Semantic Knowledge Management for Education

Andrea Kohlhase
Dept. of Mathematics and Computer Science
University Bremen
kohlhase@tzi.de

Michael Kohlhase
Computer Science
Jacobs University, Bremen
m.kohlhase@jacobs-university.de

Abstract

'Semantic technologies' are touted as the next big wave in Educational Technology and as the solution to many problems in this arena. Interdisciplinary work between the fields of Knowledge Management (KM) and Educational Technology (ET) is booming. But the crop of actual systems and semantically enhanced learning objects is still meager, maybe KM and EL they are lacking a consensus on the underlying notions e.g. of 'semantics', yielding specific problems in their interplay.

In this paper we take a look at semantic educational technologies and draw conclusions for their approach in KM. In particular, we (re)-evaluate the notions of 'semantics', 'knowledge', 'learning', their role for learning materials in ET, and how they interact with the contexts involved in the learning/teaching process. Based on this, we distill a list of conditions the underlying knowledge representation format must fulfil to support these.

As these conditions are still rather abstract, we show how they can be realized in a concrete language design, taking in our OMDoc (Open Mathematical Documents) format as a point of departure.

1. Introduction

Since the nineties the Internet and the World Wide Web (WWW) have revolutionized the way we handle information. The envisioned "service society" [CoS94] turned into a "knowledge society" [Ste94, Lie06], where distribution and communication of information are not only central issues, but also have become deeply embedded in every day life [MW07a]. The ever-growing abundance of data and their availability west of the digital divide pose not only opportunities and challenges to society, but also to Educational Technologies (ET). For the latter, the opportunities consist in access to electronic documents on a large scale anytime anywhere and more efficient communication and cooperation e.g. via e-mail, blogs, and wikis.

One answer to the evolving challenges of the web consists in the idea of the "**Semantic Web**". According to TIM BERNERS-LEE's original vision the "*Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users*" [BLHL01]. The potential of this idea is stunning, especially in educational scenarios when combined with the associated technological capacities of dissemination and communication. Even though we have powerful software systems to support knowledge work, they cannot interpret the documents on the web and therefore cannot support knowledge work at a web scale. What we need is a web of *intelligent content*, i.e. semantically enhanced learning objects and active documents that carry machine-interpretable unambiguous accounts of their meaning.

For educational scenarios, the underlying, naive thesis has been, *"If computers can understand semantics, then data can become reified knowledge, which in turn can be used as content for providing learners "anywhere-anytime" (as well as "just-in-time") with whatever they want or need to learn"*. But this impressive potential contrasts sharply with real life acceptance (cf. e.g. [DI05, p. 2] or [TS02]). In particular, learning materials that are offered and communicated with and about are still largely simple strings of characters, or worse, images for mathematical formulae or chemical compounds. Therefore, we start with the understanding of the term 'semantics'.

1.1. 'Semantics' for Education

Semantics — "*the theoretical study of meaning in systems of signs*" [Wik08] or "*the meaning or relationship of meanings of a sign or set of signs; especially : conno-*

tative meaning" [MW08]— is tackled by many different scientific communities, e.g. philosophers, linguists, pedagogues, or computer scientists. Some are thrilled and fascinated, others are awed, intimidated, and demure, but all argue that the study of semantics is important in one way or another. We may conclude, that the term 'semantics' has many distinct facets and triggers various associations. It is difficult to talk about it as people mean very different things with this rich term. For instance, 'semantics' is generally understood as "meaning", it complements the triadic language model of syntax and pragmatics, and it is strongly connected with "disambiguation", "context", or "meta-information". For computer scientists, semantics signifies *what* representational objects mean e.g. in contrast to semioticists who are interested *how* they mean something.

In this situation we are not so much interested in a definition of semantics, but in a model of semantics which can be made use of — even if it doesn't cover all its aspects. We consider the **semantics of a knowledge object** to be determined by its *structure* (how is the object built up from already known objects, how is it defined in terms of other objects) *and its context* (what do we already know about these objects, how are these objects defined, what is their relation to other objects).

If we take the "potential use" as a guiding principle for our semantic model, we have to determine where and for whom semantics can play a role for quality of use. Even though every use of semantic data eventually serves people, we need to differentiate between *direct* use by software or by people, as 'quality' takes different meanings for them. For instance, in a theorem proving system, the underlying algorithms make use of the semantic input; whereas in a mathematical tutoring system, the learning path exhibition (enabled by intelligent content) is used by a student. In the former, the user does not need to understand the underlying semantics, as her goal may have been achieved by an automated rejection of a claim. In contrast, in the latter the user wants to 'learn' and therefore needs to accept the proposed learning path in her specific situation. This can be a scenario, where she just wants to look up a fact, but may also be in a context, where one student needs to study the underlying concepts for an exam being aware of the subject from a previous lecture and another with a lack of the fundamental concepts assumed at this point.

A designer who wants to exploit semantic data has to understand the opportunities associated with them as well as the difficulties and barriers of use.

1.2. Semantic Potential in Educational Scenarios

Two dimensions for improvement stand therefore out, the *data quality* on the one hand and the *interaction quality* on the other. Analytically, both can be assigned on an abstract and a concrete level (see Figures 1 and 2). In particular, we can discern the data model and its instantiation with respect to data and the interaction model and its appropriation for interaction.

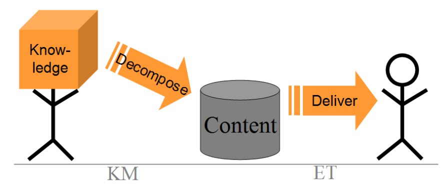

Figure 1. Data and Interaction Model

If we reformulate these aspects of semantics and digital media within an educational framework, we can speak — on the data side — of the conceptual '*decomposition of knowledge*' to store it as content in a data base and the actual process of '*capturing content*'. These tasks are mainly taken up by the field of "**Knowledge Management (KM)**". In particular, KM wants to 'capture' the data's underlying semantics in a way to get a handle for machine-support when dealing with it. From this standpoint, semantic data (or semantically enriched data) are data combined with meta-data enabling software to contextualize ('understand') it. In this sense, we will also speak of semantic data as '*machine-understandable data*'.

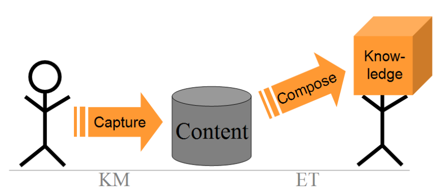

Figure 2. Formalization and Appropriation

On the interaction side, the abstract as well as the concrete level were addressed by "**Educational Technology (ET)**" researchers, where the interaction model is thought of in terms of '*delivering content*' and as '*composing knowledge*' on a concrete level. Here, technically speaking, semantic data are also data that are enhanced by information about them, but they are understood to be data that were already interpreted by humans. We will sometimes speak of '*interpreted data*' here. The main difference to the KM notion consists in the potential layer of trust. Even though 'semantic data' are basically the same for an outsider, KM designers view them as objects to be managed irrespective of their trustworthiness while designers of ET systems view them as input from a knowledgeable author evoking trust.

Interestingly, interaction quality and data quality are strongly interdependent. On the one hand, interaction quality depends rather obviously on the underlying data quality on both levels: if the data model is inadequate, the interaction model can't save it, and if the real data are of bad quality, a user's appropriation of even the best interaction model won't happen. On the other hand, a data model is always designed with a purpose in mind. This purpose assumes a built-in interaction model, particularly a human-computer relation model and with it an underlying 'Menschenbild' (idea of human), see [Hei99, p. 234]. Therefore, the data model depends conceptually on the envisioned interaction model. Moreover, concrete data instances have to be created within a system with an (explicit or implicit) interaction model. Hence, data and interaction quality are interwoven with each other on the abstract and the concrete level.

In this paper we are interested in the consequences on the abstract data quality *for* concrete interaction quality, i.e. we deal with the question what are the necessary conditions of a semantic KM data format underlying successful ET applications. We will asses these conditions for various KM formats concentrating on our OMDoc[Koh06c].

Even though we cover related issues we will not take into account the perspective of 'User Experience' (e.g [FB04, MW07a, GJ02]) which breaks the ground by dealing with motivational aspects of ET. In particular, they prepare the field so that users transform into what ET calls *learners*, who approach ET with awareness and readiness what is to come. Note that even though we are interested in interaction quality, our analysis does not take a Human Computer Interaction (HCI) perspective, which would be to care for the user *in* the using process and her 'relation' with the software resp. hardware: HCI does not consider pedagogic issues like "Bildung" or knowledge mediation.

2. Quality in Semantic Data for ET

We are especially interested, what semantic data must look like if they are intended for use in Educational Technology. Here, we do not focus on the quality of semantic data generation, but on the data format or ontology itself. Therefore we need to have a closer look at the principal objects 'knowledge' and 'learning' first to arrive at conditions for the design of KM and ET based on semantic data. For both concepts we will first review the epistemological foundations and then synthesize a conceptual model in the form of a space of knowledge and learning objects which will guide our further deliberations. The 'space' metaphor is inspired in part by SEYMOUR PAPERT in [Pap96], where he investigated different math educational approaches by relating (instead of contrasting) them within an n-dimensional space.

2.1. Knowledge

The famous (first) knowledge manager PETER DRUCKER is reported to have said that "*knowledge is between two ears and two ears only*" [Kon01], which captures the difficulties to expect when addressing knowledge from a modeller's viewpoint quite well. WERNER SESINK (a well-known media pedagogue) elucidates that 'reified knowledge' as it is offered in libraries can only be a form of intermediation of knowledge [Ses04, p. 136]. Moreover, the 'knowledge society' has already learned, that the fundamental concepts of data, information, and knowledge are not interchangeable concepts. In particular, the transitive combination of "Knowledge is created with information" and "Information are good data" and "Lots of available data" readily accepted during the Internet Bubble times cannot be held. A confirmation was given in a Delphi Study [SKMH04] concerning the future of KM. In [Kor05] KLAUS KORNWACHS critically discusses the use of the terms 'knowledge' versus 'information' and points to their "*fundamental difference*" [p. 34]. He points out that "*knowledge acquisition must be organized by knowledge itself*" [p. 36]. In particular, handling via technological systems is problematic because of this self-referentiality. Moreover, there are many critical accounts of the use of the term 'knowledge' with respect to Information and Communication Technology culminating in KM's respective "*autism*" [Lam02] (understood as "*the repetition of sentences and words without regard to their significance or the context in which they are spoken*" [ibid.]) or KM's "*nonsense*" [Wil02]. Therefore, we take 'knowledge' to mean information about that knowledge from a KM perspective , whereas we take it to mean "*factual material*" based on [DI05] or simply 'content' in ET language.

To get a better grip on the issues involved, let us start small, with the characters that make up the (textual) content of the Web: In the well-known KM model of PROBST ET AL. (see [PRR97]) they posit that glyphs, data, information, and also knowledge can be seen as stages of a pipeline shown in Figure 3 (the large circles are our's; see below for details). In particular, glyphs are just a set of *pixels* on the screen like {0;6;7;,}. A first set of rules imposed on the glyphs

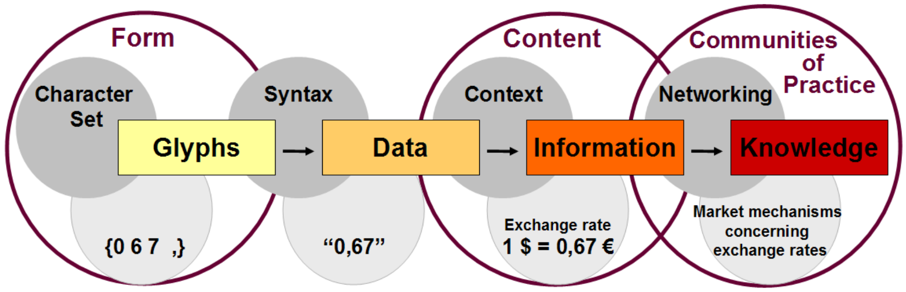

Figure 3. From Mere Glyphs To Valuable Knowledge (extended from [PRR97])

— the syntax — yields **data** which can be handled by machines like the *string* '0,67'. For obtaining meaning from such data we still need another component: the context. THOMAS H. DAVENPORT and LAURENCE PRUSAK interpret **information** as "*data that makes a difference*" [DP98]. In this view, data becomes information when a user can interpret the data in regard to a specific goal (or a local context), i.e. when they become meaningful, e.g. the *decimal number*[1] $0,67$ in contrast e.g. to an excerpt of a list of lucky numbers like "0,67,104,...". Finally, information becomes **knowledge**, if a user can interpret the information in regard to a global context like understanding the *exchange rate equation* in the area of specific market behavior with respect to change of exchange rates.

Now, what does this decomposition of the term 'knowledge' yield? On the one hand, the recognition, that information is more than a collection of data chunks, renders an extra enhancement of data via metadata annotation, i.e. semantic data, sensible. From a KM standpoint, it turns into the problem of abstracting an **ontology**, i.e. a semantic data format that structures not only data into classified data but also categorizes their interrelations. On the other hand, the recognition that knowledge is more than a collection of mere information chunks renders an intensive investigation of the 'Networking' aspect — the "*social life of information*" [BD00] — necessary. This is done for educational scenarios mainly in KM or ET subgroups within the CSCW (Computer Supported Cooperative Work) and HCI (Human Computer Interaction) communities, as well as in the current Web 2.0 discussion. Note that the boundaries start to become blurred and an inter- or transdisciplinary perspective is called for.

2.2. A Space of Knowledge Objects

If we look at the decomposition of knowledge in Figure 3 and assume a given ontology, then we recognize that we have an inscribed conceptual opportunity for separating content and form. Is it possible at all or is meaning lost if we accomplished such a separation? The starting point of our analysis is that a knowledge object is a complex entity. Our analysis here builds on our "Mathematical Knowledge Space (MKS)" as a conceptual model for mathematical knowledge based on content and form [KK05].

We differentiate between substance and accidence of a knowledge object in the Kantian tradition[2], where **substance** is the unchanging essence of an object, i.e. the totality of traits that constitute its meaning, whereas **accidence** is the object's appearance. A philosophic insight consists in the fact that these terms form a *dialectic pair*: even though an object's substance can be differentiated from its appearance, they are inseparable. Therefore, every knowledge object includes implicit formalizations (content) and explicit realizations (form), that can be interpreted as coordinates in a plane, that is structured by notions of equality. We call the latter **"substance equivalences"** as they represent meaning-conserving relations. For instance, an isomorphism $=_{log}$ between two distinct formalizations $\mathscr{G}_1$ and $\mathscr{G}_2$ of the mathematical concept 'group' is a substance equivalence, but a translation of either concept into a different natural language is one as well; we denote it with $=_{lang}$ in Figure 4.

At the left, we see the node $\mathscr{G}$, which represents the abstract concept of a knowledge object, followed by its two conceptualizations $\mathscr{G}_1$ and $\mathscr{G}_2$, which are substance-equivalent with respect to $=_{log}$ (and substance $\mathscr{G}$); we say that the $\mathscr{G}_i$ are **accidence variants**. In this example, we assume these conceptualizations to be independent of a natural language, so in another presentation step,

[1] in continental Europe

[2] There are many similar pairs, including: essence/appearance (HEGEL), matter/form (ARISTOTELES), or content/form (Mathematical Knowledge Management (MKM). Another pair often used in Computer Science is the one consisting of 'presentation' and '*re*presentation'. Principally, 'presentation' is used to describe an explicit realization (German: "Darstellung") whereas '*re*presentation' is used to describe an implicit formalization (German: "Darstellungsweise").

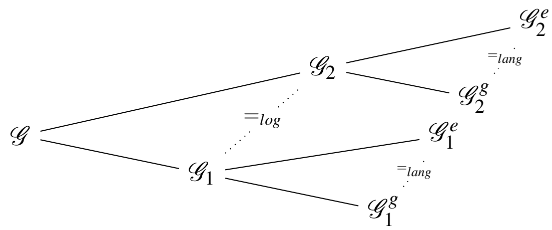

Figure 4. Knowledge Reification

we can fix that — creating an accidence variant for each natural language, in Figure 4 we have depicted two: English and German, giving rise to four accidence variants $\mathcal{G}_1^e, \mathcal{G}_1^g, \mathcal{G}_2^e$, and $\mathcal{G}_2^g$ that are substance-equivalent via the relations $=_{lang}$ and $=_{log}$.

If we combine this information with the substance and accidence relations view formulated above, we can see that Figure 4 is just the base of a tetrahedral knowledge space which we depict in Figure 5. Here, $\mathcal{G}_{log}$ is the dialectic pair consisting of the substance $\mathcal{G}$ and all (logically equivalent) formalizations $\mathcal{G}_i$ as accidences. We picture the substance and accidence relations **sub** and **acc** resp. with dashed lines.

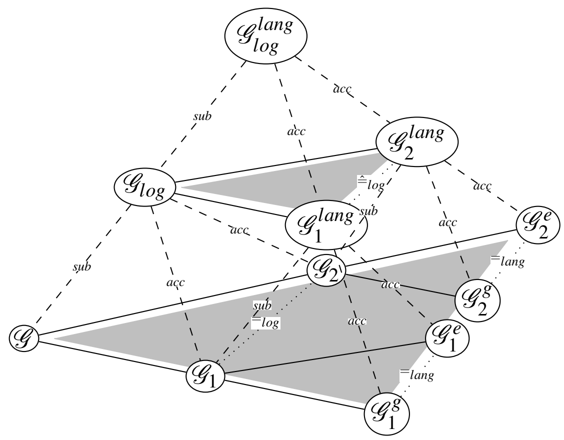

Figure 5. The Space of Knowledge Objects

2.3. Learning

The ultimate purpose for all described semantic concepts and technologies consists in re-enlivening the captured content into knowledge. In short, learning is not the composition of content as it is often thought of, but a *process of composition* enabled by the actual learner: she is *composing* knowledge and the software has to advance or trigger this hidden process. On the one hand, we have to look into a user's appropriation process and ask how semantic data can influence this process. On the other hand, we need to understand whether any software has a chance at all to cause 'learning'. Even though there is no definitive theory how learning happens, there are several well-accepted assumptions that allow us to support learning.

A user's appropriation process can be compared to that of using a library. WERNER SESINK amplifies: "*Libraries can only collect. If they weren't visited by people, who appropriate the collected knowledge, then they would transform into collection points of empty language shells*" [Ses04, p. 136]. Appropriation is done actively (but not necessarily consciously) by the user. Note that this activity does not refer to the operation of the to be appropriated object, it addresses the user's attitude and her evaluation of this object for adoption. JOHN DEWEY critically differentiates the terms 'accommodation' and 'adaptation'. The former refers to the (passive) human capability of acclimatization to circumstances, whereas latter relates to humans' (active) handling and reinterpretation of given circumstances to their own supposed advantage (from [Bel05, p. 64, 69]). In conclusion: we can rephrase **appropriation** as a concretization process of the abstractions contained in learning objects and software (see e.g. [Ses04, Sch07, Sch97]).

The very number of existing learning theories demonstrates effectively that modeling learning is a complex enterprise. They all build on distinct presuppositions in their underlying *"idea of human (Menschenbild)"* (for an overview see [Rei05, 146ff.] or [Doe]). Currently two theories are en vogue and can serve as a basis for our discussion: *Constructivism* [Pia96, MV92] builds on a knowledge *coaching model*, which considers learners as creators of their own reality. *Constructionism* [PH91], is a variant that stresses the embodied aspects of learning.

However, in order to understand learning itself, we take a more abstract stance than learning theories do. Intuitively, '**learning**' is related to a process of change: there is the experience of before and after. Formally, 'learning' is a *model of explanation* for the observation of specific changes that occur in the observed environment, which the observer accords to a (conceptual) system (following [Jün04, p. 73]). An instance of learning happens, when e.g. a student uses an ET application and she masters a subsequent online quiz on the topic and an observer (the quiz evaluation function) relates the environment (student and ET application) to a system (evaluation scheme wrt. achieving learning goals). Interestingly, SEBASTIAN JÜNGER points out, that talking about learning primarily yields information about the observer. In the example, the observer is a piece of software, that represents the designer as her "*deputy*" [dS05].

The mystery of defining learning consists in the fact that learning — contrary to popular opinion — is *no* autonomous activity with start- and endpoint. Even though we can use ‘to learn’ as an action verb: we decide to learn a topic, but we cannot cause learning, we can just experience it as such later on. We can create situations that afford learning, so-called learning scenarios, but we can not willingly generate the learning process (see e.g. [MD05, p. 30]). As a consequence, we can not model learning, as it *principally* can not be directed, not by an educator, not even by the learner herself. ET guru STEPHEN DOWNES phrases it in his well-known down-to-earth style as follows:

> “*People ask me for the analogy that I like to use for learning and what e-learning is, and I say, e-learning is like electricity, not like legos. It’s something that flows, it’s like the water system. It’s something that should be available, in the wall, where it comes out, it changes, it’s not concrete, it’s not the same thing you got yesterday - that’s what we’re really happy about with water, we wouldn’t want yesterday’s water.*” [Dow04]

Media-pedagogue KÄTE MEYER-DRAWE also points to the fact, that the very moment, in which the learning process begins, is not based on initiative, but can be considered an “*answer to retaining a (personal) standard*” [MD05, p. 34]. Critically therefore, we turn to the possibilities for setting an individual’s standards as Educational Technology applications can at most hope to manipulate these. KLAUS HOLZKAMP, argues that every human being engages in an ever-present “*inner dialogue*” [Hol95, p. 25], the result of which turns into her specific actions. The dialogue entertains the idea of at least two distinct standpoints that inform the personal standard. There are several names in the literature for this process, e.g. PAUL DOURISH calls it “*disengaging and reengaging*” [Dou03, p. 139], whereas EDITH ACKERMANN uses the metaphor of “*diving in and stepping out*” [Ack04].

2.4. A Space of Learning Objects

Instead of modeling learning itself we will now interpret the space of knowledge objects introduced in 2.2 from the perspective of how learning can be supported by ET. For this we take another look at the *front face of the MKS tetrahedron* (i.e. the triangle area between $\mathscr{G}^{lang}_{log}$, $\mathscr{G}$, and $\mathscr{G}^{g}_{1}$). Abstractly, we can see $\mathscr{G}^{lang}_{log}$ at the top as an abstract **Knowledge Object**: we can distinguish its content from its form arriving at what we call the “**Form Object**” and the “**Content Object**” — which can be recurrently subjected to the same analysis (see for the resulting view of the front face of the MKS). With the substance perspective on the Content Object we arrive at what we call the “**Platonic Object**”[3]. Successively looking down the substance branch of the tree, we arrive at more and more fundamental, abstract objects. In particular, these are increasingly liberated from their conceptualization as well as presentation. In contrast, looking down the accidence branch we arrive at more and more concrete and tangible objects. In detail, the accidence view on the Content Object leads to its conceptualization level (the “**Conceptualized Object**”), where we have a representation of the content in which certain decisions of how to think about it have been taken.

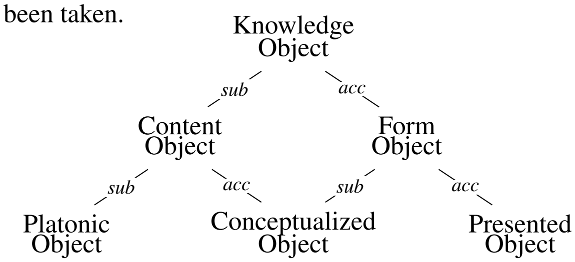

Figure 6. Learning Object Analysis Triangle

Now, let us look at the accidence aspect of the Form Object. As it becomes more and more concrete, we are lead to a presentation level and therefore to the concrete “**Presented Object**”. The substance perspective on the Form Object reveals again a conceptualization level, which by our analysis above is the Conceptualized Object. Let us clarify this with the group example: if we want to talk about what ‘the group’ really is (i.e. the Platonic Object) we have to decide on a representation (otherwise communication is impossible). This selection determines which of the possible definitions will be applied. In other words, the choice of the definition fixes the conceptualization of a group. Interestingly, so far capturing knowledge has always aimed at those knowledge objects that are “independent of everything” and *not* at the Platonic Objects themselves (possibly because we mistook them for the same).

Now we want to look at the MKS from the perspective of the learner who starts with the concrete materialization of knowledge like a certain document. From this point of view Figure 6 represents a “Learning Object Analysis Triangle”. Note that the lexical distinction between “knowledge object” and “learning object” starts to get blurred, we use the former, if we want to stress the representation aspect and the later for the application intent. The user heads for the knowledge itself — the Platonic Object — which is an author’s point of depar-

[3]The existence of such an object is not discussed, since this ontological assumption has no consequences for the conceptual model. As soon as we start reifying implicit knowledge (independent from the underlying ontology) we have to choose a form which in turn materializes the object.

ture. A reader has to differentiate between the potential content and the concrete form of a document. Depending on her personal choice what content and what form is, she understands and builds up her own knowledge. In contrast to the content author, who knows the used substance equivalence relations (and more) and actively chooses the representation of content, the recipient of knowledge has to infer the applicable equivalence relations.

We claim that the user perspective is already present in the analysis triangle of Figure 6: let us look at a student confronted with a book. It contains the knowledge in its final presented representation (*Presented Object*), but the student is aiming at an understanding of the underlying substance (*Platonic Object*). In order to decide what the content or the form is in the Presented Object, the student has to envisage a *Knowledge Object*, i.e. a potential model of the real knowledge to be learned. From this hypothetical Knowledge Object she can infer the *Content Object* and the *Form Object*. This dramatically reduces the search space of possible interpretations of the Presented Object to the presentations of the Form Object. Here, "understanding" means that the student is able to distinguish between the content of the Form Object (*Conceptualized Object*) and the Presented Object as its form.

Again, interestingly, the user generally is thought of as either modeling the Platonic Object (e.g. in case of a lecture) or the Knowledge Object (e.g. in case of an MKM system), whereas we conjecture that the user is building a Conceptualized Object as approximation of the Platonic Object. Taking this seriously might help to understand how MKM systems need to be positioned in a learning cycle.

What does this analysis have to offer for ET systems? Given the conceptual differentiation of knowledge objects discussed here, and the fan out of the presented objects[4] shown in Figure 4 we can interpret the space of learning objects as an *adaption space*, and the task of semantic ET systems as a process of

- choosing a learning path through the collection of learning objects and
- choosing an accidence variant for each of the learning object.

Together they result in a concrete **learning path** which is motivated by didactic concerns; in Figure 7 we have visualized the learning path as a gray line. Note that this particular learning path gives a self-contained exposition, as it includes all learning objects that are required by the relation denoted by the black arrows (this could e.g. be a functional dependency relation). Note as well that for each relevant learning element the learning path picks a particular representative from the substance-equivalent presentation in each of the knowledge spaces (depicted as little tetrahedra here).

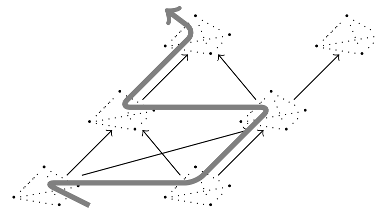

Figure 7. The Adaption Space

3. Educational Contexts, Adaptation, and Knowledge Representation

The existence and relevance of software around us is growing rapidly. Especially for educational technology we have to take this into account. This means, that as designers of technology we have to understand "*two worlds — the world of technology and the world of people and human purposes*" [Kap06, 4]. Even though this is quite an old recognition, it yields an interdisciplinary approach which is indeed difficult to accomplish in practice. TERRY WINOGRAD in [WBdYH06, p. *xvii*] strengthens this point by: "*Software is not just a device with which the user interacts; it is also the generator of a space in which the user lives.*"

As we pointed out above, Educational Technology cannot hope for automatically inducing 'learning' in a user. Even though ET aims at a much lower outcome, namely supporting the user in reaching a specific learning goal, even this cannot be handled as a *causal* relation. So, what can ET accomplish? Like a good teacher, who has a big amount of foils or social schemata at hand (i.e. ways of presenting a learning object or learning path) to guide her action for the class and for individuals, ET needs to draw on a wide variety of 'foils'. From a media theory standpoint, the advantage of using a computer for learning purposes consists in the potential of ultimate variability [Man01, p. 36] and its capability to adapt to a user's specific circumstances. Note that these include all constraints, her intrinsic ones as well as extrinsic ones like organizational burdens. Therefore, we want to take a close look at the various contexts that can be taken into account when choosing the right form for a content object. To quote KLAUS KRIPPENDORF: "*Meanings and contexts are twins, but they behave quite differently. [...] Contexts limit the number of meanings [...] the meaning of an artifact [is ...] a function of the relation-*

[4] Together, the effects we have studied in isolation in Figures 4 and 6 span the three-dimensional knowledge space in Figure 5.

ships among parts, mutually contextualized by their arrangement, and [. . .] how the whole is related to other artifacts and users' intentions." [Kri06, 59ff.]

The specific contexts we want to explore for adaptation opportunities build on the content/substance itself *and* on the learner. As the knowledge has to be mediated by software (at least ideally, see again Figures 1 and especially 2), the context of interaction between the teacher — which might be software — and learner is of interest as well. In this section we will look at the contexts from an educational perspective and relate this to our insights of the knowledge space above to prepare an analysis of the necessary capabilities of the underlying knowledge representation format in section 4. Our focus here will be to find out whether the contexts can guide engineering decisions on which parts of the knowledge to represent explicitly and which to compute on the fly in response to the needs of ET front end systems.

3.1. The Content Context

It is a common cognition that knowledge has two dimensions: Breadth and depth. We can transfer them directly to **content dimensions**. On the one hand, a larger variety of available content can potentially satisfy more learners. Note that this applies to both substance-equivalent and substance-distinct content. On the other hand, once a user has settled on the substance she is interested in, she might also want to delve 'deeper' into a topic, then the expectation criteria for use of content change from breadth to depth. Here, we mean "elaboration" and not "hierarchy" by "depth".

But we also have to take the hierarchical notion of distinct **context layers** for content into account. Content is naturally structured into various levels even though the levels themselves may not be naturally given. If we are for instance interested in cooking "Spaghetti Carbonara", we can imagine several entry layers, which trigger different learning strategies: If I already know the general picture, but have forgotten how many eggs I am supposed to use, I might call my sister for the information. But if I've never done it before, then I might look into a cookbook about pasta and go on from there. The intermediate variant would be, that I know exactly where to look for the number of eggs and accomplish it without any deviation. We see that for ET applications, the representation of the content context must be structured into levels as well to support these learning tasks. In particular, content and context must be sufficiently fine-granular to model the role of eggs in Carbonara sauce.

The interconnectedness of multiple learning objects allow to define **learning paths** through the content. Here, the local coherence of content may help to support a learner's navigation rationality. For instance, in order to prepare a pasta sauce with eggs, it is frequently assumed that one already knows how to break an egg (and according risks like spilling or crushing). The **context dependency** of content adds another aspect of these potential learning paths: imagine a search engine that indexed this paper under "cookbook" as it concluded from the frequent appearance of the name "Carbonara".

On yet another scale the context of content can (and may need to) change: when I have found a "Spaghetti Carbonara" recipe in an American cookbook, I have to translate (besides the language) all the units — e.g. 'cups' into 'grams' — before I can make use of it. This **recontextualization** is based on the human ability of accommodation [Dew33] or "coupling with the world" [Dou03] and belongs to the very basics of human learning processes. Again, the context representation has to be able to represent context *and* the various acts of recontextualization (which we can understand as the movement along substance equivalences in the knowledge space together with the necessary deviations from the learning path). An explicit representation of admissible context shifts is important as educators assign this kind of task to learners trying to understand and apply the underlying abstraction (i.e. a movement to the right in the learning object space; see Figure 4).

3.2. The Learner Context

It is a generally accepted fact that learning materials and interactions need to be adapted to the context of the learner to be effective. Depending on personal gusto, questions of layout can turn into learning hurdles, hence **customization** is relevant for the creation of a comfortable learning scenario. The **layout** of a learning object, i.e. its colors, font types, font sizes, etc. should be compatible with the learner's personal tastes: some people are alienated by high-contrast colors and others are not. Note that we cannot draw general conclusions: even though some usability engineers claim to have fail-safe recipes, we always know people who prefer things differently nevertheless.

Many technological disciplines start to address problems with modeling the object in question; here the **learner model**. But of course *in principle* human beings can*not* be fully modeled. Even if a user model is not explicitly implemented, at design time a designer has one in mind (otherwise she cannot design for interaction with users), hence we have to allow for the faultiness of this proposition nevertheless. From an educational standpoint, this has the consequence that every educational application has to be prepared for its own inadequacy. To understand the learner as an individual, autonomous human being requires technology that affords her **autonomous interaction**.

This also meshes well with the **self-referentiality of knowledge** mentioned above. Knowledge — and therefore content — is not static: it varies over time. Depending on what content is available, starting points for learning (or knowledge acquisition) differ and have to be flexible. Additionally, the learner's prior knowledge (which can change by a learning experience as well as simply forgetting) is a fundamental part of the learner's context, which must be modeled to enable effective learning: the learner very quickly gets annoyed, when her time is wasted by having to go through familiar learning objects or ones which quietly assume what isn't there.

We have seen above that the presentation of a learning object always includes a specific conceptualization of substance. Which of these available conceptualizations are used, can be decided e.g. based on the learner's **learning type** but also e.g. based on the learner's Community of Practice. A typical basic example for the distinction of the former consists in a differentiation between the 'visual type', who likes visual learning objects, versus the 'verbal type', who prefers their delivery in text form (according to the Felder-Silverman scale [FS88]). Moreover, the **personality type** can be differentiated and the software can adapt to it (e.g. using the Myers-Briggs Type Indicator [MM95] in [Jor02]). In [KK05] we use example of two conceptualizations of groups, each of which is common in a certain subfield of Mathematics. Their resp. use of one conceptualization above another — even though they are known to be equivalent — turns the choice into a practice of this respective community. Therefore, this can be considered an example of adaptation towards membership in a **Community of Practice (CoP)** [LW91, Wen99].

Note that technical representations of the learner context will take the content context into account, if only to reference it and to feed on its structure. For instance, the OMDoc-based ACTIVEMATH system [MAF$^+$03] references the content context to represent prior knowledge, and uses its dependency relations to prime a Bayesian network that calculates mastery values from user monitoring data. Generally we contend that handling of the learner's context can be much simplified by enhancing the content context and referencing it.

The context of learning naturally depends on the **situatedness** of the learner herself, her 'here-and-now', her experiences, and her expectations. Adaptation is possible here as well. For instance, learning objects can be correlated with user models that try to capture essential learning context information of an individual user like a history of visited learning objects. Another example consists in a learner's preference of operating system, or her favorite editor for interacting with learning software: in [Koh05b] the sensibility of regarding a user's past, present, and future yielded the concept of "Invasive Technology" as one adaptation factor for educational technology.

A final aspect of the user context lies in the **media** at the disposal of the learner: Mobile phones require a presentation of a learning object that is different from a large computer screen. Hence, the preparation of content has to be fitted to the output media format. Likewise the input media format has to be taken into account, e.g. an OLPC computer (see e.g. [OLP07]) has many more constraints for storing (or delivering) data than a high performance computer. We will not focus on these aspects and refer to the fields of "*Mobile Learning*", which deals with this and "*Micro Learning*" which deals with so-called micro content and explores the finest granularity of learning objects and its use.

3.3. The Interaction Context

In contrast to the two previous contexts, the interaction context is only active while the learner interacts with a particular teacher or software application and is therefore **short-lived**; we will also refer to it as the *learning/teaching* context. We view the interaction context as largely determined by the learning path played out up to the current moment, which is in turn determined by didactic strategies, interaction constraints, and a learner's actions. The time aspect is enhanced by the relevance of **meaning** in the interaction context. KLAUS KRIPPENDORF calls this "*becausality*" and bases it on the insight "*One always acts according to the meaning of whatever one faces*" [Kri06, p. 58]. In [Koh06b] we attribute this to a user's micro-perspective, i.e. her view from within, in a concrete interaction. In particular, the micro-perspective is decisive for a user's taking the action of using and thus determines her approach to software.

Again, we strive to model it on a very abstract level, delegating as much of the actual information to be content context. As we pointed out above, the underlying learning theories are quite abundant and any choice is subjective. Note that the teaching/learning context needs to be arranged according to this choice.

Adaptation of learning objects with respect to the interaction context has to integrate **organizational perspectives**. In contrast to the learner context, in which the integration is optional as it is ultimately a matter of personal choice, for the context of teaching it is a necessity. For instance, educational technologies for a university environment have to take the grading system into account. Another example of such an organizational view is the question of security within a system. Publicity of student's failures is as bad to the student as unintended free access to costly learning software to the software maker.

In blended-learning environments we have yet another set of educational requirements, which center around the educator herself. For example, if an educator aggregates learning objects, she might want to unify their layout in order to **express herself** as a consistent person *and* to supply **visual constancy** to her students. An example is the creation of a Microsoft PowerPoint presentation with the help of a "slide master". The invasive, semantic editor CPoint (see e.g. [Koh05a]) can import OMDoc learning objects which have to be fitted to the local presentation context by the aggregator.

Even though the context of teaching varies from one point to another, the **local coherence** of a learning situation (from the learner's standpoint) has to be taken into account as we consider understanding a holistic process. The well-known pedagogue JEROME BRUNER recapitulates in [Bru77, p. 12], that "*if earlier learning is to render later learning easier, it must do so by providing a general picture in terms of which the relations between things encountered earlier and later are made as clear as possible.*"

Interactivity as a feature of an educational technology has been shown to have positive effects for learning (e.g. the more interaction the better the learning). As no single system can do it all, the coordination of such (and their resulting usability) is a worthwhile goal for ET. Moreover, **cooperation** of students is generally considered fruitful for the learning process. Activities like sharing, reusing, or repurposing learning objects strengthen understanding.

4. Consequences for Semantic KM in ET

We have discussed above that data quality (the KM focus) and interaction quality (the focus of ET) are strongly interlinked. In particular, we can support ET with KM by enhancing the data quality in view of the educational contexts we discussed in the last section. Concretely, we will derive a set of requirements that would make semantic knowledge representation (KR) formats suitable as a basis for semantic ET. We will call these the **KR4ET conditions**.

Let us first consider the content context; we have argued that from an ET perspective the content context is structured into layers and dimensions. What kind of conditions can we derive from that for a KR format? We can discern abstract consequences and more concrete ones. In particular, such an abstract consequence consists in the fact, that we have to state what 'content' is to be. In other words we have to explicate its substance resp. substance equivalences, so that we can deal with content as an object. The implied objectivity is the basis for adapting it afterwards.

C1: Domain Context Modeling Setting up a context model for the domain seems evident, but we have to understand it as an agreement on what is considered constitutive and what is just nice to have for a knowledge object. As we have seen in Figure 3 and when we described the self-referentiality of knowledge, knowledge about knowledge objects relies on being able to anchor it in a semantic context that provides logical and social relations to other knowledge objects. From the learner perspective, this context and its structure must be explicitly represented to enable autonomous interaction and *enable* re-contextualization (see also **C3**).

C2: System Context Modeling The relevant context for ET is not restricted to the domain context alone. Therefore, the KR format must be ontologically neutral enough to deal with external system constraints like organizational structures, digital rights, or media types as well.

C3: Context Flexibility Only if the respective relations are made explicit, automated services can make use of them. But explicit context relations may unnecessarily fix the context relation, unless there is a way to relate contexts to each other.

Now, let us look at the more concrete consequences from the content context considerations. Here, we ask how the knowledge objects can become ET-usable (still on a very general scale). We have seen that these objects as learning objects *depend* on context and *contribute* to it — recall that substance and accidence form a dialectical pair. These dependencies and contributions have to be modeled. In other words, we like to explicate the potential accidences of the objects:

C4: Granularity of Representation Knowledge must be representable at multiple levels of granularity: from the level of a document down to the level of a single symbol. In particular, relying solely on document metadata is insufficient. Note that here the ground is laid for later access to various content layers.

C5: Referential Transparency All relevant parts of the knowledge objects represented should be referentiable and thereby retrievable by applications, either automatically or upon user request. Here, the coverage of breadth and depth of objects as content dimensions gets determined. Often this means that all knowledge elements are explicitly represented and have identifiers in an XML-based knowledge representation format.

C6: Ontological Transparency Structured collections of learning objects can serve the learner as a frame of reference for future communication and further learning. Therefore the representations of knowledge objects should include an infrastructure for ontological relations. In particular, the KR format can support the meaning-giving relation of anchoring concepts in others like already known or more primitive ones.

C7: Knowledge Object Portability Making use of a knowledge object for learning implies its potential portability, think e.g. of the accommodation process within the learning process. But when a knowledge object is retrieved, its dependencies on context should be preserved (e.g. by references), otherwise it might lose its substance qualities.

If we look back at the discussion of the space of learning objects in section 2.4 we see that learning paths play a great role in ET; we can even see aspects of the interaction context as given by the learning path. Take for instance a didactically enhanced document that introduces a new concept by first presenting a naive, reduced approximation $\mathscr{N}$ of the real theory object $\mathscr{F}$, only to show an example $\mathscr{E}_{\mathscr{N}}$ of where this is insufficient (we take $\mathscr{N}$ and $\mathscr{F}$ to be large-granular learning objects here). Then the document proposes a first "straw-man" solution $\mathscr{S}$, and shows an example $\mathscr{E}_{\mathscr{S}}$ of why this does not work in general. Based on the information gleaned from this failed attempt, the document builds the eventual version $\mathscr{F}$ of the concept and demonstrates that this works on $\mathscr{E}_{\mathscr{F}}$. Let us visualize the narrative- and content structure in Figure 8. The structure with the solid lines at the bottom of the diagram represents the content structure, where $\mathscr{N}$, $\mathscr{E}_{\mathscr{N}}$, $\mathscr{S}$, $\mathscr{E}_{\mathscr{S}}$, $\mathscr{F}$, and $\mathscr{E}_{\mathscr{F}}$ signify theory objects for the content of the respective concepts and examples. The arrows mark the conceptual dependency structure, e.g. theory $\mathscr{F}$ imports theory $\mathscr{N}$.

The top part of the diagram with the dashed lines stands for the narrative structure, where the arrows mark up the document structure. For instance, the slides sl_i are grouped into a lecture. The dotted lines between the

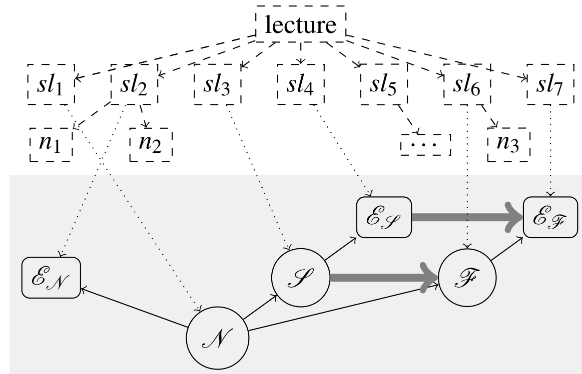

Figure 8. Concept Introduction via Straw-Man

two structures are pointers into the content structure. In the example in Figure 8, the second slide of "lecture" presents the first example: the text fragment n_1 links the content $\mathscr{E}_{\mathscr{N}}$, which is referenced from the content structure to slide 1. The fragment n_2 might say something like "*this did not work in the current situation, so we have to extend the conceptualization...*".

Stepping back from this concrete example, we can see that the situation in Figure 8 is an instance of the general setup: we can separate learning objects into two layers: A *narrative* and a *content* layer both of which consist of knowledge objects and are composed via relations (see e.g [VD04, Koh06c, KMM07b]). The presentational order of knowledge objects in documents is represented on the narrative layer, whereas the knowledge objects themselves and the ontological relations between them are placed in the content layer, which builds up the "content commons" [Tea06], i.e. a global, collaboratively authored and maintained learning resource. The connection between the narrative and the content layer is represented via *narrative relations*.

We can view this situation as an instance of the content/form distinction discussed above. The narrative structure represents the presentation, as it adds linearization and structure information.

C8: Document Representation The KR format needs to have a representation infrastructure for a wide variety of structured *documents*, including lectures, blogs, wikis, books, and essays.

C9: Discourse-Level Content/Form Infrastructure The KR format should allow the separation of content and form on the discourse level, as suggested in Figure 9: the lower level of the diagram represents the content of the knowledge (structured by the inherent semantic relations of the objects involved), and the upper part the form (structured, so that humans are motivated to concern themselves with the material, understand why some definitions are stated in just this way, and get the new information in easily digestible portions).

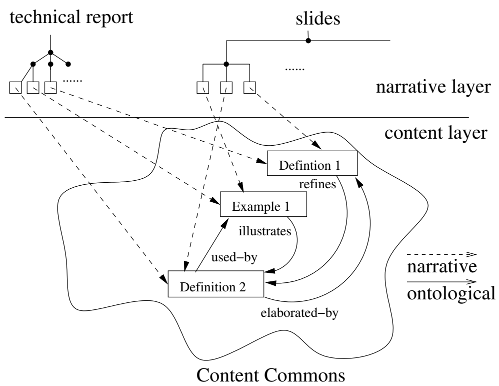

Figure 9. Narrative and Content Layers

Coming back to our example in Figure 8, we can see that the separation of narrative and content alone is not sufficient for *adaptation to a given learning/teaching context*, we also need the information that $\mathscr{S}$ is a potential straw-man example for $\mathscr{F}$, which we have indicated with the wide gray arrow in Figure 8. We need another content/form distinction here that distinguishes the root causes (i.e. the suitability as a strawman) from the particular presentation.

C10: Path-Level Content/Form Infrastructure The KR format should support the classification (of groups) of knowledge objects by their possible didactic role and relations to others, so that consistent learning paths can be derived from that. Note that the concrete classification depends on the respective learning theory, so that the information should not be realized in the representations, but attached from the outside so that different classifications and relations for different learning theories are possible.

The next two conditions concern the self-referentiality and dynamicity of knowledge we have discussed above. Knowledge representations have to deal with the dynamicity and hence to *manage change* to cope with this: Looking closely, we can see causes of change on three levels. First, the object of knowledge can change as we find out more about the world in science, or re-interpret historical development, or simply the world itself changes (e.g. the median ocean temperature). Secondly, the representation of the knowledge can change, e.g. when we correct errors in textbooks or come up with better explanations or exercises. Finally, if we use the knowledge representation format to represent the knowledge state of the user, then that changes as well over the course of a learning interaction or more generally over time — including of course that the user eventually forgets things.

C11: Terminological Extensibility One of the central aspects of learning is the extension of vocabularies by anchoring them in already-learned materials and building subsequent learning materials on the extended terminology. Therefore the KR format should provide a definitional infrastructure for extending terminologies as an integrated part of the language. Extensibility also helps to solve the bootstrapping problem (aka. the cold start problem) of learning.

C12: Management of Change Current KM systems are designed to coordinate the collaborative creation and maintenance process of document fragments and learning objects, often through the provision of a centralized repository. The focus of these systems is primarily on the documents themselves. Semantic relations between and within documents as well as effect of changes on these relations are largely neglected, although information reuse and distribution could seriously benefit by such relation management. Therefore human reviewers are needed for *management of change* to maintain consistency after modifications — a costly, tedious, and error-prone factor in document life-cycles that is often neglected to cut costs even though leading to sub-optimal results. Semantic management of change feeds on explicitly represented ontological relations that induce functional dependencies or non-interferences that allow to propagate the effects of change sets (see [MW07b] for details).

We have argued that learning more often than not is a collaborative and interactive process; this has to be supported by the KR format if it is to serve as a basis for semantic ET. For instance, STEPHEN DOWNES asks: "*What happens when online learning ceases to be like a medium, and becomes more like a platform?*" [Dow05]. For the purposes of this paper the 'platform' would be a semantic learning object management system that provisions the learning objects that make up a content commons. In the Web 2.0 era the user is increasingly being involved in creating, tagging, and aggregating the learning objects in the content commons following the paradigm of "user as *prosumer*" (i.e. as a "producer and consumer"). A similar situation emerges if we want to model the interaction context, which we can see as a dialogue "document" (see **C8**) also containing the user's answers. However, in contrast to the learning objects, which can be carefully prepared by an author — elevating them to content markup in advance — the learner contributions will either be form interactions (e.g. as answers to multiple-choice questions) or free-form text. To support learner contributions the KR format needs to ensure that ET systems can deal with such content gracefully.

C13: Semantics as Upgrade The knowledge representation format must allow a stepwise refinement of legacy documents into semantically enhanced learning objects. This is a matter of practical importance, as the depth of semantic modeling varies with the intended application and author dedication.

C14: Graded Functionality It is important that applications can degrade gracefully from high-impact services feeding on deep semantic relations to trivial services in the absence of non-trivial semantic annotations. The KR format has to provide the necessary infrastructure for this.

C15: Semantic Integration The ontology infrastructure from **C6** should be interlinked with other ontological resources (e.g. via translation or RDF extraction; see 5.2). This consequence is motivated by the fact that in larger learner communities, users will tend to use a diverse variety of tools. If they are to collaborate in a semantically meaningful way, the KR format needs to be well-integrated with competing forms of specifying semantic information.

One of the natural concerns in a content commons is to foster reuse of content, i.e. to foster a work flow using an "identify-and-reference" rather than a "copy-and-paste" procedure where possible. In ET scenarios, this is especially important since reuse — apart from reducing redundancy and thus storage costs and bandwidth — enhances the accuracy of anchoring learning objects.

C16: Structure Sharing The knowledge representation format should support structure sharing. Stronger referential transparency and portability (see **C5** and **C7**) usually allows stronger structure sharing in principle, but this must be supported by the knowledge format and the inscribed interaction design.

To implement the adaptation capabilities discussed above in ET systems, we need a source of information about the learner context, which forces us to represent it in the machine. We expect that information about learning type, periphery constraints, etc. are largely non-semantic and can be modeled with conventional user modeling technologies, so they do not have consequences for the knowledge representation and we concentrate on learner preferences and prior knowledge here.

For the *learner preferences* we will take notation preferences for mathematical formulae as a paradigmatic example. It is well-known that mathematical notations may vary, even for standard functions like binomial coefficients: depending on academic culture, $\binom{n}{k}$, ${}_nC^k$, C^n_k, and C^k_n all mean the same thing: $\frac{n!}{k!(n-k)!}$ or equivalently "the number of ways I can pull a sample of k balls from a sack of n balls".

The content/form distinction already suggests to use content representations for storage and generate adapted presentations as accidences from that. If we have a content representation for notation definitions, these can be made part of the content commons and managed with the other knowledge.

C17: Notation Definitions Having explicit notation definitions in the context can be used to simplify the representation of user notation preferences to a mere referencing scheme, as this only needs to reference and prioritize notation definitions in a "notation context" [KMM07b] for each document fragment. This is then reconciled with the author-supplied notation contexts for construction of the user-adapted document (otherwise, imagine a teacher wanting to contrast two notations common in the literature and the user model overwrites both, leaving the learner without new information).

C18: Substance Equivalences The same can in principle be done with a representation of prior knowledge and mastery levels if we have content representations of substance equivalences. Substance equivalences contain mappings with which we can generate members from substance equivalence classes from each other or from some more abstract content representation, allowing for more structure sharing and reuse.

C19: Variant Relations and Dimensions Where we lack computational representations of substance equivalences, we have to *store* all members of a substance equivalence class in the content commons. To make the substance equivalence relation explicit, we have to be able to annotate the fact that the knowledge elements are substance variants, and the dimension, which characterizes them in the equivalence class, as well. If we go back to Figure 4, then representations $\mathscr{G}_1^g$ and $\mathscr{G}_1^e$ are language variants, which means they differ in the language dimension, where the first has the value "German" and the latter the value "English". Note that we need an extensible vocabulary for language dimensions, their values, and variant relations (for instance, some are symmetric like the language translation relation, while some are not: e.g. the abbreviation relation), therefore the variant relation should be anchored in its own system context model (see **C2**).

But whether we store the accidences or generate them on the fly, in both cases we face a usability problem: we can only adapt to parts of the learner context we have already modeled from the learner's behavior in the interaction. Thus the representation of the learner context must have capabilities that allow predictions of user preferences and prior knowledge.

For predicting learner preferences like notations and familiarities (e.g. to substance equivalences) it is crucial to observe that these are not arbitrary, but the result of earlier learning situations and interaction histories. In other words, notations and substance equivalences depend heavily on the meaning-assigning practices of communities the learners are involved in. We have proposed an extensional model of such Communities of Practice based on the documents learners interact with in [KK06]. The main idea in this analysis is that many of the community-specific practices are inscribed into documents used by the community. If the documents are represented in a semantic format, the practices will have been reified and made explicit, and can therefore be harvested for a CoP model. Note that this is an extensional model that only concerns itself with the practices and their distributions over the communities, not with the social mechanisms of the communities themselves. Such extensional CoP models are exactly what we need for predictions about user preferences and familiarities — assuming of course that these are related to the practices of the respective CoPs: If a learner is associated to a CoP that prefers p_1 over p_2 and is familiar with f, then the learner will be likely to as well. This brings us to our next consequence

C20: Practices and CoPs The knowledge representation format should support the representation of all relevant practices and allow the modeling of CoPs.

But how do we get access to the documents a learner is involved with, so that we can determine CoPs and practices? Here the Web 2.0 comes to the rescue in the form of Social Tagging (ST) systems which celebrate such enormous growth rates on the World Wide Web (e.g. [GM06]). We argue that the high acceptance rates of ST are based on its meaningful interaction process with respect to conceptualization [KR08]. In particular, these systems make use of the fact that they enable an embodiment of concept development, i.e. *embodied conceptualizations*, and the underlying processes are therefore valuable for individual learning. Following [Wal06] we consider a tag as expression of the specific interest this person (with her own identity) has in the object at hand, which determines her vocabulary and that thereby provides a defining relationship in form of metadata. His "triad of object, identity, and metadata" is at the heart of private tagging approaches, and we can interpret his "dual folksonomy triad" — consisting in object, community, and metadata — as a description of the transformation process from the private to public tagging, yielding emergent folksonomies [Wal04] where the "*navigation structure is called "***folksonomy***" — short for "folks" and "taxonomy" because of its quality as a bottom-up organized, decentralized hierarchic structure*" [KR08]. In particular, (personal) conceptualization gaps can be filled with suggestions by community information. In [KR08] we suggest, that exactly the *fuzzy line between private and public* while tagging enables and enhances learning processes as dynamic folksonomies force the user to constantly go through the coupling process, thereby reflecting on the connection between meaning and tag. In terms of the knowledge space the learning process in these systems is pushed by subjective substance equivalences: tags are the assignment of meaning and the underlying assumption consists in the fact, that such titles or classifications represent substance equivalence relations of the assigned objects.

This analysis shows a way towards realizing a CoP-aware representation of the learner context:

C21: Social Tagging of Learning Objects Private tagging directly gives us the document space for deriving private practices from; social tagging gives us tag/document clouds to derive extensional CoP models.

Moreover, in the semantic arena, we can use the representation of the content context as the tag space. Learners can tag external document fragments (or learning objects) with references to their own KM-supported vocabularies (see **C11**), thus establishing (perceived) substance equivalence relations between the content context of the learner (as part of the learner context) and standard content contexts (e.g. the Wikipedia or the university curriculum). This solves one of the big conceptual problems for knowledge representation in educational technology: If we base (adapted) presentations of learning objects as a basis for learning, the learner appropriates these to form his (private) content context, how does this relate to the teacher's, the community's, or the curricular content context. The semantic ST triad (both in the primary as well in its dual form) give rise to a tight feedback loop that leads to semantic CoP-based folksonomies. Designers pay attention to their own or their intended CoP's underlying understanding of substance equivalences, as otherwise constitutive aspects of meaning might get lost in the transition. Moreover, it makes good sense to develop a model for CoPs that can be explicitly embedded in future semantic data formats to enable a broadened range of knowledge sharing practices crossing CoP boundaries. Note that a recognition of these substance equivalences will also enable ET systems to offer knowledge realized in an expert's CoP "X" fashion in a novice's CoP "Y" fashion, thereby strengthening their user-adaptability to support learning processes.

5. Towards Multi-Context Knowledge Representation for ET

We will now look at how the consequences identified in the last section can be *realized* in a knowledge representation format and how semantically enabled services can make use of the structures realizing them. We will base our discussion on our OMDoc format (Open Mathematical Documents) [Koh06c], an XML-based content-oriented representation format for scientific documents, which is now used in a large set of projects in automated theorem proving [Mül06a], eLearning [MBG+03], eScience [HKS06], document retrieval [KŞ06], and in formal digital libraries [Log06]. Note that even though the OMDoc format is originally geared towards mathematical knowledge, the concepts carry much further. We view mathematics with its explicit structure and management of context as a test tube domain which allows us to identify the relevant representational primitives. Experience shows that these are applicable at least to the hard sciences: OMDoc has been used as-is for Computer Science course materials, extended to Physics [HKS06], and a version for Chemistry is under development.

5.1. OMDoc: Open Mathematical Documents

To understand the OMDoc format, we need to distinguish it from the two main paradigms, which essentially differ in the depth of modeling of the domain knowledge, in the coverage and scalability, and in the formalisms employed.

First, the *Semantic Web* [BL98] is an approach that should be web-scalable in principle. However, the underlying context knowledge must be provided in an ontology formalism like OWL [MvH04]. This representation format is intentionally limited in its semantic expressiveness, so that inference stays decidable and web-scalable. Unfortunately, scientific knowledge can be only approximated very coarsely using this approach so far.

In contrast, the field of *Formal Methods* [Win90] use semantic formats with highly expressive knowledge representation components. They are currently only used for security sensitive applications, such as formal program verification, since on the one hand they require the commitment to a particular logical system, and on the other hand the mathematical-logical formalization needed for formal verification is extremely time-consuming.

In contrast to those, the **structural/semantic approach** taken by the OMDoc format does not require the full formalization of mathematical knowledge, but only the explicit markup of important structural properties. For instance, a statement will already be considered as "true" if there is a proof object that has certain structural properties, not only if there is a formally verifiable proof for it. Since the structural properties are logic-independent, a commitment to a particular logical system can be avoided without losing the automatic knowledge management, which is missing for semantically unannotated documents. Of course, OMDoc only supports structural plausibility checks for quality management instead of full verification. Work on the OMDoc format shows that most services in Knowledge Management do not need tedious formalization, but can be based on the structural/semantic level. It is a major aspect of our work that we do not take the all-or-nothing approach of Formal Methods where we either guarantee full correctness of a theorem, or do not give any support.

The OMDoc format builds on a semantic representation format for mathematical formulae (OpenMath objects [BCC+04] or Content MathML expressions [ABC+03]) and extends this by an infrastructure for context and domain models. OMDoc uses a four-layered structure model of knowledge.

Object level This represents objects such as complex numbers, derivatives, etc. for mathematics, molecules in chemistry, map specifiers for geo-sciences, or observables for physics. Semantic representation formats typically use functional characterizations that represent objects in terms of their logical structure, rather than specifying their presentation. This avoids ambiguities which would otherwise arise from domain specific representations.

Statement Level The (natural/social/technological) sciences are concerned with modeling knowledge about our environment, or more precisely, with statements about the objects in it. We can distinguish different types of statements, including model assumptions, their consequences, hypotheses, and measurement results. All of them have in common that they state relationships between objects and have to be verified or falsified in theories or experiments. Moreover, all these statements have a conventionalized structure, and a standardized set of relations among each other. For instance, a model is fully determined by its assumptions (also called *axioms*); all consequences are deductively derived from them (via *theorems* and *proofs*); hence, their experimental falsification uncovers false assumptions of the model. Proofs are only one example of *provenance information* that is encoded in the statement level, the trail from a measurement, via data processing, to presentation in a chart is another.

Theory/Context Level Representations always depend on the ontological context; even the meaning of a single symbol is determined by its context — e.g. the glyph *h* can stand for the height of a triangle or Planck's quantum of action — and depending on the current assumptions, a statement can be true or false. Therefore, the sciences (with mathematics leading the way) have formed the habit of fixing and describing the context of a statement. Unfortunately, the structure of these context descriptions remain totally implicit, and thus cannot be used for computer-supported management. Semantic representation formats make this structure explicit. For instance in mathematical logic, a theory is the deductive closure of a set of axioms, that is, the (in general infinite) set of logical consequences of the model assumptions. Even though in principle this fully explains the phenomenon of context, important aspects like the re-use of theories, knowledge inheritance, and the management of theory changes are disregarded completely. Hence, formalisms that have a context level use elaborate inheritance structures for theories, e.g. in the form of ontologies for the Semantic Web or as "algebraic specifications" in program verification.

Document Level The OMDoc format supports the separation documents into *narrative* and *content* layers according to Figure 9 as described in section 4. We do not claim to have invented this concept, but the OMDoc format probably implements this idea in the cleanest way; see [Koh06c, KMM07b] for details.

An important trait of the four-layer language architecture is the inherent dependency loop between the object- and theory levels mediated by the statement level: The objects obtain their meaning from the theories in which their functional components are at home, and the theories are constituted by special statements, and in particular by the objects that are contained within these statements. Experience shows that the four-level hierarchy provides a good model of the "scientific method" and indeed the whole corpus of scientific knowledge. This structure implicitly pervades scientific discourse. Making these structures explicit allows for the mechanization and automation of Knowledge Management and the unambiguous, flexible communication of mathematical objects and knowledge that is needed for meaningful interoperability of software systems in science.

Of course, some of the features discussed here are not unique to OMDoc: for instance the format CNXML [HG07] used by the CONNEXIONS project [Tea06] covers the object-, documents-, and part of the statement layer introduced above. Similarly, the LaTeX-based MMISS format [KBLL+04] covers the statement- and (parts of) the context level. Finally, the OpenMath [BCC+04], MathML [ABC+03], and CML (Chemistry Markup Language) [MR+07] provide strong object levels representation infrastructures specialized to their respective disciplines, and have a flexible mechanism of meaning assignment via a simple context layer.

5.2. KM Formats and Knowledge Models

Note that all of the formats mentioned above integrate content, context, and document markup in the form of control sequences (e.g. as XML elements or LaTeX macros) into natural language text. The specific markup reflects the various knowledge objects of the respective format and their relations among each other. We consider the **knowledge model** of the format (or **format ontology**[5]) as primary, and the specifics of the implementation e.g. in XML elements and attributes as secondary. If the knowledge models of two KM formats are compatible, we can always translate them into each other. In Semantic Web Technology, existing document models like HTML are used for the representation of learning object documents and the RDF [LS99] format is used to classify text fragments as knowledge objects and markup their relations: the text fragments are identified by URI references in subject/verb/object assertions (**RDF triples**) where the verb represents the intended relation. HENRY THOMPSON and DAVID MCKELVIE speak of **standoff markup** for this style

[5] We use this term for the (fixed set of) relations between the knowledge items identified by the KR format to distinguish it from the dynamic "domain ontology", which codifies the objects of the subject covered by the encoded learning objects.

of adding semantic information to documents externally [TM97]. The aspects of the format ontology that can be represented in a web ontology format like OWL [MvH04] can be supported by general-purpose inference mechanisms. Note that the standoff and integrated styles for semantic markup are equivalent in expressivity and their differences largely pragmatic: the former can be added to read-only documents, while the latter is more likely to be adapted while changing the learning objects. As mentioned above, we view the main contribution of these formats in their knowledge model design; if a corresponding format ontology exists, a translation to a RDF/OWL-based implementation — we speak of **RDF extraction** — is a relatively trivial exercise.

5.3. OMDoc and the KR4ET Conditions

The OMDoc format is geared towards providing an explicit context model. It represents the relevant *domain knowledge* and supports machine-supported Knowledge Management through its explicit structure. We will now see how this allows to answer the conditions from section 4:

ad C1: Domain Context Modeling The OMDoc format provides a complex infrastructure for modeling context in "theories". These group concepts and statements that give them meaning, and structure the context into a definitional inheritance hierarchy. Any representation at the object and statement level is annotated with its "home theory", which furnishes the content context. At the object level, this principle is carried to the extreme: any symbol and concept is determined by its name and home theory; thus the context of an object is modeled as the (structured) collection of the home theories of the symbols and concepts occurring in its representation.

ad C2: System Context Modeling OMDoc theories are ontologically unconstrained and allow natural language for defining concepts, but the infrastructural aspects — e.g. *definiendum*, i.e. which concept is defined, the *definiens*, i.e. by what is it defined, and the relations to other concepts — are marked up explicitly. Therefore there is no restriction on the type of material covered in the context.

ad C3: Context Flexibility OMDoc supports a notion of **theory interpretations** [FGT92, RK08] which allows concept interpretation via complex mappings and semantic views via "postulated theory interpretations". Generally, we speak of theory interpretations, if all concepts and symbols of the source theory are interpreted by those of the target theory via the translation, and the translations of all model assumptions in the source theory are fulfilled in the target theory. For instance, unit conversions give rise to theory interpretations: take a function f that maps 100° Celsius to 212° Fahrenheit, then this induces a theory interpretation, since it maps the (defining) assumption that '*water boils at* $100°C$' to the (true) assertion that '*water boils at* $212°F$' (think of "Spaghetti Carbonara"). The setup of the OMDoc theory system guarantees that theory interpretations translate true statements in the source theory to true statements in the target theory, which makes them a semantically founded instrument for transporting insights between learning situations and for recontextualization. Note that such theory interpretations account for many of the substance equivalences; in our example the inverse function f^{-1} also induces a theory interpretation, so the 'Fahrenheit' and 'Celsius' theories are equivalent in substance.

ad C4: Granularity of Representation As we have seen above, OMDoc offers markup at four levels; on the object level every symbol in mathematical formulae can be semantically anchored.

ad C5: Referential Transparency Using the referential apparatus of the underlying XML format, OMDoc allows *identifiers* and *names* on all semantically meaningful fragments. With this, we have two global (i.e. web-wide) referencing schemes at our disposal: the first allows referencing via standard Uniform Resource Locators (URLs), and the second gives us semantic referencing via *theory interpretation access paths*, see [RK08] for details and an encoding of these via URIs.

ad C6: Ontological Transparency OMDoc can be viewed as an ontology language in itself. It can define symbols and concepts and specify their relations in OMDoc documents (content dictionaries). In this approach the OMDoc content dictionaries can directly be utilized as reified learner contexts via the learning paths that cover them.

ad C7: Knowledge Object Portability
Representations of OMDoc learning objects are intrinsically portable, since all their contributions to and dependencies on context are made explicit by referential links; so they can be moved about, copied, and referenced without loss of content. Only theory-constitutive statements have

to be contained in the theory-representation, since they directly determine the meaning, separating them from the theory would radically change meaning of the theory.

ad C8: Document Representation The OMDoc format provides a simple document markup language that allows to mark up generic document sectioning hierarchies and provides a subset of the HTML text structuring elements like lists, tables, etc. With the modular design of the OMDoc language it is simple to extend this to other document models if desired.

ad C9: Discourse-Level Content/Form Infrastructure Just as for content-based systems on the formula level, there are now ET systems that generate presentation markup from content markup, based on general presentation principles, also on this level. For instance, the ACTIVEMATH system [MAF+03] generates a simple narrative structure (the presentation; called a personalized book) from the underlying content structure (given in OMDoc) and a user model.

ad C10: Path-Level Content/Form Infrastructure The OMDoc format itself only supports this by supplying the necessary preconditions: the document-level content/form infrastructure (cf. **C9**) and the fine-grained content markup (cf. **C4**). In analogy to the object-level notation definitions, which specify notation definitions using content markup patterns to trigger specific presentations [KLR07], we need statement- and even theory-level patterns for path-level document generation. But these are much more difficult to support as the knowledge objects and their didactic relations may be scattered over the content commons. Current applications that support path-level document generation [LMU01] hard-code the matching in the generation algorithm. To arrive at a scalable system we need an effective query language and content retrieval system that takes all four levels of modeling into account. We are currently working on the OMBASE system based on distributed XML database technology with the hope of achieving this.

ad C11: Terminological Extensibility OMDoc provides statement-level elements for defining object concepts and symbols, essentially enabling the authors to extend the vocabularies needed to describe objects, their properties and behaviors in *content dictionaries*, i.e. special OMDoc documents optimized as ontological references. Note that learning materials often take the form of content dictionaries by their very nature.

ad C12: Management of Change The dependency relations induced by theory interpretations and occurrences of symbols and concepts in statements as well as objects can be used for a semantically motivated management of change and distributed collaboration, which propagates changes along semantic relations.

For instance, if we change a concept definition in one learning object, then this affects all the learning objects that depend on it (their foundation und thus their meaning has changed). The conservative solution, i.e. to declare the changed learning object as a new knowledge item, leads to drastically weakened reuse factor violating **C16**. In knowledge collections encoded in OMDoc we can make use of the dependency relation to propagate the potential effects of changes (or more pragmatically non-interference of changes). We can fine-tune change propagation by semantically classifying changes, as certain document and content dependency relations are blind against certain classes of changes; see [Mül06b] for details.

ad C13: Semantics as Upgrade Like many other XML-based representation formats, OMDoc employs semantic annotations to mark up the semantic relations of text elements. In contrast to many formal approaches this leaves the choice of the depth of markup to the user. For instance, existing course materials can be migrated to a semantic collection of learning objects using an *invasive editor* [Koh05b] for OMDoc, e.g. the CPOINT system [Koh06a] for MS PowerPoint, or the STEX system [Koh07], a semantic variant of LATEX that supports translation to OMDoc.

ad C14: Graded Functionality In OMDoc-based systems grading is a simple consequence of the fact that the language primitives are largely orthogonal, hence do not interact, and are implemented modularly in the language. Therefore each of the answers to the consequences above only depends on a minimal set of representation requirements.

ad C15: Semantic Integration As OMDoc allows formal annotations, OWL statements can be embedded into OMDoc documents and then harvested by web applications [Brö07]. The integrated approach to ontologies is on marked contrast to web ontology languages like

OWL [MvH04], which specify ontological relations between web resources outside the resources themselves. Moreover, the ontological relations can be exported in the presented learning objects in the form of RDFA annotations, which reference the OMDoc system ontology [Lan07a]. From a practical perspective note that in the integrated approach ontological information is less likely to become out of sync with the underlying learning objects.

ad C16: Structure Sharing OMDoc represents scientific objects like mathematical formulae, chemical molecules, or code fragments as content representations and supplies declarative notation definitions. This measure alone directly supports the reuse of learning objects for different user communities and across learning paths, as resolves the well-known notation hurdles for reuse. For instance a mathematician can now reuse learning objects authored by an electrical engineer even though the former uses i for the imaginary unit of the complex numbers while the latter uses j. Intuitively, this approach explicitly represents objects that are as far to the left as possible in local knowledge spaces like the one in Figure 5. As these fix a minimal amount of accidences, adaption is just concretization.

OMDoc also supports reuse and sharing at a higher level by the theory interpretations mentioned above: learning objects can be re-interpreted for reuse in different contexts. Coupled with knowledge representation using the "little theories" approach [FGT92], this is a surprisingly powerful but principled reuse infrastructure, which is based on the reification of substance equivalences as theory interpretations.

ad C17: Notation Definitions The OMDoc format has an embedded language for notation definitions, which has recently been extended for the upcoming OMDoc1.8 in [KLR07]. With these we can generate any of these from the OpenMath or content MathML representation in OMDoc.

ad C18: Substance Equivalences Substance Equivalences at the statement levels can be represented by the special `alternative` relation in OMDoc. This allows to mark statement-level constructs as substance-equivalent, iff their logical equivalence can be proved in the system. Alternative definitions provide substance equivalences at the object-level, and theory morphisms at the theory level.

ad C19: Variant Relations and Dimensions OMDoc1.2 only supports a very limited set of variant relations: language translations for natural language content, and logical system variants in formal content: variants are siblings in superordinate statements, the variant dimensions are specified by the `xml:id` and the `system` attributes respectively. For OMDoc1.8 we are working on a variants module along the lines of **C19**, see [KMM07a] for details.

ad C20: Practices and CoPs OMDoc1.2 does not support CoPs, but does model some mathematical practices that can be used to identify CoPs and that allow collections of OMDoc documents to serve as an extensional CoP models [KK06]. For OMDoc1.8, we are currently extending the infrastructure for CoP modeling; see [Mül07] for first results.

ad C21: Social Tagging of Learning Objects The OMDoc1.2 does supports semantic social tagging by offering fine grained identification and markup for text fragments that can be tagged. For OMDoc1.8, we are currently experimenting with semantic tagging schemes in the Panta Rhei system, a community-aware OMDoc reader [MK07], and we will include a first tagging scheme that supports the evolution of "semantic folksonomies" in OMDoc1.8.

The OMDoc format represents only one set of concrete design decisions. Systems like CNXML [HG07] used by the CONNEXIONS project [Tea06] are based on a different set. Figure 10 gives an overview over the situation with respect to the KR4ET conditions. Note that in the last column we have added the pure Semantic Web approach — i.e. without a knowledge model expressed in a format ontology — as a baseline for the comparison (see section 5.2 for a discussion) so that we can pinpoint the contribution of the other format's knowledge model.

6. Conclusion

In this paper we have tried to understand semantic technologies for education from a foundational perspective. We take the term 'semantics' to mean "based on a collection of reified knowledge objects whose context and relation among each other are explicitly represented". In this view, the interplay of educational technologies and knowledge representation becomes central. Moreover, in this symbiosis, KM has to support the Knowledge Management needs of ET systems to support them in their intended functionality. Therefore we have (re)-evaluated the notions of 'semantics',

Cond	Title	OMDoc1.2	OMDoc1.8	CNXML	RDF/OWL
C1	Domain Context Modeling	+	+	+	+
C2	System Context Modeling	+	+	±	+
C3	Context Flexibility	+	+	+	—
C4	Granularity of Representation	+	+	+	+
C5	Referential Transparency	±	+	+	+
C6	Ontological Transparency	+	+	+	+
C7	Knowledge Object Portability	+	+	—	—
C8	Document Representation	+	+	+	—
C9	Discourse-Level Content/Form Infrastructure	+	+	±	na
C10	Path-Level Content/Form Infrastructure	—	—	—	—
C11	Terminological Extensibility	+	+	+	±
C12	Management of Change	±	+	—	na
C13	Semantics as Upgrade	+	+	+	na
C14	Graded Functionality	+	+	—	—
C15	Semantic Integration	+	+	—	±
C16	Structure Sharing	+	+	±	na
C17	Notation Definitions	+	+	—	na
C18	Substance Equivalences	±	+	—	—
C19	Variant Relations and Dimensions	—	+	—	na
C20	Practices and CoPs	—	±	—	na
C21	Social Tagging of Learning Objects	—	±	±	na
KR4ET support: + $\widehat{=}$ full, — $\widehat{=}$ none, ± $\widehat{=}$ partial, na $\widehat{=}$ not applicable (no format ontology)					

Figure 10. Language Comparison

'knowledge', and 'learning' with respect to their role for learning materials in ET, and how they interact with the contexts involved in the interaction process. Much of this analysis has crystallized around the notions of user/learner/interaction context and a structured knowledge/learning space which clarifies the roles of the content/form distinction, both of which are at the heart of semantic technologies. The provisioning of learner-adapted learning materials can now be seen as a process of choosing coherent learning paths in this (largely virtual) space of alternatives. From this analysis we have distilled a collection of twenty one conditions which an ideal knowledge representation format should meet to allow a content commons that is suitable to support semantic ET systems.

Note that we have only analyzed the KR/ET interaction from an information-theoretic point of view to be able to control the complexity of the issue. We analyze what information an ideal KR format has to represent for ET applications. We do not specify what the ideal KR format should look like, as we believe that there are a lot of ways to fulfil the conditions. These KR4ET conditions also do not suggest how a Knowledge Management system should implement the functionality to manage a content commons. We are currently experimenting with invasive editors for the MS office suite [Koh05b] and LaTeX [Koh07], a semantic wiki [Lan07b, LK07], and a community-based forum [MK07], which share OMDoc as a common KR format and are based on a shared knowledge base back end. This KR-based approach is already yielding useful synergies at the system implementation level. These allow us to experiment more readily with competing system designs which we can evaluate and compare to obtain better semantic ET systems.

We expect that the abstract formulation of the KR needs of ET systems in the form of the 21 KR4ET conditions will allow us to make different KR-based ET systems more comparable, creating a similar synergy/competition situation. We also hope that an evaluation of the underlying KR via the KR4ET conditions will prompt KM system designers to complete their systems with the functionalities they are still missing.

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.

[Ack04] Edith K. Ackermann. Constructing knowledge and transforming the world. In M. Tokoro and L.Steels, editors, *A learning zone of one's own: Sharing representations and flow in collaborative learning environments*, volume 1, pages 15–37. IOS Press, 2004.

[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.

[BD00] John Seely Brown and Paul Duguid. *The Social Life of Information*. Harvard Business School Press, 2000.

[Bel05] Johannes Bellmann. Selektion und Anpassung: Lerntheorien im Umfeld von Evolutionstheorie und Pragmatismus. *Zeitschrift für Pädagogik*, 49. Beiheft(51):62–76, April 2005.

[BF06] Jon Borwein and William M. Farmer, editors. *Mathematical Knowledge Management, MKM'06*, number 4108 in LNAI. Springer Verlag, 2006.

[BL98] Tim Berners-Lee. The semantic web, 1998. Available at http://www.w3.org/DesignIssues/Semantic.html, seen July 2006.

[BLHL01] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American Online*, 05 2001. viewed at 2005-09-26.

[Brö07] Matthias Bröcheler. The mathematical semantic web. Bachelor's thesis, Computer Science, Jacobs University, Bremen, 2007.

[Bru77] Jerome Bruner. *The Process of Education*. Harvard University Press, 196o,1977.

[CoS94] National Research Council and National Academy of Sciences. *Information Technology in the Service Society*. National Academies Press, 1994. ISBN 0309048761.

[Dew33] John Dewey. *Experience and Education*. New York: Macmillan, 1933.

[DI05] Patrick Dunn and Mark Iliff. At cross purposes: Why e-learning and knowledge management dont get along. Online at http://www.learninglight.eu, viewed at 2007/07/04, 2005. Learning Light.

[Doe] Beat Doebe. Pedagogical theories. Online at http://beat.doebe.li/bibliothek/f00048.html.

[Dou03] Paul Dourish. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, 2003.

[Dow04] Stephen Downes. Reusable media, social software and openness in education. Online at http://www.downes.ca/post/7804, September 2004. Seen on 2007-11-12.

[Dow05] Stephen Downes. E-learning 2.0. Online at http://elearnmag.org/subpage.cfm?section=articles\&article=29-1, 2005. eLearn Magazine.

[DP98] Thomas H. Davenport and Laurence Prusak. *Working Knowledge*. Harvard Business School Press, 2000 edition, 1998.

[dS05] Clarisse Sieckenius de Souza. *The Semiotic Engineering of Human-Computer Interaction*. MIT Press, 2005.

[FB04] Jodi Forlizzi and Katja Battarbee. Understanding experience in interactive systems. In *DIS '04: Proceedings of the 5th conference on Designing interactive systems*, pages 261–268, New York, NY, USA, 2004. ACM.

[FGT92] William Farmer, Josuah Guttman, and Xavier Thayer. Little theories. In D. Kapur, editor, *Proceedings of the 11th Conference on Automated Deduction*, volume 607 of *LNCS*, pages 467–581, Saratoga Springs, NY, USA, 1992. Springer Verlag.

[FS88] R. M. Felder and L. K. Silverman. Learning and teaching styles in engineering education. In *Engineering Education*, volume 78, pages 674–681, 1988.

[GJ02] William S. Green and Patrick W. Jordan, editors. *Pleasure With Products: Beyond Usability*. London and New York: Taylor & Francis, 2002.

[GM06] Laura Gordon-Murnane. Social bookmarking, folksonomies, and web 2.0 tools. *Searcher Mag Database Prof*, 14(2):26–38, 2006.

[Hei99] Helmut Heid. Über die Vereinbarkeit individueller Bildungsbedürfnisse und betrieblicher Qualifikationsanforderungen. *Zeitschrift für Pädagogik*, 2:231–244, march/april 1999.

[HG07] Brent Hendricks and Adan Galvan. The Connexions Markup Language (CNXML). http://cnx.org/aboutus/technology/cnxml/, 2007. Seen June 2007.

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

[HKS06] Eberhard Hilf, Michael Kohlhase, and Heinrich Stamerjohanns. Capturing the content of physics: Systems, observables, and experiments. In Borwein and Farmer [BF06].

[Hol95] Klaus Holzkamp. *Lernen: Subjektwissenschaftliche Grundlegung*. Campus Verlag, 1995.

[Jor02] Patrick W. Jordan. *Pleasure With Products:*

Beyond Usability, chapter The Personalities of Products, pages 19–47. London and New York: Taylor & Francis, 2002.

[Jün04] Sebastian Jünger. *Selbstorganisation, Lernkultur und Kompetenzentwicklung: Theoretische Bedingungsverhältnisse und praktische Gestaltungsmöglichkeiten*. Deutscher Universitätsverlag, 2004.

[Kap06] Mitchell Kapor. *Bringing Design to Software*, chapter A Software Design Manifesto, pages 1–9. Addison-Wesley, 1996 (2006).

[KBLL+04] Bernd Krieg-Brückner, Arne Lindow, Christoph Lüth, Achim Mahnke, and George Russell. Semantic interrelation of documents via an ontology. In G. Engels and S. Seehusen, editors, *DeLFI 2004*, volume P-52 of *LNI*, pages 271–282. Springer-Verlag, 2004.

[KK05] Andrea Kohlhase and Michael Kohlhase. An exploration into the mathematical knowledge space. In Michael Kohlhase, editor, *Mathematical Knowledge Management, MKM'05*, number 3863 in LNAI. Springer Verlag, 2005.

[KK06] Andrea Kohlhase and Michael Kohlhase. Communities of Practice in MKM: An Extensional Model. In Borwein and Farmer [BF06], pages 179–193.

[KLR07] Michael Kohlhase, Christoph Lange, and Florian Rabe. Presenting mathematical content with flexible elisions. In Olga Caprotti, Michael Kohlhase, and Paul Libbrecht, editors, *OpenMath/ JEM Workshop 2007*, 2007.

[KMM07a] Michael Kohlhase, Achim Mahnke, and Christine Müller. Managing variants in document content and narrative structures. In Hinneburg [Hin07], pages 324–229.

[KMM07b] Michael Kohlhase, Christine Müller, and Normen Müller. Documents with flexible notation contexts as interfaces to mathematical knowledge. In Paul Libbrecht, editor, *Mathematical User Interfaces Workshop 2007*, 2007.

[Koh05a] Andrea Kohlhase. Cpoint, 2005. http://kwarc.info/projects/CPoint/.

[Koh05b] Andrea Kohlhase. Overcoming proprietary hurdles: Cpoint as invasive editor. In Fred de Vries, Graham Attwell, Raymond Elferink, and Alexandra Tödt, editors, *Open Source for Education in Europe: Research and Practise*, pages 51–56. Open Universiteit of the Netherlands, Heerlen, 2005.

[Koh06a] Andrea Kohlhase. *An OMDoc Editor in MS PowerPoint*, chapter 26.14, pages 301–305. Springer Verlag, 2006.

[Koh06b] Andrea Kohlhase. The User as Prisoner: How the Dilemma Might Dissolve. In Martin Memmel, Eric Ras, and Stephan Weibelzahl, editors, *2nd Workshop on Learner Oriented Knowledge Management & KM Oriented e-Learning*, pages 26–31, 2006. Online Proceedings at http://cnm.open.ac.uk/projects/ectel06/pdfs/ECTEL06WS68d.pdf.

[Koh06c] Michael Kohlhase. OMDoc – *An open markup format for mathematical documents [Version 1.2]*. Number 4180 in LNAI. Springer Verlag, 2006.

[Koh07] Michael Kohlhase. sTeX: Using TeX/LaTeX as a semantic markup format. Manuscript, submitted to "Mathematics in Computer Science", Special Issue on "Management of Mathematical Knowledge", 2007.

[Kon01] Tony Kontzer. Management legend: Trust never goes out of style. Online at http://www.callcentermagazine.com/article/IWK20010604S0011, June 2001. Seen on 2007-10-26.

[Kor05] Klaus Kornwachs. Knowledge + skills + "x". In Klaus-Dieter Althoff, Andreas Dengel, Ralph Bergmann, Markus Nick, and Thomas Roth-Berghofer, editors, *Professional Knowledge Management 2005*, number 3782 in LNCS. Springer Verlag, 2005.

[KR08] Andrea Kohlhase and Milena Reichel. Embodied conceptualizations: Social tagging and e-learning. *International Journal of Web-Based Learning and Teaching Technologies*, (1):58–67, January-March 2008. In Press.

[Kri06] Klaus Krippendorf. *The Semantic Turn: A New Foundation for Design*. CRC, Taylor and Francis, 2006.

[KŞ06] Michael Kohlhase and Ioan Şucan. A search engine for mathematical formulae. In Tetsuo Ida, Jacques Calmet, and Dongming Wang, editors, *Proceedings of Artificial Intelligence and Symbolic Computation, AISC'2006*, number 4120 in LNAI, pages 241–253. Springer Verlag, 2006.

[Lam02] Patrick Lambe. The autism of knowledge management. Online at http://greenchameleon.com/thoughtpieces/autism.pdf, 2002. Seen on 2007-10-24.

[Lan07a] Christoph Lange. The OMDoc document ontology. web page at http://kwarc.info/projects/docOnto/omdoc.html, seen November 2007.

[Lan07b] Christoph Lange. SWiM: A semantic wiki for mathematical knowledge management. web page at http://kwarc.info/projects/swim/, seen August 2007.

[Lie06] Konrad P. Liessmann. *Theorie der Unbildung. Die Irrtümer der Wissensgesellschaft*. Zsolnay, 2006.

[LK07] Christoph Lange and Michael Kohlhase. A Semantic Wiki for Mathematical Knowledge Management. In *Emerging Technologies for Semantic Work Environments: Techniques, Methods, and Applications*. Idea Group, 2007.

To appear.

[LMU01] Paul Libbrecht, Erica Melis, and C. Ullrich. Generating Personalized Documents Using a Presentation Planner. In *ED-MEDIA 2001-World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 1124–1125, 2001.

[Log06] Logosphere: a formal digital library. web page at http://www.logosphere.org/, seen November2006a 2006.

[LS99] Ora Lassila and Ralph R. Swick. Resource description framework (RDF) model and syntax specification. W3C recommendation, World Wide Web Consortium (W3C), 1999. http://www.w3.org/TR/1999/REC-rdf-syntax.

[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+03] E. Melis, J. Buedenbender E. Andres, A. Frischauf, G. Goguadse, P. Libbrecht, M. Pollet, and C. Ullrich. Knowledge representation and management in ACTIVEMATH. *International Journal on Artificial Intelligence and Mathematics, Special Issue on Management of Mathematical Knowledge*, 38(1-3):47–64, 2003.

[Man01] Lev Manovich. *The Language of New Media*. The MIT Press, 2001.

[MBG+03] Erica Melis, Jochen Büdenbender, George Goguadze, Paul Libbrecht, and Carsten Ullrich. Knowledge representation and management in activemath. *Annals of Mathematics and Artificial Intelligence*, 38:47–64, 2003. see http://www.activemath.org.

[MD05] Käte Meyer-Drawe. Anfänge des Lernens. *Zeitschrift für Pädagogik*, 49. Beiheft, 2005.

[MK07] Christine Müller and Michael Kohlhase. panta rhei. In Hinneburg [Hin07], pages 318–323.

[MM95] Isabel Briggs Myers and Peter B. Myers. *Gifts Differing: Understanding Personlity Type*. Davies-Black Publishing, 1995. First edition in 1980.

[MR+07] Peter Murray-Rust et al. Chemical markup language (CML). http://cml.sourceforge.net/, seen January 2007.

[Mül06a] Normen Müller. OMDoc as a Data Format for VeriFun. In OMDOC – *An open markup format for mathematical documents [Version 1.2]* [Koh06c], chapter 26.20, pages 329–332.

[Mül06b] Normen Müller. An Ontology-Driven Management of Change. In *Wissens- und Erfahrungsmanagement LWA (Lernen, Wissensentdeckung und Adaptivität) conference proceedings*, 2006.

[Mül07] Christine Müller. Towards the identification and support of scientific communities of practice. In Christine Müller, editor, *JEM Workshop 2007*, 2007.

[MV92] Humberto R. Maturana and Francisco J. Varela. *Tree of Knowledge: Biological Roots of Human Understanding*. Shambhala Publications Inc.,U.S., 1992. Originally published in 1984.

[MvH04] Deborah L. McGuinness and Frank van Harmelen. OWL web ontology language overview. W3C recommendation, W3C, February 2004. Available at http://www.w3.org/TR/2004/REC-owl-features-20040210/.

[MW07a] John McCarthy and Peter Wright. *Technology as Experience*. The MIT Press, 2007. Originally published in 2004.

[MW07b] Normen Müller and Marc Wagner. Towards Improving Interactive Mathematical Authoring by Ontology-driven Management of Change. In Hinneburg [Hin07], pages 289–295.

[MW08] Merriam-Webster. Semantics — merriam-webster, 2008. [Online; accessed 7. January 2008].

[OLP07] Mit media lab & $100 laptop. web page at http://laptop.media.mit.edu, seen December2007.

[Pap96] Seymour Papert. An Exploration in the Space of Mathematics Educations. *International Journal of Computers for Mathematical Learning*, 1(1):95–123, 1996.

[PH91] S. Papert and I. Harel. Situating constructionism. In S. Papert and I. Harel, editors, *Constructionism*. Ablex Publishing, 1991.

[Pia96] Jean Piaget. *Einfhrung in die genetische Erkenntnistheorie*. suhrkamp, 1996. First edition in 1974.

[PRR97] G. Probst, St. Raub, and Kai Romhardt. *Wissen managen*. Gabler Verlag, 4 (2003) edition, 1997.

[Rei05] Gabi Reinmann. *Blended Learning in der Lehrerbildung*. Pabst, 2005.

[RK08] Florian Rabe and Michael Kohlhase. A web-scalable module system for mathematical theories. Manuscript, to be submitted to the Journal of Symbolic Computation, 2008.

[Sch97] Heidi Schelhowe. *Das Medium aus der Maschine: zur Metamorphose des Computers*. Campus Verlag, 1997.

[Sch07] Heidi Schelhowe. *Technologie, Imagination und Lernen: Grundlagen für Bildungsprozesse mit Digitalen Medien*. Waxmann, 2007.

[Ses04] Werner Sesink. *In-formation: Die Einbildung des Computers*. Number 3 in Bildung und Technik. LIT Verlag Münster, 2004.

[SKMH04] Wolfgang Scholl, Christine König, Bertolt Meyer, and Peter Heisig. The future of knowledge management: An international delphi study. *Journal of Knowledge*

Management, 8(2):19–35, 2004. Online at http://www.competence-site.de/wissensmanagement.nsf.

[Ste94] Nico Stehr. *Arbeit, Eigentum und Wissen: Zur Theorie von Wissensgesellschaften*. Frankfurt am Main: Suhrkamp, 1994.

[Tea06] Connexions Team. Connexions: Sharing knowledge and building communities. White paper at http://cnx.org/aboutus/publications/ConnexionsWhitePaper.pdf, 2006.

[TM97] Henry Thompson and David McKelvie. Hyperlink semantics for standoff markup of read-only documents. In *SGML Europe 97*, 1997.

[TS02] Sigmar-Olaf Tergan and Peter Schenkel. *Was macht Lernen erfolgreich?*, chapter 4. Köln: Fachverlag Dt. Wirtschaftsdienst, 2002.

[VD04] Katrien Verbert and Erik Duval. Towards a Global Component Architecture for Learning Objects: A Comparative Analysis of Learning Object Content Models. In *Proceedings of the EDMEDIA 2004 World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 202–208, 2004.

[Wal04] Thomas Vander Wal. Folksonomy? Information Architecture Institute Members Mailing List, July 2004.

[Wal06] Thomas Vander Wal. Folksonomies for ia. http://s3.amazonaws.com/2006presentations/OZIA/Folksonomy_for_IA.pdf, September 2006. seen on 2006-10-15.

[WBdYH06] Terry Winograd, John Bennett, Laura de Young, and Bradley Hartfield, editors. *Bringing Design to Software*. Addison-Wesley, 1996 (2006).

[Wen99] Etienne Wenger. *Communities of Practice: Learning, Meaning, and Identiy*. Cambrigdge University Press, 1999.

[Wik08] Wikipedia. Semantics — wikipedia, the free encyclopedia, 2008. [Online; accessed 3. January 2008].

[Wil02] T.D. Wilson. The nonsense of 'knowledge management'. *Information Research*, 8(1), October 2002. available at http://informationr.net/ir/8-1/paper144.html.

[Win90] Jeanette M. Wing. A specifier's introduction to formal methods. *IEEE Software*, 23(9):8–24, September 1990.