Exploiting the Annotation Practice for Personal and Collective Information Management

Guillaume Cabanac[*] — Max Chevalier[*,]**
Claude Chrisment[*] — Christine Julien[*]

[*] *Université de Toulouse, IRIT UMR 5505 CNRS*
118 route de Narbonne, F-31062 Toulouse cedex 9

[**] *Université de Toulouse, LGC ÉA 2043*
IUT Paul Sabatier, 129 avenue de Rangueil, BP 67701, F-31077 Toulouse cedex 4

{Guillaume.Cabanac, Max.Chevalier, Claude.Chrisment, Christine.Julien}@irit.fr

ABSTRACT. *Information nowadays is a capital for any organization intending to be reactive and aware of its environment. Unfortunately most modern organizations overdose on information as almost every member daily accesses, extracts and stores a growing amount of documents, i.e. vehicles for information. This situation is even deteriorating as individual efforts to organize and search for information yield poorly from the organization standpoint since diffusion mechanisms are limited. We propose a personal and collective information management architecture in order to take advantage of individual efforts, and to manage documents in a collective and sustainable way. This is based on the integration of the document lifecycle activities depicting the way people manage information and documents. The proposed architecture exploits individual efforts through interdependent processes designed on a mutual benefit scheme. These processes rely on the annotation practice, considered as a representative evidence of the way individuals interact with information.*

KEYWORDS: *Collective IS, Annotation, Information-related Activities.*

1. Introduction and Motivations

Modern organizations such as companies, R&D labs, or communities increasingly rely on Information Systems (IS) as vehicles for the information relevant to their activities. In the same time IS cause informational overdoses: the organization is seldom capable of optimally handling the collected information as a whole. This issue is twofold as it arises at individual level and a collective level. Firstly, knowledge workers—people who mainly produce and work with information in the workplace; representing 31% of the US workforce in 1995, their proportion "will continue to increase significantly into the new millennium" (Sellen *et al.*, 2003, p. 51)—have a hard time identifying, finding, and keeping relevant information related to their activities. Secondly, they rarely distribute the information they introduced into the organization, although co-workers would benefit from it since one may assume that their needs are close, or even similar regarding their activities. The cost of this twofold issue is valued at a minimum of $33,000,000 a year for a hundred knowledge workers organization (Feldman, 2004). A solution to these problems should capitalize on collective information which remains static, unproductive, and scattered in the maze of the organization otherwise. The purpose of this paper is to illuminate the issues involved in collective activities related to information and its medium, namely documents. These activities coming from the document lifecycle are presented in Sect. 2 according to individual and collective levels, thus covering the issues stated beforehand. Then Sect. 3 introduces our proposition: an original architecture integrating and exploiting the document lifecycle activities. It is based on the personal and collective annotation practice as an inter-activity vehicle for information. The main idea is that annotating a document reflects individuals' cognitive efforts (e.g. learning, arguing, correcting) while interacting with documents. In addition, the architecture provides knowledge workers with personalized assistance thanks to the processes defined in this section. Moreover, the results of an activity improves the performance of the other ones, for the individual and for the group. Section 4 outlines experiments related to the proposed information management architecture and processes; it also describes the prototype system that implements this architecture. The prototype is demonstrated with Web documents, as a common source of information for any organization nowadays.

2. Current Issues of Common Document-related Organizational Activities

Within an organization, managing collective documents properly is a performance factor: it relies on the optimization of the various activities that facilitate the access to documents, and by extension to the corresponding information. The document lifecycle (Sellen *et al.*, 2003, p. 203) gives a comprehensive view of six major document-related organizational activities—noted from ① to ⑥—that knowledge workers achieve individually or collectively, supported by the appropriate software. *Marshalling and Extracting Information* ① relies on "pull systems" such as search engines and social bookmarking (Cabanac *et al.*, 2008). *Creating, Authoring* ② and *Finalizing Documents* ③ is powered by word processors with marking and annotation capabilities. *Distribution and Work Flow* ④ may be performed manually *via* emails, mailing-lists, and posts on the intranet or automatically when exploiting workflows, recommender systems (Montaner *et al.*, 2003), and social networks (Zhang *et al.*, 2005). *Using Documents* ⑤ mainly refers to active reading, i.e. critical thinking supported by informal annotations stemming from the common paper-based annotation practice (Marshall, 1998), e.g. remarks, comments, summaries. In the digital world, readers may use a software called annotation system, such as Annotea (Kahan *et al.*, 2002). Finally, *Filing and Archiving Documents* ⑥ consists in storing documents in a Personal Information Space (PIS), mostly for finding them later, building a legacy, and sharing them (Kaye *et al.*, 2006). People commonly classify documents into thematic folder hierarchies, e.g. file system, email client, bookmark hierarchy. This latter feature enables Web users to keep and organize interesting documents in a hierarchical structure that quickly evolves as people add, on average, three to four bookmarks per navigation session (Abrams *et al.*, 1998).

The study of document-related activities ① to ⑥ reveals a plethora of existing systems, may they be individually or collectively targeted. They intend to exploit collective documents by taking into account the experience, the expertise, the activities, the contacts, etc. of knowledge workers. Since individuals benefit from the group's activities, and as the group reciprocally benefits from individuals' activities, such systems seem to be in line with a mutual profit scheme. However, we notice that each system is highly specialized: it considers only one or two—at most—partitioned activities among the six activities of the document lifecycle depicted in (Sellen *et al.*, 2003, p. 203). Moreover, these activities seem to be linear, although we think that a person does not carry on document-related activities that way. On the contrary, people can obviously search for information ①, start to author a document ②, then go back browsing again ① for going into the subject in depth. As systems are specialized, each one only considers a small part of the real work of users while neglecting the four or five other activities. As a consequence, in our view, systems do not capitalize enough on knowledge workers' daily activities, thus leading to both limited sustainability and limited long term efficiency. In order to overcome the aforementioned issues, this paper puts forward an approach based on a federated and multiuser architecture. This intends to cover the whole activities of the document lifecycle. Unlike the linear and partitioned lifecycle proposed in (Sellen *et al.*, 2003, p. 203), we intend to integrate activities and to exploit their outcome. The ultimate aim is to help each individual, which is in turn beneficial to the group.

3. Personal and Collective Annotation-based Information Management

The study of document-related activities revealed two main issues. Firstly, the document lifecycle currently involves too many systems: people need to master at least six distinct applications, thus leading to cognitive overload. Secondly, each system is highly specialized since it is designed for a unique activity. This design results in scattered and partial user profiles, leading to poor adaptation and support. To tackle these problems, we propose an original architecture providing:

– **Personal support.** The architecture relies on a unified model that federates users and their six document-related activities, thus avoiding information and user profiles scattering. Along with dedicated processes described later in this section, this federated architecture helps individuals to find ①, to exploit ⑤, to organize ⑥, to author ②③, and to distribute ④ information. This design actually enables each knowledge worker to build up his own sustainable Personal Information Space (PIS) day by day.

– **Collective support.** Our proposal is a multiuser architecture that models knowledge workers within their organization. It exploits the constituted capital (knowledge workers' PISs) by implementing automated processes based on a mutual benefit scheme. The basic idea is that people implicitly contribute while achieving their common activities. They benefit in return by receiving information relevant to their work, which is automatically extracted from the activities of the group. This approach makes the organization-wide available information—extracted or contributed by the members—profitable for the whole group, whereas it is too often unexploited in people's desk drawers and computer file systems. As a result, our approach provides any organization with a sustainable information management.

The study of document-related activities in Sect. 2 showed that the annotation practice is already part of three individual and collective activities: authoring ②, finalizing ③ and exploiting ⑤ digital documents. This motivated our choice: placing the annotation practice at the heart of our architecture, as it can also cover the three remaining activities. Next sections define the "collective annotation" concept, and show how it also federates information retrieval ①, diffusion ④, and organization ⑥ through dedicated processes depicted in Fig. 1.

3.1. *Keeping and Organizing Encountered Information* ⑥

Creating a bookmark is a common way of keeping track of an interesting document (Abrams *et al.*, 1998). Many concrete work situations require keeping not only the document but also the passages of interest (e.g. a few sentences, a definition, a picture, a schedule) along with notes. Though, creating a bookmark is rather unsuitable for this task as it can neither point to parts of documents, nor keep any reader's notes. To overcome these limits we opted for the "collective annotation" concept as defined in (Cabanac *et al.*, 2007b). The UML Class Diagram in Fig. 2 formalizes this fundamental concept. It depicts a low level of detail as we intentionally hide the attribute and operation compartments of the classes for brevity concerns. We shall explain the class diagram using a concrete scenario where a User visualizes a Resource such as a Web page, a schedule on the intranet of his company, a picture in his file system. Whenever he wants to keep a piece of interesting information, he creates an *AnchoredAnnotation*. Two situations may arise. On the one hand, if the user wants to keep the entire resource, the system needs to store its location with a GlobalAnchoring. This mimics a classical bookmark storing the URLs of documents. On the other hand, the user may want to keep parts of a resource only, e.g. two non contiguous sentences. To handle this case, an alternative is to modify the resource by adding markers refer-

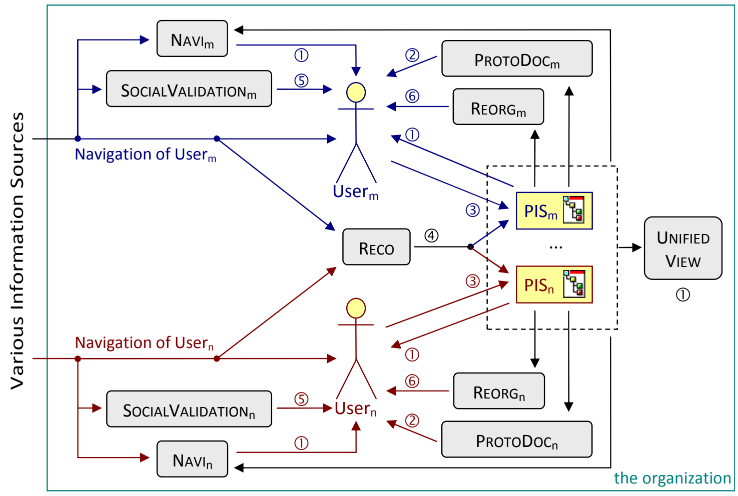

Figure 1. *Global view of the proposed architecture depicting an organization with two knowledge workers—it can be generalized to any number of users. Arrows represent data flows between the users, their PIS, and the six interconnected processes. Labels ① to ⑥ refer to the activities of the document lifecycle (Sellen* et al.*, 2003, p. 203).*

encing the beginning and the end of the user's selection. This only works when facing modifiable resources, which is a quite unusual case with public documents. That is why we opted for another alternative which consists of storing an anchoring point (*LocalAnchoring*) that locates the user's selection unambiguously. To handle the various resource formats, the *LocalAnchoring* abstract class must be refined; for HTML documents it can be an XPointer that expresses the selection path in the document object model (DOM) for instance.

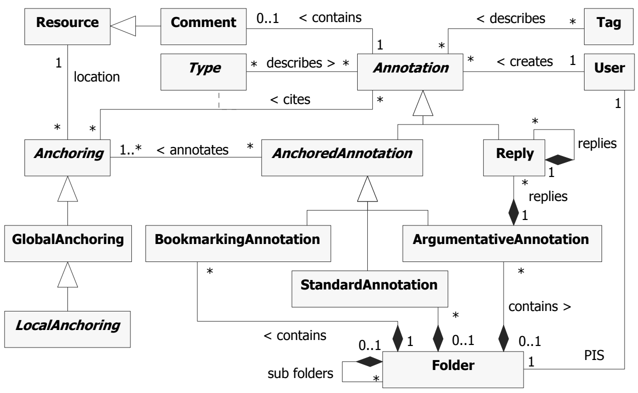

Figure 2. *UML Class Diagram of the proposed "collective annotation" concept.*

Previous classes model objective data that an annotation system infers from the document part selected by the user. The remaining classes model the subjective information introduced by the user who creates an annotation. Any *Annotation* may contain a Comment without restriction on the media or on the format, e.g. rich text, an audio recording. Being a subclass of Resource, a Comment or any part of it can be annotated in turn: this design allows the creation of recursive annotations. In addition to a comment, the annotator (i.e. the user creating the annotation) may indicate references to other (parts of) resources. Moreover, he can associate such citations as well as the annotation itself with *Type*s, thanks to the "$<$ cites" and "describes $>$" relationships. The *Type* abstract class describes the annotator's intent by giving an overview of its meaning: subclasses may cover taxonomies of objectives (e.g. comment, example, question), of actions (e.g. to do, to read), of opinions (e.g. refutation, neutral, confirmation), or any domain-specific concepts (e.g. business intelligence: partner, product, competitor, etc.). Providing organizational members with such taxonomies may help them to describe encountered information with a common ground, so as to improve their understandability. As opposed to assigning a predefined *Type* coming from a fixed vocabulary, the Tag class enables users to describe an annotation with their own words, as promoted by the social bookmarking approach, cf. Sect. 2. When it comes to annotating a resource, we discern three main user objectives that the proposed model takes into account: keeping and organizing information, note-taking, and discussing. Firstly, a BookmarkingAnnotation enables knowledge workers to keep and organize information. This kind of *Annotation* refers to the common bookmarking practice (Abrams *et al.*, 1998) while allowing a finer-grained anchoring: it is anchored to parts of a resource whereas a classical bookmark concerns the entire resource. In order to get access to his kept information, each User owns a PIS structured as classical bookmarks, i.e. a hierarchy of Folders. One reason to provide a hierarchy comes from the study (Jones *et al.*, 2005) which underlines that knowledge workers' need to classify information into hierarchies "to get things done." The second purpose of an *Annotation* is achieved by the StandardAnnotation class that enables to take notes on a resource without necessarily requiring its classification. Proofreading during activity ③ generates many annotations (corrections, misprints, etc.) that the annotator does not need to classify in his PIS; what is really essential for him and his co-workers instead is to view these annotations while re-reading the annotated document. Finally, the third purpose of collective annotation is to discuss in the context of the documents. Such a debate is initiated by an ArgumentativeAnnotation; later other readers may express their standpoints by formulating a Reply to this annotation, or to Replies recursively, thus forming a discussion thread similar to the Usenet ones.

Summing up the architecture design, a User creates an *AnchoredAnnotation* to keep encountered information. He may organize it the way he wants (e.g. by topic, by project, by date) within his personal Folder hierarchy (PIS). To achieve the automated processes mentioned earlier, we endow the proposed architecture with the additional model depicted in Fig. 3 (note that the two models complement each other; the latter is commented throughout this section). The proposed architecture is designed along with the six processes depicted in Fig. 1. Concerning the activity ⑥, the REORG process aims to reduce the high cognitive load involved when reorganizing a PIS (Abrams *et al.*, 1998). It takes as input the user's PIS for suggesting him a thematic classification, then the user can accept it partially or entirely. The proposed algorithm (Chevalier *et al.*, 2004) is based on a Hierarchical Agglomerative Clustering (Jardine *et al.*, 1971) which requires the Indexation of the Resource contents.

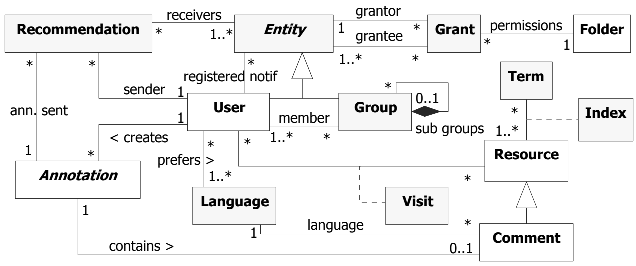

Figure 3. *Complementary UML Class Diagram for the proposed system.*

The previous section depicted the twofold role of an annotation. For *individual* activities its main purposes are to support critical thinking in-context (i.e. not on a separate sheet) while reading documents, and to keep interesting information in the reader's PIS. This is respectively achieved by StandardAnnotation and BookmarkingAnnotation. Regarding *collective* activities, annotations can be shared so as to support collaborative work: readers get previous readers' comments and feedback, they can participate in in-context debates through discussion threads as well (ArgumentativeAnnotation). One essential issue about annotation systems in general concerns their scalability. When displayed within documents at the exact place they were authored, they might disturb the reader. Empirical evidence shows that the difficulty in reading a document increases with the number of displayed collective annotations. For instance, the video `http://www.irit.fr/~Guillaume.Cabanac/annotation/demoAmaya.wmv` demonstrates this issue with the W3C Annotea/Amaya annotation system (Kahan *et al.*, 2002). We propose two complementary ways of reducing readers' efforts. As a first adaptation to User needs, our architecture hides any Comment that is not expressed in a Language they chose (Fig. 3). This avoids displaying utterly incomprehensible information to users. The second adaptation regards ArgumentativeAnnotations and the debates they may spark off. When many debates are anchored to a given document, the reader may consult each one in turn. Given a debate, deducing its participants' global opinion mentally enables to evaluate the "social validity" of the ArgumentativeAnnotation. Although necessary for critical judgment, this evaluation requires cognitive efforts while first identifying argument opinions, and then synthesizing opinions recursively upwards in the discussion thread. We intend to relieve readers of this burden thanks to the SocialValidation process that we defined in (Cabanac *et al.*, 2007b). It mimics individuals by synthesizing Reply opinions to obtain the social validity of the ArgumentativeAnnotation. This value is gradual as it ranges from "refuted" to "confirmed," it represents the global opinion expressed in the considered discussion thread. Given this process, the second adaptation we mentioned consists of informing users about the degree of consensus (resp. controversy) of each debate. As a result, they can focus on stabilized information, or on ongoing discussions where people have not found a consensus yet.

3.2. *Creating and Finalizing Documents* ② ③

Much of the explicit knowledge of an organization is held in the documents it produces, since knowledge workers spend a great amount of time authoring documents. A common task consists of extracting the essential facts from several documents, so as to synthesize them in a report (Feldman, 2004). Regarding the proposed architecture, *BookmarkingAnnotation*s are a sure way for a reader to collect nuggets of information along with his Comments and interpretations. For a specific project (e.g. Stock Exchange daily analysis) the user may create a dedicated Folder in his PIS gathering all the relevant annotations he wants to keep. When dealing with collaborative search and analysis, the implicated *Entities* (either Users or Groups of users) can create annotations and retrieve them from a shared Folder provided they obtained the appropriate Grant. In order to assist knowledge workers in harnessing the collectively collected information, the PROTODOC process drafts a document from any selected Folder of the PIS. This proto-document encompasses each annotation, the contents of its *Anchoring* within the annotated Resource, its social validity, and the information provided by the annotator (Comment, citations, etc.). The user can use his favorite word processor to complete and rework this draft afterwards.

3.3. *Improving Collective Information Retrieval* ①

The proposed architecture empowers the two classical modalities at the heart of information retrieval: searching and browsing. Regarding the search modality, we proposed to complement it by taking into account collective annotations in (Cabanac *et al.*, 2007b). The basic idea is to use readers' contributions, namely their annotations as a "social feedback" to improve IR recall (by retrieving more documents relevant to the query) as well as IR precision (by retrieving relevant documents only). Concerning IR recall, "the vocabulary problem" states that a user's query rarely ($< 20\%$) contains the same terms as the required documents (Furnas *et al.*, 1987). We suggest matching the query with annotations in order to indirectly find documents, and even passages when dealing with contextual IR—aiming to answer a query with passages instead of complete documents. Concerning IR precision, readers' Comments allow the disambiguation of annotated documents, and the integration of complementary terms in the indexing process. In addition, the social validity of discussion threads sparked off by ArgumentativeAnnotations may lead to characterize a Resource as trustworthy, controversial, popular, alive, outdated, abandoned ... Taking these indicators into account allows the adaptation of the search engine to users' preferences. As regards to the browsing modality, the NAVI process recommends documents to the user, provided they are relevant to his current navigation, see Fig. 1. This recommendation process is original in many respects. Firstly, since recommendations come from each organizational member's PIS, it exploits co-workers' ability to find and classify interesting documents. By doing so, long-time retrieved documents and then forgotten in individuals' folders are automatically proposed to other people. We put forward the hypotheses that *i*) a document inserted in a folder was considered interesting by his owner, since he achieved a cognitive effort for selecting the most appropriate folder in his PIS. Moreover, *ii*) other knowledge workers that share similar interests may be interested in such a document. A second original feature concerns the algorithm that matches recommendable documents with the one that the user is currently viewing. Classical methods are content-based: they recommend documents according to their text only. Conversely we proposed usage-based similarity metrics (Cabanac *et al.*, 2007a) to evaluate how closely knowledge workers use two given documents. Provided that people store together the documents they find similar (for any reason: their topic, their author, etc.), we stated that the more two documents get closely classified by the most people, the more they are usage-based similar. The proposed metrics is a tree walking algorithm that accesses knowledge workers' PISs. The NAVI process exploits it to recommend usage-based similar documents coming from co-workers PISs to each user during his navigation stage. The user may also view recommendations on a map representing the knowledge workers' documents organized according to the usage-based metrics, thanks to the UNIFIEDVIEW process detailed in (Cabanac *et al.*, 2007a). This allows the exploration and discovering of the capitalized documents that knowledge workers introduce daily into the organization.

3.4. *Distributing Collective Information to Knowledge Workers* ④

The study (Feldman, 2004) reports how a poor collective information diffusion leads organizations to a terrifically counterproductive and costly outcome: waste of time, information recreation, etc. We propose to capitalize on the information kept by knowledge workers, which often stays dormant in their computers otherwise. To do that, we offer Users the capability to send *manual* recommendations to the other *Entities* he knows. A dual feature enables Users to register for notifications concerning other *Entities*' documents. By doing so, one can proactively specify whose documents shall interest him, akin to Web syndication via RSS feeds. As underlined in Sect. 2 manual diffusion is limited by various human factors: the sender's social network, his willingness to share when information is commonly perceived as power, the implied cognitive efforts ... To overcome these limits, we propose to complement manual diffusion with *automatic* diffusion thanks to the RECO process as depicted in Fig. 1. It considers each document that enters the organization (i.e. retrieved by any member) as a candidate one for recommendation. As a result, it exploits collective information retrieval to help each member. In a word, the RECO process fully explained in (Chevalier *et al.*, 2004) works as follows. Each candidate document is first indexed, cf. the Term and Indexation classes. Then its thematic similarity with Users' PIS Folders is evaluated: each folder is represented as a classifier built by extracting features from its documents. Finally the candidate document is recommended in the folder having the best similarity value, provided that it exceeds a dynamic threshold. This threshold ensures that recommendations don't overload knowledge workers.

4. Ongoing Experiments and Prototype Development

As a first step towards validating the proposed architecture, we experimented with two processes among the six proposed ones in an organization of 14 researchers whose PISs contained 4,079 documents (resp. 486 folders) with an average of 291 documents (resp. 34 folders) by user. Five users were asked to evaluate the information recommended by the NAVI process as they browsed the Web, following a fixed navigation. This experiment detailed in (Chevalier *et al.*, 2004) showed that the more a user browses the Web, the more the recommendations he gets are accurate. We also experimented with the RECO process through the TREC 2001 OSHUMED/MeSH collection in order to compare different strategies for recommending information in a folder hierarchy like a PIS. Finally, we are currently experimenting with the SOCIALVALIDATION process depicted in Sect. 3.1. The purpose of this experimentation is to evaluate how close the proposed algorithm (Cabanac *et al.*, 2007b) is to human perception of consensus in argumentative discussions. We designed a protocol in compliance with the experimental psychology standards for Internet-based experimenting (Reips, 2002). An online Java WebStart software (cf. `http://www.irit.fr/~Guillaume.Cabanac/expe`) allows the participation of worldwide volunteers informed via mailing lists, e.g. the ACM SIGCHI. This experiment in progress launched in April 2007 has been raising 179 inscriptions, 118 of whom started and 51 finished the experiment. The contributed evaluations are currently under study.

A second ongoing work concerns the development of a prototype which implements the personal and collective information management architecture introduced in this paper. This proof of concept software called TafAnnote (cf. `http://www.irit.fr/~Guillaume.Cabanac/TafAnnote`) is designed on a two-tier model. The client side is a toolbar for the Mozilla Firefox Web browser that gives access to the supported features: annotation creation, discussion thread support, PIS management and annotation search. The client side communicates through a HTTP connection with the server side that stores annotations. Whenever a user requires a Web page, the browser retrieves its contents and sends its URL to the annotation server. Finally the client side merges each fetched annotation with the retrieved document model (DOM): annotations are displayed in-context according to their respective anchoring points. TafAnnote currently supports the HTML format by storing anchoring points as XPointers. Annotating different document formats is a challenge as specific anchoring techniques, i.e. subclasses of the *LocalAnchoring* are required. We addressed this problem in the decisional systems context by defining a dedicated anchoring technique (Cabanac *et al.*, 2007c) for the datawarehouse resources, namely multidimensional schemata and tables.

5. Conclusion and Future Works

This paper investigated how knowledge workers achieve the six activities of the document lifecycle (Sellen *et al.*, 2003, p. 203). We reviewed both individual and collective prominent approaches and systems dealing with each activity. This state-of-the-art study revealed two main issues. Firstly, the great diversity of systems implies data scattering and cognitive overload for the user, who has to master six systems in all, i.e. one per activity. Secondly, the activities are represented as linear and partitioned, although people do not behave this way. As a matter of fact each system is highly specialized in a unique activity, this leads to partial user modeling and limited support by extension. In order to overcome these issues, we proposed to federate the lifecycle activities into a unified multiuser architecture. Its ultimate purpose is to support each user for his daily document-related tasks. An original aspect of our proposal is that the organization is the real source of support. Indeed we exploit each knowledge worker's Personal Information Space (PIS) to help any user; such an individual assistance improves the performance of every user, which in turn improves the organization itself as a whole. The architecture we modeled is based on a key concept that we related to each activity: collective annotation as an evidence of knowledge workers' intellectual work. In addition we defined six automated processes represented in Fig. 1. They help each user to reorganize his PIS (REORG), to evaluate the social validity of argumentative annotations (SOCIALVALIDATION), to draft a proto-document from encountered nuggets of information kept thanks to annotations (PROTODOC), to discover collective documents relevant to his navigation (NAVI) or to his longterm interests (RECO), and to get a UNIFIEDVIEW of the knowledge workers' documents. This collective information management architecture is currently under experiment as the SOCIALVALIDATION process is the object of an Internet-based experiment rallying worldwide participants.

Perspectives for this work mainly concern further validating the proposed architecture and processes. We shall use methods from the cognitive sciences to investigate whether knowledge workers improve their efficiency in daily activities, and evaluate the trade-off between added constraints and benefits as proposed in (Millen *et al.*, 2003). Less HCI-related issues concerning digital annotation must also be addressed. One of them concerns the anchoring on various document formats, another deals with "robust" anchoring on evolving resources (e.g. modified documents), and a third one refers to scalability issues of the current client-server architecture as well as annotation visualization.

6. References

Abrams D., Baecker R., Chignell M., "Information Archiving with Bookmarks: Personal Web Space Construction and Organization", *CHI'98: Proceedings of the conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 41–48, 1998.

Cabanac G., Chevalier M., Chrisment C., Julien C., "An Original Usage-based Metrics for Building a Unified View of Corporate Documents", *in* R. Wagner, N. Revell, G. Pernul (eds), *DEXA'07: Proceedings of the 18th International Conference on Database and Expert Systems Applications*, vol. 4653 of *LNCS*, Springer, pp. 202–212, September, 2007a.

Cabanac G., Chevalier M., Chrisment C., Julien C., "Collective Annotation: Perspectives for Information Retrieval Improvement", *RIAO'07: Proceedings of the 8th conference on Information Retrieval and its Applications*, CID, May, 2007b.

Cabanac G., Chevalier M., Chrisment C., Julien C., Soulé-Dupuy C., Tchienehom P., "Web Information Retrieval: Towards Social Information Search Assistants", *in* T. Kidd, I. Chen (eds), *Social Information Technology: Connecting Society and Cultural Issues*, Information Science Reference, chapter 16, pp. 217–251, March, 2008.

Cabanac G., Chevalier M., Ravat F., Teste O., "An Annotation Management System for Multidimensional Databases", *in* I.-Y. Song, J. Eder, T. M. Nguyen (eds), *DaWaK'07: Proceedings of the 9th International Conference on Data Warehousing and Knowledge Discovery*, vol. 4654 of *LNCS*, Springer, pp. 89–98, September, 2007c.

Chevalier M., Chrisment C., Julien C., "Helping People Searching the Web: Towards an Adaptive and a Social System", *ICWI'04: Proceedings of the 3rd International Conference WWW/Internet*, IADIS, pp. 405–412, 2004.

Feldman S., "The high cost of not finding information", *KM World magazine*, vol. 13, num. 3, pp. available online `http://www.kmworld.com/Articles/PrintArticle.aspx?ArticleID=9534`, March, 2004.

Furnas G. W., Landauer T. K., Gomez L. M., Dumais S. T., "The Vocabulary Problem in Human-System Communication", *Commun. ACM*, vol. 30, num. 11, pp. 964–971, 1987.

Jardine N., van Rijsbergen C. J., "The Use of Hierarchic Clustering in Information Retrieval", *Information Storage and Retrieval*, vol. 7, num. 5, pp. 217-240, 1971.

Jones W., Phuwanartnurak A. J., Gill R., Bruce H., "Don't Take My Folders Away!: Organizing Personal Information to Get Things Done", *CHI'05 extended abstracts on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 1505–1508, 2005.

Kahan J., Koivunen M.-R., Prud'Hommeaux E., Swick R. R., "Annotea: an open RDF infrastructure for shared Web annotations", *Comp. Netw.*, vol. 32, num. 5, pp. 589–608, August, 2002.

Kaye J. J., Vertesi J., Avery S., Dafoe A., David S., Onaga L., Rosero I., Pinch T., "To Have and to Hold: Exploring the Personal Archive", *CHI'06: Proceedings of the conference on Human Factors in computing systems*, ACM Press, New York, NY, USA, pp. 275–284, 2006.

Marshall C. C., "Toward an ecology of hypertext annotation", *Hypertext'98: Proceedings of the 9th conference on Hypertext and hypermedia*, ACM Press, New York, NY, USA, pp. 40–49, 1998.

Millen D. R., Fontaine M. A., "Improving Individual and Organizational Performance through Communities of Practice", *GROUP'03: Proceedings of the international conference on Supporting group work*, ACM Press, New York, NY, USA, pp. 205–211, 2003.

Montaner M., López B., de la Rosa J. L., "A Taxonomy of Recommender Agents on the Internet", *Artif. Intell. Rev.*, vol. 19, num. 4, pp. 285-330, 2003.

Reips U.-D., "Standards for Internet-Based Experimenting", *Experimental Psychology*, vol. 49, num. 4, pp. 243–256, 2002.

Sellen A. J., Harper R. H., *The Myth of the Paperless Office*, MIT Press, Cambridge, USA, 2003.

Zhang J., Ackerman M. S., "Searching For Expertise in Social Networks: A Simulation of Potential Strategies", *Group'05: Proceedings of the international conference on Supporting group work*, ACM Press, New York, NY, USA, pp. 71–80, 2005.