Combining Scientometric and Content Analysis Methods for Identifying Core Concepts and Action Principles of Information Systems Development

Nik Hassan  
*University of Minnesota - Duluth*

Lars Mathiassen  
*Georgia State University*

Follow this and additional works at: [http://aisel.aisnet.org/amcis2009](http://aisel.aisnet.org/amcis2009)

**Recommended Citation**
[http://aisel.aisnet.org/amcis2009/487](http://aisel.aisnet.org/amcis2009/487)
Combining Scientometric and Content Analysis Methods for Identifying Core Concepts and Action Principles of Information Systems Development

First author’s name
Affiliation
E-mail address

Second author’s name
Affiliation
E-mail address

ABSTRACT
This study suggests refining and further developing the body of knowledge (BoK) framework for information systems development (ISD) proposed by Iivari et al. (2004) using a combination of scientometric and content analysis approaches. The paper synthesizes descriptive concepts and actionable principles that scholars in highly-cited ISD articles agree to be fundamental to ISD research and practice. The results of the study highlights the attention given by IS researchers to behavioral and contextual factors over and above methodological and technical factors.

Keywords (Required)
Information systems development body of knowledge (BoK), IS design, scientometrics, content analysis

INTRODUCTION
The goal of this study is to demonstrate the benefits of combining scientometric and content analysis approaches to support the conceptual development of the information systems development (ISD) field of study. It begins by arguing for the use of scientometrics coupled with a form of content analysis for identifying core concepts and important historiographical events. A pool of 3,046 articles from 34 highly-ranked information systems (IS) journals is used as the source for identifying ISD articles. Subsequently, as an illustration of the proposed method, many useful concepts and actionable principles are extracted from the five most-cited ISD articles from this pool. The results of this content analysis are used to refine and further develop the body of knowledge (BoK) framework for ISD proposed by Iivari et al. (2004).

APPLYING SCIENTOMETRICS AND CONTENT ANALYSIS FOR SUPPORTING CONCEPTUAL DEVELOPMENT OF INFORMATION SYSTEMS

Ever since Clark (1957), Westbrook (1960) and Price (1963) proposed measuring the quality of any scientific work by the amount of usage it attracts, citation counts have been used in numerous studies to identify important scientific papers. The number of citations that a paper receives serves as an efficient and useful proxy for the paper’s quality and significance. Although the results of citation counts may be slightly skewed by questionable citing behaviors such as self-citing, citing of close friends, and citing of prominent work to confer status, the impact of these behaviors can be controlled so as not to diminish the value of identifying highly-cited works (Price, 1963; Wade, 1975). Citation counting is one of many techniques available from a relatively young field called “scientometrics” for studying the history and sociology of science.

A direct result of using citation counts for analyzing the historiography of a publication is the ability of identifying key events that can be considered historically significant. Such knowledge-related events can be said to have a major impact on the literature and thinking at the time (Garfield, 1955). This notion of the importance or significance of the publication is based on the qualitative assessment done by the citing author of the value and importance of the cited work. The study of publications with high “impact factors” can be extended not only to the life work of the author of that publication, but also to the journal, book or institution where the event was recorded. Based on Small (1978), this study extends the concept of “impact” down to the context surrounding the citations and the concepts cited in those citations. Scientometric techniques enable researchers to start this process of introspection by going both backwards and forwards from any reference to follow the historical and sociological development of the subject in question. Using this method, a “critical path” of discoveries and major research events can be documented enabling researchers to identify and agree on a representation of the IS field. Garfield (1964) applied this technique to compare the historical description of the discovery of the double helix by Isaac Asimov (1963) against actual citation data from the authors involved in the discovery. The analysis found additional techniques and concepts that contributed to the discovery from 17 articles and 31 additional author citations not mentioned in Asimov’s book. The effectiveness of using citation counts for measuring the quality of work is verified in several other ways.
Evidence from historiographical studies has shown that the number of honors conferred on authors correlates well with the number of citations they receive (Cole & Cole, 1967). Garfield and Marlin (1968) used citation counts to successfully predict Nobel Prize winners in 1969. Data from the Science Citation Index shows that Nobel Prize winners are highly-cited with total citation counts fifty times more than the average scientist. Following Garfield’s tradition, since 1989 Thomson Reuters continues to successfully predict who’s likely to receive the Nobel Prize (Thomson Reuters, 2008).

Content analysis is more well known as a means of “making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use” (Krippendorf, 2004, p. 18). In the context of supporting the conceptual formation of academic fields such as IS, content analysis is useful in conceptualizing and making inferences from the corpus of research texts such that patterns, analytical constructs and the development of scholarship can be identified and evaluated. In the case of the IS field, the identification of important or “core” concepts of IS becomes especially relevant. Because the development of indigenous theory in IS continues to be a major issue (Gregor, Grover, Lylytinen, Saunders, & Niederman, 2008) (Gregor, 2006) , what is even more critical is the development of IS concepts that feed into the development of these theories (Markus & Saunders, 2007). Conceptual formation is the bedrock of knowledge and science (Carnap, 1995; Hempel, 1956). Without a cogent foundation for the invention and development of unique concepts that belong to IS, the field will continue to be dependent on other disciplines. Scientometrics has always played a key role in identifying and analyzing core concepts in the natural and social sciences as well as in the fields of humanities. For example, in biochemistry, many of the classic publications that led to the discovery of insulin did not even mention the term “insulin.” However, the phrase “internal secretions of the pancreas,” which was used by the scientists that discovered insulin described the same concept (Banting & Best, 1922). Other concepts that linked “diabetes mellitus” to pancreatic defect was known to scientists 30 years before insulin was invented (Garfield, 1970b). The combination of citation analysis and content analysis resolves the problem of the proliferation of terminology by linking the concept via citations written in a publication to similar concepts written by other researchers regardless of what terms or words are used.

This combined method also supports the efforts of conceptual development by providing an objective way of indexing key word and subject entries (word concepts) in publications. Instead of relying on a very small number of subject indexers, who subjectively choose their preferred set of key words for the index, the citation index utilizes the virtual army of indexers—the scientists themselves—who would in effect index a publication from their point of view, interpreting the terminology for everyone else by including the cited reference in their own publications. The result of this manner of indexing is relatively stable because as new ideas, concepts and subjects are studied and discovered, instead of updating the subject index to include such discoveries, the citation index would only need to add referencing publications (Garfield, 1979). The subject index generated by the citation index becomes a virtual dictionary of concepts that are, by all practical purposes, agreed by scholars publishing in a particular field. For IS, the combination of scientometrics and content analysis therefore presents a useful starting point from which scholars can identify concepts and refine them for inclusion into indigenous theories.

**METHODOLOGY**

The goal of this study is to demonstrate how we can identify and synthesize core concepts and actionable principles from the literature specifically in the area of ISD. Morrison and George (1995) found that more than 50% of research in IS lies in the area of ISD. If this is true, the results of this study would go a long way in characterizing the core concepts of ISD itself. The combination of scientometric and content analysis approaches are effective in identifying these core concepts and principles because: (1) a cited paper has provided an intellectual foundation for citing works (Price, 1963), (2) it is fair measure of research communication and activity (Garfield, 1973) (3) the most-cited papers represent the body of papers that has had the most impact and influence on the discipline (Garfield, 1970a, 1971), (4) the number of citations avoids the inherent bias of a single authority’s opinion of the quality of the paper (Westbrook, 1960), and (5) the citations symbolize the concepts and actionable principles that have been found to be most useful by other researchers (Small, 1978).

As in any scientometric study, the most critical step is choosing the articles to be included in the pool for citation analysis (Chua, Cao, Cousins, & Straub, 2003). The journal rankings provided by Saunders (2007) supplied 34 titles for journals popular among IS authors. Only the titles ranked with average rank points of 30 or higher were picked and they also needed to be indexed and available for analysis in the ISI Web of Knowledge.

The complete pool of 34 journals used in this study is shown in alphabetical order in Table 1.
At the time of the analysis, the ISI Web of Knowledge provided data for “source articles” or articles in the database published between 1999-2007 inclusively. Source articles published earlier than 1999 are not included in this study. From the same 34 journals, articles that have been cited at least 40 times were selected. Finally, the bibliographic information including titles, times cited, and abstracts of 3046 highly-cited articles in the allied computing field were documented into NVivo\(^1\), a content analysis software.

Because the 3046 highly-cited articles are not all ISD articles, we identified ISD articles using the following steps:

\(^1\) NVivo Ver. 8, QSR International, http://www.qsrinternational.com
1. A qualitative analysis was performed (using NVivo) on the titles and abstracts of these articles using the coding supplied by Iivari et al. (2004) shown in Table 2. Variations of the terms were included in the search criteria. Wild card characters were used (e.g., implement*) for each of the terms.

2. The coding by Iivari et al. (2004) did not include terms involving business process reengineering or redesign, so variations of these codes were included to ensure that articles discussing such topics would be captured.

Based on the frequency of occurrence of all the possible combinations of the coding in the title, keywords and abstracts of the pool of articles, the NVivo software ranked relevant ISD articles according to relevance. Livari et al.’s (2004) coding terms included the term “acceptance.” The use of this coding term captured many highly-cited articles referring to the Technology Acceptance Model (TAM) or user acceptance. These articles are not categorized as ISD articles because they study post-development and adoption-related issues rather than systems development or design. Only the top 100 most-relevant articles were chosen for further analysis. These 100 most-relevant articles were ranked according to the number of times they were cited. A qualitative process was employed to exclude “pure” computer science articles focusing on hardware or software processing of symbols which ignores the social, “pure” management articles that don’t involve any technology component, or “pure” operations research-type articles that focus on models and algorithms. This process follows from several studies that identify the IS field as the intersection of the social and the technological (Lee, 1999; Sidorova, Evangelopoulos, Valacich, & Ramakrishnan, 2008; Vessey, Ramesh, & Glass, 2002). The filtering resulted in the five most-highly cited articles in ISD shown in Table 3.

<table>
<thead>
<tr>
<th>Article</th>
<th>Number of times cited</th>
</tr>
</thead>
</table>

Table 3: Five Highly-Cited ISD Articles

Using NVivo, a qualitative analysis of these five articles was performed to extract as many core concepts and actionable principles that could be fitted into the Iivari et al. (2004) ISD BoK framework. If a particular concept did not fit into any of the categories, a new concept category was created. Iivari et al. (2004) mapped articles from MIS Quarterly and Information Systems Journal to test their framework. Instead of using just any article from IS journals, this study chose articles that the majority of IS scholars agree, as reflected by their citations, were most useful to them and reflected the level of interest and activity that warranted these articles to be cited. We suggest that using this approach, it is possible to synthesize concepts and principles that reflect a truly distinctive and representative BoK of ISD.

RESULTS AND DISCUSSION

The results from the analysis are shown in tables 6-11. Each table represents an ISD category extracted from the five selected articles. The first column in each table describes the core concepts and propositions referenced by the articles. The second column describes any actionable principles suggested or implied by the articles related to each core concept.
Descriptive Concepts and Propositions

- Each application type demands a different set of skills, knowledge and requirements specifications
- Individual developers understand limited areas of the application
- Domain knowledge tend to be spread thinly among developers
- Thin spread of application domain worsens in embedded systems (increased interdependencies)
- Differences in individual talents impact project performance
- Projects have competing development perspectives
- Design work accomplished through expert interaction with others
- Expert power - the ability to influence a group through superior knowledge
- Erosion of manager’s technical knowledge
- Changes in application domain weakens a company’s technical maturity and foundation for making sound management decision
- Coordination process for multi-company projects are more complicated than single company projects

Actionable Principles

- Treat each application type set as a different project
- The larger the system, the greater need for deep integration of different knowledge domains
- Substantial time commitment is required for learning application domain
- Avoid building embedded systems -- reduce interdependencies
- Identify expert/exceptional designers or “project gurus”
- Integrate competing perspectives using expert insights
- Communicate domain knowledge to as many developers
- Use expert designer to exert authority
- Engage with technical staff to improve technical knowledge
- Undertake additional learning in case of major changes to business applications
- Overcome company-specific models by coordinating among co-contractors

Table 4: Application Domain Knowledge

Iivari et al. (2004) identified application domain knowledge as one of the five areas of knowledge critical to ISD: (1) domains of IS applications, (2) application domains, (3) domains of IS development processes, (4) technology domain and (5) domain of inter- and intra- organizational context. Although Iivari et al. (2004) focused only on the domains of IS development processes, this study finds that the intricate relationships among all the five domains demand a more integrated approach. Therefore, this study combines the domains of IS application and application domain because the different domains of IS application are likely to be closely linked with application domains. For example, a transaction processing systems (an application domain) are likely to be similar regardless of which domains of IS application the systems are put to use. Similarly, decision support systems (another application domain) will also have the same characteristics regardless of which domain the systems are supporting. The concepts and related actionable principles for application domain knowledge found in the five articles mostly focus on the social and behavioral context related to the expertise required for ISD. The interaction of expertise available with the size and characteristics of the project create challenging dynamics which need to be managed in order for the ISD project to be successful.

Descriptive Concepts and Propositions

- Business vision and process objectives guide process redesign
- High-impact business processes
- Measurement of business processes critical for future improvement
- IT works by leveraging business processes
- Business process prototypes highlight problem areas

Actionable Principles

- Develop clear business vision and process objectives
- Identify high-impact business processes
- Measure business processes
- Identify IT levers
- Build prototypes to help with development process

Table 5: Organizational Alignment
The organizational alignment category concepts and principles are extracted mostly from Davenport and Short’s (1990) classic article on business process redesign and reengineering. The alignment discussed in this category focuses on taking advantage of IT capabilities to improve business processes.

<table>
<thead>
<tr>
<th>Descriptive Concepts and Propositions</th>
<th>Actionable Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Gap between computational structures and required behavior in application domain</td>
<td>- Map required application domain behavior with computational structures</td>
</tr>
<tr>
<td>- Learning costs</td>
<td>- Add time and resources for learning costs</td>
</tr>
<tr>
<td>- Changing customer needs</td>
<td>- Prioritize requirements</td>
</tr>
<tr>
<td>- Customers undergo learning and enlightenment</td>
<td>- Anticipate learning and enlightenment</td>
</tr>
<tr>
<td>- Customers misunderstand tradeoffs between requests, capabilities of existing technology, delivery schedule and cost</td>
<td>- Educate customer of change request impact</td>
</tr>
<tr>
<td>- Customers underestimate impact of change request</td>
<td>- Negotiate conflicting requirements and trade-offs between proxy and real customer</td>
</tr>
<tr>
<td>- Conflicting proxy requirements</td>
<td>- Define clear project mission</td>
</tr>
<tr>
<td>- Project mission</td>
<td>- Tightly coordinate requirement between different components</td>
</tr>
<tr>
<td>- Different interpretation of requirements among different components</td>
<td>- Track unresolved issues</td>
</tr>
<tr>
<td>- Unresolved design issues</td>
<td>- Avoid formalizing level of detail</td>
</tr>
<tr>
<td>- Granularity and level of detail of requirements</td>
<td>- Find a balance between stability and getting the right design</td>
</tr>
<tr>
<td>- Trade-off between getting requirements right and stabilizing requirements</td>
<td>- Accommodate change as a natural process</td>
</tr>
</tbody>
</table>

Table 6: Requirements construction

The requirements construction category focuses on the “wicked problem” (DeGrace & Stahl, 1990) issue of ISD. Different causes of requirement fluctuations and changes resulting in a gap between real and planned (or unplanned) behavior of the system are described in this category.

<table>
<thead>
<tr>
<th>Descriptive Concepts and Propositions</th>
<th>Actionable Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User interface design</strong></td>
<td><strong>Design software to be modular to improve comprehensibility, flexibility and changeability, and manageability</strong></td>
</tr>
<tr>
<td><strong>Database Design</strong></td>
<td><strong>Encourage use of information hiding, independent development</strong></td>
</tr>
<tr>
<td><strong>Software Design</strong></td>
<td></td>
</tr>
<tr>
<td>- Modularization and decomposition</td>
<td></td>
</tr>
<tr>
<td>- Information hiding</td>
<td></td>
</tr>
<tr>
<td>- Independent development</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Architectural design

The framework suggested by Iivari et al. (2004) considered each component (database, user interface, user support, system controls and monitors) as separate ISD performance processes. The five articles suggest that these areas are perhaps elements of the overall system architecture. If this is accurate, the concepts and principles for these elements will differ based on the characteristics of the element. For example, database design may require different concepts of data modeling which will be distinct from general software design. On the other hand, the design of user support systems, user interface and system controls and monitors are essentially similar to software design. Hence, this study considers all of them as part of the overall architecture and groups them in the same category.
### Descriptive Concepts and Propositions

- Differences in communications between layers create communication problems
- Documentation is a weaker form of communication
- Each developer interacts with different communications nets
- Representations and models provide common reference for communications
- Project reviews provides effective communication channel
- Boundary spanners are effective communication channels
- System partitioning impacts communication and coordination
- Company level reviews not effective for communicating design problems
- Different reporting structures impede sufficient communications
- Direct customer interface improves communications
- Multiple customers or interfaces complicates the project

### Actionable Principles

- Encourage more interaction between layers
- Encourage more face-to-face communications among developers
- Identify communication nets and integrate information among different communication nets
- Use standard representations and models for interaction not for static documentation
- Use project reviews as informal communication channel
- Engage a boundary spanner to open channels of communication
- Partition system according to development effectiveness rather than user requirements
- Use informal channels to communicate company-level design problems
- Employ a single reporting chain of command
- Avoid intermediaries between developers and customers
- Reduce the number of customer interfaces

**Table 8: Project organizing**

The largest category extracted from the five articles relate to project organizing and project management. This category focuses on issues surrounding communication and coordination of IS projects, structures that facilitate the progress of complex ISD projects, and the fit between the different development tasks and the management of developers.

### Descriptive Concepts and Propositions

- Behavioral factors have more impact on software productivity than tools or methods
- Application domain knowledge acquired through relevant experience
- Expert coalition – individuals sharing the same design model
- Dominant expert coalition often takes control of development
- Competing coalitions-formed usually by representatives from different companies or different departments

### Actionable Principles

- More focus on behavioral factors than tools or methods
- Hire expert designers with relevant experience
- Use expert coalition to take control of project direction
- Form coalitions early in the project phase
- Negotiate with competing coalitions

**Table 9: People management**

Both project organizing and people management categories overlap in many ways. However, issues relating to people management may transcend the individual project or even company boundaries, therefore, a separate category is maintained taking into consideration the multi-layered nature of ISD.
Descriptive Concepts and Propositions | Actionable Principles
---|---
- Upstream processes exert tremendous impact on software productivity, quality and costs throughout the life cycle | - Focus more on upstream development processes
- Tools and technologies produce low impact on software productivity | - Focus more on behavioral and communication factors rather than tools and technologies
- Increased user involvement leads to increased system usage | - Increase user involvement in the development process

Table 10: Method management

Perhaps the most interesting observation from the five articles is the lack of impact found from methodologies and tools on the productivity and performance of ISD. Instead of relying on methodologies and tools, the five articles suggest focusing more on behavioral and communication factors.

Descriptive Concepts and Propositions | Actionable Principles
---|---
- Social context impact how CASE tools impact development process | - Plan the use of CASE Tools to fit the intention and actions of the developers. If the intent is incremental change revolves around existing systems. If the intent is radical change, major incentives may be required to motivate transformation
- Tools that facilitate sharing, support change, exploration and discovery offer the most benefit to ISD | - Use tools that facilitate sharing, change, and that support exploration and discovery

Table 11: System Development Tools

The five articles suggest that although tools have shown to have produced minimal impact on the productivity of ISD, tools that support communication and coordination will most likely show a positive impact. Tools that focus only on diagramming, documentation or the generation of code are unlikely to produce expected benefits.

LIMITATIONS AND CONCLUSION

This study is limited to only five of the most cited articles. Consequently, many of Livari et al.’s categories could not be analyzed. The choice of the most relevant articles related to ISD from which to select the most cited articles also impacts the results. The test search used by the NVivo software on the article titles, keywords and abstract may not identify and rank the most relevant ISD articles, therefore the threshold for the “most relevant” articles should include more than 100 used in this study so as not to exclude highly cited articles from the “most relevant” pool. Despite all these limitations, this study presents a good start to the process of synthesizing core concepts and actionable principles for the ISD BoK by combining scientometric and content analysis approaches.

REFERENCES