Rendezvous *with KnowGenesis*

By: Saurabh Kudesia

> "I'm certainly proud to see that level of readership for and support of our project. We're constantly evaluating data that help us to keep the TC Library site a valuable resource for users."

Geoffrey Sauer

Assistant Professor, Rhetoric and Professional Communication, Iowa State University
(*http://engl.iastate.edu/programs/rhetoric/*)
Director, Studio for New Media
(*http://newmedia.engl.iastate.edu/*)
Director, the EServer
(*http://eserver.org/*)
Director, the EServer Tec Comm Library
(*http://tc.eserver.org/*)

Geoffrey Sauer is an assistant professor in the Rhetoric and Professional Communication Program in the English Department at Iowa State University.

His research studies the history of publishing, particularly the recent history (for the most part, post-1979) in all variety of media (including music, film, video and software distribution, as well as traditional monographs and serials). He teaches courses in rhetorical theory, film theory, multimedia production, intellectual property and cultural theory.

He is the director of the EServer, a nonprofit online publishing venture in the arts and humanities (by some estimates, the most visited humanities website in the world). It has editors, organized into collections to publish works to approximately two million readers per month. It hosts a variety of online resources, including websites about American culture (such as Bad Subjects and Cultural Logic), literature (such as The David Mamet Review and The Thoreau Reader), art (such as Cultronix), and technical and professional communication (such as The EServer Technical Communication Library). His work with online publishing dates back to 1990, and has led him recently into the study of content management systems and the rhetorical and design theories implicit within SGML-derived publishing standards (such as XML, XSLT, XPath and XHTML).

He is also the director of the ISU Studio for New Media, an interdisciplinary research institute organized to support, further, and coordinate work with digital media currently done by individuals across multiple departments at Iowa State University.

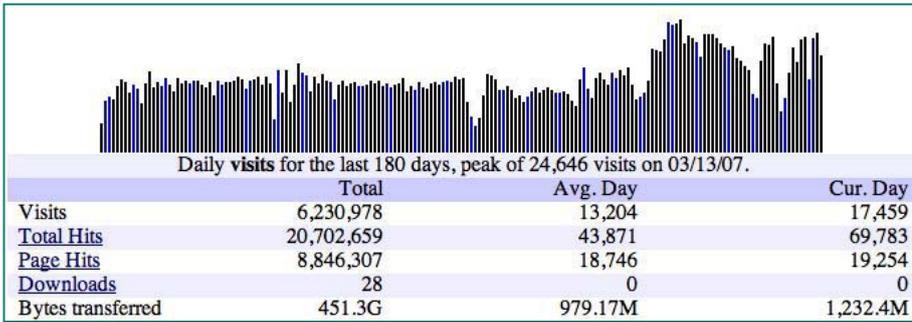You can always reach Professor Sauer at *webmaster@eserver.org*.

Saurabh: What constitutes a successful online library?

Geoffrey: I'm not sure we can generalize extensively about online libraries. I've been involved as a webmaster or consultant with dozens, including university library websites, sites run by academic presses, EServer collections, websites for both small and large Society for Technical Communication (STC) communities, as well as sites for the IEEE Professional Communication Society, the Association of Teachers of Technical Writing (ATTW), and Council for Programs in Technical and Scientific Communication (CPTSC). They've all varied significantly, in terms of their audiences, purposes, and user needs.

But I suppose I could say that a highly successful online library tends to recognize inchoate needs among a set of people (often a geographically-dispersed or heretofore ill-organized group), and produces a resource that presents a vision of the community which enables current members to recognize something about themselves—something that enables collaborative knowledge use, production, and collaboration.

It is true, however, that when you actually administer an online library, you can be more specific by finding some quantifiable measures of success in its server log data. The EServer Technical Communication Library website, for instance, serves an average of just more than 17,500 visitors per day—over 80,000 "hits" per day.

To some degree, that may be considered a form of success itself. It's difficult to compare websites' popularity against one another (few sites release their log analyses to the

Daily **visits** for the last 180 days, peak of 24,646 visits on 03/13/07.

	Total	Avg. Day	Cur. Day
Visits	6,230,978	13,204	17,459
Total Hits	20,702,659	43,871	69,783
Page Hits	8,846,307	18,746	19,254
Downloads	28	0	0
Bytes transferred	451.3G	979.17M	1,232.4M

public—see *http://tc.eserver.org/about/ recent.lasso* to see the small portion of the 150-page log analysis the TC Library releases to the public). But there are some means by which to compare the success of an online library project against its peers.

One of these is Alexa.com, the division of Amazon which measures the popularity of a large number of Internet domains (those which it finds to be among the 100,000 most popular). Alexa's comparison of websites in technical communication suggests that the TC Library site is today the most popular website in the world for technical communication issues. If you go to *http://www.alexa. com/browse?&CategoryID=2389* you can see a list of Alexa's top-ranked sites in "technical writing:"

If you go to *http://tc.eserver.org/about/ alexa.html*, you will see a chart that displays Alexa's measure of popularity of some of the most popular technical

communication websites, in terms of the most popular by page views over the past three months:
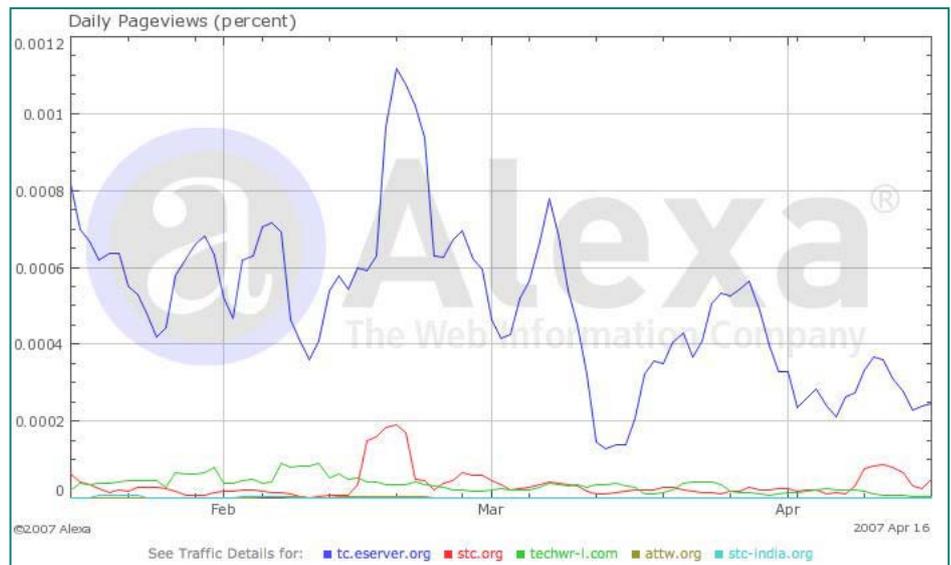
I'm certainly proud to see that level of readership for and support of our project. We're constantly evaluating data that help us to keep the TC Library site a valuable resource for users.

Saurabh: What sort of planning went into designing and implementing the EServer library? Can you share some of the lessons learned during the process?

Geoffrey: EServer.org was founded in 1990 as an online resource in the arts and humanities, originally as a Gopher, FTP, and telnet resource, and later, in 1993, as a collection of websites. Today it publishes fifty websites, including some of the most popular in drama, fiction, poetry, rhetoric, and other fields.
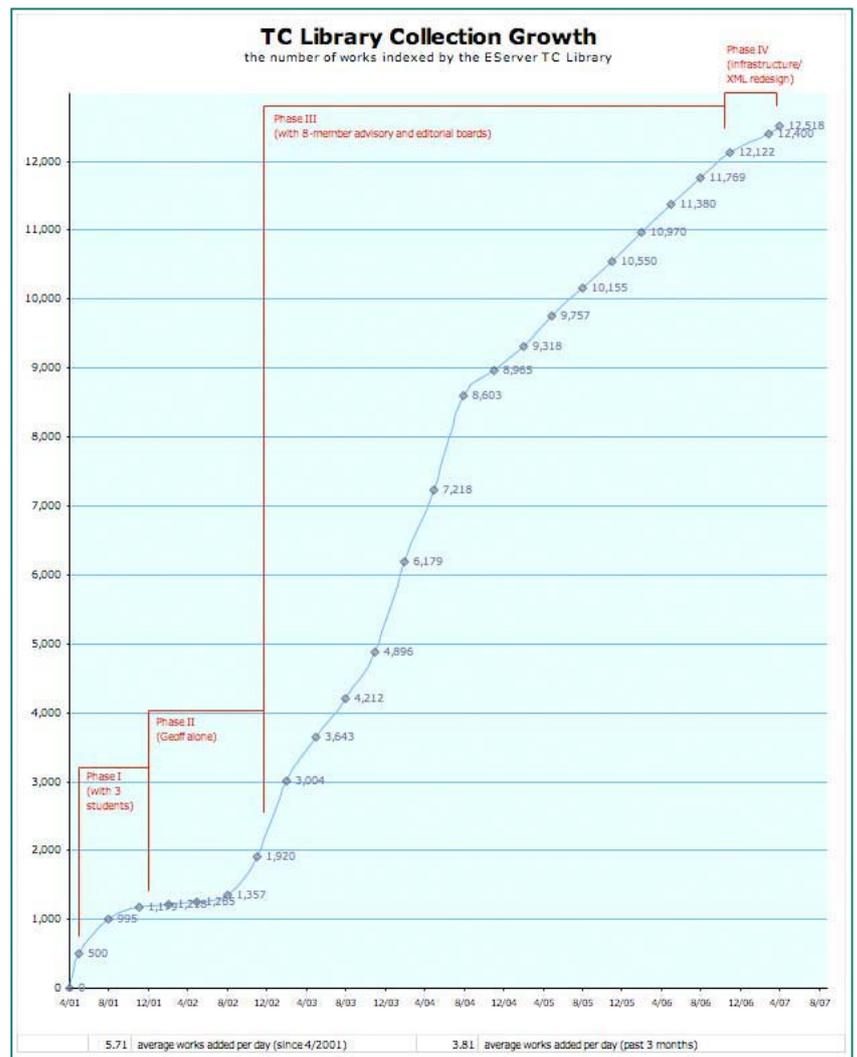
The EServer Technical Communication Library, one of these fifty EServer collections, was launched in 2001 by myself and three undergraduate students. I had taught a seminar on university research (I was the chair, at the time, of the University of Washington's University Library Committee), and the students were dismayed to find how difficult it was to research issues in technical and professional communication using traditional library indexes. They needed to search a variety of library indexes from a range of tangentially-related fields in order to find what few articles they could about technical communication issues. They also completed that assignment concerned that there was quality work in industry not included in the research they were

able to find.

So we four undertook to build a free online index of writings available directly online from peer-reviewed scholarly journals, as well as resources posted by workplace practitioners. The goal was to produce a resource that would contribute to professionals' understanding of quality work already done by others in the field, in ways that would both help practitioners benefit from scholarly research they currently didn't know about, and to help academic researchers learn to incorporate writings by workplace technical communicators about the issues currently facing our field.

By June 2002, my students had graduated, leaving the TC Library site with approximately 1,000 works in our index and twelve top-level categories.

For ten months or so, after that the site had limited growth, but in October 2002 I recruited academics and practitioners I respected to join the editorial board and the advisory board for the site. The editorial board would supervise the development of new features for the system and addition of new works to the database. The advisory board would decide issues of policy and governance for the project.

TC Library Collection Growth
the number of works indexed by the EServer TC Library

Phase IV (infrastructure/ XML redesign)

Phase III (with 8-member advisory and editorial boards)

Phase II (Geoff alone)

Phase I (with 3 students)

5.71 | average works added per day (since 4/2001)   3.81 | average works added per day (past 3 months)

The editorial board worked to develop a more usable and accessible interface to the website, a four-tier system of categories for works, a system that permitted our users to rate and review any work in our index, and a cadre of student volunteers who diligently worked to enter new works into our index.

You can see a chart of the collection growth from the founding until the present.

Since the surge of growth in the collection, our visibility has continued to increase. Today the 12,450 works indexed in our collection generate more than 60,000 web pages of information, including details about each work, as well as listings of works by category, author, publisher, year published, and a wide range of other metadata categories.

In late 2002 I was recruited by Rebecca Burnett to leave the University of Washington-Seattle and join the RPC faculty at Iowa State University. (Because EServer.org is a nonprofit corporation, it moved with me, as I am its director and the chair of its board.) In 2003 we relocated the EServer to Ames, Iowa (in STC region 4, the same region as India). The Rhetoric and Professional Communication program here at ISU has 21 tenure-line faculty in rhetoric and technical communication studies, which makes it (to my knowledge) the largest program

in our field in the world. Because of the large and diverse group and the support for technological ventures such as the TC Library, the site has prospered here.

The standard for promotion and tenure purposes (necessary to retain one's position at a research university) at most schools is "impact on the field," which most often means research publications in peer-reviewed scholarly journals which are well-regarded by reviewers. But the sophistication of the program here has accepted my work with the TC Library as offering "impact" on the field of technical communication comparable to that of traditional scholarly articles or books, which makes ISU an excellent home for me and for the site for the foreseeable future.

My specialty in information technologies in the practice of technical communication has suited the interests of some of our faculty colleagues and many graduate students, and the level of support for the project here can be seen in how the TC Library site has prospered since 2003.

At present, we are in a phase of intensive work redeveloping the XML infrastructure of the site and the database infrastructure which powers the site. When we release updates to the site in late 2007, it will be much easier to import large numbers of posts into our index, and our website interface will benefit from a more consistent XHTML and AJAX interface.

Saurabh: One of the biggest challenges in library planning is to provide users with access to information that has been evaluated, organized and preserved in the most useful format. How did EServer library, which relies on end users to submit and classify the data, is able to handle this issue?

Geoffrey: Coordinating the work of volunteers is a complex undertaking. When we first designed the TC Library, we had two choices: we could build the site with a simple set of fields (easy for novice volunteers), or with a complex, fully-articulated set (harder for volunteers, but more useful in the long term).

We looked at the Usable Web site (*http://usableweb.com/*), which was until 2003 an extremely popular online resource in the fields of usability and web design. Usable Web kept author, title, URL, and category fields for almost 1,000 works in web usability. It was therefore quick and easy for volunteers to add new works to its index, but the site was also limited in some ways as to how one might search for sites, or export the data for use in academic research or citation. Its simplicity was high, which was both an advantage and disadvantage.

We then looked at highly structured and complex

systems, such as the web interfaces to ProQuest and Project Muse. These followed scholarly archiving best practices, offering full support for library Z39.50 and MARC metadata standards, but offered forms and interfaces far more confusing than most web volunteers might be willing to use to add items to a catalogue.

Our editorial board had a subcommittee which developed an interface with a middle level of sophistication, one with sufficient features to require a bit of dedication from online volunteers, but not one with so many features that it would be prohibitively complex to use. This has worked well for us, overall.

We currently have a small committee looking at future needs for adding work that has been evaluated and reviewed prior to inclusion in our system. We have a team working on XSL transformations to permit our system to import metadata in the Journal Archiving and Interchange 1.0 DTD, which will permit us to import articles published in a number of scholarly journals directly into our database. And with the experience from this project, we'll be able to develop a system that will parse RSS 2.0 XML data for rapid (but not wholly automatic) import of feeds from journals such as the *IJTC* into our index.

Then, in addition, we're developing AJAX versions for a next-generation version of our "add a site" and "update an entry" forms. We expect those to be tested in summer 2007, for release to the public in late 2007. To some degree those will assist by allowing simple sites to be added quickly, while more complex works may use more sophisticated interfaces for things such as multiple authors and complex/

multiple categories.

Saurabh: The necessity for a more comprehensive understanding of user needs, objectives, and behavior in employing digital library systems was stressed repeatedly as a basis for designing effective systems. Do you think that enough attention was paid to these issues during the planning phase of EServer library?
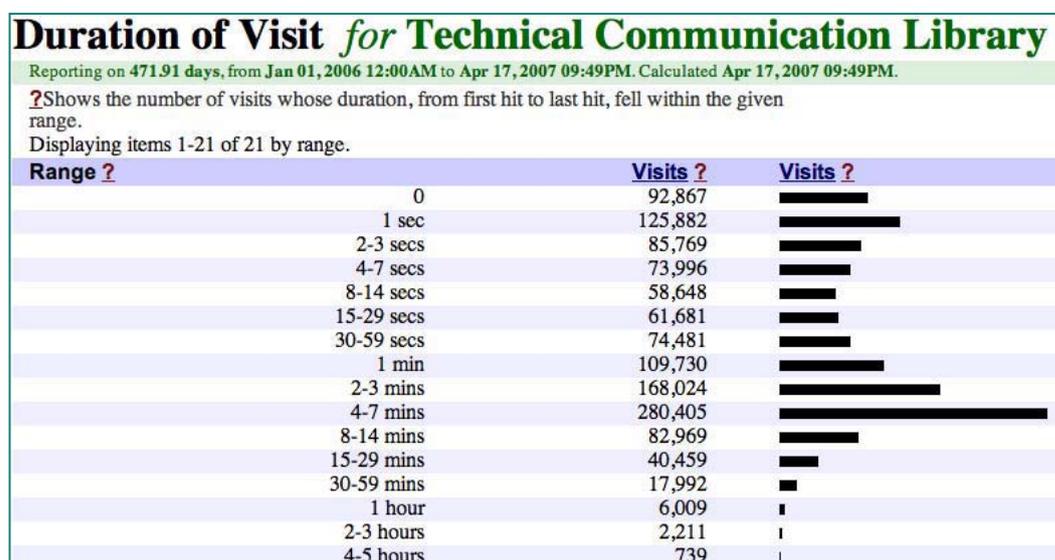
Geoffrey: Of course not. It's impossible to pay enough attention to the usability of a website. One devotes as much time as one can. Usability, user-centered design, interaction design, accessible design and experience design methods require significant resources of time and energy from contributors. The TC Library is fortunate to have editorial and advisory boards with notable experts in these methods of design, but because the site is entirely a volunteer undertaking, there are limits to the resources we can devote to the usability of our interfaces. Add to this that our readership is extremely diverse, from highly skilled academics entirely within the profession of technical communication to workplace practitioners who haven't heard the term before, but find our site from an internet search. It's difficult to model our site for traditional personas or user-centered design schemata.

We're of course always interested in improving our usability. We'd certainly welcome participation from your readers, if they're interested in undertaking usability studies or redesign efforts to help us improve our systems.

Saurabh: The use of more sophisticated models of user behavior and needs in long-term interactions with digital library systems is a potentially fruitful area for research. How do we replicate these models to match the library design close to user requirements?

Geoffrey: The log analyses generated by the EServer help us to discern interesting patterns of usage. Among these are histographs of the time spent by visitors on the site.

My colleagues on the editorial and advisory boards and I are very fond of this chart, as it implies that visitors are doing what we had hoped: browsing the TC Library site, in order to explore to find works they might otherwise not have known. Considering that more than 50% of our visitors come to us from Internet searches, we consider the number of visitors spending four to seven minutes (and other similar data) very useful in measuring user behavior.

Duration of Visit *for* Technical Communication Library

Reporting on 471.91 days, from Jan 01, 2006 12:00AM to Apr 17, 2007 09:49PM. Calculated Apr 17, 2007 09:49PM.

? Shows the number of visits whose duration, from first hit to last hit, fell within the given range.

Displaying items 1-21 of 21 by range.

Range ?	Visits ?	Visits ?
0	92,867	
1 sec	125,882	
2-3 secs	85,769	
4-7 secs	73,996	
8-14 secs	58,648	
15-29 secs	61,681	
30-59 secs	74,481	
1 min	109,730	
2-3 mins	168,024	
4-7 mins	280,405	
8-14 mins	82,969	
15-29 mins	40,459	
30-59 mins	17,992	
1 hour	6,009	
2-3 hours	2,211	
4-5 hours	739	

We then have similarly complex data available to us, based upon such information as: the number of pages browsed based upon which search phrase visitors who come to our site from specific phrases in major internet search engines; the most popular routes through our website; and specific data about patterns of site visits based on country of origin.

I presented a paper in 2005 to an Association for Business Communication panel about patterns to the search phrases in searches of the TC Library index by users in the United States, China, and India, in order to discern what those differences may tell us about the state of technical TC idioms for the industry in all three locations. The paper reviewed TC Library log data, and looked at works such as articles from *Indus* and *Intercom* in order to discuss how practitioners employed technical communication jargon in their writing about the field. The findings from the TC Library data for that study were highly useful, and led me to argue for a change in scholarly publication for global audiences in our field.

We believe that if the site continues to grow in popularity that it may help those in our field to research a number of similar topics of interest to the profession as a whole.

Saurabh: What are the best approaches to balance the diversity of participation models and tailoring presentation of content and nature of services to audience needs?

Geoffrey: *Balance* is always the operative term here. In the case of the TC Library, we have an active eight-member editorial board and an active mailing list which permits a variety of our editors to experiment with interfaces that foster new sorts of participation by our users.

Saurabh: Every aspect of the library must be planned with scalability in mind. Of course, this will clearly have major implications for infrastructure definition and design. Are we moving rapidly towards an infrastructure that can support and facilitate research with this common vision?

Geoffrey: Scalability is, in some ways, a bugaboo of the dot-com revolution. While there might at one time have been the prospect of building a service that could become among the most popular websites in the world, it seems far less likely today than in the past. We don't see the prospect of our site growing in popularity by a factor of ten or twenty as a realistic goal. Technical communication simply isn't that large of a field, as yet.

The TC Library website would certainly benefit from a RAIC infrastructure that would have regional servers in different locations around the world (to serve our visitors more quickly). We would always like newer hardware and more sophisticated software tools to enable our system to improve response times and scale for larger audiences. The EServer always publicizes its wish list for donations to enable certain equipment we would like (see *http://about.eserver.org/wish_list.pdf*). But at present, we see standard workstations as being more than capable, when well-planned and well-managed, to serve the scale of audiences likely for any site with a TC focus.

That said, if any of your readers has $20,000 they'd be willing to donate to help us upgrade our system to a more powerful, more scalable hardware platform, we'd be delighted to talk with them. :)

Saurabh: How can digital library projects strive for approaches that incorporate high functionality and extensibility?

Geoffrey: We must constantly work for it by attending to both the details of the site and the "big picture" of the field of electronic publishing as a whole. This is, to some degree, why I feel the organization of the TC Library site is well-suited to the tasks. We have academics who research and publish in the fields of information design, and who therefore study the larger implications of emerging technologies, but the actual practice of running our server ourselves also keeps us grounded in the technical

intricacies necessary to appreciate the detailed work done by our volunteers who help us improve the site.

Saurabh: Integrating the strengths and limitations of purely computer-based technologies for describing objects and repositories, and the appropriate roles for the efforts of human librarians and subject experts in the digital library context as a complement to these technology-based approaches is a challenging issue. What are the best approaches to achieve this integration?

Geoffrey: In my experience, the best approach is to build close connections between technical communicators and librarians. My own mother is a tenured academic librarian, and my sister is a professional database developer. We've worked closely with librarians in every stage of developing the TC Library website. When we first built the site, I worked closely with the head of the Engineering Library at the University of Washington, as well as a number of librarians and colleagues in the Information School at UW. Since we

relocated to Iowa State University, I have worked closely with academic librarians here about our user interface and our site's ability to interoperate with Library Management Systems (websites that aggregate academic indexes' content into searches). It is always useful to work closely with the experts in research when building services such as online libraries. We've been very lucky to have the level of support we have.

Saurabh: What more services and functionalities are you planning to add to the EServer library in the coming years?

Geoffrey: We've been working for the past year intensively on XML syndication systems, including RSS, the Journal Archiving and Interchange DTD, and AJAX (Asynchronous JavaScript and XML) interfaces to our database. This will permit us to import new articles automatically from a variety of sources that regularly post new articles (particularly the peer-reviewed scholarly journals). I look forward to seeing the results of these endeavors by late 2007.

Saurabh: Any message to our readers?

Geoffrey: I have been a subscribed reader of the *KnowGenesis IJTC* since I first discovered issue two in 2006. I look forward to cooperating with the authors, editors, and readers of the journal as we develop future versions of the TC Library website.

I suppose I'd encourage any *IJTC* readers who would be interested in helping us to improve the services, interfaces, or collections of the EServer TC Library to contact me at *webmaster@eserver.org*. We're always looking for partnerships, and *IJTC* readers may help us to plan for features and capabilities of our site that hadn't yet occurred to the members of our boards.