Knowledge-rich indexing of learning objects
Kabel, S.C.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Summary

Learning object technology is a recently developed field of study that concerns re-use of existing material for instructional aims. In the world of training and instruction, re-use of existing material is considered as the solution to prevent high development costs, and to make lesson material available to the general public. It so happens that developing lesson material is necessary again and again in a changing context, but often confiscates much time. For this reason it is too expensive to develop lesson material (of a certain quality) entirely from scratch each time, certainly because it becomes obsolete after some time. Developing and using lesson material, however, come within reach for everyone if material can be built from re-usuable components, also called learning objects. There is an enormous quantity of material available on the Internet which can be used for instructional aims, for example images, video material, presentations, etc. Unlike the usual sequential structure of textbooks, the Internet pre-eminently facilitates an “examining” way to learn, in which it is not necessary to stick to a fixed order of teaching material. Also within companies and institutions often much material exists that can serve as a basis for instructional aims or information needs, and for many re-use is a promising prospect.

A number of processes underlie re-use, as outlined in Chapter 1. In short these are: analysis, storage, retrieval and finally re-use of material. Because existing material not always has the desired re-usable format, it can be necessary to analyze this material first and segment it in fragments. The grain size of the resulting fragments is important thereby. Small of pieces material, for example a single image, are easier to apply again in several contexts. Larger pieces material (for example a web site) are, because they have more context, less easily fit in elsewhere. The resulting fragments must therefore be as small as possible, but still cover a coherent subject. Nevertheless larger pieces material must of course also be accessible for re-use. Then the fragments must be indexed and stored. Automating aforesaid processes can be very lucrative, certainly when much material is available. Publicly available learning objects are generally stored in learning object repositories, so that they can be distinguished from other material on the Internet. The indexing ensures that material can be retrieved. Retrieved material is finally re-used, for example in lesson material, or to gain knowledge in an other way.

Re-use of material requires however in the first place a rich indexing of that material, to be able find it and to support the aim of re-use. Because there is a large quantity of diverse material available, and because people have divergent information needs, a rich indexing is of vital importance to make the processes underlying re-use run smoothly. A rich index illuminates the material from several perspectives, where task-supporting aspects play an increasingly important role. Recently a standard is available, IEEE LOM (2002), to index learning objects. This standard is widely used, often slightly adapted, and makes re-use of (lesson) material by a wide public possible.

Although within limited time much has been reached, the standard on some points has not been given full details and has not been structured enough, and in each process still a number of problems occur (cited in Chapter 1). The central hypothesis in this thesis is that knowledge-rich indexing is necessary to solve these problems, and optimize re-use. A
knowledge-rich index must be based on formally represented knowledge in an ontology. Instead of an open indexing-vocabulary, or one that consists of a flat list words, an ontology has a number of characteristics by which it is considered as a knowledge-rich indexing approach. Essentially an ontology consists of a number of unique concepts associated with synonyms and abbreviations, which is organized in a hierarchical structure. Properties of concepts, and relations, can be expressed. For example, a concept “car” (synonym “automobile”), has property “color”, and has a relation with another concept “owner”. Ontologies pre-eminently allow to represent several perspectives on the material to be indexed in an organized way. The central question is thus: how can the processes underlying re-use be improved by means of a knowledge-rich index based on ontologies? This question is answered by means of theoretical research, a case study, and several empirical investigations.

In a study of the literature (Chapter 2) a number of methods and techniques which are used in the processes underlying re-use are reviewed. First several types of material and several grain sizes are defined. Then several ways to index are discussed; it is explained what an ontology-based indexing framework can add to an index, and relevant standard indexing frameworks are compared with each other. Also problems which arise in each separate process are commented, and it is explained how ontologies can contribute to possible solutions.

An example of knowledge-rich indexing vocabulary is given in Chapter 3. A case study outlines in detail how ontologies are applied in each process, and how ontologies support the transition between processes. A large European research project is described that can be seen as a characteristic example in which making existing material re-usable is necessary. Three technical domains in which maintenance training for complex systems is given have been examined: repairing cars, weapon systems, and programming and configuring systems to control traffic. Within these institutions or companies the main sources to develop lesson material are the often extensive technical manuals delivered with the system. As such, these manuals are not suitable for instruction. Developing and regularly updating lesson material costs much time and effort. To support the processes (analysis, storage, retrieval and re-using material) a number of software applications which are based on ontologies has been developed. With these applications large quantities of source material can be automatically analyzed and segmented in smaller fragments. These fragments are, also automatically, indexed using ontologies, and are stored in a database. Developing lesson material is supported by a software application with which a blueprint of a lesson can be made, which subsequently can be filled in with relevant material from the database. Retrieving material is extensively supported by a rich index. This whole shows how, by application of ontologies, the transition from source material to lesson material supported.

After it has been shown that the use of ontologies is feasible, a number of experiments is described in which conceptual and economic variables concerning the use of ontologies have been investigated. The impact of the use of a knowledge-rich indexing vocabulary on manually indexing material, and on retrieving material and composing lesson material have been investigated. To be able to calibrate a knowledge-rich approach, a point of reference had to be created. A knowledge-rich, ontology-based method has been compared to a method based on an unstructured set of keywords, which served as a benchmark. Also relatively inexperienced persons took part in the experiments, to be able to depart from the worst conceivable situation. It has been assessed that this imitates a realistic situation,
because standard indexing-frameworks for (lesson) material have only just reached maturity. At the same time it can be expected that this experience increases rapidly. As in practice people gain experience, better results on the investigated variables can be expected. A number of issues has been investigated. 

Because indexing is mostly still done manually by people, it can be expected that differences between indexers occur. This “noise” in indexes results in inaccurate and incomplete search results, and must therefore as much as possible be prevented. It has been investigated if the use of ontologies contributes to more consistency between different indexers (Chapter 4). This research shows that the structure in ontologies contributes to a higher consistency between the values which different indexers ascribe to the same material and with use of the same indexing vocabulary. In addition, the medium type plays a role. Images and video material generally was less consistently indexed than textual material and audio material, possibly because the first require more interpretation. The type of property that must be described is also of influence on consistency. Concrete properties (such as “type of representation”) showed few differences between values ascribed by indexers. The more abstract the property (for example “type of knowledge”, or “semantic density”), the more inconsistencies occur. Generally approximately 25% noise in indexes can be counted on, assuming relatively inexperienced indexers and the use of a structured indexing vocabulary.

In a next investigation the added value of a rich index for retrieving material from a database, and for creating lesson material has been measured (Chapter 5). With added value is meant: the efficiency and effectiveness of retrieval of material, the relevance of material that is used in lesson material, and the quality of the lesson material. Quality of the lesson material is then determined in terms of soundness (such as completeness and correctness) and readability (such as coherence and redundancy). From this research becomes clear that also for retrieval and re-use of material the use of ontologies has an added value. Material could be retrieved more efficiently and more effectively, and more relevant material was used in the lesson material. The use of ontologies also contributes to a better quality of the lesson material. The product was generally more complete and more correct, but in this research the use of ontologies did not directly contribute to a better readability of the product.

In two final experiments the economic feasibility of a knowledge-rich approach has been measured (Chapter 6). In this research, the indicator of costs was the time which is necessary to carry out (part of a) task. It has been investigated how much time is necessary for indexing material, and how much time is necessary for retrieving material from a database, and for composing lesson material. This has been done to find out if the costs of a knowledge-rich indexing approach can be regained, and if so, when? The results indicate that the costs are especially high in the beginning, but costs are saved during composition of lesson material. More specifically, indexing as well as specifying queries for a search engine cost much time. However, because of these time investments, determining the relevance of a collection of retrieved fragments went rapidly, and also creating lesson material was done in a short time.

A method based on only a collection of keywords was on the whole at least as fast, but the separate steps showed the opposed pattern. Indexing cost little time, specifying a query too. However, because of inaccurate search results determining the relevance from a set of retrieved fragments cost much time, and so did composing lesson material from retrieved fragments. Characteristic of a knowledge-rich method is that the point of interest moves
from the contents of the material, to the metadata. In other words, instead of always having to examine the contents of material to determine the relevance of it, relevant material is collected on the basis of indexes. A knowledge-rich method can be considered as an investment. The decision to follow that approach depends on several factors. In general it can be assumed that when precise search results are important, an investment in a knowledge-rich index eventually pays itself back.

In Chapter 7 conclusions are formulated. The investigations show that the use of ontologies in many respects has a positive impact on the processes which underlie reuse. Although a method based on a set of keywords served as a benchmark in the investigations, both methods complement each other in reality. The investigations have been conducted in controlled setting. More study into the daily practice is necessary to further improve and simplify the handling of material and the use of indexing vocabularies. For example a study into the use of (standard) indexing vocabularies for indexing learning objects, and for retrieval of learning objects from current learning object databases, or a study into the application of a theory to design learning objects.