

A Personal Information and Knowledge Infrastructure Integrator

K. Andrew Edmonds, James Blustein* and Don Turnbull**

Psychology Department, Clemson University, Clemson, SC 29631

*Faculty of Computer Science, Dalhousie University, Halifax, NS

**School of Information, University of Texas at Austin

Email: andy@uzilla.net, jamie@cs.dal.ca, donturn@ischool.utexas.edu

Abstract

The Next Big Thing is being grown organically, cultivated by software developers and pruned by personal Weblog publishers. The rising Weblogging space of the Internet is looking more like traditional hypertext than the Web of the 1990s. The ways in which Weblogging has evolved beyond the previous limitations of the Web as hypertext, and the ways Weblogging is evolving towards common-use hypertext destined to play a critical role in everyday life, will be explored. We have a vision of a universal information management system built on extending the traditional hypertext framework. In our utopian future, everyone will use tools descended from today's blogs to structure, search and share personal information, as well as to participate in shared discussion. We begin by expressing a vision of common-use hypertext for information management and interpersonal communication. This vision is grounded in the rapid evolution of Weblogs and known issues in information systems and hypertext. The practical implications of who will use these systems, and how, is expanded as usage scenarios for Weblogs now and in the future. After recapping the current issues facing the Weblogging community, we look to the long-range implementation issues with optimism. Our system is forward-looking yet realistic. The activities the system will support are extrapolated from recent developments in the online community, and most of the sketches of implementation are based on current approaches. It is of more than passing interest that the features we extrapolate were all described by Nelson as early hypertext ideals. Of particular interest is that the features are now being implemented because of perceived immediate need by communities of interest.

1 Introduction

The profusion of always-on personal computing networking, personal digital assistants (PDAs) and Web-enabled mobile telephones is evidence that many people live in a world where access to information and personal communication is hardly limited by geographical boundaries. The Web is both metaphorically and technically the force that mediates these transfers of data and personal communications. In fact, thanks to the ubiquity of the Web today the metaphor of networks of information nodes, a concept first pioneered by hypertext theorists, is the most commonplace understanding of how information is stored, accessed and created. This understanding is being extended and improved upon by a new generation of software applications, protocols and dedicated users who are putting a more individual, granular spin on creating and accessing information through the use of Weblogs (blogs). The blogging community is vibrant and not restricted to technical elite. Special purpose (but easy to use) software enables people with no particular computer expertise to publish ideas, facts, engage in discussion, and build online directories of resources about a Web's worth of topics.

The Weblogging world has already extended the Web into a more robust hypertext system. Rich site summary (RSS) and related developing forms of XML-based syndication enable transclusive – a process of including one document or part of a document inside another – properties with automated and semi-automated reciprocal linking functionality that is moving towards traditional hypertext backlinks. This blogging system has developed organically through mostly open standards and is fueled by what are essentially citation tracking systems. Blogs themselves are a collection of electronically preserved information where the content itself (its format and structure) is the context of the system (Table 1). Moreover, using standard Web browsers as the composition, editing and viewing mechanism makes the document itself the interface; an evolving personal information hub augmented in value by its relationships to other hubs on the Web. This infrastructure and its trajectory can be seen as a supplementary system, a meta-level above the static and major media areas of the Web.

Our mission in this work is to synchronize the progress made in the Weblog world with longstanding hypertext research and provide an understanding of how the Weblogging phenomenon could be taken forward to truly represent, if not advance, general hypertext functionality as envisioned by its originators, including Bush, Nelson and Engelbart. In his seminal work, [Bush \(1945\)](#) predicted the use of the memex as an information discovery and navigation system, but neglected to focus on the network effect of being able to link in and leverage the power of links created by others to help with identifying and contextual understanding of the storehouses of information he envisioned accessing. Just as search engine capabilities have expanded in recent years due to concepts related to mining link context ([Brin and Page 1998](#), [Kleinberg 1998](#)), blogs expand on the memex's design ideals to make documents, links, news feeds and annotations the glue that is transforming the Web into a hyperlinked, multi-perspective environment. As people become more accustomed to blog-like functionality, their natural proclivity for collecting and commenting on information (either explicitly or implicitly) can prove altogether new methods for both finding and interpreting inter-linked information on the Web.

The main surprise with this phenomenon is the massive appeal of communication using hypertext. From technology enthusiasts to politicians, teenagers to entertainment personalities, personal iterative publishing has become a major trend in the last five years. The acquisition of a leading blogging tool by Google ([Gilmore 2003](#)) [1] and the incorporation of Weblogging into AOL's services ([Brockman 2003](#)) conveys the extent to which this phenomenon has grown.

The first main Internet trend for self-expression was the personal home page, with its list of favorite links and text about likes and dislikes. Soon this trend developed into individual Web sites, where many individuals, small groups or independent businesses could explain themselves to anyone with an interest. Weblogging extends this trend of self-expression to dynamic, almost continually prolific linking and commentary about life and any kind of information on the Web. The proliferation of

alternative linking and distribution methods allows users to both stream information to one another for reading, but also to weave a dense network of links throughout the Web with their own personal perspective and preferences as one hub.

This new form of a hypertext is markedly different from the defining period of the Web in the late 1990s, and hints at the future of common-use hypertext. This paper explores how blogging is embracing the ideals of hypertext as seen in Xanadu (Nelson 1990), Bush's (1945) memex, and the accumulated research of the hypertext and hypermedia communities (Engelbart 1962 and the Bootstrap Institute). The key benefits seen by end-users will be identified. By projecting these developments into the future, we explore the potential impact and the biggest technological hurdles to accomplishing common-use hypertext.

Table 1: Weblogging adapts hypertext features

Nodes:	Encapsulated units of content in any <u>MIME</u> type format identifiable by <u>W3C</u> -compliant protocols and data structures.
Transclusion:	<u>XML</u> <u>RSS</u> -based syndication distributes content across multiple venues.
Link types:	Static and dynamic <u>URIs</u> for tracking and addressing comments, posts and news with time/date stamps and associative properties identifiable using information retrieval methods.
Backlinks:	Trackback and <u>HTTP</u> -referrer linking provides bidirectional links.
Annotation:	A core attribute of many blog posts and the syndication format is a link, often directly connectable to a content-level <u>XML</u> node.

2 The Next Big Thing

If the trends we identified above continue, the future will include virtually everyone using a technology evolved from today's blog software to manage and share information about topics of their choice in a dense network of personal, corporate and aggregated information services. Some people, for instance, will want to have their favorite cooking recipes available wherever they are, while other people will be more interested in sharing political tracts and opinion about current events. Corporations will hire writers to create blog-like network presences that are like today's soap operas, in attempts to promote brand loyalty through social hegemony. The systems we foresee, and the implementation of which we outline below, will easily fulfill all of these needs for everyone with access to today's technology.

The rest of this article is structured thus: we begin by expressing a vision of common-use hypertext for information management and interpersonal communication. This vision is grounded in the rapid evolution of Weblogs and known issues in information systems and hypertext. The practical implications of who will use these systems and how they will benefit follows. A more detailed exploration of Weblogs now and projected into the future includes case studies of usage scenarios. After recapping the current issues facing the Weblogging community, we look to the long-range implementation issues with optimism.

Our vision of the Next Big Thing, while forward looking, is grounded in current practice and demonstrated need.

3 Grand Vision

We introduce our vision of the Personal Information and Knowledge Infrastructure Integrator (*PIKII*) of the future through a series of scenarios. Following these fictional descriptions of how we expect the system to be used, and issues relating to their use, we discuss implementation issues in more detail.

In the common-use hypertext of the future, the world will bear some marked similarities to the current world of Weblogging. People will use hypertext structures to manage their personal information, be it in the form of diaries, platforms for political campaigns, records of research projects (akin to laboratory notebooks), Web clippings (schraefel and Zhu 2001), or networked photo scrapbooks that can then be shared with others and open to collaboration with others. Services will help interested people connect with one another through citation tracking, update monitoring, transclusion and aggregation, and social networking.

3.1 Scenario A: Managing Information of Diverse Types

We consider a hypothetical user named Alice who is planning to purchase a house. She decides the best way to manage the glut of possible useful information is to create and maintain what we will refer to as a home-blog about the buyer preparations she'll need to work through: loans, agencies, state and municipal information, neighborhood opinions, etc. She can either put this information in gradually or link to appropriate Web sites, news postings, and other related blogs and individual blog postings. To solicit advice from others, she can selectively make her content public or enable access only to select family members, realtors, and potential neighbors. Gradually, over time, this personal blog can serve as a living information repository about the house purchase process.

Later, after the house is purchased, we imagine that Alice will decide to extend the use of the blog to plan for home improvements. She may take and post photos of the house and yard, including plans for improvement, proposed schedules, and links to home improvement tips on the Web. Again, selective publishing of this information may serve to elicit comments from others and serve as the basis for her progressive portfolio of house improvements. In sum, Alice's blog can serve as the centerpiece of her home management information system, with the potential to continually evolve through progressive postings as well as comments and links from others in her circle of access. Such circles might be termed tightly-knit communities or internetworked information communities by other authors.

By using off-the-shelf Weblog technology, Alice's information center (her home-blog) can be accessed through almost any Web-enabled device. For example, Alice could refer to her blog via a wireless PDA when shopping for materials or to illustrate

some concept to receive more exact advice when undergoing a project. The blog can also serve as a troubleshooting platform for asking for direct kinds of help or to comment on products or services and show their results to a possibly wide audience.

3.1.1 Implications of this scenario

As system capabilities grow, traditional browsers will become both more expansive with functionality for knowledge production and refined for information consumption, including improved integration with other data sources (Bernstein 2003). Personal indexing will enable more fluent access and retrieval, even in the face of massively increased amounts of data. Integrated publishing and annotation tools will be central to the browser experience, as well as other aspects of personal computing, even at the operating system level. The open nature of Web protocols and formats should promote the increased adoption of hypertext for all manner of personal computing tasks, while extending these personal communications for collaboration with the already networked information provided on the Web.

These systems will enable:

- connections between like-minded individuals and groups
- partitioning and identification of communications by audience
- recollection and retrieval for personal information access

The systems that will evolve from today's blogs will become part of the personal information infrastructure: everything will be stored in these formats, be it Palm-like data or personal records and archives. Usage of hypertext will span the domains and activities shown in Table 2.

Table 2: Domains and activities

Domains	Activities
Business	Learning
Financial	Exploring
Medical	Discovering
Entertainment	Reading
Vocation	Commenting
Avocation	Presenting
Research	Synthesizing/summarizing/organizing
Argumentation	Advertising, promoting, and lobbying
Politics	Debate

Later in this article we consider how we expect today's technology to morph into The Next Big Thing. Below we consider the short-term prospects (section 5) and then speculate on longer term changes and capabilities (section 7). Of particular relevance to this scenario is the discussion of the technology underlying blogs today (section 4) and how Alice would be authoring her blog (section 4.2).

3.2 Scenario B: Passive Information Sharing and Active Privacy Protection

Our first scenario (section 3.1) can easily be accomplished with tools that are readily available today. However, it required the user to find all of the relevant information and organize it herself. Our next scenario explores the effect of the additional power to search in other people's blogs for information.

Privacy can be a major concern. Perhaps a teenager may not want anyone to know that they ever enjoyed listening to certain types of music or watching a certain film. Such information need never be revealed to anyone. In the system we imagine, it will be possible to search portions of other people's blogs for data, if the data owner has given permission. In cases where permission has not been given we anticipate that probabilistic referring agents will help. We will explain both of those concepts in turn, using an example.

Say Bob is a hypothetical user of the system we envisage evolving from blogs: a PIKII. Bob wants to give someone a kit to begin doing a craft, such as bead knitting, as a surprise gift. In a common gift-giving situation, Bob is not selecting the item for himself and cannot question the intended recipient to determine exactly what is needed. Time is essential as Bob needs the present for the next day and cannot spend much time searching. Bob uses his PIKII to search over his friends and finds that Francis, who is in the same book circle, is a professional bead knitter. Bob and Francis are in a community of interest, so Francis has chosen to allow Bob to have access to basic information in her PIKII. Bob searches for detailed information about bead knitting kits and discovers little that he can use given Francis's advanced skills in the domain.

Bob's next step is to use a probabilistic referring agent to find someone else he could ask for advice (or in terms of using our system exclusively, someone whose PIKII he could search). Here is where our system is so radical and yet sane: if Bob had the time then he might communicate directly with Francis to ask if she could recommend someone else he could inquire of. But faster than direct contact, Bob is able to access the semi-private network of Francis, gaining pointers to individuals who have granted her access to their private data stores. An automated search of those resources verifies the utility of the information available. It will then be up to Bob, or more likely a process in Bob's system, to contact those people to find the information he is seeking. With an automated system the entire process would appear seamless to Bob and, with unlimited

resources, it would also be quick.

3.2.1 Implications of this scenario

While Alice's primary concern about using such a system might be to ensure that access to her financial data was restricted to her domestic and financial partners, Bob desires the granting of access to be handled seamlessly and relies on a personal connection with a domain expert to enable a bounded search of available resources.

The identification of introductory material on bead knitting is a challenge that today's Google is capable of handling, given both a semantic and an additional limiting beginner keyword. A PIKII that understands your history with a topic might be able to generate such limited queries automatically. In addition, domain specific metadata produced by a network of specific expertise exceeds the capabilities of today's Internet and could provide a concise synopsis of the subcategories in this craft.

Below, we discuss how people who use today's systems form and maintain communities (section 4.3). While speculating about the near future of blog-like technology we discuss recommender systems (section 5.1), the technology most like the example presented immediately above, and how we expect the communities of interest to grow (section 5.2). Further implications of this scenario include rights management (who has commercial ownership of the intellectual property represented by the data in their PIKII).

3.3 Scenario C: The Importance of Temporal Context in Making Sense of Content

This next scenario illustrates some important points about our design:

- the importance of users' trust in sharing information through the type of system we envision;
- the difference between dimensions of search;
- the importance of data being presented to users in an order that will best help them to make sense of it.

Chas is a PIKII user. His physician has just told him that he should have surgery to treat a serious medical condition. Chas wants to learn more about the condition so he can decide what treatment is best for him, and to choose another physician he can trust for a second opinion. He does a Web search and quickly finds a recent article in a medical journal about new treatments for his condition. Chas has no specialized medical training and therefore finds parts of the article difficult to understand. He uses a Web-based glossary to find definitions of key terms but still does not feel confident that he understands enough of the article to base a decision on it.

His next step is to find someone else who has information about his condition that can help him: specifically someone who can help him to understand that article. A search with his PIKII turns up many leads. Some of those possible sources are commentaries by others on the article or their experience with the same medical condition, and some are online communities of people with the condition. He investigates the communities but finds that all of them are funded by drug companies. Chas still does not know enough about his diagnosed condition to trust that the information he finds is unbiased.

Trust is an essential issue when evaluating the quality of information. If users do not feel they can rely on sources of information to give them sufficiently accurate information and to keep confidences, then users will be unlikely to use those sources. This property applies equally to commercial information providers and informal contacts. It does not matter to Chas if he does not trust a potential source of information to keep his inquiries about his medical condition private because they do not use up-to-date electronic privacy screens or because his health insurance company owns them. It only matters that he would not feel confident trusting them.

One of the leads Chas finds is a *trail*: a sequence of links that someone else followed and found useful about a topic, or topics. Chas notes that this trail ends with the article he is trying to comprehend and is not authored by anyone with an obvious bias. The trail may be a pre-prepared sequence, often called a *tour* (Trigg and Weisner 1986), an unedited record of links followed by someone else (Bush 1945), or most commonly followed links (Pausch and Detmer 1990, Wexelblat and Maes 1997, Chi et al. 2000).

As Chas reads the documents in the trail, he makes notes for himself about the trail and the documents in it. Notes about significant terms used in the documents are entered into his glossary so he can easily refer to them when reading other documents. Those glossary entries, in effect, span multiple documents. Furner et al. (1999) determined that hypertext editors often do not agree on what links should be made. Their observations support the view (Blustein and Staveley 2001) that readers make the most sense out of documents by making their own links and annotations, although experience shows that people still learn by following trails made by others.

3.3.1 Implications of this scenario

Chas' scenario involves strong issues of privacy and trust when seeking and evaluating needed information. Chas needs to get an overview of a large amount of diverse information. The traditional hypertext trail is a missing piece in our current Web, as is the ability to assimilate various bits of information into a personal, annotated history. Although today's blog authors use Weblogs to keep track of information, because they lack access control mechanisms they are not yet suitable for users like Chas. Current blog technologies also lack strong facilities for extended knowledge building.

Tague-Sutcliffe (1995) coined the terms 'ideal chain' and 'optimal retrieval chain' to describe the sequence of documents that a reader must encounter to satisfy their need for information. In Chas's case he needs to apprehend various parts of medical and biological background before he is prepared to comprehend what is in the document that has the information he needs. Tague-Sutcliffe made clear that information needs are dynamic and, to an extent, personal where a property of *informativeness* measures the power of a trail to provide needed information. Personal and temporal relevance are obviously important factors in that measure. The system we foresee will necessarily use that measure in some way to order posts into

the most useful sequence for the individual reader at the stage they are reading them.

In scenario B ([section 3.2](#)), Bob also found value in a custom query for content appropriate to his level of expertise. Adaptive hypermedia work has established a strong precedent for methods to customize content to a user. But in a world where one's personal information space begins to approach the scope of the entire Internet of today, the [PKII](#) will have to work implicitly. Monitoring of engagement with new content is a key step in supporting the recording of useful trails ([Claypool et al. 2001](#)).

4 Weblogs Today and Tomorrow

4.1 Reading Blogs

Weblogs combine push and pull delivery methods. Dedicated Weblog reading software, called aggregators, enables the low-latency presentation of push models, but the medium is inherently on-demand, as in the pull model. The automated presentation of push might become important with more robust models of user interest. One style of interface provides a newsreader style experience while another orders posts reverse chronologically in an html page. Aggregators also vary in when, if and how they present the original content versus a standard [XML](#) rendering.

[Nelson \(1990\)](#) describes a property called *transclusion* as a process in which part of a document may be in several places. The most transclusive of the aggregator designs is the reverse chronological ordering that merges information from multiple sources into a newspaper-like listing. Frequent polling by search engines and aggregators keeps the fragments up-to-date with edits. While the simple representation of a single author's Weblog posts is more aptly termed syndication, the rise of merged [XML](#) documents from multiple authors on related topics approaches Nelson's vision for transclusion in a way that user's find useful. A key issue that the current Web tool set has dealt with is preserving authorial credit.

Having finally achieved separation of content from presentation on the Web, [RSS](#) enables content to be flexibly distributed and recombined. Services such as [Feedster](#) offer keyword-based search over [RSS](#) items creating logical composites of content. Other services focus on link tracking, enabling a mapping of content across blogs. Readers find new Weblogs through links from other blogs, called *blogrolls*, and topical directories, such as [PhDWeblogs](#). We have more to say about blogrolls in [section 4.3](#).

4.2 Authoring Systems

A key enabler for Weblogging has been the ease of use of authoring tools. By alleviating the need to create navigation and automating structured markup with simple template systems, the barrier to publication has become negligible. Using a template-based system, blog tools automatically take care of creating most navigational links too. The evolution of personal content management systems ([CMS](#)) brings us closer to truly accessible publishing as a characteristic of the Web.

The Weblog community also allows non-blog owners to contribute to discussions. Users may also comment on the actual Weblog post pages and advanced systems distribute these comments in [XML](#). While comments on Weblog posts lack some of the advantages of traditional hypertextual annotation, bloggers find the process captivating and the phenomenon is spreading to new applications. [Selfe and Boese \(2003\)](#) have used the Moveable Type ([MT](#)) content management system to publish a document with each chapter as a blog entry, allowing chapter-level annotation and bidirectional linking.

Integration with browser mechanisms and related software in [PDAs](#) and mobile phones will make it easy for users to reference their experiences and quickly access those of others ([Bernstein 2003](#)). In the browser, this integration might take the form of coupling of bookmarks and history with content authoring. Additional fluency in creating links, augmented with (automatic or edited) metadata, is clearly needed. Tracking a conversation across numerous Weblogs can be a difficult task and hypertext work has shown how link types, for instance, can help create useful overviews.

Additional support for metadata about posts has significant use after authoring, but the challenge is in making the specification easy. One Weblogging system, LiveJournal [[2](#)] supports a sort of node type for the emotional state of the post, and it finds wide use. The lazy web [[3](#)] serves to collect project ideas, a sort of node type. The site is a blog using the [MT](#) system and supports comments, a form of annotation, and trackback, a mechanism for creating bidirectional links.

Weblogs currently serve as a sort of bookmark system for some, but this utility would be greatly enhanced by the ability to publish trails as described earlier (in scenario C, [section 3.3](#)). The information value of a document often depends, in part, on the user and the context in which the user encounters it. The order in which previous documents were presented contributes much to the informativeness of the current document. Current hypertext work is tackling this problem ([Pratik et al. 2003](#)), though the notion is longstanding ([Bush 1945](#)).

4.3 Connections between Bloggers

A common page element for Weblog HTML pages is the blogroll, a list of related blogs. Blogrolls order the blogs by last update and even offer titles of recent posts. In addition to site-level links, individual posts create a network of related links. Two systems exist currently for promoting bidirectional links. Trackback is a simple HTTP notification system in which a linking page requests a reciprocal link. The system was introduced in the [MT](#) system in June 2002 by SixApart ([Trott 2002](#)) and has been adopted widely. It was simply a good idea with a simple implementation using open standards, and it works in moderated and unmoderated forms.

In our vision, people will be able to connect with huge, Web-scaled or small circle-of-friend groups who share a common interest. A key appeal of the Weblogging phenomena is the nature of communication as a medium for sharing and self-expression. Social networking, already currently abuzz with Friendster [[4](#)], Ryze [[5](#)], and other distributed Friend-of-a-Friend ([FOAF](#)) efforts, is thriving. Interaction among individuals and users can be expressed with any number of traditional hypertext link types ([Conklin and Begeman 1988](#)) and are one possible area of extension to make this set of relationships more robust. Link types enable more useful high-level views and add a personal filtering that current online

directories cannot.

4.4 Popularity

Popularity is important to blog authors as it determines their influence in areas of importance to them. The popularity of blogs and other Web pages is most often measured in terms of how easy a Web page is to find with the Google search engine. Google, the most popular search engine on the Web today ([Sullivan 2003a](#), [Sullivan 2003b](#)), uses a technology known as *PageRank* to determine the ranking of results.

PageRank is determined primarily by link popularity. Unlike most other search engines that return ranked results based solely on the terms found in the Web pages listed in the results, Google's results are based on the contents of other Web pages. Pages that contain the terms in the query are considered to be about the topics those terms represent, and the pages that are specifically linked to by those pages are the results returned by Google. *PageRank* has long been considered a form of currency (see [Walker 2002](#), for instance).

PageRank tends to make it easier to find the most popular sites about particular topics [6]. However, with finer granularity of indexing (and querying) it will be easier to find blogs that have more focused appeal. [Pu et al. \(2002\)](#) reported that the average query length (in English and Chinese) is roughly two words, which is in accord with [Nielsen's \(2001\)](#) finding (for English only). However, when using interfaces that promote natural language queries, the length (and possibly specificity) of queries is much greater ([Losee and Paris 1999](#), [Franzen and Karlgen 2000](#)). Personalized query augmentation based upon a model of one's interest ([Pitkow et al. 2002](#)) is one technology to increase the effectiveness of these searches.

Still, *PageRank* relies on an impoverished notion of the link compared with early hypertext systems. The National Education Association came under fire from critics for linking to an external site in the period following the September 11 attacks on the USA [7]. This type of occurrence and the use of the *PageRank* algorithm by Google create a situation in which non-affirming links can be mistaken and inadvertently increase the reach of the targeted content.

5 Short-term PIKII Opportunities

5.1 Recommender systems

Currently there are three ways of obtaining recommendations for books, films, courses, etc., from communities of interest: asking outright, searching published comments, and using recommender systems. This example will concentrate on film but it is general enough to include other recommendations as well.

The first method will always be impractical for people who want immediate recommendations or want to canvas large communities of interest.

The second method only works if members of the community have actively recorded their opinions and reviews. Today people often resort to reviews by film critics to determine which films to see, but the quality of film reviews is highly variable and is often extremely subjective. When reviews are subjective, the person seeking the recommendation must decide how closely the critic's opinions match their own. This situation is often highly unsatisfactory. Furthermore film critics, even good ones, review only a small fraction of available films. The main advantage of the first two methods is the richness of the information that is available from descriptions created by people. However, that very richness requires much time to read and understand. The third method, using a recommender system, requires more detailed investigation.

To examine the third method, using a recommender system, we will use the example of the Movie Lens project ([Good et al. 1999](#)). That project uses anonymous reviews from everyone in the system's database to predict how much one will enjoy a film. The prediction of how much one user will enjoy a particular film is based on other users' ratings of that film. The other users whose ratings are used must have similar ratings for films that the target user has rated.

The recommendations are anonymous: no user can determine which other users gave specific ratings. However, the system has a group feature that allows users to share their ratings (if for instance a group wants to see a film together and wants help in selecting which one to see).

Two drawbacks of recommender systems such as Movie Lens are that:

1. recommendations are not nuanced (a rating is for an entire film and there is no easy way to determine why the rating was assigned);
2. recommendations do not adapt to the rater's changing opinions (a movie that earns a high rating when the rater is a child may not be as well received when the rater is a young adult).

We expect systems such as we described in scenario B ([section 3.2](#)) to develop from needs such as we have described here. The system we foresee will manage a user's data for a lifetime, if not longer, and will enable the recording and use of sense-making features so that, for instance, the user can revisit opinions of a film from 15 years earlier and understand their former state of mind. By including versioning and annotation, the system will allow users to update their records.

5.2 Online Communities Of Interest

Online communities are one of the killer applications of the Internet ([Grossman 1987](#), [Rheingold 2002](#)). We consider a single scenario of problem solving and communication in a huge field of application.

The Mozilla open source development community is already massively hypertextual. Tools exist which transform source code, check-ins and bug reports to HTML. In the last year a robust blogging community has emerged as well as tools for monitoring updates and transcluding excerpts. This blogging community supplements the existing Usenet and bulletin board systems.

Members of this online community span roles from core developers to end-users, and quality assurance volunteers to add-on developers. We will consider this last part of the community for our speculation. The Mozilla suite is also a cross-platform, multilingual application development platform with the reference implementation of the Web browser that is its flagship product. Developers using this toolkit are referred to as Mozilla Application Developers (MAD).

The MAD community suffers from a lack of adequate documentation of the underlying platform, forcing developers to seek out the personal knowledge of other developers for complex efforts. These efforts often require reference to other developers, source code and previous discussions.

An introduction of a new developer by an experienced MAD developer to an original author of a toolkit (the author) might occur with partial transclusions from both of the developer's personal blog/PIKII spaces. The request would arrive, not as an email, but as a request for information in the author's to-be-attended workspace for the Mozilla project. Automated content analysis between the nodes related to the information request and the author's personal historical record would confirm the relevance of the request and that an appropriate level of searching of the public computer network had occurred prior to the personal request. This automated processing and the personal relationship between the more experienced developer and the author, or a more general notion of community karma, would place the request at a priority level. The best known use of community karma is at [Slashdot](#), a community-moderated bulletin board system for distributing and discussing news reports.

If the author had previously answered this question but did not remember where, he would form a search based on multiple attributes, for example, keywords and a Web location reference. If the previous answer had been close but not an exact response to the new request, a new node might be registered with a bidirectional link to the original answer, typed as an elaboration link.

5.3 Business Use of Weblogs

The technologies described to this point have obvious implications for businesses or commercial purposes. Weblogs should move towards being the common format for corporate knowledge exchange. Each individual's work can be published with permissions for particular group members, for internal corporate consumption or eventually edited and approved for external use as altogether new information or as additional content for commercial Web sites. Business desktop operating systems will gradually evolve into content management and creation toolkits, using open Web standards to network and store both personal and corporate business data. These new formats for data access and storage will enable a more open development path for extending systems, no proprietary lock-in and extensible, customizable interfaces at the client or content level.

Corporate portals could be transformed into RSS reader interfaces with dynamic data selected by each user in association with their work responsibilities and interests and then augmented with the recommender technologies proposed earlier. Opening an organization's hierarchy to one of information sharing would encourage users to comment and improve any information item via their own networked information space, to be shared with others interested in the same topics or working on similar work projects. More blogging and linking will create a social capital in the organization, akin to Gatekeepers ([Allen 1977](#)) where those who are sources of information often continue to acquire more information through networking (both physical and informational) gradually enhancing both their value to the organization and amongst their peers. These new technologies can network both the organization and improve the physical and virtual links between employees, businesses and their customers.

6 Summary of Current Issues

The amount of change required to move from the current technology to our vision of technology will include: systems for access control, search and relevance estimation, navigation and personal information organization, and metadata for both links and nodes.

RSS, the XML syndication format most common in Weblogs, has suffered from format forks in its development. Although RSS is an important move in terms of altering the granularity of publishing to traditional hypertext node levels, the amount of metadata present in this format is clearly insufficient and there are heated debates about the next generation of formats. The adoption of the Resource Description Format (RDF), an XML format capable of representing full graphs instead of the simple trees of XML and using namespaces, shows promise for extended metadata in syndicated Weblog content. Developments in this area could bootstrap conversational tracking and increase the effectiveness of distributed hypertext discussions. Version 2.0 of RSS ([Winer 2003](#)) incorporates comments and provides pointers to the URI to add new comments.

Early efforts at syndicating news through RSS are being used as a bootstrap to enable more fundamental transclusive functionality. Topical RSS collections, crafted by reading and publishing software as well as search engines, are approaching realization of an infinite number of composite documents. Controlling access to these collections in precise and editable ways is one of the areas most in need of development.

Granular and easily modifiable access control to personally crafted content collections is needed. Today we can search the content of blogs that have been broadcast with RSS-based tools such as Technorati ([Sifry](#)) and [Feedster](#), but we have no tools that enable automatic content negotiation between users' software agents or even between two users. The closest to the vision described in scenario B ([section 3.2](#)) is that some blog authors choose to restrict who can read their blogs (so they are semi-private) but those blogs are not available for indexing or content negotiation with RSS-based tools. This area will require much progress if the sharing of private information in PIKIIs is to occur as we expect it will. At least some of the necessary impetus will come from the business cases being made for the development of the semantic Web ([Berners-Lee et al. 2001](#)).

Tracking conversations in the blog world is augmented by an array of dedicated services. Traditional hypertext metadata like link and node types as well as the personal network attentional management features mentioned here are key for retaining usefulness as the amount of blog content in the world grows. Other types of sequences, like those of buying and caring for a house, will also need to be represented in consistent and machine readable ways.

In addition to opportunities to move the infrastructure of the Web forward, work on browser clients is also progressing ([Phelps and Wilensky 2001](#)). Notions of adaptive clients, perhaps starting with the adaptive homepage ([Anderson 2002](#)) and

incorporating richer revisitation support (Tauscher and Greenberg 1997) into a personal information manager are needed to manage the massive growth of information. The recent release of the Mozilla browser from Netscape and the creation of the Mozilla Foundation (Decrem 2003) provides a world class Internet client for customization and experimentation.

7 Looking Ahead

An enriched personal history of interaction with any networked information, organized by time, location or activity will add much-needed context to ubiquitous computing and its potential for always-on history collection. This history will be available in the universal information manager for user controlled contributions to a spectrum of distributed access, from private to public and dynamic to archival. Already the practice of moblogging (i.e. the use of digital camera-equipped cell phones to take and share photographs taken anywhere [8]) is expanding the abilities of personal information collections. Moreover, this expansion of digital information collection leads to a multimedia-rich world of individual history, shareable with family, friends and others as permitted. Flexible recombinations of media will allow the easy assemblage of interlinked hypermedia scrapbooks in the PIKII: to catalog the interactions of subsets of people, places and activities enabled by automatically created metadata at the time of media creation, through subsequent interaction and by explicit tagging.

Systems that generate and use implicit tagging and information classification are also key elements of the PIKII. Just as Google uses popularity and relevance measures to sort and rank Web information, authoring tools will enable the use of information annotation in appropriate metadata dimensions to add information about a link or node of information. Such link type information might be, at its simplest, an affective score or a value along a more sophisticated dimension such as typing the rhetorical relationship. This information, when combined with personal history, information content, the interaction with a peer's data (expressed in any number of ways from a blog post, shared access to personal information or popularity measures), will be key factors that help make information searching more personally relevant.

Beyond singular units of information, the PIKII will provide interfaces for mapping discussions distributed across the Internet and could be the catalyst for widescale adoption of link types in more traditional discussion systems. Affective components of link types may dominate the social aspects of Weblog communication due to simplicity in authoring and dynamic typing through the explicit and implicit methods previously noted. While transclusion and annotation have formed the basis for widespread adoption of hypertext for Weblog communication, the proposed link and node type additions, as well as more general metadata improvements, will facilitate the intertwining of information, but with an intelligence to help manage attention and provenance.

In many ways, this article aligns with a subset of the goals of the semantic Web space (Berners-Lee et al. 2001), which also promises utility for metadata-enriched information about everyday events. In an ideal world, service providers and vendors, software tools and agencies would offer information in standardized, metadata-enriched, machine readable formats suitable for semantic Web intentions. Many chores might be automated, as in the arrangement of health care for example.

Expanding from the semantic Web, a system of successful micropayment schemes may arise, whether they be karmic and barter schemes or involve actual funds transfer that may drive the received value of both preparing and accessing this semantically-enriched information. Information exchanges with knowledgeable experts and the distribution of favors through a Friend-of-a-Friend network may prove to be more valuable and more popular than micropayments. As we have seen, a key to the widespread adoption of Web information to date is the ability to connect openly with individuals and groups who share common interests, a trend that should continue.

This combination of personal, aggregate and networked contextualizing of information nodes and their linking methods has wide-ranging potential for many dimensions of personal knowledge management efforts. The critical need for personal information management and publishing is to bring the fluency that Weblogging software has created for publishing to the process of connecting and integrating information, leading to a storehouse of personal knowledge.

8 Conclusion

We have a vision of a universal information management system built on a hypertext framework. In our utopian future, everyone will use tools descended from today's blogs to structure, search and share personal information as well as to participate in shared discussion. Just as Nelson (1990) envisioned a network where everything is deeply intertwined, we propose that not only everything, but everyone can belong to several, possibly overlapping and discordant, intertwined communities of interest. These communities will form dense networks of information linkage, allowing many types of structured and unstructured content to continually expand and weave even more interconnected webs of relationships.

People are motivated to communicate many aspects of their lives to many different audiences. The rapid growth of Weblogging has affirmed the appeal of hypertext and validated the notion of individuals as content producers. The availability of personal hypertext systems, with support for granular control over sharing nodes, will increase this adoption for both Weblog authors and readers.

The growth in the amount of digitally captured and hypertextualized information in the coming years will be even more astounding than the growth of the Web over the past ten years. There are significant technical challenges to overcome, but the standards-based organic growth of Weblogs and the Internet shows methods by which these challenges might be overcome. Rejecting the Web as not-hypertext is missing the point. The Web is an incubator for a continuously evolving system of content, user interests and supporting technologies.

Acknowledgements

The authors wish to thank the anonymous reviewers as well as Scott Johnson, Andria Burdette and Helen Ashman for valuable feedback on this work. Blustein notes that the name PIKII was partly inspired by the notion of a "pocket Kim" (a wondrous wisdom-dispensing device that helps make sense of your world), which was in turn inspired by Kim Kofmel.

Request for Feedback

Reader feedback is important to us and we invite you to share your thoughts [via email](#) or via [trackback](#) at the [Topic Exchange](#).

To support distributed discussions, each paragraph in this work has an id attribute. For example, the [second paragraph under heading 3.1](#) has the id p3.1.2.

References

- Allen, T. J. (1977) "Information needs and uses". In *Annual Review of Information Science and Technology*, Vol. 4, pp. 3 - 29
- Anderson, Corin R. and Horvitz, Eric (2002) "Web Montage: A Dynamic Personalized Start Page". *Eleventh International World Wide Web Conference*, Honolulu, HI, May <http://www2002.org/CDROM/refereed/468/>
- Berners-Lee, Tim, Hendler, James and Lassila, Ora (2001) "The Semantic Web". *Scientific American*, 284(5):34 - 43, May <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809FC588FF21&catID=2>
- Berners-Lee, Tim (1999) *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web* (HarperCollins)
- Bernstein, Mark (2003) IPodlings, 18 November http://markbernstein.org/Nov0301.html#note_35207
- Blustein, James and Staveley, Mark (2001) "Methods of Generating and Evaluating Hypertext". In *Annual Review of Information Science and Technology*, Vol. 35, chapter 6, edited by Martha E. Williams (American Society for Information Science and Technology)
- Brin, S. and Page, L. (1998) "The anatomy of a large-scale hypertextual web search engine". *Proceedings of the 7th International WWW Conference*, pp. 107 - 117 <http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm>
- Brockman, Katherine (2003) "America Online Members Capture The Spirit of America - In Pictures". AOL/TimeWarner press announcement, 1 July http://media.aoltimewarner.com/media/press_view.cfm?release_num=55253250
- Bush, Vannevar (1945) "As We May Think". *The Atlantic Monthly*, 176(1):101-108, July <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>
- Chi, Ed H., Pirolli, Peter and Pitkow, James (2000) "The scent of a site: a system for analyzing and predicting information scent, usage, and usability of a Web site". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, The Hague, The Netherlands, pp. 161 - 168 <http://citeseer.nj.nec.com/chi00scent.html>
- Claypool, Mark, Brown, David, Le, Phong and Waseda, Makoto (2001) "Inferring User Interest". *IEEE Internet Computing*, November/December <http://www.cs.wpi.edu/%7Eclaypool/papers/ui/>
- Conklin, Jeff and Begeman, M. L. (1998) "gIBIS: A hypertext tool for exploratory policy discussion". *ACM Transactions on Office Information Systems*, 6(4):303-331, October
- Decrem, Bart (2003) "Mozilla.org Announces Launch of the Mozilla Foundation to Lead Open-Source Browser Efforts". Mozilla.org, 15 July <http://www.mozilla.org/press/mozilla-foundation.html>
- Engelbart, Douglas C. (1962) "Augmenting Human Intellect: A Conceptual Framework". Summary Report AFOSR-3223 under Contract AF 49(638)-1024, SRI Project 3578 for Air Force Office of Scientific Research, Stanford Research Institute, Menlo Park, CA, October <http://www.bootstrap.org/augdocs/friedewald030402/augmentinghumanintellect/ahi62index.html>
- Franzen, Kristofer and Karlgren, Jussi (2000) "Verbosity and Interface Design" SICS Technical Report T2000:04 (presented at AAAI Spring Symposium 1997) <http://citeseer.nj.nec.com/313985.html>
- Furner, Jonathan, Ellis, David and Willett, Peter (1999) "Inter-linker consistency in the manual construction of hypertext documents". *ACM Computing Surveys*, 31(4es) <http://polaris.qseis.ucla.edu/jfurner/csurv00.pdf>
- Gilmore, Dan (2003) "Google Buys Pyra: Blogging Goes Big-Time". SiliconValley.com, February 15 <http://weblog.siliconvalley.com/column/danqillmor/archives/000802.shtml>
- Good, Nathaniel, Shafer, Ben, Konstan, Joe, Borchers, A., Sarwar, B., Herlocker, Jon and Riedl, John (1999) "Combining Collaborative Filtering with Personal Agents for Better Recommendations". In *Proceedings of the 1999 Conference of the American Association of Artificial Intelligence (AAAI-99)*, pp. 439-446 <http://citeseer.nj.nec.com/good99combining.html>
- Grossman, Wendy (1987) *net.wars* (NYU Press)
- Kleinberg, J. (1998) "Authoritative sources in a hyperlinked environment". *Proceedings of the Ninth ACM-SIAM Symposium on Discrete Algorithms*. Also appears as IBM Research Report RJ 10076, May 1997 <http://citeseer.nj.nec.com/87928.html>
- Kuhlman, Ashby (2002) When is a link an endorsement? September 6 <http://www.ashbykuhlman.net/blog/2002/09/06/0546>
- Losee, Robert M. and Paris, Lee Anne H. (1999) "Measuring Search-Engine Quality and Query Difficulty: Ranking with Target and Freestyle". *Journal of the American Society for Information Science and Technology*, 50(10):882-889 <http://www.ils.unc.edu/%7Elosee/par/paril.html>
- Nelson, Theodor Holm (1990) *Literary Machines*, edition 90.1 (The Distributors, 702 South Michigan, South Bend, IN 46601-3122)
- Nielsen, Jakob (2001) "Search: Visible and Simple". *Alertbox*, 13 May 2001 <http://www.useit.com/alertbox/20010513.html>
- Pausch, R. and Detmer, J. (1990) "Node Popularity as a Hypertext Browsing Aid". *Electronic Publishing: Origination, Dissemination and Design*, 3(4):227 - 234, November <http://cajun.cs.nott.ac.uk/compsci/epo/papers/volume3/issue4/ep035rp.pdf>
- Phelps, Thomas A. and Wilensky, R. (2001) "The Multivalent Browser: A Platform for New Ideas". *Proceedings of Document Engineering*, Atlanta, GA, November

Pitkow, James, Schütze, Hinrich, Cass, Todd, Cooley, Rob, Turnbull, Don, Edmonds, Andy, Adar, Eytan and Breuel, Thomas (2002) "Personalized Search". *Communications of the ACM*, 45(9), September

Pratik, Dave, Karadkar, Unmil P., Furuta, Richard, Francisco-Revilla, Luis, Shipman, Frank and Dash, Suvendu (2003) "Navigation, Path-centric browsing, Navigation metaphors, Directed paths, Walden's Paths, Path Engine". In *Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia*, Nottingham, UK, August

Pu, Hsiao-Tieh, Chuang, Shui-Lung and Yang, Chyan (2002) "Subject Categorization of Query Terms for Exploring Web Users' Search Interests". *Journal of the American Society for Information Science and Technology*, 53(8):617-630

Rheingold, Howard (2002) *The Virtual Community: Homesteading on the Electronic Frontier*, revised edition (MIT Press)
<http://www.well.com/user/hlr/vcbook/>

schraefel, m.c. and Zhu, Yuxiang (2001) "Interaction Design for Web-Based, Within-Page Collection Marking and Management". In *Proceedings of the Twelfth ACM Conference on Hypertext and Hypermedia*, Aarhus, Denmark, August

Selfe, Cynthia and Boese, Christine (2003) The Clemson Laptop Program: Insiders' Perspectives <http://www.nutball.com/laptopresearch/>

Sullivan, Danny (2003a) "Nielsen NetRatings Search Engine Ratings". *Search Engine Watch*, 23 February
<http://www.searchenginewatch.com/reports/article.php/2156451>

Sullivan, Danny (2003b) "comScore Media Metrix Search Engine Ratings". *Search Engine Watch*, 16 August
<http://www.searchenginewatch.com/reports/article.php/2156431>

Tague-Sutcliffe, Jean (1995) *Measuring Information: An Information Services Perspective* (Academic Press)

Tauscher, Linda and Greenberg, Saul (1997) "How people revisit web pages: empirical findings and implications for the design of history systems". *International Journal of Human Computer Studies*, 47(1):97-138 <http://ijhcs.open.ac.uk/tauscher/tauscher-01.html>

Trigg, Randall H. and Weisner, Mark (1986) "TEXTNET: A network-based approach to text handling". *ACM Transactions on Office Information Systems*, 4(1):1-23, January

Trott, Mena and Trott, Ben (2002) Feature: TrackBack, 10 June http://www.movabletype.org/trackback/archives/2002_06.html

Walker, Jill (2002) "Links and power: the political economy of linking on the Web". In *Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia* (ACM Press), pp. 72 - 73

Walker, Jill (2003) "Definition of weblog". To appear in *Routledge Encyclopedia of Narrative Theory*, 28 June 2003 version
http://huminf.uib.no/%7Ejill/archives/blog_theorising/final_version_of_weblog_definition.html

Wexelblat, Alan and Maes, Pattie (1997) "Using History to Assist Information Browsing". In *RIA0'97: Computer-Assisted Information Retrieval on the Internet*, Montreal

Winer, Dave (2003) *RSS 2.0 Specification* <http://blogs.law.harvard.edu/tech/rss>

Links

Bootstrap Institute <http://www.bootstrap.org/>

Feedster <http://www.feedster.com>

PhDweblogs <http://www.PhDweblogs.net/>

moblogging.org, Mobile blogging resources <http://www.mobloqqing.org/>

Ito, Joi (2003a) Moblogging, Blogmapping and Moblogmapping related resources <http://radio.weblogs.com/0114939/outlines/moblog.html>

Ito, Joi (2003b) Adding More Information to Links http://joi.ito.com/archives/2003/03/14/adding_more_information_to_links.html

Slashdot, Open Source Development Network <http://www.slashdot.org/>

Seitz, Bill, WebSeitz/wikilog <http://webseitz.fluxent.com/wiki/FrontPage>

Sifry, Dave, Technorati, Sifry Consulting <http://www.technorati.com/>

Notes

[1]

We have more to say about Google's power in [section 4.4](#).

[2] LiveJournal is at <http://www.livejournal.com/>. See especially the LiveJournal Moods Web page <http://www.livejournal.com/moodlist.bml>

[3] lazy web is at <http://www.lazyweb.org/>. Instructions for use are as follows: Do you have an idea that you think others might be able to solve? Make a LazyWeb request by writing it on your own blog, and then sending a Trackback ping to the new url:

<http://blog.mediacooperative.com/mt-pi.cgi>

[4] Friendster is a Web-based service that creates informal groupings of people based on the descriptions of their interests and those of other people they are grouped with. Friendster is at <http://www.friendster.com>

[5] Ryze is a Web-based service designed to help people create personal networks for many reasons but mainly to help them advance their careers. The Ryze is at <http://www.ryze.com>

[6] We are indebted to Cathy Marshall for her observation that Google tends to consolidate power for high-ranking Web sites.

[7] For an account of the NEA issue, see [Kuhlman \(2002\)](#). More Weblog discussion on the notion of no-endorsing links may be found at [Ito \(2003b\)](#).

[8] See [directory of mobile bloggers](#). [Ito \(2003a\)](#) provides a list of technical requirements for moblogging adoption.