Towards the Identification and Support of Scientific Communities of Practice*

Christine Müller
Computer Science, Jacobs University, Bremen, Germany
c.mueller@jacobs-university.de

This works applies the economic theory of *communities of practice* (CoP) to the area of science and scientific education. A document-centered view on CoPs is taken, i.e. it is assumed that *scientific practice* is inscribed in the content, structure, and appearance of documents. However, scientific practice is also encoded in the meta view on these documents, i.e. the community's rating, discussions, and reviews of document. It is claimed that explicating and comparing both types of practice allows for identifying various scientific communities and sub-communities. In order to test these intuitions, the collaborative and interactive reader `panta rhei` is implemented. `panta rhei` supports its users to easily find, discuss, and rate information. The system infers the users' virtual practice from their ratings, annotations, and browsing behavior in the system. Moreover, `panta rhei` analyzes the users' collection of imported scientific document in order to extract their inscribed practice. Based on the comparison of practices, `panta rhei` makes predictions on the users' CoP membership, i.e. their social networks and implicit relations among each other. In addition, `panta rhei` utilizes community information to analyse the relevance as well as the user's preferred content, structure, and appearance of scientific documents. This paper focuses on the discussion of scientific CoP from a document-centered perspective. It further introduces the current features of `panta rhei`'s educational release and the system's extension towards a community tool.

1 Introduction

The explosive growth of information available online and the rapid increase of scientific information critically degraded the ability of scientists to keep informed about areas of interest other than their major research areas. The problem of information overload has been addressed by several approaches, such as providing personalized information by recommender systems, improving search by faceted browsing environments, and facilitating the assessment of information by means of annotations. Nevertheless, none of the existing methods provides a complete and/or satisfying solution to the problem, especially in the field of scientific education. This work aims at diminishing the overload of (new) scientists by applying the economic theory of Communities of Practice (CoP) [LW91] to science and scientific education.

By definition, communities are groups of people that collaboratively develop knowledge, share experience, and build up a common identity. One occurrence of communities are online communities which are groups of people that share similar interests and exchange experiences via websites, blogs, or mailing

*I would like to thank Michael Kohlhase, Normen Müller, Achim Mahnke, Marc Wagner, and Melanie Hartmann for their contributions to my work. Moreover, I want to thank the members of the KWARC group for the fruitful discussions and continuous feedback.

lists. In contrast, communities of practice are informal practice-oriented groups of people with similar tasks. They are further described by a specific set of practices, which are shared by their members to achieve these tasks.

An educational CoP An example for an educational CoP is a course established by its members, i.e. the students. The course can be described by a set of requirements set by the professor, which define the *practice of the course*. For example, the students of a computer science lecture have to acquire technical terms and definitions, learn how to solve mathematical problems as well as how to prove mathematical statements, and improve their programming skills. While interacting with the professor and with each other, they adapt to the practice of the lecture but also contribute with their individual habits. In consequence, students also influence and change the course with their individual habits and thereby make it a dynamic and emergent structure.

A scientific CoP Another example refers to a scientific scenario. For example, the participants of a conference share certain interest in the same research field. They submit papers that they believe to be related to the conference's sessions and attend various talk hoping to get in contact with similar interested researchers. The more experienced and established a scientist, the better his overview of which communities relate to his practice and the better his ability to transfer his work to other communities. However, we face difficulties in distinguishing scientific communities due to their emergent and dynamic nature: Scientific communities continuously change, they arise and disappear again. Moroever, many scientific communities overlap so that firm boundaries can not be drawn.

The concept of CoPs has been widely used to describe the process of social learning that occurs when people who have a common interest in some subject or problem collaborate over an extended period to share ideas, find solutions, and build innovations. Several existing CoPs have been analyzed [Wen05]. This work intends to identify CoPs based on pre-defined practices or a given set of members. Moreover, ways of how to support and establish CoPs shall be find. To model CoPs and CoP membership, this paper proposes an *extensional approach* [KK06] and *intensional approach*.

Extensional The extensional approach defines CoPs as *informal practice-oriented groups of people with similar tasks*. In particular, an extensional CoP model identifies pactice of a CoP based on an explicit specification of its members, e.g. registering a student for a course: The students are aware that they are attending a lecture and that they need to fulfill certain requirements. Their CoP membership in terms of the registration to a course is obvious. However, they may not be aware of the practice they need to adapt to. An extensional model helps explicating the practice required for the course as well as the practice of the students, helping them to adapt to the professor's expectations. Nevertheless, students are not necessarily aware of their (intensional) membership to other (sub)communities in the course. These can be defined by their background (field of study), their origin (their cultural heritage and language) or their interest (interest in web design or hardware). Helping students to establish and find these (sub)communities and/or supporting existing once, also improves the learning and knowledge exchange within the course community.

Intensional The intensional defines CoPs, as communities *described by a set of practices performed by its members in order to solve similar tasks*. An intensional model identifies members of a CoP and the CoP itself based on sets of shared practices of various users. For example, the participants during a conference are not always aware of the fact that they belong to a specific (sub)community in the research community. Especially, newcomers face difficulties in overviewing the wide field of science and identifying communities that relate to their practice, especially when talking about interdisciplinary research scenarios. This works aims at facilitating scientists to explicate their practice and at comparing scientific practices among researchers to improve the knowledge exchange in scientific communities.

This paper takes a document-centered view on CoPs, i.e. it is assumed that *scientific practice* is inscribed in the content, structure, and appearance of documents. Moreover, scientific practice is also encoded in the meta view on these documents, i.e. the community's rating, discussions, and reviews of document. It is claimed that explicating and comparing both types of practice allows for identifying various scientific communities and sub-communities. In order to test these intuitions, the collaborative and interactive reader `panta rhei` [Mül07c, MK07, Mül07b] has been implemented, which integrates extensional and intensional CoP models. The identification of practice (in order to either describe extensional CoPs or to identify intensional CoPs) is based on the analysis of imported scientific documents as well as the observation of virtual interactions such as ratings, annotations, and browsing. Moreover, `panta rhei` will provide questionnaires in order to model user preferences and portfolios, which can be used to find further similarities among users. The identification and comparison of practices allows `panta rhei` to make predictions on the users' (extensional and intensional) CoP memberships, i.e. their social networks and implicit relations among each other. The identified CoPs and the respective practice are further used to provide services such as an elaborated *semantic search* and *adaptation* of content, structure, and appearance of scientific documents.

2 Document-Centered View on Communities of Practice

[KK06] propose the application of the economic theory of *communities of practice* [Wen05] to the area of mathematics. According to their discussions, *mathematical practice* is *inscribed* into *documents*, e.g. by selecting specific notations or referencing other mathematical publications. Analyzing a *collection of documents* will potentially lead to *clusters of shared practices*, i.e. (extensional) *communities of practice*, that define the mathematical practice of researchers. In addition, the authors emphasize the importance of collaborative ratings on documents e.g. in terms of their relevance, soundness, and significance. This *meta information* is an important dimension for the identification of practice and CoPs.

Moreover, [KMM07b] introduce the notion of *documents as interface to knowledge* and emphasized their importance in order to analyse and identify scientific practice. A three-layered document model for the semantic mark-up language Open Mathematical Document (OMDoc) [Koh06] was proposed: a *content layer*, a *narrative layer*, and a *presentation layer*, whereas all layers encode different practices that can be used to identify and support CoPs as exemplified below.

Presentation The presentation layer explicates the *notation practice* of the author as well as his layout preferences e.g. in terms of *color*, *font*, or *border*.

Content The content layer may be used to identify communities of practice by analyzing the relation and meaning of document fragments. Based on semantic representation language, such as OMDoc, we are e.g. able to explicate the meaning and dependencies of scientific ideas as well as the relation between document fragments. Moreover, by tracing the references of an author's document, we can provide insights into his background, interest, and focus in the context of the current document.

Structure The narrative layer, i.e. the structuring and selection of content, can help to identify further CoPs an author relates to. For example, the structure of a lecture encodes how the professor suggests to study his course. In contrast, tracking a student's way of digging through the lecture material allows to identify his way of learning about the new material. Alternatively, the structure of a student's proof or the structure of his homework encodes how a student solves a certain problem.

Meta The meta information on documents e.g. in terms of their relevance, soundness, and significance, can be used to analyse the collaborative view on documents.

The following sections provide examples for each document dimension, i.e. the presentation (cf. section 2.1), the content (cf. section 2.2), the narrative structure (cf. section 2.3), and the meta dimension (cf. section 2.4), as well as their application to identify and support CoPs.

2.1 Notation Practice

[KMM07b, WM07, Mül07b] define *notation practice* as the selection of an adequate presentation for symbols. In this sense, an author's *notation practice* is her individual way of selecting notations, which she acquires throughout her life and which is influenced by a number of factors. Moreover, several examples of notation practice and approaches of how to categorize and describe it are provided.

Explicating the Notation Context of Mathematical Documents [KMM07b] propose a document model that facilitates the explication of a document's notation context, i.e. the selection of appropriate presentations for all symbols in the document. By separating the notation context from the structure and content of a document, the authors aim at identifying the author's selection of notation, i.e. his notation practices. This facilitates both the identification of CoP that share specific notation preferences and, conversely, the adaptation of documents to the notation preferences of identified CoPs, e.g. groups of readers and co-authors. Furthermore, explicating a document's notation context allows for referencing and, particularly, reusing contexts, which reduces the author's workload during the authoring process.

Communities of Notation Practice in PLATΩ The PLATΩ [pla, WM07] system provides a collaborative environment for authoring and reading scientific documents. Based on the *semantic annotation* of documents, PLATΩ is able to extract the semantics of the mathematical knowledge contained. These annotation have to be provided for all semantic parts in the document, e.g. theories, definition, theorems, proofs, or formulae. In particular, PLATΩ separates the meaning and the notation of mathematical objects, such as symbols and formulae. This allows the system to provided *dynamic notation context* for arbitrary scientific content. Considering the notation context as a dynamic parameter separated from the content of a mathematical document, facilitates both the identification of the author's *notation practice* and of *communities of practice* (CoP) that share specific notation preferences, and conversely, the adaptation of documents to the notation preferences of identified CoPs, e.g. groups of readers and co-authors. In the scope of the PLATΩ system, the authors speak of two types of communities: explicit and implicit communities.

explicit CoPs Are manually created in the PLATΩ system by entering a community name and its members, e.g. a working group. Optionally, users can enter the notation practice of the CoP, e.g. by selecting a specific library in PLATΩ. Alternatively, the PLATΩ system captures the members' notation choices, extracts the notation practice of the CoP, and creates a respective library. The latter approach can be compared to the author's understanding of an extensional CoP model. The library is either maintained by the members of the group or by the PLATΩ system, which continuously observes and updates the CoPs notation choices. The CoP membership of a user, is used to e.g. adapt documents he reads and writes. Separating the presentation of documents, from their content and structure, allows the consistent use of notation within CoPs as well as the collaboration of author's across CoPs.

implicit CoPs PLATΩ also aims at supporting existing CoPs, i.e. the system observes and extracts the notation practices of all of its members to find CoPs across various explicit CoPs based on the authors' individual notation practice. This approach relates to the author's understanding of an intensional CoP model. The implicit CoP membership is valuable whenever an author opens a document with notations which are currently not known/ specified by his explicit CoPs. In this case, PLATΩ can draw on the notation specification in the author's implicit CoPs or, if no specification can be found, fall back to the default notation in the document.

2.2 Document Content

The content layer can be used to identify communities of practice by analyzing the relations and meaning of document fragments. Moreover, by tracing the references of an author's document, we can provide insights into his background, interest, and focus in the context of the current document.

Representation of Document Fragments and Ontological Relations The content layer is the representation of "information units and their ontological relations" [Mül07d], i.e. of the relations and meaning of document fragments. Based on semantic representation language, we are able to explicate the meaning and dependencies of scientific ideas as well as the relations between document fragments. Among other approaches such as OPENMATH [BCC+04], MATHML [ABC+03], and CNXML [CNX07], OMDOC [Koh06] provides an elaborated mathematical ontology, which categorizes and relates mathematical objects. For example, when introducing symbols in their documents, the markup facilitates authors' to reuse existing definitions and theories and, in particular, to relate their current discussion to other scientists. Tracing the fragments' ontological relations across documents of various authors allows to identify communities of practice based on the reused theories and definitions. Moreover, OMDOC provides a *document ontology* [Doc], which allows to categorize and relate document fragments of rather structural nature such as 'sections', 'definitions', and 'examples'. Comparing the type of content by means of the document ontology, can also be used to identify an authors' practice. For example, based on the number and types of document fragments in an author's document his writing practice can be described as 'demonstrative' (due to many examples), 'formal' (due to many definitions and formulae), or 'informal (due to many text fragments), and 'visual' (due to many figures).

Document Fragments and Variants Documents are interface to scientific knowledge and are created and adapted for a specific context defined e.g. by its audience (or community). In consequence, COPs can be analysed by comparing the content of a document and by categorizing authors into COPs based on similar text fragments in their documents. [KMM07a] present an infrastructure for representing and managing alternative document fragments, which are referred to as variants. Moreover, they extended OMDOC with the specification of variant dimensions, such as the language (German, English), the area of application (mathematics, physics, computer science), or the level of detail (short or long version). By interpreting the selection of concrete variant dimensions and their values as *shared practice*, the authors want to identify *communities of practice* based on similar preferred variant contexts. Vice versa, they aim at describing COPs by their most frequently used variant dimensions for various concrete documents.

Author support The modeling of COPs based on variant contexts will, in particular, support the variant selection process: Instead of requiring a variant context for each user, COP models allow for reusing existing specifications for users of the same community. For example, let us assume that students of a course are members of one COP. A teacher could now specify two different courses, e.g. his General Computer Science (CS) lecture for Mathematicians and the CS lecture for Physicians. Depending on a student's subscription to a course, the selection process can now reuse the variant information in order to instantiate abstract course material, i.e. generating appropriate (concrete) lecture notes for both communities, i.e. the mathematicians and physicians (cf. [KMM07a]).

User support Learning environments that are able to handle variants of documents can offer services such as the description of the documents, e.g. in terms of its language and audience. Moreover, they can provide readers with additional (variant) document fragments or adapt a given document based on the user-specific background. In consequence, the abstract course given by the professor can be interpreted as community, while the concretizations of the course (e.g. for mathematicians or computer scientists) are sub-communities or instantiations.

Tracing references The references between documents (fragments) can be used to identify communities build up of nodes (the author) that are tied by uni- or bidirectional references to each others publications. The references are interpreted as practices which encodes the author's background, interest, and focus in the scope of the concrete document. Although, author's might be able to keep an overview of the focus of their related work, they can be supported by e.g. visualizing the reference-network. Moreover, similar to the "fiend-of-a-friend" approach, the referenced-based interrelation between scientists can be used to improve recommendations and search ranking. For example, documents which are "closely" related to the author's current context, i.e. the current document he is reading or writing, could be ranked higher than "more remote" publications. Moreover, reference could be typed, i.e. the reference annotation can be enriched by further semantics of the relation between document. For example, a document fragment 'proves', 'refutes', 'challenges', or 'agrees' with the referenced document fragment.

2.3 Narrative Structure and Browsing practice

The narrative layer, i.e. the structuring and selection of content or also referred to as "representation of the presentational order of information units" [Mül07d], can help to identify further CoPs. For example, the structure of a lecture encodes how the professor suggests to study his course. In contrast, tracking a students way of digging through the lecture material allows to identify his way of learning about the new material. Alternatively, the structure of a student's proof or the structure of his homework encodes how a student solves a certain problem.

Monitoring Browsing Practice In the beginning of the new semester, a professor sets up a specific structure for his lecture, e.g. a General Computer Science lecture (GenCS). He defines the order of several sections he wants to discuss while taking into account their dependencies: For example, he decides to introduce *graphs* before talking about *trees*, since trees are a special class of graphs. The professor's practice of how to present his lecture is inscribed into his lecture notes, e.g. the table of content. Although some students might adapt this structure to study the material, they will most likely access the material in different orders, e.g. depending on their current assignment they have to solve. When presenting the course in an online learning environment, the students' browsing practice can be tracked and analyzed e.g. by considering the log files and session data. Findings from the students' browsing behavior can be used to improve recommendations: For example, the system can recommend the 'most popular page' based on the number of hits of the page. Moreover, the system can support students in browsing the material, e.g. by recommending the 'next page' based on the previous browsing of other students. When tracking the students' interaction, the students' context should be taken into account. For example, in an educational scenario we can assume that the context is defined by the current assignment.

Analyzing the Structure of Mathematical Exercises The analysis and comparison of exercises allows for the identification of problem solving practice of students. The example below, presents an exercises from the analysis in [SDB07]:

st[0]:	$(R \circ S)^{-1} = \{(y,x) \mid (x,y) \in (R \circ S)\}$
tu[0]:	This statement is correct.
st[1]:	$(R \circ S)^{-1} = \{(y,x) \mid \exists z (z \in M \wedge (x,z) \in R \wedge (z,y) \in S\}$
tu[1]:	This formula is also correct.
st[2]:	$(R \circ S)^{-1} = \{(y,x) \mid \exists z (z \in M \wedge (z,x) \in R^{-1} \wedge (y,z) \in S^{-1}\}$
tu[2]:	This is correct. You are on a good way.

st[0]:	One needs to show equality between two sets.
tu[0]:	That's right! How do you proceed?
st[1]:	I use the extensionality principle.
tu[1]:	That's right.
st[2]:	Let $(s,r) \in (R \circ S)^{-1}$. According to the definition of the inverse relation it then holds that $(r,s) \in (R \circ S)$.
tu[2]:	That's right!

The proof in the dialog on the left is based on a pure "rewriting" style, whereas the dialog on the right employs the extensionally principle. However, both solution paths result in a correct solution for the exercise. They constitutes to possible ways of how to solve the problem. The structure of a proof possibly allows for describing the problem solving practices of users, e.g. in terms of the language (natural or formal) or level of detail of the proof.

Explicating a Documents Structure As mentioned above, [KMM07b] have extended the OMDoc document model in [Koh06] to facilitate the explication of a document's notation context, i.e. the selection of appropriate presentations for all symbols in the document. Nevertheless, the separation of content and structure has already been addressed in the former document model of OMDoc1.2:

OMDoc1.2. contains two OMDoc types in order to mark up the knowledge contained in a mathematical document and its structure: Content OMDocs are "knowledge-centered documents that contain the knowledge conveyed in a document" [Koh06]. In contrast, narrative OMDocs are used to "reference the knowledge[-centered documents] and add the theoretical and didactic structure of a document" [Koh06]. The combination of the narrative structure and the (mathematical) content of a document as the formal representation of a document model, has been defined by Normen Müller as NarCons, i.e. two-dimensional graphs consisting of a **nar**rative layer and a **con**tent layer [Mül06], and are a suitable model that allow for the explication of a document's structure in separation of its content.

In order to model the structure of exercises, [GT07] describe an attempt, which is applied by the learning environment ActiveMath [MAF+01]. The exercises are represented as finite state machines, in which the authors can visually manipulate the graph of an exercise, i.e. the alternative paths for solving an exercise.

2.4 Rating and Annotations

Besides the analysis of documents, we can also focus on the interaction of users via community platforms such as forums, blogs, newsgroups, or wikis in order to identify their ratings, reviews, and annotation of documents. In particular, this allows for identifying the relevance for and acceptance of scientific documents by a specific CoP.

Moreover, the interaction between users can be valuable in order to identify social networks, i.e. social structures made of nodes (the individual scientists) that are tied the indirect and direct interaction via community platforms. For example, user rating and social network can be used to implement collaborative recommendations of documents, e.g. by basing recommendations on the rating of other users that are 'peers' of the user and rate items similarly. As opposed to content-based recommendation approaches, collaborative recommendations measure the similarity of user-specific ratings instead of keyword frequencies in documents. Thus, they can be applied to any kind of content, even to items that are dissimilar to those known by the user (cf. [AT05]).

3 *panta rhei* — Features and Outlook

The interactive and collaborative community reader *panta rhei* has been built to gain intuitions on the applications of CoPs to science and research. In its first version, *panta rhei* is based on the educational scenario, in which the system combines marked-up course material, the narrative course structure, and a course forum for the General Computer Science (GenCS) lecture at Jacobs University[1]. The system currently supports the browsing, annotation, discussion, search, and rating of course material, which will be described in the following section.

3.1 The educational prototype

Currently, the GenCS lecture is accompanied by a course forum that is used to distribute the course notes, homeworks, and announcement. Students use it to discuss the GenCS lecture, to ask questions, and to solve their homework. They are supported by our Teaching Assistants (TAs) who contribute to the discussions. It has been a limiting factor to the discussions that one cannot explicitly point to the section in the lecture or homework since these are only available as attachments in PDF format. In consequence, students have to describe their problem in the forum hoping that the TAs and other students will find the referred slides in the PDF and provide appropriate answers. Moreover, in order to reply to all upcoming questions and to assure whether their answers did help the students, TAs have to frequently check the forum postings.

Figure 1 and Figure 2 present two screen-shots of the system: The *table of content* on the left displays the sections and subsection of GenCS, i.e. the narrative structure of the course and the list of assignments.

[1] For information on the system's implementation and architecture please refer to [Mül07c, MK07]

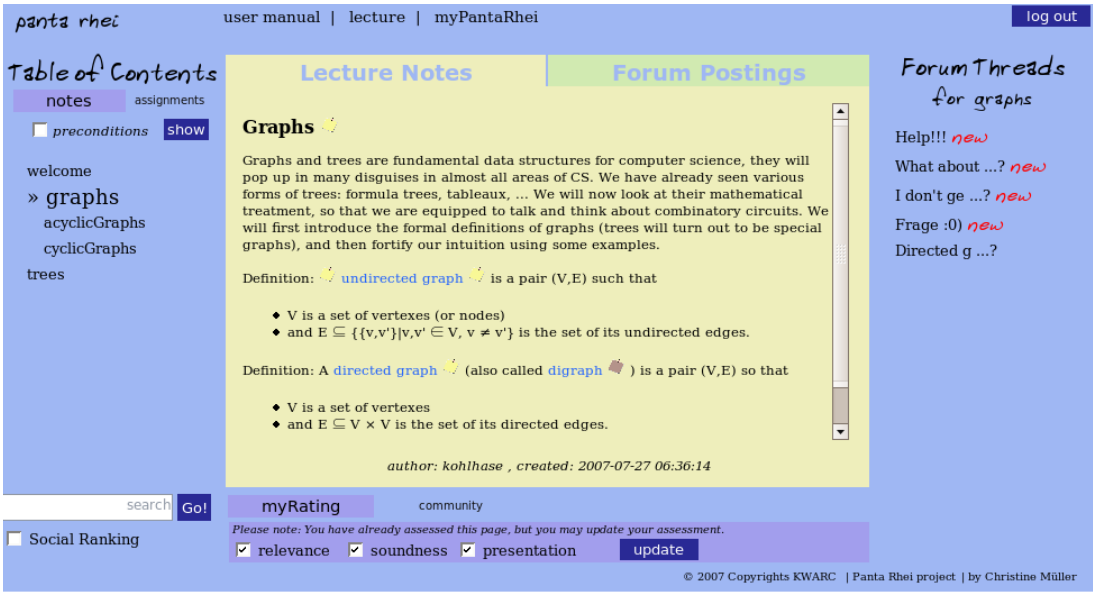


Figure 1: The `panta rhei` content

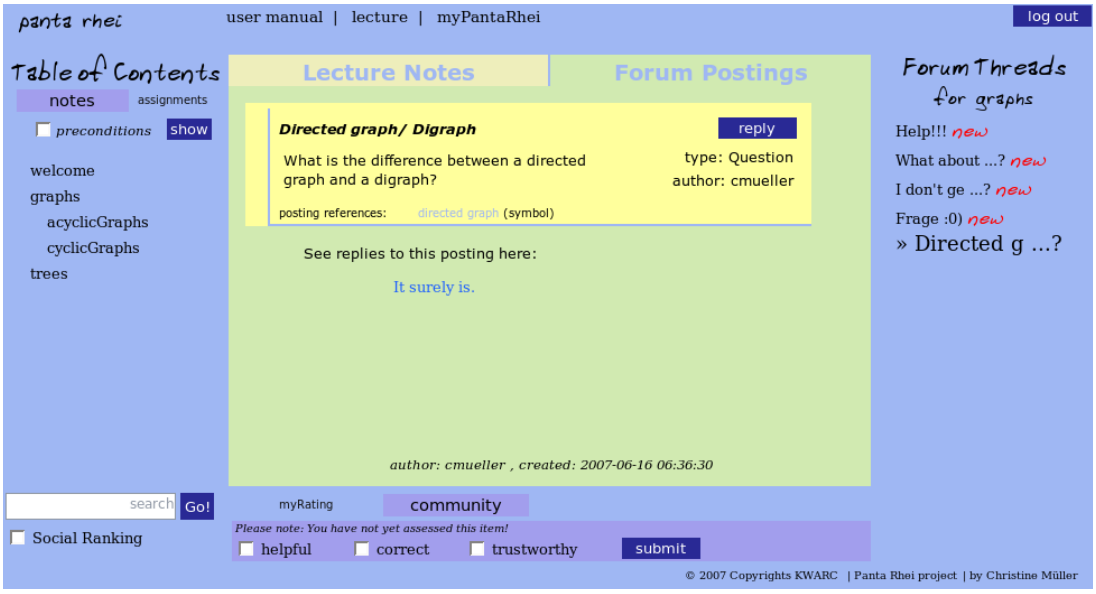
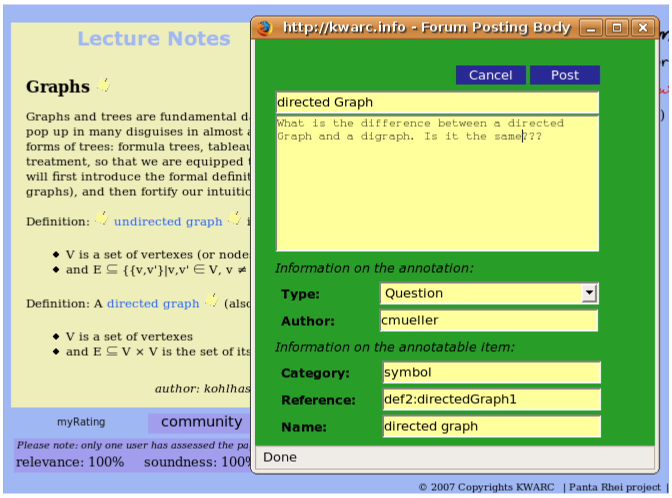


Figure 2: The `panta rhei` forum

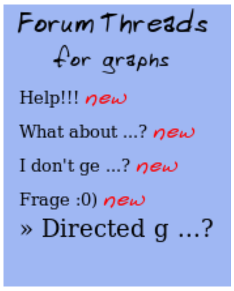
Depending on the selected tab, the main section in the center of `panta rhei` displays the hyper-media presentation, i.e. a *course* page or assignment (Figure 1), or the discussion media, i.e. a *forum* posting (Figure 2) used to discuss the lecture and homework. The section on the right-hand side displays a list of *forum threads* that are linked to the course page and which `panta rhei` can rank and filter. Moreover, users can rate course pages and forum postings by using the rating form below the main section.

Features `panta rhei` supports students' activities, e.g. the *browsing of course material*, the *annotation* of course material or forum postings as well as the *search for postings and course material*, and thus facilitates the discussion of course material between students and TAs. Additionally, the system allows to *rate* of forum postings (helpful, correct, trustworthy) and course pages (relevance, soundness, and presentation). In [KK06], these ratings or *value judgments* have been introduced as essential property in order to model a community of practice.

Annotation of course materials To post a question or comment referencing an information item (e.g. a concept, its definition, or a claim), students can to click on a specific locator button in the form of a *post-it* note. This displays a pop-up, given the author, the type of annotation, as well as further meta-information about the annotated item on the page. An information item that can be annotated in this way is referred to as **annotation item** (ANIT). After submitting the annotation, `panta rhei` creates a respective forum posting that is linked to respective ANIT. In the figure to the right the student decided to pose a question about the *concept* of a *directed graph* on the *graphs* course page.

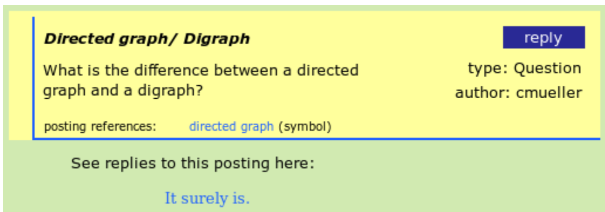


Searching for questions in the content Students can choose to solely view all postings that refer to a page or its sections. This is currently implemented by links on the page, i.e. by selecting a respective link on the page, `panta rhei` returns a list of threads that are linked to the respective section ranked by their creation timestamp (cf. figure to the left).

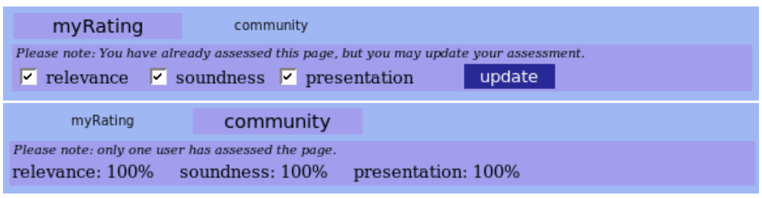


Moreover, users can set the display preferences of the lists. For now, user can choose to highlight all postings before a certain timestamp (i.e. 'yesterday', '3 days ago', 'a week ago', 'semester start') or to hide all postings after the timestamp. For example in the figure to the right, the list of posting has been filter: Only postings posted within the last three days are displayed. All postings that have been created since the user's last login are additionally marked as new.

Browsing the forum After selecting one of the *forum threads* in the list of postings linked to the current ANIT, the tab of the center switches to the *forum* view (cf. Figure 2). Students can now reply to the posting or browse the posting thread by clicking on the replies below the posting. The figure to the right displays a question about *directed graphs* and a link to an answer to the posting.



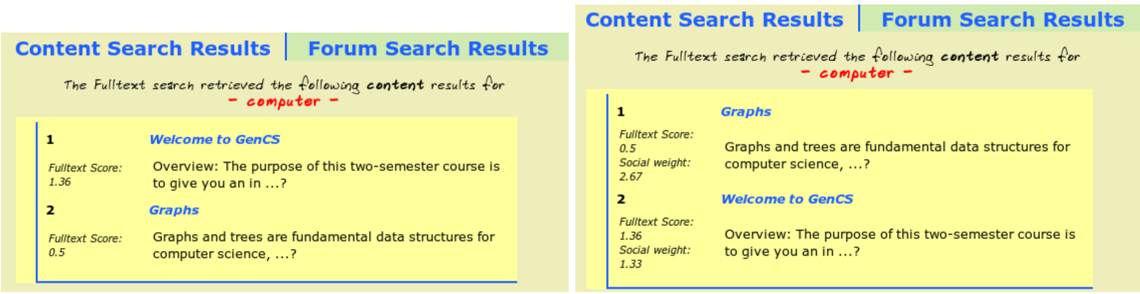
Rating content and forum postings The figure below displays a rating form for the course content allowing to rate the relevance, soundness, and presentation. In addition, users may view an community rating, i.e. an average rating of all users. Forum postings are rated in three different categories: helpful, correct, and trustworthy.



Searching content and forum `panta rhei` implements a simple full text search on content and postings. Moreover, the system provides an optional (simple) social ranking of search results based on the overall rating of pages and forums. The social rank is computed by computed as an average of the social ranks of all users, whereas the social rank of a user is an average of the three rating categories (e.g. relevance, soundness, and significance) for each user.. Moreover, the social rank is weighted with the number of ratings. The formula below computes the user-specific rank (R) as sum of all category values (V) divided by the number of categories (c), overall users and weighted by the number of users (u), who rated the item.

$$R = \frac{\sum_{i=1}^{u}(\frac{\sum_{j=1}^{c}(V_{ji})}{c})}{u}$$

The figure below displays the results for the full text search on the course content for the query string *computer*: The similarity measure of the search is displayed on the left of the results and leads to the ordering of results. However, when ranking the results based on the community rating as below, the page 'graphs' has a higher rank and is thus displayed first.



3.2 Extending `panta rhei`

Notation Selection and Adaptation. The integration of a *Notation Selector* such as the one by Smirnova/Watts [Not07, SW06] in order to allow users to enter their notation preferences will be evaluated. Furthermore, this work aims at facilitating the adaptation of documents according to the user's or his community's preference settings.

One possible way is to implement a PLATΩ interface[2]. While `panta rhei` will take charge of the user model and description of the required notation context, PLATΩ would take over the adaptation of the document. In order to do so, `panta rhei` needs to provide PLATΩ with the notation context, e.g. implemented by connecting the respective notation grammar with the user's notation choice as well as the document content and its relation to other documents, which are referenced or used by the document. For example, in order to adapt an *assignment* which includes a specific formulae, `panta rhei` also needs to provide the document which defines the symbols in the formulae. Moreover, PLATΩ needs to know about the semantics of the formulae, i.e. instead of simply provding the presented formula (e.g. represented in Presentation MathML [ABC+03]) `panta rhei` would also need to provide the respective meaning (e.g. represented in Content MathML [ABC+03]). Alternatively, PLATΩ could parse the original document (including Presentation MathML) only, identify the respective meaning (facilitated by the additionally supplied documents with e.g. the definitions of symbols) and adapt the document to the new notation.

Narrative Structure. The current prototype facilitates the definition of narrative structure for different user groups by creating a community profile, i.e. a preference setting, which is reused by a specific group of users. However, `panta rhei` still lacks a user friendly interface that allows authors to create and modify narrative structures as well as preference settings.

[2] I'd like to thank Marc Wagner for his following ideas on the integration of PLATΩ and `panta rhei`

In addition, this work intends to track the user's browsing behavior, i.e. his interactions. In the educational case study, interactions of students will be stored with the current context, i.e. the current assignment they have to solve. It is assumed that in the week of the assignment student's focus on the solution of the problem. In consequence, their interaction and ratings can be linked to the assignment. Comparing the interaction threads will possibly allow to identify the 'most viewed' or 'most relevant' pages in the context of the assignment. For example, these pages can be recommended to a next-year student, who is solving the same assignment.

Content. The prototype supports different variants of documents, e.g. based on the user's preference settings the German or English version of the course is displayed. Nevertheless, all content and its description have to be manually entered into the database and the file system. There is currently neither an import functionality nor an automatic mark-up of the content. Moreover, the documents have to be manually annotated, i.e. an HTML tag has to be added behind each "annotation item" (ANIT) in order to allow for its discussion. Nevertheless, import and automatic mark-up are preconditions in order to motivate author's to use the system since they are most likely not willing to provide semantic annotations the description of content as well as the annotation of their documents.

Content Markup of HTML files [NH07] proposes an annotation/ indexing pipeline for the marking up of HTML or PDF Formats. The first phase of the pipeline constitutes of the automatic annotation of the document layout, i.e. the extraction of "attribute-name/attribute-value" pairs based on e.g. specific HTML tags. In the second phase, these annotation are enriched by semantics gained from an ontology of the content and structure of the content, e.g. the OMDOC ontologies. In the third phase, the extracted data is stored in an index, used for processing of semantic search. Knowing about the content and structure, `panta rhei` could allow user's to simply indicate the type of content they want to discuss, e.g. symbols, definition, section-headings. However, in order to be a good source for automated annotation and markup, the imported documents have to possess a well-defined structure and restricted grammar. If this is not the case, `panta rhei` could offer to parse and annotated the documents based on a list of keywords given by the authors. The author will evaluate which approach to apply best for marking up of imported HTML documents in `panta rhei`.

Content Markup of OMDOC **files** Documents in the OMDOC format do not require the "layout annotation" [NH07] above, since they are allready contain semantic annotation which can be directly extracted and used for the content markup in `panta rhei` as well as semantic search. For example, OMDOC could be extended by a respective ANIT tag which would be used to automatically mark up the document during the OMDOC2HTML transformation.

OMDOC **Import** The goal is to import OMDOC[3] directly in `panta rhei` in order to extract the semantics of its content and structure and, based on this, to mark up the document for discussion in `panta rhei`. However, in the first step the author is working on an import for documents based on a light-weight OMDOC document model [KMM07b]. The OMDOC format is used as a wrapper for the (HTML) document's section. The narrative part of the document is used to define the narrative structure, which references the HTML snippets in the content part of the document. `panta rhei` will provide an XML parser, which extracts the *narrative structure*, the content as well as its description (e.g. author, date of creation, language, area of application) of the document. After parsing the document, its content snippets are inserted into the file system; while its description as well as narrative structure are stored in the respective database tables.

Meta. The prototype currently implements a rating and annotation feature, as described in section 3.1. The rating is currently based on a simply computation: `panta rhei` computes an average of all ratings

[3]The OMDOC can either be provided by the OMDOC repository, OMBase, directly or, alternatively, from the OMDOC-based semantic wiki SWIM [Lan07]

weighted by the number of users that rated the posting. In future version, the rating shall take a user's peer and communities into account. Moreover, this work aims at correlating a rating with the author of an entry, i.e. a trustworthy posting increases the trustworthiness of the user. In consequence, a trustworthy user's rating will have a greater impact than a user with lower trust, when computing an item's rating.

4 Conclusion

This paper discussed the notion of *Communities of practice* in science and educational science from a document-centered perspective. Examples for an scientific and educational community have been provided. Moreover, an extensional and intensional modeling approach was proposed. The document-centered views on communities has been described by four dimension: the presentation, content, and narrative structure of a document as well as a meta perspective on the document itself. Based on the dimension, various approach of how to identify and how to support communities have been provided. In addition, the collaborative and interactive reader `panta rhei` was introduce. The current features of the educational release of the system have been presented. Based on the four dimension, extension of `panta rhei` have been approach.

As a next step, the discussion on CoPs and extensions of the `panta rhei` system will be evaluated and discussed. A road map for the system will be written, in which the implementation of the selected features will be described.

5 References

[ABC+03] Ron Ausbrooks, Stephen Buswell, David Carlisle, Stéphane Dalmas, Stan Devitt, Angel Diaz, Max Froumentin, Roger Hunter, Patrick Ion, Michael Kohlhase, Robert Miner, Nico Poppelier, Bruce Smith, Neil Soiffer, Robert Sutor, and Stephen Watt. Mathematical Markup Language (MathML) version 2.0 (second edition). W3C recommendation, World Wide Web Consortium, 2003. Available at `http://www.w3.org/TR/MathML2`.

[AT05] Gediminas Adomavicius and Alexander Tuzhilin. Toward The Next Generation Of Recommender Systems: A Survey Of The State-Of-The-Art And Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[BCC+04] Stephen Buswell, Olga Caprotti, David P. Carlisle, Michael C. Dewar, Marc Gaetano, and Michael Kohlhase. The Open Math standard, version 2.0. Technical report, The Open Math Society, 2004. `http://www.openmath.org/standard/om20`.

[CNX07] CONNEXIONS. Project homepage at `http://www.cnx.org`, seen February 2007.

[Doc] The cnx and omdoc document ontology.

[GT07] George Goguadze and Jan Tsigler. Authoring interactive exercises in activemath. In Librecht [Lib07].

[Hin07] Alexander Hinneburg, editor. *Wissens- und Erfahrungsmanagement LWA (Lernen, Wissensentdeckung und Adaptivität) conference proceedings*, 2007.

[KK06] Andrea Kohlhase and Michael Kohlhase. Communities of Practice in MKM: An Extensional Model. In Jon Borwein and William M. Farmer, editors, *Mathematical Knowledge Management, MKM'06*, number 4108 in LNAI, pages 179–193. Springer Verlag, 2006.

[KMM07a] Michael Kohlhase, Achim Mahnke, and Christine Müller. Managing variants in document content and narrative structures. In Hinneburg [Hin07].

[KMM07b] Michael Kohlhase, Christine Müller, and Normen Müller. Documents with flexible notation contexts as interfaces to mathematical knowledge. In Libbrecht [Lib07].

[Koh06] Michael Kohlhase. OMDoc – *An open markup format for mathematical documents [Version 1.2]*. Number 4180 in LNAI. Springer Verlag, 2006.

[Lan07] Christoph Lange. SWiM: A semantic wiki for mathematical knowledge management. web page at `http://kwarc.info/projects/swim/`, seen August 2007.

[Lib07] Paul Libbrecht, editor. *Mathematical User-Interfaces Workshop 2007 at MKM*, 2007.

[LW91] Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation(Learning in Doing: Social, Cognitive and Computational Perspectives S.)*. Cambridge University Press, 1991.

[MAF+01] E. Melis, E. Andres, A. Franke, G. Goguadse, P. Libbrecht, M. Pollet, and C. Ullrich. ActiveMath System Description. In Johanna D. Moore, Carol Luckhard Redfield, and W. Lewis Johnson, editors, *Artificial Intelligence in Education*, volume 68 of *Frontiers in Artificial Intelligence and Applications*, pages 580–582. IOS Press, 2001.

[MK07] Christine Müller and Michael Kohlhase. panta rhei. In Hinneburg [Hin07].

[Mül06] Normen Müller. An Ontology-Driven Management of Change. In *Wissens- und Erfahrungsmanagement LWA (Lernen, Wissensentdeckung und Adaptivität) conference proceedings*, 2006.

[Mül07a] Christine Müller, editor. *1st Workshop on Scientific COmmunities Of Practice (SCOOP-2007)*, 2007.

[Mül07b] Christine Müller. Lectora – towards an interactive, collaborative reader for mathematical documents. `http://kwarc.info/cmueller/papers/Mueller_ResearchProposal_2007-03-14.pdf`, March 2007. Research proposal for a Ph. D. thesis.

[Mül07c] Christine Müller. Panta rhei. Project Home Page at `http://kwarc.info/projects/panta-rhei/`, seen August 2007.

[Mül07d] Normen Müller. Towards an Ontology-Driven Management of Change. Exposé of PhD research proposal, June 2007.

[NH07] German Nemirovskij and Eberhard Heuel. From semantic document annotation to global search facilities for personalised study programmes. In Müller [Mül07a].

[Not07] Notation Selection Tool, seen February 2007. available at `http://www.orcca.on.ca/MathML/NotationSelectionTool/`.

[pla] Plato: Interactive mathematical authoring.

[SDB07] Marvin Schiller, Dominik Dietrich, and Christoph Benzmüller. Towards computer-assisted proof tutoring. In Müller [Mül07a].

[SW06] Elena Smirnova and Stephen M. Watt. Notation Selection in Mathematical Computing Environments. In *Proc. Transgressive Computing 2006: A conference in honor of Jean Della Dora (TC 2006)*, Granada, Spain, 2006.

[Wen05] Etienne Wenger. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, 2005.

[WM07] Marc Wagner and Christine Müller. Towards community of practice support for interactive mathematical authoring. In Müller [Mül07a].