

Uncovering the digital “x” phenomena in the IS field: A text analysis approach

JOAQUIN RODRIGUEZ, GABRIELE PICCOLI

Abstract As digital computers permeate an increasing number of activities, the term digital is associated with more processes, entities, and objects (e.g., digital economy, digital natives). This trend reflects in Information Systems research, with “digital” appearing before existing research concepts (e.g., digital infrastructure), often substituting for “IT”. In this study, we contribute to the delineation between digital “x” and IT “x” research by investigating the degree of usage of the terms and uncovering the most widespread digital “x” constructs. Using text analysis techniques, we can analyze a broad set of journals and a long time period. Early results, based on 17 years of *MIS Quarterly* publications, confirm the increasing use of the term digital. Conversely, digitize and digitalize, the two main forms of the term digital, are rarely used and do not show any significant increase in usage over time. While preliminary and not concluding, our results confirm the increasing popularity of the term digital. However, a limited number of digital “x” terms appear to connote novel constructs and the bulk of the research does not address the socio-technical process of digitalization that is central to the IS discipline. We discuss the implications of our early results and provide a research agenda.

Keywords: • Digital “x” • Digitalization • Digitization • IT “x” •

1 Introduction

As the computer mediation of activities continues unabated, the term "digital" is associated with an increasing number of processes, entities and objects. Digital economy, digital natives, and digital media are small sample of the terms that are now common use in academic research, the media, and everyday conversations. In Information Systems research the digital "x" concept is now pervasive. The main examples are digital innovation (Yoo et al., 2010), digital business strategy (Bharadwaj et al., 2013), and digital options (Sambamurthy et al., 2003) – concepts that have gained widespread use. A panel at the International Conference of Information Systems (ICIS) 2017 noted that digital "x" concepts often mirror well-established IT "x" ones, such as IT innovation or IT strategy, raising the question of whether we are putting "old wines in new bottles" or producing novel contributions (Baiyere et al., 2017).

The question is critical for the field, because academic research is founded on agreed upon constructs that lay the conceptual foundation of a field of inquiry (Kuhn, 1970). Knowledge needs this foundation to advance, and new concepts are needed when new phenomenal emerge. Conversely, just relabeling constructs to follow fads leads to a proliferation in terminology that threatens knowledge accumulation and transfer between and within disciplines.

In this work we contribute to the discussion of whether digital "x" is a fruitful re-framing of established IT "x" concepts. Specifically, we first provide a set of definitions of key constructs for digital "x" terminology. Second, we propose a methodology that leverages text analysis techniques to assess the degree of usage of new terms and the identification of the digital "x" concepts.

2 Core Definitions

In this section, we ground the definition of the term "digital" and we introduce its most common forms.

2.1 Digital

The term digital, as opposed to analog, indicates the discrete representation of signals or data using a sequence of finite number of values. We become familiar with digital computation as kids, when we learn to count using our fingers. It was computing pioneer George Stibitz, doing research on computational machines during World War II, who popularized the term "digital" to describe systems or devices based on discrete representation of signals and/or data (Williams, 1984).

Although digital can be used to describe any system or device that uses a discrete representation, it was conceived and became widely adopted in relation to general purpose computing devices following the Von Neumann architecture (Von Neumann, 1945). While computers are human-made "devices that can be instructed to carry out arbitrary sequences of arithmetic or logical operations automatically" ("Computer," 2018), it was the invention of the transistor and integrated circuits that ushered in the digital electronic computer. Subject to "Moore's Law" (Moore, 1965), these devices experienced exponential growth in computation power, thus becoming pervasive. As digital electronic computers permeated all aspects of the economy over the last five decades, we observe two phenomena. First, the term digital becomes

associated with objects (e.g., digital media) and phenomena (e.g., digital marketing) in which digital computers play an increasingly central role. Second, the term digital is associated with the outcomes derived from the use of electronic computers. Digital transformation, digital revolution, or digital innovation is common terminology to refer to the consequences of embedding digital computers in activities and processes.

2.2 Digitize and digitalize

When investigating the adoption of the term digital we need to consider its two main forms, digitize and digitalize. In mainstream parlance the two terms are used interchangeably ("Digitization," 2018). Conversely, they should have different connotations. Digitizing is the technical "process of converting analog signals into a digital form, and ultimately into binary digits." (Tilson et al., 2010, p. 749) The conversion of a paper book into an e-book is an example of digitization. Digitalization is a "sociotechnical process of applying digitizing techniques to broader social and institutional contexts that render digital technologies infrastructural." (Tilson et al., 2010, p. 749) For example, the Amazon Kindle radically changed the acquisition process of books for millions of people through the employment of digitizing techniques, which ultimately rendered the platform infrastructural. The Kindle platform does not only digitize the book content but also serves as an infrastructure for the creation of new services such as the possibility to loan books from public libraries directly to Kindle devices.

Although it is possible that researchers use the terms interchangeably as advocated by their general meanings, it is crucial to study their usage separately for two reasons. First, it helps determine if the literature prefers one or the other - assuming they are used interchangeably. Second, establishing different connotations of the terms presupposes that they are not already widely adopted. It would be daunting to reverse this trend if they are used interchangeably.

3 Methodology

We adopt text analysis techniques to determine the extent of use of the terms digital, digitize and digitalize, and the identification of the digital "x" concepts as used in IS literature. Although the use of text analysis techniques does not yet enable the categorization of advanced conceptualizations (Orlikowski and Iacono, 2001), they have proved valuable to synthesize large amounts of text and to uncover trends in research areas and topics (Blei and Lafferty, 2006; Sidorova et al., 2008).

We analyze the full text of articles published in the IS basket of eight journals¹ from 2000 to 2018 using R (R Core Team, 2017). The timeframe considered is constrained by the availability of native Pdfs enabling the extraction of text without optical character recognition (OCR). While pragmatic, this choice is also appropriate because it covers a long enough range of years to capture the emergence of the terms and concepts we are interested in.

In the remainder of this document we report results for 17 years of *MIS Quarterly* papers (from 2000 to 2017) for a total of 861 papers. We included in our analysis all types of manuscript categories, including editors' comments, and special issues².

¹ Basket of eight IS journals (<https://aisnet.org/general/custom.asp?page=SeniorScholarBasket>).

² Results from multiple journals will be available for presentation at the conference.

First, we downloaded the Pdf files for each paper from the AIS electronic library adopting a file naming convention that allows identifying the year and journal of publication. Pdf files are then parsed and read into R data structures using the "pdftools" library (Ooms, 2017). We preprocess the articles by removing any text from the references section on, thus excluding exhibits and appendices. We then tokenize the text and remove stopwords using the library "tidytext." (Silge and Robinson, 2016). The tokenization process separates hyphenated words in its single word components, therefore creating multiple tokens. The tokenization and stopword removal process reduced the number of tokens from 866,846 to 486,806 with 31,097 unique tokens. Terms are stemmed, using the SnowballC library (Bouchet-Valat, 2014), when analyzing the digital "x" bigrams. Stemming enable the consolidation of plurals and singular forms, as well as any forms that belong to the same root. The stemming process was not applied to the terms digital, digitization, and digitalization because they would stem to the same root. Finally, the data is converted into a tidy format (Wickham and others, 2014) that becomes available for analysis³.

4 Early results

4.1 How prevalent is digital, digitize, and digitalize

As expected, the usage of the term digital significantly increases over time. We measure usage as the percentage of the total keyword frequency over the total word count of the articles in a given year. Therefore, usage conveys a measure of relative importance of the analyzed word over the total length of the articles. Furthermore, the measure accounts for differences in paper lengths and the number of papers published in each year. From 2000 to 2012 the usage remains relatively low, however 2013 represents an inflection point and we observe a significant increase afterward (Figure 1). The spike in 2013 is determined by the *MIS Quarterly* special issue on Digital Business Strategy, however, the prevalence of the term remains higher in the following years than those before 2012. The term "digitize" registers a slight increase over the years; however, its use remains quite limited. Digitalize, on the other hand, is rarely used and we see no sign of significant increase in usage. When compared to the term IT, we observe how IT and digital have opposite trends, with the usage of IT decreasing over time, whereas digital is on the rise. This result, if confirmed across the basket of eight journals, could signal a substitution effect from "IT" to "digital."

4.2 Digital "x" concepts

Our analysis confirms that the number of unique digital "x" concepts that appears every year in *MIS Quarterly* is on the rise (Figure 2). From 2000 to 2007 the number of digital "x" concepts used in the papers was 7.75 on average. Afterwards, the number of concepts increased significantly, reaching its peak in 2013 with more than 120 digital "x" concepts in use, for a

³ At this stage of our research, we parse the Pdfs entirely. Therefore, keywords, titles, and abstracts are combined with the rest of the document body. As Pdfs do not contain document structure information, identifying different sections of the document require ad-hoc solutions that might not be consistent across all articles over the considered years.

total of 1,253 occurrences over 23 different papers. From 2013 to 2017, the number of concepts used is on average 100.

The frequency of the digital "x" terms follows a strong positive skew distribution. 64% of the digital "x" concepts appear only in one paper during the analyzed period, such as digital opportunity or digital leash. Nearly 29.8% of the concepts were present in more than two and less than ten papers. Concepts present in more than ten papers where only 6.2% of the total.

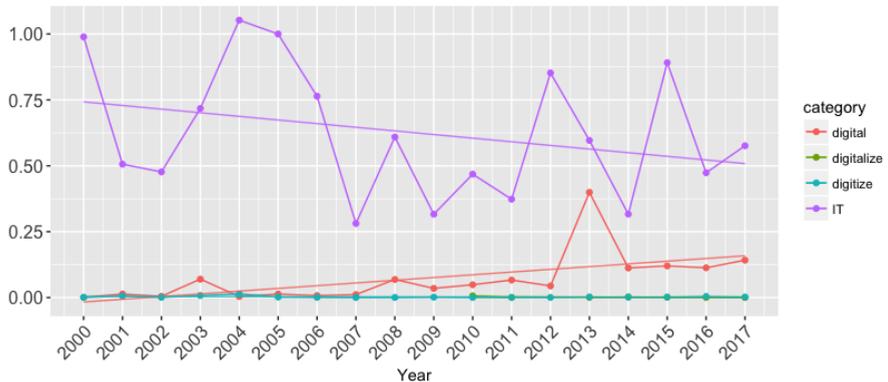


Figure 1: Yearly average *usage* of the keywords

This corroborates the hypothesis that researchers favor the use of digital "x" terms when they want to refer to the consequences of using digital computers (e.g., digital crime) or to those domains or contexts in which digital computers are employed (e.g., digital dashboard). As a consequence, only a small percentage of the digital "x" terms introduced are intended as new IS constructs. We define a construct as a "concept with added meaning, deliberately and consciously invented or adopted for a special scientific purpose" (Kerlinger, 1973). In the remaining cases "digital" is used as an adjective to characterize the digital nature of objects or process (e.g., digital game).

We identify 15 digital "x" concepts present in at least ten papers which we interpret as IS constructs (Table 1). Therefore, in Table 1 we do not include terms, such as digital economy or digital age, as they are not IS constructs. Analyzing the terms associated to digital we can identify some common areas. Digital innovation, business strategy, options, capabilities, processes, and assets are concepts more likely to relate to the implications of the employment of digital computers to organizations. In contrast, we can reasonably associate digital infrastructure, platform, network, and market to the research of inter-organizational relationship and markets. We can speculate that digital services and artifacts are related to the investigation of novel forms of services or artifacts.

5 Discussion

This research is inspired by the timely and important call to delineate the differences between the streams of research of IT "x" and digital "x" (Baiyere et al., 2017). We are not yet in position to provide definite answers, however, some interesting insights emerge from our early analysis. First, our early results confirm the increasing popularity of the term digital. However, we

observe a minimal increase of the term digitization and negligible use of digitalization. The scarce popularity of terms digitization and digitalization might indicate that work in the *MIS Quarterly* discipline has not yet focused on process as a focal issue in the analysis. Rather the work in this space started with a focus on the consequences of digitization (e.g., digital options).

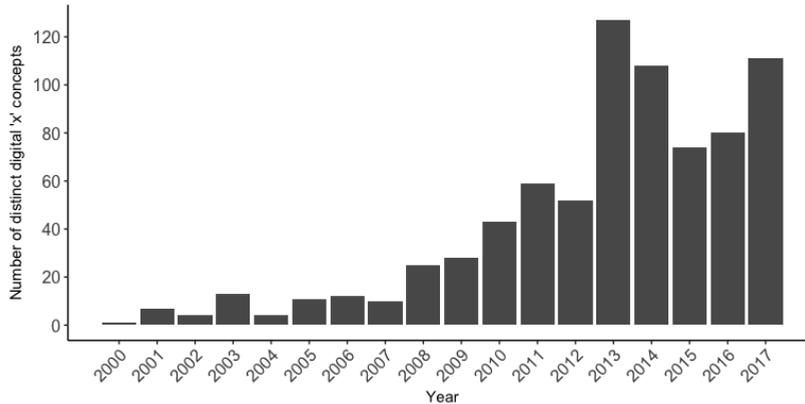


Figure 2: Yearly number of distinct digital “x” concepts

Table 1: Number of papers where the digital “x” concept is present

Digital “x” concept	digital innovation	digital artifact	digital business	digital infrastructure	digital platform	digital divide	digital option	digital capability	digital network	digital environment	digital asset	digital market	digital process	digital service	digital trace
N. of papers	24	22	22	21	20	18	18	15	14	12	11	11	11	11	11

Secondly, although we agree that the study of digitization as a technical process is not relevant to the IS discipline, it is more concerning that there is limited interest in the socio-technical process of digitalization. This result is consistent with previous research that shows how IS research tends to study the IT-artifact as an exogenous black box (Akhlaghpour et al., 2013; Orlikowski and Iacono, 2001). However, the results may also reflect the delay typically associated with publication in top journals and broadening our analysis to conferences may help balance the results.

Third, we provide some initial evidence showing that many digital “x” terms (almost two thirds) are sparsely used (e.g., digital agreement). These terms are contextual to the analysis. Only a limited number of digital “x” terms appear to connote novel constructs. We identify the 15 most recurrent digital “x” constructs published in the *MIS Quarterly* since the year 2000. Our results indicate that, while the number of digital “x” concepts used in IS research is

significant and growing, the percentage of those concepts that are prominent in the literature is limited.

Fourth, there are several digital “x” concepts that may not have an IT “x” counterpart (e.g., digital divide, digital trace, digital immigrant). Therefore, we might speculate that these concepts represent new phenomena that the literature did not address before – new wine if you will.

Finally, we showcase a systematic and scalable text analysis approach for literature analysis.

5 Future research

In future research, we will broaden the number of analyzed journals and the timespan considered. A more comprehensive analysis will determine whether the trend we observe in our preliminary results is consistent across journals. Moreover, a cross-journal analysis, and the inclusion of conferences, will uncover time-lag effects and differences across journals. Adopting a longitudinal perspective in the analysis of core digital “x” constructs will establish whether once introduced they become increasingly used or they tend to vanish. More importantly, our research agenda aspires to conduct a literature review of the most prevalent digital “x” concepts and to compare them to established IT “x” construct, to address the question of whether digital “x” is mere substitution for IT “x” or truly “new wine.”

References

- Akhlaghpour, S., Wu, J., Lapointe, L., Pinsonneault, A., 2013. The ongoing quest for the IT artifact: Looking back, moving forward. *Journal of Information Technology* 28, 150–166.
- Baiyere, A., Grover, V., Gupta, A., Woerner, S., Lyytinen, K.J., 2017. Digital“ x”-A new tune for IS research or old wine in new bottles? Thirty eighth International Conference on Information Systems, Seoul 2017.
- Bharadwaj, A., Sawy, E., A, O., Pavlou, P.A., Venkatraman, N.V., 2013. Digital Business Strategy: Toward a Next Generation of Insights (SSRN Scholarly Paper No. ID 2742300). Social Science Research Network, Rochester, NY.
- Blei, D.M., Lafferty, J.D., 2006. Dynamic topic models, in: *Proceedings of the 23rd International Conference on Machine Learning*. ACM, pp. 113–120.
- Bouchet-Valat, M., 2014. SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library. Computer, 2018. The free encyclopedia. (2004, July 22). FL: Wikimedia Foundation, Inc. Retrieved March 14, 2018 from <https://en.wikipedia.org/wiki/Computer>.
- Digitization, 2018. The free encyclopedia. (2004, July 22). FL: Wikimedia Foundation, Inc. Retrieved March 14, 2018 from <https://en.wikipedia.org/wiki/Digitization>.
- Kerlinger, F.N., 1973. *Foundations of behavioral research*. New York, Holt, Rinehart and Winston [1973].
- Kuhn, T.S., 1970. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Moore, G.E., 1965. Cramming more components onto integrated circuits, *Electronics Magazine*.
- Ooms, J., 2017. pdfutils: Text Extraction, Rendering and Converting of PDF Documents.
- Orlikowski, W.J., Iacono, C.S., 2001. Research commentary: Desperately seeking the “IT” in IT research - A call to theorizing the IT artifact. *Information Systems Research* 12, 121–134.
- R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Sambamurthy, V., Bharadwaj, A., Grover, V., 2003. Shaping agility through digital options: Reconceptualizing the role of Information Technology in contemporary firms. *MIS Quarterly* 27, 237–263.
- Sidorova, A., Evangelopoulos, N., Valacich, J.S., Ramakrishnan, T., 2008. Uncovering the intellectual core of the information systems discipline. *Mis Quarterly* 467–482.
- Silge, J., Robinson, D., 2016. tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *JOSS* 1. <https://doi.org/10.21105/joss.00037>
- Tilson, D., Lyytinen, K., Sørensen, C., 2010. Research Commentary—Digital Infrastructures: The Missing IS Research Agenda. *Information Systems Research* 21, 748–759. <https://doi.org/10.1287/isre.1100.0318>
- Von Neumann, J., 1945. First draft of a report on the EDVAC.
- Wickham, H., others, 2014. Tidy data. *Journal of Statistical Software* 59, 1–23.
- Williams, B.O., 1984. Computing with electricity, 1935-1945. University of Kansas.
- Yoo, Y., Henfridsson, O., Lyytinen, K., 2010. Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research* 21, 724–735. <https://doi.org/10.1287/isre.1100.0322>