Communities of Practice and Semantic Web Stimulating Collaboration by Document Markup

Christine Müller

Department of Computer Science, University of Auckland, New Zealand
Computer Science Department, Jacobs University Bremen, Germany
c.mueller@jacobs-university.de
http://kwarc.info/cmueller

Abstract. We believe that mathematics is the language of science and has paved the way of many innovations. However, mathematical research is often said to be "non-practical" and "hard to digest". Furthermore, experts have access to highly specialized results, but are often less aware of applications outside their own community.

Our work promotes the exchange of *highly-specialized knowledge* between individuals with different mathematical background. We draw on modern *representation formats* to mark up structure and meaning of mathematical texts to reduce *barriers of communication*. Moreover, we integrate modern web technologies to build an adaptive, active, and collaborative environment, in which users engage in a *community of practice*.

Keywords: content markup, communities of practice.

1 Introduction

Mathematics is one of the oldest disciplines and the basis for most scientific and industrial innovations; mathematicians have paved the way to new scientific inventions allowing others to develop mathematical methods further and to provide a practical use. However, mathematical research is often said to be "non-practical" and "hard to digest" as many struggle to interpret mathematical writings. Furthermore, mathematical experts have access to highly specialized results, but are oftentimes less interested and aware of applications outside their own community. Although interdisciplinary collaborations increased over the years, most communication of mathematical knowledge is based on documents that solely present the solutions of problems; mathematical publications do not provide detailed insight in methods, deadlocks, and practices involved in the problem solving process or further examples and illustrations. Oftentimes, mathematical results are only partial articulation; many details are left out and only reside in the expert's mind. Professional mathematicians can rediscover the missing steps and resolve ambiguity by drawing on their mathematical background and experiences, while less experienced readers and novices to the respective mathematical field may struggle.

Our work promotes the exchange of *highly-specialized knowledge* between individuals with different mathematical background by improving the access to

J. Yang et al. (Eds.): UNISCON 2009, LNBIP 20, pp. 432–437, 2009.
© Springer-Verlag Berlin Heidelberg 2009

technical documents as well as stimulating online collaborations. We draw on *mathematical content markup formats* to mark up structure and meaning of technical texts to reduce *barriers of communication.* Moreover, we integrate modern web technologies to build an adaptive, active, and collaborative environment, in which users engage in a *community of practice.*

2 Methods

Communities of Practice (CoPs) [24] are groups of people who share an interest in a particular domain. By interacting and collaborating around problems, solutions, and insights they develop a shared practice, i.e. a common repertoire of resources consisting of experiences, stories, tools, and ways of addressing recurring problems. The theory assumes that learning is a *collaborative activity* rather than the reception of factual knowledge or information. By participating in a community of practice novices gradually increase their engagement with the community and expertise, while moving from the outside towards the inside.

Content-Oriented Representation bridge (fully formal) mathematical input languages and presentation markup languages. Formal languages, such as CASL [1], enforce full formalization of knowledge but provide sophisticated services, such as automatic program verifications or proof checking. In contrast, presentation-oriented languages, such as TEX or LATEX, provide informal and flexible ways to express knowledge concepts but only facilitate limited machine support such as type-setting or keyword search. Content-oriented representation formats for mathematics, such as MATHML [23], OPENMATH [17], sTEX [11], or OMDOC [10], do not enforce full formalization but are more tedious to write than presentation-oriented languages. Content markup *explicates the structure and meaning of documents*, which does not facilitate automatic proving or verification but much stronger services than informal formats (see Sect. 5). This work builds on the XML-based, web-scalable *Open Mathematical Document Format* (OMDOC [10]), which serves as *content markup format* and *ontology language* for mathematical documents on the World Wide Web.

3 A Case Study on Community of Practice

The theory of *community of practice* allows us to focus on the *social* and *practice-oriented* nature of mathematics. Despite a common belief that mathematical practitioners prefer isolation and self-study, we observed that they are *highly collaborative* and *active in their community.* Mathematical collaborations are essential for any stage of mathematical practice: from identifying a challenging problem, becoming acquainted to new problem-solving methods, up to the verification of results and peer-review of publications.

We also observed that mathematics is divided into several *heterogeneous* sub-communities. Although outsiders may get the impression that mathematical practitioners form a homogeneous, unified community and share the same practices all over the world, they actually form various sub-communities that differ in their *preferred notations*, *basic assumptions*, and *motivating examples*. We believe that different mathematical practices hamper understanding and constitute *barriers of communication*.

Our case studies are a mixture of literature analysis, student group discussions, and interviews with professional mathematicians of various disciplines. For a general understanding of mathematical practice we drew on literature such as [8,4,3,19,21]. Further research focused on specific practices such as the choice of mathematical notations [22], basic assumptions and logical foundations [20], and the choice of typical examples [7]. We have gained intuitions on notations based on a web survey and an analysis of our Computer Science course materials with student volunteers (see [14]). We are currently conducting interviews with experts from various mathematical fields (including randomness, complexity theory, computability, and group theory) to gain insights in potential add-on services throughout the document life cycle, i.e. the writing, review, publication, search, and finally study and reuse of mathematical publications.

4 Novel Techniques for Technical Communications

Based on our observations in Sect. 3, we distinguish two types of services: Services for *reducing barriers of communications* (Sect. 4.1) and services for *stimulating technical collaborations* (Sect. 4.2).

4.1 Services for Reducing Barriers

We believe that different mathematical practices, such as mathematical notations and background assumptions, hamper communication, even among professional mathematicians. For example, C_k^n (Russia), C_n^k (France), and $\binom{n}{k}$ (Germany) denote the same concept, while $\mathbb{N}$ can be defined as *the set* $\{1, 2, 3, \ldots\}$ (positive integers in number theory) or *the set* $\{0, 1, 2, 3, \ldots\}$ (non-negative integers in set theory and computer science). The first example can cause unnecessary misunderstandings as users are actually referring to the same concept (the binomial coefficient) but know different presentations. The second example can cause pitfalls as users may believe to talk about the same concept but actually define it differently. Not knowing the proper constraints and side effects of a concept can cause inconsistent reuse as well as error-prone applications.

Markup of the *implicitly inscribed* structure and meaning allows us to *explicate practices* involved in the production of mathematical texts, potentially, reducing *ambiguities* and *barriers*. In the following, we illustrate how the markup of mathematical notations can facilitate the adaptation of alternative presentations and increase the *accessibility of documents* for the (novice) reader.

The structural markup of mathematical notation is specified in the two webstandards, OpenMath and MathML, based on which we specified a framework for the context-aware configuration of mathematical notations [9]. In [16] we have extended our approach with user (and community) modeling techniques to facilitate system-driven adaptation based on the readers' profiles and situations. The notation framework has been implemented in the Java library JOMDoc [6] and will be integrated in our web-based reader *panta rhei* (see Sect. 5). During the adaptation of notation JOMDoc can enrich technical documents with interactive triggers, which can be interpreted by the Javascript Framework JOBAD [5] to provide an (inter)active reading environment.

The study of mathematical notations has revealed insights into other scientific and industrial scenarios, which we can now better understand and support. For example, consider a *needs requirements life cycle*: Specifying what a project should accomplish can be challenging. It's easy to misunderstand requirements if they are not clearly articulated. We do not claim that we can manage all factors of the requirements gathering process, but we can reduce some of the pitfalls. For example, based on the markup of technical concepts, we can explicate whether a client, salesman, or programmer refer to the same requirements or whether they denote different concepts similarly. Vice versa, we can identify whether varying labels actually refer to the same knowledge concept.

4.2 Stimulating Technical Cooperation

Adaptation of documents and explication of meaning can reduce misunderstandings, but do not necessarily stimulate cooperation and learning, which is defined as a *collaborative activity* that requires participation in a *community of practice* (see [24]). Traditional mathematicians prefer face-to-face collaborations with personal brainstorming and discussion using blackboard and chalk. Nevertheless, with the increase of global collaborations, more and more mathematicians draw on modern information technology. However, online communication is still tedious due to the nature of the mathematical language, a mix of formal notations and natural language text (see [16]). Mathematicians thus call for more efficient discussion and collaboration spaces as well as review and publication facilities.

5 Implementation

In [15], we introduced the interactive and collaborative reader *panta rhei* [18], which facilitates readers to challenge, discuss, and rate online documents. Based on the sTEX-2-OMDoc-2-XHTML workflow, authors can write their documents in their preferred LaTeX editor (using the semantic sTEX package [11]) and publish their results in the web reader: LaTeXML [13] provides the conversion from sTEX to OMDoc, while XSLT stylesheets [10] facilitate the transformation from OMDoc to XHTML. During the conversion semantic identifiers and metadata (the content markup) are preserved, which improves the web-accessibility of the imported documents. Markup of narrative structure allows us to adapt the size and navigation of documents (see [12]), markup of concepts allows semantic search and easy cross-linking, enhancement with action triggers facilitate interactivity (see [5]), while distinction of content and form supports different visualizations, such as the adaptation of notations (see [9]).

The *panta rhei* system integrates the imported contents with easy-to-use web facilities. Displayed documents are "read only", but all users can input their opinions, questions, and answers via a forum and annotation facility. Moreover, implicit and explicit user modeling techniques are applied to personalize the adaptation of content (see [16] for details). Fig. 1 displays the two constituents of the system: *panta* (the user interface) and *janta* (the backend service). *panta* implements the discussion, annotation, and tagging facilities and gathers information on the user's notation preferences, while *janta* takes over all the content and user data handling as well as adaptation (please see the *panta rhei trac* [18] for details).

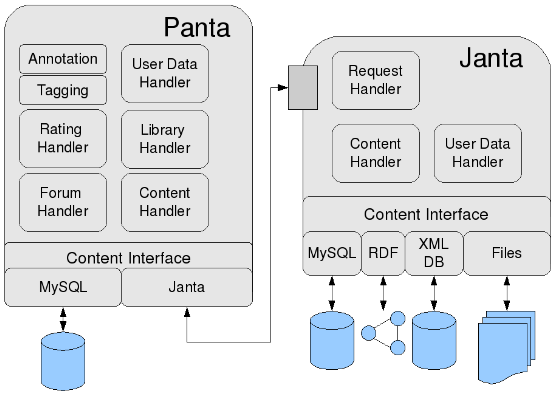

Fig. 1. System Architecture

The usability of the *panta rhei* system will be evaluated based on the *Cognitive Dimensions approach* [2].

6 Conclusion and Outlook

Our work promotes the exchange of technical knowledge by reducing barriers of communication and by providing a web-based collaboration space. We build on the semantic representation format OMDoc, which marks up the structure and meaning of mathematical knowledge and improves its searchability, display, and personalization on the Web. We integrate the rich representation layer (of OMDoc material) with discussion, interactivity, and animation facilities to provide an adaptive, active, and collaborative environment, in which users engage in communities of practices to exchange technical knowledge more efficiently. We are developing a prototypical implementation to illustrate and evaluate our approach. Both, conceptual work and implementation, are still on-going. We will soon release a proof-of-concept prototype of the *panta rhei* system, based on which, we can start our evaluation.

Acknowledgments. We would like to thank Andrei Aiordachioaie, Stefania Dumbrava, Josip Dzolonga, Jana Giceva, Michael Kohlhase, Christoph Lange, Darko Makreshanski, Dimitar Misev, Normen Müller, Alen Stojanov, and Jakob Ücker for their contribution to the panta rhei, JOMDoc, and JOBAD project. This work was supported by JEM-Thematic-Network ECP-038208.

References

1. CoFI (The Common Framework Initiative). In: Mosses, P.D. (ed.): CASL Reference Manual. LNCS. IFIP Series, vol. 2960. Springer, Heidelberg (2004)
2. Green, T., Peter, M.: Usability analysis of visual programming environments: A cognitive dimension framework. Journal of Visual Languages and Computing 7, 131–174 (1996)

3. Hardy, G.H.: A Mathematician's Apology. Cambridge University Press, Cambridge (1992)
4. Heintz, B.: Die Innenwelt der Mathematik. Zur Kultur und Praxis einer beweisenden Disziplin. Springer, Wien (2000)
5. JOBAD framework – JavaScript API for OMDoc-based active documents (2008), `https://jomdoc.omdoc.org/wiki/JOBAD`
6. JOMDoc Project (2008), `http://omdoc.org/jomdoc`
7. Kerber, M., Melis, E., Siekmann, J.: Analogical reasoning with typical examples. SEKI-Report SR-92-13, Unversität des Saarlandes (1992)
8. Kitcher, P.: Mathematical naturalism. In: Kitcher, P., Aspray, W. (eds.) History and Philosophy of Modern Mathematics, pp. 293–325. University of Minnesota Press (1988)
9. Kohlhase, M., Müller, C., Rabe, F.: Notations for Living Mathematical Documents. In: Autexier, S., Campbell, J., Rubio, J., Sorge, V., Suzuki, M., Wiedijk, F. (eds.) AISC 2008, Calculemus 2008, and MKM 2008. LNCS (LNAI), vol. 5144, pp. 504–519. Springer, Heidelberg (2008)
10. Kohlhase, M. (ed.): OMDoc – An Open Markup Format for Mathematical Documents [version 1.2]. LNCS (LNAI), vol. 4180, pp. 25–32. Springer, Heidelberg (2006)
11. Kohlhase, M.: sTEX: Using TEX/LATEX as a semantic markup format. Mathematics in Computer Science. Special Issue on Management of Mathematical Knowledge (in press, 2008)
12. Kohlhase, M., Müller, C., Müller, N.: Documents with flexible notation contexts as interfaces to mathematical knowledge. In: Libbrecht, P. (ed.) Mathematical User Interfaces Workshop (2007)
13. Miller, B.: `LaTeXML`: A LATEX to xml converter. Web Manual seen (September 2007), `http://dlmf.nist.gov/LaTeXML/`
14. Müller, C.: A Survey on Mathematical Notations. KWARC report, Jacobs University Bremen (2008)
15. Müller, C., Kohlhase, M.: Panta rhei. In: Hinneburg, A. (ed.) LWA Conference Proceedings, pp. 318–323 (2007)
16. Müller, C., Kohlhase, M.: Context Aware Adaptation: A Case Study on Mathematical Notations. Research reports, Centre for Discrete Mathematics and Theoretical Computer Science, University of Auckland (November 2008)
17. OPENMATH Home. seen (March 2007), `http://www.openmath.org/`
18. The panta rhei Project (2008), `http://trac.kwarc.info/panta-rhei`
19. Polya, G.: How to Solve it. Princeton University Press, Princeton (1973)
20. Rabe, F.: Representing Logics and Logic Translations. Ph.D thesis, Jacobs University Bremen (2008)
21. Ruelle, D.: The Mathematician's Brain. Priceton University Press (2007)
22. Smirnova, E., Watt, S.M.: Notation Selection in Mathematical Computing Environments. In: Proceedings Transgressive Computing 2006: A conference in honor of Jean Della Dora (TC 2006), Granada, Spain, pp. 339–355 (2006)
23. W3C. Mathematical Markup Language (MathML) Version 2.0 (Second Edition) (2003) Seen (July 2007), `http://www.w3.org/TR/MathML2/`
24. Wenger, E.: Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press, Cambridge (2005)