On Challenges for Information Management Technology

Won Kim, School of Information and Communication Engineering, Sungkyunkwan University, Suwon, S. Korea

Abstract

Today information management technology faces two major related challenges. One is to tame the information and options explosion that are upon us. Another is to support the information needs in the ubiquitous environments that are being created. These two challenges have received considerable attention from various segments of information management technology research community. Some of the subjects of research have been addressed sufficiently, while other subjects still require considerable research. In this paper, I review and analyze the challenges, and offer some directions for some of the subjects of research, so as to help marshal the creative energies of the corresponding segments of the research community for faster solutions to the challenges.

1   INTRODUCTION

Today people suffer through two major maladies brought about by the advances in information technology and wide adoption of it during the past few decades, from ever-powerful personal computers to the Internet. These are information explosion, and options overload in computer systems and electronic devices people use.

There are many elements of information technology that have contributed to today's information explosion. These include semi-conductor, display, digital storage, personal computers, networking, communications, multimedia processing, and digital convergence, among others. These have enabled the storage, processing, and transmission of all types of information using computers. The Internet has already become an indispensable source of information of all types for the masses. As such, corporations, governments, and various for-profit and non-profit organizations post information and advertise on the Internet. The Internet has also released people's apparent pent-up desire for self-expression and comradeship with people they do not even know, in the form of user-created contents, blogs, posting of documents of all types, opinions about all types of things, and responses to questions from others. This has made the World-Wide Web an even richer source of information, and yet at the same time has greatly aggravated the

information explosion problem. The abundance of types of information available has also made it ever more difficult for people to wade through them to find precisely what they need.

Advances in information technology have also made relatively inexpensive mobile devices and consumer electronic devices available for the masses. Digital convergence has made it possible for data from different types of devices, such as digital cameras and mobile phones, to be treated equally as blobs of digitized data. This, along with the competitive pressures in the market, has resulted in ever growing lists of options (or features) for hand-held mobile devices and consumer electronic devices. However, the general public has found the options provided either unnecessary or difficult to utilize because of the limited size of the display and the limited means of input and output on many of these devices. To make matters worse, the user interfaces of the computer systems and electronic devices have not been designed with adequate consideration of the usability for the general public. As a consequence, most PC users, despite having had to shoulder the burden of administering the PCs, have also to call for help. The general public has to carefully read the user manuals to be able to operate electronic devices such as washers and dryers, microwave ovens, VCR/DVD recorders, digital cameras, digital televisions, rice cookers, the clocks built into various home appliances, etc.

The widespread use of mobiles devices, besides bringing about the options overload problem, has also resulted in the creation of ubiquitous computing environments for various applications. Increasingly, people's daily business of living, working, entertaining themselves, learning, interacting with their worlds, and gathering information is being conducted on the go with the mobile computers and electronic devices. Information of various types needs to be stored, retrieved, and transmitted among computers and devices, both mobile and stationary, via various types of networks.

Information management technology is, as the name says, technology for managing information, having evolved from file management, information retrieval and database management technologies to encompass customer relationship management, supply chain management, enterprise resource planning, data and application integration, multimedia processing, data mining, and Web personalization and recommendation, among others. In my view, two of the most worthy and pressing areas of research and development in information management today are the taming of information explosion and options overload, and support for ubiquitous computing environments. Certainly, these two areas of research have received lots of attention from various segments of the information management technology research community, and research into many of the subjects has resulted in widely used commercial products. However, in my view, various subjects within these areas require significant additional research and development. I also feel that the pace of advances in these subjects has lagged behind the pace of the problems mounting. The objective of this paper is to highlight the importance of addressing the problems, and marshalling the creative energies of the researchers working in these areas. I will review the various aspects of the challenges facing information management technology, summarize current approaches to addressing them, and offer directions for some of them.
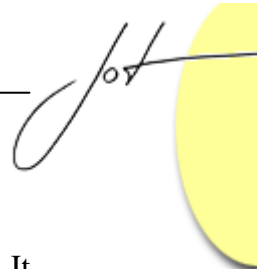
2 THE INFORMATION EXPLOSION PROBLEM

Information explosion does not mean simply that there is just too much information. It manifests itself in two different ways: accessibility problem and relevance problem. Often, despite the fact that lots of information are stored in computers, it is very difficult or even impossible for computers to access them. Further, even when it is possible for computers to access all the information, a lot of it is not really what the people accessing them need. Table 1 summarizes the dimensions of the information explosion problem and current solutions.

First, I will examine the accessibility problem. Broadly, there are two sub-dimensions to the problem. One is distributed information; that is, information is stored in different computers and managed by different information management systems (i.e., file systems or database management systems) or by different applications. This in turn has two cases, depending on whether the existence, location and access requirements of some elements of the distributed information are known or not known. An example of the latter case is all the Websites that have not been indexed by the search engines, and therefore are not accessible to most people. For the former case, there are various well-established solutions on the market. They include data warehousing, data integration (or federation), application integration, integrated content management, supply-chain management, customer relationship management, enterprise resource planning, etc. These solutions, with the exception of data integration, address the application development issue and performance issue by assembling into a central repository all necessary information from disparate information sources. Data federation addresses mostly the application development issue by logically assembling necessary information, while all information remains with disparate information sources.

Another dimension of the accessibility problem is the semi-structured data and multimedia data. While such data as records in a relational database tables, and records in files are regarded as structured data, such data as emails, XML documents, and all types of forms used by corporations and government branches, etc. are regarded as semi-structured data. Semi-structured data has an underlying structure; however, some components of the data are free-form text or multimedia data whose semantics are not known to the systems that manage the data. Multimedia data includes photographs, satellite images, video clips, audio clips, television broadcasts, etc. If semi-structured data or multimedia data are not manually tagged and classified (usually manually and sometimes semi-automatically), it becomes difficult or impossible for computers to search them or match them with given sample data. Much research has been done, and there are commercial products for automatically matching images, such as fingerprints and faces; matching audio, including voice, music, and sound; recognizing anomalies in images and audio; creating indexes for fast search and matching of images, audio, and video; etc. Some Internet search engines provide facilities for image search. Research has also been done, and even commercial products are available, for enabling automatic classification of semi-structured data, automatic keyword extraction and even summarization of free-form text.

problem	issues	sub-issues	current solutions
accessibility problem	distributed information	locations known	data warehousing, data integration, application integration
		locations unknown	search engine indexing
	semi-structured data and multimedia data		automatic keyword extraction and summarization of text, automatic classification of data, automatic tagging of multimedia data, indexing of multimedia data
relevance problem	search		Internet search engines, personalization, context awareness
	profiling		e-commerce recommendation engines, personalization, context awareness

Table 1: Summary of the Dimensions of the Information Explosion Problem and Current Solutions

Now I will examine the relevance problem. This refers to the retrieval and presentation of too much information that is not relevant to the needs or intentions of the people seeking such information, even when the information access problem has been addressed. Again, there are two, related, sub-dimensions to the problem. One is the search issue arising from the inability of the search mechanism to find and rank relevant information from stored information that it can access. An example of this is the Internet search. If a person uses a search keyword that is too simple, he may be inundated with Web pages that have nothing to do with what he wanted in the first place. If he uses too-specific a keyword combination, he may not get any result. Much research has been done to make search keyword specification match people's intentions, and also to refine the results of Internet search. Internet search engines have become considerably better at finding more relevant information and prioritizing the search results. Various techniques, such as the page rank algorithm, have been incorporated into Internet search engines to increase accuracy of the search results. Personalization techniques, such as collaborative filtering and content filtering, have also been introduced to increase accuracy of search. Some aspects of context awareness techniques, in particular, the location and time, have been proposed to increase relevance of information delivered to people.

The second sub-dimension the information relevance problem is the profiling issue. This issue arises, often with information that computer systems automatically capture or generate, such as information access history, product purchase history, profiles of people segmented on the basis of various attributes (e.g., gender, income level, education level, ethnic origin, religion, etc.), and so on. An example is a person visiting an e-commerce Website and receiving recommendations for products or services – either when he explicitly requests such recommendations or by default. In general, there is a lot more
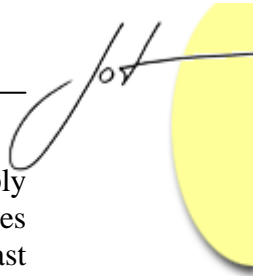
products and services than the person is potentially interested in, and receiving sharply limited relevant information can be useful to the person. The recommendation engines that e-commerce Websites employ have become more sophisticated during the past several years in matching recommendations to the purchase histories and/or profiles of the Website visitors. Here, too, some personalization techniques and some aspects of context awareness techniques can increase the accuracy of recommendation.

Despite the significant advances that have been made thus far, in my view, much additional research is needed to address the information explosion problem. The subjects of research include the entire relevance problem; automatic classification and indexing of semi-structured data and multimedia data; automatic content identification of multimedia data; and the integration of Web data.

3 THE OPTIONS OVERLOAD PROBLEM

The options overload problem today is straining the usability of computer systems and electronic devices. There are three dimensions to the problem, including too many functions, too many categories of information, and too many combinations of modality of presentation to the users. These are summarized in Table 2.

First problem is the number of functions of a system or a device exceeding the number of keys (or buttons, switches, etc.) on the system or device. If a computer system or an electronic device only provides a small number of functions, say less than 10, it may often be possible to manifest these functions on the input mechanism, and allow people to invoke any desired function with the press of a single key (or button, switch). Unfortunately (or fortunately?), many of the computer systems and electronic devices today come with a rich set of functions -- too many to be mapped one-to-one to the keys on their input mechanisms. For example, all of the functions of the digital television or television recorder cannot be mapped to the remote control and the instrument panel on the television set, and require a menu that displays operating instructions on the television screen. Various electronic devices, ranging from the microwave oven and the rice cooker to the automobile navigator and digital camera, map their functions to the input and output mechanisms of the devices themselves. Because the number of functions usually exceeds the number of keys on the system or device, the general public often does not know to what functions some of the keys on the system or device are mapped, unless they consult the user manuals.
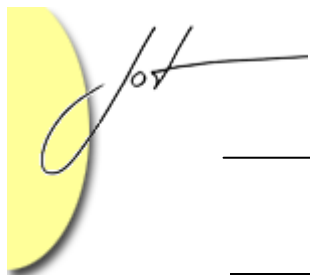
sources	problem
functions	exceeds the number of keys on the input mechanisms
categories of information	exceeds the size of comfortably presentable list
combination of modality of presentation	complex, inconsistent, non-intuitive combination

Table 2: Summary of the Sources of the Options Explosion Problem.

The second problem is the number of categories of information exceeding the number of elements in a short list that can be presented to and comfortably comprehended by people. Many Internet portals have a bewildering mix of a search keyword box, a bunch of service options, news snippets, announcements, a list of popular search keywords, etc. Many Websites require the visitors to navigate a deep sequence of pages to get to the information they seek or need. Information-rich electronic devices, such as the digital television and television recorder, provide menus that display on the screen. Often the information architecture of the menu is inadequately designed, and as a result, the general public does not know how to operate the devices, and cannot even find the operating instructions for some of the functions.

The third problem is the ineffective combinations of possible modalities of information presentation to people. There are lots of modalities available to the user interface designer, including font size, font style, color, symbols, icons, image, short descriptive text, voice, sound, light, 3-d graphics, even duration of pressing (on a button), the swipe of a magnetized card, RFID, biometrics, motion, tilting, etc. The user interface designer must select a good combination of presentation modalities suitable for the system or device of interest, without getting befuddled by the large number of modalities at his disposal.

The options overload problem must be addressed by great user interfaces. Some of the popular systems and electronic devices today, including Apple's iPOD and TiVos' television program recorder, have received high marks for the design and usability of their user interfaces. However, the usability aspects of the user interfaces are the Achilles heel of a majority of today's computer systems and electronic devices. The design of the user interfaces of these systems and devices should be significantly redone with the view to making them so simple and intuitive as to ultimately obviate the need for the user manuals. As a part of the approach to achieving this goal, the user interfaces may also be customized by taking into account the usage history, profiles, and preferences of the users of the systems and devices. It is interesting that personalization or customization represents a common goal in the solution to the information explosion problem and the options overload problem.
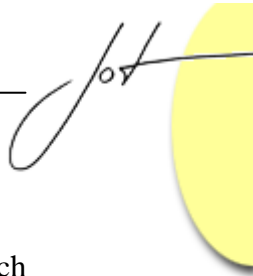
4 SUPPORT FOR UBIQUITOUS COMPUTING ENVIRONMENTS

A ubiquitous computing environment is a distributed computing environment. Much research has been done during the past three decades on data management in distributed systems, addressing such issues as query optimization and query processing, catalog management, transaction management, reliability, maintaining consistency of replicated data, etc. Much of the results can be adapted fairly easily to ubiquitous environments. However, several new factors have to be taken into account.

1. Mobile handheld devices and stationary electronic devices now become "computers" in a distributed system of computers that have been presumed in earlier research. Some of these "computers" have limited processing power, limited memory, and no hard disk drive.
2. The Internet is now injected as a data source for many of these new "computers." This and the above considerations make the traditional synchronization techniques, such as the two-phase commit protocol, largely inoperative, and make security guarantee more difficult.
3. The types of information are no longer limited to alphanumeric data, and now a good portion of the information to be managed includes semi-structured data and multimedia data.
4. Some of these "computers" have limited input and output mechanisms, exacerbating the already serious user interface (options overload) problem.

The above considerations point to security, privacy, synchronization, query processing, and user interfaces for the tiny input and output mechanisms as worthy research subjects. Many of the subjects of research for addressing the challenges of information explosion and options overload are directly applicable to supporting the information management requirements of ubiquitous environments.

5 CONCLUDING REMARKS

In this paper, I reviewed and analyzed the two major challenges facing the information management technology today; namely, the information and options overload problems, and support for ubiquitous environments. As the pace of advances in research to meet these challenges is lagging behind the pace of the problems becoming ever more intractable and the spread of ubiquitous environments, I am hopeful that researchers working on various relevant subjects will find practical solutions faster.

ACKNOWLEDGMENTS

About the author

Won Kim, Professor and Univeristy Fellow with the School of Information and Communication Engineering at Sungkyunkwan University, Suwon, S. Korea. He is Editor-in-Chief of ACM Transactions on Internet Technology (www.acm.org/toit). He is Global General Chair of the Human.Society@Internet International Conference. He is the recipient of the ACM 2001 Distinguished Services Award, and is an ACM Fellow. He can be reached at wonkim@skku.edu