Context Meta-data	SVM RF LR	F-score: 73%
This article's approach (Deep-CENIC)	<b>Ideational</b> <b>Non-ideational</b>	Information systems / IT business value review articles (N = 1256)	Syntactic Semantic Context Word embeddings	<b>End-to-end BLSTM</b>	F-score: 83%

Note: Support Vector Machine (SVM); Random Forest (RF); Long Short-Term Memory (LSTM); Bi-directional Long Short-term Memory (BLSTM); Logistic Regression (LR).

<sup>a</sup> Association for Computational Linguistics (ACL) Anthology dataset.

standard in citation classification. While the dataset has been shown invaluable for means of evaluation and comparability in many citation classification studies [25,26,50], its heterogeneity in article genres and research topics, and the limited sample size allow for little empirical insight into the growth of knowledge in a field.

Features used in automated citation classification are largely focused on contextual, structural, and meta-data-based properties of research articles and the syntactic and linguistic features of citations. The actual content of citing sentences, citation contexts, and cited CAs are rarely considered. The content of citing sentences is by and large only exploited through the identification of cue words that signal certain citing intentions. Thus, the rich content of citing sentences is reduced to a limited set of cue phrases. While traditional machine learning techniques used in previous citation classification studies are constrained in their use of natural language as direct input features, one more recently developed set of analytical methods – deep learning methods – has proven to be extremely effective in text recognition and NLP.

### 2.3. Deep learning and text classification

Deep learning is a new type of machine learning that utilizes multiple stacked layers of artificial neural networks to progressively extract higher level features from the raw input data [35]. State-of-the-art deep learning architectures, based on artificial neural networks, have been successfully applied to research areas including NLP [16]. Deep learning approaches use multiple hidden layers, potentially modeling complex non-linear relationships more effectively and outperforming traditional machine learning models.

While supervised deep learning models include many different architectures such as convolutional neural networks, recursive neural networks, and reinforcement learning approaches, the class of networks that is most effective for classification of sequential data, such as text

data, is Recurrent Neural Networks (RNNs). RNNs model connections between neural nodes along a sequence as found in time series data or textual sentence structures. This architecture enables such approaches to account for sequential features. Specifically, Long Short-Term Memory (LSTM) networks have recently broken records in improved machine translation, language modeling, and text classification [35]. LSTMs are a type of RNN that includes feedback loops, allowing data to persist over many network iterations and thereby enabling the discovery of long-term dependencies. Since RNNs store information from previous iterations, they can learn by keeping information of each word even though the word is distant from the current iteration. It thus makes RNNs the preferred deep learning architecture for sequence learning, NLP, and text classification.

Despite remarkable successes of deep learning approaches, which have been proven to outperform traditional machine learning approaches in contexts such as recommender systems [18], fraud detection [65], and fake news detection [37], they have not yet been considered by researchers in citation classification (cf. Table 1). This paper explores the potential of deep learning architectures in ideational impact classification incorporating both syntactic, semantic, and contextual derivative features and word embeddings based on the actual content of the citing sentences.

### 3. Deep-CENIC: classifying ideational impact

Our objective in this study is to develop an approach that can classify the ideational impact of a RA based on the uptake of its concepts by CAs and thus support researchers in searching for papers that have taken up ideas and concepts of prior work. We therefore start by outlining our techniques for data collection and coding procedure, explain the construction of features extracted from the data, and present the classification framework.

### 3.1. Data collection and coding procedure

To develop and evaluate the ideational impact classification approach, we collected a corpus of documents comprising cited review articles (RAs) on ITBV and their CAs. Our full sample of RAs is based on the set described by Wagner, Prester, Roche, Benlian and Schryen [63]. They originally identified 214 standalone RAs that have been published in a set of 40 major IS journals between 2000 and 2014. We chose RAs because they (1) aggregate the key theories, concepts and ideas of a discipline, (2) address the main questions and problems and summarize the major issues and debates, and (3) synthesize the fragmented body of knowledge of a discipline into a coherent whole [51,54]. Although some RAs do not go beyond summarizing the state of a field [20], methodologists and editors have emphasized the need for RAs to make a substantive contribution to theory [49]. Thus, in many fields, RAs present a promising genre for making theoretical contributions and advancing scientific knowledge. For these reasons, RAs represent an interesting paper genre for analyzing ideational impact and the growth of knowledge. Furthermore, RAs have been shown to receive considerable numbers of citations [43]. Such high numbers of CAs pose problems for researchers when searching for literature on a particular concept, because the manual effort required to sift through vast amounts of research to find relevant articles is increasing substantially [34]. Therefore, RAs present an ideal case to develop automated citation analysis approaches, which reduce the manual effort required in filtering for the most relevant literature. From this dataset, we chose RAs in the domain of ITBV, which leaves us with a total of 24 RAs. The ITBV literature is concerned with the impact of investments in particular IT assets on the performance of organizations and other economic entities [53]; it is a major research topic for IS researchers. The ITBV domain is mature enough to provide sufficient number and diversity of RAs (e.g., theoretical RAs and meta-analyses) as well as a high number of CAs.

We conducted a citation analysis in the form of a forward search to find all CAs potentially using the knowledge developed in the cited RAs. This resulted in the identification of approximately 30,000 CAs. Since the generation of an annotated corpus for the training of the machine learning classifiers requires a manual coding of every paper, we decided to filter the CAs for papers that have been published in journals included in the Senior Scholars' Basket of Journals. This set of eight journals is widely acknowledged as a basket of top journals in the IS field and recognizes topical, methodological, and geographical diversity. Thus, the final dataset comprises 1256 CAs published in journals included in the Senior Scholars' Basket of Journals, in addition to 24 RAs on ITBV. To test the robustness of our model against a broader set of publication outlets, we collected an extended test set from the original 30,000 CAs. A list of these articles referred to in this paper is included in the supplementary online appendix.

Based on the distinction provided by Hassan and Loebbecke [22] between the ideational dimension and other dimensions of citations, we employed a coding scheme that classifies citations which facilitate the growth of knowledge and therefore have ideational impact. Factors we

considered when coding for ideational impact include the extension of a theory developed in the RA, synthesizing or closing of research gaps identified in the RA, and responding to a research agenda developed in the RA. Table 2 illustrates some representative citing sentences and coding decisions. Ideational impact was coded by manually analyzing each of the 1256 CAs. We coded ideational impact when we found an explicit and direct attribution of the concepts developed in the RA. Thus, when coding ideational impact it is necessary to consider the text of both the cited and CA to judge whether a citation represents concepts from the CA [58]. Inter-rater reliability between the two coding authors was sufficient with a Cohen's Kappa value of 0.89.

### 3.2. Feature set construction

For our deep learning approach, we developed a feature set based on the citation classification literature<sup>2</sup>. We included syntactic, semantic, and contextual dimensions of citing sentences as shown in Table 3.

We operationalize a set of syntactic features, because we expect

**Table 3**  
Feature set.

Feature	Description (Data Source)	References
<b>Syntactic Features</b>		
Textual type	Number of textual citations (CA)	[62]
Standalone reference	Number of 'standalone' citations (CA)	–
Position in sentence	Position of the reference in citing sentence (CA)	[28]
Comparative/superlative clauses	Number of comparative and superlative clauses (CA)	[28]
Personal pronouns	Number of personal pronouns (CA)	[28]
POS patterns	Appearances of POS patterns (CA)	[11]
<b>Semantic Features</b>		
Title/abstract similarity	Semantic similarity of the titles and abstracts (RA & CA)	[19]
Citation sentiment	Sentiment of the citing sentence with regards to the RA (CA)	[2,28]
RA knowledge contributions	Knowledge developed in the cited RA as a prerequisite for ideational impact (RA)	[54]
<b>Contextual Features</b>		
Position within full text	Number of citations appearing in the different sections of the paper (CA)	[11,28]
Citations toward the RA	Total number of RA citations (CA)	[28]
Citing sentence variety	Number of different citations in the citing sentence/context (CA)	[11,28]
Citing sentence density	Focal citations divided by total citations in the sentence (CA)	[11,28]
Total number of references	Total number of references in the CA's bibliography section (CA)	[28]
Total citations	Total number of citations in the CA (CA)	[28]
Weighted citation count	RA citations divided by total citations (CA)	[28]
Self-citation	At least one author of the RA and CA is identical (RA & CA)	[60,64]

**Table 2**  
Impact types and example citing sentences.

Impact Type	Example Citing Sentence	Rationale for Coding
Ideational	"Drawing on the resource-based view (RBV) of the firm as an overarching framework and prior research ([...] Melville et al. 2004 [...]), we propose three reasons to explain why overall IT investments are likely to have a positive association with accounting profits." – Mithas et al. [121 p. 207] "This decision was based, in part, on work suggesting that our understanding of the BVIT would benefit from the use of primary data to empirically examine the link between IT and firm performance ([...] Wade and Hulland 2004)." – Nevo and Wade [122 p. 408]	CA draws on the RBV to develop the concept of IT investments. CA follows the proposed research agenda.
Non-ideational	"DeLone and McLean (1992) provide a thorough overview of the main research in the quest for the key success factors of that time." – Bartis and Mitev [75 p. 113] "Information technology (IT) that promises to enhance organizational performance costs companies millions of dollars to implement (Kohli and Devaraj 2003)." – Xue et al. [155 p. 400]	RA is cited as an exemplary review on the topic. RA is cited to highlight the business impact of the topic.

Note: Dataset references available in the supplementary online appendix

citing sentences that signal ideational impact to adhere to a specific sentence structure. By analyzing the *textual type* of a citation, we distinguish between the two major syntactic types, namely textual (i.e., author name outside the reference marker) and non-textual (i.e., simple reference marker) citations [62]. Because we filtered the CAs in our dataset for papers that appeared in the eight journals included in the Senior Scholars' Basket of Journals, we fit our citation identification approach to the particular journal guidelines on reference formats. Thus, we were able to capture author-date citations (e.g., Harvard style) as well as numeric citations (e.g., Vancouver style). We further look at whether the citation 'stands alone' or whether it is a part of multiple references grouped together in one reference marker. Additionally, as proposed by Jochim and Schütze [28], we extract the absolute position of the citation within the citing sentence. We also developed features based on the grammatical structure and POS sequences of the citing sentences including the use of *comparative and superlative clauses* (e.g., more, best), first- and third-person *personal pronouns*, and *POS patterns* signaling certain citation functions [11].

We further operationalize a set of semantic features. Topic relatedness between citing and cited papers is an important indicator of ideational impact. We operationalize topic relatedness between cited and CA as *title and abstract similarity* [19], and measure it based on latent semantic analysis (LSA) [33]. Another semantic feature that has been proven useful for citation classification is text *sentiment* [2,28]. Sentiment analysis extracts affective and subjective information from documents that express the author's attitude toward the text [45]. Hence, we include measures for citation sentiment to control for authors' attitude toward the RA, which is expected to show more emphasis when authors select citations that carry conceptual and organic implications [2,28]. Lastly, we consider the manually coded *knowledge contributions* of the RAs as a prerequisite for ideational impact on the CAs. In essence, we consider the following types of knowledge contributions [54]: synthesis, adoption of a new perspective, theory building, theory testing, identification of research gaps, and provision of a research agenda. This feature is important because CAs can only build upon knowledge contributions made by the cited RA [54,58].

The development of our contextual features is primarily based on citation metadata. Because the location of a citation has been shown to be "the most reliable information on citation function one could obtain from the paper directly" [11 p. 625], we extract both the *location* of the

citations in the full text of the CA as well as the total number of *citations toward the cited RA*. We further extract the number of different references cited in the citing sentence (i.e., *citing sentence variety*) and the proportion of citations toward the RA against total citations in the citing sentence (i.e., *citing sentence density*). We also include the *total number of references* in the bibliography section and the *total number of citations* of all cited papers. This is done to derive a *weighted citation count* across all references. Because citing one's own research might indicate a higher probability of re-use of intellectual material from previous work, we include a feature indicating a *self-citation* [64].

### 3.3. Content-enriched deep learning approach

We apply the features listed in Table 3 in a novel deep learning approach, we call Deep-CENIC (**Deep** - Content-ENriched Ideational Impact Classification) to predict an RA's ideational impact. It is composed of two components: a deep neural network (DNN) utilizing syntactic, semantic, and contextual features and a bidirectional LSTM (BLSTM) utilizing the actual content of the citing sentences. The deep learning architecture of our ideational impact classification approach is depicted in Fig. 1.

The proposed approach starts by creating word embeddings from the words of all citing sentences by efficiently mapping semantic information onto a dense distributed representation in the form of a 256-dimensional vector. The distributed representation is learned based on the usage of words in citing sentences. This allows words that are used in similar ways to result in having similar representations, naturally capturing their meaning. Let  $S$  be a citing sentence representing the sequence  $[w_1, w_2, \dots, w_T]$ . Variable  $w_t$  denotes word  $t$  in the citing sentence. The objective for training the word embedding layer is to, given a word  $w_t$ , maximize the average log probability as presented in the objective function  $L$  in Eq. (1).

$$L = \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t) \quad (1)$$

Parameter  $T$  denotes the number of training words. Parameters  $-c$  and  $c$  are limits of our context window with  $w_{t+j}$  denoting the words surrounding  $w_t$  within the context window. We use a context window of five as recommended by the developers of the word embedding model [16]. We use the Skip-gram method to build the word embedding,

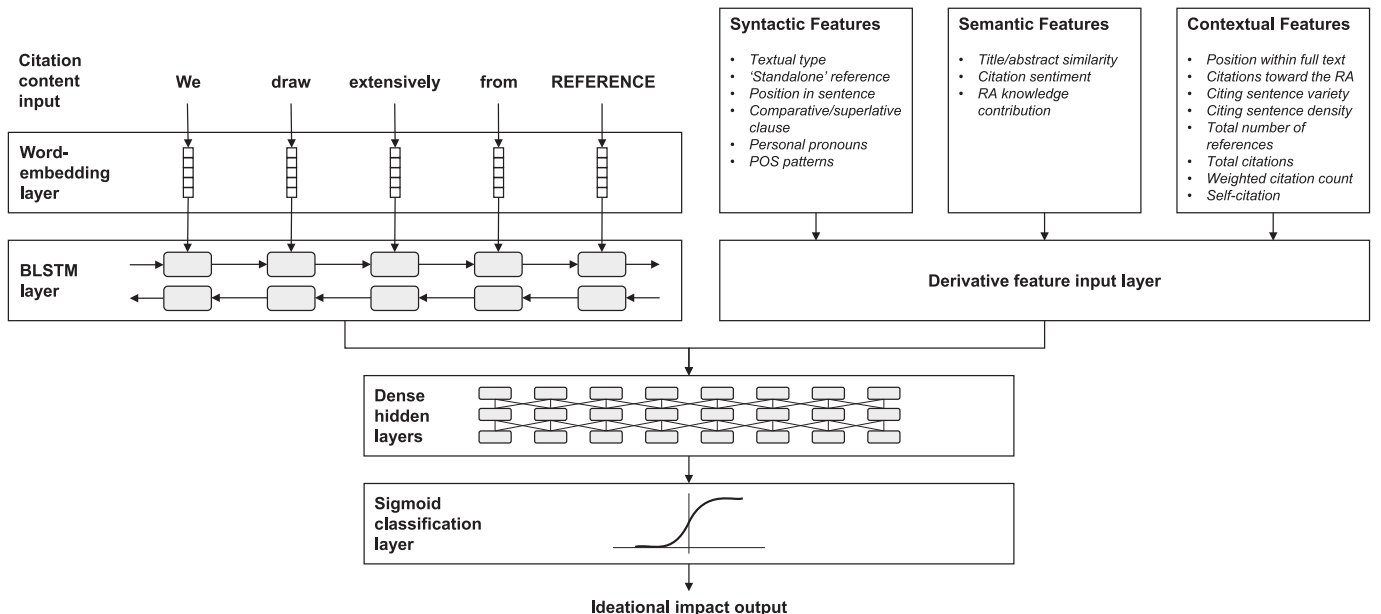


Fig. 1. Deep-CENIC architecture.



because it is designed to predict the context of a word and therefore performs better in predicting uncommon words of academic writing.

To effectively classify the semantic meaning of citing sentences, we used the dense vector representations as the input for a BLSTM deep learning model. BLSTMs are an extension of traditional LSTMs that train two instead of one LSTM on both the original input sequence and the reversed input sequence. We use the bidirectional structure because citing sentences follow the syntactic and semantic rules of the English language as well as structural rules with regards to common citation practice. Words in the sentence may exhibit long-distance semantic dependencies regardless of the word order, which the bidirectional structure is able to capture. For example, reference markers can stand at the beginning or end of a citing sentence, while the referenced concept or idea can be named at the opposite end.

We employ a BLSTM that takes the 256-dimensional word embedding as the input to feed its 64 LSTM units. An LSTM unit consists of a memory cell, which keeps track of the dependencies between the word embeddings in the input vector  $x_t$  and three gates, which steer the flow of information inside the unit: an input gate with activation vector  $i_t$ , an output gate with activation vector  $o_t$ , and a forget gate with activation vector  $f_t$ . At each step in the citing sentence sequence, the LSTM takes both the last hidden state and the word embedding as the current input to compute the cell state vector  $c_t$  and the current hidden state with vector  $h_t$ . The weight matrices  $W$  and  $U$ , and bias vectors  $b$ , which need to be learned during training, determine how the gates operate. The computational learning steps taken in the LSTM unit are summarized in Eqs. (2)–(6).

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f); \quad (2)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i); \quad (3)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o); \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c); \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

The second part of the Deep-CENIC architecture utilizes the complementary features set comprising the derivative properties of the citing sentences. The syntactic, semantic, and contextual features described in Section 3.3 serve as the 17-dimensional input vector for each citing sentence. These derivative features are merged with the 64-dimensional output of the BLSTM layer.

The DNN structure of Deep-CENIC consists of three stacked densely connected neural network layers. Each neural network layer consists of 64 units. The first neural network layer receives input vector  $x$  representing the 81-dimensional input sequence  $[x_1, x_2, \dots, x_{81}]$ . Each component of the input vector corresponds to one component of the weight vector  $w$ , which need to be learned during training. Thus, the summation of the product of the individual  $x_i, w_i$  pairs is equivalent to the dot product of the input and weight vector. The bias added dot product is activated through a rectified linear unit. The computational learning steps taken for each densely connected neural network layer  $l$  are summarized in Eq. (7).

$$y^{(h)} = \max(0, (x \odot w_l) + b_l) \quad (7)$$

$$p(y = j|x) = \frac{e^{(x \odot w_o) + b_o}}{e^{(x \odot w_o) + b_o} + 1} \quad (8)$$

Finally, after the third densely connected neural network layer, a single Sigmoid layer (Eq. (8)) is computed to predict a single output value – that is, the ideational impact classification. Variable  $y$  denotes the predicted ideational impact type. Variable  $x$  denotes the output of the last densely connected layer or the input to the final Sigmoid layer. Vector  $w_o$  denotes the weight parameter and variable  $b_o$  denotes the bias parameter. The Sigmoid layer is a frequently used method for binary

classification that maps  $n$ -dimensional real-valued inputs  $x$  to a  $j$ -dimensional real-valued output  $y$  computing a value between 0 and 1. In our Deep-CENIC model, the Sigmoid function produces the probability of ideational impact type  $j$  (ideational impact or non-ideational impact) given the input  $x$ .

#### 4. Evaluation to examine classification performance

##### 4.1. Benchmark models and evaluation metrics

We select three classes of machine and deep learning methods as benchmarks to evaluate our Deep-DENIC approach: a discriminant machine learning model, a decision tree model, and deep learning models.

Support Vector Machine (SVM) is a type of discriminant machine learning model that has been prominently used in citation classification and shown promising performance in various studies [11,25,50]. The SVM classifier predicts whether a RA has an ideational impact on a CA or not, given the set of syntactic, semantic, and contextual derivative features. In line with earlier research [25], we use all derivative features presented in Section 3.3 as the input for the SVM classifier.

Random Forest (RF) models, a type of decision tree models, are the top-performing traditional machine learning models for citation classification [25,62]. Therefore, we benchmark our proposed method against a RF classifier trained on the same derivative feature set used in the SVM model. We used scikit-learn, a Python library for machine learning, to implement the benchmark models.

As our model is an enhanced deep learning model that combines content-based word embeddings and syntactic, semantic, and contextual derivative features, we benchmark our proposed Deep-CENIC model against two standard deep learning models: DNN and BLSTM models. The DNN represents a sequential architecture of three fully connected hidden layers identical to the DNN utilized in the Deep-CENIC approach, but without the additional word embeddings. The BLSTM model classifies ideational impact exclusively based on word embeddings identical to the architecture used in the Deep-CENIC model, but without taking derivative features into account.

We adopted precision, recall, and F1 score as the evaluation metrics because they are commonly used in text classification studies with binary prediction models [11,28,50]. Precision assesses the proportion of citations the model classified as ideational impact citations that actually symbolize ideational impact. Recall measures the proportion of actual ideational impact citations in the dataset that the model can identify. F1 score is a comprehensive measure of accuracy integrating both precision and recall. Researchers are usually interested in optimizing either precision or recall, since there is an inverse relationship between the two measures. In our study, we aim to identify as many ideational impact citations as possible. Extracting more ideational impact citations could lead to a more holistic assessment of a discipline's cumulative tradition. Such improved understanding of ideational impact and its role in building a cumulative tradition could provide us with a new tool to evaluate the growth of knowledge of a discipline. Therefore, we weigh recall higher than precision in the context of our study.

##### 4.2. Evaluation of ideational impact classification

We evaluate our models on the annotated dataset, with 1256 RA-CA pairs and 3493 citing sentences. We used 80% of the data as the training set and 20% as the test set. To avoid overfitting, all the evaluations were performed in ten-fold cross validation using 10% of the training set as the validation set. We repeat the training procedure for each model 50 times and report the average performance on the test set in Table 4.

As Table 4 shows, our proposed Deep-CENIC model achieves the highest precision (84.28%), recall (82.59%), and F1 score (83.36%). Although the other machine and deep learning models were, overall, able to achieve high precision values, our Deep-CENIC model has the

**Table 4**  
Evaluation of ideational impact classification.

Model	Precision	Recall	F1 Score
SVM	62.55%	35.97%	45.24%
RF	63.56%	38.22%	47.50%
DNN	52.22%	48.03%	49.60%
BLSTM	62.62%	62.89%	62.65%
<b>Deep-CENIC</b>	<b>84.28%</b>	<b>82.59%</b>	<b>83.36%</b>

most salient advantage in the recall. Because of this our model also outperforms all other models in terms of the F1 score. Considering recall is more important in citation classification and Deep-CENIC also reaches high precision values, Deep-CENIC is the best approach for ideational impact classification. The substantial improvement in recall could help identify most of the ideational impact citations representing papers that extend concepts and ideas of the cited paper, thus providing an effective filtering mechanism for ideational impact citations. Comparing the approach to other citation classification studies, our proposed Deep-CENIC model outperforms earlier work based on traditional machine learning models [50,60,62] and performs on par with the state-of-the-art models [25].

Our proposed Deep-CENIC model improves the most popular model for citation classification (RF) in the recall by 44.37% and more recently applied models (BLSTM) by 19.69%. The improved recall enables Deep-CENIC to identify 68 more ideational impact citations than RF and 23 more than BLSTM. These ideational impact citations cover 301 citing sentences in 96 pairs of cited and CA in our test dataset. This increase in performance accounts for 60.05% of cited and CA pairs that represent ideational impact.

To test the significance of Deep-CENIC's performance improvement, we conducted pairwise *t*-tests based on 50 iterations of the training and testing procedures of each model to compare the performance of the Deep-CENIC model against benchmark models. The results indicate that our proposed Deep-CENIC model significantly outperforms all the benchmark models ( $p < 0.001$ ). Table 5 shows the pairwise *t*-test results.

Deep-CENIC significantly outperforms all the benchmark models because of its unique architecture that combines both citation content word embeddings of the citing sentences and syntactic, semantic, and contextual derivative features. The model can identify ideational impact based on the content of the words in the citing sentences. We perform a sensitivity analysis that shows the individual effectiveness of the syntactic, semantic, and contextual derivative features and the citing sentence-based word embeddings compared to the combination of both models in Deep-CENIC (Table 6). The significance tests are shown in Table 7.

The comparison between Deep-CENIC, which combines original word embedding and syntactic, semantic, and contextual derivative features, and the individual models utilizing only one of the two feature sets shows the effectiveness of combining both models in our Deep-CENIC architecture. Deep-CENIC significantly outperforms both the derivative feature-only and the original word embedding-only models in terms of precision, recall, and F1 score ( $p < 0.001$ ). Our results, therefore, demonstrate the relevance of the actual content of citing sentences for automated ideational impact classification.

**Table 5**  
Pairwise *t*-tests for Deep-CENIC against the baseline models.

Model Comparison	$\Delta\mu$ -Precision	$\Delta\mu$ -Recall	$\Delta\mu$ -F1 Score
Deep-CENIC vs. SVM	21.73%*	46.61%*	38.12%*
Deep-CENIC vs. RF	20.71%*	44.37%*	35.86%*
Deep-CENIC vs. DNN	32.05%*	34.56%*	33.76%*
Deep-CENIC vs. BLSTM	21.65%*	19.69%*	20.71%*

\*  $p < 0.001$

**Table 6**  
Sensitivity analysis of individual models.

Model	Precision	Recall	F1 Score
Derivative features model (DNN)	52.22%	48.03%	49.60%
Original word embedding model (BLSTM)	62.62%	62.89%	62.65%
<b>End-to-end learning model (Deep-CENIC)</b>	<b>84.28%</b>	<b>82.59%</b>	<b>83.36%</b>

**Table 7**  
Pairwise *t*-test for sensitivity analysis of individual models.

Model Comparison	$\Delta\mu$ -Precision	$\Delta\mu$ -Recall	$\Delta\mu$ -F1 Score
End-to-end learning model (Deep-CENIC) vs. derivative features model (DNN)	32.05%*	34.56%*	33.76%*
End-to-end learning model (Deep-CENIC) vs. original word embedding model (BLSTM)	21.65%*	19.69%*	20.71%*

\*  $p < 0.001$

#### 4.3. Robustness analysis

To analyze the robustness of our results, we tested the classification performance of the Deep-CENIC model on another dataset. This is to evaluate potential overfitting of our model that has been trained on a relatively narrow set of journals with specific reference guidelines and institutionalized citing practices. Thus, our model could potentially perform worse when classifying CAs that appeared in a broader set of journals. To test the robustness of Deep-CENIC, we coded a random sample of 150 CAs from the more than 30,000 original citations. CAs in this sample have been published in 40 major IS journals as well as conference proceedings published by the Association for Information Systems. We followed the same coding approach as for the original dataset.

Our Deep-CENIC model achieves similar performance on the broader test set when measured by precision (73.31%), recall (80.49%), and F1 score (76.73%). Although the performance on the broader set of publication outlets is lower than for the original dataset, it is nevertheless performing reasonably well on a dataset drawn from a different set of publication outlets. Considering recall is more important in our study, it is interesting to note that the recall on the broader test set is higher than the precision. This means that Deep-CENIC can identify more ideational impact cases in the broader test set relative to the data set it was trained on. The results of our robustness analysis, therefore, show that our model can identify ideational impact within the broader IS discipline.

#### 5. Illustrative ideational impact analysis of ITBV-RAs

We conducted an illustrative analysis to demonstrate Deep-CENIC's ability to support ideational impact studies and explore its potential for evaluating the growth of knowledge. In the first part of the analysis we used Deep-CENIC as a recommender system for academic literature to automatically identify ideational impact citations. In the second part of the analysis we built on the automated analysis and conducted a manual qualitative analysis of knowledge growth in the ITBV domain.

##### 5.1. Comparison of ideational impact and non-ideational impact

We applied Deep-CENIC as a search tool to filter our dataset of ITBV-RAs for citing relationships that were identified as ideational impact cases. Fig. 2 charts the distribution of ideational and non-ideational impact for our dataset of 1256 citing relationships from which Deep-CENIC identified 326 as ideational impact cases. Ideational impact is significantly lower than the overall citation impact (i.e., the sum of ideational and non-ideational impact), averaging 25.95% of annual citations (95% CI: [20.40;31.50]).

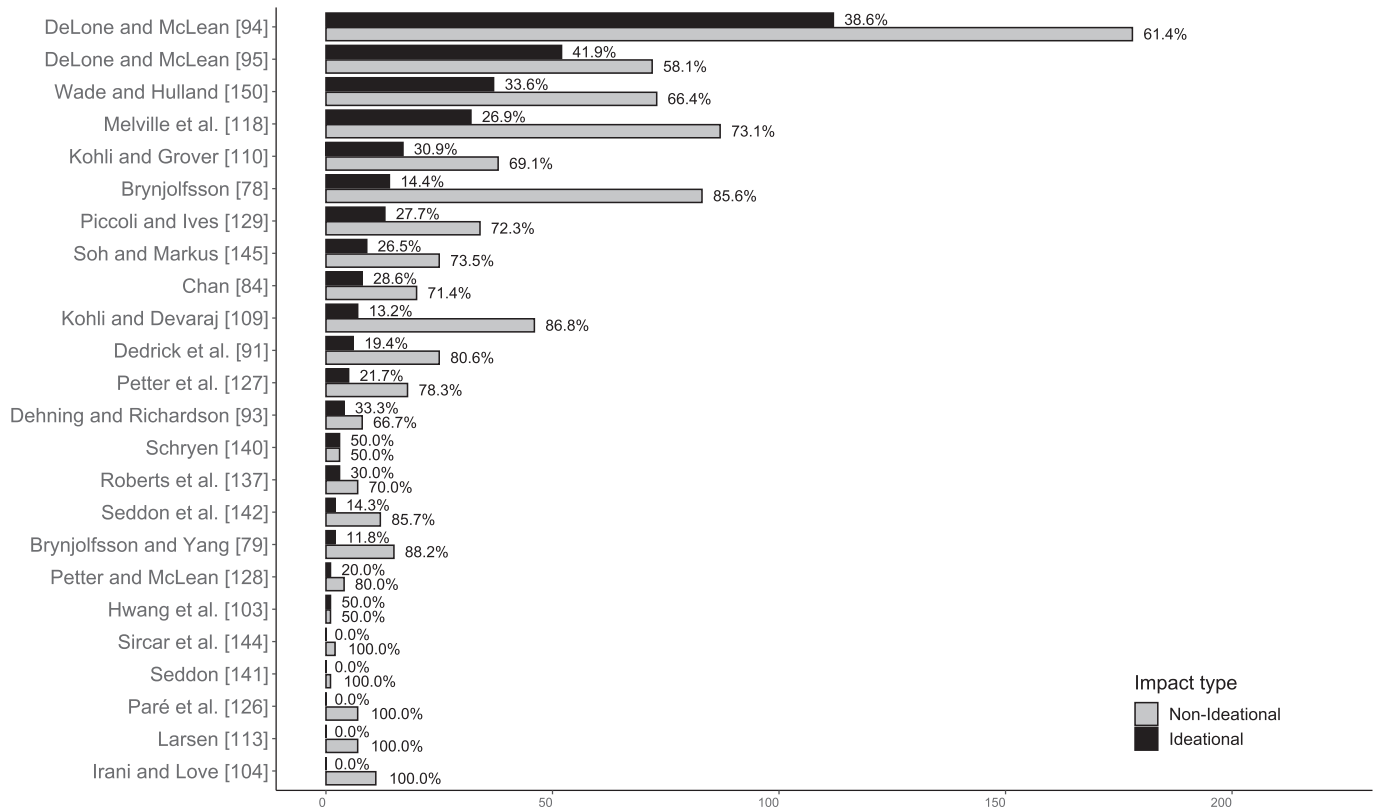


Fig. 2. Ideational vs. non-ideational impact comparison (dataset references are available in the supplementary online appendix).

The comparison of ideational and non-ideational impact citations as displayed in Fig. 2 reveals several striking differences when RAs are ranked according to ideational impact. For example, although Brynjolfsson's [78] paper on the productivity paradox of IT is considered a classic in the ITBV literature it has received relatively few ideational impact citations. Further, the analysis illustrates the possibility that low-impact papers exert high proportions of ideational impact, or that high-impact papers exert low proportions of ideational impact. For example, the two papers by DeLone and McLean [94, 95] on IS success differ exactly on this point. The later publication contains a higher proportion of ideational impact citations compared to the earlier version. These differences suggest that the extended model offered in the more recent paper represents a strong impulse for the growth of knowledge on IS success. In the next part of our analysis we expand on this aspect and show how Deep-CENIC provides a valuable starting point for exploring the growth of knowledge in the ITBV domain.

## 5.2. Analysis of knowledge growth based on Deep-CENIC

As described in Section 2 above, the broader value of analyzing the ideational dimension of citations lies in the potential to elaborate on how concepts imparted from and onto the cited text [22] evolve throughout diverse research streams. The use of this category of citations will have a major impact on the growth of knowledge in the IS field, and in particular, on addressing the lack of indigenous theories in the field. Theory is one of the major types of knowledge contribution that happens to be an important goal of any research effort as well as how that research is evaluated [1]. The IS field has been debating the nature and role of these knowledge contribution for some time with intense debates regarding whether or not a theoretical core is necessary [39,66] and disagreements concerning whether or not the field can speak of native theories [17,67]. Because scientific investigations self-select evidence for the construction of knowledge [32], citations and references to previous studies play a critical role in how the knowledge

is constructed. The Deep-CENIC approach proposed uses the context surrounding the citations to classify its function. Once citations are classified as ideational, it is possible to further analyze qualitatively if the pattern of citations in the citing article offer knowledge contributions that are original, conceptual and organic to the research area [24].

To go beyond a classification of citations into ideational and non-ideational, we conducted such a manual qualitative analysis of the citation contexts extracted and recommended by Deep-CENIC. Instead of ranking individual RAs as discussed in the previous section, the objective of this analysis is to map the development of key concepts in the ITBV domain. The qualitative analysis unfolded in three steps. First, we analyzed the citation contexts that have been classified as ideational. The citation contexts included the citing sentence, the sentences immediately surrounding the citing sentence, and the broader context of the citation within the paragraph, section, as well as the entire paper. Second, we extracted the key concepts and ideas that were referred to in the context of the citing sentences. This task was performed by two of the authors and, in case of disagreements, reviewed by a third author (all three authors had considerable experience with the ITBV literature and its conceptual foundations). Third, we grouped commonly occurring concepts together for each RA. Table 8 shows the top RAs in terms of ideational impact, the key concepts that were cited from these RAs, and example citing articles referring to those concepts.

The findings help to understand the key concepts that define the IT business value domain as they have been taken up by subsequent research. Although all cited concepts revolve around the core theme of IT business value, multiple sub streams emerged from this analysis including IS success, the resource-based view of the firm, IT value (co-) creation, and IS strategy. Further, our analysis revealed several dimensions of how CAs cumulatively build on the knowledge developed in RAs. For example, some CAs adopt major parts of an individual RA's conceptual framework including concept relationships, some CAs synthesize and integrate concepts from multiple RAs, and other CAs explore new theoretical relationships by drawing on individual concepts of one



**Table 8**

Key concepts cited from the ITBV-RAs.

Review articles	Key concepts cited	Example citing articles
DeLone and McLean [94]	Information quality	[100, 117, 124]
	System quality	[100, 132, 153]
	IS use	[74, 81, 156]
	Individual impact	[108, 139, 147]
	Organizational impact	[112, 153, 156]
Melville, Kraemer and Gurbaxani [118]	IT resources	[86, 87, 92, 106]
	Organizational impact	[80, 143, 151]
	Industry characteristics	[77, 102, 143]
	Competitive position	[96, 111]
	IT investments	[98, 121]
	Business agility	[131]
	IT resources	[73, 96, 121, 123]
Wade and Hulland [150]	IT capabilities	[72, 87, 106, 116]
	IT assets	[122, 123, 141]
	Industry characteristics	[120]
Delone and McLean [95]	Business agility	[116]
	IT use	[74, 82, 146]
	User satisfaction	[76, 107, 148]
	System quality	[71, 85, 88]
	Service quality	[85, 105, 115]
Brynjolfsson [78]	Information quality	[85, 124, 139]
	Temporal lag effects	[125, 135]
Kohli and Devaraj [109]	Organizational performance	[89, 118]
	IT resources	[118]
	IT utilization	[133]
	IT payoff	[138]
	Value cocreation	[99, 101, 133]
Kohli and Grover [110]	IT capabilities	[116, 133]
	Organizational performance	[131]
	IT-strategy alignment	[154]
Piccoli and Ives [129]	IS strategy	[86, 97, 130]
	IT competencies	[83]
	IT resources	[114, 119]
	IT assets	[122]
	Business agility	[136]
Soh and Markus [145]	IT value creation	[80, 149, 156]
	IT value	[90]
	IT assets	[109]
	IT capabilities	[134, 152]

Note: Dataset references available in the supplementary online appendix.

RA.

Fig. 3 uses the literature that draws on the resource-based view of the firm as an illustrative example to map the development of concepts in that sub stream of the ITBV literature. The display presents the two illustrative RAs on the left and example CAs on the right. Each line represents the flow of one concept through the citation network. The lines start with the original concept of the RA on the left and end with the key concepts that the CA develops based on the original concept. Although not representative of entire body of literature on the resource-based view of the firm, the knowledge flow diagram illustrates several ways in which RAs contribute to the growth of knowledge. For example, it shows how the concept of IT resources and IT capabilities are central to this stream of literature (represented by the two thickest flows). These two concepts are integrated with many other concepts such as IT assets and IT investments. It also shows how some concepts in the CAs are very close to the original concept of the RA (e.g., IT investments, dynamic capabilities), whereas other ideas have been developed substantially (e.g., synergistic relationships, organizational agility, operational alignment). Apart from the meaning of concepts, we map how some concepts have been developed soon after the publication of the RA (e.g., resource complementarity), while other concepts took longer to be developed (e.g., digital business strategy, business process agility, operational alignment).

The nature of the relationship between those concepts can be inferred using a deeper qualitative study of the associated RA and CA. For

example, Chen et al. [87] propose and test several propositions involving business process agility by expanding the concepts of IT capabilities and organizational performance from Wade and Hulland [150] with how business processes enable IT resources from Melville et al. [118]. These propositions were developed by applying the existing rules of discourse surrounding IT capabilities, resources, and firm performance, and by drawing from various sources without extensive manipulation. Such use of the ideational dimension of citations is part of what Hassan and Serenko [24] call the conceptual citation pattern. In this way, various concepts are applied to add to the existing knowledge of ITBV.

## 6. Discussion

### 6.1. Implications for ideational impact classification

We developed a content-enriched deep learning approach to identify ideational impact and support researchers in searching for academic papers that have taken up ideas and concepts of prior work. The deep learning model we call Deep-CENIC applies NLP techniques to extract a range of features and combine those with a deep learning architecture using word embeddings based on the content of citing sentences. The Deep-CENIC model identifies ideational impact from a large amount of automatically extracted in-text citations at a high level of performance. As our results indicate, content-enriched word embeddings can represent the semantic meaning underlying the varied vocabularies and concepts referred to in citing sentences that may signal ideational impact. The Deep-CENIC model can capture ideational impact based on the actual content of citations, thus addressing the limitation of existing citation classification models, which are primarily based on derivative and meta-data-based features [25,50].

The ideational impact classification approach can be generalized to evaluate cumulative tradition in many other research genres, topic areas, and disciplines. In contrast to other approaches such as literature reviews or meta-analyses, Deep-CENIC can analyze large literature sets across different paradigms and theoretical models. It can therefore complement existing approaches that rely on hypothetico-deductive models and well-defined constructs [33]. Furthermore, it goes beyond traditional one-dimensional citation analyses based on count data [12,13,27] by taking the context and content of citing sentences into consideration. The proposed model could, therefore, be utilized to support analyses and decisions in a range of applications, including research rankings, scientometric analyses, and search tools based on citation data, among others. Research evaluation, for instance, in most institutions relies on citation indices (e.g., the h-indices and Journal Impact Factor), which are typically based on overall citation numbers [42]. Distinguishing ideational from non-ideational impact could foster the development of new measures of citation impact that are less susceptible to well-known weaknesses of citation analysis and more in line with actual knowledge development [3,9]. Furthermore, tools for academic literature searches have been dominated by two main approaches: a keyword-based search applying specific subject terms to reduce literature to a topic and a citation-based search starting from a key document to identify related literature citing the key document. This latter way of using citations as a search tool [22] is effective and benefits from the availability of readily accessible citation indices such as those provided by Google Scholar and Web of Science. Future research could build on our classifiers to implement new search tools that are based on ideational impact and enable scholars to find relevant research more efficiently.

### 6.2. Implications for IS research and the ITBV literature

We developed the Deep-CENIC approach in the context of ITBV-RAs, which synthesize a prominent research stream in the IS field. We explored how the identification of ideational impact citations can help

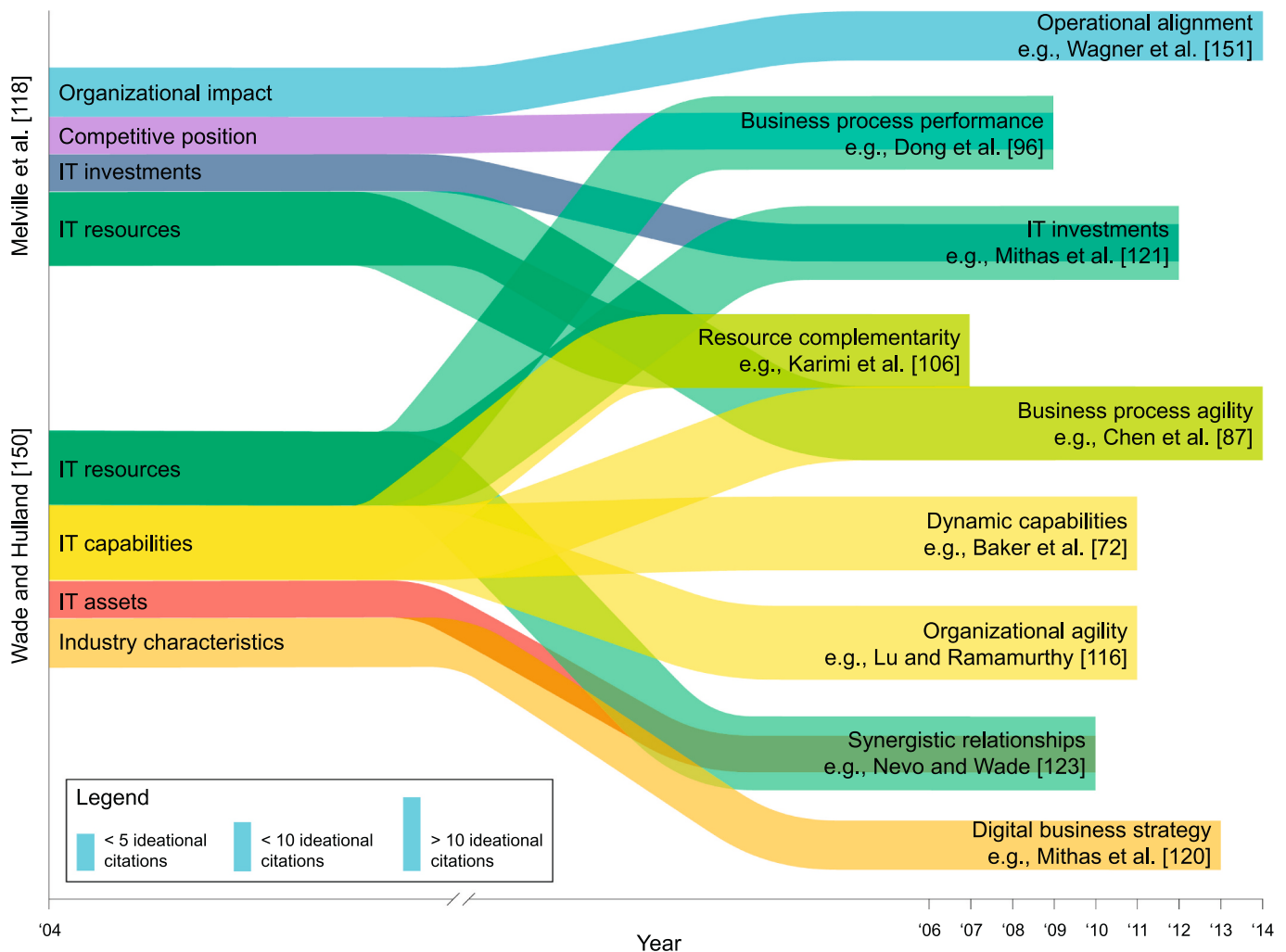


Fig. 3. Illustrative knowledge flow diagram (dataset references are available in the supplementary online appendix).

us trace the key concepts and ideas developed in a field, as they flow through expanding citation networks. By identifying the ideational impact of RAs, we showed their important role for concept development and the growth of knowledge of a field. While several methodologists and editors have argued for this quality of RAs [49,51], our study provides empirical evidence for RAs' "value for the field" [51 p. 242] in terms of knowledge growth. We show that, at least in IS research, RAs do not only serve as summaries of past accomplishments but help develop original knowledge contributions, thereby stimulating ideational impact. At the same time, our analysis demonstrated that ideational impact numbers are significantly different from overall citation numbers. Traditional citation count data is therefore insufficient as a proxy for ideational impact and unable to capture the meanings within the CAs based on the concepts adopted from the cited articles. Our results therefore caution against associating overall citations with knowledge impact.

After identifying CAs that communicate or elaborate on ideas and concepts imparted from and onto the cited text, we presented the RAs that have the highest ideational impact in the ITBV domain. Furthermore, filtering for those CAs that developed the original concepts and ideas further, we extracted the key concepts of the ITBV domain. With substantially less manual effort, our analysis of the qualitative citation patterns underlying the citation dataset revealed several different ways in which concepts have been used. Such automated approaches can therefore provide a valuable starting point for analyzing the growth of knowledge in the IS field. For example, they could inform debates

around its reliance on reference disciplines, its theoretical core, and the state of indigenous theories in the field.

The ideational impact analysis illustrates the potential of Deep-CENIC for supporting literature search and selection techniques. We do not suggest, however, to use Deep-CENIC as a substitute for hermeneutic research practices. The goal of this paper is not to encourage researchers to cite references without reading them, but to support researchers who want to know which references have taken up ideas and concepts of prior work. In fact, our paper directly addresses the problem of frivolous citing behavior because it can help to identify papers containing a large proportion of perfunctory citations. In turn, the illustrative analysis discussed in Section 5 is based on an in-depth manual qualitative analysis of citation contexts that required reading the full text of papers to interpret and make sense of the literature. The main advantage of Deep-CENIC lies in its potential for reducing the manual effort required by recommending academic literature that is most relevant. Thus, our aim in developing Deep-CENIC is not in automating the interpretation of researchers, but rather in supporting and augmenting those efforts by excluding literature that does not build on the concepts and ideas one is interested in.

### 6.3. Limitations and future directions

Our study has several limitations. First, we are among the first to explore the ideational dimension of citations. Although we developed a consistent understanding of what ideational impact means in the context

of our dataset, it may not be the only way of operationalizing ideational impact. We call for future research to develop further our conceptualization of ideational impact and to further define operational criteria for distinguishing citations. A more detailed understanding of the target variable will eventually lead to more robust classifiers and could lead to the development of multi-class classification that goes beyond a binary distinction of ideational and non-ideational impact.

Second, our Deep-CENIC model is based on a relatively simple deep learning architecture and limited set of derivative features. In the future, we plan to implement new features to increase classification performance and experiment with other classifiers to develop more robust models. Promising paths to develop new features include, for example, deep semantic features based on the full text of citing and cited papers. Such extensions could also lead to further automation and potential identification of concepts and concept relationships required for knowledge flow analyses [6,23].

Third, our Deep-CENIC model is trained on a particular domain. Although we tested the robustness of our proposed model on a broader set of publication outlets including major IS journals and conferences, our dataset is still limited to RAs on ITBV. We are confident that our approach can be generalized to identify ideational impact in a range of genres and topics; however, similar to comparable approaches [e.g., 34], this would require training data specific to new application domains. Future studies could develop such datasets for other research genres, topic areas, and disciplines to compare ideational impact across different research streams.

## 7. Conclusion

Identification of ideational impact is a critical issue when evaluating research impact, searching the literature, and analyzing citation data. Our objective was to develop an automated approach for classifying ideational impact and to design a system that supports researchers in searching for academic papers. We developed a high-performance deep learning model (Deep-CENIC) that considers both content-based word embeddings and syntactic, semantic, and contextual derivative features to identify citations symbolizing the impact of ideas and concepts from the cited paper. Evaluation results show that our Deep-CENIC model outperforms all the baseline models in classifying ideational impact citations. Success in developing approaches able to address the task of reliable, large-scale identification of ideational impact for papers, research streams or even entire disciplines – as this paper found possible for RAs in the ITBV domain – has the potential to make the process of evaluating the cumulative tradition of entire disciplines reproducible and transparent. Hence, this study could help achieve what Merton suggested in his early works on the science of science: “Having access to cumulative opportunity for scholarly work is one thing; seizing that opportunity and putting it to effective use is quite another” [41 p. 93].

## Notes

1. Although the terms citing sentence and citation sentence are often used interchangeably in the scientometric literature, in this paper, we refer to the sentences that contain citations as citing sentences to avoid confusion around the legal term “citation sentence” [8 p. 3–4].

2. Our code and dataset are available at <https://github.com/julianprester/deep-cenic>.

## Acknowledgements

We are grateful for the comments received during presentations of earlier versions of this paper at the University of Regensburg, University of New South Wales, and the International Conference on Information Systems. We also thank the editor and reviewers for their invaluable comments and suggestions. The research is supported by a grant of the German Science Foundation (DFG) for the research project

“Epistemological Advances Through Qualitative Literature Reviews in Information Systems Research”.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dss.2020.113432>.

## References

- [1] P.J. Ågerfalk, Insufficient theoretical contribution: a conclusive rationale for rejection? *Eur. J. Inf. Syst.* 23 (6) (2014) 593–599.
- [2] A. Athar, Sentiment analysis of citations using sentence structure-based features, in: *Proceedings of the 49th Meeting of the Association for Computational Linguistics*, Portland, OR, 2011.
- [3] S. Bhattacharjee, J.R. Marsden, H. Singh, An approach to identify influential building blocks and linkages in an information resource network, *Decis. Support. Syst.* 52 (1) (2011) 217–231.
- [4] L. Bornmann, H.-D. Daniel, What do citation counts measure? A review of studies on citing behavior, *J. Doc.* 64 (1) (2008) 45–80.
- [5] L. Bornmann, R. Mutz, C. Neuhaus, H.-D. Daniel, Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results, *Ethics Sci. Environ. Polit.* 8 (1) (2008) 93–102.
- [6] L. Bornmann, K.B. Wray, R. Haunschild, Citation concept analysis (CCA): a new form of citation analysis revealing the usefulness of concepts for other researchers illustrated by exemplary case studies including classic books by Thomas S. Kuhn and Karl R. Popper, *Scientometrics* 122 (2) (2020) 1051–1074.
- [7] D.O. Case, G. Higgins, How can we investigate citation behavior? A study of reasons for citing literature in communication, *J. Am. Soc. Inf. Sci.* 51 (7) (2000) 635–645.
- [8] Columbia Law Review, Harvard Law Review, University of Pennsylvania Law Review, Yale Law Journal, The Bluebook: A Uniform System of Citation, The Harvard Law Review Association, Cambridge, MA, 2015.
- [9] Y. Dang, Y. Zhang, P.J.-H. Hu, S.A. Brown, H. Chen, Knowledge mapping for rapidly evolving domains: a design science approach, *Decis. Support. Syst.* 50 (2) (2011) 415–427.
- [10] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, C. Zhai, Content-based citation analysis: the next generation of citation analysis, *J. Am. Soc. Inform. Sci. Technol.* 65 (9) (2014) 1820–1833.
- [11] C. Dong, U. Schäfer, Ensemble-style self-training on citation classification, in: *Proceedings of the 5th International Joint Conference on Natural Language Processing*, Chiang Mai, Thailand, 2011, pp. 623–631.
- [12] S.B. Eom, Mapping the intellectual structure of research in decision support systems through author cocitation analysis (1971–1993), *Decis. Support. Syst.* 16 (4) (1996) 315–338.
- [13] S.B. Eom, S.M. Lee, J.K. Kim, The intellectual structure of decision support systems (1971–1989), *Decis. Support. Syst.* 10 (1) (1993) 19–35.
- [14] E. Garfield, Citation indexes for science: a new dimension in documentation through association of ideas, *Science* 122 (3159) (1955) 108–111.
- [15] E. Garfield, Citation analysis as a tool in journal evaluation, *Science* 178 (4060) (1972) 471–479.
- [16] Y. Goldberg, A primer on neural network models for natural language processing, *J. Artif. Intell. Res.* 57 (2016) 345–420.
- [17] V. Grover, K. Lyytinen, R. Weber, Panel on native IS theories, in: *Special Interest Group on Philosophy and Epistemology in IS (SIGPHIL)*, Orlando, FL, 2012.
- [18] Y. Guan, Q. Wei, G. Chen, Deep learning based personalized recommendation with multi-view information integration, *Decis. Support. Syst.* 118 (2019) 58–69.
- [19] C. Guo, Y. Yu, A. Sanjari, X. Liu, Citation role labeling via local, pairwise, and global features, in: *Proceedings of the 77th Annual Meeting of the American Society for Information Science and Technology*, Seattle, WA, 2014.
- [20] C. Hart, *Doing a Literature Review: Releasing the Social Science Research Imagination*, SAGE Publications, London, UK, 1999.
- [21] N. Harwood, An interview-based study of the functions of citations in academic writing across two disciplines, *J. Pragmat.* 41 (3) (2009) 497–518.
- [22] N.R. Hassan, C. Loebbecke, Engaging scientometrics in information systems, *J. Inf. Technol.* 32 (1) (2017) 85–109.
- [23] N.R. Hassan, J. Prester, G. Wagner, Seeking out clear and unique Information Systems concepts: a natural language processing approach, in: *Proceedings of the 28th European Conference on Information Systems*, Marrakech, Morocco, 2020.
- [24] N.R. Hassan, A. Serenko, Patterns of citations for the growth of knowledge: a Foucauldian perspective, *J. Doc.* 75 (3) (2019) 593–611.
- [25] S.-U. Hassan, M. Imran, S. Iqbal, N.R. Aljohani, R. Nawaz, Deep context of citations using machine-learning models in scholarly full-text articles, *Scientometrics* 117 (3) (2018) 1645–1662.
- [26] S.-U. Hassan, I. Safder, A. Akram, F. Kamiran, A novel machine-learning approach to measuring scientific knowledge flows using citation context analysis, *Scientometrics* 116 (2) (2018) 973–996.
- [27] C.W. Holsapple, L.E. Johnson, H. Manakyan, J. Tanner, An empirical assessment and categorization of journals relevant to DSS research, *Decis. Support. Syst.* 14 (4) (1995) 359–367.
- [28] C. Jochim, H. Schütze, Towards a generic and flexible citation classifier based on a faceted classification scheme, in: *Proceedings of the 24th International Conference on Computational Linguistics*, Mumbai, India, 2012.

- [29] D. Jurafsky, J.H. Martin, *Speech and Language Processing*, Pearson, Upper Saddle River, NJ, 2009.
- [30] D. Jurgens, S. Kumar, R. Hoover, D. McFarland, D. Jurafsky, Measuring the evolution of a scientific field through citation frames, *Trans. Assoc. Comput. Linguist.* 6 (2018) 391–406.
- [31] P.G. Keen, MIS research: reference disciplines and a cumulative tradition, in: *Proceedings of the 1st International Conference on Information Systems*, Philadelphia, PA, 1980.
- [32] K.D. Knorr-Cetina, *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*, Pergamon Press, Oxford, UK, 1981.
- [33] K.R. Larsen, C.H. Bong, A tool for addressing construct identity in literature reviews and meta-analyses, *MIS Q.* 40 (3) (2016) 529–551.
- [34] K.R. Larsen, D. Hovorka, A. Dennis, J.D. West, Understanding the elephant: the discourse approach to boundary identification and corpus construction for theory review articles, *J. Assoc. Inf. Syst.* 20 (7) (2019) 887–927.
- [35] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [36] B.A. Lipetz, Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators, *Am. Doc.* 16 (2) (1965) 81–90.
- [37] M.G. Lozano, J. Brynielsson, U. Franke, M. Rosell, E. Tjörnhammar, S. Varga, V. Vlassov, Veracity assessment of online data, *Decis. Support. Syst.* 129 (2020).
- [38] C. Lu, Y. Ding, C. Zhang, Understanding the impact change of a highly cited article: a content-based citation analysis, *Scientometrics* 112 (2) (2017) 927–945.
- [39] K. Lyytinen, J.L. King, Nothing at the center? Academic legitimacy in the information systems field, *J. Assoc. Inf. Syst.* 5 (6) (2004) 220–246.
- [40] R.K. Merton, Priorities in scientific discovery: a chapter in the sociology of science, *Am. Sociol. Rev.* 22 (6) (1957) 635–659.
- [41] R.K. Merton, J. Gaston, *The Sociology of Science in Europe*, Southern Illinois University Press, Chicago, IL, 1977.
- [42] J. Mingers, L. Leydesdorff, A review of theory and practice in scientometrics, *Eur. J. Oper. Res.* 246 (1) (2015) 1–19.
- [43] J. Mingers, F. Xu, The drivers of citations in management science journals, *Eur. J. Oper. Res.* 205 (2) (2010) 422–430.
- [44] R. Mitkov, *The Oxford Handbook of Computational Linguistics*, Oxford University, Oxford, UK, 2005.
- [45] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments, *Decis. Support. Syst.* 53 (4) (2012) 675–679.
- [46] M. Moravcsik, Life in a multidimensional world, *Scientometrics* 6 (2) (1984) 75–85.
- [47] M. Moravcsik, P. Murugesan, Some results on the function and quality of citations, *Soc. Stud. Sci.* 5 (1) (1975) 86–92.
- [48] D.E. O'Leary, The relationship between citations and number of downloads in Decision Support Systems, *Decis. Support. Syst.* 45 (4) (2008) 972–980.
- [49] C. Post, R. Sarala, C. Gatrell, J.E. Prescott, Advancing theory with review articles, *J. Manag. Stud.* 57 (2) (2020) 351–376.
- [50] F. Qayyum, M.T. Afzal, Identification of important citations by exploiting research articles' metadata and cue-terms from content, *Scientometrics* 118 (1) (2019) 21–43.
- [51] F. Rowe, What literature review is not: diversity, boundaries and recommendations, *Eur. J. Inf. Syst.* 23 (3) (2014) 241–255.
- [52] D. Schall, A multi-criteria ranking framework for partner selection in scientific collaboration environments, *Decis. Support. Syst.* 59 (2014) 1–14.
- [53] G. Schryen, Revisiting IS business value research: what we already know, what we still need to know, and how we can get there, *Eur. J. Inf. Syst.* 22 (2) (2013) 139–169.
- [54] G. Schryen, G. Wagner, A. Benlian, G. Paré, A knowledge development perspective on literature reviews: validation of a new typology in the IS field, *Commun. Assoc. Inf. Syst.* 46 (2019) 134–186.
- [55] T. Silva, Z. Guo, J. Ma, H. Jiang, H. Chen, A social network-empowered research analytics framework for project selection, *Decis. Support. Syst.* 55 (4) (2013) 957–968.
- [56] H. Small, Cited documents as concept symbols, *Soc. Stud. Sci.* 8 (3) (1978) 327–340.
- [57] H. Small, On the shoulders of Robert Merton: towards a normative theory of citation, *Scientometrics* 60 (1) (2004) 71–79.
- [58] L.C. Smith, Citation analysis, *Libr. Trends* 30 (1) (1981) 83–106.
- [59] J.M. Swales, Citation analysis and discourse analysis, *Appl. Linguist.* 7 (1) (1986) 39–56.
- [60] S. Teufel, A. Siddharthan, D. Tidhar, An annotation scheme for citation function, in: *Proceedings of the 7th SIGDIAL Workshop on Discourse and Dialogue*, London, UK, 2009, pp. 80–87.
- [61] D. Truex, M. Cuellar, H. Takeda, R. Vidgen, The scholarly influence of Heinz Klein: ideational and social measures of his impact on IS research and IS scholars, *Eur. J. Inf. Syst.* 20 (4) (2011) 422–439.
- [62] M. Valenzuela, V. Ha, O. Etzioni, Identifying meaningful citations, in: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, San Francisco, CA, 2015, pp. 21–26.
- [63] G. Wagner, J. Prester, M.P. Roche, A. Benlian, G. Schryen, Factors affecting the scientific impact of literature reviews: a scientometric study, in: *Proceedings of the 37th International Conference on Information Systems*, Dublin, Ireland, 2016.
- [64] X. Wan, F. Liu, Are all literature citations equally important? Automatic citation strength estimation and its applications, *J. Assoc. Inf. Sci. Technol.* 65 (9) (2014) 1929–1938.
- [65] Y. Wang, W. Xu, Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud, *Decis. Support. Syst.* 105 (2018) 87–95.
- [66] R. Weber, Editor's comments: theoretically speaking, *MIS Q.* 27 (3) (2003) iii–xii.
- [67] R. Weber, Evaluating and developing theories in the information systems discipline, *J. Assoc. Inf. Syst.* 13 (1) (2012) 1–30.
- [68] Y. Xu, X. Guo, J. Hao, J. Ma, R.Y. Lau, W. Xu, Combining social network and semantic concept analysis for personalized academic researcher recommendation, *Decis. Support. Syst.* 54 (1) (2012) 564–573.
- [69] G. Zhang, Y. Ding, S. Milojević, Citation content analysis (CCA): a framework for syntactic and semantic analysis of citation content, *J. Am. Soc. Inf. Technol.* 64 (7) (2013) 1490–1503.
- [70] X. Zhu, P. Turney, D. Lemire, A. Vellino, Measuring academic influence: not all citations are equal, *J. Assoc. Inf. Sci. Technol.* 66 (2) (2015) 408–427.

**Julian Prester** is a Scientia PhD scholar at the University of New South Wales. His research focuses on digital work, digital nomadism, and the theories and methods used by researchers in the Information Systems discipline. His work has been published in international journals, including *Communications of the Association for Information System and Computers & Security*, and conferences such as the *International Conference on Information Systems*. He regularly serves as a reviewer for IS journals and conferences.

**Gerit Wagner** is a postdoctoral fellow at HEC Montréal. He is interested in literature reviews, the impact of research methods, and digital platforms for knowledge work. His work has been published in international journals, including *Decision Support Systems*, *Computers & Security*, *Communications of the Association for Information Systems* and conferences such as the *International Conference on Information Systems*. He is a member of the Association for Information Systems and he regularly serves as a reviewer for information systems journals and conferences.

**Guido Schryen** is a full professor and holds the Chair of Management Information Systems and Operations Research at Paderborn University, Germany. His research interests cover quantitative decision support, benefits of information systems and services, literature reviews, and IT security. He has published both quantitative and qualitative research in international journals, including *European Journal of Information Systems*, *Communications of the Association for Information Systems*, *European Journal of Operational Research*, *Computers & Security*, *OR Spectrum*, *Communications of the Association for Computing Machinery*, and *Business & Information Systems Engineering*. He is Associate Editor of *Business & Information Systems Engineering*.

**Nik Rushdi Hassan** is Associate Professor of Information Systems (IS) and Chair of the Management Studies Department at the Labovitz School of Business and Economics (LSBE), University of Minnesota Duluth, USA. He is currently Associate Editor for the History and Philosophy Department of the *Communications of the AIS* and Senior Editor of *Data Base Advances in Information Systems*. He is one of the Editors of a recent special issue of the *European Journal of Information Systems* on Philosophy and the Future of the IS Field and has published in the *Journal of the Association for Information Systems*, *Journal of Information Technology*, *European Journal of Information Systems*, *Data Base Advances in Information Systems*, *Information Systems Management Journal*, *Communications of the AIS*, *Journal of IS Education*, *Informing Science Journal*, *Review of Accounting and Finance*, *Journal of Documentation* and in *The 2018 Routledge Companion to Management Information Systems*.