

Division of Research
School of Business Administration
The University of Michigan

June 1991

**INFORMATION RETRIEVAL IN BUSINESS:
AN UNMET CHALLENGE**

Working Paper #663

Michael D. Gordon*
The University of Michigan

*Computer and Information Systems, Graduate School of Business,
University of Michigan, Ann Arbor, MI 48109, Michael Gordon @ub.cc.umich.edu
UserLB63@umichub.bitnet

FOR DISCUSSION PURPOSES ONLY
None of this material is to be quoted or
reproduced without the expressed permission
of the Division of Research

Copyright 1991
The University of Michigan
School of Business Administration
Ann Arbor, Michigan 48109-1234

ABSTRACT

The management of textual information is an important problem for business, yet it is a topic that the academic IS community has virtually ignored.

This paper presents the findings of interviews with thirty nine managers and staff in in fourteen firms. Two underlying questions guided these interviews:

- 1) Is information retrieval (i.e., the management of text and other information not manageable by traditional DBMSs) effective within businesses?
- 2) What technological and procedural remedies may help solve business information retrieval problems, and are they being tried in practice?

The findings from these interviews provide an accumulation of examples documenting the prevalence and severity of business information retrieval problems. Following a discussion of these findings, we look to related literature to help devise a research agenda aimed at improving the practice of information retrieval within business by understanding its nature in a more complete, scientific way.

1. Introduction

The academic Information Systems (IS) community has paid scant attention to the problem of information retrieval -- the ability to store, retrieve, and disseminate "unstructured" information such as reports, memos, newspaper and journal articles, electronic mail, notes or conversations with suppliers, etc. Yet, such unstructured information¹ is more common in organizations than is concise, "structured," tabular information for which traditional database management methods are appropriate and extensively studied.

Two possibilities explain these facts: First, despite the volume of unstructured information, it presents little problem to business. Or, second, the academic IS community has largely overlooked an important problem.

This paper presents the results of interviews conducted in organizations to examine unstructured information handling behavior and its impact on the firm. These results support the second explanation. Failing to manage unstructured information results in the loss of intellectual product, duplication of effort, information overload, and excessive storage costs. The consequences for business are impaired planning and decision making and an overall decrease in competitiveness and profitability.

However, despite the academic IS community's lack of attention to information retrieval in business, considerable attention has been paid by other disciplines to related problems. In particular, computer science and information science have investigated information retrieval for quite some time, mainly from a technical, laboratory perspective. Additionally, some progressive firms are beginning to develop solutions for dealing with their own information retrieval problems.

The goal of this paper is to improve the practice of information retrieval within organizations. To this end, the first half of the paper points out the difficulties with information retrieval that firms currently are experiencing. The second half seeks to better understand and improve upon these problems by establishing their interplay with research from related disciplines. Ten researchable propositions are developed from this analysis. Each pushes us toward a

¹We focus in this paper on electronic and paper documents. We use the term "unstructured" information to also include photographs, computer graphics, images, etc. to which keywords, textual descriptions, or other descriptors can be applied. Thus, they too become manageable by information retrieval methods.

more complete scientific understanding of information retrieval within organizations. As such, each represents an important step on the path toward making information retrieval truly beneficial in the work place.

Section 2 discusses the difference between managing data (with a database management system) and managing unstructured information and indicates the lack of attention the academic IS community has paid to the latter problem. Section 3 describes the way in which the interviews were conducted. Section 4 describes, in interviewees' own words, the problems (and associated costs) with unstructured information that organizations face today. Section 5 discusses information retrieval research issues that should be pursued to improve information retrieval practice.

2. Textual information in organizations: A neglected IS problem

Database management systems are appropriate and powerful tools for storing, manipulating, and retrieving "structured" data. However, text is much more prevalent in organizations than is structured data, and it must be managed by different methods.

Consider personnel and inventory data. Both the high degree of structure of such data and the determinacy of their values make database methods appropriate for their management. The record structure of such data captures all essential characteristics. By assigning one set of fields to personnel records (employee id number, employee name, department, date of hire, etc.) and a different set to inventory records (part number, quantity on hand, etc.), we can precisely frame requests for information. For instance, we can ask to list employees by date of hire, a request that makes no sense for inventory records. Additionally, the fact that data have determinate values provides us with "content addressable" access to information. For instance, we can find all part numbers with inventory below 100 units. Together, well structured data and determinate values provide us with "all-and-only" access: all the records we need and only those records.

But, textual information is not so highly structured, nor does it have equally determinate values. A document (stored electronically or in paper files) is ordinarily represented by several fields. Some provide *content* information, usually in the form of keywords (such as, "Fourth generation language," or "microcomputers"). Others give factual, *contextual* information about the document (such as document author, date of publication, etc.).

There are, however, no guaranteed effective methods for assigning content terms (keywords) to documents in a determinate way: Perceptions of content vary greatly among individuals, and language provides countless ways to express similar ideas. For example, someone looking for documents on "office automation" won't be furnished with documents described by the keyword "local area networks," even though such networks are integral to automated office work. Similarly, documents containing both the subject designations "learning" and "computers" might be relevant to either improving education with the help of computers or to making computers more intelligent -- topics that have little to do with each other. Thus, a searcher looking for documents using this pair of search phrases will identify documents he or she will not find relevant. Further, the contextual information associated with documents is not complete enough for selecting particular documents.

This means that database management models which logically address the retrieval of documents in all-and-only fashion are not appropriate. Instead different models of retrieval must be used to govern retrieval of documents (Blair, 1984).

Though, classically, the information retrieval problem involves finding references to relevant journal articles or books in a library, information retrieval is also a business problem. Any kind of report, memo, or meeting minutes is manageable by information retrieval methods. So, too, are filed exchanges of conversations, electronic mail messages, word processing files, written procedures for accessing certain tasks, passages within computer conferences, or even photographs, graphic or image data that are indexed by keywords or described by accompanying text. In fact, "unstructured" information occupies a much larger position in conducting business than does the "structured," tabular data manageable by database management methods. Whether or not firms have explicit plans, procedures, and software support, they are constantly dealing with unstructured information to keep their businesses going. In short, most business information is in the form of documents.

Ironically, the IS research community has largely ignored the business information retrieval problem. Eleven leading research journals in IS and management were electronically searched to determine the coverage they gave to business information retrieval, and a twelfth was searched manually.² (See Table 1

²The electronic searches employed the full text of documents' abstracts, their titles, and keywords. Searches were not made specifically to find documents on topics related, but not central, to information retrieval, such as executive information systems.

for the journals searched.) All searches covered the five year period from February 1985 through January 1990. All told, these twelve research journals published just under three thousand articles during this five year period. Queries were developed by two information specialists in an attempt to uncover all articles in the twelve journals pertaining to business information retrieval. In all, only fifteen articles were found to have direct relevance to the problem of business information retrieval. Of these, all but two were technical and without a business or managerial focus. (Ten of the relevant articles appeared in *Communications of the ACM*, the most technically oriented computer journal of the twelve searched, and a journal with a stronger computer science than IS following.) In contrast, the trade press has begun to give considerable attention to business information retrieval. For instance, an electronic search of *Computer World* and *Datamation* for articles on information retrieval covering just the fourteen months beginning January 1989 found nineteen relevant articles. Jointly, the statistics concerning trade press versus research press coverage attest to the relevance of the problem of information retrieval for business and the academic IS community's neglect of this problem.

The remainder of this paper focuses on unstructured information retrieval as actually practiced in large corporations.³ We focus on the management of internal textual information (as opposed to external information available by subscription), since many solutions pertinent to problems with the latter are beyond a firm's direct control. We observe firms' information retrieval practices and see the penalties they pay for failing to manage better their unstructured information. Finally, we help define research areas that having strong bearing on helping businesses with their information retrieval problems.

3. Method

In this section, we describe the firms that provided data for this study and the methods used to gather those data.

3.1 Sample

All twenty-four firms participating in a semi-annual forum for information systems executives were invited to participate in this study. This group was targeted because of member firms' expressed interest in effectively managing their

³For a full discussion of information retrieval theory and methods see van Rijsbergen (1979) or Salton and McGill (1983).

business information. As a result, it was felt that these firms would be most likely to have successful information systems for business information retrieval. Fourteen firms accepted the invitation and participated in this study. By conducting interviews with all consenting firms, it was hoped we would obtain a broad description of the current state of business information retrieval practice in major corporations. Collectively, the interviewed firms represent both manufacturing and service industries, and publicly and privately held companies. Participating firms were large organizations, with revenues often exceeding a billion dollars annually (see Table 2).

Altogether thirty-nine people were interviewed, and all interviews were audio tape recorded. Participants held jobs from the executive level down to the clerical level, though most interviewees were upper level managers (see Table 3). Various functional areas were represented, including legal, purchasing, planning, and information systems. Interviews ran from forty-five minutes to several hours, and each interview was completed in a single session. Although an attempt was made to interview a cross section of people at a variety of firms, participants interviewed constituted neither a random nor stratified sample from the companies or industries represented.

3.2 Choice of method

These interviews were conducted in an effort to elicit illustrative examples which reveal the current state of practice of unstructured information management, especially management of text. In-depth questioning, conducted in person without a structured questionnaire, allowed the exploration of issues that could not be fully anticipated. Two underlying questions guided this research:

- 1) Is information retrieval effective within businesses?
- 2) What remedies (technologies and human information handling activities) may help solve business information retrieval problems, and are these remedies being tried in practice?

All individuals interviewed experienced information retrieval difficulties. Even the best organized individuals felt the need for better tools and better means of sharing. Yet, people often felt reluctant, hesitant, or embarrassed admitting their information retrieval failures, as if their difficulties indicated a deficiency with a trivial problem they ought to have mastered. In contrast, we present an accumulation of examples of information retrieval failures to indicate the

prevalence and severity of this generally neglected problem. In short, we have taken a qualitative approach which is appropriate for providing a description of how information retrieval problems can interfere with efficient and effective business and for generating theory, (Eisenhardt, 1989).

We have not attempted a statistically rigorous analysis or traditional IS case study analysis (Benbasat et al., 1987). Instead, this work should be regarded as an attempt to delineate a set of problems generally neglected by academic IS, show their importance, and help improve businesses' practice of information retrieval by pointing out research necessary to correct their current problems.

4. Interview results

Findings from the interviews are grouped into four categories: problems with searching and losing; problems with sharing; problems with information overload and storage volume; and organizational issues for effective business information retrieval. For each category, we discuss associated IS research questions in section 5.

4.1 Searching and losing

People (and organizations) store documents so they can later make use of them. But searching for missing information takes time, and losing information wastes ideas, evidence, and know-how. Both can severely affect a firm's performance and even threaten its survival. Unfortunately, these costs are too common an occurrence.

4.1.1 Losing time

Professional workers are estimated to spend 25% of their time distributing, filing, and retrieving documents (Yourdon, 1986), and some professional groups spend more (McNurlin, 1989; Black, 1990). A difficulty in locating information affects all functions and all strata. A manager of research and development described spending an entire day looking for a strategic planning document he had completed about six months earlier. An attorney summed up the situation concisely: "We have a significant problem locating documents. We waste a lot of time." Even the executive level is affected. One vice president explained that certain searches were simply too complicated to be delegated. As he explained: "Today I looked with my secretary for all documents [on a particular topic] for over

an hour. We wanted more [text, drawings, and diagrams] but just didn't know where to look. [Such occurrences] arise several times per week."

The reason for all this effort is that certain information is mandatory for business to be conducted effectively. Consequently, a manager of end user computing described conducting the following search with another high level manager: "The two of us looked for a letter from my boss to the manager of operations. We each looked for one [entire] week. ... And his secretary also helped for half a day."

4.1.2 Loss of proof, fact, or experience

Observed a purchasing manager: "There have been many instances where you've got to prove it by pulling hard copy. The problem begins when it's the [three year old] paperwork you're looking for, not the [current paperwork]." One of the failed corporate-wide paper hunts he mentioned involved a \$100 million anti-trust finding against an industry from which his company bought raw goods. All U.S. customers of that industry had to prove how much they had purchased during the seven years covered by the finding. "We had to send in reams of documents. Without the [purchase] orders, you're lost. And [we couldn't] find everything [we needed] ... from [our] archives."

The biggest costs to organizations of unlocateable documents involve disrupting or or delaying the ability to refer to needed information. A document contains dates and facts which must be known for proof or verification. The reasoning behind some action, a carefully reasoned policy statement, or an explanation of technical operation may be found in passages of certain documents. A vice president in charge of operations added that "Some of that stuff [that I've filed] I will get called on a couple years later and [asked] 'Why did we do that?' or 'What was the approval process?' And I need [to know]...I would be [in] very sad [shape] if these [files] went away because I periodically have to call on them [to address building problems, agreements, systems], and it helps keep things from being innuendo and keeps things down to the fact." At the CIA, where great volumes of unstructured intelligence information is processed, intelligence analysts consider their personal files their most important source of intelligence information (Lancaster, 1978).

A document can encode intense, sustained intellectual activity for which individuals are highly trained and well paid. Such knowledge is part of the information backbone of an organization. An analyst described "dissecting a computer program for [many] hours to solve some really critical problems" because

the documentation was lost or thrown out. With the documentation, the problem could have been solved much more quickly.

4.1.3 Contorted work

Patterns of work develop around the knowledge that finding documents is difficult. A high level manager was discussing trying to assemble all the relevant information from others in his firm in trying to decide an issue: "I don't do that too much. A lot of that may be that we know it would be almost impossible to do." Thus, the technical decisions he makes for the company don't benefit from pooled, collective experience and wisdom. More subtly, a search may be concluded on the mistaken assumption that all relevant materials have been located. As an attorney said: "[The documents we need to support our work] generally turn up where they should. But, then, we only look there." In fact, even with the support of electronic information retrieval systems, searchers conclude searches far before they have retrieved all the relevant information on a topic -- even when they are confident they have retrieved it all (Blair and Maron, 1985).

A manager of strategic planning described his difficulty in preparing plans, since 25% of the information he needed came from other people's files. He described his uneasiness about not knowing if his plans reflected his organization's true concerns since he was never sure if he had all the information he needed.

Failing to locate needed information can cause after-the-fact recollection from memory to replace more accurate documentation of experience. One individual explained his need to redo from memory a missing, six month old project plan needed by a corporate Vice President. After several days of attempting to reconstruct missing facts and figures the document was recreated: "Let's just say I got away with it," although the re-creation was nowhere as complete or accurate as the original. "That does happen with regularity." Another manager expressed similar uneasiness, explaining that he usually had to re-think and rewrite a lost document every few months. To anyone who has ever re-read something he or she has written, the point is clear: There is an immediacy to thought that makes even your own thinking seem foreign a few months later. To recreate lost or missing documents by relying on memory, often without notes, is risky business, sometimes putting a firm or an individual in a legally vulnerable position.

Of course, inefficient and ineffective patterns of work affect work groups or entire organizations. For instance, the Texas State Board of Insurance, Workers' Compensation Division has used paper files exclusively to support 95% of its policy reviews and rate-making decisions. As its incoming paperwork grew 50% to 5

million pages of information in 150,000 file folders, the Board's adherence to manual paper processing methods cut its throughput in half (Black 1990).

In summary, the impact of yesterday's documents on today's work is illustrated by the remarks of a senior level manager responsible for business tactics in several technical areas. When asked if there was ever a document he had looked for but been unable to find, he replied, "I never found it many times. Some would have made my life a lot easier."

4.2 Sharing

Falsely, some people feel they can have under their control all the information they need without sharing. For example, as Lancaster (1978) described: CIA analysts charged with producing "finished" intelligence require comprehensive information about some topic in which they specialize: some aspect of science and technology, medicine, economics, etc. Despite their comprehensive need, these analysts felt they could personally assemble, file and locate all the information they would require. Studies showed they were wrong: analysts' personal files were not complete, and "major value" items were actually stored elsewhere in the organization. As a result, the CIA maintains company-wide intelligence files for both archival and current awareness purposes.

An organization can do what no individual possibly could, and sharing of information is vital to coordinate and execute its various activities, from operations through strategic planning. We focus in this section on problems surrounding sharing information.

4.2.1 Out there ... but where?

Few organizations maintain active, departmental or corporate-wide files to support workers' overlapping information needs. An attorney said: "In the legal department here, everybody writes documents of some type that can be used again. A sales agreement, a distribution agreement, you name it. An agreement for the purchase of sale of the business. But I'll be damned if I know who wrote what, on what subject, or where they are. But I know it's out there. So I've got to hunt these guys down." A director of administration, whose job includes being a disseminator of information, gets information "catch as catch can." Different departments inform her differently, some with regular reports, some erratically, and some not at all. Several times per week she must spend over an hour tracking down information that others failed to provide her so she can answer someone's request.

In fact, sharing can be impossible even when one knows who has already done work that could have direct bearing on a current situation. A vice president interviewed said, "Today I needed some files maintained by [someone who was no longer with the organization]. I couldn't use his files. Yet, you could have immediately gotten what you wanted from them if he was there." Such problems arise through hirings, firings, job reassignments, and corporate reorganizations.

4.2.2 Consequence: replicated resources

Some firms recognize that duplicating information wastes resources. This sentiment was expressed by an information systems executive in a manufacturing organization: "We have replicated information systems groups [within our organization], and even within one of our operating companies at different locations. It would be nice [for one group] to say 'Is there any other work going in this area?'" Actually, such duplication of effort may never even be suspected. In one large company, several separate groups independently formed committees to evaluate the acquisition of main frame-based text retrieval software, at considerable cost to each group. When such duplication was discovered after the fact, the possibility of sharing knowledge and, thus, reducing expense had passed. "We never even knew it," one committee member said. Further, even if one suspects that others across the organization need to be informed about a change to a document or about some newly gained information, it may be impossible to tell who should be notified. As an example of effective resource usage, a group of 48 energy researchers reported average savings of just under \$1300 for every report they read (with maximum savings of \$1.5 million) (Repo, 1987). These savings came from being able to avoid having either to repeat investigations or gather together information already available. With information systems designed for broadcasting and archiving information, costs from duplicated effort are avoidable. Further, projects formerly requiring too great a research expense can be undertaken by exploiting the knowledge already existing within the organization.

4.2.3 Consequence: losing opportunity and courting disaster

Companies fail to share documented findings of studies (for instance the evaluation of text retrieval software) and changes in policy or practice (engineering change notices are often communicated erratically). Information from the field (obtained by salespeople, customers, etc.) has a difficult time finding its way to those who can make use of it (Gilad and Gilad, 1986). On occasion, lack of shared information can potentially precipitate a catastrophe. Senior engineers' formal memoranda to management about sticking cooling valves possibly leading to a meltdown at Three

Mile Island pre-dated the actual accident by two and a half years (Burns, 1987). Their memoranda, submitted on the wrong form, were misdirected and ignored, as were their further urgings to take action. So were notices that the poor human-factors design would be a severe handicap during a crisis -- another prediction that was borne out during the accident.

4.2.4 Consequence: Lost work product

Organizations have their own way of doing things, procedures they follow, and resources they use to get their work done. Often, this knowledge is tacit, a fact which can painfully come to light during times of change. An attorney, describing the situation in his firm, concluded, "We do not do a very good job of capturing the work product. It's hard to know how to do your job if you're new or someone leaves. It would be very helpful just to know where to look." An information system that lets you "know where to look" can alleviate a good amount of disruption during transition.

4.3 Overload and volume

Our age of information has spawned an overload of information -- both paper and electronic. American business deals with 400 billion paper documents, a number which is growing by 70 billion a year (Yourdon, 1986).

4.3.1 Extent and toll

The purchasing manager who "start[s] every day with a six to eight inch lump of paper...You just don't know where to start," and the information planning executive who says "It takes me at least an hour a day just to sort my paper" experience the same problem: it takes time and effort just to filter irrelevant incoming information so that work can begin. This overload can exact its toll. Ackoff (1967) has discussed the "misinformation" that comes from too much information.

Overload by paper is being accelerated by the ease with which one can prepare, revise, and print manuscripts. The advent and widespread use of technologies such as word processing software, computer typesetting, and laser printers on local area networks mean that a greater number of easy-to-produce, nice-to-look-at documents will be distributed internally in the form of reports, memos, etc. In addition, professional quality newsletters, solicitations, product announcements, etc. will flow increasingly from the outside to further overwhelm business with paper.

Similarly, electronic transmission of documents is an accepted method of communication in many businesses today. Electronic mail and facsimile transmissions, especially in conjunction with distribution lists which automatically send a single message to a pre-selected group of people, have spawned a new kind of overload: electronic junk mail (Denning, 1982). Posting of messages to computer bulletin boards and in computer conferences create new sources of information to contend with.

4.3.4 Storage Volume and Cost

How much unstructured information do people and organizations store? And what are the costs of their doing so? As we will see, excessive storage costs add to problems we have already discussed surrounding losing, failing to share, searching for, and being overloaded by information.

Secretaries, managers, and those on the shop floor store documents. A vice president interviewed maintains 80 file drawers of documents, each about three-fourths full. An organization that measured what it was maintaining discovered five years worth of records amounting to eleven linear miles of stacked paper.

This is a business age in which it costs four cents to file every document, \$30 to process a purchase order, and \$68 dollars for a single misfile (Swartz, 1986). As we have seen, American business deals with 400 billion paper documents, a number which is growing by 70 billion documents per year. This proliferation of paper is accompanied by the increasing ease with which one can electronically create and distribute multiple copies of documents using word processors and electronic mail networks.

Training in the area of storage and retrieval is almost nonexistent. In response, people tend to over-retain. One planning manager reported being "afraid to throw away a couple of old filing cabinets [worth of information]. And a lot is duplicates of others' files." Estimates are that at least fifty percent of a company's records are duplicates, twenty five percent of remaining records are worthless, and 85% of all filed documents are never referred to again (Swartz, 1986). Such behavior comes from hoping that personal retention and organization of large amounts of information will make needed information accessible, despite the penalties: increased storage costs and, what is worse, more difficult retrieval due to increased volume.

Certainly, savings from reduced storage are admirable. But we must not measure the effectiveness of information management solely in terms of cubic feet of storage and costs of filing paper and electronic documents. We must always

consider the uses to which documents are put and the ways in which inattention to their management can disrupt people's work.

4.4 Organizational issues

By design or default, companies manage their unstructured information. The impetus for improving information management in most organizations is the realization that neglect costs money.

Some progressive firms are beginning to make it easier to store, locate, disseminate, and share information. These activities parallel efforts abroad at providing retrieval for unstructured information. The Japanese are replacing the custom design and development of subcomponents of new products by searching for existing subcomponents catalogued along any of several retrieval dimensions, including: part geometry, the process used to make the part, the functionality of the part, etc (Gunn, 1981). Additionally, MITI has sponsored a major national project, Sigma, to create a national software library where Japanese companies may obtain free subroutines (Haavind, 1986).

But, effective retrieval of unstructured information remains a fantasy in most U.S. organizations. Companies typically employ computer systems for unstructured retrieval only after they dramatically demonstrate their benefits or because equally dramatic consequences could have been avoided by their use. A consultant to a manufacturing firm described "an unbelievably successful lobbying effort [using text retrieval software] ... which saved \$700 million dollars ... [and] changed the way [they] did business. They scanned on O.C.R. ten years worth of documents and took them out of the file cabinets." Another large company involved in a suit brought against the entire chemical industry had a surprisingly small settlement against it compared to its smaller co-defendants. Head counsel immediately attributed this savings to his company's ability to locate textual evidence necessary to support its legal claims. A less happy outcome surrounded the "uninformed lower level manager who incinerated some of the archives. [The company] found during litigation that they couldn't defend. They changed their business, [making] wholesale use of electronic information now."

The lines are drawn in many corporations over "old ways" vs. "new ways." An attorney in one corporation described "the enormous amount of textual information [we] manipulate," adding "It's absolutely essential to automate." Head counsel at another large firm stated he was still unconvinced that computers had any place in law. His feelings echoed those related by a information specialist at a

facilities management organization describing a firm resisting the computer:

"We've been doing business the same way for seventy-five years. Why change now?"

Several implicit, but fallacious, economic arguments hamper the use of information retrieval technology to make management, and business in general, more effective and efficient. First, information technology costs money, but non-computerized information systems are free. A consultant describing a firm with massive amounts of stored paper explained that they were keeping "eleven linear miles at three and one half cents a page ... and they had no idea what it cost to store, let alone ... retrieve, let alone the value [of being able to effectively retrieve it to help them conduct their business]." Second, the costs of information technology are real, but not the benefits. The information systems staff at one firm interviewed was asked if text retrieval software that could contribute a ten percent advantage in winning a \$700 million lawsuit could be easily sold to management. "[No, because] we look to hard savings. The DASD [direct access storage device] requirements [would] choke our hardware planners who [would say] 'Can you really justify the cost?'" Similarly, equating information retrieval benefits with reductions in storage costs ignores that fact that cheap storage (on fiche, for example) can reduce the number of retrieval cues associated with documents and worse a searcher's prospects of finding relevant information (Blair, 1984). Third, benefits must have a strong local effect to be truly meaningful. However, information storage, retrieval, sharing, and filtering affect many people, diffusing their benefits across the corporation. In fact, as several information systems managers suggested, information storage/retrieval/sharing systems can have their greatest effects across departmental boundaries within a given organization by allowing better exchange among people who would ordinarily be isolated from each other. As a result, it is hard to find a vigorous champion for such systems, let alone someone to incur the expense of deploying one.

5. Information Retrieval Research for IS

The challenge for IS research is to assess more completely the extent of existing business information retrieval problems, to better understand the preliminary steps being taken to deal with these problems by some innovative firms, and to appropriate or adapt successful retrieval methods studied and developed by other disciplines. In this section, we make suggestions for research relating to

these issues. Our research suggestions take the form of questions, novel remedies, and brief sketches of proposed research.

The issues and questions we raise point to research areas that are important for IS to investigate. We ignore important research issues of a strictly technical nature that are actively being investigated by other disciplines. In some instances, the appropriate investigation should be qualitative (e.g., case study); in others, more traditional quantitative methods would be preferred. No matter what methods are followed, the aim of these investigations should be to develop both a descriptive and normative theory for business information retrieval. In this effort, we present a set of researchable propositions. We group together research topics to parallel the presentation in earlier sections of the paper.

5.1 Searching and losing.

Benchmarks. We have seen many examples of problems with finding information. Still, the question remains: How big is the information retrieval problem for business? At present, we have no satisfactory answer.

The question of retrieval effectiveness has been raised in the computer and information sciences, but without a business focus. In one recent study (Blair and Maron, 1985; Blair, 1990), for instance, we find that electronic, full-text retrieval provided fewer than one relevant document in five. In other words, with a medium-size document database (40,000 legal documents), being able to retrieve a document by searching for given patterns within its complete text leaves over 80% of the documents relevant to an inquirer's need unretrieved.

To understand more completely the extent of the problem, we need to repeat such investigations to consider a variety of influencing factors. First, business information retrieval solutions are not likely to not scale up. The chief trial lawyer in a major, eight year lawsuit involving 22.5 million documents discussed his firm's wholly unsatisfactory information retrieval despite using a variety of retrieval methods: full-text, manual coding, controlled vocabularies, etc. Although he had no formal evidence, he suspected considerably worse retrieval than Blair and Maron's (upper bound) hit rate of one relevant document in five. The explanation for such a phenomenon is that the language problems that beset smaller scale retrieval -- such as the seemingly indefinite number of phrases to express a given concept, and the the wide variety of meanings attributed to any given word or phrase (Swanson, 1960; Blair, 1990) -- can be more severe for larger systems where the combinatorial explosion of language can more completely come into play.

Everyday experience illustrates this phenomenon. For instance, knowing your personal library, it may be perfectly sensible for me to ask for your book about "database management systems." In the context of a research library, a similar request would flood me with books, most of which would be irrelevant.

Second, the consistency of language will vary in different organizational contexts. For instance, in disciplines such as engineering, law, or even information systems, there is a technical vocabulary that helps to limit the linguistic variety we see in less technical areas such as advertising or corporate strategy. As a result, we expect more effective information retrieval for disciplines that use language in a more consistent way (Lancaster, 1979).

Third, information retrieval for informal information will be less effective than for traditional "archival" information. Within a firm, information retrieval may help manage a variety of textual information, from SEC filings and newspaper and magazine articles, to official internal reports and policy statements, to hastily worded memos. With less formal information, we expect more linguistic variety and, again, more difficulty with retrieval. More formal documents provide fuller detail (e.g., referring to James Smith as "Smith" or "Mr. Smith," for example, and not "Smitty," "James," "Jimmy," "J.S." etc.) and more complete context (avoiding such phrases as "the unfortunate incident which you're already aware of" in favor of more precise descriptions). We have the potential today to allow the retrieval of word processed documents which are already in electronic form as well as to scan and use OCR on existing paper paper documents to allow them to be retrieved. Thus, we need to know more about how effectively we can retrieve such documents, many of which will be quite informal.

Overall, we conclude that

Proposition 1: Information retrieval is a significant problem for business (with fewer than 20% of sought documents being found). Further, the extent of the problem worsens

- a) with increasing size of the document collection;
- b) with a reduction in the technical vocabulary associated with the users of a document collection; and
- c) the less formal the information being stored.

To date, studies of information retrieval effectiveness have almost always been based on small scale document collections (generally fewer than 5,000

documents) that have been specially constructed or isolated from larger collections for experimental purposes (see, for example, Sparck Jones, 1981). Further, there has been no attempt to focus on differences in retrieval effectiveness for different business functions or contexts, or differences arising from formal versus informal retrieval. We view studies aimed at exploring these issues as being important benchmarks. Such studies should provide data familiar to information retrieval researchers -- such as recall-precision curves⁴ -- as well as evidence relating these data to the costs organizations incur for a) failures to retrieve relevant information and b) retrieving information that is not relevant.

Naming. At the heart of the solution to problems of business information retrieval is improving inappropriate and/or inconsistent naming of documents. A secretary described her exasperation in searching for documents for her boss, explaining that, to her, his file system was incomprehensible: To file, he "picks a different name every time [for the same topic] or simply circles the name on a memo he receives from various people on some project ... [Retrieval becomes] really difficult because ... there may be ten files in there [on the same topic] under ten different file headings." More dramatically, but in the same spirit, a company conducted a half million dollars of research unnecessarily because of improperly indexed (named) patents (Repo, 1987)

Research involving naming documents (i.e., describing their subject contents) needs to be done with respect to the analysis, development, and application of supporting tools for imposing structure on a vocabulary of terms. Blair and Gordon (1990) have described the fallacy of regarding business information retrieval as a direct extension of the problem of retrieving documents from a library, the differences in the two forms of retrieval including attitudes regarding what information is important to retain, the costs associated with errors in retrieval, etc. In the same way, we must investigate whether traditional methods used in libraries, such as controlled vocabularies and thesauruses (Lancaster, 1986) are as effective for freer-form business information retrieval. Because language difficulties so greatly interfere with effective retrieval (Blair, 1990), we suspect that

⁴Recall is defined as the proportion of relevant documents that is retrieved. Precision is defined as the proportion of retrieved documents that is relevant. A recall-precision curve describes performance of a system over time by plotting precision for various level of recall.

Proposition 2: Naming conventions, such as those supported by controlled vocabularies and thesauruses, will improve retrieval in comparison to situations in which document descriptions and query terms may be constructed without any restrictions.

Analysis should also consider how business terminology is most effectively structured: In a strict hierarchy, in a network? Differently from individual to individual, or department to department? Further, what instruments need to be developed to analyze the effectiveness of such a structure short of implementing a retrieval system that uses this structure and then testing its effectiveness? The development of structuring tools must also consider the type of software best able to fulfill the tool's function. A hypertext system, for example, may be effective in allowing navigation among a set of interrelated concepts. On the other hand, rich hypertext linkages may hide the overall structure that best serves business information retrieval.

Knowledge-based methods. Newer methods of retrieval are being tried in the work place. For instance, the National Library of Medicine has developed a sophisticated form of vocabulary control that relies on knowledge-based indexing (Humphrey, 1989). Using hierarchically arranged knowledge of medicine, the system suggests, restricts, and automatically applies certain index phrases to documents based on other phrases applied already. For instance, the subject "human" is automatically associated with a document indexed by the term "child." Similarly, a document could not be indexed to indicate the medical complication "biopsy," since the system knows a biopsy is a diagnostic method and not a type of complication. In these ways, subject phrases are applied more consistently, with deeper semantic interrelationships, and with less effort on the part of indexers. Retrieval is made more effective by reducing the problems of linguistic indeterminacy.

For a second example, the RUBRIC system has devised a special set of rules to support the need for comprehensive information retrieval by a group of governmental intelligence analysts (Tong et al., 1987). For this group: 1) important results depend on their ability to locate information; 2) the information covers a relatively narrow topic and is required on a recurring basis; 3) the users' information requirements are very well understood; and 4) the needed information is likely to contain specialized terminology that indicates the content of documents. Rules for each RUBRIC user are custom built. Some rules help define highly

specific concepts the user is interested in, and others help indicate how specific words in document texts should serve as evidence for the occurrence of a topic. The system applies these rules to the full text of a document to predict the document's relevance.

One can easily imagine similar knowledge-based information retrieval systems for other information seekers with specialized needs: financial analysts, market analysts, patent attorneys, etc. Since knowledge based methods focus on smaller areas of discourse, they overcome some of the linguistic difficulties that surround more broadly conceived information retrieval systems. In addition, knowledge-based strategies help move information retrieval from an inductive (van Rijsbergen, 1979) to a deductive basis. With this shift, information retrieval is based on a simpler, more computationally tractable model of inference. Overall,

Proposition 3: Knowledge-based methods of business information retrieval promise to outperform more traditional methods.

Further, knowledge-based information retrieval systems promise to surpass traditional methods in effectiveness if they are targeted for the right group of users. As Strassmann (1985) argues, the greatest chance for successful deployment of information technology comes from having high performance workers exploit the technology and share their experiences of success with others. Knowledge-based systems should be developed for such groups of people. Careful study is required to determine which individuals (and business functions) can most greatly benefit from such support.

The development of knowledge-based information retrieval systems depends on extensive knowledge acquisition. This activity is far from completely understood in the development of *traditional* expert systems, for which the goal is to devise a set of production rules that embody the expertise to *solve* some problem. For information retrieval, the knowledge to be acquired pertains to defining topics of interest, specifying sources and formats of information, identifying textual cues that suggest certain topics, describing relationships among ideas, etc., instead of problem solving. Thus,

Proposition 4: Traditional knowledge acquisition methods for devising a set of production rules are unlikely to be unsuccessful.

Rather, new methods of knowledge acquisition will be required to support knowledge-based information retrieval requirements, especially if one envisions such a facility being useful and available to a large number of information workers with a variety of different specialized needs.

5.2 Sharing

Information retrieval systems promote both ad hoc sharing (for information needs that do not recur regularly) and coordinated sharing (for information needs that are known in advance to recur frequently). We consider both types of sharing.

Ad hoc sharing

Extent of problem Problems with sharing often come to light by accident when, for example, two individuals or two groups unexpectedly discover that they have independently produced, gathered, assembled, or analyzed the same information. On the other hand, the failure to make such discoveries in no way suggests that companies are successful in sharing information. Rather, it may simply indicate that an accidental discovery has not occurred.

Swanson (1987) demonstrated how key linkages in the published medical literature had never been made, some of which suggest treatments for untreatable diseases. Such bibliographic isolation within the medical community, for whom access to information is vital for treatment and scientific advancement, suggests the scope of the problem within the business community. In medicine, there is a formal, published body of archival information to which all parties have free access. Further, despite different fields of specialization, the medical community's common scientific and educational background and members' pursuit of common goals improves the prospects of sharing and unintended discovery. In contrast, business sharing is impeded by the dispersion of information throughout an organization and by different focuses and goals among individuals in different departments or different divisions. Further, a published, scientific literature rewards contributions of information for group consumption and scientific custom demands recognition of previous, related contributions. In business, however, political and personal motivations may discourage sharing. In sum,

Proposition 5: Failures to share are frequent occurrences in business. Such failures impose significant cost on organizations due to duplicated effort or actions that are taken by people who are less than fully informed.

Descriptive research can help document the extent of such failures as well as the most likely participants in such failures. We also need empirically validated communication models that predict the degree and effectiveness of sharing as more people become both potential contributors to and recipients from shared information bases.

Methods. Several of the organizations interviewed are investigating systems for information sharing or are in the early stages of developing them. Applications include an online corporate-wide information base providing shared access to a Controller's manual covering corporate financial procedures, shared text bases providing departmental access for managerial decision-making, and shared text information bases to support legal departments and technical divisions. Non-text projects include information bases for retrieving engineering drawings and a facility for locating and retrieving appropriate subsets of statistical databases to provide input for graphics-aided decision making. Development of these shared repositories requires attention to information quality and information base size.

Quality standards can be established in either of two ways (Warren, 1985). Information to be shared can be examined for its factualness, comprehensiveness, clarity of presentation, timeliness, potential relevance, and other factors. Alternatively, information quality can be inferred from usage patterns: High quality documents tend to refer to other documents of high quality, as do reputable authorities. Inferring quality from empirical usage data is likely to be easier than assessing a document's factualness, comprehensiveness, etc., although the latter method may relate more directly to a searcher's perception of quality. Further, when selection of information for a shared information base is based on quality, these two standards may lead to the selection of different materials for sharing. Research should compare the effects of each form of standard on selection, compare both sets of selected documents with what it is important for firm members to actually share, and address the relative costs of maintaining quality in these two ways. Selecting information for sharing based on quality standards is especially important when we consider that information users are more strongly influenced by ease of access than by quality of information (O'Reilly, 1982). In other words, they seek higher quality information over lower quality only when it is equally accessible. With electronic storage and retrieval, however, we have the opportunity to make most accessible the information that is also of highest quality.

Size of a shared information base is another important concern. Unlike database management systems for managing "factual" data, larger, more comprehensive textual information bases increase the likelihood of a searcher retrieving non-relevant information (Lancaster, 1979). On the other hand, smaller information bases can fail to be comprehensive enough to support the tasks they are to support. A careful assessment must be made to better understand these competing pressures. Such assessments will most likely require empirical investigation of a system's effectiveness, though simulation studies (Gordon, 1990) may be effective, too.

These concerns with quality standards and information base size suggest that

Proposition 6: When developing a shared information base, information quality and information base size are both important factors for success.

Finally, although I may *guess* what information of mine you might need, I will rarely *know* what information of yours will truly help me. Thus elicitation methods which ask users to share whatever information they feel others can use take an indirect approach to establishing a shared information base. Instead, methods of elicitation for establishing shared information bases for work groups or organizations should be developed and studied in an effort to permit sharing without overwhelming. For instance, an expert system could incorporate the quality standards and the stated information needs of a particular *group*. Then, whenever a group member completed a document, he or she would be queried by the system regarding its quality and information content. Saved documents (i.e., those available for sharing) would be those of high quality that the system deemed to match the information needs of others.

An even more challenging task is the prospect of finding key linkages within a firm without any advance preparation. For instance, it would be useful for a firm to identify all employees with expertise in a given area or locate any internal study already completed regarding a particular market. Swanson (1988, 1989) has begun exploring a systematic process for making such discoveries for archived, indexed publications. In essence, a "seed" concept initiates an iterative bibliographic search aimed at discovering a pattern of ideas related to the seed. Each of these is used to uncover other ideas, and the convergence of these ideas ultimately serves as a source of evidence concerning new, undiscovered linkages to the seed concept. This tack has

been successful to a degree, but requires elaboration and extensive research, especially if one envisions appropriating such a method for use in discovering linkages among less formal sources of information within a firm (including informal reports, memos, and the generally untapped information people keep in their heads).

Coordinated Sharing

A recent trend in information retrieval is the development of optical image processing systems that can assist in coordinating information transactions. For instance, FileNet's system establishes high speed communication networks that link massive storage devices and employees working at high resolution workstations. The system centrally stores a single optical image of a scanned document to which many people can have access and begins to automate the processes of filing, storing, routing, and basing actions on documents. For instance, the system can be instructed that only when two letters of reference plus a statement of credit worthiness from a bank have been input to the system for a customer should his or her loan be evaluated. Rules to encode such paper flows (written as a FileNet WorkFlo "script") mean that a loan officer never needs to look to see if a particular application is ready to be processed nor ever needs to look for a complete set of materials ready for her to work on. Instead, the application and required information is automatically gathered together and routed to her (optically), ready for her action, only when -- and as soon as -- it has arrived at the system.

With the advent of such systems for sharing information and coordinating work, research is required to identify those factors important in insuring their successful usage. As a starting point,

Proposition 7: A list of success factors for optical image processing includes applications in which:

- 1) supporting documents are easily categorized by type (so that each type may be managed by an appropriate work flow script)
- 2) such information is of high volume (on the order of ten thousand pages a day or ten million pages in storage) (McNurlin, 1989)
- 3) each document requires various actions to be taken by many people (thus presenting problems with coordination)
- 4) each action involves multiple documents (more coordination problems)
- 5) there is a known sequence to these actions (to permit script writing)

- 6) the information initiating each action is created or arrives at the system at unpredictable times.

Research regarding these factors for success might be conducted by formal analytic models and/or field study.

5.3 Overload and Volume

Filters can help sort information by subject or context and are under development for electronic communications and information available publicly online. For instance, the Lens system uses personalized rule-bases to direct messages, text, etc. to appropriate users and to establish priorities among documents (Malone et al., 1987). In this way, the most important messages and documents can be brought to a user's attention first, and others can be automatically discarded. The system relies on the information it manages being partially decomposed into structured fields. Filtering mechanisms not relying on such decomposition are available, too. For instance, filtering full-text documents can be assisted by exploiting certain rules of grammar that help distinguish different meanings constructed from similar vocabulary (such as "expert systems" vs. "systems expert") (see, for example, Meltzer and Haas, 1989; or Salton and Smith, 1989). We expect then that

Proposition 8: Text filters will improve the precision (percentage of retrieved documents that are relevant) of business information retrieval systems by screening from searchers irrelevant documents that contain surface similarities to relevant documents (such as similar wording or phrasing).

Other media are now as cheap as paper for storage. Because of their convenience and durability, optical technologies must be given very serious consideration as a storage medium for replacing both paper and magnetic memory. One firm interviewed justifies putting on CD-ROM any information it needs to distribute thirty or more copies of. Other firms are following suit. Digital scanning and OCR capabilities increase the attractiveness of such means of storage and distribution. In fact,

Proposition 9: The reduction in volume of paper and the improved patterns of communication, information access, and decision making make CD-ROM a cost effective storage and transmittal medium for widely distributed archival information.

5.4 Organizational and behavioral issues

The existence of a tool to support information retrieval does not ensure it will be used or used properly. In part, users' tastes will dictate which tools will be used most often. Assessments can be made of which methods of retrieval the business community would actually use. For instance, would full-text retrieval be better accepted than keyword-based retrieval? Would a vector-based retrieval system (Salton and McGill, 1983) be used? Similarly, research can indicate if users would electronically make personal annotations to a shared text file, or whether they would provide feedback that could be incorporated by an adaptive system (Gordon 1988a, 1988b). Such assessments might take the form of prototype development or well validated questionnaire. Further, adoption of a tool may be based on various factors, such as perceived usefulness and perceived ease of use (Davis et al., 1989). Such relationships deserve exploration in the context of business information retrieval. Even with information retrieval tools available, searchers may be confused about their proper use. For instance, well educated, well trained searchers still have difficulty in correctly using Boolean connectives (Borgman, 1985).

In addition, problems involving ineffective and inefficient human information handling procedures and bloated organizations contribute to the problems of storing, sharing, and finding information. Currently, for instance, an engineering change notice can involve up to 450 communications, a customer request within a manufacturing firm up to 120, and one within a service firm 300. Ninety percent of the information communicated is not new but simply a reformulation for internal purposes (Strassmann, 1985). Thus, it would not be surprising to learn that better procedures and more streamlined organizations can help with business information retrieval. It is important, however, to determine what information retrieval problems would still remain if these steps were taken so that we may understand more completely when and how to use technology to combat them. In short,

Proposition 10: Ineffective information retrieval can stem from unused tools, misused tools, and lax procedures in addition to the lack of proper tools.

Research that apportions blame among these factors can help direct efforts at improving the problem. Such evidence will most likely come from cases studies. For a simple example, one can study the effect of a personal bibliographic retrieval system on the amount or quality of work an employee produces. Or, the effectiveness of information retrieval can be compared for managers who a) develop procedures to stay personally involved with the storage of the documents they need for their work (as Strassmann (1985) recommends) versus b) those who rely exclusively on delegating this task to others.

Finally, the advent of optical image processing systems should provide an intellectually rich environment in which to examine the interplay of system development and organizational change. Technologically, such systems are evolving and still under development. Competing technological pressures come from having to store millions of pages of information, retrieve and produce page-per-second, print-quality screen images, and transmit all this information over computer networks. Organizationally, the effects of such systems on how work gets done are too complex and too subtle to be understood before such technology is installed and operational (Runyan, 1990). Still, approximately one half of 400 surveyed companies are at least thinking about an imaging pilot project (May, 1990).

Traditional methods of requirements analysis fall short of measuring how fundamentally these systems can transform an organization as well as their effects on intangibles such as customer service. Thus, new methods are being explored in an attempt to learn concurrently about an evolving technology and the new forms of business it will support -- even as system deployment is fully under way (Runyan, 1990). Organizations that are undertaking efforts to fundamentally change the way they do their work (such as the United States Patent and Trade Office with its \$800 million effort to establish an Automated Patent System (Runyan, 1990)) can become laboratories for IS research. From them, we have much to learn about managing the IS risk associated with technical and organizational uncertainty, and developing information systems that truly transform organizations.

6. Conclusion

Modern business depends on information and supporting technology to function, let alone compete effectively. Larger organizations, more complicated organizational structures, as well as an increasingly complex environment result in an ever increasing volume of information with which an organization must deal. Around the globe, information work comprises the largest single employment category in industrial societies, consuming, for example, two-thirds of labor expenditures in the United States (Strassmann, 1985). Working with information is the central activity for many people in various organizations, including managers, consultants, lawyers, technologists, and employees involved with research and development, strategic planning, etc.

Properly used, information retrieval systems can provide information workers with more relevant information than they have dreamed possible (or considered economically feasible) and information in the form best suited to their needs: text, graphic, image, etc. At the same time, new patterns of work can develop from information systems that better support information sharing and reduce the information overload with which we too often deal today.

In the next decade, management of unstructured information will come to the fore, enabling retrieval, dissemination, analysis, and synthesis of disparate information (Marchand, 1985). The management of unstructured information will better support professional, managerial, and operational activities by focusing on the content of information itself and how it is used and valued in the organizational setting.

Yet now, organizations pay too little attention to the management of unstructured information by devoting space to house information they don't use, time to locate information they can't find, and brain power in needlessly revisiting already solved problems. Though we are overwhelmed by information, we can't seem to access what we need.

The challenge for the academic IS community is to recognize the business information retrieval problem, to better understand the related work performed by other academic disciplines, and to conduct research aimed at integrating information retrieval technology and business information retrieval practice. We need theory that provides a fuller understanding of why some firms perform better than others in their information retrieval, a more complete categorization of business information retrieval problems, and clearer guidelines for using technology

to help solve these problems. In these ways, we may make information retrieval serve organizations to the fullest extent possible.

Acknowledgments:

I thank Randy Cooper of the Decision and Information Science Department at the University of Houston for his extensive comments. I also thank Dave Blair, Dennis Severance, and Michael Lougee, all from the University of Michigan, and Fred Gordon of the *Plain Talk Investor*, Northbrook, Ill.

REFERENCES

Ackoff, R.L. "Management Misinformation Systems." *Management Science*. December, 1967, pp. B147-B156.

Benbasat, Izak, Goldstein, David K., and Mead, Melissa. "The Case Research Strategy in Studies of Information Systems." *MIS Quarterly*, September, 1987, 369-386.

Black, Nancy Hollen. "Escape from the Paper Pit." *Datamation*, April 15 1990, pp. 86-88.

Blair, David C. "The Management of Information: Basic Distinctions." *Sloan Management Review*. Fall 1984. pp. 13-23

Blair, David C., and Maron, M.E. "An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System." *Communications of the ACM*, Vol. 28, No. 5, March 1985. pp. 289-299.

Blair, David C., *Language and Representation in Information Retrieval*, Elsevier Science Publishers. Amsterdam, 1990.

Blair, David C. and Gordon, Michael. "The Management and Control of Written Information: Growing Concern Amid the Failure of Traditional Methods." *Information & Management*. Forthcoming.

Borgman, Christine L. "The User's Mental Model of an Information Retrieval System." *Proceedings of the Eighth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 268-273, Montreal, June 1985.

Burns, Christopher. "Three Miles Island: The Information Meltdown." *Information Management Review*, Vol. 1, No. 1, 1985. pp. 19-25.

Davis, Fred D., Bagozzi, Richard P., and Warshaw, Paul R. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models." *Management Science*, 35(8), 1989, pp 982-1002.

- Denning, Peter J. "Electronic Junk," *Communications of the ACM*, March, 1982.
- Eisenhardt, Kathleen M. "Building Theories from Case Study Research." *Academy of Management Review*, 14(4), 1989, pp 532-550.
- FileNet Corporation, Costa Mesa, California.
- Gilad, Tamar, and Gilad, Benjamin. "Business Intelligence--The Quiet Revolution." *Sloan Management Review*. Summer, 1986. pp. 53-61.
- Gordon, Michael D. "The Necessity for Adaptation in Modified Boolean Document Retrieval Systems," *Information Processing and Management*, 24(3), 1988a, 339-347.
- Gordon, Michael D. "Probabilistic and Genetic Algorithms for Document Retrieval." *Communications of the ACM*, Vol. 31, No. 10, 1988b, pp. 2109-1218.
- Gordon, Michael D. "Evaluating the Effectiveness of Information Retrieval Systems Using Simulated Queries." *Journal of the American Society for Information Science*. To appear. 1990.
- Gunn, T. *Computer Applications in Manufacturing*. Industrial Press, Inc. 1981.
- Haavind, R. "Tools for Compatibility." *High Technology*, August 1986. pp. 34-42.
- Humphrey, Susanne M. "MedIndEx System: Medical Indexing Expert System," *Information Processing and Management*, Vol. 25, No. 1, 1989, pp. 73-86.
- Lancaster, F.W. *Toward Paperless Information Systems*. Academic Press. New York. 1978.
- Lancaster, F.W. *Information Retrieval System -- Characteristics, Testing, and Evaluation*. 2nd Edition. New York: Wiley, 1979.
- Lancaster, F.W., *Vocabulary Control for Information retrieval*, Second Edition, Information Resources Press, Arlington, VA. 1986.

Malone, Thomas W.; Grant, Kenneth R.; Turbak, Franklyn A.; Brobst, Stephen A.; and Cohen, Michael D. "Intelligent Information Sharing Systems." *Communications of the ACM*, Vol. 30, No. 5. May, 1987. pp. 390-402.28.

Marchand, Donald A. "Information Management: Strategies and Tools in Transition." *Information Management Review*. Vol. 1, No. 1, 1985. pp. 27-34.

May, Thornton. "Justifying the Image." *Datamation*, April 15 1990, pp. 82-84.

McNurlin, B. C., Ed., *I/S Analyzer*, Vol. 27, No. 5, May 1989.

Meltzer, D.P., Haas, S.W. "The Constituent Object Parser: Syntactic Structure Matching in Information Retrieval." *Proceedings of the Twelfth Annual International ACMSIGIR Conference on Research and Development in Information Retrieval*, Belkin, N.J. and van Rijsbergen, C.J. Eds., pp. 117-126, Cambridge Mass, June 1989.

O'Reilly, Charles A. "Variations in Decision Makers' Use of Information Sources: The Impact of Quality and Accessibility on Information." *Academy of Management Journal*, Vol. 25, No. 4, 1982. pp. 756-771.

Repo, Aatto J. "Economics of Information," *Annual Review of Information Science and Technology*, M.E. Williams, ed., Vol 22, 1987, pp. 3-36.

Runyan, Linda. "PTO's Inventive Ways with Imaging." *Datamation*, April 15 1990, pp. 92-95.

Salton, Gerard, and McGill, Michael J. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Co. New York. 1983.

Salton, Gerard, and Smith, Maria. "On the Application of Syntactic Methodologies in Automatic Text Analysis." *Proceedings of the Twelfth Annual International ACMSIGIR Conference on Research and Development in Information Retrieval*, Belkin, N.J. and van Rijsbergen, C.J. Eds., pp. 137-150, Cambridge Mass, June 1989.

Sparck Jones, Karen, Ed. *Information Retrieval Experiment*. Butterworth and Company Publishers, 1981.

Strassmann, Paul A. *Information Payoff: The Transformation of Work in the Electronic Age*. The Free Press. New York, 1985.

Swanson, Don R. "Searching Natural Language Text by Computer." *Science*, 132(3434), 1960, 1099-1104.

Swanson, Don R. "Two Medical Literatures that are Logically but not Bibliographically Connected." *Journal of the American Society for Information Science*. 38(4), 228 -233. 1987.

Swanson, Don R. "Unnoticed Connections in the Literature of Medicine: Implications for Knowledge Representation and Natural Language Searching." Unpublished manuscript distributed at the American Society for Information Science Mid-Year Meeting, Ann Arbor, MI, 1988.

Swanson, Don R. "A Second Example of Mutually Isolated Medical Literatures Related by Implicit, Unnoticed Connections." *Journal of the American Society for Information Science*, 40(6), 1989.

Swartz, Herb. "For the Record." *Skylines*. Feb. 1986, pp. 10-11.

Tong, R.M., Appelbaum, L.S., Askman, V.N., Cunningham, J.F. "Conceptual Information Retrieval using RUBRIC" in *Proceedings of the Tenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New Orleans, 1987, 247-253.

van Rijsbergen, C.J. *Information Retrieval*, Second Edition. Butterworths. London. 1979.

Warren, Kenneth S. *Selectivity in Information Systems*. Praeger Publishers. New York. 1985.

Yourdon, Edward. "Paper Chase: Keeping up with Office Productivity." *Computer World*. July 21, 1986. pp. 53-58.

Academy of Management Journal
Administrative Science Quarterly
California Review of Management
Communications of the ACM
Decision Sciences
Harvard Business Review
I/S Analyzer (formerly EDP Analyzer)
IBM Systems Journal
Information and Management
Management Science
MIS Quarterly
Sloan Management Review

Table 1: Journals Searched for Articles on the topic of Business Information Retrieval

Annual Revenues of Participating Firms

Revenues	Number of Firms
\$900 Million - 1 Billion	1
\$1 - 2 Billion	2
\$2 - 3 Billion	1
\$9 - 10 Billion	1
\$11 - 12 Billion	1
\$22 - 23 Billion	1
Data Not Available	7

Table 2. Annual revenues for seven of fourteen participating companies.

Job Classifications for Subjects

Category	Number
Administrative	3
Attorney	5
Clerical	5
Executive	4
Manager	14
Technical	8

Table 3. Job classifications for thirty-nine subjects. Technical jobs include hardware and software specialists, corporate librarians, research scientists, archivists, and records retention personnel.