

# Electronic Documents

## New Roles and Value-Added Contents

Chern Li Liew and Schubert Foo

Nanyang Technological University  
School of Communication & Information  
Division of Information Studies  
31 Nanyang Link, Singapore 637718

**Abstract.** The progressive research and development in computing and information processing technologies have stimulated the emergence of a new range of electronic documents that are expected to fully exploit the electronic medium's basic properties of added interactivity and malleability. Coupled with the affiliation with metadata and additional layers of e-contents, e-documents are engaging a new paradigm of users who are actively seeking, selecting and constructing information more than they are 'receiving' information contained within e-documents and collections of such documents. This paper reports on the pertinent findings of a research that intentionally took an immense leap in the design and empirical evaluation of an information environment for a new generation of e-documents that is endowed with a set of innovative interactions and electronic enhancements. Apart from presenting a list of essential features of future e-documents that will serve as useful design guidelines, the research advocates that the way forward is not to promote e-documents either as replacements or as alternatives to their print counterparts, but rather to continually strive to focus on the needs of users and on the tasks they will likely seek to accomplish. Future work on designing e-documents should thus focus on designing for media complementarity, and aim to further develop and incorporate tools and features that are most helpful to the user community.

## 1 Introduction

Few industries have experienced a degree of change that equals the profound transformations that are now rocking the publishing industry. These changes, brought on

by new developments in science and technology, are global in scale. There are three main driving forces behind this progress. The first is computing power. The second force, data communications and networking, has become commonplace in the same time-frame. The third factor, often underestimated, is information content. A vast proportion of scholarly literature is now being generated or converted in electronic format, stimulating the progress and acceptance of technology-mediated communications. Since the early 1990s, digital audio-visual media and computer-based documents are fast becoming an influential medium for both professional and scholarly communication [1, 2]. During the past few years, it has become apparent that there is an urgent need for innovative approaches to support the use of e-documents. User interface mechanisms that support multiple information gathering, analysis and communication tasks are greatly needed. New paradigms and technologies must engage much more than merely allowing users to 'see' information. Users must also for instance, be able to manipulate it to focus on what is relevant, to re-organise it to create new information, and to share and communicate their new discoveries to other users. This paper examines the idea of the new roles e-documents play in scholarly and scientific communication and explores the potential of electronic value-added contents offered by e-documents.

## 2 Interactive E-Documents: Effects & Potential

E-documents may contain and organise many forms of interactive media, including text, numbers, still photographs or illustrations, animation, visualisation of spatial or numeric information and digital audio-video materials. A review of currently existing e-journals show that these examples of e-documents are moving towards the direction of being interactive systems, offering users various interactive features to interact with information flexibly. Regardless of the economics of e-publishing, there is no doubt that e-publishing will allow fully enabled information discourse. There are no added costs to producing a colour image rather than a black-and-white one. Videos can be created and saved in a digital files that can then be viewed on any computer that has the appropriate software. The same is true for sound.

More important and revolutionary perhaps, is that e-publishing brings data into the hands of the users in a way that facilitates interaction and discovery. The best way to illustrate this is by providing an example drawn from the field of chemistry. Chemists are very interested in the structure of a molecule, and one way of obtaining this is via x-ray crystallography. Once a crystallographer has determined the structure, she must then select a single view of this molecule for presentation in the printed article. On a printed page, this is a static image. A reader of this work may be interested in a view of the back side of the molecule. Unfortunately, the reader is forced into some very time-consuming manipulations to obtain the necessary view, either by entering the structure by hand into her own molecular viewing program or by obtaining the structure from a database and

feeding it into the viewer. With e-publishing, what is published is not necessarily the static image but the co-ordinates of the molecule itself. The reader can use a browser that will automatically direct the co-ordinates into the appropriate viewer, and the reader can then manipulate the structure to her heart's content. This concept of interactive manipulation of data is possible in all disciplines – be it the 3D image of a brain tumour obtained in an MRI (magnetic resonance image), the scattering tracks made by micro-particles, a matrix of wind speeds in a hurricane or the DNA sequences of a dinosaur. The interactivity and electronic enhancements travel far beyond what is capable in traditional print. This is, perhaps the most compelling reason for the shift to e-publishing.

Nevertheless, users think the same about e-documents as they do about any new developments. Unless users can perceive a real advantage to e-documents, where they can do more or better with these digital documents than they do with paper, users would be reluctant to use them extensively. This paper looks at the likely form e-documents would take in the light of user needs and requirements. The findings and recommendations are drawn from a research undertaken to examine what users require and desire in a new generation of enhanced and value-added e-documents [3].

### **3 A Research on E-Documents: Enhanced Interaction and Value-added Contents**

The research was an exploratory effort to achieve the aim by first, eliciting user needs and requirements, and looking at related present state-of-the-art developments in e-documents that were engaging leading researchers in pushing forward thinking in this subject. Following these preliminary reviews and investigations, a set of characteristics deemed desirable for an environment and its information objects in order to support enhanced interaction and value-adding of e-documents was derived [4]. The properties were subsequently used as the basis for the design of a proposed information environment (PROPIE), a platform to demonstrate how creative and effective e-document use could be supported [5]. Providing a suite of novel features and interactive tools that could be flexibly combined, PROPIE allowed users to apply multiple novel ways to intuitively query and navigate information in e-documents. The querying and browsing processes in PROPIE were supported by various interactive and visualisation techniques. A series of empirical evaluation of various aspects of the environment was conducted to obtain representative users' feedback. Data were collected in three tiers of evaluation through focus groups and a series of questionnaires that collected both quantitative and qualitative data from a total of 83 graduate students. The conclusions were that top-ranked features of such e-documents were those not available in their print counterparts – querying, navigation and visualisation that contributed to enhanced access, advanced interaction and value-adding contents.

## 4 Discussion of Features and Functionality

The features and functionality identified as essential in the study are discussed in this section in conjunction with established findings from other researchers. A portion of the materials in this section has been published earlier [6].

### 4.1 Integrated Access to Multiple Collections/Items

A multiple workspace e-document environment supports the integration of both the user's personal workspace(s) and the broader information world. The tight integration simplifies the task of consolidating the user's intermediate and final information-seeking tasks. Findings of the *Electronic Survey of Max Planck Society Researchers* [7] and the *Café Jus Project* [8] have also repeatedly expressed the need for an integrated access interface to multiple collections of journals as well as to other information sources. Pop-up windows could be used to further support an 'overview + details' viewing and the switch between collections and items should not distract users from their main task. Bates [9] has discussed the importance of supporting the evolving and interactive search process users engage in while using online information via her now-classic 'berrypicking' model. Other studies by Xie [10], Robins [11], Ingwersen [12] and Belkin and colleagues [13] have also emphasised the importance of recognising the need for online systems to support the iterative and interactive search/navigation processes users engage in.

### 4.2 Interactive Visualisation

The fundamental mode of display within the e-document environment should be graphical, giving a visualisation of for instance, the structure of the user's search/navigation, a visualisation of the various conceptual items of the user's information-seeking tasks and a visualisation of the search/navigation session. These visualisations are constructed on the basis of initial user input, and are intended for direct manipulation (DM), modification and selection by the user. This form of presentation and interaction responds to various issues. One crucial one is that it provides a natural structure for showing a user's current state of search/navigation, and the relationship of that to what has come before (and perhaps to what might follow). Another is that it provides a powerful framework of browsing [14]. DM in a coherent manner could also facilitate the spontaneous dynamics of spatial layout in e-documents. Designers for instance, could consider doing away with the idea of constraining document movement to 'active' areas such as title bars and scroll bars. The whole document and the environment should be made 'active' to support intuitive, effort-less navigation within the virtual information space.

### 4.3 Multiple Representations

Multiple representation of information objects and their organisations should be a part of the e-document environment so as to support a variety of task and information-seeking perspectives. Users should be able to manipulate various aspects of the representations and views directly, and be able to make a choice of suitable views according to their needs. Some researchers such as Freeman and Gelernter [15] have argued that a time-ordered approach (e.g. timeline, calendar) to organising information is more suitable than a spatial approach (e.g. map [16], forager [17], 3D trees [18], fisheye [19] and galaxy [20]). Their time-ordered system, 'LifeStreams' stores documents in streams, time-ordered 1D lists of information. Similar time-ordered lists could be found commonly in applications such as email lists or a Web history list. The findings of the current research suggest that the two approaches are complementary rather than exclusive. The advantages of both should be exploited. Although not evaluated in the current research, there could also perhaps be a combination of a spatial arrangement of information with a chronological navigation mechanism such as that demonstrated in the 'Time-Machine Computing' [21]. O'Hara and Sellen [22] have discussed the importance of helping users to gain a sense of the overall structure of a document.

### 4.4 Seamless Access to Other Information Sources

The e-document environment should support seamless access to other value-adding features while maintaining the context of the original document. These sources could include thesauri, a look-up glossary/definition, translation tools and additional details on images/diagrams beyond that is available in the original document. The accuracy, authenticity and comprehensiveness of such information sources must be governed. These value-adding contents should exist as separate layers of components easily accessible and governed by some kind of access control mechanism.

### 4.5 Collaboration with Other Users

The e-document environment should facilitate the communication and collaboration amongst users. Throughout the environment, users should be provided with a suite of interactive tools not only for querying, navigation, and organisation but also for authoring. There could be for instance, tools that allowed a user to annotate while using a document, to view other users' comments and annotations and to interactively communicate with other users via for instance, embedded email functionality within a document. The e-document system should also allow users to search, sort, select and filter the attached markings. A few earlier studies have pointed out the importance of recognising annotating as an integral part of using e-documents. Maurer [23] suggests that users should be able to add notes to

e-documents at every point much in the same manner as some people tend to highlight areas of text, or much as some tend to scribble notes in the margins of a document. According to Marshall [24], users are not adding details so much as extracting only pieces of information of interest. Users are eliminating mentally, then physically the bulk of information that is not relevant, while at the same time recording their own interpretations and thoughts. The study conducted by O'Hara and Sellen [22] has also pointed out the need for markings such as comments and annotations to be functionally distinct from the 'base' documents. They should also exist as separate layers of value-adding components attached to a document and be governed by some kind of access control mechanism. Wilensky [25] has highlighted four types of functionality to support an adequate digital annotation facility: It must be possible to place an annotation *in situ*, the supported types of annotation must be highly expressive and extensible, annotations must be essentially format and platform independent and it must be possible to annotate openly and distributedly.

#### 4.6 Links to Related Works

Visual representation of an item or collection in context to other related objects is important in supporting the context-driven evolution of a user's interest. These representations could for instance, map an item of interest to related items (which include related works cited, works that cited the item and latest works found related or similar to the item) and present their relationships using various visualisation. These tools should also provide links to these related items, providing details on users' demand. The *ISI<sup>®</sup> Web of Science<sup>®</sup>* [26] has demonstrated the use of a variety of links in e-journals through Reference Links, 'Time Cited Links and Related Records<sup>®</sup> links'. Recently, Van de Sompel and Beit-Arie [27] have also explored the possibility of open reference linking in the Web-based scholarly information environment using the Open URL Framework so that links are possible across distributed information resources, irrespective of their providers. Alerting of new related items should also be supported in the e-document environment. The alerting could initially be based on contents and topics, and later on user-defined profiles or user's navigation history. Collaborative information-recommender systems have recently gained much attention. One example of the recent systems is *SurfLen* [28] that uses a quality-based data mining of association rules. Other such systems include *Alexa* [29] where recommendation is largely based on a user's rating and explicit proposals, *GroupLens* [30], a recommender system in the Usenet news domain that uses collaborative filtering, and *Ringo* [31] which makes personalised recommendations for music albums and articles based on a dynamic profile of each user's interest. Most of the available systems provide dynamic, real-time analysis for document clusters, with innovative mining and clustering topologies for displaying of results.

#### 4.7 Interaction with 'Live' and Multimedia Objects

The e-document environment should support interaction with 'live' and multimedia objects. Based on the findings of the current research, these interactions could include:

- The ability to check out a demo file (e.g. a 3D complex chemical structure) without losing context of the original task (supported by the use of multiple workspaces with navigation being via a combination of context-sensitive menus in one frame and hypertext links and demo files within another frame).
- To interact with a live graph and equations whereby users can input their own data sets to the interactive parameters or to interact with parts of author's data and observations by for instance, submitting a query to the author's programme, and
- To have a thumbnail image of all diagrams or pictures available within a document, and to access detailed descriptions and explanations of these images beyond that available in the original document.

#### 4.8 Backtracking

The e-document environment should support backtracking of users' search/navigation sessions and records. Studies have shown how users often engage in moving down levels of granularity as they home in on target clusters of information, or up as they backtrack to seek elsewhere. Information-seeking is a temporarily prolonged process with the users often moving back to previous work done and building on it. All objects within the e-document environment hence, should come with representations of their history in a visualisable, malleable mechanism. The notion of user's path through digital information and their use is at least, as old as Bush's notable MEMEX system [32]. Work done by users to solve problems in e-documents should leave footprints. These traces could be made accessible to future users who could take advantage of the work done in the past to make their problem solving easier. Such history records could be organised and presented to users using time-ordered and spatial approach to enable the archival of the records, and to enable 'time-travelling' - chronological navigation over archived information and users could spatially arrange items in ways meaningful to them or remove items no longer needed.

#### 4.9 'Focus + Context' Viewing and Exploration

A 'body & clone' and the 'multi-clone' analogies should be considered in the environment to support 'focus + context' viewing and exploration. Context is maintained as object 'bodies' are left (highlighted) in the original (home) space while



the 'clones' are drawn out and explored in a separate (action) space. Object 'bodies' always appear in its original form and context. An object may pass through other objects in the space and have its position and form altered in a separate space. Using this analogy, users could also move and slide objects (focus) back and forth more intuitively and easily between 'home' and 'action' space. This also allows the user to switch between focus sets to quickly locate objects that are relevant to current task. Users could also interact with more than one object (or set of objects) to for instance, compare, to reveal pattern, to discover new relationships and doing all these while maintaining context of the whole information space. It also allows users to examine multiple object sets based on non-spatial properties (i.e. they may not be located close to each other in the original space).

#### **4.10 Controlling the Amount of Information Displayed**

Users should be given control over the environment, for instance, to be able to minimise or enlarge, activate and deactivate the workspaces according to users' personal needs and preferences. According to the findings of the current research, users also prefer to have information presented to them in different levels of details. A focus group conducted for the *NSF/DARPA/NASA Digital Library Initiatives (DLI) Project* at University of Illinois [33] has earlier reported similar findings. The study involved some faculty members who reportedly placed great importance on the process of stepping from 'a little information' to 'more information' – e.g. from titles and authors, to introduction and section headings, to full-text. The e-document environment should support this process by breaking a document into various sections and sub-sections.

#### **4.11 Adaptive and Adaptable Options**

An e-document environment should consider incorporating certain adaptive and adaptable options to guide users in their use of the many tools and functionality within the environment that is intended for different tasks and needs. These help systems could for instance: suggest adaptation of menus to users' needs according to user's tasks at hand, recommend tools and explain their use, communicate to users why certain recommendations and adaptations have been done, providing suitable instructions and interventions, and providing additional support and interface simplification for novice users or users with lower spatial ability.

#### **4.12 Workspaces/Windows Auto-Management**

Generally, the focus group findings confirmed the need for managing the workspaces and pop-up windows [34] was found useful. The groundwork for this VW approach



was the gradual reduction in screen real estate requirements for unused workspaces (windows) as time proceeded. This reduction strategy released space for the active workspace(s) while reducing screen contention and unwanted clutter.

## 5 What of the Future?

Publishing is a product of human yearning, and today people yearn for something more than printed text. If a new format comes along, be it the e-document or something else that is demonstrably superior to documents on paper, people can be counted on to embrace it. The participants' feedback show a growing interest in e-documents, and users expect them to offer advanced forms of interactive functionality and other value-adding features that their print counterparts cannot offer. It is evident that for e-documents to survive and thrive, they must be different from their print antecedents and must play a new role. Designers and publishers of e-documents must aim for media complementarity rather than media redundancy – they must fully exploit the electronic medium's basic properties – with 'interactivity' as the primary characteristic of new technologies. With interactive features, end-users are recognised as active participants in the communication process. Users will be increasingly seeking, selecting and constructing information more than they 'receive' the information contained within e-documents. It is however, reasonable to conclude that with e-documents, the ultimate goal is not simply to provide artefacts, but to provide access to information users can use that contributes to their work. The way forward is obviously not to promote e-documents either as replacements or as alternatives to their print counterparts, but rather to continually strive to focus on the needs of users and on the task they will likely seek to accomplish. Future work on designing e-documents should aim to further develop and incorporate tools and features that are most helpful to the user community. There are a number of areas where ongoing and future work should be aimed. It is crucial to continue to undertake studies to find the kinds of functionality and interfaces that will be most important to users. More extensive testing and evaluation of new interface technologies and paradigms in laboratory or field experiments, or other empirical studies involving people who are representative of the user communities should also be undertaken.

## References

1. Lynch, P.J. and Jaffe, C.C. 1990. An introduction to interactive hypermedia. *Journal of Biocommunication* 17 (1), pp. 2-8.
2. Jessup, M.E. 1993. Update in biomedical visualization: the professional communicator's role. *Journal of Biocommunication* 19(3), pp. 2-7.

3. Liew, C.L., Foo, S. & Chennupati, K.R. 2001a. A user study of the design issues of PROPIE: A novel environment for enhanced interaction and value-adding of electronic documents. *Journal of Documentation* 57 (3), pp. 377-426.
4. Liew, C.L. & Foo, S. 1999. Derivation of interaction environment and information object properties for enhanced integrated access and value-adding to electronic documents. *Aslib Proceedings*, 51 (8), pp. 256-268.
5. Liew, C.L., Foo, S. & Chennupati, K.R. 2000. A proposed information environment for enhanced integrated access and value-adding to electronic documents. *Aslib Proceedings*, 52 (2), pp. 58-75.
6. Liew, C.L., Foo, S & Chennupati, K.R. 2001b. Towards a new generation of electronic documents. *The Journal of Information Science* 27(5), pp. 327-342.
7. Rusch-Feja, D. and Siebesky, U. 1999. Evaluation of usage and acceptance of electronic journals, *D-Lib Magazine* 5 (10) (October 1999) Available at <http://www.dlib.org/dlib/october99/rusch-feja/10rusch-feja-summary.html>
8. Woodward, H., Rowland, F., McKnight, C., Pritchett, C. and Meadows, J. 1998. Café Jus: an electronic journal user survey, *Journal of Digital Information* 1 (3) (October 1998) Available at <http://jodi.ecs.soton.ac.uk/Articles/v01/i03/Woodward/>
9. Bates, M. 1989. The design of browsing and berrypicking techniques for the online search interface, *Online Review* 13(5), pp. 407-424.
10. Xie, H. 2000. Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society of Information Science* 51 (9), pp. 841-857.
11. Robins, D. 2000. Shifts of focus on various aspects of user information problems during interactive information retrieval, *Journal of the American Society for Information Science* 51(10), pp. 913-928.
12. Ingwersen, P. 1996. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory, *Journal of Documentation* 52(1), pp. 3-50.
13. Belkin, N.J., Oddy, R.N. and Brooks, H.M. 1982. ASK for information retrieval: Part I. Background and theory. *Journal of Documentation* 38(2), pp. 61-71.
14. Shneiderman, B. 1998. *Designing the user interface: strategies for effective human-computer interaction*, 3rd edition. Reading: Addison-Wesley.
15. Freeman, E. and Gelernter, D. 1996. LifeStreams: A Storage Model for Personal Data, *SIGMOD Record* 25(1), pp. 80-86.
16. NewsMaps – explores landscapes of information. 1999. Available at <http://www.newsmaps.com>
17. Card, S.K., Robertson, G.G. and York, W. 1996. The WebBook and the Web Forager: An information workspace for the World-Wide Web, *Proceedings of ACM Conference on Human factors in Computing Systems: Common Ground*, April 13 - 18, pp. 111-117.
18. Robertson, G.G., Mackinlay, J.D. and Card, S.K. 1991. Cone Trees: animated 3D visualizations of hierarchical information, *Proceedings of ACM Conference on Human Factors in Computer Systems*, New Orleans, April 27 - May 2, pp. 189-194.
19. Furnas, G.W. 1986. Generalised fisheye views, *Proceedings of ACM Conference on Computer Human Interaction*, pp. 16-23.
20. Rennison, E. 1994. Galaxy of news: an approach to visualizing and understanding expansive news landscape, *Proceedings of the ACM symposium on User Interface Software and Technology*, Marina del Rey, CA, November 2 - 4, pp. 3-12.

21. Rekimoto, J. 1999. Time-machine computing: a time-centric approach for the information environment, Proceedings of the ACM symposium on User Interface Software and Technology, Asheville, CA, pp. 45-54.
22. O'Hara, K. and Sellen, A. 1997. A comparison of reading paper and online documents, Proceedings of ACM Conference on Computer Human Interaction, Atlanta, GA, pp. 335-342.
23. Maurer, H. 2001. Beyond classical digital libraries. In Chen, C.C. (ed). Global digital library development in the new millenium. Beijing: Tsinghua University Press, pp. 165-173.
24. Marshall, C.C. 1997. Annotation: from paper books to the digital library, Proceedings of the 2nd ACM International Conference on Digital Libraries, July 23 - 26, 1997, Philadelphia, PA, pp. 131-140.
25. Wilensky, R. 2000. Digital library resources as a basis for collaboration work, Journal of the American Society of Information Science, 51(3), pp. 228-245.
26. Atkins, H. 1999. The ISI<sup>®</sup> Web of Science<sup>®</sup> – links and electronic journals, D-Lib Magazine 5 (9) (September 1999) Available at <http://www.dlib.org/dlib/september99/atkins/09atkins.html>
27. Van de Sompel, H. and Beit-Arie, O. 2001. Open linking in the scholarly information environment using the Open URL Framework. D-Lib Magazine 7 (3) (March 2001). Available at <http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>
28. Fu, X., Budzik, J. and Hammond, K.J. 2000. Mining navigation history for recommendation, Proceedings of Intelligent User Interface, January 9 - 12, New Orleans, LA, pp. 106-112.
29. Alexa. 2001. Available at <http://www.alexa.com>
30. Resnick, P., Neophytos, I., Mileth, S., Bergstrom, P. and Riedl, J. 1994. GroupLens: an open architecture for collaborative filtering of Netnews, Proceedings of Conference on Computer Supported Cooperative Work, pp. 175-186.
31. Shardanand, U. and Maes, P. 1995. Social information filtering: algorithms for automating 'Word of Mouth', Proceedings of ACM Conference on Human Factors in Computing Systems, May 7 - 11, Denver, CO, pp. 210-217.
32. Bush, V. 1945. As we may think. Atlantic Monthly (July 1945). Available at <http://sunnyoasis.com/services/aswemaythink.html>
33. Bishop, A.P. 1998. Digital libraries and knowledge disaggregation: the use of journal article components, Proceedings of ACM International Conference on Digital Libraries, June 23 - 26, Pittsburgh, PA, pp. 29-39.
34. Miah, T. and Alty, J.L. 2000. Vanishing Windows – a technique for adaptive window management, Interacting with Computers 12, pp. 337-355.