SIG/CON: The Next 25 Years

Joseph Janes
School of Information and Library Studies, University of Michigan, Ann Arbor

Introduction

It is an honor to have been asked by our esteemed founder, Llewellyn C. Puppybreath, III, to contribute this exploration of the future of SIG/CON. As the only SIG within ASIS to claim the entire body of the Society as members, this duty bears special responsibility. Dr. Puppybreath, sadly, was unable to write this piece himself; he was suddenly and unexpectedly called off for jury duty somewhere out west, and when last I heard from him, he told me it would be quite some time before he was able to once again take up the reins of scholarly research in the information sciences. I feel sure we will all be enjoying his company at future gatherings of the SIG at ASIS meetings to come.

It is a double honor to have this piece of speculation appear in the *Journal*. It has now been 10 years since a piece worthy of SIG/CON has appeared in *JASIS*. Not since the now legendary paper on the Dillion (or was it Dillon?) Hypothesis of Titular Colonicity (Perry, 1985) has another work of sufficient stature been included within these covers. Nonetheless, I feel sure that more are to come.

In this paper, I will endeavor to describe what I think are the most important areas of research which should and will be part of the SIG/CON agenda for the next twenty-five years. Of course, I am susceptible to the same pitfalls and foibles of all prophets and prognosticators —things will undoubtedly not work out as I describe, and it is indeed quite possible that none of what I discuss here will come to pass. To generations future who may read this and scoff at my Delphic abilities, I have but one thing to say: hard cheese.

Having passed on that little tidbit of wisdom, I will proceed to highlight some possibilities for new work in some familiar areas of information science, followed by thoughts about newly emerging areas of research.

Traditional Information Science

I have always maintained that the most important challenges facing information science surround the understanding and processing of *natural language*. The problems of ambiguity, synonymy, conflation, and so on not only keep thousands of librarians and information scientists employed, but also make research in almost every area of our field damned difficult. The natural language processing community has tried valiantly over the last few decades to develop automated approaches to working with natural language, but frankly, they've come up pretty dry. It is an enormously difficult problem.

I therefore suggest they take another tack. Rather than trying to build large lexicons and thesauri to *describe* and understand natural language, they should develop small-scale, easy-to-understand languages which then would be *prescribed* for use in the writing of journal articles and the statement of queries. Perhaps a model to follow would be that of Newspeak, as described by Orwell (1949). “Doubleplusgood” is certainly a model of clarity and lacks ambiguity, and it is undoubtedly true that restricting the range of vocabulary couldn't do much damage to many journal articles.

The study of the retrieval of information is one of the core areas of information science research, and will continue to be central in the future. *Information retrieval* research, however, should take on a new form. Fans of the Terry Gilliam film *Brazil* (Milchan, Cassavetti & Gilliam, 1985) will recognize this opportunity for expansion immediately. In this inspired vision of a future dominated by concerns about the collection and dissemination of information, Gilliam addressed the future of IR: Information Retrieval is the name given to the secret police interrogation force. Clearly, this is a significant opportunity for the IR community to share its expertise with a new and wider audience, and provides all new perspectives on such well-known and thorny

problems as relevance feedback, query updating, inverted files, as well as, of course, full text retrieval.

The *mathematical* branch of information science has made many significant contributions, but there have been some complaints over the years regarding the complexity of some of the formulas and developments in this work, particularly in articles appearing in *JASIS*. This criticism is of course unfair, and in fact, our colleagues of a mathematical bent should continue to explore ever more intricate and detailed equations. While other fields are content with such trivialities as

$$x^n + y^n = z^n$$

and even things as simpleminded as

$$E = mc^2$$

our intrepid colleagues should continue to explore formulas of ever more mind-altering dimensions. The following, taken from a recent article, is the merest suggestion of the uncharted territory ahead:

$$(WX)(WX)^T(WX)X^T = (RM^TkCP^x)(P^xC^TM^Tk\,R^T)(RM^Tk\,CP^x)X^T$$

Who said mathematicians should have all the fun?

In the field of *user studies*, one of the most significant problems has always been getting the cooperation of the right users. Obviously, if an insufficient or unrepresentative sample is chosen, or if the subjects are uncooperative or downright stupid, the results of any beautifully crafted study will suffer. Therefore, we will need to find a way around this problem. I suggest SIG/CON members explore this idea: find a group of highly educated, willing, bored people (there are many of these in Ann Arbor sitting around coffee shops, wearing black, drinking espresso and misquoting Kafka) and pay them to be professional users, thus eliminating the problem.

An interesting side effect of this maneuver could be that using such intelligent, willing and pliable subjects will improve the results of all studies, eliminating those tedious problems of users who are bored, do not follow directions, do not really know what their information needs are, or are just too unintelligent to care.

Of course, the field of *relevance research* is near and dear to me, and it was my honor to present my first paper at a SIG/CON meeting on this very topic. My work on relevance judgments of identical twins separated at birth has become justly famous (Janes, 1993). The whole field of relevance work, however, has moved in the last few years much more towards the user-centered notion: that to fully understand the processes by which people make decisions about information which is presented to them, we must understand the complex nature of the user, their need, and the interaction with the information items. However, we are still plagued by a significant difficulty, that of individual differences. Recent research has begun to identify patterns of judging behavior across different kinds of users, but there is still a great deal of idiosyncratic behavior at work here. Frankly, to track down these kinds of picky peccadilloes of relevance judgment will be tedious in the extreme and take forever. Who needs it?

Thus, I suggest the following elegant solution: build a new and different information retrieval system for every person on Earth. This will solve the individual differences problems—crafting such systems will allow us to tailor them to the whims and needs of each person, and has the wonderful side effect of providing a tremendous employment opportunity for information scientists to build several billion systems. I will ask only a small royalty from each design for coming up with the idea.

Research into the performance of *online searchers* is also of particular interest to me; my second SIG/CON paper comparing the online searching performance of a class of library students and an infinite number of monkeys was very warmly received in 1994 (Janes, 1994). With the increasing prevalence of CD-ROMs and other electronic methods of searching for information, there has been some concern about the future of online searching as a professional activity and, thus, as an area for research. While I think a significant decline in online searching is unlikely, it might be wise for researchers to turn their attention as well to other, related problems. A prime example is the search for anything decent to watch in a 60- (or, heaven spare us, a 500-) channel cable world. Clearly, there are patterns of searching here to rival any search tactics or berry picking previously identified. Are there operationalist and conceptualist channel surfers? Are linear or random techniques more useful? Gender differences alone could occupy years of fruitful research and be a positive boon to humanity. This is a fine opportunity for the application of well-established information science methodologies in the lucrative field of entertainment marketing.

An area which should undergo dramatic expansion is *bibliometrics*. Understanding the nature of scientific

communication in a distributed networked environment is critical, and many of the tools developed in the print world could be brought to bear on the 'Net. The problem, though, is the name. Is there anything worse sounding than "bibliometrics?" We have an image problem here, and one I think can be solved with the introduction of a new singing group. What R.E.M. did for sleep research, U2 did for the surveillance community, and Alabama did for, well, Alabama, Zipf and the Bradfords can do for bibliometrics. Their first album, *Impact Factor*, alone, could change the face of the field forever (auditions could be at a future SIG/CON session).

We should not neglect our associates in the field of *classification* research. If documents were not appropriately described, organized, and classified, the rest of us would have precious little to do. Here is a golden opportunity for some interdisciplinary research. With the new theme in information retrieval research mentioned earlier, it would be a natural collaboration for classification researchers to develop schemes by which to decide which documents should be classified and how—which ones should be marked Secret, Top Secret, Eyes Only, and the highest level (used only for the most important and highly relevant documents), Burn Before Reading.

The Shock of the New: Hot Topics in the 90s

In the last few years, several new and exciting areas of research and development have opened up in the information science arena. It is certain that even more work will be done in these areas; here are some provocative ideas:

Intriguing current work in *visualization* in information retrieval hopes to allow users to see a graphic display of the results of their searches, thus allowing them to see the information in totally new ways. Proponents of these ideas take on almost Tantric tones: See the information . . . See the information . . . Be the information . . . Be the information . . . It is a short step from this to a much more interesting multimedia approach, and to a system which takes a much more active role in guiding users through databases and information resources, and which will have an exciting new interface. The documentation for the Tommy system (a descendent, no doubt, of Thomas) will also be much simpler than Dialog, for example; it appears in its entirety here (Townshend 1969):

"See me. Feel me. Touch me. Heal me."

Of course, the biggest research area to arise in the past couple of years is *digital libraries*. I am somewhat skeptical about the future of these, to be honest. Perhaps I do not understand it that well, and I certainly do not wish to do a disservice to my friends and colleagues working on this problem, but I actually do not think there will be a large societal call for research dealing with large collections of fingers. Perhaps other body parts would be more attractive: pedal libraries (collections of feet), for example, or optical libraries (small collections of eyes have already been started, of course, by forward thinking hospital libraries). The idea of classifying fingers by size, shape, length and color of nail, presence or absence of knuckle hair, and so on leaves me rather cold. I know this must be an important area, since the National Science Foundation and others are pumping millions of dollars into it, but the appeal and the utility escape me. Perhaps further research will clarify matters somewhat.

On the other hand, research into *geographic information systems* has direct social relevance. Yes, the current research in storing and retrieving geographically-specific information is all very valuable and wonderful, but the real future lies in finding out how different people seek geographic information, especially why men will not ask directions or admit that they are lost. Allied development work, building information systems to allow men to find their way without seeming to ask, would also be of great value.

There has also been much work, at least in the computer science world, about *virtual reality*. Although somewhat unsettling to those of us who have not gotten *this* one down yet, let alone dealing with other realities, it nonetheless opens up intriguing opportunities for the information science community. Virtual worlds will, by necessity, have virtual books. These virtual books will be stored and accessed via virtual libraries. These virtual libraries will be visited by virtual patrons and be staffed by virtual librarians. The processes which go on here will, consequently, be studied by virtual information scientists. Voilà! Since these virtual folks will be doing all the work, the rest of us can all go home and relax and finish that volume of Proust we've all been promising ourselves. (Of course, things will not change dramatically for educators in library and information science, who have been engaging in virtual work for some time.) This research should be given top priority right away. Some of that NSF money should slide over here, if we know what is good for us.

The world of *distributed networked information systems* has been moving much too fast for much serious research to be undertaken around it. New protocols, new data structures, new services, even whole new ways of storing and thinking about information arise almost daily. What is a serious researcher to do, other than surf aimlessly and read Dilbert every day? One phenomenon, though, does bear further scrutiny. It would appear that the early days of giving cutesy names to Internet-based information tools (à la gopher, archie, veronica) has given way to a more sinister collection of spiders, worms and other nasties. Does this reflect some deeper ,dark side to the World Wide Web? Or worse yet, a conspiracy to scare off the timorous? This is a problem worthy of Oliver Stone, or at least Scooby Doo and the gang.

The design of *image* databases presents many subtle and difficult problems associated with the organization and retrieval of images—there are no obvious ways to create abstracts, any textual descriptors or index terms do not do justice to the complexity of the images, and searching based on shape, color, and so on will be very hard indeed. The obvious solution, therefore, is to admit an honorable defeat, give up and let people flail around blindly. This saves a lot of our time and trouble all the way around, and users can certainly not do any worse than they would if we were to develop intelligent retrieval for image files.

A few of our colleagues in the archival world are now becoming concerned about the problems surrounding *archives of electronic objects*—old data files, tapes, and even hardware. How will we store, preserve, document, arrange, and describe these items for future generations of researchers? How will we authenticate official electronic communications? How will we decide what parts of the record to keep for posterity and what to dispose of forever? Oh dear. These are challenging questions, and especially important because the fate of such an increasing fraction of the human record hangs in the balance. One hopes that the archival world will adopt its usual cautious behavior, and wait a few hundred years or so to see what happens before doing anything too rash.

Conclusions

Of course, we shall all be surprised by something totally new and different and unpredictable which will arise in the next few years and become an important focus for research. I am confident that the SIG/CON community will take on these problems with the same finesse and panache that has been the hallmark of our SIG over the last twenty-five years. Indeed, I feel certain that there will continue to be a sufficient supply of strange goings on in all areas of information science research to keep SIG/CON going for a very long time indeed.

References

Milchan, A. and Cassavetti, P. (Producers) & Gilliam, T. (Director). (1985). *Brazil*. [Film]. London: Brazil Productions.

Orwell, G. (1949). *Nineteen eighty-four*. New York: Harcourt Brace Jovanovich, Inc.

Perry, J. (1985). The Dillon hypothesis of titular colonicity: an empirical test from econological sciences. *Journal of the American Society for Information Science*, *36*, 251-258.

Townshend, P. (1969). *See me, feel me*. [Song]. From Tommy. [Opera]. London: Fabulous Music, Ltd.