

Enhanced Annotations

Harald Krottmaier
Institute for Information Processing and
Computer Supported new Media (IICM)
Graz University of Technology, Austria
hkrott@iicm.edu

Abstract: Usage of annotation features is common and implemented in many information systems such as digital library systems or learning environments. The Journal of Universal Computer Science (running since 1994) was one of the first digital library systems, which offered this interactive feature to their users. Unfortunately, features already available in modern knowledge management (KM) systems are not integrated in annotation facilities. KM-features will support users in the creation of annotations and they would make a much more structured discussion about the content of an article possible. In this paper we are going to show why the integration of a very common KM feature (similarity search) is necessary. It will be shown that this simple additional feature will enhance the process of annotating scientific papers dramatically.

Introduction

Annotations are usually small text documents attached to a published article. Ideally an annotation is attached to a fragment of an article. Thereby additional information about this specific part of the content is added. An annotation is a kind of *electronic post-it* and may start a discussion about the annotated part.

In a traditional environment with printed material annotations are also very common. Personal owned books, journals and especially learning material is “personalized” with markers, pens and bookmarks (Marshall, 1997).

There are much more advantages in the electronic environment than in the traditional one because annotations may be shared. Authors may get additional input from readers of their contributions. If these annotations are typed, i.e. marked with additional attributes such as *question* or *answer*, the author, but also other readers, may benefit from this feature.

Let us now explore some situations where authors as well as readers benefit from annotations attached to an article fragment. Let us assume the following types of annotations: *comment*, *question*, *answer*, *problem*, *solution* and *advice* as implemented in the Journal of Universal Computer Science (J.UCS, 2002) and some other annotation-aware systems like Annotea (Kahan et al., 2001).

Motivation

An article which has been published is often immutable (“dead documents”, Nürnberg and Hicks, 2002). As in many professional electronic publishing systems (especially in the scientific community) this annotation feature is the only way how an author may clarify a previously published article.

In a traditional – exclusively printed – environment there is no feature like this available! “Letters to the editor” may be used to clarify some previously published ideas. However, it is not very likely, that *all* users who read the original contribution will also read this specific letter to the editor.

Many publishers offer electronic discussion forums or news groups, where readers are able to discuss a publication. Unfortunately, this discussion is very often separated from the electronic content. There is no hint when reading the content that there is a discussion about the document or a specific part of the document.

On the one hand, many annotations of type *question*, *problem* or *advice* to a specific document may indicate that the document is not well-written. This fact should influence the author when writing upcoming articles and should improve the style of future papers. On the other hand, readers, who are going to work with an article with such kinds of annotations, would probably decide not to read the article first but start with the questions and problems.

The style of the article and the open questions obviously not answered in the article are usually a bad reputation of the article. This implicit fine-grade rating of articles is much more sophisticated than the common number-based rating system of articles available in many systems.

The visualization process of an annotated article must be very flexible. In addition to private annotations it should be possible to switch annotations on and off, list articles not just in alphabetic order but also by the implicit ranking of the articles.

Improvements

Tools for creating annotations must be easy to use. No additional software should be installed on the client side. Therefore it is quite difficult to annotate arbitrary document types. To give an example: PDF is very often displayed with Adobe Acrobat Reader. Unfortunately, it is not possible in this software to annotate PDF-documents. Although Adobe announced that upcoming versions of the software will support this feature, it will take time till the new version is widely available on every platform. Some ideas on how to overcome this problems are discussed in (Krottmaier and Helic, 2002).

In currently available systems it is possible to write arbitrary comments to articles. No quality-control is implemented in many systems, therefore any kind of annotation – even violations of the “netiquette” – may be published. Technically it is possible to check an annotation against a list of “black words” before it is published. If there is a match, an editor should review the annotation. User management is therefore a requirement for a system supporting annotations.

Before uploading an annotation, the system must check if there is a similar mini-discussion available in the system. Highly sophisticated similarity matching algorithms are already available for many platforms and should be used to support the user in creating annotations. A similar discussion available in the system or in some other part of the document or even in another document should be taken into account before this new annotation is published.

Beside this “simple quality control” many other mechanisms must be implemented to guarantee quality in the system. Users submitting annotations should be rated by other users. This mechanism is well known in discussion forums and it should be available in every annotation aware system, since annotations are a kind of discussion about a document or about document fragments. Highly rated users should be honored by the system.

Many publishing systems or learning environments provide content in different electronic formats. In the Journal of Universal Computer Science content is offered in PostScript, PDF and HTML. Since the contents are equivalent, a discussion in the PDF document must be synchronized with a discussion in the HTML formatted content. Therefore the annotation feature should be independent of the storage location of the documents and must reflect to the *contents* stored in these documents.

Annotations should also be active to the participants of a discussion. This may be implemented by sending emails to the authors of an article or readers of the article who subscribed to upcoming discussions or similar discussions. Ideally a document will answer questions automatically as described in (Heinrich and Maurer, 2000). This notification mechanism must be highly configurable, i.e. users should be able to stop notifications and review their user-profile in a regular manner.

Conclusions and Future Work

Electronic annotations are a powerful feature. Many implementations are already available. If many users work with this interactive feature the published material will automatically be improved over time. Therefore every involved participant – every author and every reader – will benefit from this annotation feature. This paper listed some additional necessary features to make annotations more powerful and to allow high-quality, peer-reviewed discussions about documents and document fragments.

An annotation feature is already implemented in J.UCS. At the moment it is possible to attach typed annotations (such as “Question”, “Answer” etc.) to content stored on the system. Authors of annotations may be informed by email (on request) of further discussions about this document. Currently we are implementing more granularity when attaching annotations to a document. Therefore it will be possible to annotate paragraphs, sentences and even words of documents. Quality control as explained in the previous section will be available.

References

Heinrich, E. and Maurer, H. (2000). Active Documents: Concept, Implementation and Applications. *Journal of Universal Computer Science*, 6(12): 1197–1202.

http://www.jucs.org/jucs_6_12/active_documents_concept_implementation.

J.UCS (2002). Journal of Universal Computer Science. <http://www.jucs.org>.

Kahan, J., Koivunen, M.-R., Prud’Hommeaux, E., and Swick, R. R. (2001). Annotea: An open rdf infrastructure for shared web annotations. In *Proceedings of the WWW10 International Conference*.

Krottmaier, H. and Helic, D. (2002). More than Passive Reading: Interactive Features in Digital Libraries. In *Proceedings of E-Learn*, Montreal, Canada.

Marshall, C. C. (1997). Annotation: From paper books to digital library. In *ACM DL*, pages 131–140.

Nürnberg, P.J. and Hicks, D.L. (2002). Models for Publishing Academic Hypertexts In *Proceedings of elpub2002*, Karlovy Vary, Czech Republic.