Exploring the Structure of Software Development Research: A Preliminary Text Analysis

Full Paper

Sridhar Nerur
University of Texas at Arlington
snerur@uta.edu

VenuGopal Balijepally
Oakland University
balijepa@oakland.edu

Abstract

Organizations rely on software development projects to achieve a competitive advantage, increase operational efficiency, and to ensure compliance. This research employs topic modeling and bibliometric techniques to analyze abstracts from four top IS journals to identify dominant themes in software development. The distribution of topics across the years and the underlying topic clusters are presented as well.

Keywords

Software development, intellectual structure, text analytics, software projects

Introduction

It is widely acknowledged that Information Systems/IT–when designed well–can confer a number of advantages on organizations. These include operational excellence, as exemplified by companies like Walmart, and strategic differentiation, as evidenced by the extraordinary performance of companies such as Google, Netflix, and Apple, to name but a few. The excellent works of Peter Weill and Jeanne Ross (Weill and Ross 2005) on Enterprise Architectures and IT Governance make it abundantly clear that the benefits commonly associated with IS/IT can hardly be gained without an "optimized core", a robust, standardized IT platform that can lead to reusability and enhanced predictability. Therefore, the capabilities that an organization builds around its software development practices are critical to the operational efficiency as well as the responsiveness and competitiveness of a firm. Given this backdrop, it would be interesting to see how much importance the academic community gives to software development. In this exploratory study, we use bibliometric and text analytics to enumerate the key topics of software development that have been studied by scholars in the IS field.

IS academic research typically lags technology trends in the business world in terms of currency of topics covered. However, over time the topics in academic IS research tend to mirror the technology trends evident in the business world. Therefore, understanding the latent intellectual or conceptual structure of software development research in IS should provide us with a window into the technological and methodological trends of the practitioner world to help us reflect on and improve the relevance of our research. Such an endeavor should also help us to understand the relative importance of topics and research streams and alert us to the research gaps that need to be addressed. Understanding how much importance IS academic world has accorded to the methods and practices of software development should help us comprehend the priorities inherent in IS research and guide us in our future research efforts.

Advances in text analytics provide us with an efficient way to tease out the underlying structure of any corpus of knowledge. Text analytics complements rather than serving as a substitute for other techniques (e.g., bibliometrics, network analysis, meta-analyses, qualitative studies) that are typically used for understanding the intellectual underpinnings of a field. This exploratory study assesses the intellectual structure of software development in terms of the key thematic areas that are latent in the extant IS literature.

The remainder of the paper is organized as follows. The next section discusses prior literature, followed by a description of the methodology used in the study. Subsequently, the results and their implication are discussed. Finally, limitations and conclusions are presented along with directions for future research.

Exploring the Intellectual Structure of a Field

Articles are repositories of knowledge relating to topics within a field. The knowledge aggregated across the articles in a field defines the ontology of the domain. Thus, articles and, by extension, their authors are often the units of analyses when exploring any field or discipline. Exploring the intellectual structure of a field is of enduring interest to academics across disciplines for multiple reasons. First, it helps understand how various topics within the filed emerge, evolve and possibly fade over time. Second, it helps compare and contrast the evolution of thoughts and ideas in the field across multiple time periods to better understand trajectories and anticipate future topics of inquiry. Third, it helps highlight the under-researched topics/areas where future inquiry could help bridge knowledge gaps. As the purpose of our research is to explore the intellectual structure of software development research in IS, we review two approaches – bibliometrics and text analytics – that have frequently been employed to elucidate the underlying themes in a discipline.

Researchers have primarily used two broad bibliometric approaches to understand the intellectual structure of fields, namely, Author Co-citation Analysis (ACA) (McCain 1986; White and Griffith 1981), and Document Co-citation Analysis (DCA) [e.g., (Ramos-Rodríguez and Ruíz-Navarro 2004)]. Not surprisingly, documents and authors are the units of analyses in DCA and ACA, respectively. These techniques have been used to elucidate the underlying themes or sub-fields of disciplines such as information science (White and McCain 1998), strategic management (Nerur et al. 2008; Ramos-Rodríguez and Ruíz-Navarro 2004), macroeconomics (McCain 1984), international management (Acedo and Casillas 2005), consumer research (Hoffman and Holbrook 1993) and information systems (Culnan 1986; Culnan and Swanson 1986). In the domain of agile software development, Dingsøyr et al. (2012) made a preliminary effort to understand the intellectual structure of agile development research and its underlying themes.

The quantitative analysis of citation data is at the core of most bibliometric approaches. The limitations of citations have been well documented in the literature. One of the key issues in using citations is that they tend to ignore the context in which they occurred. Furthermore, the semantic content of the abstracts or articles is completely ignored when citations alone are considered. Text analytics can overcome this shortcoming and can provide additional insight that one might not get with biblimetrics alone. Taken together, bibliometrics and text analytics are likely to yield a far richer account of the intellectual landscape than either one by itself.

Overview of Text Analytics Approach

The proliferation of social media (e.g., Facebook, Twitter, LinkedIn), blogs and news outlets is generating enormous amounts of unstructured data (i.e., text, videos, audio, images). As a consequence, tools and techniques are evolving to harness the immense potential of data to afford actionable insights. Text analytics and/or text mining is being increasingly used as a research tool to detect underlying semantic patterns in such data. Current text mining tools can retrieve text from multiple file formats, such as PDF (Portable Document Format), XML (Extensible Markup Language), JSON (JavaScript Object Notation) and tsv/csv (tab or comma separated values). After retrieving the raw text, it needs to be pre-processed and prepared for analysis. The preparation typically involves operations such as conversion to a common case (e.g., lowercase), removal of stopwords (e.g., commonly occurring words such as “of”, “the”, “for”), and stemming to reduce similar words to their roots (e.g., reducing words “development”, “developing” and “develop” to their root form) so that they are counted as equivalent words (Miller 2015; Weiss et al. 2010).

The subsequent involves generating a vector of numbers based on either word frequencies or weights/scores denoting the importance of different words. One technique often used for the latter is Term Frequency – Inverse Document Frequency (tf-idf), which assigns higher weights to rare words as well as to words that frequently occur in a few documents but not across the entire corpus (Manning et al. 2008). The vector of numbers resulting from tf-idf vectorizer is then used to create a Term-Document

Matrix (TDM), where terms and documents are represented as rows and columns, respectively. For instance, an element TDM_{ij} is a frequency count for the occurrence of i_{th} word in j_{th} document. The TDM is used to create a distance (e.g., cosine similarity) matrix, which can then be used for comparing, and clustering documents based on similarity. The TDM may also be subjected to Non-Negative Matrix Factorization (NNMF), a popular technique for decomposing large multivariate data sets (Hoyer 2004; Lee and Seung 1999; Nerur and Balijepally 2015).

The NNMF procedure requires a-priori specification of the number of topics desired in a corpus (say, f). The procedure then factorizes the original TDM matrix (say having t x d dimensions) into two smaller matrices A and B (say, with dimensions of t x f and f x d, respectively). An iterative procedure processes A and B to approximate the underlying factorized matrix producing a result set containing words grouped into desired number of topics (Hoyer 2004; Lee and Seung 1999; Nerur and Balijepally 2015).

Latent Dirichlet Allocation (LDA) is an alternative procedure for deriving topics from any corpus (McCallum 2002). LDA also requires a-priori specification of the number of topics. TopicMapping (see Lancichinetti et al. (2015)), a technique based on network approach, is sometimes used to determine the optimum number of topics. We used LDA and TopicMapping approaches for topic analysis in this study. The next section presents the results and related discussion.

Results & Discussion

Data Collection

As stated earlier, the objective of this research is to assess the intellectual structure of software development research in IS to understand its important topics and seminal authors. Specifically, we focused on unraveling the research themes in software development research published in four top IS journals (i.e., MIS Quarterly (MISQ) and Information Systems Research (ISR) which are US based and Information Systems Journal (ISJ) and European Journal of Information Systems (EJIS) which are Europe based) during the last over a quarter century, i.e., from 1990 to 2015. As an exploratory effort, we chose these four journals because they appear among the AIS scholars' basket of journals and consistently rank high among IS journals.

For each journal group, we retrieved abstracts from the Web of Science database that satisfied our search criteria[1]. In all we had 290 abstracts with 146 from the two European journals and 144 from the two North American journals. Figure 1 shows the year wise distribution of the articles involved. As evident from the figure, there is a general increase in the articles relating to software development research during the past decade (i.e., since 2006) compared to previous decades (i.e., prior to 2006). This alludes to the increasing research interest generated by agile development methods, which are a significant departure from prior plan-based methods (e.g., waterfall method). In addition, the advent of phenomena such as offshoring and open source software development in the last decade or so seems to have contributed to this enhanced interest in software development research.

In this research, we used text analytics to uncover the underlying semantic structure of software development research. We used the following steps to analyze the text in the abstracts:

- Preprocessing: We removed commonly occurring English words (i.e., stop words) and other irrelevant terms from the abstract as also punctuation and digits. Stemming made it difficult to identify some of

[1] TS=("software development" OR "software project management" OR "software developer" OR "software coding" OR "programming" OR "programmer" OR "systems analysis" OR "system design" OR "object-oriented" OR "software testing" OR "software engineering" OR "software architecture" OR "UML" OR "unified modeling language" OR "program comprehension" OR "code generation" OR "software maintenance" OR "software testing" OR "code testing" OR "acceptance testing" OR "program testing" OR "software quality" OR "software specification" OR "design pattern" OR "programming language" OR "traditional development" OR "water-fall method" OR "plan-driven method" OR "software tool" OR "database design" OR "database management") and SO=("MIS Quarterly" OR "Information Systems Research" OR "European Journal of Information Systems" OR "Information Systems Journal") and PY=(1990-2015)

the words and hence we proceeded to analyze the text without stemming or lemmatization. We used a Python program to do the preprocessing.

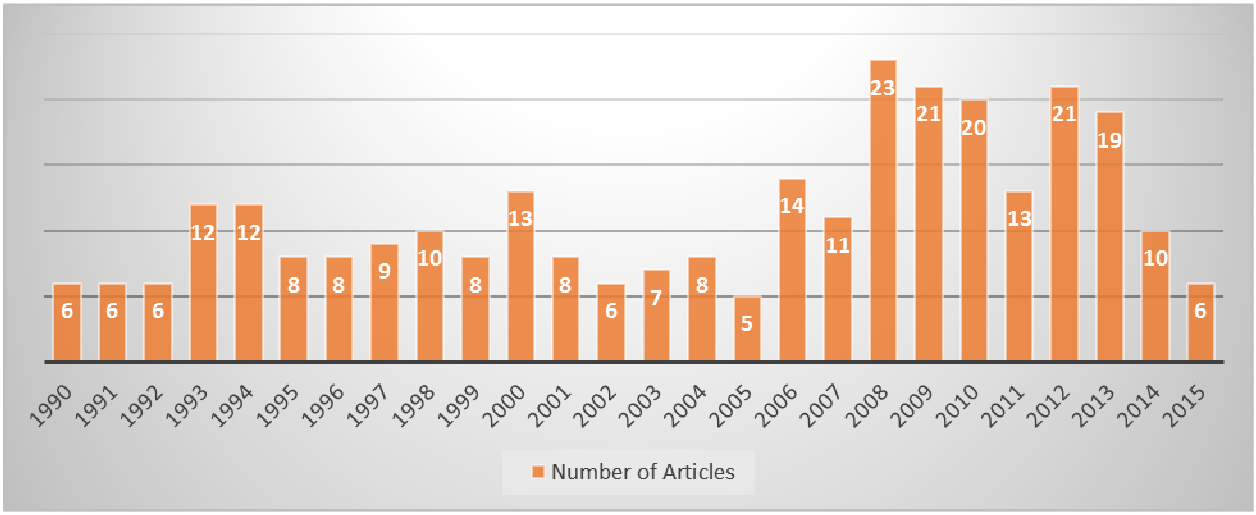

Figure 1. Distribution of Software Development Research Articles Across Years

- We subjected the preprocessed text to topic modeling. Specifically, we used Mallet's implementation of Latent Dirichlet Allocation (LDA) for deriving topics from all our abstracts (see McCallum (2002)). LDA requires that we specify the number of topics in advance, which makes it somewhat difficult because the researcher may not have an idea of the optimum number of topics reflected in all the abstracts. To determine the optimum number of topics, we used a technique called TopicMapping (see Lancichinetti et al. (2015)) that relies on a network approach to optimize the number of topics. We then used Mallet's topic modeling software to derive eight topics.
- In addition to performing an LDA, we also subjected our data to bibliometric analysis using an excellent software package called Vosviewer.

As mentioned earlier eight topics were derived based on a distribution of topics across our corpus and the way the words were distributed across the topics. In addition to extracting the topics, Mallet's topic modeling software provides the following:

- The key words associated with each topic, in descending order of importance.
- The extent to which each topic is reflected in each abstract. This file was then used to derive a year-to-topic matrix that shows the importance of each topic for each year of our analysis. The year-to-topic matrix was then used to generate cluster and heat maps.

Results

A topic denotes a research theme within a field. Applying LDA analysis on the abstracts we extracted eight underlying topics and related words as summarized in Table 1. We provided suitable labels for the eight topics based on the words associated with them.

As evident from Table 1, distributed agile development, software analysis and design were the top two topics explored by IS researchers followed by offshoring and open source software. With the advent of agile manifesto (AgileAlliance 2001), there has been renewed interest in software development research in general and agile methods in particular. Some important agile themes explored included distributed development (Cummings et al. 2009; Vlaar et al. 2008; Wakefield et al. 2008) and adoption of agile methods (Fitzgerald et al. 2006) (Vidgen and Wang 2009). Similarly, the top software analysis and design (SAD) sub-topics included user engagement in systems development (Champion et al. 2005; Metcalfe and Powellrt 1995), system usability issues (Fisher 1999), creativity in systems development (Cooper 2000), and CASE methodology implementation (McChesney and Glass 1993). Table 2 displays the top articles relating to the top four themes of software development research identified through text analysis.

Topic	Top 20 words associated with topic	Topic Label
1	development teams agile team software work practices distributed project isd agility offshore organizations communication projects coordination differences collaboration global members	Distributed agile (Agile)
2	systems information development system design process software analysis requirements model case data technology management organizations knowledge organizational business users implementation	Software analysis and design (SAD)
3	contract vendor contracts firms projects offshoring outsourcing client providers costs offshore model vendors pricing knowledge clients data cost price service	Offshoring vendor contracts (Contracts)
4	source open community network oss social members developers participation online model communities theoretical integration capital team communication activities content features	Open source software (OSS)
5	innovation product firms innovations software adoption process system assimilation errors case products order strong diffusion leaders dimensions cost base radicalness	Innovation diffusion (Innovation)
6	software development quality model process relationship task time fit applications levels developers maintenance higher knowledge approaches effect design efficiency cognitive	Software effectiveness (Quality)
7	conceptual models methods database modelling modeling model design computer analysts objectoriented techniques programming data designers language problem understanding concepts learning	Software modeling (Modeling)
8	project projects control risk management factors managers success formal risks controls client organizational informal isd impacts clan importance goals strategic	IS project management (PM)

Table 1. Key Topics of Software Development Research Between 1990 and 2015

Figure 2 shows the heat map of the topics distributed across the years along with cluster groupings. A close examination of the heat map reveals that software analysis and design (SAD) has been the most enduring topic across time periods. The heat map also reveals the ascendance of SAD and agile topics in the recent past (i.e., 2008 to 2013). Incidentally, the topic of innovation diffusion has been much less emphasized during the same period.

SNo	Topic	Authors	Title	Journal	Year
1	Distributed Agile	(Cummings et al. 2009)	Crossing Spatial and Temporal Boundaries in Globally Distributed Projects: A Relational Model of Coordination Delay	ISR	2009
2	Distributed Agile	(Vlaar et al. 2008)	Co-creating understanding and value in distributed work: How members of onsite and offshore vendor teams give make demand and break sense	MISQ	2008
3	Distributed Agile	(Wakefield et al. 2008)	A Model of Conflict Leadership and Performance in Virtual Teams	ISR	2008
4	Distributed Agile	(Fitzgerald et al. 2006)	Customizing agile methods to software practices at Intel Shannon	EJIS	2006
5	Distributed Agile	(Vidgen and Wang 2009)	Coevolving Systems and the Organization of Agile Software Development	ISR	2009
6	Software Analysis & Design	(Metcalfe and Powellrt 1995)	Information: A perceiver-concerns perspective	EJIS	1995
7	Software Analysis & Design	(Champion et al. 2005)	Client-Led Information System Creation (CLIC): Navigating the gap	ISJ	2005
8	Software Analysis & Design	(Fisher 1999)	Improving the usability of information systems: the role of the technical communicator	EJIS	1999
9	Software Analysis & Design	(Cooper 2000)	Information technology development creativity: A case study of attempted radical change	MISQ	2000
10	Software Analysis & Design	(McChesney and Glass 1993)	Post-implementation management of CASE methodology	EJIS	1993
11	Offshoring Vendor Contracts	(Gefen and Carmel 2008)	Is the world really flat? A look at offshoring at an online programming marketplace	MISQ	2008
12	Offshoring Vendor Contracts	(Chul Ho et al. 2013)	Contracting Information Security in the Presence of Double Moral Hazard	ISR	2013
13	Offshoring Vendor Contracts	(Gefen and Carmel 2013)	Why the first provider takes it all: the consequences of a low trust culture on pricing and ratings in online sourcing markets	EJIS	2013
14	Offshoring Vendor Contracts	(Jayanth et al. 2011)	Vendor and Client Interaction for Requirements Assessment in Software Development: Implications for Feedback Process	ISR	2011
15	Offshoring Vendor Contracts	(Moreno and Terwiesch 2014)	Doing Business with Strangers: Reputation in Online Service Marketplaces	ISR	2014
16	Open Source Software	(Chen et al. 2013)	Continued Participation in Online Innovation Communities: Does Community Response Matter Equally for Everyone?	ISR	2013
17	Open Source Software	(Yuqing et al. 2012)	Building member attachment in online communities: applying theories of group identity and interpersonal bonds	MISQ	2012
18	Open Source Software	(Scheffel et al. 2011)	An Experimental Comparison of Linear and Nonlinear Price Combinatorial Auctions	ISR	2011
19	Open Source Software	(Brian and Xiaoqing 2012)	The Cross-Purposes of Cross-Posting: Boundary Reshaping Behavior in Online Discussion Communities	ISR	2012
20	Open Source Software	(von Krogh et al. 2012)	Carrots and rainbows: motivation and social practice in open source software development	MISQ	2012

Table 2. Top Five Articles Relating to Top Four topics identified by Text Analysis

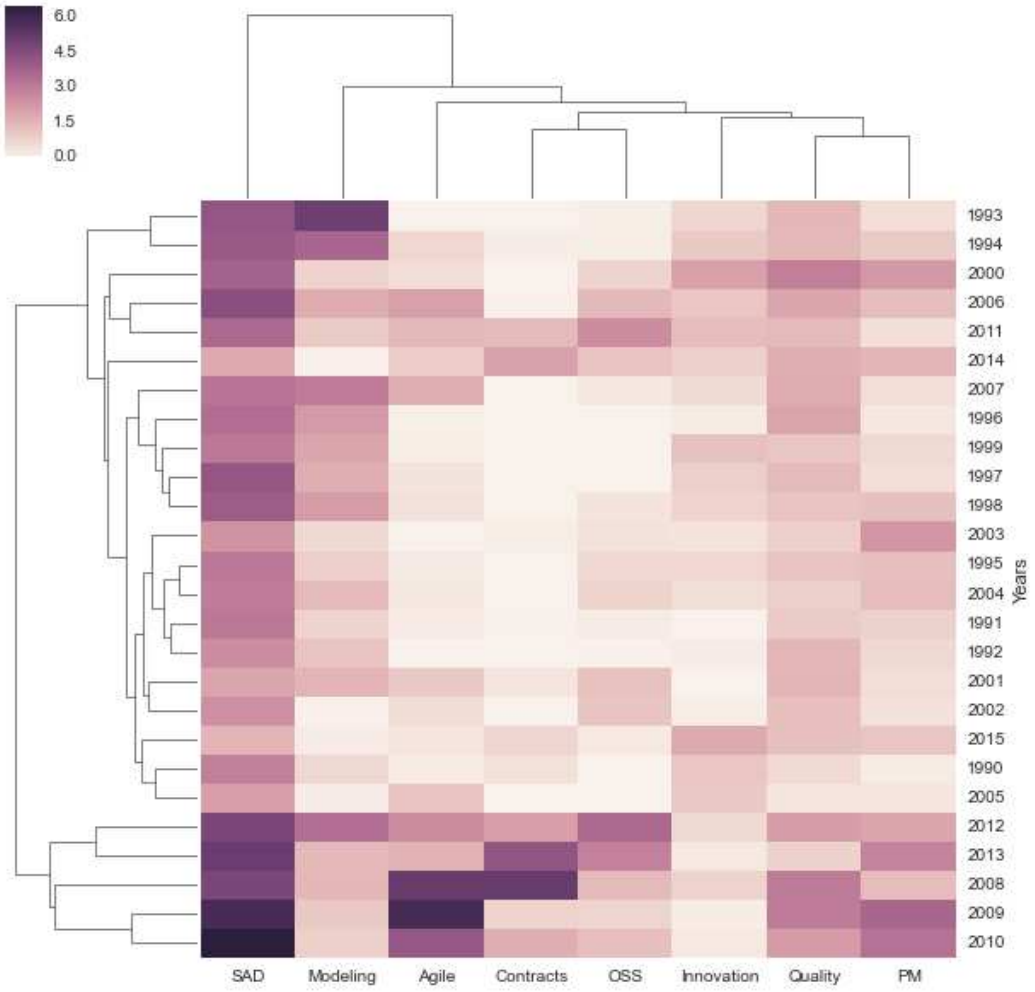

Figure 2. Heat Map of Software Development Research Topics Across Years

An excellent tool called VosViewer (Van Eck and Waltman 2011) was used to obtain: a) clusters of words based on their co-occurrence frequency (Figure 3) and b) clusters of first authors based on their co-citation frequencies (see Figure 4). Figure 3 displays the linkages between the top words from the abstracts. These are derived from the frequency of co-occurrence of words in the corpus. In addition to showing the linkages between words, the bubble chart also shows the changes in the vocabulary over the years. For instance, database design, representation, and designer were the top words found in abstracts during the nineties, while words such as open source project, vendor, service, and risk permeated the vocabulary during the past decade.

Figure 4 illustrates author clusters based on the co-citations of their papers with each color denoting a specific cluster. For instance, the authors shown in brown (e.g., Boehm, Brooks, Yourdon etc.) are frequently cited together for topics related to software development methods and practices. Similarly, the authors shown in blue (e.g., Hirschheim, Lyytinnen, Klein etc.) are frequently cited together for topics related to development paradigms and philosophical approaches.

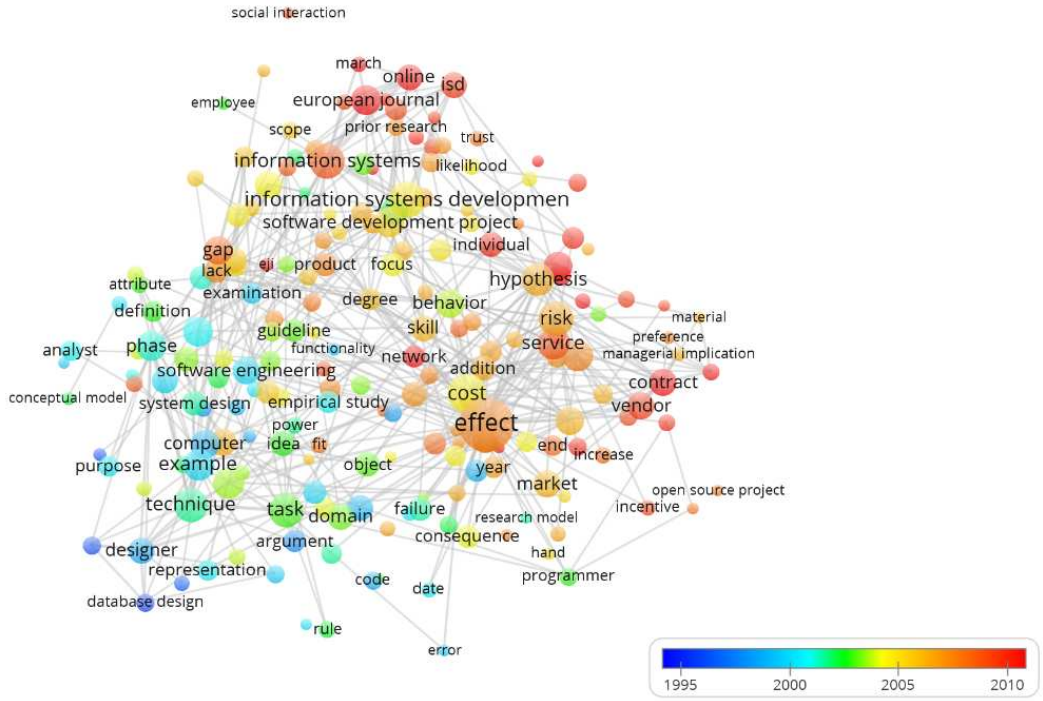

Figure 3. Word Linkages based on co-occurrence of words

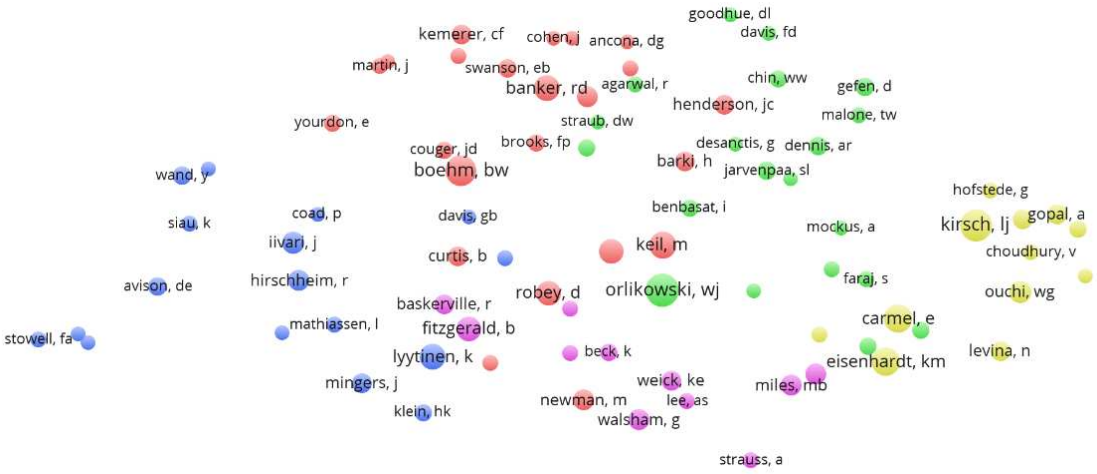

Figure 4. Author Clusters based on Bibliometric (Co-Citation) Analysis

Limitations

Our study has certain inevitable limitations. First, as an exploratory effort, we limited our search to only four IS journals. As these journals appear in the AIS scholars' basket of journals and consistently rank at the top among IS journals, we deem them to be acceptable for our research purpose. Second, we used only abstracts of articles, which are imperfect substitutes for the complete papers they summarize. Topic analyses involving complete papers would undoubtedly be a more comprehensive and rigorous approach to understand their latent semantic structures. Third, the abstracts that we retrieved are specific to the search terms we had used. Any changes to the search terms used will have some effect on the number of abstracts retrieved. Finally, since our focus is on the intellectual structure of disciplines, we have only looked at bibliometrics and text analytics. As a consequence, other pertinent approaches, such as qualitative studies or meta-analyses, have been excluded from our discussions. However, we believe that these limitations do not significantly detract from the general findings of this study.

Conclusion and Future Direction

This paper presented a preliminary topic analysis of software development research published in some top IS research outlets. Our study makes several contributions. First, it illustrates the potential of text analysis in unraveling the underlying structure of a corpus. Second, it illuminates how text analytics could complement other approaches (e.g., bibliometrics) used for comprehending the intellectual structure of a field. Third, it reveals the key research themes within software development. A future extension to this research may involve: a) examining full papers rather than just abstracts, b) sampling a larger journal set than the four journals examined here, and c) using other techniques (e.g., network analysis) to complement the findings from the text analysis to provide a richer interpretation of the structure of the domain involved.

REFERENCES

Acedo, F. J., and Casillas, J. C. 2005. "Current Paradigms in the International Management Field: An Author Co-Citation Analysis," *International Business Review* (14:5), pp. 619-639.

AgileAlliance. 2001. "Manifesto for Agile Software Development." Retrieved Jun 1, 2013, from http://www.agilemanifesto.org

Brian, S. B., and Xiaoqing, W. 2012. "The Cross-Purposes of Cross-Posting: Boundary Reshaping Behavior in Online Discussion Communities," *Information Systems Research* (23:3), pp. 993-1010.

Champion, D., Stowell, F., and O'Callaghan, A. 2005. "Client-Led Information System Creation (Clic): Navigating the Gap," *Information Systems Journal* (15:3), pp. 213-231.

Chen, Z., Jungpil, H., and De, P. 2013. "Continued Participation in Online Innovation Communities: Does Community Response Matter Equally for Everyone?," *Information Systems Research* (24:4), pp. 1112-1130.

Chul Ho, L., Xianjun, G., and Raghunathan, S. 2013. "Contracting Information Security in the Presence of Double Moral Hazard," *Information Systems Research* (24:2), pp. 295-311.

Cooper, R. B. 2000. "Information Technology Development Creativity: A Case Study of Attempted Radical Change," *MIS Quarterly* (24:2), pp. 245-276.

Culnan, M. J. 1986. "The Intellectual Development of Management Information Systems, 1972-1982: A Co-Citation Analysis," *Management Science* (32:2), p. 156.

Culnan, M. J., and Swanson, E. B. 1986. "Research in Management Information Systems, 1980-1984: Points of Work and Reference.," *MIS Quarterly* (10:3), p. 288.

Cummings, J., Espinosa, J., and Pickering, C. 2009. "Crossing Spatial and Temporal Boundaries in Globally Distributed Projects: A Relational Model of Coordination Delay," *Information Systems Research* (20:3), p. 420.

Dingsøyr, T., Nerur, S., Balijepally, V., and Moe, N. B. 2012. "A Decade of Agile Methodologies: Towards Explaining Agile Software Development," *Journal of Systems & Software* (85:6), pp. 1213-1221.

Fisher, J. 1999. "Improving the Usability of Information Systems: The Role of the Technical Communicator," *European Journal of Information Systems* (8:4), p. 295.

Fitzgerald, B., Hartnett, G., and Conboy, K. 2006. "Customising Agile Methods to Software Practices at Intel Shannon," *European Journal of Information Systems* (15:2), pp. 200-213.

Gefen, D., and Carmel, E. 2008. "Is the World Really Flat? A Look at Offshoring at an Online Programming Marketplace," *MIS Quarterly* (32:2), pp. 367-384.

Gefen, D., and Carmel, E. 2013. "Why the First Provider Takes It All: The Consequences of a Low Trust Culture on Pricing and Ratings in Online Sourcing Markets," *European Journal of Information Systems* (22:6), pp. 604-618.

Hoffman, D. L., and Holbrook, M. B. 1993. "The Intellectual Structure of Consumer Research: A Bibliometric Study of Author Cocitations in the First 15 Years of the Journal of Consumer Research," *Journal of Consumer Research* (19:4), pp. 505-517.

Hoyer, P. O. 2004. "Non-Negative Matrix Factorization with Sparseness Constraints," *J. Mach. Learn. Res.* (5), pp. 1457-1469.

Jayanth, R., Jacob, V. S., and Radhakrishnan, S. 2011. "Vendor and Client Interaction for Requirements Assessment in Software Development: Implications for Feedback Process," *Information Systems Research* (22:2), pp. 289-305.

Lancichinetti, A., Sirer, M., Wang, J., Acuna, D., Körding, K., and Amaral, L. 2015. "High-Reproducibility and High-Accuracy Method for Automated Topic Classification," *Physical Review X* (5:011007).
Lee, D. D., and Seung, H. S. 1999. "Learning the Parts of Objects by Non-Negative Matrix Factorization," *Nature* (401:6755), pp. 788-791.
Manning, C. D., Raghavan, P., and Schütze, H. 2008. *Introduction to Information Retrieval*. Cambridge University Press.
McCain, K. W. 1984. "Longitudinal Author Cocitation Mapping: The Changing Structure of Macroeconomics," *Journal of the American Society for Information Science* (35:6), pp. 351-359.
McCain, K. W. 1986. "Cocited Author Mapping as a Valid Representation of Intellectual Structure," *Journal of the American Society for Information Science* (37:3), pp. 111-122.
McCallum, A. K. 2002. "Mallet: A Machine Learning for Language Toolkit." from http://mallet.cs.umass.edu
McChesney, I. R., and Glass, D. 1993. "Post-Implementation Management of Case Methodology," *Eur J Inf Syst* (2:3), pp. 201-209.
Metcalfe, M., and Powellrt, P. 1995. "Information: A Perceiver-Concerns Perspective," *Eur J Inf Syst* (4:3), pp. 82-92.
Miller, T. W. 2015. *Modeling Techniques in Predictive Analytics with Python and R: A Guide to Data Science*. Upper Saddle River, NJ: Pearson Education.
Moreno, A., and Terwiesch, C. 2014. "Doing Business with Strangers: Reputation in Online Service Marketplaces," *Information Systems Research* (25:4), pp. 865-886.
Nerur, S. and Balijepally, V. 2015. "Understanding the Structure of Agile Software Development Using Text Analytics: A Preliminary Analysis," *Twenty-first Americas Conference on Information Systems, Puerto Rico*.
Nerur, S. P., Rasheed, A. A., and Natarajan, V. 2008. "The Intellectual Structure of the Strategic Management Field: An Author Co-Citation Analysis," *Strategic Management Journal* (29:3), pp. 319-336.
Ramos-Rodríguez, A.-R., and Ruíz-Navarro, J. 2004. "Changes in the Intellectual Structure of Strategic Management Research: A Bibliometric Study of the Strategic Management Journal, 1980-2000," *Strategic Management Journal* (25:10), pp. 981-1004.
Scheffel, T., Pikovsky, A., Bichler, M., and Guler, K. 2011. "An Experimental Comparison of Linear and Nonlinear Price Combinatorial Auctions," *Information Systems Research* (22:2), pp. 346-368.
Vidgen, R., and Wang, X. 2009. "Coevolving Systems and the Organization of Agile Software Development," *Information Systems Research* (20:3), pp. 355-376.
Vlaar, P. W. L., van Fenema, P. C., and Tiwari, V. 2008. "Cocreating Understanding and Value in Distributed Work: How Members of Onsite and Offshore Vendor Teams Give, Make, Demand, and Break Sense," *MIS Quarterly* (32:2), pp. 227-255.
Van Eck, N.J., & Waltman, L. 2011. "Text mining and visualization using VOSviewer," *ISSI Newsletter*, 7(3), 50–54.
von Krogh, G., Haefliger, S., Spaeth, S., and Wallin, M. W. 2012. "Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development," *MIS Quarterly* (36:2), pp. 649-676.
Wakefield, R. L., Leidner, D. E., and Garrison, G. 2008. "A Model of Conflict, Leadership, and Performance in Virtual Teams," *Information Systems Research* (19:4), pp. 434-455.
Weill, P., and Ross, J. 2005. "A Matrixed Approach to Designing It Governance - a Simple One-Page Framework Can Help Companies Allocate It Decision Rights and Accountabilities So That Individual It Decisions Align with Strategic Objectives," *MIT sloan management review* (46:2), pp. 26-34.
Weiss, S. M., Indurkhya, N., and Zhang, T. 2010. *Fundamentals of Predictive Text Mining*. Springer London.
White, H. D., and Griffith, B. C. 1981. "Author Cocitation: A Literature Measure of Intellectual Structure," *Journal of the American Society for Information Science* (32:3), pp. 163-171.
White, H. D., and McCain, K. W. 1998. "Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972–1995," *Journal of the American Society for Information Science* (49:4), pp. 327-355.
Yuqing, R., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., and Kraut, R. E. 2012. "Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds," *MIS Quarterly* (36:3), pp. 841-864.