

Extending Annotations to Make them Truly Valuable

P. Korica

Institute for Information Systems and Computer Media
Graz University of Technology, Graz, Austria

pkorica@sbox.tugraz.at

H. Maurer

Institute for Information Systems and Computer Media
Graz University of Technology

and

JOANNEUM RESEARCH

Graz, Austria

hmaurer@icm.edu

N. Scerbakov

Institute for Information Systems and Computer Media
Graz University of Technology, Graz, Austria

nsherbak@icm.edu

Abstract: Although the basic idea of annotations is rather old most web material does not allow for annotations. Even if it does, annotations are usually little used. However, research shows that annotations are potentially an important part of digital libraries and particularly eLearning: they can help users to learn more easily by (collaboratively) adding comments, links and bookmarks, thus personalizing web material to fit specific learning needs. In this paper we will present a new and more general approach to annotations that removes obstacles encountered with traditional annotations, increases their usefulness and provides substantial enrichment for digital material, particularly in eLearning systems. The simple but powerful idea is a strict separation of the actual content (web material) with all annotations plus other activities relegated to special workspaces. Such workspaces behave like a special layer attached to every 'unit' of digital documents. A unit can be a page of a document, a slide of a Power Point presentation, a picture or an audio file, etc. Any workspace can be used or ignored by all persons authorized to do so, but is invisible to all other users. Arbitrary many workspaces for arbitrary many groups of users can be attached to each unit.

The need for annotations

When using ordinary printed media like a book, university lecture notes, etc. we can write notes in the margins or between the lines of the document, we can highlight parts which seem important to us with e.g. a felt-pen, we can leave bookmarks, add sketches of drawings and much more. Knowing that this helps users with printed media it is surprising that such functions are rarely provided in connection with digital media. After all, we would expect digital documents to provide at least the same functionality as printed versions (Maurer 1997).

The idea to provide such functionality is quite old and has been implemented in some systems (Maurer 1996). Approaches allowing highlighting can be found even earlier in e.g. all versions of the so-called "PC Bibliothek" such as (Duden 1995). Thus it is interesting to note that the concept that annotations provide an ideal concept for collaboration of groups of users has not been widely understood, so far. We claim that this is because "littering of nice material" with personal remarks visible to all or a group of users is something few persons feel comfortable doing. By removing this barrier we believe annotations will become a valuable and much used tool.

Annotations are basically notes attached to documents. In first implementations, annotations were just textual notes regarding one document (Andrews et al. 1994). Soon systems started to provide much more

functionality (Maurer 1996). Nowadays it is possible to attach arbitrary multi-media documents and links, for example in Hyperwave Information Server (Hyperwave 2005) as annotations. In the web-based teaching system WBT Master developed and used at Graz University of Technology (WBT 2005a) it is possible to annotate any kind of material. Such material can consist of papers done by students, parts of a digital library, other user contributions, games etc.. To add a note (an annotation) clicking a simple "Comment"-button suffices. One can annotate using plain text or complex HTML encoding, can dynamically upload local files and attach them to annotations or link annotations with other objects of the system or link to an URL. Annotations in WBT Master can be modified by their authors (Helic et al. 2004b).

Annotations can be used as a kind of bookmark, too. Again, the analogy with printed material is helpful: if we want to assure that we can easily find certain pages of a book on re-reading, we often just write a list of all relevant page numbers on one of the first pages of the book. We can do exactly the same by writing the page numbers as annotations at the beginning of the document: indeed, the advantage of digital media becomes apparent: while in the printed book we have to turn to each of the desired pages, in the digital version the page numbers may well be clickable, eliminating the manual search for a particular page.

Annotations can be particularly useful for eLearning purposes: they facilitate learning for individual users (by highlighting or adding notes that will make the repetition of material easier), but even more importantly can be a powerful tool of collaboration. Users can add textual comments, links or some other multimedia content to a document that can be helpful for others.

We claim that by extending the concept of annotations a much more powerful system can be obtained that is indeed used extensively, as first experiments with WBT Master (Helic et al. 2004b) prove convincingly.

Past experience with annotations

In the past, the opportunity to annotate web material has only been provided by few servers, and where provided it has usually not been used to a large degree. There are a number of reasons that explain this fact. Note that an annotation can either be added for just the individual creating the annotation (usually called 'private annotation'), or for a certain well-defined group of users ('group annotation'), or for everyone having access to the information that has been annotated ('public annotation'). It is worthwhile to consider the three types in turn.

When a public annotation is made it can either be created anonymously (the server hosting the material does not know who is making the annotation), or the person performing the annotation is known to the server. Anonymous public annotations seem to be the obvious way to go for, since usually when surfing the web the identity of users is not known to the site at issue. However, in this case slanderous remarks, statements violating the law of some country etc. can be the result. This is not only undesirable, but also dangerous to the owner of the site at issue: in most countries the owner of a server is responsible for its content. (Indeed, in many countries even in the realm of printed material a similar situation holds: publishers of e.g. newspapers are usually responsible even for what is written in 'letters to the editor' even if it is stated quite clearly that 'the opinions expressed are not necessarily those of the owners of the publication'!). Thus, if anonymous public annotations are permitted some kind of mild censoring has to take place, and to be on the safe side even before the annotation is publicly available. This creates two problems: first, there must be persons responsible for looking at each annotation made before "releasing" it, and secondly annotations are not visible instantaneously, but possibly with a substantial delay. The lost spontaneity of annotations discourages many potential writers, and makes annotations for rapidly changing material (like in newspapers or such) meaningless. Thus, anonymous public annotations will only make sense if the material at issue is reasonably stable in time, and if one accepts the fact that many annotations will never be made since users would really want to see their comment immediately. As a consequence, few sites offer anonymous public annotations. Some manage, never the less, to create over time a reasonable amount of 'interaction by means of annotations' as can e.g. be seen by the large site (AEIOU 2005 and Maurer 1994) that contains a fairly static description of many aspects of Austria. It is worthwhile to compare the high-quality material resulting from a solid data-base with annotations that are edited in AEIOU (AEIOU 2005) with the more heterogeneous quality of Wiki-Systems (Wiki 2005) and with efforts in digital libraries (Dreher et al.2004).

The alternative to anonymous public annotations is to make sure that persons creating an annotation have to identify themselves to the system. The obvious obstacle in this case is: how can the system be sure that

an identification submitted is genuine? If the system just asks for an email or address or such, the information may well be a fake. Even worse, persons may actually use existing email addresses and create annotations that look like they come from a certain person, yet they come from someone else. Thus, the identification process is a complex issue in itself and will make the wide use of annotations, where authors are known to the system, difficult. The usual trick is of course to ask for an email, but to let the system generate a password and send it to that email. Unfortunately, if Person A wants to pretend to be person B, it is fairly easy to obtain an email account of the type Bxx@yy (where xx is some number and yy is one of the many free mail-service providers like Yahoo or Hotmail or such). Clearly, annotations written by some one with email address Bxx@yy can easily be misunderstood as being authentic opinions of Person B, although B has nothing to do with the entries at issue. Note that even if the identification process does work properly, issues remain: the owner of the site is still responsible for the content and is just slightly better of than in the anonymous case: if sued, the owner of the server can at least in turn sue the person having created the annotation! The situation is even more complicated due to the fact that some persons creating annotations may be hesitant to destroy the nice layout of web-page, or may not even want their name to appear! More on this has been discussed already a decade ago in (Flynn et al.1995).

At the very other end of the spectrum are private annotations. This clearly requires again identification, but in this case a valid username and a password may be enough: after all, the fact that a private annotation exists would only show up if a user is properly identified. It is thus our belief that private annotations should be allowed by all web servers, would help users and could also benefit the operators of servers: users could choose a username that does not give them away (i.e. the system would not know who is creating a private annotation) yet the system could still extract interesting usage information. We believe that private annotations would slowly be used by more and more persons as the public is getting used to this functionality. There is one inherent danger, however: operators of servers might be tempted to ask as username (or in addition to the username) for a valid email address, maybe even with the explanation that in this fashion the password can be sent if ever forgotten! All private annotations would then not only be visible to the system (they always are, of course) but can then also be associated with real living persons (or at least email addresses). This can lead to undesirable profiling. Observe in passing that private annotations are provided in a number of software packages from Power Point to many text processing and graphic packages!

In between public and private annotations are group annotations. We are aware that a number of systems such as Hyperwave (Hyperwave 2005, Andrews et al. 1994) and WBT-Master (Helic et al. 2004b) offer group annotations. The basic idea is that for every unit (e.g. page of a document) any user can choose an identity, and then invite others to join, creating a community of users that help each other by adding notes, links, etc. It has been pointed out before that such functionality is particularly useful for eLearning systems where users collaboratively enrich existing material to the extent necessary for their own purposes. Hence this should be one of the basic features of any eLearning environment. The drawback of the usual annotation mechanisms (where annotations become a visible part of the annotated material) is twofold: a pleasing layout of web pages or such can be destroyed by a multitude of group annotations, and links or inline images may shift in an undesirable way.

Summarizing, public annotations seem to be of limited use, hence the concept is not very visible when surfing the web. Private and group annotations should be supported much more than they currently are, yet if they are supported this should be done in the generalized form that we are proposing as the core of this paper, and which we describe in the next section.

The solution: Group work spaces attached to material at a fine granularity level

In what follows we assume that the material on the server at issue is divided into small 'units': the definition of what constitutes a unit is left to the author. Typically, a unit may be an HTML page, a page in a Power Point presentation, a page in some other kind of document, but units of even smaller granularity are conceivable. Such smaller units could be paragraphs of a page, pictures, figures, tables of numerical or other entries, etc. It is such units to which information can be attached, i.e. units can be, so to speak, annotated. However, in our system instead of annotations we use a more general approach: with each unit an arbitrary number of a separate workspaces corresponding to different user groups are associated.

Before we describe what such workspaces consist of let us first describe how such workspaces are introduced, and how they appear to the user.

With a single exception with which we deal at the end, our notion of group work spaces depends on the fact that users are identified. Identification usually occurs by just choosing a name and password (hence the system does not know the real identity of the user), or by using as name an email with an initial password assigned and sent by the system (hence assuring that the system and other users are aware of the identity of the person within the limits explained earlier). In both cases they also choose what is called a 'group name'. A typical group name might be 'Math 201 students' for a number of students of second year mathematics at a college, or 'Secretaries of xxx' indicating that a group of secretaries are collaborating on something, etc. There are also other ways of identification that we will deal with a bit later. Note that in the two identification variants mentioned, users remain on their own (i.e. only they will have access to certain functionalities) until other users are invited to join, such users having the possibility to decline or to accept the invitation. Of course the usual functions of changing passwords, of inviting further persons, or of deciding to quit a group, to rename a group etc. are provided for all users in identified mode.

Any identified user can be 'enabled' or 'disabled'. For disabled users, the material on the server appears as it would without annotations and without our group workspaces. The only option for disabled users is to either switch to anonymous mode or enabled mode. (In anonymous mode, the only option is to switch to identified mode). If users are enabled, material on the server changes its appearance in two ways.

First, each unit shows a button of the type 'Define work space', allowing to define a new workspace for all users belonging to a group with a name to be specified (this is necessary, since a user may belong to more than one group!). The typical implementation of this is as follows: when clicking at the button "Define work space" a list of all groups the user belongs to is shown, the desired group is selectable by one click. After creating a new workspace the user enters the workspace and has a number of facilities available described below.

Second, each unit shows a list of names of groups for workspaces created earlier. By clicking at one of the entries the workspace chosen is entered.

Once entering a workspace, three parts of the workspace become visible, the 'info-part', the 'question-part' and the 'discussion-part'. The info-part contains comments, links, or more complex documents as inserted here by other members of the group; new information can be added, if desired, using a simple special purpose editor. The discussion-part contains a discussion forum (Helic et al.2004a), if one of the users has started such a forum and can be perused by reading or adding to it as one would expect. The question part contains questions asked by other members: it allows to answer questions posed or to add new questions; all such questions have to relate to the unit at issue. Note that the question part is really a list of 'frequently asked questions': however, all the questions are relevant only to a small unit (hence keeping the granularity fine avoids long lists of frequently asked questions). The fact that thus the list of questions remains small assures that after a few users have asked questions concerning the unit at hand "all reasonable questions will have been asked", hence the system 'stabilizes' rapidly: new members of the group wanting to ask a question are likely to find their question already answered: thus, documents turn into 'active documents' as introduced in (Heinrich & Maurer 2000). Observe also that a long list of questions associated with a unit may well indicate that the unit at issue needs improvement, hence the system will send an email to the person responsible for the unit if a threshold parameter is exceeded.

It should be clear that material enriched in the way described is ideal for collaboration and for eLearning, but also gives a completely new twist to digital libraries. Suppose for instance that a paper (or single web page) on some geographical aspect of Austria is made available for all high school students in some grades of a cluster of high schools: it may well turn into a crystallisation point of very interesting discussions and collaborative efforts! This is certainly our experiences with a number of classes in computer science at Graz University of Technology, Austria.

Let us mention in passing that identification may also be handled by accepting all users coming from a certain domain or set of IP addresses as a group. Still more extreme, even anonymous users may be allowed to use workspaces: in both cases some of the problems described earlier concerning anonymous annotations of course do arise.

Finally, the space allocated for this paper does not allow us to discuss all issues to the depth needed, nor do we think have we solved all problems emerging, if our approach is used on a very large scale. Just to give an idea of the multitude of interesting situations arising let us mention a few.

It is not clear how to handle situations when a person belongs to a large number of user groups: in enabled mode more work spaces may become visible that might be desired. To merge all work spaces a user is involved in might be one solution: however, no successful way of merging discussion forums is known to the authors of this paper. An other alternative may be that users can disable certain groups they belong to at certain points in time. It may also be reasonable to not just have names for groups, but also for workspaces, and then allow to disable/enable some of them if desired. The role of the first person opening a group (or a workspace?) is unclear: should such persons have special privileges, like being able to edit contributions by others? When questions are asked in a question-part, it surely would make sense to send the questions to all people knowledgeable in that area. How to handle this best? Should teachers and tutors have a special role in such workspaces? Etc., etc.: only further experiments will show the right balance between usability and complexity!

Conclusion

Annotations are a very flexible and powerful tool for customizing almost all kinds of information. Unfortunately, their potential concerning digital media has not been realized to any extent, yet. We believe that a more general concept of annotation that separates content from material added later changes the situation dramatically to the better: our ideas presented in this paper and first experiments with systems such as (WBT 2005) are very definitely supporting this statement.

References

- AEIOU (2005) <http://www.aeiou.at> (2005-04-16)
- Andrews, K., & Kappe, F., & Maurer, H., & Schmaranz, K. (1994). On Second Generation Hypermedia Systems. *Journal of Universal Computer Science*, 0 (0), 127-136.
- Dreher, H., & Krottmayr, H., & Maurer, H. (2004). What we Expect from Digital Libraries. *Journal of Universal Computer Science*, 10 (9), 1110-1122.
- Duden (1995). Duden: *Die deutsche Rechtschreibung*. Dudenverlag, Mannheim.
- Flynn, B., & Maurer, H. (1995). Levels of Anonymity. *Journal of Universal Computer Science*, 1 (1), 35-47.
- Heinrich, E., & Maurer, H. (2000). Active Documents: Concept, Implementation and Applications. *Journal of Universal Computer Science*, 6 (12), 1197-1202.
- Helic, D., & Maurer, H., & Scerbakov, N. (2004a). Discussion Forums as Learning Resources in Web-Based Education. *Advanced Technology for Learning*, 1 (1), 8-15.
- Helic, D., & Maurer, H., & Scerbakov, N. (2004b). Knowledge Transfer Processes in a Modern WBT System. *Journal for Network and Computer Applications*, 27 (3), 163-190.
- Hyperwave Information Server (2005) www.hyperwave.com (2005-03-24)
- Maurer, H. (1994). The AEIOU Hypermedia Project. *Computer Animation'94*, IEEE Computer Society Press, 192-196.
- Maurer, H. (ed.) (1996). *HyperWave: The Next Generation Web Solution*. Addison-Wesley Longman, London.
- Maurer, H. (1997). What we want from WWW as Distributed Multimedia System. *VSM'97*, 1997, IEEE, Geneva. 148-155.

WBT Master (2005a) <http://coronet.iicm.edu> (2005-03-24)

WBT Master (2005b) http://coronet.iicm.edu/wbtmaster/courses/community01_library.htm (2005-03-24)

Wiki Wiki Web (2005) <http://c2.com/cgi/wiki?WelcomeVisitors> (2005-03-24)