## SPECIAL SECTION

### Bringing Genre into Focus

15] Introduction  
by Luanne Freund and Christoph Ringlstetter, Guest Editors of Special Section

17] Why Information Has Shape  
by Andrew Dillon

20] Stalking the Wild Web Genre (with apologies to Euell Gibbons)  
by Mark A. Rosso

23] Situating Relevance Through Task-Genre Relationships  
by Luanne Freund

27] Practical Aspects of Automatic Genre Classification  
by Christoph Ringlstetter and Andrea Stubbe

31] LIS and Genre Between People, Texts, Activity and Situation  
by Jack Andersen

### FEATURE

35] Research Interviews for Library and Information Professionals  
by Andrew P. Carlin
We begin this issue with ASIS&T news and activities. ASIS&T 2008 President Nancy Roderer brings us up-to-date with her President’s Page on a number of important projects that are engaging the attention of the Board of Directors at present, including the re-compete of the contract to publish our journal (JASIST) and the search for a new JASIST editor. The Board has also approved our trying to create a new global alliance among information societies. Nancy also mentions the work of the placement task force and of the task force considering whether ASIS&T should become involved in the accreditation of information science programs.

In another important development the Board has recently approved an open access policy for JASIST that allows authors to post preprints of their articles on their personal websites or the repositories of their institutions (but not subject repositories). Since we are a society with members heavily engaged in information research, including research about scholarly communication, ASIS&T and our Journal publisher, Wiley-Blackwell, have decided to study the effects of such a move. To implement this program, they jointly administered a baseline survey in January of this year, which Margeaux Johnson and Nancy Roderer discuss in their article “ASIS&T Scholarly Communication Survey” in Inside ASIS&T.

Apart from activities at the national level, it is always interesting to look at the ASIS&T local chapter events, where so many innovative ideas are nourished. Also in this issue of Inside ASIS&T, Aaron Louie of the Pacific Northwest Chapter (PNC) reports on ASIS&T’s first “unconference,” in this case a “BarCamp” called “InfoCamp,” jointly sponsored by the PNC and the University of Washington’s ASIS&T Student Chapter. Don’t know what a BarCamp is? Here’s your chance to find out, and, more importantly, to discover how your chapter also might host a low-cost, exciting unconference.

Our special section might be a good topic for a BarCamp – except that BarCamps don’t have topics. Those of us who have had much exposure to library cataloging are sensitized to genre (and form) and their use of subject headings and classification systems. Catalogers, at least, also know the complexity and difficulties of genre, while all librarians deal with it in the arrangement of their collections. Archivists and records managers are also keenly aware of genre or form in records, which is usually closely related to the type of transaction being documented by the record creator and thus of great importance. And all of us are very much aware of the new genres that are being created by the digital environment – the homepage, the FAQ, the “about,” the blog, the wiki and so forth. What should the role of genre be in information retrieval on the web? Is genre information potentially a cost-beneficial addition to web search engines as a primary or a limiting attribute? Is it well enough defined for such broad use? Can it be identified automatically? How can its purpose – its link to the activity that produced the document – be exploited to add another dimension to retrieval and to the study of user behavior? These questions and others are the subjects of this issue’s special section: Bringing Genre into Focus, edited by Luanne Freund and Christoph Ringlstetter. Six authors treat both sides of the debate and address some of the practical problems.

Finally, if you want to sell your managers on adding genre to your website search engine, Ted Sienknecht and Andrew Carlin have some excellent advice for you. In our Information Architecture (IA) column Ted lays out how to use information advantage to your advantage to justify any IA innovation, while in a feature article Andrew Carlin provides some very practical, nuts-and-bolts guidance on how to make any interview, whether for academic research or for requirements analysis, a successful and productive experience for all concerned.
Fellow Information Professionals:

The year is going by very quickly for me, with much good work being done by a large and dedicated group of ASIS&T volunteers. My thanks to all of them!

This has been a very busy year for people working on our publications. While the work of the JASIST RFP Task Force, the JASIST Editor Task Force and the Publications Committee is still proceeding, it is very appropriate to thank Trudi Hahn, Gary Marchionini and Sam Hastings respectively for their leadership of these activities.

Speaking of ASIS&T publications, I should refer you to an article in Inside ASIS&T on a recently completed scholarly communication survey of information science authors and ASIS&T members. This survey was done in December 2007, just before our new green road open access policy went into place for JASIST, and was done to establish a baseline so that we can track differences as authors and readers react to the new policy. Interestingly, almost all survey respondents had some knowledge of open access policies, while to date only about a third had published in an open access journal. Many thanks to University of Maryland student Margeaux Johnson for her invaluable help in analyzing the survey.

The survey results provide a reminder of the substantial number of information professionals outside the United States, including a growing number outside of the western world.

At the April meeting of the ASIS&T Board, the group had the welcome chance to review and approve plans for a Global Alliance of Information Sciences, Technologies and Service Societies proposed by the International Relations Committee. The alliance is planned as an informal partnership with other information organizations around the world, and our first letter of invitation will go out shortly. This activity was initiated by member Michel Menou in his time as IRC chair and is being continued under the committee leadership of Kendra Albright.

Yet another area in which members have been active this year is rethinking our placement services. Here a task force led by Abby Goodrum has been looking at ways that our current services in this area could be expanded to better serve both employers and applicants at the Annual Meeting and throughout the year. The task force is initially concentrating on academic applicants and jobs, but will move beyond that group to all information professional positions in time.

I hope you have been wondering how the Information Professional Task Force, my primary presidential initiative, is going. This group’s recent activities have included conducting a review of U.S. information programs and accreditation practices and talking with other information associations about potentially joining us in an effort to create a new accreditation process for master’s level information programs. With our partner the Council on Library and Information Resources, we will be holding a meeting of interested groups in early summer. Key individuals on these efforts include long time ASIS&T member Ann Prentice as well as two University of Washington students, Samantha Becker and Bo Kinney, who served as contractors to the task force to carry out the review.

As I recognize the work of our many volunteers, I should also point out that there are ample opportunities for more members to join in. Would you like a job, large or small?

Have you participated in ASIS&T before, but are without an assignment at the moment? Are you new to ASIS&T and wondering if participating might enhance what you get from the organization? (It would!) There are a number of ways to volunteer, including a number described in a link from our society home page at www.asis.org. In addition to the opportunities listed there, we have a number of active task forces (short-term groups working on a specific project) that you might join. Both Don Case, ASIS&T president elect (dcase<at>email.uky.edu), and I (nroderer<at>jhmi.edu) would be delighted to hear about your interests and suggest an assignment that would fit your needs.
News from ASIS&T Chapters

Ching-chih Chen, professor in the Simmons Graduate School of Library and Information Science and an international leader in digital library research and development, is the 2008 recipient of the American Library Association Beta Phi Mu Award. The honor is presented annually to a library school faculty member or individual for distinguished service in librarianship education. Chen is a longtime member of ASIS&T, where she has served in numerous local and national capacities in service to the field of information science.

The award jury responsible for Chen’s selection was “extremely impressed” with her “profound impact on the field of global librarianship,” as well as her devotion to her students and her record of publications and presentations.

The International Calendar of Information Science Conferences (http://icisc.neasist.org/) is a nonprofit collaboration between the Special Interest Group/International Information Issues (SIG/III) and the European (ASIST/EC) and New England (NEASIST) chapters of the American Society for Information Science and Technology, with the additional support of Haworth Press.

ASIS&T Meetings Continue to Set the Tone for Information Professionals

For nine years now, the ASIS&T Information Architecture Summit has proven itself as the premier gathering place for information architects. The ninth annual summit, held in April in Miami, attracted over 600 attendees presenting, discussing or just absorbing cutting-edge information on social networking, gaming, patterns, tagging, taxonomies and a wide range of IA tools and techniques. As one attendee put it, “Jared M. Spool’s opening keynote on Journey to the Center of Design and Andrew Hinton’s closing plenary on Linkosophy were amazing bookends for a great conference.”

Next up for the IA community is the fourth EuroIA, this year to be held in Amsterdam, September 26-27. EuroIA will explore the theme of Redrawing the Map, between countries and online as we look to forge new international alliances to adapt traditional deliverables to the needs of a Web 2.0 world.

And then get ready for the 10th IA Summit, March 18-22, 2009, in Memphis. But it’s not just the IA community benefiting from cutting-edge conferences sponsored by ASIS&T. The ASIS&T Annual Meeting continues to address the broader fields and boundaries of information science and practice. Coming up in late October in Columbus, People Transforming Information – Information Transforming People will use the lens of the human and social condition to seek a better understanding of the impact of the new information environment made possible by the digital world.

Check the ASIS&T website regularly for updated information on the ASIS&T conferences.
Designing a User-Centered Conference for User-Centered Information Professionals: The Story of InfoCamp Seattle

by Aaron Louie

In 2007 the ASIS&T Pacific Northwest Chapter and the ASIS&T University of Washington Student Chapter set out to revolutionize our regional annual meeting. Our vision was to create a highly collaborative, vibrant atmosphere where practitioners and students could share ideas and strengthen the local community of practice. The result was InfoCamp Seattle – an unconference that encouraged all to participate and invigorated both professional and student chapters.

What’s Wrong with Normal Conferences?

Serial conference attendees know the drill. We identify the one must-attend annual conference, beg our bosses to let us go, then drop a couple thousand dollars on airfare, hotel room and registration, hoping to be partially reimbursed someday. Once we get to the conference, we desperately hope that they’re offering something of professional value, because this event is our one chance a year to meet people in our field and gain crucial training. Pity those poor souls who work for local government, non-profit organizations, small companies or themselves – they may never get support from their employers to attend the big conferences.

In the library and information science community, there are limited – and often cost-prohibitive – venues for social interaction and professional development. Our field is constantly evolving, and those without a substantial travel budget or professional education program are left behind.

But it’s not just the lack of money or networking opportunities. Most conferences focus on presenting major findings and hosting panels of leading thinkers in the field. This conference format, while useful and important, usually lends itself to a one-way mode of information sharing. Most attendees are only allowed the five-minute question and answer period following a presentation in which to share their thoughts. Since topics and presenters are determined three to six months in advance (mostly to allow time to print the programs and reserve venue space), much of the content of a traditional conference runs the risk of being out-of-date by the time it is presented. In addition, big names and popular topics are often given priority, as the conference must appeal to as many people as possible. This configuration results in a conference that is of general practical value and provides discipline-defining vision but may fail to meet any single attendee’s particular needs.

Finding a Solution

In the Pacific Northwest Chapter we began asking our colleagues: How can we fill this gap? Can we do something at a smaller scale, on the cheap, that can provide specific, practical, up-to-date value to all participants?

As a first step, we started up a regular informal social event for the Pacific Northwest Chapter to discover what user-centered information professionals in Seattle and Portland wanted in a conference.

Starting in April 2007, we organized a monthly meet-up at the Elysian Pub, where we sat down to drink and converse with our fellow librarians, information architects, professionals and students. What we discovered is that there’s a constant need for informal information sharing and development of best practices within specific industries. Conversation at our meet-ups often included tips, tricks and comparing notes about common context-specific challenges. Participants traded business cards and contact information, forming new friendships and...
making important professional connections. We observed that the greatest value to attendees of these informal gatherings was in establishing local ties and brainstorming immediate, practical solutions to everyday problems.

The mix of people who came to our meet-ups was surprising. In any given month, we encountered professionals and students from a wide range of industries. Many of them had never attended a major conference or held a membership in a professional organization. The reasons they gave for this lack of participation? They were too busy working or had no travel budget. What would they rather have? A local, open, no-pressure venue where they could collaborate with other people who were facing the same kinds of challenges they were.

At this point, we decided to design a conference that was affordable and timely, encouraged social interaction and allowed all to participate. We wanted a format where any attendee could present work in progress with little preparation. This informality would encourage all to share nascent ideas, emerging technologies and practical suggestions without fear of rejection by a committee or advance scrutiny by peers. We would also allow groups to evolve organically without preparation or approval. All we would do is provide space, time and attention for group members to form and collaborate.

Fortunately, we didn’t need to look far for alternative conference models. In recent years, collaborative, open conferences have become increasingly popular. The common element is that the attendees create the content, usually day-by-day, at the conference. This species of conference is generally known as an “unconference,” of which there are several common variants. One of these is known as a “BarCamp,” created in response to FooCamp, an invite-only unconference for Friends of O’Reilly (the publisher of many books for computer and information professionals). Since then, hundreds of unconferences have been created in a myriad of industries and disciplines. In fact, Seattle was host to MindCamp, an unconference on technology issues, which Corprew Reed, the secretary and treasurer of our chapter, had helped organize.

Our counterparts at the University of Washington student chapter were also excited to hold an unconference. Rachel Elkington, then vice-chair of the student chapter, tells the following story:

At one of the social mixers [at the 2007 IA Summit], I talked to Nick Finck (from Seattle) and John Allsop (from Australia). They told me about the whole BarCamp phenomenon, which is basically ad-hoc conferences where everyone participates in doing everything so that there is no admission charge and all ideas are allowed to be expressed and exchanged.

Andy, the chair of the student chapter, added this:

From the student perspective, we want to provide our peers a chance to engage the professional community … This was really a chance for the students to plug into the professional community and especially to get to know some of the more dynamic and engaged personalities in that community. This was a chance to engage in discussion without having to be obviously identified as a student. I really feel this format levels the playing field in that regard, because it is so egalitarian.

We decided on a two-day conference over a weekend in the autumn before the ASIS&T Annual Meeting. Following the example of other BarCamps, we decided to call our unconference “InfoCamp,” in reference to the central concern of ASIS&T members and its allied practices: information. We modified the format slightly, adding an opening keynote for each day.

Following the BarCamp format, our schedule would not be decided beforehand. No speakers or topics would be pre-selected. We would create a theme, invite the right people and let the attendees decide what they wanted to talk about. By design, it would be participatory and user-centered, encouraging input, discussion and debate from everyone who attended. Another common practice in BarCamps, which we adopted, is to solicit sponsorship to cover costs in order to minimize the registration fees for attendees.

Because no sessions would be set before the conference, the participants could talk about the latest developments in their field and the most pressing current issues. And we could focus on providing the right atmosphere to make it all happen.
Making It Happen

Most of our challenges involved selecting and solidifying the date and venue. The sponsors, vendors, publicity and so on were all dependent on those two variables. However, selecting a date is not as easy as one might think. We needed to ensure that school would be in session in order to reach the students. We also needed to consider the other conferences and events happening within our professional community that might conflict. Given these constraints, the date was narrowed down to two weekends in October.

Finding an appropriate venue that was open on either of those two dates was a difficult undertaking. We finally found our venue, though – the Youngstown Cultural Arts Center. It was the perfect size and had parking. However, we were only able to reserve certain rooms in the building during certain times of the day, which turned out to be problematic for scheduling sessions.

Once the date and location were settled, everything else seemed to fall into place. Essential to this success were the committee members and volunteers we added to the team. Kristen Shuyler, a librarian at Seattle University, took on most of the publicity tasks, sending emails to newsgroups and posting on event websites. I created the posters, t-shirts and name badges. Corprew Reed handled venue relations. Andy Szydlowski scheduled the caterer and handled all the equipment. And Rachel Elkington organized a cadre of volunteers who took care of buying food and office supplies, setting up on the day of the event and helping us run the event.

On the day of InfoCamp 2007, over 100 participants arrived, and all our planning and preparation was put to the test. Thankfully, we had plenty of help. Jack Baur, one of our volunteers, recalls this scene:

During the event, I primarily helped get the kitchen ready and the food and coffee into the hands of people who needed it. The set-up was very ramshackle and hurried, but I think that aspect contributed to the high-energy feeling among the staff. We were all so committed to InfoCamp’s success...

The attendees ranged from students to librarians to usability professionals, and several of our sponsors sent representatives. The open BarCamp format proved to be the right approach as the conference unfolded. People enthusiastically signed up for sessions, and all the sessions I attended were alive with discussion and active participation. Our keynote speakers for the two days, Nick Finck and Bob Boiko, gave inspirational and rousing talks in the morning to get people thinking and talking.

ZAAZ, one of the corporate sponsors, held a user interface design mini-event – “Interface-off” – that featured dueling laptops, a series of information architecture and usability challenges and a DJ. A group of librarians spontaneously created a track on library issues, including sessions on technology challenges in rural public libraries and practical uses of social software in libraries. Several disparate sessions were consolidated at the last minute with other sessions with similar topics, such as a combined session on user experience consulting techniques and library reference interviewing. Jack Baur describes the community that emerged:

The atmosphere was open, excited and convivial. The focus was on how much we all had in common as information people no matter where we were working and recognizing the intersections of our work. I got lots of great ideas and met lots of great people. The transparency and openness of the organization really added to that community as well. Everyone’s ideas were valid and everyone’s help, no matter how small, was appreciated. We were all in it together: it really was our conference.

During the closing session of the conference, one participant said that the unconference format was far more valuable than other high-cost training seminars she had previously attended. This comment was met with enthusiastic applause by all in attendance, and we knew we had taken the right approach.

It was such a success, we’re doing it again this year! InfoCamp Seattle 2008 is scheduled for September 27-28 at the Youngstown Cultural Arts Center. For more information, visit http://infocamp.info. For photos of InfoCamp 2007, visit www.flickr.com/photos/tags/infocamp/.
**Tips for Planning Your Own Local InfoCamp**

**Decide on the date & venue first.** All other logistical considerations flow from having the days and location set in stone. This will be, by far, the greatest expenditure. Find a space that is large enough to accommodate up to 200 people, with plenty of individual spaces to use as break-out rooms. Look for community centers, schools and libraries if budget is tight.

**Don’t choose a theme.** It’s not totally necessary to set a slogan or tagline for your InfoCamp. A distinct graphical motif will suffice. The real value of the unconference format is in allowing the themes to emerge organically.

**Pursue sponsors early,** often. Find local businesses and organizations who are recruiting information architects, librarians, taxonomists, usability specialists and other in-demand information professionals. Give them a proposal describing the perks and benefits to them of sponsoring your conference. We offered sponsors links on our website, an exhibit table at the conference and free attendance.

**Make registration cheap.** If you’ve gained enough sponsors, you can afford to charge very little for registration. This low cost allows students and practitioners in not-for-profit organizations to afford to attend your conference. You can streamline the registration process by working with BrownPaperTickets.com, a nearly-free online registration and payment processing website that caters to nonprofits.

**Publicize using social media.** First, set up a website with details about the date and location of your InfoCamp. Then leverage every kind of Web 2.0 social media site you can find to publicize the conference. Create an event and profile for the conference on Facebook and invite everyone in your professional network. Post your event on Yahoo! Upcoming, Meetup.com and other event announcement sites. Post to the local and regional chapter email lists of professional organizations, such as ASIST SIGIA-L, IA Institute, ACM SIG-CHI, ACRL and so on.

**Delegate, delegate, delegate.** Don’t try to plan a conference alone. Find colleagues who are willing to own some part of the process. Also contact your local university’s library school, communications department or human-computer interaction program to recruit volunteers to help you on the day of the event. Offer free (or nearly free) registration as an incentive for students to get involved. Once you’ve delegated, step back and give your team room to be creative!

**Provide free Wi-Fi access.** An essential element of the success of an unconference is the ability for any participant to access the Internet at any time. Such access allows people to blog while at the conference, upload photos in real time, share new sites, look up a reference, download files they’d like to share and so on. Make sure you have a fast connection and multiple access points.

**Supply plenty of coffee and food.** Make sure to provide coffee first thing in the morning and food for lunch, so participants don’t have to leave the site. If you’ve chosen the right venue, they’ll let you bring in your own food and/or caterer, further keeping costs low. Many BarCamps also offer a bar and evening social events to motivate participants to stick around and socialize throughout the conference.

**Provide the framework, but not the content.** Structure the schedule with plenty of slots for multiple sessions throughout the day. However, leave all but the first welcome session blank. Don’t worry – it will fill up.

**Prime the participants.** Find a dynamic and provocative speaker to give a pep talk at the beginning of each day of the conference. Also be prepared with a few topics to give participants an idea of the possibilities for sessions. Have each person on your planning committee ready with a recent (less than one week old) development or inflammatory debate topic in the unlikely event that no one signs up to host a session. You’ll have everyone clamoring to sign up for a slot in no time!
The Northern Ohio Chapter of ASIS&T (NORASIST) took a look at Digital Preservation Basics: How to Ensure Long-Term Access to Your Digital Assets as the theme for an April meeting. Karen Gracy, assistant professor in the School of Library and Information Science at Kent State University, planned a presentation to help attendees understand the critical problems of digital preservation, identify the risks to digital collections and devise short- and long-term strategies for keeping collections accessible into the future.

The recently formed Carolinas Chapter of ASIS&T (cc:asis&t) discussed Institutional Repositories: The Great Debate with Helen Tibbo, School of Information and Library Science at the University of North Carolina, and Kevin Smith, Duke University. In the spirit of classical oration and political debate, the two speakers were to square off on such topics as scholarly communication, copyright, institutional memory, digital libraries, metadata, intellectual property and digital preservation.

IN MEMORIAM

Homer J. Hall

The family of Homer James Hall informs us that the 96-year-old research chemist and information scientist, as well as civic leader, father, grandfather, great-grandfather and friend, died peacefully in his sleep in Exeter, New Hampshire, on April 26. Homer was born in Uniontown, Pennsylvania, and lived most of his childhood in Marietta, Ohio. After graduating from Marietta College in 1931, he went on to Ohio State University where he earned his Ph.D. in chemistry. Homer worked for Esso Standard Oil of New Jersey for 41 years, where, drawing on his ability to read chemistry research reports in seven different languages, he published research abstracts for chemists and chemical engineers to help them stay current in their field. During World War II, he earned patents for the invention of detergent jet fuel. After retirement, he continued to work in the field of information science as a member of the American Society of Information Science and Technology.

In recent years, Homer served on several ASIS&T committees, including Awards and Honors, Constitution and Bylaws, Leadership, Planning and Professionalism. He was also an officer in Special Interest Group/Information Analysis and Evaluation.

Homer was married to Juliet McCarrell Leiper for 65 years until her death a few years ago. He is survived by four children, seven grandchildren; 16 great-grandchildren and his 99-year-old sister. In addition to his wife, he was preceded in death by his parents, a son and his brother.

Memorial contributions may be made to Amnesty International USA, 5 Penn Plaza, 16th Floor; New York, NY 10001.

Sandra Tung

Sandra Tung, formerly of Santa Monica and Altadena, California, died April 23. Sandra retired last year from her position as business information manager for business development at Boeing. Previously she worked at Rockwell, Savage Information Services and NASA Applications Center. Sandra was a longtime member of the Los Angeles Chapter of the American Society of Information Science and Technology (LACASIS). During the 1980s, she helped make LACASIS a stronger, more vibrant organization. A talented leader with a “can-do attitude,” Sandra understood the value of working with new members to insure the next generation of leaders would be strong. She was active in LACASIS for over 15 years, serving as treasurer, marketing/database coordinator, Program Committee member and Awards Committee member. In 1994 she was recognized for her service and won the Outstanding Member Award.

One of Sandra’s legacies is the continuing success of LACASIS as a professional organization. She will always be remembered as a LACASIS superstar.

Sandra is survived by her children Doug, Diana and Irene; three grandchildren and a loving extended family.

Donations may be made in her name to San Gabriel Valley Habitat for Humanity, 770 N. Fair Oaks, Pasadena, CA 91103 (www.sgvhabitat.org).
As of January 2008 the *Journal of the American Society for Information Science & Technology* (JASIST) will follow the “green road” open access model – authors publishing in the *Journal* (and only authors) may post preprints to their own or their institution’s repositories with links to the final article and to the online journal.

The ASIS&T Board of Directors arrived at this decision after a detailed analysis that weighed the logistical, economic and philosophical issues involved. With the JASIST publisher, Wiley-Blackwell, we agreed to institute this policy on a trial basis and to work together to monitor the impact of the new policy. We were interested broadly in understanding what effects the new policy would have on submissions, readership and subscriptions, and we agreed to conduct baseline and subsequent surveys to track changes in these areas.

The first survey was conducted online in late 2007 and addressed four areas:

- Who are our members, authors and potential authors?
- What are the publication trends among these groups?
- What level of access is there to journals among these groups?
- Are our members, authors and the information science community aware of and/or participating in the open access movement?

This article reports generally on the results of survey. Two additional articles in the *Bulletin* in the coming months will continue discussion of the survey results, focusing on the subsets of respondents who were ASIS&T members and open access authors.

### Groups Addressed by Survey

There were three groups we wanted to reach with the survey: current ASIS&T members, JASIST authors and information science researchers. These three groups are not separate and distinct: many individuals belong to all three categories.

Blackwell collaborated with ISI to determine contact information for authors currently publishing in the fields of “information science & library science,” “library science” and “information technology and communication systems.” They also determined contact information for authors who had published articles in *JASIST* or similar publications. The peer-reviewed publications selected for the survey were the *Annual Review of Information Science and Technology*, *Information Processing and Management*, *Information Research, Journal of Documentation*, *Journal of Information Science*, *Library Information Science and Research* and *Scientometrics*. From these contacts, Blackwell distributed the survey to 3,740 researchers. Additionally, ASIS&T distributed the survey to a randomly selected group of 2,414 members, some of whom may have overlapped with the authors identified by Blackwell and ISI.

We received 581 responses of which 348 (or 59.9%) were from ASIS&T members.

ASIS&T and Blackwell developed the survey questions jointly. A number of questions were taken from surveys created by Ian Rowlands and his collaborators.

[1] [2]

### Who Are Our Members, Authors and Potential Authors?

Information science is traditionally seen as an area of western scholarship, but our survey results indicate that a growing number of scholars resides outside the western world. While most of the respondents were from North America...
(71.9%) and Europe (11.4%), 12.9% identified themselves as being from other areas of the world. Regionally, 7.4% were from Asia, 2.8% the Middle East, 2.2% Australia and 0.5% South America. In addition to these respondents, we received several write-in comments from researchers working in Africa, a region that was accidentally omitted from our responses. As one commenter pointed out, this distribution may reflect “the imbalance in scientific rating globally.” The discipline of information science is growing globally and we must consider the implications of scholarly communication and publication models in a global context.

The majority of respondents (70.2%) identified themselves as employees of colleges or universities. This result is not surprising, considering that academic positions require publication for promotion and tenure and that our survey solicited authors who had published in premier journals in the field of information science. We did see a small number of responses from researchers in other institutions, including commercial organizations (6.7%), government (5.9%), research institutes (3.3%) and self-employed individuals (3.8%).

The respondents were also asked how many years of research experience they had. The two most common responses were 1-5 years (24.6%) and 21+ years (23.8%). This spread seems to indicate that there are many researchers at both ends of the spectrum, when it comes to research experience, with fewer people in the middle. It also underscores the need for mentoring in our field as the reins pass from one generation of researchers to the next. Overall, the discipline seems to be undergoing a transition. There are increasing numbers of new researchers and contributions coming from increasingly diverse regions.

**What Are Current Publication Trends among Information Science Researchers?**

The majority of those surveyed, 68%, submit research papers to peer-reviewed journals. The following are the top 10 journals considered by the 395 people indicating that they submit research papers:

- **Journal of the American Society for Information Science & Technology** (197 responses)
- **Information Processing & Management** (92 responses)
- **Journal of Documentation** (65 responses)
- **Library and Information Science Research (LISR)** (50 responses)
- **College and Research Libraries** (49 responses)
- **Scientometrics** (39 responses)
- **Journal of Academic Librarianship** (35 responses)
- **Journal of Information Science** (35 responses)
- **Library Quarterly** (31 responses)
- **Journal of the Medical Library Association** (24 responses)
- **Portal: Libraries and the Academy** (24 responses)

Authors from JASIST, Journal of Documentation, Library and Information Science, Scientometrics and Journal of Information Science were solicited for this study so these results may not be reflective of overall trends in the field.

Even though the same authors rated impact factor (IF) as an important criterion when choosing the journal to which they would submit their publications, only three of the 10 journals with the highest impact factor in the field of “information science & library science” as listed in Journal Citation Reports (JCR) appear on the list above. They are JASIST (IF 1.555, ranked 6th by JCR), Information Processing and Management (IF 1.546, ranked 7th by JCR) and the Journal of Documentation (IF 1.439, ranked 9th by JCR). The five journals with the highest impact factors in JCR do not appear in the list of the top 10 journals that the authors in our survey consider when submitting. Of these five only two, MIS Quarterly (IF 4.731, ranked 1st by JCR) and the Journal of the American Medical Informatics Association (IF 3.979, ranked 2nd by JCR), appear within the top 25 journals considered by our survey respondents.

The authors who submit research papers to peer-reviewed journals were asked how many journal papers they had written in the past three years. The most common responses were two (16.7%), three (14.7%) and four (12.7%). In the past three years 61.5% of authors had published one to five articles, while 22.8% had published 6-10 articles and...
12.2% had published more than 10 articles. It seems that, for the most part, the authors in our survey averaged one to two peer-reviewed journal articles per year.

When asked how many peer-reviewed journal articles they had published overall in their careers, the responses seemed to be consistent with the number of years of research experience. That is, the most common responses were at the low and high ends of the scale with fewer responses in the middle. The most common responses were three articles (6.6%) and 30+ articles (6.6%). Overall, 32.2% had published 1-10 peer-reviewed articles in their careers, 22.3% had published 11-20 articles, 32.4% had published 21 or more articles in their careers and 13% did not respond.

What Level of Access Do ASIS&T Members and Information Science Researchers Have to *JASIST* and Journal Literature in General?

Overall, the respondents had very good access to journal literature. When asked, “How would you describe your current level of access to the journal literature?” 75.4% responded “good” or “excellent.” The most frequent response was “Good; I have access to most journals I need” with 45.1% of the total. This result was followed by 30.3% responding “Excellent; I have access to all the journals I need.” Only 4.8% rated their access level to journal literature as “poor” or “very poor.” The high level of access to journal literature in our sample group may be related to their professional careers in colleges and universities.

Of those surveyed 80.8% currently read *JASIST*. The most common frequency of use was every issue (33.6%), followed by one or two times a year (21.7%), every other month (17.4%), never (12.4%) and less than once a year (8.1%). A small number of users, 4.3%, responded “I would use it if I were able to access it easily.”

The respondents were also asked how they accessed *JASIST*. They were asked to indicate all the ways they accessed the journal. In these responses 62.8% used a membership subscription to read the journal in print or online while 42.5% indicated that they accessed the journal through a library or library service, such as interlibrary loan. The most common format used was electronic, with 60% indicating that they access the journal either through their library’s electronic license or through online member access. Among those surveyed, there did not seem to be problems accessing the journal.

Are ASIS&T Members, Authors and the Information Science Community Aware of and/or Participating in the Open Access Movement?

There was an overwhelming awareness of open access (OA) with 95.7% responding that they knew about open access journals. Of those, 60.4% knew “a lot” or “quite a lot” about OA. Only 4% responded that they knew nothing at all in regard to open access journals. Even though most of the respondents were aware of OA journals, the majority, 64.7%, had never published an article in one.

Overall, there was a positive attitude towards open access as evidenced by both the survey responses and comments. Respondents were asked to rate traits they associated with OA journals on a scale of 1-5, with 1 meaning that they “do not associate” the trait with OA and 5 meaning that they “very strongly associate” it with OA. Very few respondents thought of open access as “radical,” “ephemeral” or “not archived properly.” These qualities averaged 2.2, 2.5 and 2.3 respectively (or “associate a little”) on the scale. The concept of the “author pays to publish” was not associated with the concept of open access either. It rated an average of 2.3 or “associate a little” on the rating scale. This result is surprising since the gold road, or full open access model, often relies on author fees for publication.

At the end of the survey the respondents were asked if there were “any other thoughts or experiences they would like to share about the topics of this survey.” One hundred twenty-three people responded with further comments. Approximately 48% (59 responses) commented on open access specifically. These comments fell into three categories: people who were passionately in favor of OA, people who were hesitant about OA and people who liked the idea of OA but did not want publication fees. Many of the comments were strongly supportive of OA, asserting that “open access is no longer why or whether, but when and how” or “open access is vital for the future of publishing.” These comments often
pointed out the accessibility benefits of OA and the need for change and emphasized that scholarly societies should be leaders in the OA movement. There were also several respondents who were hesitant or skeptical about OA, and they provided comments such as, “I’m not sure that the rush to open access is so sensible” and “I am not yet persuaded that a viable economic model for open access can be sustained.” The third group was positive about the merits of OA, but was opposed to the idea of authors paying for publication. A typical comment in this group was, “I am strongly in favor of open access journals, but not if they charge substantial fees to authors.”

Conclusion

This survey gave ASIS&T the opportunity to answer four questions:

Who are ASIS&T members, JASIST authors and information science researchers?

We found that the face of information and library science is changing. Researchers from Asia, the Middle East and Africa are becoming more active in the discipline. Also, there are many researchers who are either beginning their careers or who have 21+ years of research experience, with fewer researchers in between.

What are the publication trends among these groups?

Of those surveyed 68% submit articles to peer-reviewed journals. It appears that on average the authors in our sample group published one to two papers a year.

What access levels do they have to journal literature in general and JASIST specifically?

In answer to this question 75.4% of the respondents indicated that they had “good” or “excellent” access to journal literature and 80.8% read JASIST. The majority, 60%, accessed the journal on line either through their library’s institutional subscription or through a personal membership.

Are they aware of and/or participating in the open access movement?

There was an overwhelming awareness of open access with 95.7% of respondents indicating that they had some knowledge of open access journals. However, only 29.4% of authors had published in an open access journal.

This survey establishes a baseline for the further survey data we will solicit after the JASIST open access policy has been in place for a period of time.
POSITION DESCRIPTION

Editor-in-Chief

JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY (JASIST)

NATURE OF POSITION: The Editor’s foremost responsibility is to ensure the high quality and quantity of papers published in JASIST. Working in collaboration with the Editorial Board, the Editor’s additional responsibility is to set the vision and scope for the Journal in a manner consistent with those of the Society. The Editor solicits high quality papers that fit the scope of JASIST, and oversees the activities of the Associate Editors, Guest Editors, and Editorial Board to ensure that submitted papers are peer reviewed by appropriately qualified persons. The Editor ensures prompt communication with authors to accept or reject manuscripts or to request revisions in response to referees’ reports. The Editor, in consultation with the Society, appoints and renews Editorial Board members for an agreed term ensuring that the Board’s composition is sufficiently international and broad in scope to maintain JASIST’s stature worldwide within its field. The Editor communicates at least annually with the Editorial Board concerning the development of the Journal, editorial strategy, submissions and promotion. The Editor assists the Publisher in promoting JASIST by advising on publicity, and promoting JASIST wherever possible through contacts and at conferences attended. The Editor is responsible for the intellectual workflow (e.g., using the electronic manuscript system, sequencing papers in issues, etc.) and ensuring that the instructions to authors are followed (e.g., that papers are original and have not been published elsewhere; that papers are not defamatory or otherwise unlawful, and that appropriate illustrations and tabular matter, permissions, and assignments of copyright are included). The Editor receives some support for editorial assistance from the Publisher.

BACKGROUND REQUIREMENTS:

a) Education: Graduate degree.

b) Skills Required: The Editor must be a leader who has strong motivational, interpersonal, and communication skills. He or she must be highly motivated to publish a successful journal and must recognize the broad scope of information science and technology. A global perspective, flexibility, and diplomatic skills are required to encourage diverse and creative contributions and to arbitrate controversial issues and points of views. Knowledge of electronic communications and manuscript management systems is essential. The Editor must be familiar with electronic publishing trends, issues of originality and reuse, open access points of view and publishing economics. In addition to oversight abilities, the Editor must be willing and able to delegate responsibilities.

c) Term: The term for this position is five years. A limit of two terms is imposed.

d) Relevant Experience: Substantial knowledge of JASIST and the field of information science and technology and a strong record of scholarly publication. Prior editorship experience and/or editorial board experience with a comparable scholarly, technical, or scientific (refereed) journal is highly preferred.

APPLICATION PROCEDURE

Applications will be considered beginning on March 1, 2008 and until the position is filled. Send applications to: JASIST Editor Search Committee c/o Richard Hill ASIS&T 1320 Fenwick Lane, Suite 510 Silver Spring, MD 20910, USA

The application package should include, but not be limited to, the following:

- Applicant’s vision of information science: the directions it should take; its hot, warm, and cold areas;
- Applicant’s publishing & editorial experience;
- Reasons why the applicant is specially qualified as Editor-in-Chief of this publication;
- Applicant’s view on current issues and strategies with the publication: backlog, special issues, etc.;
- Applicant’s ideas on ways to improve the readability of the publication: special initiatives, survey articles, etc.;
- Applicant’s ideas on methods to shorten reviewing time and strengthen the review process;
- Experience with electronic manuscript management systems;
- Applicant’s vision of JASIST as an international journal;
- Applicant’s vision of JASIST Editorial Board structure and function;
- Applicant’s understanding of ASIS&T, and conception of the relationship of the Journal to ASIS&T members;
- Other plans for improvement.

The applicant must also provide a detailed resume, listing all past and present affiliations, editorial positions held, and activities in professional societies and technical conferences. Further, the applicant should provide a complete list of publications, honors and awards received, and other information deemed relevant to the Editor-in-Chief position.
Introduction: Bringing Genre into Focus
by Luanne Freund and Christoph Ringlstetter, Guest Editors of Special Section

Most of us spend considerable amounts of time seeking information, navigating and carrying out tasks within complex digital information environments such as the Internet, digital libraries and workplace intranets. These environments are filled with an amazing array of documents, expressions of the human impulse to communicate by sharing our knowledge and experience with others.

From the standpoint of an information professional it is easy to lose sight of the creative and expressive aspects of documents, given the decades-old practice of working with surrogates: subject headings, keywords, abstracts and bags-of-words, rather than the full text of documents in their original form. While these surrogates are capable of representing the language and concepts contained in a document to varying degrees, they do not do a good job of representing the document as a whole and provide very little contextual information as to why a particular document was created, what purpose it was meant to serve and how it might be used. But information environments and systems have changed dramatically over the past 10 to 15 years, and increasingly we have access to complete, full-text documents in digital form. Alarming, one approach to these digital documents is to consider them simply as containers, which add nothing of value to the information content. Proponents of this approach would have us “liberate” the content from documents, pushing and streaming it through networked information spaces to pop up in other locations and in different forms.

The concept of genre, which is the focus of this special section, offers a strong counter-argument to this content-centric approach, suggesting that we have much to lose by abandoning established document conventions. Genre has been used since ancient times to classify texts and in the modern world is a prominent feature of almost all large collections of human-authored documents. Genres are naturally occurring patterns of communication, characterized by common elements of form, function and content, which arise out of recurring situations and communication needs and are recognizable within communities. Some of the most highly structured and familiar genres, such as scholarly articles, have long and fascinating histories and have migrated, relatively intact, from print to digital format. Others, such as Frequently Asked Questions (FAQs) are newborn and surprisingly robust web-based genres. Genres exist because they add value to the communication process – otherwise, they would neither have emerged nor been sustained over time.

In this special section, a group of authors have come together to present and discuss the benefits and challenges of focusing on genre within information science and technology research and practice. Genre-based approaches are applicable in a wide range of areas: knowledge management, information architecture, information retrieval, information systems, archives and records management, and information interaction and use. The idea for this collective effort emerged from a panel on genre and information retrieval presented at the ASIS&T Annual Meeting in Milwaukee in 2007, which

Luanne Freund is an assistant professor at the School of Library, Archival and Information Studies at the University of British Columbia in Vancouver, Canada. She can be reached at luanne.freund@ubc.ca.

Christoph Ringlstetter is a postdoctoral fellow at the Alberta Ingenuity Center for Machine Learning (AICML), University of Alberta (Canada). He can be reached at Ciskristof@at@yahoo.de
prompted a lively exchange of ideas among panel members and conference attendees. The scope of this special section extends beyond that of the original panel and delves more deeply into the value proposition of focusing on genre.

In “Why Information Has Shape,” Andrew Dillon presents a thoughtful essay on how and why genres emerge through language use as patterns that facilitate communication. He emphasizes the value of genre in providing cues and structures to support easier information processing, comprehension and navigation. Focusing on genre, according to Dillon, would provide a framework for thinking about information use, a vital but under-studied area in information science.

Mark Rosso turns his attention to the value of labelling genres to support web retrieval in his “Stalking the Wild Web Genre.” He draws upon his dissertation research on genre recognition and presents a lively and detailed list of the pros and cons of genre-based web retrieval. The prevalent role that genre already plays in our everyday information practices is one of the strongest arguments in favor, and the practical difficulty of genre classification is one of the main obstacles. However, Rosso leaves it up to the reader to decide.

In “Situating Relevance Through Task-Genre Relationships,” Luanne Freund offers an alternative perspective on the use of genre in information retrieval. Freund makes the argument that genres carry the stamp of the situations in which they were created and for which they were intended, which makes them more or less suited to the tasks in which searchers are engaged. This relationship between tasks and genres is valuable as a means of determining situational relevance, which search systems to-date have largely ignored.

Christoph Ringlstetter and Andrea Stubbe discuss “Practical Aspects of Automatic Genre Classification.” Problems of defining an appropriate genre palette, the selection of expressive features and the choice of an effective classifier are discussed. The proper engineering of these challenges is a pre-condition to many of the system-based applications of genre that are under discussion.

In the wrap-up article to this section, “LIS and Genre: Between People, Texts, Activity and Situation,” Jack Andersen presents a broader case for the value of focusing on genre. Rather than emphasizing the role that genre plays in supporting recognition, retrieval, processing and use of textual information, he argues that thinking about genre would result in a fundamental and much-needed shift in the focus of LIS research and practice toward understanding how and why documents are produced, used and organized in the context of human activities and work.
Why Information Has Shape
by Andrew Dillon

Language, in both written and spoken forms, contains many subtleties that a member of a speech community slowly learns to decode and use. Grammatical rules and vocabulary occupy most attention early on, but as we become more and more experienced, we learn to use language constructively and in patterns that serve as a bridge between sender and receiver. In written language, where the normal restrictions of working memory capacity can be overcome through cueing, headings and the ability to retrace previous utterances, humans have evolved a range of genres that serve as patterns for communication over lengthy discourse.

Genre studies can become very technical, but the basic functions of genre are not difficult to grasp. As stylized patterns of communication adopted within groups, genres serve to establish the parameters of a communication process and in so doing prime the participants to deliver and anticipate receiving certain types of information. Conforming to these expectations enables easier processing of the information flow and also signals progress through the communicative passage to participants, both temporally and semantically.

These naturally emerging regularities of language that we term genre influence human information processing in a regular, repeatable manner. Humans are natural pattern extractors – our brains seem to seek regularity in complex data, and we are known to impose patterns where they don’t even exist. I used to regularly demonstrate this phenomenon to classes by presenting them with truly random letter or number sequences and asking students to guess the next two elements in the series. Given sufficient time, students would invariably produce answers with sometimes tortuous arguments of why their sequence logically worked, and some would not accept that the sequence had no underlying logic or pattern. Such pattern-seeking dispositions help us organize a world of data coming at our senses continuously. In this way, it is not hard to understand the importance of genre in information activities. When attuned to their existence, a reader, a listener or a viewer has a mental framework into which she can put incoming data, providing context and cues for comprehension, priming her to expect further data of a certain type and drawing attention to components of the communicative process where data may be missing.

We have studied the process of making sense of informational forms when they conform or violate such expectations of genre. For example, in one study, we developed an online newspaper that contained real text from a daily newspaper, grouped into typical story types (current affairs, local, national, sports, etc.), but was presented in ways that reflected either typical newspaper layout and form or ways that violated such expectations. When people were asked to interact with these digital newspapers continuously for several days, we observed their navigation and checked their comprehension for content. Our results showed that the initial response to the genre-violating form was quite negative, and their performance revealed greater navigational difficulties and poorer ability to locate specific information. Over the course of several days, their performance improved but they never performed as well as readers exposed only to the more typical, genre-conforming newspaper.

It seems, therefore, that people respond to structural regularities in language that create a sense of shape within an information space and people employ these cues, often without conscious awareness, to guide their exploration and comprehension. Since this appears to be a natural component of information use, we should not be surprised that our routine communications evolve conventional forms. At one level, the forms give rise to visual cues and structures (headings, layouts, access mechanisms) that are easily understood.
However not all conventions lend themselves to instantiation so simply, and one finds the cues buried within the text in the form of language employed, the high level order of content and narrative, and the form of information provided at various parts of an extended document or exchange.

It follows that the perception of genre conventions and the ability to interpret their uses depend on the experience of a user. Given the subtle nature of many genre conventions, multiple and repeated exposure to a form is often required to understand its application. Conversely, understanding and using genre conventions appropriately might be considered a defining characteristic of expertise in a domain. This observation is certainly true of advanced communicative forms such as academic articles, scholarly books and technical reports. In a study of scientific article users, we showed that experts could read an isolated paragraph of text from a journal and estimate with almost perfect reliability in which section of the journal article this text belonged. Novice readers of this literature were unable to do so. However, it is also likely that we can discriminate users of various digital environments on the basis of their familiarity with genre.

Given the obvious potency of genre conventions in information use, it might seem natural to consider genre as a cue for retrieval, but I am less convinced this technique has real value. Certainly we can tag documents with genre labels rendering them identifiable on this basis, but it is not clear that such a label would add significant information to an already tagged set of documents. Knowing that a journal article appears in a scholarly publication is important, but this determination can come from many existing sources independent of a formal genre descriptor. But genre attributes can add significant value as navigation aids within a document, and if we were able to determine a finer grain of genre attributes than those typically employed, it might be possible to use these as guides for information seekers. For this to work, we would need a way of describing aspects of the information shape that reflect underlying narrative and semantic flow, allowing us to use these, relatively unambiguously, at the document level so as to support refined searches within a document type.

Such analyses offer clear research targets, and we have attempted to do something of this form, at least at the visual level, with a search interface for poetry that presents a template to the user in the form of a verse structure, with a series of blank lines representing the layout of a title and verse. The user can input any term in any space and the search interface interprets accordingly to find the term in the title, the first line, other than first line but somewhere in the verse, etc. Poems have relatively regular physical structures that allow this, and the user response is generally positive. But extended narratives offer a more complex challenge given their more uniform physical formats and the need for a more semantic reflection of their shape.

You can appreciate the importance of semantics by reflecting even on your own use of information space. While significant research efforts are expended on retrieval and navigation, people do not retrieve documents for the purpose of navigating them. They retrieve them to use, to be informed, to understand, to solve a problem or for similar reasons, and they navigate these documents to gather meaningful content for such tasks. Thus, retrieval and navigation are elements of a task, but they are not the complete task, and I often wonder if we have lost sight of this reality in our research efforts. There seems to be a whole literature given over to these task elements without an equivalent level of effort to understand how meaning is extracted or how people make sense of the documents they are navigating through. Information use such as reading, comprehending and comparing involves both spatial and semantic processing interchangeably if not simultaneously, as people create working models of the space in which they are working. People learn to recognize regularities in the spatial form (layout, headings, length and so forth) and in the semantic form (introduction, method, discussion, analysis and so forth) and then employ both elements to guide their process. In the field of human-computer interaction there have been decades of work on the spatial aspects of designing the best interface mechanisms to aid navigation, but precious little work addressing the semantic elements, which can actually aid a user in comprehending the documents being used. This is an imbalance in research that should be corrected in the years ahead.

In all this we really ought to recognize the incredible sophistication of cognitive processing that underlies the evolution of our world’s information space. At this moment we sit on the inherited forms of documents refined continually by creators and users, members of speech and practice.
communities, that has resulted in forms that allow experienced users to convey and extract meaning rapidly or to leverage properties of the document in support of comprehensive processes that would exhaust our limited cognitive architecture were we to rely on real-time communication only. With new digital forms possible, we can witness in our world now the costs of violating genre and the process of enabling new genre forms. As we have shown in our research, the genre of a web home page took hold very quickly in the 1990s – in a matter of years, when most scholars of the paper world were convinced genre only emerged over decades.

The apparent chaos of digital communications is superficial and may reflect a bias in perception by those schooled in the paper domain. It is not enough to apply genre conventions from paper to digital space, the transfer is not always neat, and we should realistically expect native digital genres to evolve through the community of users most actively communicating through these channels. Such genres will evolve and as they do, we may lose genre forms from the pre-digital era. This transformation is the shape of information, and the continuous tension between fixed and fluid is part of the natural order.
Stalking the Wild Web Genre
(with apologies to Euell Gibbons)
by Mark A. Rosso

Genres – they are not the offspring of the taxonomist’s notebook or traditional thesauri, but rather, they originate from people’s everyday speech and usage of information. Genre names are useful for referring to communication in the recurring situations in which we find ourselves – whether it’s looking for a job/hiring someone (a resume) or acknowledging a thoughtful gift/having that gift acknowledged (a thank-you letter). How can this everyday terminology be useful for web retrieval?

Imagine being able to tell your web search engine the types of pages you are looking for and even types that you don’t want to see: that ability is the vision of incorporating genre into the interface of web search engines. Genre could be part of the query formulation or a description of each result on the search results page. It’s an idea that’s been around for a while. But we’ve yet to see an implementation of retrieval by genre in any web search engines. Research issues have included the identification of specific web genre labels and genre classification by automated algorithms. For the last decade or so, academicians have been pursuing the idea with mixed results so far. Is it time to give up the pursuit, or is there still uncharted territory out there to discover?

Let’s look at the pros and cons of this somewhat controversial research area.

Reasons to Investigate Web Retrieval by Genre

The following are among the reasons to investigate web retrieval by genre:

- **Genre is used to retrieve physical documents.** People normally store and retrieve documents by genre. We all have places to store our books, old income tax forms, receipts, etc. Certainly, libraries and archives make good use of genre as well. We will return to this point at the end of the article.

- **The web has already spawned unique genres, while many traditional genres have migrated to the web.** While poems, recipes, newspapers and other formats are now on the web in versions still similar to their print lives, totally new web genres have sprouted: homepages, blogs, FAQs and wikis. Information seekers should be able to exploit the obvious differences between these document types in the digital realm.

- **There is a need for non-topical search descriptors.** Keywords that we type into Google are typically representative of topic or subject, and we all know that search needs are more encompassing than that. Although a savvy searcher can often use keywords to indicate the contextual aspects of an information need, a document’s contextual information is generally not easily predicted by the words it contains. It makes intuitive sense that our ability to precisely describe our information needs must increase as the web continues to grow. We can be pretty certain that our patience to sift through 10 or 20 search results will not grow at the same pace.

- **Studies have shown that people can recognize the genre of digital documents.** What’s missing with these new genre are aspects of **physical form** – they’re all flat on your screen, whereas in the physical world, we distinguish recipes on 3-by-5 index cards from the morning newspaper without any thought. Luckily, research has shown that folks can distinguish certain types of digital genre pretty well. In fact, independent user studies conducted in Sweden, Germany and the United States for the purpose of eliciting the web genres that people conceptualize have identified similar genres [1]. It would seem that shared genre knowledge of web pages has a somewhat cross-cultural basis!

- **Genre is a compact way to describe a document.** With the burgeoning global trend to mobile search from our personal devices, novel ways are needed to present search results on small screens. In just one or two words, a genre label can say a lot about a webpage.
People think this is a good idea. Some of the most popular tags for web pages on the social tagging site delicio.us are genre labels, such as blog, howto, tutorial, news and research. Participants in an experiment that studied the evaluation of genre-labeled web search results said they liked having the genre label there [2]. Lots of researchers, including those who have written in these very pages before, have focused their research agendas on web genre [3]. Several classifications for digital objects, including the Dublin Core, have fields to describe genre-esque attributes such as document type or resource type. FaceTag, a prototype collaborative tagging tool for the web (described in a Bulletin issue last year [4]) uses resource type as one of 10 proposed facets. The Hawaii International Conference on System Sciences (HICSS) has sponsored a mini-track on digital genres for many years now. Finally, the National Science Foundation awarded a team of academic researchers a $150,000 grant in 2005 to explore the use of genre in web retrieval, and there is currently another much larger grant proposal under consideration.

Reasons that Web Retrieval by Genre Might Not Be a Great Idea

The following are among arguments suggesting that web retrieval by genre might not be a workable notion:

- **Not all web pages are the result of typical, recurring situations.** What makes genres so useful is that they embody time-tested communicative actions and reactions to circumstances in which we human beings often find ourselves. However, it makes sense that some web pages are not products of typical situations, or if they are, the searcher may have no idea what that typical situation is. See the next point.

- **Any given web user is not a member of many of the unlimited number of genre user groups represented on the web.** Have you ever clicked on a link on a search results page and gotten something, but you had no idea of what it was? The web gives us access to all sorts of documents that we would never have had the opportunity to view without it. Unfortunately, that also means we can see pages that were never intended for us or that we cannot even remotely understand, lacking the context in which they were developed. If these pages do have genres, they are not ones that have any meaning for us. Their labels may not be of much use to us (or many users) for search purposes.

- **Genre information is already present in many web search results.** Studies have shown that information already present in search results – in the title, the snippet or the URL – oftentimes indicates the genre of the web page. For example, we’ve all seen the tilde in a URL which strongly suggests a personal page. With all those implicit genre cues out there, the benefit of adding a genre label to search results is diluted.

- **Genre is a moving target.** Research in genre theory has shown very clearly that genres evolve over time because the circumstances in which we find ourselves and the acceptable responses to these circumstances change, too. This evolution is especially true today and especially true for the web. Genre research done in 2004 reported on study participants who didn’t know what a blog was. There would be much less of that today in 2008, I’m sure. Why is this genre evolution a problem? It means that the genre labels used to describe web pages and the algorithms used to automatically classify web pages would need to be updated on an ongoing basis. This would be a lot of work for a search engine feature whose benefit is presently unclear.

- **Users may not understand the relationship between genre and their search (or there may not even be a relationship).** During my dissertation research, several subjects made inaccurate relevance judgments based on genre labels in the search results [2]. What they were looking for was on the target web page, but the subjects were of the opinion that a personal page or a blog (as advertised in the search result’s genre label) wouldn’t be the place to find what they were looking for. This skepticism may be a product of the moving target phenomenon noted above. If genres are evolving, it makes sense that people will catch on at different rates, leading some to make wrong decisions based on past assumptions or observations.

- **Even if useful genres can be articulated, it is unclear that machine classifiers are up to the task.** Although many genres have distinctive forms that classifiers could take advantage of, some don’t – they are documents that typically contain paragraphs of text, distinguished primarily by the type or style of verbiage. A news article with a title
and text can look like an executive summary – until you start reading. Given the large number of documents and document types, and the diversity of authors and tasks, web genre may just be too heterogeneous for classifiers to handle at an acceptable level of error.

- **Even if machine classifiers are up to the task, who’s going to annotate all that training data?** Unless we can make a game out of it, like Luis Ahn’s ESP Game, this annotation is an expensive proposition – especially if these moving targets need to be continually maintained! Maybe type social tagging (think del.icio.us) can come to the rescue here…

- **Even if implementation is possible, genres could be spammed easily.** Think personal pages that are just there to sell you life insurance. More seriously, especially if advertising were a genre that one could exclude from search results, many pitches would go underground and masquerade as other genres.

### Conclusion

I can sum up the search for the elusive web genre in three steps. First, appropriate genre labels and definitions must be determined and validated through user studies to ensure that the terminology is indeed understood by users as indicating familiar genres. Second, it needs to be shown that the genres identified are useful for web search tasks. Finally, the genres must be predictable by machine algorithms as there are too many pages on the web for search engine companies to classify by hand.

Genre is not the only research area that is held back by a lack of usable training data. In fact, some researchers [5] have called for the establishment of a discipline called **annotation science** (www.itl.nist.gov/iaui/894.02/minds.html).

The objective is to develop processes for identifying what needs to be annotated, the annotation process itself (both manual and automatic) and the embedding of best practices into tools to help streamline the process. Genre researchers should definitely take part in this effort.

I want to leave you with a story [6] that was relayed to me by someone who managed an engineering department’s project files for a large corporation in the early 1970s. The company’s capital construction projects, ranging in value from a few million dollars to more than $100 million, on average generated a lateral file drawer for every $2 million of a project. That’s a lot of documentation. Under the original file organization, retrieval success for engineers’ requests for project materials was only about 60%. To make matters worse, it was impossible to tell if the materials needed in the other 40% were misfiled or never even received to begin with. After studying the patterns of engineers’ requests, the manager observed that more than 90% of the time the engineers invoked document type as part of their retrieval request. Consequently, she initiated an experiment to re-organize the project materials by genre and then by company name (equivalent to corporate author) within genre. Retrieval success quickly improved to 96%, and it could be determined with certainty that the other 4% was never received. Organization by genre was the key to success. Ongoing analysis of incoming documents ultimately identified more than 80 unique document types which were applied as the primary organizing structure of all new project files.

The lesson to be learned from this story is that genre can be a powerful hook into the relevance of a document. And, as far as the ever-growing web is concerned, Web searchers may soon need all the hooks they can get.
Situating Relevance Through Task-Genre Relationships
by Luanne Freund

One of the most discussed concepts in information science is relevance, a relationship in which one thing has a direct bearing on another. We assume that people seek out relevant information to deal with a question or problem, and we design information systems to identify the most relevant information for any given query. Yet assessing the relevance of information is far from straightforward in real-life situations. Several decades of user studies have shown that people use a wide range of criteria to select one information source over another, such as subject matter, style, format, novelty, quality and authorship. Furthermore, perceptions of relevance change from person to person over time and from one situation to the next. In information seeking, as in yard sales, one man’s treasure is another man’s trash.

The complexity and subjectivity of real-world relevance led early information retrieval researchers to focus on the relatively static and measurable concept of logical or topical relevance (is a given document about the topic of a searcher’s information need?), rather than the dynamic and subjective concept of utility or situational relevance (is a given document suited to a particular searcher’s situation at a given point in time and space?). As Cooper noted in 1971

“The system designer cannot currently do much to ensure that only the most credible material will be retrieved: such a goal lies largely beyond the state of the art at present. He cannot gauge the relative importance to the user of different component statements either…Logical relevance is almost the only factor in utility which the designer does know how to deal with very effectively at present. This suggestion, if true, would help to explain why topic relevance and not utility has received the most attention in the literature.” [1, p. 36]

However, it is now almost 40 years later. Search systems have developed rapidly since the 1970s, and yet the state of the art still pays very little attention to the impact of the searcher’s situation on relevance. Search engines retrieve thousands of documents for every query and depend upon the searcher to sift through and find the treasures among the trash. Despite their indiscriminating nature, search systems are the most effective tools available to deal with the swelling sea of digital information. In the face of our growing dependency upon them, it is essential that we begin to develop systems capable of discriminating between documents that are simply on topic and those which are genuinely useful to the searcher. In other words, it is time for information professionals and system designers to come to grips with situational relevance.

Matching Documents to a Searcher’s Situation

In order to design systems capable of measuring and making use of situational relevance, we need to be able to model the searcher’s situation, identify the associated document features and find a way to match these up. By and large, topical relevance is measured by comparing search queries with documents and identifying the degree of term overlap. The strength of the query-document match varies according to the frequency and location of the overlapping terms. If we apply the same logic to situational relevance, we should look for evidence of situational overlap between searchers and documents. The concept of situation is not as well-defined as topic or subject, but refers to some combination of personal, social, organizational and physical factors that exist at a given point in time and influence a...
person’s behavior. Of these factors, the task in which a searcher is engaged has been identified as a dominant feature of the situation. So one way in which overlapping situations are expressed in information seeking is that, in a given task situation, a searcher will prefer document genres designed to support that task. For example, a homeowner installing a dishwasher is likely to prefer a set of instructions rather than a product review or technical specifications. In the workplace, a project manager preparing a marketing presentation for a line of dishwashers is likely to prefer promotional materials or other marketing presentations over user manuals or installation instructions.

Of course, these genre preferences are secondary and supplementary to topical relevance. Information that is not about the specific brand of dishwasher in question will be largely irrelevant, regardless of the searcher’s task and the document genre. And, if only a handful of documents about the specific dishwasher are available, then distinguishing between them on the basis of situational relevance would not add much value. Nevertheless, given that most search queries are very brief (one to two terms) and retrieve many more documents than searchers need or want, techniques to further refine search results are sorely needed. The remainder of this article will expand upon this notion of using task-genre relationships to determine situational relevance.

Task as a Situational Model for the Searcher

Tasks are activities undertaken to achieve goals. All of us are engaged in tasks, both assigned and self-motivated, in the course of our daily lives. These may be work or leisure tasks, for example, researching a new drug under development or designing a poster for a community dance. In the course of carrying out work or leisure tasks, problems and questions arise that prompt more specific tasks in the form of information tasks and search tasks – activities focused on the acquisition and use of information. Some common types of information tasks are learning, making decisions, finding facts, carrying out a process and solving problems. So, for example, in order to design a poster (work task), it might be necessary to find facts (information task) on the location and venue for the dance and to learn (information task) to use a graphic design software application. These information tasks are motivated by and nested in the broader work and/or leisure tasks. Together they provide a task-framework for information behavior.

The value of focusing on such a task framework is that it provides a means of distinguishing patterns of information behavior at an intermediate level instead of generalizing at the level of the entire population or trying to tease out differences among individuals. One of the main problems with search engines is that they are designed based on the assumption that everyone who submits the query “border collie” wants to see the same set of results (in fact, they assume that we all want to see the Wikipedia page of that name). The other extreme is personalization systems, designed on the assumption that each individual has unique and abiding personal preferences for “border collie” documents, based on his or her interests and abilities. The former approach is excessively generic and the latter is likely to require more intrusive personal data collection than most of us would be willing to allow. A middle ground is to assume that searchers’ preferences vary according to the task situation. A set of common tasks associated with the query “border collie” might include purchasing, caring for and learning about, and different sets of results would be best suited to each. So the task framework of a search can serve as a simple model of the searcher’s situation, which takes us part way in coming to grips with situational relevance.

Genre as a Situational Model for Documents

But how can we determine which results are best suited to these task scenarios? This is where genre steps up to the plate. Genre variation is a prominent feature of most large, mixed-content digital document collections, such as the Internet, organizational information spaces and digital libraries. Although the concept of genre has been around for a very long time, current genre research focusing on these large digital collections emphasizes the social and communicative functions of genre, defined as “typified communicative actions characterized by similar substance and form and taken in response to recurrent situations” [2, p. 299]. Genres emerge within communities of authors and readers as rhetorical devices to facilitate and enrich text-based communication. Because genres arise out of recurring situations, they carry the stamp of those situations and convey additional contextual information to the reader.
Consider the executive summary as a genre. It emerged within organizational settings to meet the need to summarize and highlight the major points of much longer reports. Without even looking at it, the reader knows a number of things about the document—why it was written, the audience for whom it was intended, that it is part of a larger document, that it will be relatively easy to read and brief and, importantly, that it presents a carefully constructed scale version of the truth. So genres serve as cues, informing a familiar reader about the situation in which a document was created, the intent(s) of the author, the structure of the document, the style of language and the uses to which it is suited. Within any given community or domain, repertoires of genres emerge in response to common needs and situations, become more familiar and formalized through use and may evolve or die out in response to changing needs. So like task, genre serves as a situational model. It offers a means of subdividing large sets of documents about a given topic into intermediate groups based on genre types, which vary as to their fit with a given situation.

The Relationship of Task and Genre

The next step is to determine in what sense task and genre “overlap.” Of course, there is a certain obvious connection between them, which is clear from some of the examples above. Who wouldn’t prefer a nice, clear set of step-by-step instructions to a 250-page user manual when trying to install a new piece of software? However, to move beyond these obvious individual cases and consider applying this idea to system design requires a deeper look at the nature of task-genre associations.

Our studies of task-genre associations in a software engineering work environment provided evidence that task and genre are, indeed, related. One study examined a database of shared documents, which had been indexed by the employees using task and genre descriptors. We found that the descriptors overlapped in non-random patterns throughout the dataset. For example, documents with the descriptor “cookbook” (step-by-step instructions) also tended to be assigned the descriptor “software installation” more often than other task descriptors. Another study asked participants to assign usefulness scores to 16 genres with respect to five information tasks. Again, scores varied significantly across tasks and genres, and it was interesting to see that for each task type, a different genre received the highest average usefulness score, suggesting a strong alignment (see Table 1).

<table>
<thead>
<tr>
<th>Information Task Types</th>
<th>Genre Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>fact-finding task</td>
<td>product documentation</td>
</tr>
<tr>
<td>decision making task</td>
<td>best practices</td>
</tr>
<tr>
<td>procedural (how-to) task</td>
<td>cookbooks</td>
</tr>
<tr>
<td>problem solving task</td>
<td>technotes</td>
</tr>
<tr>
<td>learning task</td>
<td>manuals</td>
</tr>
</tbody>
</table>

Based on these studies, we were able to identify some patterns in task-genre associations:

- Genre repertoires of varying sizes are associated with specific work and information tasks and, at a broader level, with various professional roles performed in the workplace.
- Task and genre are associated on the basis of the level of detail in the work and the information content. High level, conceptual tasks, such as design or planning, are associated with genre types with a broader scope and more abstract content; the opposite is true for low-level technical tasks.
- There are varying degrees of strength and scope in the associations of a given genre with workplace tasks. For example, software manuals were found to be useful in almost all task situations, while demos were associated with a very limited number of tasks.
- Some genres are more strongly associated with information tasks and others with work tasks. Some were not strongly associated with any tasks in this setting.
- Some genres, such as design patterns, clearly emerged to support a particular work task (software architecture), while other genres were co-opted from other contexts and used in task situations for which they were not originally intended.
These observations all point to the complex, dynamic and somewhat messy nature of the task-genre relationship. Therein lies its strength, as an organic expression of real-world information practices; and therein lies its weakness, as finding a way to harness this relationship is a challenging proposition.

Making Use of Task-Genre Relationships

Returning to situational relevance, it seems evident that an association exists between a searcher’s task frameworks and document genre types that represent some measure of situational relevance. However, there is still considerable work to be done before this approach can be implemented in search systems, not least of which is to develop more robust methods for automatic genre classification. We also need to find a scalable method of measuring strengths of associations between tasks and genre types in a given domain. These tasks are not trivial, but as the articles in this special issue indicate, considerable research attention is being turned toward genre, and the same is true of task. In fact, there are a number of reasons why I am more optimistic about the role that genre can play in information retrieval than some of my fellow authors.

Genre can be used explicitly in a search system by allowing searchers to either pre-select genre limits when they issue queries or post-select genre categories to limit or cluster their search results. Either method requires that the searcher be familiar with the genre types in the collection, be cognizant of which genre will best serve his or her purpose and be willing to spend the time to provide this input. It also requires that the collection be accurately classified by genre and that genre type labels be familiar and recognizable to users. I am not confident that these conditions can ever be met satisfactorily, and as Dillon notes in his article in this section, the benefits of such a system over the status quo are not likely to be substantial.

However, as I have suggested here, genre can also be used implicitly, through its association with task, to influence the ranking of search results based on situational relevance. This approach requires that the searcher’s task be specified or inferred, but does not require that the searcher be familiar with specific genres or labels. In this type of system, genre classification can be fuzzy and approximate, since genre only serves as one of many sources of evidence used to rank documents, rather than as a display or filtering feature that requires a high level of accuracy. I believe that this type of search system is not unobtainable, and given that we have already built a prototype using this approach, it may not be too far over the horizon.

Resources Mentioned in the Article

Practical Aspects of Automatic Genre Classification

by Christoph Ringlstetter and Andrea Stubbe

In the field of automatic text processing the technical term genre refers to the partition of documents into classes of documents with similar function and form. Genre represents an independent dimension, ideally orthogonal to topic. Traditionally, most work in the area of text classification from a practical as well as from a theoretical perspective has focused on the problem of how to recognize thematic domains. However, given a user’s information need, even prior to content, the genre of a document leads to a first coarse binary classification of the recall space into immediately rejected documents and those that require further processing.

Depending on the information task at hand, each genre can represent a class of documents that should be filtered. For example, cooking recipes represent a kind of “noise” if someone needs to find articles about the economic outlook on fish breeding; a person might be interested only in prose about the Spanish Civil War, another only in military documents.

In cases like these, a genre-triggered search can deliver significantly higher precision than a simple keyword search. If the documents are not tagged initially and the document base is too big for manual annotation, we need an automatic classification system.

Schema of Genres

The concept of classifying documents into different genres is based on an explicit schema of genres: a hierarchical or flat organized list of labels for the genres of a certain domain, short descriptions and, desirably for each genre, an expressive collection of example documents.

In our opinion, a general schema is not practical. The schema depends on both the document repository and the information task. For example, a retrieval system for a digital library will involve other genres than a general search engine for the Internet. Different granularity levels of a schema have to meet the requirements of different application scenarios. While disputed for more general domains, schemata for established document repositories such as the news domain or the documents of a certain company are rather concise. For professional domains such as the scientific area, users have internalized a catalog of genres, a schema that is rapidly learned by newcomers.

Hierarchical organized schemata, as compared to flat lists, have the advantage that the granularity of classification can be adapted to the information task and that different levels of classification errors can be distinguished. An example for a hierarchical view on document genre is given with the branches for the high-level genres journalism and literature as proposed in our genre hierarchy for web retrieval [1]:

- **journalism container**: commentary, review, marginal note, interview, portrait, news, feature, reportage
- **literature container**: poem, prose, drama

Features for Classification

Given a target schema, a kernel issue behind document classification is the selection of features from reference documents, that is, the training corpus. For the majority of applications, the selection of features is done manually. While often global feature sets are used, from a practical perspective, we propose specialized features for each genre. In an iterative
process, all training documents for a given genre are investigated to identify important characteristics and sometimes defining clues.

Many different kinds of features can be considered, including form, vocabulary and parts of speech, complex patterns and combinations of all these. Form features can be further divided into statistical clues such as average line length or number of sentences, document structure, the formatting of the text and, for web documents, HTML meta-information such as content-to-code-ratio. Vocabulary features include specialized word lists as well as dictionaries, for example, positive adjectives or the most common English words. Also multi-word lexemes, signs (emoticons) or phrases (such as “to whom it may concern” in letters) can be helpful. Patterns include more complex units such as repetitions of characters, dates or bibliographic references. Combinations of these features result in high level structures. For example, a casual style of writing can be recognized by the number of contractions (such as won’t) and the use of vague, informal and generalizing words (such as roughly) that are held in lexical background resources. The occurrence of some kind of agents can be recognized through dialog features (as only agents can speak), pronouns, names and living entities. Sometimes it is also necessary to distinguish different styles of writing or structure within genres; commentaries, for example, can either be polemic pamphlets or more objective documents, showing the pros and cons of a topic.

Classifiers

If one looks into the specialized literature on genre recognition, machine learning approaches with big global feature sets are widely proposed. Unfortunately, for these so-called supervised methods, massive annotated training data are a preliminary. If training data of that amount are available, support vector machines (SVM) are the best performing classifiers [2]. Several open source implementations can be integrated with reasonable effort into scalable systems.

When only smaller training sets are at hand, manual feature pruning helps to restrict the impact of artificial statistical correlations. In this case, simpler classifiers implemented as decision trees are competitive. The feature list for each document class can be pruned by classification performance on the training corpus. For our implementation, we evaluate candidate features for all classes of the specified schema and try to separate the training files of the chosen genre from the other files by determining thresholds that maximize precision and recall for those features and their combinations. If the use of a certain feature leads to a performance improvement, it is added; otherwise, it is discarded. This process can be automated, but for clear schemata manual pruning also leads to reasonable results. The iteration is terminated when the classifier reaches values for recall and precision of a chosen percentage on the training corpus, values that depend on the information task and on user expectations.

For either choice of classifier the system needs to be adaptive to changes in the information space and capable of exploiting available user information. Emerging new genres have to be easily integratable into the classification architecture. Furthermore, changes in the gestalt of established genres have to be acknowledged (genre shift).

In addition to these dynamic elements of the genre palette itself, available user data should be exploited to improve classification performance. We proposed a learning algorithm that employs user behavior in a feedback loop to improve classification performance [3]. Several levels of cooperativeness were distinguished and led to different perspectives in the utilization of available user data. We developed a model that also can be used for the case of a silent interface to retrieve data for classifier improvement from navigation behavior on the retrieved documents.

With regard to the implementation, many different packages are available to serve as the core of a genre classification system. The WEKA package is a well documented JAVA implementation of the main machine learning algorithms. It is available at www.cs.waikato.ac.nz/~ml/weka. Thorsten Joachims provides highly efficient C implementations of different SVM classifiers that have been used for genre classification: http://svmlight.joachims.org/. An experimental implementation of genre specific classifiers, including feature sets and example documents, can be found at www.cis.uni-muenchen.de/~andrea/genre/. Sven Meyer zu Eissen will provide a Firefox add-on at www.uni-weimar.de/cms/medien/webis/research/projects/wega.html that will enable a broader community to test genre qualified web search.
Performance

A still critical issue of automatic genre recognition is performance. It is reasonable to suppose values for precision of up to 75% (precision: correctly classified documents as compared to all classified documents) at 50% recall (recall: classified documents as compared to all relevant documents in a corpus). Note that these values vary considerably between certain genres but can be used as a clue to decide whether an application can benefit from automatic genre classification. As always, a dilemma between recall and precision exists: achieving a higher number of correctly classified documents has to be paid for with lower precision.

In experiments with our simple classifiers, on a corpus with 1,280 example documents organized in 32 different genres, we reached a precision for the classification into original classes of 72% with an overall recall of 54%. As mentioned, the prediction quality differs considerably between certain genres. In our case ranging from an F1 value of 14.7% for marginal notes, 81% for FAQs and 100% to empty web documents (F1: a measure that is used to set recall and precision into proportion). Genres with a definite structural appearance, such as directories, poems, FAQs and forums, involve certain form features and because of these features are much better recognized than average.

A problem of measuring performance arises because many documents do not belong clearly to only one class. When we consider documents as correctly classified that did not end up in their original, intended class but in a class that would also be well-justified if multiple classes were allowed, the precision for our experiments rose to an average of over 80%. Depending on the task and on user acceptance a document may be suitable for an alternative class if it either is a mixture of genres (like a presentation in form of a timeline) or if it contains a certain amount of material that belongs to a different genre. For example a scientific report with a great deal of statistical information might be classified as statistics or a presentation with a great amount of programming code might be classified in that latter category. These few examples already shed light on the problems with evaluation statistics.

Comparing our own results to previously published work, the small size of our training corpora and the high number of possible classes should be emphasized. In one study [4] that uses a training corpus with 10,000 documents and only seven genres, an F1 value of 89% was reached, which sharply decreased with the reduction of training documents. In another study [5] a Bayesian classifier was used to classify documents into nine of the genre classes represented in the Brown Corpus. A recall of 58% and a precision of 62% were reported. Karlgren and Cutting document [6] the influence of the number of genres on classification quality, with a decline from 73% precision using four different genres to 52% when they used all 15 Brown Corpus genre categories.

In summary, we can state that for the moment a recall of 50% and a precision of 75% seem to be realistic if one classifies over a standard web genre set. If only specific, well-structured genres have to be recognized, these numbers improve dramatically. The same is true if only a few clearly separable genres form the document base. Whether genre classification is effective in practice depends on the task and on the users’ openness to advanced search technology in general.

Practical Application

Whether the proposed application is web search or access to the internal documents of an organization, the usual interface has to be enhanced to give the user the possibility to restrict his document search on certain genres. At the same time the additional information on the results has to be appropriately communicated to the user.

A genre attribute could be introduced as an additional optional criterion for experienced searchers, analogous to the filetype attribute most of the current search engines provide. Another possibility that was proposed in several prototype implementations is a navigation tree that visualizes the underlying genre schema. As for the result of the document search, the genre of a document could be communicated with a genre marker in the heading of the snippet text. To enable an explicit feedback functionality, the result page has to be extended – for example with radio-boxes where the user can provide input on the genre of a presented document. This feature is used to collect data for evaluation statistics or the incremental improvement of classifiers.
Many variants of the sketched interface are conceivable with a completely silent interface as an extreme minimum in the spectrum of interaction. This is an issue especially if more complex search tasks have to be carried out. Since “most users are reluctant to do additional work” [7, p.469] for web search the most realistic variant is the silent interface that minimizes the cognitive load of the user. Desired genres have then to be deduced from the query combined with locally or globally aggregated knowledge about the user. The feedback of the user is derived from his observable navigation on the result set.

The output functionality of genre-qualified information access should be adapted to the task and the user expectations. Thinking of a standard search engine interface, the deletion of documents not falling into the target class is a much stronger choice than a simple re-ranking algorithm. If only applied within the chunks of the standard output such as the pockets of 10 documents, even with a precision of only 50%, a subjective improvement of the search experience is reachable.

**Conclusion**

Practitioners should be aware that genre recognition for the foreseeable future will be error-prone. Depending on the informational environment, users tend to be differently indulgent toward false positives or wrongly discarded documents. After the implementation of a genre-based retrieval system, a thorough evaluation on an independently constructed evaluation corpus should be conducted to measure system performance. The final decision for or against the launch of such a system can only be made with respect to the daily information need of the users. The initiatives within the information science community for a broadly acknowledged schema of document genres and a serious test suite for automatic genre recognition, guided, for example, by the work of Marina Santini, will lead the way to substantial progress in the next years.

**Resources Mentioned in the Article**

A couple of years ago I taught a master’s course in knowledge organization. In the Danish context, students in a master’s program work full-time and study on a half-time basis. Most of the students had educational backgrounds as librarians, and all of them held jobs that had to do with knowledge organization work. On a particular day the topic was genre theory. At the outset I asked the students what they were working with in their jobs. The orchestrated reply was, “We work with information!”

I asked them, “Information – just like that?”
“No,” they said. “We work with various forms of information.”
“Do you mean various forms of texts, I asked?”
“Yes,” they replied. “Now you got it! We work with annual reports, schedules, records, project plans, strategy reports, memos, agendas and other sorts of texts performing certain actions in our work places.”

I replied that these forms of texts are what I would call genres, and that I had no idea why they call them information. The point is that the particular workplaces where they work will probably not refer to all these forms of texts as “information,” but rather by their names (annual report, projects plans). This exercise is a first step in making people think in genre terms, to call texts and other forms of communication (genres) by the name they are given by the users of these texts. The next step is to begin to see what sorts of actions these genres accomplish on behalf of their users in a given setting – that is, to see genres as means to an end. Thus, genres are forms of communication used by people in particular contexts and situations to accomplish something particular in the activities with which these people are involved.

How do we frame genre in the context of library and information studies (LIS)? Let us assume that LIS as an academic field is the study of (1) how and with what means professional, scholarly, cultural and social knowledge as materialized in documents (print or electronic) is communicated in society and (2) what function libraries and other similar knowledge organizing institutions or activities have, or are supposed to have, in these communications. Professional, scholarly, cultural, organizational and social communication accomplished by documents comprises a diversity of genres: work orders, newsletters, recipes, bibliographies, articles in literary-cultural journals, surveys, chronicles, technical reports, governmental reports, scholarly articles, book reviews and so on. Many people educated with a background in LIS probably work in some way with such genres.

As LIS is interested in how knowledge in documents and other artifacts is organized, genre theory is a productive perspective. Studying genres reveals that genres and human activity are important organizing factors of communication and knowledge. It tells us what kinds of genres various people and institutions in various communicative contexts prefer. A genre view of these communicative activities provides a means to systemically examine document production and use and the organization of document production and use. Thus, using genre as an analytical concept in LIS would help us understand how professional, cultural and social communication is carried out by a diversity of people via genres. It would also enable understanding of how genres organize activity, texts, knowledge and people.

Moreover, a genre perspective would not only focus on the people and the production and use of various documents, but would also stress the typified activities, the work, in which people are discursively engaged. From a genre point of view, then, document production and use is not viewed as an end in itself but in terms of how documents help people do their work. If
LIS by means of theoretical and empirical studies can gain a more thorough understanding of situated and typified document production and use, then we may also be able to have a more informed understanding of the ways in which information systems help or do not help people do their work. This goal implies that we must view information systems not as an end in themselves but rather as a means to an end. We must focus on what information systems do, as tools to be used in goal-directed activity. It may be that such a view has been present in LIS but formulated in other ways. However, what has been missing is the genre perspective.

But why should LIS want to know about genre? A genre-theoretical approach to LIS research problems provides an opportunity to connect to other theoretical discourses, in particular in the humanities. Moreover, genre can provide LIS with a concrete object of study. In fact, Phil Agre has suggested that genre, rather than information, be used as a focal point in LIS [1]. Genre theory is not about text types in isolation, but rather about the fact that recognizing (as both producer and user) a particular text type means recognizing a particular communicative situation and activity in which that type of text (genre) is used to accomplish a given task. Bazerman [2, p. 16] puts it this way:

Genres help us navigate the complex worlds of written communication and symbolic activity, because in recognizing a text type we recognize many things about the institutional and social setting, the activities being proposed, the roles available to writer and reader, the motives, ideas, ideology and expected content of the document, and where this all might fit in our life.

Take the example of the digital library. The digital library is a textual place on the Internet which helps us “…navigate the complex worlds of written communication and symbolic activity…” When we recognize a given digital library as such, we also know what to expect of it, that is, what it can do for users and what users can accomplish by using a digital library, as compared to other textual places on the Internet that perform similar actions. A genre understanding of digital libraries thus provides a means of understanding matters of knowledge organization, communication of symbolic activity and information seeking. Generally, genre broadens our understanding of how knowledge is organized and communicated by means of recorded discourse as articulated through some generic form.

What is most important to know about genre is that, first and foremost, genre studies are not primarily concerned with text types, but with how different human activities involving the use of texts become typified as a consequence of the production and use of recorded discourse by different agents. Typified human activities reveal what kinds of paths and access to knowledge are considered legitimate or appropriate in particular contexts, and thus also what kinds of information seeking strategies are employed and why they are employed the way they are. By taking up a genre approach, LIS researchers and practitioners are capable of producing an understanding of the institutional and social setting in which the communication and organization of knowledge, culture or information take place. It can also lead to an understanding of why this activity has the shape it has and what impact it may have on human activity.

Many user studies have been carried out over the years trying to determine how people use the kinds of information and information systems they do. But rarely has emphasis been put on the artifacts that materialize and configure information, and the effect of artifacts on the use of information (see [3]). Although genre is much more than text type, the text as a social and material object is nevertheless important for theoretical and methodological reasons. Therefore, a genre approach to LIS would imply a greater focus on the study of texts and on how and why they accomplish what they do on behalf of the people producing and using texts in human activities. In short, user studies would be genre studies. Traditional user studies have focused on how people seek information and what sources they use and do not use. Important insights have been gained here. But the starting point has been people’s information need and consequently how they have tried to fulfill this need. Genre studies would start with the genre and then move backwards to ask questions like, how has this particular text come to look and be used like it has and by whom? To which actions, or goals, is the text intended to contribute? Who is involved in producing and using this text? What kind of larger textual and activity system is the text a part of?
Thus, to study a genre is to study how knowledge is regulated, codified and altered by people and their communicative activities. One might even say that a particular genre defines an information need. For instance, a fishing guide cannot fulfill the need for information concerning how to prepare a sushi dinner. The fishing guide and its informative potential is not unlimited. This limitation exists because a genre, and the form of knowledge it materializes, is defined and constrained by the activity it is used to accomplish.

A major issue in genre theory is how to recognize and understand various types of texts and what they accomplish in different human activities based on the production and use of texts. In this way genre theory is concerned with much more than mere text types and their formal textual features. An approach that only studies text types and their formal qualities does not allow for inquiry into the discursive activities performed by texts.

Generally speaking, the concept of genre covers the characteristics that differentiate texts (verbal or written) from each other. But this differentiation is not a matter of recognizing purely textual and formal features. To recognize a particular text type is to recognize a particular communicative situation and activity in which that type of text (genre) is used to accomplish a given task. In our everyday interaction with texts as producers or recipients, genres are means of orientation. Our knowledge of genres determines the means and modes of the production and use of texts and what we expect of them in professional as well as everyday life. Genre becomes a kind of textual existentialism and hence connected to literacy, as the more we know about the communicative activities we are involved in, the more we also know how to understand and use the texts produced by these activities. This also explains the view on texts and contexts in genre theory. Texts and contexts are not considered two distinct categories but the text is integral to both context and action. To produce and use a text is to be situated in a context with socially and historically developed typified activities of producing and reading texts.

For example, a scientist cannot just invent his own way of reporting experiments to colleagues, if he wants to be taken seriously. He has to know how to apply some conventions for presenting the experiment setup, evaluation methods, results, etc. On the other hand, the reader of a scientific article has to know something about the tradition of reporting such experiments in order to understand and evaluate the contents of an experimental article. This was what Bazerman [4] showed in his study of the experimental article in science. Bazerman paid particular attention to how this genre emerged historically and how it was, and is, shaped by the recurrent typified communicative activities of writers (scientists reporting their experiments) and readers (evaluating and learning from the experiments of others) and how the genre shaped the knowledge producing activity into a typified activity (the writing and reading of an experimental article) as a product of its history and agents. Thus Bazerman’s study demonstrates how we may conceive of genre as “typified rhetorical actions based in recurrent situations” [5].

In connection with genre and written communication the concept of typification becomes even more critical. Given that writers and readers are separated in time and space, a means is needed to communicate appropriately and avoid, or minimize the risk of, misunderstandings. In a typified communicative activity certain actions are carried out in certain situations following certain forms of communication, leading to the ability to recognize and understand particular standardized practices and activities [6]. Thus, typification is what allows one to recognize and identify a particular context and its particular forms of communication, i.e., its genres.

End Remarks
Genre makes us see how communication and knowledge are structured, stabilized and codified by people, situations and activities. To observe genres in action is to observe the work they are performing on behalf of their users. Thus, we gain an understanding of how and why certain people are using certain forms of communication in certain recurrent situations. We learn to see how it is that practices become stabilized, i.e., typified. More importantly, we learn to see the knowledge work performed by librarians and other information professionals as part of a larger activity in which knowledge is produced circulated, stored and used.

Authors Note: This article is a heavily condensed version of my recent article for the Annual Review of Information Science and Technology [7].
Resources Mentioned in the Article


Research Interviews for Library and Information Professionals

by Andrew P. Carlin

The open-ended interview, as a research method, is a highly productive tool for eliciting information. Unfortunately, there is very little guidance for information professionals on its actual use. This article looks at some decisions that will affect the progress and outcome of interviews. As you’ll never be able to recreate a particular conversation again, the objective of this article is to anticipate problems so that your interviewing experiences go more smoothly. While these notes are tailored toward research interviews, some are transferable to product design and user evaluation sessions also.

Librarians as Researchers

Increasingly, librarians are being encouraged to engage in research [1]. The issue of whether librarians should engage in research isn’t new [2]; nor is it a straightforward development of the librarian’s status [3, 4].

There are important reasons why librarians should be competent in doing research: to conduct research for themselves for their own professional development; to conduct research on behalf of their employers or parent organizations; to justify their roles and demonstrate the value of information specialists to the organization; and to be able to appraise research in order to apply evidence-based library research to their own practice.

Before embarking upon a research project, you need to consider carefully the research question: What is the problem to which you need answers? Then consider what the most appropriate method would be in order to find these answers. Such methods may include using direct observations, questionnaires, focus groups and structured or unstructured interviews. Using unstructured or open-ended interviews as a research method does not mean performing “soft” research since there is now recognition that the aims of evidence-based practice do not disqualify insights from “qualitative” research [5].

While librarians can build upon their expertise in teasing out information from users in reference interviews – where librarians often has to present users with alternatives to discover what they are looking for – the research interview requires the librarian to resist putting words into the user’s mouth. Instead, the librarian is seeking the user’s understanding of a topic, rather than providing candidate lines of inquiry for the user to choose among topics.

There are lots of resources that discuss interviewing and developing rapport with your interviewee and that help with the analysis of “qualitative data” [6, 7, 8], but not so much on how to get started. A research primer for librarians [9] gives little practical detail on how to do interviews. For the moment, I shall assume that you know the research question; you know what you are going to ask (an interview schedule); and you have identified, approached and made appointments with people to interview. Beyond these steps, I’ll give some practical tips on doing interviews.

Recording

Before embarking on your interviews, remember that the decisions you make at the outset will influence, to a great extent, what sort of data you can collect and what you are able to do with the data [10]. With this in mind, you are in a stronger position to select different options for data analysis. For example, “Do I record the interview or take notes?” “What do I do with the recording?”

Andrew P. Carlin is a librarian at St Columb’s College in Derry, Northern Ireland. He can be reached by email at acarlin865<at>stcolumbs.com.
I recommend that you record your interviews. A recording is much more reliable than your memory of what an interviewee said, is more accurate than notes taken during the interview and may contain details that you will regard as important later (indeed, you may not even have noticed at the time). There are ethical implications of recording interviews, however. Ask permission before switching the recorder on. Tell interviewees the purpose of the interviews and the use envisaged for your project. Ensure that you give a thorough commitment to confidentiality – that everything they say is confidential, that you are the only person who will have access to the tape, that they will not be identifiable from any information they provide. If permission to record is granted, record your opening statement and promise of confidentiality as part of the interview. If you don’t have permission to record the interview, don’t record it. If you do have permission, remind the interviewee that it will be recorded before commencing the interview.

Once you have a finished recording, you will find yourself asking, “What do I need to transcribe?” Do you transcribe only the interviewee’s talk, only the talk that pertains to your project or a few sound bites? Transcribe everything, including your questions and prompts. How – and how much – you transcribe will affect how you analyze the interview and therefore affect the final analysis. If you only transcribe selectively, this will prevent you from considering the interviewee’s understandings in entirety. A full transcription also allows you to judge how much you have influenced an interviewee’s responses.

Obviously, you don’t want to listen back to the recording and realize you’ve done most of the talking. Try to minimize the amount you talk. Knowing that you will be transcribing all of the talk will help you limit the amount of interruptions you make; use head nods rather than continually saying “Mm” or “Uh-huh” during answers.

Test Your Equipment

Before you set off for the interview, make sure your recording machine is working properly, that its batteries won’t go dead. Carry spare batteries with you just in case. Also, make sure you have spare tapes or memory cards. Take the plastic wrapper off any spare media you have; you don’t want to have to fiddle with packaging during an interview, nor (worse) have to pause before continuing the interview, nor (worse still) miss anything your interviewee says. What they have to say, in their own words, is paramount.

Position the microphone carefully so that it picks up both your interviewee’s and your own voice. This skill requires practicing with your recording equipment, to establish the optimum distance according to background noise. Be prepared to move the microphone during the interview: you can’t assume that because you have arranged interviews by appointment that they will be conducted in a quiet room away from all the noise. If your interest is software or system development, consider using a video camera also, positioned to capture user-routines and screen work [11], which can be matched with the audio files.

Be prepared to go on walkabout. Some of my most arresting interview experiences have been peripatetic. The shame about these impromptu walking tours was not having a video camera to record them; I only had an audio device. So you can only listen to someone saying “Look at that,” etc. Note-taking is problematic when on the move, too! Just make sure you keep the microphone near the interviewee (it doesn’t have to be too close – testing your equipment beforehand is vital) and keep a commentary of what’s going on. This commentary is part of the interview: “Where are we now?” “What’s going on here?” “Tell me what I’m looking at.”

Treat the interview as a conversation rather than forcing a question-answer format throughout. Indeed, not everything you say has to be a question (“Tell me about that”); nor do all your questions have to be too specific. This restraint helps avoid introducing bias into questions or asking leading questions [12]. Try asking something like “So what’s going on here?” Your interviewees are then in a position to interpret what you were asking about, rather than you determining what sort of answer they provide. You don’t have to ask too many questions anyway since people are only too glad to have the opportunity to talk and be listened to.

While a certain level of competence is required to talk with software designers and systems engineers, don’t be afraid to ask what you might perceive as stupid questions. Stupid questions make sure you get details that are taken for granted and sometimes elicit valuable information that might have been missed otherwise. Saying “You’ll have to explain that for me” is
not playing dumb; it’s getting the interviewee to tell you what’s going on. At any rate, your interviewees are your local experts on the matter you are researching, and you want to hear their stories.

During the Interview – Listening and Note-Taking

Listen carefully to what your interviewees say, and show them that you’re listening attentively. Even when you are recording the interviews, make sure you take notes. When talking with your interviewees for 30 minutes (then turning the tape over hurriedly) you can get lulled into a false sense of security by the recording machine. However, the machine might not record everything, or it might have stopped working. So you must take notes during the interviews, but not so that note-taking interferes with listening or with the interview itself.

Make a note of anything you want to return to at the end of the interview: rather than interrupting the flow you can ask for clarification later. If you have a loose-leaf notebook, select one sheet (the top one is most useful) and write “notes-to-self” on it. This avoids you having lots of points needing clarification that you can’t find, as they’re all collected together on one page.

Whatever your interviewee has to say is relevant and is of interest. Frequently, you will hear interviewees say phrases such as “I don’t know if this is relevant, but…” or “You’re probably not interested in how […] works.” Reassure them that everything they say is important to you. It is.

At the End of the Interview

After you’ve clarified any loose ends, ask the interviewees if they have any questions about your project. Answer these queries as best you can. Thank interviewees for their time. Be fulsome in acknowledging how helpful their insights will be to your research. Offer to send a copy of the recording you’ve made. If this offer is taken up, do send one.

After the Interview

As soon as you get back to your workstation, you must make a backup of the recording – do so before listening to the recording, before transcribing or before any analysis. Work from the copy, not the original [10].

Write an appraisal of the interview while it is still fresh. What were the most surprising points to come out of the interview? Did anything go particularly well or particularly badly? Was there anything you would do differently or wouldn’t do again? Did the interviewee draw your attention to anything you hadn’t considered before and that you could introduce as a topic in future interviews? Detailing your own post-interview impressions will benefit the corpus of interviews for your current project and assist you in developing your interview skills for future projects.

Always remember that regardless of how careful you are with your interview schedule, every interview is different. In light of your interviews, you may find that you need to revise your research question, or you may discover that there is a more important topic for research than the one you set out to investigate.

---

Resources Cited in the Article

You’d think that I would know better after more than a decade working at the intersection of information management/strategy and user experience. But I’m still surprised that I’m surprised when I hear of an executive dismissing an information architecture (IA) study out of hand. Every day we design and organize information systems tailored to particular user groups and needs. So while we practice the principle of fitting the solution to the customer, I wonder if sometimes we inadvertently stumble in getting the executive recognition of the value of IA and how it is a core component of customer experience. In this column, I’d like to propose an approach which has been shaped in consultation with peers in the strategy, customer relationship management, IA and user experience communities. Let’s see how systematically using real-world customer data can support our findings, introduce new issues requiring attention and provide a built-in business case to invest in the skills and solutions information professionals can bring to bear.

Speaking to Executives and Other Challenges

Many have written (better than I) about the disruptive technologies and changes the IA domain faces. Search as the primary means of wayfinding; social networks and trust economies; user-generated tags and folksonomies; advances in computer-based classification; entity extraction and semantic technologies – all, depending on one’s point-of-view, threaten our value and existence or offer new tools and domains for optimized information architecture. As if that challenge weren’t enough, practitioners across the user/info-centered continuum (human-computer interaction, information architecture, interaction design, new product design, technical communication, user experience and so forth) face a constant battle in translating our findings and recommendations into a form that resonates with finance-focused executives. To the extent that good executives are, among other things, good at assessing the cost/benefit tradeoffs of alternatives and allocating resources based on real-world data, we need to speak their language if we (and the customers we speak for) want a fair outcome. To that end, I propose we adopt information advantage in our information architecture efforts.

Information Advantage Defined

Recognizing that the term information advantage has specific meanings in the defense and business strategy communities, I propose a non-bellicose, IA-specific definition: We achieve information advantage when we optimally and systematically use real-world customer data to architect information solutions that yield the desired customer experience.
At its heart, information advantage provides a way to manage and improve customer experiences with IA products such as catalogs, websites or software. It provides a structured methodology for identifying, synthesizing and acting on data that drives the success (or failure) of those customer interactions. Information advantage synthesizes existing and new customer data to

- discover and quantify trends, issues and things requiring further attention;
- provide evidence for/against existing findings (for instance, small-sample studies); and
- provide real world, quantitative justification for action (business case).

How to Embrace Information Advantage

To flesh out the anecdote from the introductory paragraph, I was surprised that an executive didn’t get the importance of updating the company’s public website based on findings from testing the site IA with a group of users. While I consider it significant that 10 of the 12 participants had a problem with the breadcrumbs, the executive simply couldn’t extrapolate (or believe the finding could be properly extrapolated) to their customer base and bottom line. Yet later, when the executive was given data indicating there were thousands of contacts with their contact center on this same issue, the message hit home. Multiply the actual number of customers experiencing an issue with the associated cost(s), and your justification for additional study or remediation is right there in dollars and cents. We thus supplement existing tools and techniques (and previous findings) with volumes of data that support them and/or justify particular action.

The 50,000-foot view of how to approach this customer information synthesis looks like this:

1. Identify and discover potential sources of feedback
2. Obtain and clean up data
3. Aggregate and synthesize data
4. Identify and prioritize trends/issues
5. Develop and execute remediation plan

This list is all fine and good, but it doesn’t appear to differ much from other methodologies (at least it shares face validity in that sense). That’s why we’ll need to address specific tools and techniques to help us get there. First let’s consider what kinds of information should be leveraged.

Sources of Data Driving Information Advantage

Chances are, if you support e-{something}, you’re already using data from web analytics and/or search engine logs to help you refine the user experience with your site. But just as customers have no qualms about jumping from channel to channel to suit their particular task needs and interaction preferences, they increasingly look to (and contribute to) information outside the control of your company when considering how or whether to do business with you. To be effective in using the increasingly user-generated content to achieve information advantage, we must look beyond traditional sources and methods of collecting customer feedback and use relevant information from sources such as the following:

- Records from customer contact centers and help desks (calls/emails/chats)
- Solutions that monitor forums, blogs, wikis and similar sites for product mentions and sentiment
- Social network/tag cloud analysis
- Satisfaction surveys (especially open-ended questions that can be mined/classified)
- Bug reports and feature requests
Automated web quality assurance solutions
New and existing IA/usability/customer studies
Debriefing trainers, field service, sales and other personnel who have direct face-offs with the customer
User forums (may be local or product-specific)
Customer-attended conferences
Warranty registrations and product field- (or factory-) service
Automated testing and A/B testing solutions
Web 2.0-aware market research/focus groups (txt-based panels, anyone?)
Professional (3rd party analyst) and amateur reports and reviews
Design “playgrounds” with users

For better or worse, this list is illustrative, not comprehensive. We must address the challenges of multiple sources, data formats, structures/definitions and the sheer volume of data with a robust toolkit.

Tools of Information Advantage
Our IA toolkit containing card sorts, taxonomies, prototyping, pluralistic walkthroughs and other techniques has helped and will continue to help us build and refine solutions. But increasingly, we must rely on technology to help us effectively manage technology and data. The volume of data alone – compare the few hundred terms in a website to the tens of thousands of tags the same number of pages may yield through a social bookmarking tool – requires new approaches and computer-assisted analysis. Random sampling and statistical techniques, data/text mining, automated classification/extraction and instrumentation are a few examples of how we can tie this information into our core IA work. Because this method cost-justifies itself readily, a quick and dirty manual review to start things off is typically sufficient to yield results that justify additional staff, resources and time to ramp up the scope and effectiveness of your information advantage effort.

The Power of IAs Practicing Information Advantage
Not only does a smartly done synthesis of customer feedback help identify the trends, issues and items meriting further attention, but it also provides a built-in business case for why these things need to be addressed. This business case is music to the ears of executives who only hear numbers and data. The idea of cross-domain synthesis is not new (nor do I presuppose to have perfected a comprehensive model for obtaining information advantage). I trust, though, that with our expertise and a focus on information advantage, IAs can overcome the dual challenges of adapting to disruptive technologies and securing the ear of executives who value the customer but need just a bit more data before they see the bottom line: information advantage equals success.