DOI:10.1145/3442186

Vinton G. Cerf

Half-Baked High-Resolution Referencing

IN THE PAST, I have written about digital preservation. I would like to turn to a related topic that I will call *high-resolution referencing*. In conventional print publication media, it is possible to cite books, chapters, papers, sections, pages, paragraphs, and even sentences. One reason this is possible is that these media *fix* the work indelibly. Of course, one must have the correct version of the publication in hand, so to speak, since pagination is a function of font size, for example. In the World Wide Web, the Hypertext Transport Protocol and the Hypertext Markup Language serve the needs of users to refer to Web pages and can do so with considerable precision by using features of extended URLs to reference specific sections of Web pages. URLs referencing anchor points within a Web page offer what I will refer to as a high-resolution reference. Of course, if the Web page has been changed, such references may fail with the too familiar “404 page not found” or similar error message.

In the world of Google Docs, and other document processing systems, it is often possible to keep track of the time sequence in which edits have been made so as to “undo” an action or to return to a previous version of the document. This leads me to wonder whether time resolution, in addition to space resolution, might be an interesting functionality to instantiate. A reason this may be of interest is Web page references are beginning to show up in print and other media with the annotation “retrieved <date>” included. While this information is helpful, a later reader may not find what the reference intended if the Web page has evolved since it

was referenced by the writer. One might imagine a construct in which the document (Web page, PDF?) includes timestamped edit information such that the version of the document at a given date/time might be reconstructed. Since editing can be a messy process, one supposes the writer, interested in capturing versions, might want to identify at what point a document should be “versioned.” This is not unlike existing mechanisms for keeping track of software versions by “checking out” and “checking in” versions of source code. This could become metadata for the document in the same sort of way that breakpoints and periodic backups allow for recovery to a known condition in a lengthy computation.

Assuming for a moment that this would be an interesting capability, it remains to figure out how to imple-

I wonder whether time resolution, in addition to space resolution, might be an interesting functionality to instantiate.

ment it for various cases. In the case of Google Docs, the internal representation appears to allow the document to be reconstructed in its entirety upon fetching, from its initial instantiation and subsequent editing. This suggests a versioning record could be as simple as recording a date/time at which the document is at “version X” for some value of X. A reference to “version X” of the document would reconstruct all edits up until the date/time at which version X was “marked.” It seems equally feasible to export a document in a variety of formats including Web page HTML including an indication of which version it represents.

It is not clear to me whether one could incorporate such time-based mechanisms within an HTML or PDF document without incurring either overhead for generating and storing every “version” or reconstructing the entire object every time the object is retrieved as happens with Google Docs. Assuming that time or version-based citations are feasible and useful, there comes the question of how to generate the references. Generating these citations sounds like a non-trivial exercise and tools are emerging to assist authors with the generation of citations and for readers to use them. One set of tools created by Frode Hegland and his collaborators can be found at <https://www.augmentedtext.info>.

I am sure readers of this column will have a lot to teach me about floating half-baked ideas.

Vinton G. Cerf is vice president and Chief Internet Evangelist at Google. He served as ACM president from 2012–2014.

Copyright held by author.