

Enhancements of Personal Information

David Maier, Alon Halevy, Marcia Bates, Harry Bruce, Ben Bederson, Mel Knox
(Recorder)

Personal information as initially encountered can often be very raw, fragmentary, or partially relevant; it may come from disparate sources with differing format and structures. Hence there have been many proposals to enhance it in various ways to make it more useful for the task at hand, to improve later findability, or to record and reuse human analysis and judgment connected with it. Enhancements typically involve adding more data to personal information or adding links between previously unconnected pieces of personal information, but can involve deletions or removal of extraneous relationships. This breakout group discussed the kinds of enhancements that have been considered (or should be considered), the variety of reasons for enhancing personal information, and the issues that arise in devising enhancement methods. We recount our discussions on each in turn.

Examples of Enhancement

Enhancements to personal information can be done on individual pieces of information, or on collections of data. They can be performed (semi-) automatically by some system or manually by a user.

We first consider examples of enhancements to individual documents. The most common example of an enhancement to a single document is to annotate it. For example, we may add some annotations on the content of a particular photo, or comments on the context of a particular file in our directory. Judgments and rating of documents can also be useful for later recall and ranking in searches. Annotations can be attached to a document as a whole, or to some passage or other information element within a document. An annotation may be as simple as highlighting a phrase, or as complex as interlinear markup of a foreign text with pronunciation, literal translation and idiomatic translation. Summarization is also a form of enhancement. Summaries of personal information items can later provide us quicker ways of recalling their content. Another important example of enhancement is providing the lineage or provenance of a particular document (or internal element of a document). The lineage may point to the document from which it was derived, the method by which it was derived, when it was created or modified, the instrument it was captured with (e.g., camera), or other activities that were being performed in parallel with the one relevant to the document at hand. Of course, cleaning data items (e.g., fixing name spellings) can enhance their quality later on, as can stripping headers or formatting that are no longer relevant.

A second class of enhancements involves adding links between related items of personal information. One example is linking the entry in one's contacts file to emails from the particular person, or even to the papers co-authored with that person. Such annotations enable use to easily cross application boundaries when we browse our personal

information (in the spirit of the Memex vision). The group speculated on linkages between digital items and physical artifacts, so that, for example, rearranging sticky notes on a wall reorganized corresponding elements in an outline, or corrections written on a paper printout are reflected in the electronic version of a document.

Enhancements that involve collections of information can come in at least two varieties. The first is clustering data items in a semantically meaningful way. For example, grouping one's email into different *activities*, or clustering all the photos that involve a particular person, place or event. A second class of enhancements has the goal of improving the efficiency of locating particular data items. These will include novel indexes on the data or access methods for browsing the data (e.g., a new "virtual" directory structure that brings together information from multiple other directories). Adding a glossary of definitions used in a particular data set can also significantly enhance its readability in the future.

Finally, a more exotic kind of enhancement endows data items with dynamic capabilities, giving additional functionality beyond the provided by the applications for creating or viewing them. As one example, you may consider using data items as input an engine that selects advertisements that are likely to be relevant to you based on your personal information. A different kind of example is to add computational power to email. For example, an email message request votes or selections from the receivers, and provide capabilities for the receivers to vote and to automatically tabulate their replies. One can also imagine a module that hides or selected portions of a document depending on who is viewing it.

Reasons for Enhancing Personal Information

There are multiple reasons for enhancing personal information.

1. **Improve the quality of the information:** Information as received is often raw or unvetted and needs to be enhanced to make it more easily accessible and understandable by its users. Examples include cleaning data, excerpting from a web page, summarizing a document and trimming headers or inclusions from an email message.
2. **Reminding qualities:** Enhancements to personal information can help the user remember what they were doing when they first inspected or created the data, thereby providing context for (re)locating other relevant data.
3. **Efficiency:** Adding index structures enables more efficient search into our personal information. A different form of efficiency is to provide additional access paths to data by creating virtual directories (or other forms of super-imposed structure) that can support browsing or navigation.
4. **Add missing information:** It often happens that certain information is simply missing or would simply not be part of the application originating the data, such as annotations on photos or lineage information. Similarly, adding links between disparate data items fills in gaps in our personal information, such as connecting two email addresses used by the same person.

5. **Repurpose the information:** Our personal data can often be leveraged for different tasks than were originally intended, and it is often necessary to enhance the data so it is relevant to other tasks (adding keywords or ratings, making document structure explicit, XXXselecting subsets).
6. **Record results of human analysis:** an obvious reason to enhance personal information is to record the results of machine or human analysis and then carry them forward with the data (e.g., cleaning references to people or articles in one's personal information, marking questionable sections of a web page, flagging useful functions in a user guide).

Issues for the Mechanics of Enhancements

We discussed several issues regarding the implementation of enhancement modules and architectures for incorporating them. The following are some of the issues that came up and suggested principles for building enhancement tools. Note that not all of these are relevant in every enhancement context.

1. The enhancements should be available with the data later on – i.e., they “move with the data.”
2. We should be able to access or reconstruct the original (un-enhanced) data.
3. Addition of enhancements should link data items across different media, and should not end at the edge of the screen.
4. Enhancements should be optional: Their creation should not necessarily be completely automated, and a user should always be able to reject a proposed enhancement.
5. Like the data itself, the enhancements should be searchable.
6. Enhancements should consider the internal structure of documents (e.g., spreadsheets, email messages), in order to associate with the appropriate granularity (column, cell; header field, body paragraph).
7. There is a need for a global scope for information beyond the actual data elements themselves. For example, we need to be able to store a link between a paragraph in a PDF file and a row in a database, even though the link is not stored with either of these data items.