

8-25-1995

The Role of Electronic Documents in Workflow Management

Zhiping Dong
University of Rochester

Rajiv Dewan
University of Rochester

Abraham Seidmann
University of Rochester, seidmannav@ssb-facstaff.ssb.rochester.edu

Follow this and additional works at: <http://aisel.aisnet.org/amcis1995>

Recommended Citation

Dong, Zhiping; Dewan, Rajiv; and Seidmann, Abraham, "The Role of Electronic Documents in Workflow Management" (1995).
AMCIS 1995 Proceedings. 179.
<http://aisel.aisnet.org/amcis1995/179>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 1995 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

The Role of Electronic Documents in Workflow Management

Zhiping Dong, Rajiv Dewan and Abraham Seidmann
William E. Simon Graduate School of Business Administration
University of Rochester, Rochester, NY 14627
seidmannav@ssb-facstaff.ssb.rochester.edu

Document management issues are getting abundant attention from managers and MIS researchers. According to a recent study by the Delphi Consulting company, the 1994 market comprising document-centered software, workflow, and image routing was worth \$4.5 billion. While electronic documents hold the promise of effective support for a reengineered organization, a significant number of operational and informational aspects need to be studied as companies focus more on the ways users consume documents. This is a shift from the traditional emphasis on the document creation process. In addition, the earlier document management systems were designed mainly to automate internal processes in paper-intensive companies such as banks, insurance companies and healthcare providers. In contrast, the current business interest lies in 'compound documents' that involve multiple types of media such as images, video clips, word-processed text, graphics, and even spreadsheets. With these 'compound documents' users can call upon a common repository for multiple output-types.

In this paper we are looking at several research issues that exploit the latent structure of documents and permit creation of customized views. We analyze the generic structure of the document life cycle and critically examine the capabilities and limitations of the information flow models typically employed at each stage. Our analysis points at the central role of integrated work flow management techniques as a key enabler of Business Process Reengineering. While paper documents may be well suited for conventional organizations with a rigid mode of operation, electronic documents better support the reengineered organizations with team-oriented case-groups that require greater flexibility and support for group works [1]. Other factors such as increased time based competition and the increase in the volume of documents due to outsourcing and supply chain management have added urgency to the need for a change. All of these factors point at the need for a fresh approach to document management in organizations.

There are three basic ways of representing documents electronically: Image, formatted and marked-up text. Document Image is the oldest form of electronic document representation. Documents are digitized using scanners and the images are compressed and archived on large capacity memory devices. Digitized documents take much less physical space than storing paper, and they can be viewed on screen, zoomed, cropped and cleaned up using image processing software. Images may be indexed by certain key fields at a later stage. Documents that include graphics integrated with text or

handwritten annotations are prime candidates for image systems. With proper indexing, effective storage, retrieval and sharing is available for imaged documents [2]. However, the storage and telecommunications requirements of high resolution color documents make them economically infeasible for many applications. An application of image system is USAA's (United Services Automobile Association) large scale, optical disk-based operational image system. This system is a success story of combining imaging and workflow technologies [3].

Electronic imaging of documents has the advantage of capturing hard copies into a digital medium, but its support for modification and retrieval is limited. On the other hand, documents created using a word processor or formatter such as Microsoft Word or WordPerfect are created in a digital form to begin with and hence can be managed better using software tools. The formatted text forms the basis of the second kind of electronic documents. These are designed for specific hardware and software systems. While the use of formatted text representation supports personal document creation and storage, it is not as well suited for group work and inter-organizational document exchange. There are ways of making formatted text documents more portable, such as using Adobe Acrobat [4], but these systems do not help with other problems such as search and retrieval when many documents are involved.

To overcome the problems resulting from the lack of structure and context-sensitive search capabilities, several companies have introduced mark-up languages. Mark-ups are an extension of formatting marks placed in documents. This represents the third form of electronic documents. Some mark-ups, such as Nroff and TeX, are called *specific mark-ups* because they describe the appearance of the document and provide only a limited parsing of the document into headers, footers, body, sections, etc. *General Mark-up* languages extend this to identify other user defined fields and entities such as invoice date, shipping date, etc. Having this elementary level description of documents allows for detailed parsing and effective retrieval of documents. Standard Generalized Markup Language (SGML) [5], is the overarching standard among mark-up languages. SGML offers the way to address information at the element level. This provides support for structure-based search such as "Get all the invoices dated in October 1987 with shipping dates in September 1987". SGML facilitates data interchange among different machines regardless of differences among systems, devices, languages, or applications involved. SGML documents can be kept in an integrated information base, for example, invoices can be associated with orders, inventories, and payment. This feature combined with element level parsing helps in search and retrieval. While SGML has been around for several years it has not gained as of yet a wide acceptance. Its value will be realized only once the users base exceeds a critical mass in terms of the relative document exchange traffic.

The three major forms of electronic documents mentioned above have a significant impact on the ways in which users select to manage the documents' life cycle comprising creation or reception from an outside entity, processing, storage, retrieval and publishing. Businesses face different problems in processing documents at each stage of the document life cycle. A typical case in point is the health care industry. By many

estimates, a significant proportion of a medical doctors' time is spent in filling out the forms and documents. The document creation problem gets more involved when the document is created collaboratively by a team dealing with a particular business case. Coordination and control of successive and concurrent changes create the kind of problems that arise in data sharing. Except, that these problems are exaggerated by the nature of document processing. While structured data, such as in a relational database, can be easily segregated at the record level, documents created using conventional technologies cannot. These documents are often managed at a file level, thus making the tasks of configuration control and change notification a lot more difficult. Software Engineering techniques used for managing software production in which large numbers of programmers produce a tightly knit product can be used in conjunction with electronic documents to provide more effective group support.

Storage and retrieval of new documents can also pose vexing problems especially for document intensive industries such as Banking. Some surveys have found that businesses in such industries use 25 to 50% of their office space for storage of paper documents. Closely coupled to the storage problems are the search and retrieval tasks. The lack of structural and contextual information about the content of documents exacerbates the search and retrieval problems. For instance, a typical invoice record may have a number of date fields for shipping, invoicing, and for the purchase order origination. Effective parsing of this invoice into its sub-components is a requisite for handling queries addressed at certain date attributes such as "Find all the January invoices pertaining to parts shipped during last December." These issues are dealt with easily using a structured markup language such as SGML to define the structure of the document. Such marked up documents can be searched on context and appearance.

Traditional MIS departments in many corporations focus on transaction driven file structures and have for many years ignored the area of document management. As a result document management has become disjoint from the corporate Information Systems. With the increased emphasis on business process innovation, document management has moved to the forefront of information systems arena. In addition, recent developments in document representation technologies provide managers with difficult choices when it comes to designing their administrative systems. The three electronic document representation systems, discussed above, are clearly distinguished by their capabilities and limitations from the users' perspective. Our on-going research on document management indicates that each one of these representation modes has a role in different organizational settings and difference phases of document usage. The key parameters relevant to determining the applicability of each of these representations are the mode of origination, search and retrieval requirements, indexing, information volume per document, inter-relationships with other documents, authentication and encryption requirements, workflow integration and monitoring.

References

1. C.A. Ellis, S.J. Gibbs and G.L. Rehn, "Groupware: Some issues and experiences," *Communications of the ACM*, January 1991.

2. Tony Davey, "Document Image processing: an overview," *Text Retrieval: Information First*, Proceedings of the Text Retrieval Conference of the Institute of Information Scientists, 1992.
3. C.A. Plesums and R.W. Bartels, "Large-scale Image Systems: USAA Case Study," *IBM Systems Journal*, Vol. 29, No. 3, 1990.
4. Steve Cunningham, "Electronic Publishing with Adobe Acrobat," *Computer Graphics* Vol. 28, No. 1, February 1994.
5. *The Standard Generalized Markup Language (ISO 8879)*, C.F. Goldfarb, Editor, International Organization for Standardization, Geneva, 1986.