Building Blocks For Rapid Development of Information Seeking Support Systems

Gary Geisler
School of Information
University of Texas at Austin
1 University Station, D700
Austin, TX 78712-0390
512 232-2980
geisler@ischool.utexas.edu

ABSTRACT

This paper proposes that to help accelerate progress in our understanding of how to best develop information seeking support systems in an increasingly wide range of domains, a coordinated effort to develop, publicize, and use *building blocks*—tools, services, system frameworks, data sets, sample collections—would make it easier for system developers to more quickly create and evaluate new systems and allow these developers more time to focus on potentially useful novel features. An explanation of this rationale is followed by a preliminary list of potential building blocks and a discussion of the challenges associated with encouraging use and development of building blocks.

1. INTRODUCTION

As Marchionini [2] suggests, systems designed primarily to present ranked retrieval results—no matter how effective the underlying retrieval mechanism might be—are not sufficient for many information seeking tasks. There are clearly many situations in which finding specific information resources might be part of the user's workflow but not the central activity. The learning and investigation tasks described in Marchionini [2] suggest a wide range of activities beyond entering search terms and evaluating the returned results, as do the "design principles" described by Shneiderman [5] in his analysis of how to better develop creativity support tools. These activities include exploring information by different means, such as graphical visualizations, results tightly-coupled to user interaction (dynamic queries), clustered and flexible categorization; personalizing and storing resources found useful through user histories, annotations, tagging, and persistent user-based lists and collections; and communicating and sharing information and discoveries with other people.

The April, 2006 special issue of the *Communications of the ACM* focusing on exploratory search highlights some of the systems and tools developed to support a broader range exploratory search tasks. White et al. [7] point out in their introduction to this issue that while most of these systems and tools are prototypes restricted to specific domains, information technology has advanced to the stage where we can begin to develop these types of systems and tools to support exploratory search more broadly. This paper builds on that premise, along with the ideas and suggestions for creativity support tools put forth by Shneiderman [5], and argues that one factor in stimulating progress toward better information seeking support systems might be a coordinated effort to encourage the production and use of freely-available, easy-to-use, *building blocks*—tools, services, system frameworks, data sets, sample collections—that will enable system developers to more rapidly assemble and evaluate new information seeking systems in new domains.

2. WHY BUILDING BLOCKS?

Consider a simple but recent, real-life example: The author teaches a graduate-level course on digital library principles and development to information and library science students, the large majority of whom have neither a computer science background nor programming expertise beyond, perhaps, a scripting language such as PHP. Figure 1 shows a timeline students developed for the Spring, 2007 course, in which the class project was to develop a digital library of archival materials from the production of the film *Gone With the Wind*. Figure 2 shows a timeline developed by the students in the course the following year (this time focusing on archival materials associated with the Alfred Hitchcock film *Rebecca*). The first example shows part of one of a series of linked but static graphics; somewhat useful for users of the digital library, but limited. In the second example, the students used SIMILE [6], an open source timeline tool developed at MIT. Because SIMILE is free, well-documented, and designed to be used with minimal effort, creating the *Rebecca* timeline required no more time and effort than the static timeline that was done for the *Gone With the Wind* digital library. It is clear, however, that as a prototype system to be evaluated and used by potential users, a digital library with the sort of functional, interactive timeline provided by SIMILE will be much more useful as a realistic demonstration and for obtaining constructive feedback in an iterative system development process.

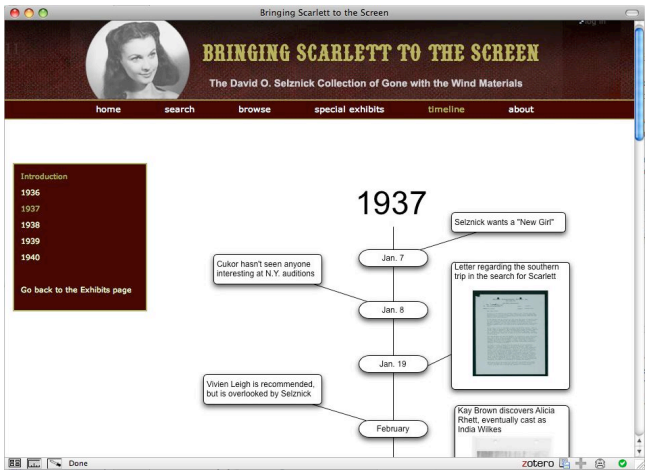

Figure 1. Static timeline from 2007 course

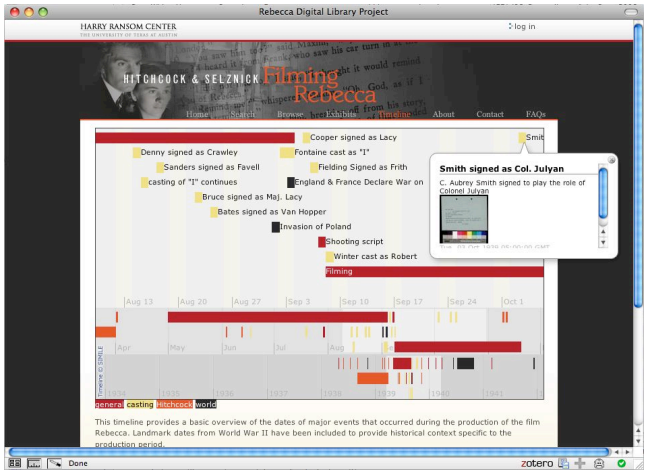

Figure 2. SIMILE-based timeline from 2008 course

While the example above is admittedly simple and anecdotal, it is important to recognize that as information systems increasingly become Web-based, the developers of these systems are increasingly more likely to resemble the students in the digital library course example than expert programmers with computer science degrees. The growing quantity of information resources becoming available in digital form naturally leads to a corresponding interest in creating focused collections of these information resources in a broadening range of disciplines and domains. The growth of digital libraries and digital media collections is one example of this, as are the efforts by publishers to provide increasingly feature-rich systems around their content. While, for instance, the New England Journal of Medicine might have the resources to hire computer science graduates to develop useful new features for their medical resource-based information seeking system [3], there will be many other situations (in archives, museums, and libraries, for example) where the people responsible for creating new information seeking systems do not have particularly strong programming backgrounds and are, in some cases, hired more for their domain expertise. Building blocks could help these developers better contribute to the creation of useful systems in new domains.

Building blocks for information seeking support systems could benefit more than those without sophisticated programming expertise, however. Even for experienced programmers, developing new system tools and features takes considerable time and effort. As more digital information resources and resource collections become available, there will be a broadening range of domains for which we might create useful information seeking support systems. While the literature in HCI and related fields shows no shortage of exemplary systems created for specific domains, it seems likely that most of these were developed over extended periods of time with relatively good grant or industry funding. As we face unprecedented opportunities to work with new collections of resources and new types of users and tasks in new domains, it is imperative that we be able to develop and evaluate systems for these domains more quickly than in the past. To create truly useful information seeking support systems in novel domains, and indeed to stimulate progress towards a general understanding of how we can create better systems for information seeking, we need to be able to rapidly and iteratively develop and test new ideas. Building blocks would help even experienced programmers by enabling them to rely on existing tools where appropriate and devote more of their creative energies to potentially useful new and novel features.

3. ISSS BUILDING BLOCKS

The SIMILE timeline tool described earlier is considered to be an example of a "building block" in this paper because it has been expressly provided to be easily used and adapted by others in their own systems. A developer of a new information seeking support system who believes a timeline might be beneficial to the intended users of that system does not have to design and write original code to provide this feature. A small investment in time to install and configure the timeline tool enables the developer to add a potentially useful feature to the system and move on to the next. Although the ideal scenario where the developer can pick and choose among a wide range of building blocks for a new system is not without significant challenges, some of which are discussed later, a growing availability of similar tools, along with a favorable climate towards open source development and improvements in supporting technologies, suggests that a much broader set of building blocks is not an unrealistic goal.

After many years, it seems safe to say that the open source software movement has made significant inroads into mainstream computing and that the open source model has led to rapid improvements in a range of software applications. Many of the existing tools and services that might be used as building blocks in information seeking systems are open source or at least made freely available. We've also seen funding agencies in recent years encourage, or in some cases require, products of grant-funded projects be made freely available, as well as programs that have directly funded projects to develop open source or free tools and services. In other words, there seems to be growing general support in the research community for the idea that new tools and services should be developed with the intent that they be freely shared and used by others.

At the same time, information technology has progressed to a point where an increased focus on creating tools that are written to be easily shared with and adapted by others is quite reasonable. A wide range of free, customizable communication tools such as wikis and blogs provide many possibilities for sharing information and collaborating on projects. We have easy-to-use methods for organizing and working with data, through powerful and free database management systems such as PostgreSQL and MySQL or well-defined standards for structured data, such as XML, RDF, and JSON. And lightweight scripting languages such as Python, PHP, and Ruby on Rails make it easy to work with databases and structured data to create sophisticated tools and services, making it much more feasible to develop tools and services that can be easily adopted and adapted by others.

So if the time is right to focus on building blocks for information seeking support systems, what might these building blocks be? A preliminary and incomplete list of ideas for the sorts of tools and services from which information seeking support systems could be more easily and rapidly built includes:

- **Sample data sets and content collections:** While systems are often created in response to a need to make more accessible an

existing set of data or collection of information resources, there are also situations where a researcher or developer wants to explore the feasibility or effectiveness of an idea for a new system or feature. In these cases the developer often simply needs a structured collection of resources, any resources. Rather than forcing the developer expend time and effort to assemble a collection of content simply to have something to test with (which often results in the use of unrealistically small collections), a selection of pre-existing data sets and resource collections would not only save the developer time but enable the system to be developed with a more robust set of content.

- **Image, video, audio processing tools:** Multimedia content is increasingly central to digital libraries and other information resource collections. To be made useful in a system, however, this multimedia content often needs processing in various ways. Tools that can help automate and batch process tasks such as format conversion, resizing, preview and summary surrogate generation, and even feature extraction, would greatly increase the speed at which high-quality multimedia-rich systems can be developed.
- **Record cataloging and metadata:** Most information seeking systems rely on rich metadata to describe and distinguish resources and to use as a basis for search, browse, and other system features. Managing resource records—adding new records, modifying record attributes, deleting records—is often done in ad-hoc, cumbersome, and inefficient ways. Tools that provide simple cataloging interfaces and ways to bulk import, export, or convert between data formats could help make system development more efficient.
- **Search engines:** This is an active area of research with much potential for enabling the system developer to choose, based on appropriateness to the domain in which the system is being developed, from different engines with different strengths. The key is being able to easily install and configure an engine and connect it to the rest of the system.
- **Tools for creating browsing interfaces:** Flamenco [1] is a great example of a tool in this category; similar tools based on other organizational schemes would enable a wider range of possibilities for metadata in different domains.
- **Information visualization tools:** While there have been many excellent information visualization projects, examples, and commercial products, there is arguably a need for simpler, lightweight visualization tools that would enable a system developer to integrate visualizations—as a supporting feature, not the focus of the system—into an information seeking system by selecting a type of visualization and performing a simple mapping of visualization attributes to data set attributes. Prefuse [4] is a promising example, though its basis in Java arguably creates some drawbacks in terms of lightweight building blocks.
- **GIS and mapping tools and services:** Relatively easy-to-use mapping APIs from Google, Yahoo, and others have resulted in a plethora of recent examples showing the value of integrating geospatial views of information resources into systems. More polished open source versions of these would make useful building blocks for many systems.
- **Timeline tools:** In many systems chronological-based views provide users with useful ways of finding resources of interest and understanding relationships between resources. SIMILE has been mentioned as a good example of a building block for this purpose, but many other ways of presenting and interacting with resources based on time attributes are possible.
- **Web services:** There are a growing number of easy-to-use methods and APIs that enable a developer to integrate dynamic information from other Web sites or services, such as weather data, news headlines, or Amazon search results.

4. ENCOURAGING USE OF BUILDING BLOCKS

As mentioned above, there are already available excellent examples of tools and services that could be considered building blocks for the development of information seeking support systems. More are being developed all the time. However, because useful tools are developed in many different types of environments, by people with varied motivations, and publicized in different venues, it isn't easy for a system developer to know whether relevant building blocks might be available for a given project and which might be the most useful.

To facilitate the use of building blocks and thereby promote more rapid development of a wide range of information seeking support systems, it might be beneficial to develop a dedicated repository of information about and links to information seeking system related building blocks. This repository could include documentation, examples, and best practices related to information seeking support system development as a way to educate and help developers and encourage the growth of a cohesive research and development community around this topic. A rough example of how this repository might take shape is the Yahoo! Design Pattern Library [8], which provides documented design patterns and stencils for Web development, many of which are linked to code modules that provide dedicated functions.

Creating and maintaining this sort of resource has many challenges not addressed here, but the benefits of doing so are clear: system developers could visit this centralized resource when working on a project, read about available building blocks, determine which might be most useful and relevant for the particular project, and navigate to the source of the building block to download the code and documentation.

5. ENCOURAGING DEVELOPMENT OF BUILDING BLOCKS

The promise of building blocks hinges, of course, on the development of useful tools and services designed to be easily used and adapted by system developers. On the one hand, many tools and services that can be useful for information seeking systems are already being developed, without a focused effort to encourage or organize these efforts. On the other hand, there are many tools and services that could have much more practical value to the HCI, IR, and related research communities if there was an increased focus on making them better documented and sharable. And there are certainly a wide range of needed tools, both in areas discussed earlier and in areas not mentioned in this paper, that do not yet exist but would be very feasible to develop today.

Although most developers of tools and services, at least in the academic and industry research communities, are probably sympathetic to the idea of contributing the work they do to the

community and seeing it used in systems developed by others, there are certainly obstacles that challenge the vision of information seeking system building blocks outlined in this paper. Two of these challenges—receiving recognition for non-traditional intellectual contributions and compatibility of developed tools and services— are discussed here briefly.

Developing useful code is time-consuming; developing code that can be adopted and adapted by others requires even more time, and providing illustrative examples and documentation of the code more time still. Quantifying and assessing the quality of this effort is perhaps more difficult than, for example, evaluating the contribution of a peer-reviewed publication, but it is clear that we are in a transition period where there are increasingly varied ways for those in academia and industry to make significant intellectual contributions to science and society. Better mechanisms for recognizing tool and code contributions in the tenure and promotion process, for example, would surely enable some researchers to spend more time developing useful, sharable tools and code without feeling pressure to spend that effort on more clearly publishable work.

If a selection of appealing building blocks are available to a system developer but these building blocks vary in the programming languages they are written in or the operating systems they support, the value of these building blocks to the developer is diminished. It is difficult to develop a system with tools written in different languages, and not feasible to mix tools that only run on incompatible operating systems. For example, there are useful, freely available tools available today that are based on Microsoft technology (often for good reasons); while these might be potentially good building blocks for systems that can reasonably require Windows, they have limited value for developers interested in creating information seeking systems for intended audiences that do their work on other OS platforms.

6. CONCLUSION

It is clear that systems that enable users to better explore a collection of information resources, to discover new ideas and make connections among those resources, to collect and save resources, and to communicate and share reflections and ideas about the resources with others will become increasingly valuable in many different domains in the near future. We are only in the earliest stages of understanding what might be the most useful components and features of effective systems that support these activities, and how these features might vary in systems intended for different domains. This paper suggests that a focused effort on developing, publicizing, and using building blocks—sharable, adaptable, and well-documented tools and services—could increase the rate at which we are able to develop and evaluate new systems and thereby help accelerate progress in our understanding of how to best provide information seeking system support in different domains.

7. REFERENCES

[1] Hearst, M. (2006). Clustering versus faceted categorization for information exploration. *Communications of the ACM 49*, 4, 59-61.

[2] Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM 49*, 4, 41-46.

[3] New England Journal of Medicine beta site. beta.nejm.org

[4] Prefuse Information Visualization Toolkit. www.prefuse.org

[5] Shneiderman, B. (2007). Creativity support tools: accelerating discovery and innovation. *Communications of the ACM 50*, 12, 20-32.

[6] SIMILE Timeline. http://simile.mit.edu/timeline/

[7] White, R.W., Kules, B., Drucker, S. and schraefel, m.c. (2006). Supporting exploratory search: Introduction. *Communications of the ACM 49*, 4, 36-39.

[8] Yahoo Design Pattern Library. http://developer.yahoo.com/ypatterns/