From Implicit to Explicit Knowledge: A Tool for Preserving and Sharing Mental Links in Science

Philipp Andelfinger
Steinbuch Centre for Computing
Karlsruhe Institute of
Technology (KIT)
76131 Karlsruhe, Germany
philipp.andelfinger@kit.edu

Matthias Keller
Steinbuch Centre for Computing
Karlsruhe Institute of
Technology (KIT)
76131 Karlsruhe, Germany
matthias.keller@kit.edu

Holger Kühner
Steinbuch Centre for Computing
Karlsruhe Institute of
Technology (KIT)
76131 Karlsruhe, Germany
holger.kuehner@kit.edu

Hannes Hartenstein
Steinbuch Centre for Computing
Karlsruhe Institute of
Technology (KIT)
76131 Karlsruhe, Germany
hannes.hartenstein@kit.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
CSCW'14 Companion, Feb 15-19 2014, Baltimore, MD, USA
ACM 978-1-4503-2541-7/14/02.
http://dx.doi.org/10.1145/2556420.2556476

Abstract
In this paper we propose MentaLink, a tool for preserving and sharing mental links in science. MentaLink is intended as a public knowledge base of links between scientific publications. Contributors can collaboratively define and edit links between entire articles or specific text passages. MentaLink builds on the idea of typed links, allowing authors to explicitly specify how the publications relate to each other, e.g., whether they share the problem statement, whether one publication is built on another or whether they contain contradictory results.

Author Keywords
Collaborative knowledge base; mental links; scientific publications

ACM Classification Keywords
H.5.4. Information interfaces and presentation (e.g., HCI): Hypertext/Hypermedia

Introduction
"*There is a growing mountain of research. […] The investigator is staggered by the findings and conclusions of thousands of other workers— conclusions which he cannot find time to grasp, much less to*

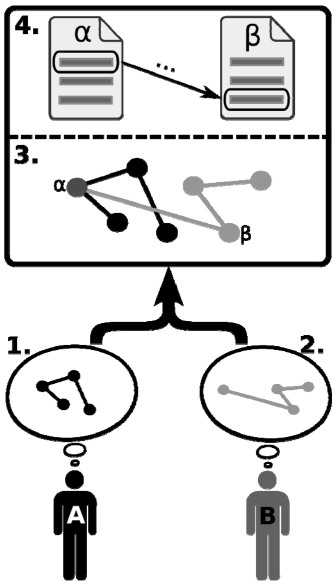

Figure 1 – Combining researchers' mental models of the relationships between publications in a given field to build a shared knowledge base.

remember, as they appear." Vannevar Bush wrote more than 70 years ago [1]. To solve this problem, he proposed a machine that allows associative linking of documents. His vision of the MEMEX apparatus paved the way for the concept of hypertext and, eventually, the WWW. However, the problem he describes is still unresolved today. And, just like Vannevar Bush, we argue that a tool for managing associative links would increase the efficiency of research. Mental links play an important role in the research process: *why do we participate in scientific exchange and attend conferences*? Because we benefit from learning about what others do if we can link their work mentally with our own. *Why is finding relevant literature difficult and time consuming?* The answer is that it is hard to identify the links between our own work and the work of others. Every researcher builds her own mental model about how the relevant publications relate to each other in his field and sees only part of the full picture.

Concept

We propose MentaLink, a tool for preserving and sharing mental links in science that combines different aspects of existing approaches into one solution:

First, it allows researchers to share user-generated links between scientific publications. In contrast to, e.g., Zotero [2], specific text passages can be linked besides entire articles.

Second, it implements the idea of typed links [3], allowing the community to explicitly specify how the publications relate to each other, e.g., whether they share the problem statement, whether one publication is built on another or whether they contain contradictory results. Compared to the scientific references defined by a publication's authors, a more fine-grained and complete network of associations will evolve.

Third, it is a collaborative knowledge base that facilitates achieving consensus by providing voting and discussion tools, similar to Wikipedia.

Fourth, it is a tool for visualizing relationships between publications in a research field as a graph (like, e.g., [4]) that can be traversed and explored.

We illustrate our vision based on an example (cf. Figure 1). Consider a researcher A reading a research paper she is not yet familiar with. While reading the paper, she will typically put the paper in the context of other works based on other literature already known to her and on references made by/to the paper.

In effect, the researcher is building a mental model of the relationships between publications within the research topic at hand (1.). This model goes far beyond what is visible through the reference graph alone: first, relationships may not be reflected by explicit references, e.g., if the original authors were not aware of the publication during writing. Second, documents may be related in a variety of ways: e.g., for disputing a claim, as support for an argument, or by shared methodologies. Third, relationships in the mental model may pertain to specific passages of a document instead of the whole document.

As the researcher's mental model of the relationships between publications in a given field grows more detailed, it becomes easier and easier to contextualize additional publications.

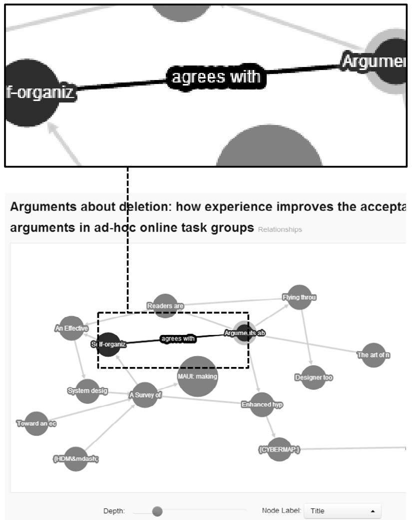

Figure 2 - the graph view shows user-contributed relationships between publications.

Consider now a researcher B exploring the same field of research. Usually, the researcher will duplicate much of researcher A's work when building his own mental model of the research field (2. in Figure 1). However, he may also be aware of relationships and relevant publications not part of A's mental model. Our aim is to allow A and B to combine their individual mental models to form a shared knowledge base. Based on the envisioned system, further researchers investigating the given field will be able to explore a graph-based overview of the relevant literature and its inter-relationships (3.) and to view the full text of pairs of documents side by side and explore their relationships in detail (4.). We hereby aim to allow researchers to build on the existing knowledge. As researchers add more and more relevant publications and their relationships the value of the shared knowledge base increases.

In addition to the basic facilities described above, we state a number of challenges that are brought up when the system should be deployed:

C.1) Researchers must be able to trust in the validity of the displayed relationships between publications.
C.2) The system must respect the copyright of all publications.
C.3) As most scientific publications are available in a read-only PDF format, relationships cannot be added to the document itself. Still, relationships should be anchored on the document in a way to allow for text-based processing of marked passages.

Implementation

MentaLink is a tool which allows researchers to manage relationships between publications collaboratively. Our prototype of MentaLink is implemented as a web application and offers functionalities to explore relationships between publications in a graph view, to upload publications as PDF files, to view and contribute relationships by using a dual PDF view and to discuss and rate relationships.

The graph view (cf. Figure 2) shows an overview of multiple publications interconnected by relationships which were already entered by the MentaLink community. The view serves as a starting point for research as it allows viewing all relationships of a paper of interest at a glance. A "zoom-out" function - also showing papers which are more than one hop away – helps the researcher to explore the environment of a publication. The node size indicates how many relationships for a publication were already provided by the community.

The dual view (cf. Figure 3) enables researchers to add and view typed relationships between two publications on a per text-block basis. This view shows the source and target publications of a relationship side by side. Source and target text blocks of a relationship are highlighted in the same color. To add a relationship, the researcher marks the source and target text blocks and indicates the relationship type.

MentaLink addresses the challenges stated in Sec. "Concept": The option to up- or downvote and discuss a relationship helps researchers to estimate the validity of this relationship (C.1). To avoid copyright infringement, a researcher has to prove possession of a publication by uploading it before he is able to view it in the dual view (C.2). Source and target text blocks of a relationship are stored separately from the documents

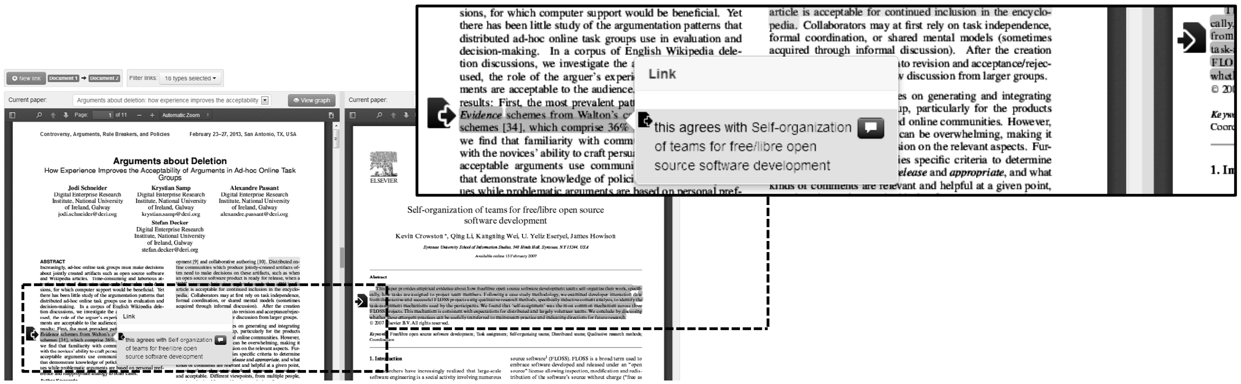

Figure 3. A link of type *agreement* between two publications in the dual view

itself (C.3). MentaLink identifies text blocks by storing and matching their full content; additionally, rectangular areas can be defined within a document to refer to non-text elements like figures.

Discussion

Although the MentaLink prototype implements the essential features required to be useful, when considering real-world deployment, a number of issues remain to be addressed.

First, in our experiments and the example cases presented here, a relatively small number of publications and corresponding relationships are visualized in the graph view. In a real-world setting where there might be large numbers of relevant publications and relationships, it might be necessary to allow users to limit the number of visible entities based on, e.g., the number of hops from a reference document or a relevance measure based on vote counts.

A second issue is the question of reaching critical mass w.r.t. to user counts. Initially, the system contains little information and users will have to exert effort populating the knowledge base with little value to be gained in return. A possible way of approaching this issue is by initially providing the plain reference graph as a basis for further refinement by the users. The reference graph could be gathered automatically either from the publications themselves or from existing databases on the web.

Acknowledgement

The MentaLink prototype was developed in a practical course at the Karlsruhe Institute of Technology (KIT) in the summer term 2013. We would like to thank the following students, who also contributed many ideas: André Becker, Andreas Bender, Leon Burkard, Daniel Deckers, Alexander Eckert, Simon Flaig, David Rieger, André Simmert, Florian Stolz, Martin Sturm, Felix Zapf-Schramm

References

[1] Vannevar Bush, "As We May Think," *Atl. Mon.*, Jul. 1945.

[2] http://www.zotero.org/.

[3] D. Shotton, "CiTO, the Citation Typing Ontology," *J. Biomed. Semant.*, vol. 1, no. Suppl 1, pp. 1–18, Jun. 2010.

[4] C. Chen, "CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature," *J. Am. Soc. Inf. Sci. Technol.*, vol. 57, no. 3, pp. 359–377, 2006.