Understanding IT Innovations through Discourse Analysis

Ping Wang, Chia-jung Tsui, Kenneth R. Fleischmann, Douglas W. Oard, and Lidan Wang
College of Information Studies, University of Maryland, College Park, MD 20742
{pwang, ctsui, kfleisch, oard, lidan}@umd.edu

The dynamic information field is characterized by the constant ebbs and flows of innovations in information technologies (IT). Accordingly, managing information and formulating policies in the iField require understanding IT innovations – what they are and will be, who develops and/or adopts them, and how innovations may be effectively developed, implemented, and used. Despite a relatively sustained research literature on IT innovations [2], our knowledge of innovations is still inadequate to effectively inform strategic information management and policy-making in the iField. For instance, the field is filled with numerous buzzwords and acronyms, making it hard to differentiate true progress from mere change. And most research and practice are focused on highly popular innovations such as Web 2.0 and cloud computing; little is known about why only some innovations come to be popular while others do not. The lack of understanding is in part caused by limited research designs that focus on only one or a few innovations, owing to the difficulty in analyzing large-scale data on multiple innovations.

The present study seeks to address these limitations by offering a theoretical foundation and an analytical method for understanding the dynamic interactions among IT innovations. Theoretically, we posit that innovations emerge and evolve in an ecosystem. Each innovation can be likened to a species competing with or supporting others in a resource space. One important resource that every innovation relies on is *attention* from people and organizations. A certain innovation requires a certain type of attention. For example, the innovation of computer-aided software engineering (CASE) asks for attention mainly from system analysts and programmers. Their attention may also be "nutritious" to the innovation of object-oriented programming (OOP), but not so much to customer relationship management (CRM), which thrives on the attention from a different group of people. Because CASE and OOP "consume" the same type of attention (i.e., from the same group of people), the two innovations are related. Innovations may be related for other reasons as well. For example, different innovations may be developed to solve similar problems, require common knowledge for understanding the problems or similar skills to implement the solutions, or share the practices or roles to be affected by the innovations. To the extent two innovations are related, attention may flow from one to the other. The relationship between a pair of innovations may take on different forms: They may compete with each other or they may complement each other.

The flows of attention among innovations are both reflected and enabled by discourse – what people have said and written about the innovations. While the discourse about an innovation sometimes manifests human actions undertaken on behalf of the innovation, often the discourse itself is a form of human action, e.g., to make sense of, promote, or denounce the innovation. Therefore, analysis of discourse about multiple innovations can help us understand the evolution of innovations and the relationships among innovations. There are numerous outlets of discourse in the innovation ecosystem including books, magazines, conferences, blogs and wikis, and so on. Our strategy is to examine both the content and volume of discourse, making it possible to understand the contexts in which volume patterns emerge and evolve and content originates and changes.

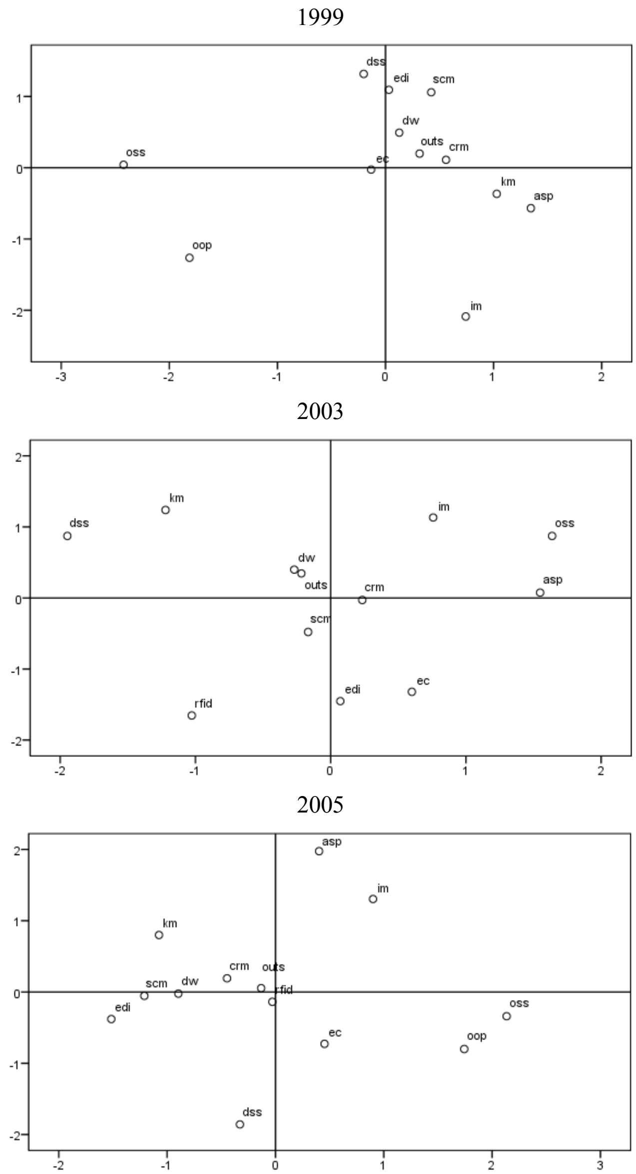
Specifically, we have thus far downloaded all articles published in *InformationWeek* during a ten-year period (1998-2007) from the Lexis/Nexis online database. *InformationWeek* is used here as an exemplar outlet of the IT innovation discourse. We are in the process of acquiring content from other outlets including academic publications and informal sources such as blogs.

We examined the subjects that had been automatically assigned to the *InformationWeek* articles in Lexis/Nexis and tallied the number of articles for each subject. After reviewing the frequency table of the subjects, we eliminated the subjects whose labels are general terms (e.g. "children") and the subjects that refer to topics other than IT (e.g., "organized crime"). Then from the remaining IT subjects, we randomly drew a sample of 13 IT innovations with various degrees of popularity (Table 1). Each innovation corresponds to a unique subject. To examine the volume of the discourse on these innovations, we drew a line chart to visualize the evolving popularity of each innovation, measured by the number of articles in the subject corresponding to the innovation. To examine the content of the discourse on these innovations, we applied *Kullback–Leibler (KL) divergence*, a measure of the difference between two probability distributions [3], to the measurement of the difference between innovation discourses.

Table 1. Selected IT Innovations

IT Innovation	Acronym/Abbreviation
Application service provider	ASP
Customer relationship management	CRM
Data warehouse	DW
Decision support systems	DSS
Electronic commerce	EC
Electronic data interchange	EDI
Instant messaging	IM
Knowledge management	KM
Object-oriented programming	OOP
Open source software	OSS
Outsourcing	OUTS
Radio frequency identification	RFID
Supply chain management	SCM

First, we extracted all paragraphs containing at least one of the 13 IT innovation labels from the *InformationWeek* articles for each year. Probability distributions were constructed as language models for those paragraphs. KL divergence yields an asymmetric 13x13 matrix (since KL divergence is not commutative). We symmetrized the matrix by averaging the elements for each innovation pair to produce a symmetric dissimilarity measure between each pair of innovations. Next, we applied *multidimensional scaling* (MDS) [1] to the matrix in order to visualize the distances between innovations in a 2-dimensional space. MDS is a statistical information visualization technique that assigns a location to each item in a lowdimensional space based upon a matrix of item-item (dis)similarities.



In some years, there were no articles about certain innovations.
Figure 1. MDS for KL Divergence of IT Innovations

Figure 1 shows three MDS plots generated by SPSS for 1999, 2003, and 2005, in which the visual proximity of IT innovations is indicative of similarity. In the figure, some IT innovations are close to each other, suggesting that the language used to describe them is somewhat similar. For example, Customer Relationship Management (CRM), Data Warehouse (DW), and Supply Chain Management (SCM), three typical enterprise software innovations, are close to each other. In contrast, Application Service Provider (ASP) and Decision Support Systems (DSS) are far away from each other throughout the years.

To see whether content similarity has anything to do with the popularity of the innovations measured by the discourse volume, we present those three closely related innovations (CRM, DW, and SCM) and those two distant innovations (ASP and DSS) in Figure 2. The popularity of each IT innovation is indicated by the yearly counts of articles that carry the innovation label as a subject in Lexis/Nexis. In Figure 2, peaks for all three enterprise software innovations (i.e., CRM, DW, and SCM) occurred around year 2000. Together they experienced a significant drop between 2000 and 2002. And since then their popularities have been declining. On the other hand, ASP and DSS have largely uncorrelated trends. These examples suggest that a relationship may exist between content similarity and innovation popularity. It is worth noting that, while we have presented here an example of positively correlated trajectories of closely related, somewhat complementary innovations, negative correlation may exist among the trajectories of closely related *competing* innovations.

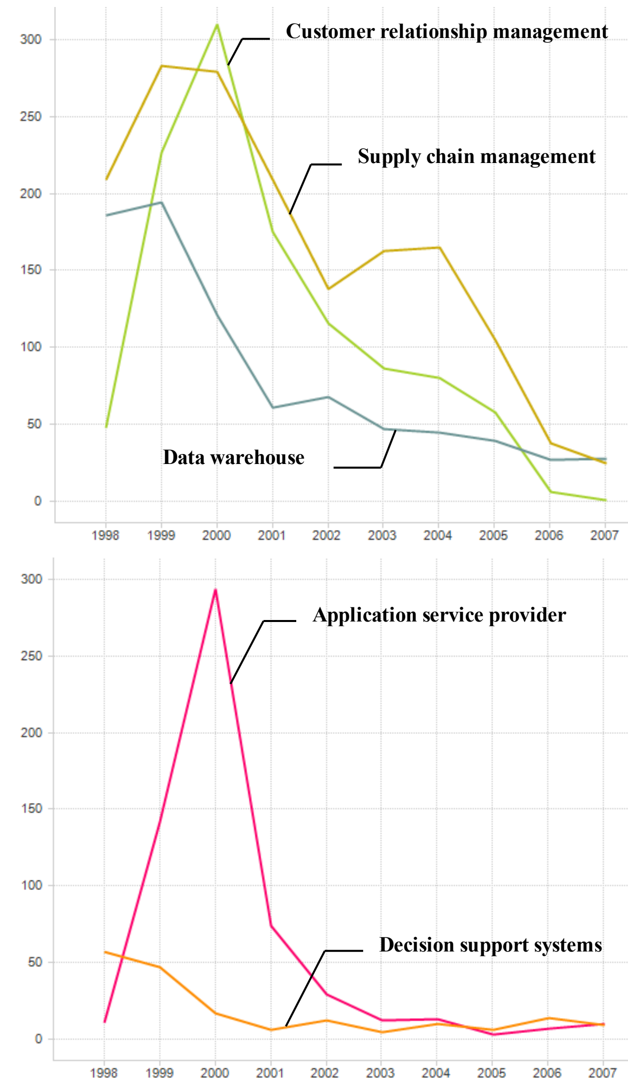


Figure 2. The Popularity of IT Innovations

The MDS plot derived from KL divergence as illustrated in Figure 1 is a promising technique. It helps information researchers and practitioners monitor multiple innovations and their dynamic relationships over time. On this innovation "radar screen," innovations close to each other may be treated as clustered groups for analysis. This approach, when employed longitudinally, also helps us understand the evolution of innovations. An innovation's older discourse can be compared with its current discourse. We can also compare the discourse on older innovations with that on newer innovations, detecting the new wine in an old bottle and/or the old wine in a new bottle. Additionally, this approach is scalable, suitable for analyzing a large number of innovations over a long period of time.

Presently, we are collecting data from a diverse set of discourse outlets and applying our methods to approximately sixty innovations over a 20-year period. A larger dataset will enable us to develop and test some hypotheses on the complex relationship between innovation popularity and the content of discourse.

In summary, our study offers an ecological theoretical framework and a scalable analytical method for information researchers, managers, and policy-makers to monitor and understand IT innovations in the information field. The key insight here is that discourse both reflects and constructs technological dynamics in the iField and thus such dynamics can be studied and understood through discourse analysis.

Keywords
Information technology innovation, discourse analysis, popularity, information management, information policy

ACKNOWLEDGMENTS
This study is based upon work supported in part by the National Science Foundation under Grant IIS-0729459. We thank Tiffany Chao, An-Shou Cheng, Philip Resnik, and Asad Sayeed for their help!

REFERENCES

[1] Borg, I. and Groenen, P. 2003. Modern Multidimensional Scaling: Theory and Applications. *Journal of Educational Measurement, 40* (3). 277-280.

[2] Fichman, R.G. 2004. Going Beyond the Dominant Paradigm for Information Technology Innovation Research: Emerging Concepts and Methods. *Journal of the Association for Information Systems, 5* (8). 314-355.

[3] Kullback, S. and Leibler, R.A. 1951. On information and sufficiency. *Annals of Mathematical Statistics, 22* (1). 79-86.