A Semi-Auto Text Mining Approach for Literature Review: An Example From IT for Entrepreneurship

Yousra Harb, Yarmouk University, Jordan*
Yanyan Shang, University of Tampa, USA

ABSTRACT

The emergence of powerful information technologies (IT) has changed innovation and entrepreneurship in significant ways. Research in IT for entrepreneurship is relatively new, and there is a growing interest from academics for further studies on investigating this area of research. This study reviews research carried out in the domain of IT for entrepreneurship. A total of 1,005 papers, published between 1980 and 2021, were used to uncover the latent topics addressed in this domain. A topic modeling (LDA) algorithm was used to automate the process of extracting the initial research topics from the data. The results show that IT for entrepreneurship studies are classified into six aspects of research: entrepreneurship initiative and innovation, strategy, business process management and operation management, entrepreneurship education, industry analysis, and business model. The results raise awareness of IT-associated entrepreneurship areas of research, provide useful insights for future research, and inform practice in this domain of study.

KEYWORDS

Digital Entrepreneurship, Entrepreneurship, IT, Literature Review, Topic Modeling

INTRODUCTION

Entrepreneurship, defined as the process of innovation that brings new products and services to the market (Bandera, Bartolacci, & Passerini, 2016), has drawn increasing attention in the Knowledge Management (KM) literature. Most KM literature on entrepreneurship has focused on entrepreneurial activities in the late stages of the KM process. However, a study in the field of KM and entrepreneurship shows that knowledge socialization is the most mature knowledge stage (Bandera, Bartolacci, & Passerini, 2016). We are interested in the early stages of the KM process, and we investigated how the innovation enablers affect the field of entrepreneurship. Waves of technological changes bring fascinating opportunities for innovators and entrepreneurs (Steininger, 2019; Secundo et al., 2020). Advances in information technology also serve as a significant enabler for innovation and economic growth (Boeker et al., 2019). Entrepreneurs are one of the change agents, and they play a vital role in economic growth based on information technologies and information communication technologies.
The reasons for adopting IT are to generate growth, stay competitive, and enhance innovation abilities (Nguyen, 2009). The business model is one of the main tools for absorbing technological perspectives of innovations. Particularly, entrepreneurs design and implement new business models through the enabling of information technology capabilities (Rai & Tang, 2014). In fact, businesses that are based on IT use different business models. This should be considered, otherwise the business may miss profitable opportunities (Hafezieh et al., 2011). IT is also a form of digital technologies used by organizations which play an essential role in providing means to enhance information and knowledge capture and sharing, lower the cost of production and labor, and add value to products and services and thus enhance the quality (Nguyen, 2009; Olsson and Bernhard, 2020). In the era of the knowledge economy, IT and ICT could be used as a source of knowledge by firms to enhance their performance (Javed et al., 2020). The literature also highlighted that digital technologies (e.g., such as big data, business intelligence, cloud computing, mobile services, and internet of things) facilitate entrepreneurial activities and significantly contribute to the digital economy due to the ability of technologies to support create activities (Abubakre et al., 2020). By adopting IT and digital technologies, the firm also initiates new ways of doing business such as digital business models and digital entrepreneurship (Elia et al., 2020).

The crucial difference between digital and traditional entrepreneurship is in the way the entrepreneur makes and markets a product (a good or a service) (Hafezieh et al., 2011). Specifically, digital entrepreneurship relies on digital media tools and IT in the pursuit of entrepreneurial prospects (Giones, 2017). The degree of digitalization is reflected in the digital nature of the product, digital marketing, digital distribution, digital interactions, and in the internal activities of a firm (Hull et al., 2007). Moreover, in this context, IT was also proven to influence the organizational design in terms of job design through supporting the autonomy within the organization and thus influencing employees’ attitudes toward involvement in entrepreneurial activities; lateral linkage design concerns with enhancing analytical and design capabilities; and decision-making design by taking the advantage of both centralization and decentralization design (Muñoz et al., 2016).

In this study, we attempt to uncover the directions of IT-associated entrepreneurship by analyzing the titles and abstracts listed in 1005 publications from 1980 to 2021. This study is different from the traditional Systematic Literature Review (SLR) in several ways. Firstly, the traditional SLR approach relays on the researchers’ subjective opinions in selecting the included studies. This study proposed a method that utilizes a more objective approach by pre-defining the explicit criteria to retrieve qualified studies. Secondly, due to the aforementioned limitation of researchers manually reading articles, the traditional SLR approach includes a fairly small number of articles, usually a couple to a few hundred (e.g., Steininger, 2019). Our proposed semi-auto approach can ensure a better coverage of literature by reviewing over thousands of studies.

The remainder of the paper is organized as follows: Section 2 discusses the background knowledge and existing literature. The research method is then illustrated in Section 3. Section 4 presents the model results and findings. Section 5 provides discussion and analysis, while the final Sections conclude and detail an agenda for future study.

BACKGROUND

The nature of IT-associated entrepreneurship (such as digital entrepreneurship) enables the “entrepreneurial opportunities being created and pursued through the use of technological platforms and other information communicating equipment” (Antonizzi & Smuts, 2020). As technology is cultivated and advances, IT-associated entrepreneurship may evolve into many categories, such as exploiting new business opportunities in novel information technologies, identifying new business models caused by the development and advancement of IT, and creating or updating organizations (ICT) since it may generate new opportunities for development (Elia, Margherita, & Passiante, 2020; Steininger, 2019; Olsson & Bernhard, 2020).
with advanced IT utilization. More and more attention has been drawn to how IT serves as an enabler for entrepreneurial activities (von Briel, Davidsson, & Recker, 2018). For example, Kuester, Konya-Baumbach, & Schuhmacher (2018) studied the adoption of Internet-enabled service innovations for start-ups. Elia, Margherita and Passiante (2020) investigated how digital technologies reshape the entrepreneurial process and proposed a descriptive framework of a digital entrepreneurship ecosystem. Due to the importance of information technology to entrepreneurs and the great advantages it offers them, a holistic perspective about the research streams of IT for entrepreneurship is needed to further understand this area of research and provide new directions for research about the link between IT and entrepreneurship.

A systematic review of academic literature in a specific area can be undertaken with the purpose of summarizing the empirical evidence of the benefits and limitations of a method, identifying research gaps to suggest areas for further investigation, and/or providing a framework to appropriately position new research activities (Kitchenham, 2004). A systematic review synthesizes existing work in a manner that is fair and thorough, which provides important insights about the effects of some phenomenon across a wide range of settings and empirical methods (Kitchenham, 2004). An SLR about entrepreneurship for several contexts already exists: these include entrepreneurship as practice (Claire et al., 2019), entrepreneurial ecosystems (Cao and Shi, 2020), academic entrepreneurship intentions (Neves & Brito, 2020), women entrepreneurship (Rashid & Ratten, 2020), and knowledge management and social entrepreneurship (Maalaoui et al., 2020). However, there has been little done in recent years to review the status of the literature on IT-associated entrepreneurship.

We carefully searched for systematic literature review studies on IT-associated entrepreneurship. Steininger (2019) conducted a systematic review of information systems, entrepreneurship, and general and strategic management literature using 292 articles to provide an overview of IT and entrepreneurship research. Other research studies have a more specific focus. For instance, Zaheer, Breyer, and Dumay (2019) provided a structured review of digital entrepreneurship literature by analyzing 133 articles in this discipline. Olanrewaju et al. (2020) systematically reviewed 160 articles published between 2002 and 2018 in the social media and entrepreneurship domain. Secundo et al. (2020) also conducted a structured literature review about the influence of social media technologies on entrepreneurship activities and processes. These studies provided excellent insights in a smaller scope and more specific area within the IT-associated entrepreneurship field. The articles that were reviewed in these studies are less than 300. Our study is the first attempt to survey a large corpus of IT and entrepreneurship articles. In addition, it employed a semi-auto approach, a topic modeling algorithm, in conducting this literature review.

Using a text mining approach to review literature is not an entirely new idea. Several research studies reviewed literature in various domains using article metadata. In management information systems, Delen and Crossland (2008) applied a text mining approach to analyze the titles and abstracts of 1123 articles. Abu-Shanab and Harb (2019) conducted text mining on article keywords in the e-government area of research. Harb and Abu-Shanab (2020) employed text mining (LDA topic modeling) on article title, abstract, and keywords in a knowledge management context. Bragge et al. (2012), analyzed articles’ keywords using text mining in the multiple criteria decision-making area of research. In the field of operation research and management science, Gatti, Brooks, and Nurre (2015) used LDA topic modeling to uncover the latent topics of abstracts from 37 related journals. Worth noting, abstract data contains a high density of words, therefore making it more appropriate for an LDA algorithm (Gatti et al., 2015). Given the high volume of research articles returned by hitting online databases, reading and open coding the content of each paper is a challenging and intensive labor task (Moro, Cortez, & Rita, 2015). Therefore, automating the process of reading and open coding the content enable the research to extract useful topics from large data without being contaminated by any preconceptions (Yu, Jannasch-Pennells, & DiGangi, 2011).

The main goal of this paper is therefore to uncover the directions of IT-associated entrepreneurship by analyzing the titles and abstracts listed in 1005 publications from 1980 to 2021. By deploying a
topic modeling algorithm, our study can encompass a large number of articles and perform a thorough literature review.

METHOD

According to Webster and Watson (2002), a literature review of prior studies is a critical step to analyze a given subject and synthesize previous research. Such an approach would strengthen a particular field of study and provide important clues on current and future research directions. In fact, it is an essential approach to identify trends of research that can unveil new studies and discoveries (Levy & Ellis, 2006).

This study followed the structure of an SLR in terms of searching databases and the selection of articles. However, we employed an automated text mining approach to analyze the research area of entrepreneurship and IT, identifying the research trends and growth of knowledge, and providing directions for future research in this promising research area. The main research questions of this study can be expressed as follows:

**RQ1:** What are the main research topics within information technology for the entrepreneurship literature?

**RQ2:** What are the future research directions in the field of information technology for entrepreneurship?

The first research question aims to define the current status in the literature about the research streams of IT for entrepreneurship. This would provide an improved understanding about the role of IT in the field of entrepreneurship. The second research question provides insights for researchers in the field of IT for entrepreneurship.

Afterward the next step is to define the research protocol that involves the identification of the literature sources, the methods, and the tools used for the analysis of retrieved articles. Figure 1 presents the complete process of the literature review.

The following sections detail the process of conducting this review.

---

**Figure 1. Process of the literature review**

- **Formulating the research questions**
  - **RQ1.** What are the main research topics within information technology for entrepreneurship literature?
  - **RQ2.** What are the future research directions in the field of information technology for entrepreneurship?

- **Searching databases**
  - Sources: Web of Science, Scopus, EBSCO, and ABI
  - Keywords identification: ("entrepreneur**" AND "computerized information") OR ("entrepreneur**" AND "Information system") OR ("entrepreneur**" AND "information technology") OR ("entrepreneur**" AND "information and communication technology") OR ("entrepreneur**" AND "computer* technology") OR ("entrepreneur**" AND "infotech") OR ("entrepreneur**" AND "computer-based informat**") OR ("entrepreneur**" AND "computer-human")

- **Selection of publications**
  - Time frame: 1980-2021
  - Document type: Articles
  - Exclusion criteria: Articles published in conferences, books, book chapters, theses, and reports

- **Analysis of content**
  - Extract titles and abstracts of the collected articles
  - Use textmining: LDA algorithm
  - Name the discovered topics
  - Output: Aspects of IT-associated Entrepreneurship
Searching Databases and Selection of Publications

Interrelating the information technology field with entrepreneurship is an interdisciplinary topic and requires cross-domain work in the areas of information technology, information systems, and entrepreneurship to investigate the use of information technology for business management, and four prominent databases that include business management journals were used to source literature: Web of Science, Scopus, EBSCO, and ABI. We included only refereed journal articles in the search result. This approach is consistent with previous studies on entrepreneurship that have included only journal articles (e.g., Olanrewaju et al., 2020).

To investigate IT for entrepreneurship, the selection of the paper was performed by searching the keywords (“information technology”, and “entrepreneurship”). However, we noticed that studies in this domain also use other IT-related terms such as information systems, information communication technology (ICT), or computerized information. Therefore, we included keywords reflecting information technology terms. The final search query was:

((“entrepreneurship” AND “computerized information”) OR (“entrepreneurship” AND “Information system”)) OR (“entrepreneurship” AND “information technology”) OR (“entrepreneurship” AND “information and communication technology”) OR (“entrepreneurship” AND “computer technology”) OR (“entrepreneurship” AND “infotech”) OR (“entrepreneurship” AND “computer-based information”) OR (“entrepreneurship” AND “computer-human”).

We limited the search to the abstract to have focused results and retrieve relevant articles. As stated, only journal articles published in English were selected. In line with prior SLR studies, we excluded articles published in conferences, books, book chapters, theses, and reports (Olanrewaju et al., 2020). This resulted in 408 hits in Web of Science, 596 hits in Scopus, 342 articles from EBSCO, and 305 articles from the ABI database. Through manual screening and after removing duplicate articles, the shortlist contained 1005 articles. The time period under investigation is between 1980 and 2021. Figure 2 presents the distribution of articles by year.

Analysis of the Content: Topic Modeling

As the number of articles is large, extracting useful information by manual coding might be a time-consuming and tedious task (Harb, Shang, & Al-musa, 2020). Hence, text mining was utilized to
facilitate the process of extracting useful topics from the literature. In this approach, we adopted a topic modeling (LDA) algorithm to unveil hidden topics from a collection of documents (Blei, Ng, & Jordan, 2003). LDA is a generative probabilistic topic modeling algorithm that outperforms over other well-known topic models.

The preprocessing of the data included the following transformations: (1) we removed the stop words and we also extended the stop words to include some other noisy words such as (‘among’, ‘use’, ‘subject’, ‘edu’, ‘use’, ‘every’, ‘also’, ‘one’). Moreover, to accurately define the research topics using LDA we excluded the following keywords (“entrepreneur*, technolog*, and informat*). These keywords were excluded in order to obtain more focused topics. We noticed these keywords happened largely in the returned topics by the algorithm as they were used to search for articles in the field of IT for entrepreneurship. (2) Lemmatization was performed on the documents, and (3) The documents were also converted to lower case. We selected the Latent Dirichlet Allocation (LDA) model from the Gensim package available in Python software as it is one of the most important probabilistic used algorithms for topic identification (Moro et al., 2015). LDA is a generative probabilistic topic model from a collection of documents. In this generative process, each document is modeled as a multinomial distribution over k topics and it models each topic as a multinomial distribution over vocabulary V.

Choosing the right number of topics in LDA is an important task. Literature has proposed some measures to evaluate the quality of the generated topics such as predictive likelihood and coherence measures (Syed & Spruit, 2017; Wallach, Murray, Salakhutdinov, & Mimno, 2009). Past studies have shown the predictive likelihood may not yield interpretable topics from human judgment (Syed & Spruit, 2017). Moreover, when the generated topics are used to understand the trends or the development of a particular field of study, it is important to apply measures that serve this purpose (Syed & Spruit, 2017). Thus, to serve the motivation of this study that pertains to understanding the development of the IT/IS and entrepreneurship field, we used topic coherence measures to uncover the coherence of a topic. Each generated topic consisted of words, and the topic coherence measures score a topic by measuring the semantic similarity of the top N words in the topic.

Creating LDA Models

The following procedures were followed to create LDA models in this study: (1) sensitivity tests were performed to determine the model number of topics (k). These tests were performed in sequence over a validation corpus set, (2) for every LDA model created (541 in total), the coherence score was calculated. To determine the optimal number of topics, we investigated the coherence scores that correspond to varying numbers of topics. In line with the previous studies (e.g., Wallach et al., 2009), we selected the model that yielded the highest coherence score (see Figure 3). In this case, we chose k=14. We then trained the final model using the selected k.

Labeling Process

The process of selecting a representative name to interpret a list of words under each topic generated by LDA is important to specify what constitutes the homogeneity of a topic (Harb and Abu-Shanab, 2020). In the literature, assigning labels to topics is usually a manual process, especially when labeling demands domain knowledge to achieve high labeling quality (Chang et al., 2009; Harb et al., 2020). In the present study, the generated topics were manually labeled based on the association of words in each generated topic. To ensure the quality of the labeling process and avoid bias, the authors independently labeled the topics learned. We agreed to use certain criteria in the naming topics process such as the association of the words in a topic and the frequency of the words in the original dataset. We found a substantial agreement in the results of the naming topic process. The first topics that resulted from the LDA algorithm were used to illustrate the labeling process:

Innovation and knowledge management (innovation, system, management, knowledge, new, business, factor, development, firm, based).
In this topic, the ten retrieved words were used to find out the theme of this topic and assign a representative title to it. We found the words (innovation, knowledge, management, new, development) are all related to innovation and knowledge management in business. Moreover, looking into the retrieved articles in the dataset that were used to discover the topics by LDA, we found 89 articles addressed the innovation topic and 45 articles investigated the knowledge management-related area of research.

Thus, the aforementioned topic was given the “Innovation and knowledge management” label. It is worth mentioning that the inspection of the generated topics is not mutually exclusive. Specifically, some words appear in more than one topic and thus some existing topics reveal some overlaps (Abu-Shanab & Harb, 2019).

RESULTS

This section reports and discusses the results of the LDA topic modeling. The LDA topic modeling conducted in the previous section resulted in a list of top words, whose latent meanings can be regarded as topics. A list of top words per topic is provided, as well as a score that indicates the weighted presence of the top word per topic (topic loading). Table 1 shows the top words generated from the topic modeling and the topics that are extracted from the titles and abstracts of the selected literature. With removing the keywords that were used to search for the articles, the topics are focused on different aspects of IT-associated entrepreneurship. As such, we have a better understanding on what types of research have been done in the field.

14 topics discovered from the topic modeling were mapped to 6 aspects of IT-associated entrepreneurship research. To generalize the topics to different aspects of IT-associated entrepreneurship, we go through the meaning of each topic, then assign the closely related topics to the same category. For example, topics “innovation and knowledge management”, “innovation and business development”, and “initiative and technology adoption” are focusing on topics that related to starting up a new business or creating a new method/business model for business development or improvement; therefore, we generalize these topics to the entrepreneurship initiative and innovation aspects of IT-associated entrepreneurship.
ANALYSIS

This section discusses the findings and implications of our study. As shown in Figure 4, entrepreneurship initiative and innovation, business model, and business process management and operation management have received considerable attention by previous literature in IT-associated entrepreneurship, whereas entrepreneurship education, strategy, and industry analysis are not well-covered areas of research in the domain.

Entrepreneurship Initiative and Innovation

About 26% of the selected literature investigates areas about IT-associated entrepreneurship initiative and innovation. Entrepreneurship initiative refers to the ability to turn ideas into action. “It includes the ability to plan and manage projects in order to achieve objectives, but also creativity, innovation, and risk taking” (Ribeiro, Santos, & Martins, 2019). There are general two ways to initiate an entrepreneurial project: creating new ideas or adopting technologies or methods for imitating the new combination of production already available (Relivingmbadays, 2012). We found that literature in this area was largely interested both in the adoption of information technology (IT) and information and communication technology (ICT). For example, studies (Asongu and Tchamyou, 2016; Elshaiekh,
Further, under the entrepreneurship initiative and innovation research area, the number of research studies published on knowledge management (KM) and entrepreneurship has been increasing significantly. In addition, recent research studies pay considerable attention to the

<table>
<thead>
<tr>
<th>Aspect of IT-associated Entrepreneurship</th>
<th>Weight</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entrepreneurship initiative and innovation</td>
<td>26.70%</td>
<td>Innovation and knowledge management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Innovation and business development</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Initiative and Technology adoption</td>
</tr>
<tr>
<td>Strategy</td>
<td>14.30%</td>
<td>Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Entrepreneurial strategic orientation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Labor market and policies</td>
</tr>
<tr>
<td>Business process management and operation management</td>
<td>16.60%</td>
<td>Business process management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operational and administrative management</td>
</tr>
<tr>
<td>Entrepreneurship education</td>
<td>11.30%</td>
<td>Entrepreneurship education</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Woman entrepreneurship and Education</td>
</tr>
<tr>
<td>Industry analysis</td>
<td>11.70%</td>
<td>Small and Medium-sized Enterprises</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industry and healthcare sector</td>
</tr>
<tr>
<td>Business model</td>
<td>20.60%</td>
<td>Business model and social impact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E-commerce business model</td>
</tr>
</tbody>
</table>
adoption of digital technologies to further enhance KM activities in entrepreneurial entities (e.g., Audretsch et al., 2020). The growing interest in this topic shows the importance of KM for the formation and growth of businesses. In the digital age, KM is seen as a critical resource for entrepreneurial entities operating in a highly competitive environment. The natural outcome of KM and entrepreneurial behavior is capturing innovative ideas, achieving pro-activeness, and creating long-term competitive advantage (Bandera, 2016).

Future research could focus on the new role of information technology in the process of initiating or innovating. We suggest that future studies in this area can develop frameworks to show how information technology can be helpful in incubating entrepreneurial initiation such as capturing new ideas and visualizing or prototyping new ideas. Future research may also investigate mixed perspectives in entrepreneurship (innovation, adopting new technologies, knowledge management) in different industries.

Strategy

Strategy refers to a plan of action or policy designed to achieve a major or overall aim. About 14% of the selected literature investigates areas of strategical IT-associated entrepreneurship: for instance, the data analysis, the entrepreneurial strategic orientation, and the labor market and policies. Entrepreneurship scholars have genuinely attempted to understand what opportunities to bring into existence, and how future goods and services are discovered and exploited to create and grow new ventures (Meyer, Necks, & Meeks, 2017). Studies in this area investigated business activities across the value chain to understand how these activities are pertinent to effective organizational strategies. Strategic entrepreneurship has been highly of interest to scholars, and many excellent studies have been produced. We found that most studies in this area focused on economic opportunities. We suggest that future study may consider investigating non-profit organizations and/or the philanthropic responsibilities of an entrepreneurial company.

Business Process Management and Operations Management

Business process management (BPM) combines knowledge from information technology and management sciences and applies it to operational business processes (van der Aalst, 2013; Weske, 2007). BPM and Operations Management (OM) are closely related to an organization’s productivity and cost control, and thereby, has received considerable attention. Studies regarding BPM and OM focus on the role of information technology adoption in business process improvement. For example, Sacco and Strait (2004) explored the adoption of IT infrastructure to facilitate the transition of a nonprofit organization to pursue an entrepreneurial path. Wu, Cegielski, Hazen, and Hall (2013) conducted an empirical study to investigate the factors that affect the organization’s propensity to adopt cloud computing technologies.

We suggest that future studies interested in this area can investigate how to utilize information technology to facilitate and improve the process of entrepreneurship itself. Entrepreneurship is a journey which starts from a new creative idea and a whole process to convert the idea into a business. Information technology can be adopted not only to improve the business model, but also the entire journey of converting the business idea into an enterprise.

Entrepreneurship Education

The disruptive role of IT in today’s business environment stresses the importance of integrating entrepreneurship education into higher education (Boldureanu et al., 2020). This integration requires significant steps from higher education including the introduction of theoretical courses on entrepreneurship for undergraduate and graduate students, training sessions, and events promoting entrepreneurship. Several research studies investigated the impact of students’ entrepreneurship education on entrepreneurship intention (e.g., Ndofirepi, 2020; Jena, 2020) and in creating new venture creation, innovation, and development (Galvão, Marques, & Ferreira, 2020). Some other
studies investigated another perspective in this area such as designing entrepreneurship courses in universities or introducing new program in entrepreneurship in universities (e.g., Duval-Couetil, Ladisch, & Yi, 2020). Future research studies would benefit from exploring the current teaching methods to foster sustainable entrepreneurship in universities in developed and developing countries and identify ways for improvement.

**Industry Analysis**

Industry-level research was a popular topic in this area. Articles related to this topic investigated the role of IT as a central part for value creation. By exploring the dataset in this study, we found that researchers studied IT-associated entrepreneurship related themes in several industries such as education, healthcare, tourism and hospitality, manufacturing, agriculture, banking, telecommunications, construction, transportation, and the IT and software industry. Other investigated domains include government, financial, service, food, airline, music, fashion, and semiconductor. Amongst all the industries that were studied, we found that education was well-covered in IT-associated entrepreneurship-related research, as well as tourism and hospitality, healthcare, and software and IT industries. Consequently, there is room for research in other industry sectors.

**Business Model**

Studies in this area of research are concerned with investigating the disruptive role of IT in business. IT and ICT adoption enable entrepreneurs to find new ways of doing business (Olsson & Bernhard, 2020). Companies may need to transform themselves in significant ways to succeed in this digital world. Moreover, research studies addressed the role that IT plays in transforming entrepreneurship. Particularly, IT is a catalyst for new forms of innovation, creating new business models, new customer experiences, and new types of products and services (Nambisan, Wright, & Feldman, 2019).

Further, the adoption of IT and digital technologies transforms business models and creates digital business models that enable value creation and capture. The interplay between digital technologies, digital business models, and digital innovation is still under explored and warrants further future research (Steininger, 2019). Also, future research may focus on some organizational issues related to implementing digital business models such as digital readiness and the implications of the transformation of business models on individuals, organizations, and industries.

**CONCLUSION**

A literature review is useful and helpful in examining how well an area of research has been studied. However, traditional ways of conducting literature reviews involve a time-consuming manual process and a large amount of subjective reading. In this study, we proposed to use a semi-auto method to perform a literature review. The LDA topic modeling approach can mine the topics quickly and fairly in a relatively objective and scientific way. We believe it can help researchers perform a thorough yet less time-consuming literature review. Further, we utilized LDA topic modeling to analyze 1005 articles in the interdisciplinary area of IT-associated entrepreneurship. The study results show a strong interest of researchers in different IT- entrepreneurship related topics. Business models and entrepreneurship initiative and innovation were among the well-investigated areas of research, while entrepreneurship education, strategy, and industry analysis research topics warrant further future research. The findings of this study suggested several insights on future studies in the area.

Our study is not without limitation. First, this study is limited by the sample selection. The focus was primarily on journal articles. Conference papers, books, and other journal articles published in other databases were not included in the analysis. Therefore, the extracted topics in the IT-associated entrepreneurship interdisciplinary field might not reflect topics in types of publications that were excluded from this study. Second, this study analyzed part of the article’s metadata. However, in
some cases titles and abstracts do not reflect all the content of the paper. Hence, future research can extend this study by employing full text analysis.

Implications and Opportunities for Further Research

This study has methodological, theoretical, and practical contributions. Methodologically, it shows the applicability of employing text mining (topic modeling) to automatically extract the initial research topics from a large data set.

Theoretically, the aspects of IT-associated entrepreneurship developed in the current study contributes to the theoretical foundation of IT and entrepreneurship research. The results obtained provide a comprehensive picture about the uses and impacts of IT for entrepreneurship.

Practically, the results of the current study provide a useful quality reference source. Emerging entrepreneurs can further understand IT uses and impacts in the business context. The aspects of IT-associated entrepreneurship also provide researchers and educators with an ontology of IT and entrepreneurship related topics. This would help cover the deficiencies in research and provide a guide for future research.

The obtained results allow us to suggest future lines of research. For instance, we encourage researchers to further investigate entrepreneurship education and its role in empowering women entrepreneurial engagement. Although educational attainment of women is rising, their participation in entrepreneurial activities is low, especially in many developing countries. Thus, future studies can further explore ways of fostering entrepreneurship culture among women to increase their entrepreneurial participation. The results of the current study also call for better understanding of digital entrepreneurship, its role in process transformation, and its determinants and readiness factors.

ACKNOWLEDGMENT

This paper is a revised and expanded version of a paper entitled ‘Linking Information Technology and Entrepreneurship: A Literature Review’ presented at 54th Hawaii International Conference on System Sciences (p. 5016).
REFERENCES


Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Joint Technical Report, Computer Science Department, Keele University (TR/SE0401) and National ICT Australia Ltd.


Yousra Harb is an Associate Professor in MIS from the Department of Information Technology at YU. Her research interests are in the areas of knowledge management, data analytics, healthcare analytics, e-commerce and e-government. She received her PhD in Information Systems from Dakota State University (DSU), USA, and MS in Information Systems, Management Information Systems (MIS) from DSU, USA and Yarmouk University (YU), Jordan.

Yanyan Shang is an Assistant Professor of Information and Technology Management at the Sykes College of Business, The University of Tampa. Her research broadly lies in management strategies, social media, and data modeling and analysis. Shang has published a peer-reviewed article in the journal IEEE Transactions on Engineering Management, and several peer-reviewed conference publications and presentations, including ICIS, AMCIS, HICSS, INFORMS, and MAM conferences.


