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.
6 The costs and benefits of a knowledge-rich index

6.1 Introduction

A knowledge-rich indexing approach has several advantages with respect to other methods, but requires effort and time. The question addressed in this chapter is whether this time is worth the investment, and, if so, when the investment is regained. Indexing learning objects with metadata is an obstacle for many companies and institutions. Nevertheless the potential of richly indexed learning objects, together with technological developments that make this possible, drive many to index their material. Usually the investment in indexing is made based on the expectation to save time and costs later on. Indexes are provided because it is anticipated that the material can be found quickly when searched for, and time is saved by re-using existing material. However, an insight into the costs and benefits of these efforts is often lacking because learning object technology is relatively new. In the dynamic environment of learning object technology, cost estimations are popular and pricing models for online content are being developed (see for example Darimont, 2002).

There are various ways to calculate cost, and there is no consensus on one ideal way. The return on investment analysis, commonly called ROI (Kruse, 2003), is one of the most popular ways to measure the financial returns on e-learning. It allows calculating the return on a training investment, for example an on-line sales training program. Total benefits are estimated by looking at reduced travel expense, reduced time away from job, increases in measurable business metrics (for example, sales, quality, service, inventory turns, etc.). Costs for e-learning include course development expense, student salaries for time in training, instructor salaries (if synchronous training), and any hosting or help desk fees. However, to accurately estimate these variables is difficult. The availability of indexed content may be assumed in an estimation. "Course development expense" may include for example the design and development of a 5-minute chunk of learning material, but it may exclude (or not explicitly represent) the cost of indexing the content that was used for developing the learning object. By necessity, since learning objects cannot yet be automatically indexed with semantically rich instructional properties, indexing is often done manually. In order to provide insight into these costs, the cost-benefit relation of indexing and re-using content is investigated in more detail. The goal is to provide insight in the costs and benefits in terms of time of the steps in the process from manual indexing of small learning objects (document fragments) to their re-use in instructional material, illustrated in Figure 6-1 (derived from the conceptual model of material handling processes in Chapter 1).

![Figure 6-1 Steps in the process from indexing to re-use.](image-url)
The process starts with indexing fragments, and continues with re-use\(^{10}\): searching and selecting fragments to compose instructional material from.

In this study two methods of indexing and retrieval are compared: a keyword and an ontology-based method. In the keyword setting, fragments are indexed with keywords using a free vocabulary. In the ontology setting indexing is based on ontologies (structured, fixed vocabularies). A set of ontologies represents different viewpoints on the fragments. These viewpoints are reflected in fragment attributes that concern the Topic, Representational Type, Description Type, Knowledge Type and Instructional Role.

Retrieval in the keyword setting is based on elaborate but flat lists of domain related keywords, except for text that is retrieved based on full text. In the ontology setting, retrieval is based on the set of ontologies (using attributes Topic, Representational Type, etc.).

Time is the main indicator of costs. The time to index fragments with ontologies is seen as a (one-time) investment to enhance the re-use process and save time. This is based on two theoretical assumptions. First, because a fragment is annotated on several attributes, it is assumed that indexing with ontologies takes longer than with keywords. Secondly it is assumed that re-use based on ontologies takes less time than keyword-based re-use, because a search from different perspectives yields precise results, and the task-related indexes support lesson material composition (see also Chapter 5). It is hypothesized that the larger the amount of time invested in indexing, the less time is necessary for re-use (retrieving fragments and composing lesson material). If it is correct that re-use based on ontologies saves time, the longer indexing time will eventually be compensated. The hypothesis is illustrated in Figure 6-2, showing the break-even point at which extensive indexing could start to pay off.

\[ \text{Figure 6-2 Break-even point of cost-effective indexing and re-use.} \]

In order to test the hypothesis, two questions are formulated:

- How much time does it take to index a fragment manually, with keywords and with ontologies?
- How much time does it take to re-use a fragment based on keywords and on ontologies?

If the hypothesis that the larger the amount of time invested in indexing, the less time is necessary for re-use holds, the following question must be answered:

\(^{10}\) In this chapter, the term “re-use” is used to refer to the retrieval and re-use processes in the conceptual model in Chapter 1.
How often does a fragment have to be re-used to compensate for extensive indexing with ontologies?

Additionally, the following question is formulated:

- Where can time be saved in the indexing and re-use process?

These questions are investigated in two experiments.

6.2 Scope of the research

The experiments are conducted in a controlled setting. The experiments are set up as realistic as possible, however several choices were made that can influence the scope of this research. The first is a choice of subjects, who should represent the "average indexer, retriever and instructional designer". Similar to previous experiments, a worst possible situation is simulated by selecting relatively inexperienced subjects, to prevent that time is underestimated. The subjects selected for these experiments used Internet on a regular basis, which indicates they were used to using keyword-based search engines like Google. In the keyword situation simulated in this experiment, subjects had to use keywords to annotate and retrieve material. Given that subjects were familiar with a keyword-based approach, they are expected to index and retrieve material reasonably quickly.

In the ontology situation simulated in these experiments, subjects had to use a knowledge-rich annotation structure to annotate and retrieve material. It is expected that, although people increasingly use (instructional) annotation structures, there is generally not much experience with a knowledge-rich approach. Because an underestimation of time necessary to use a knowledge-rich annotation structure could have negative consequences when an investment is made, the experiments are designed such that a relative pessimistic estimation of time is made, which in real life can only turn out to be shorter when people gain experience. In both the keyword and ontology situation simulated in these experiments, lesson material had to be created from the earlier selected set of fragments. Although they were familiar with for example creating Microsoft Powerpoint presentations, the subjects participating in the experiments were relatively inexperienced compared with professional instructional designers. These subjects are therefore expected to need relatively much time to create lesson material, when compared to for example a teacher who does this on a daily basis.

Besides the choice of subjects, a second factor can influence the scope of this research. A number of decisions in the design of the tasks were made to make the tasks workable for the inexperienced subjects. Similar to experiments described in Chapter 4, indexing was done with a limited number of terms. Keyword indexing was done quite freely. To reduce the complexity of the task of indexing with ontologies on multiple attributes this was done per attribute subsequently for all fragments that had to be annotated. The annotations were not corrected for quality. Furthermore composition of lesson material was done with a limited number of fragments, and pedagogic and editing guidelines were provided.

Sections 6.3 and 6.4 describe the experiments measuring indexing and re-use time. In Section 6.5 the results are interpreted in terms of the cost-benefit relation of indexing and re-use, and bottlenecks in the indexing and re-use processes that take much time are identified.
6.3 Measuring indexing time

The costs of indexing with keywords and with ontologies were measured in terms of time. Additionally the major sources of costs are located by measuring indexing time per attribute (Representational Type, Topic, Description Type, Knowledge Type, Instructional Role) and per medium type (picture, sound, text, video).

Like in previously described experiments, the domain of gorillas was used because it is a publicly known and accessible domain where differences in prior knowledge of subjects were shown to be minimal.

6.3.1 Method

Subjects had the task to annotate a set of 16 multi-media fragments about gorillas and the time to annotate a fragment was measured. A fragment was annotated on 6 attributes: one Keyword attribute to measure keyword indexing time, and 5 attributes to measure ontology indexing time. The 5 attributes represent different viewpoints on a fragment (Representational Type, Topic, Description Type, Knowledge Type and Instructional Role). Each of these 5 attributes corresponds with an ontology (see the annotation structure for fragments about gorillas described in Chapter 4). The attribute values (the terms that were used for annotation), are either keywords typed in by subjects or concepts selected from ontologies. A maximum of 5 terms per attribute was allowed in order to provide a guideline for subjects.

The 16 fragments, 4 of each medium type, were annotated with Keywords first (to prevent an influence on keywords coming from the ontologies), and subsequently on Representational Type, Topic, Description Type, Knowledge Type, and Instructional Role, in this order. To prevent a bias in the order in which fragments were annotated the fragments were presented in random order. The time a single person needs to annotate a fragment with both kind of attributes, keywords and those based on ontologies was measured, because it was anticipated that this represents a realistic situation. This design implicates that the subjects see the 16 fragments 6 times in succession. As will be elaborated below, to reduce the possible effect of the time necessary to view a fragment (for the first time), this “view”-time is not included in the indexing time that is measured.

The prior knowledge of subjects that could influence the indexing task is instructional experience, knowledge of the topic and familiarity with search engines. These variables were assessed using a questionnaire.

The following subsections describe the variables, subjects, materials and procedure.

6.3.1.1 Variables

The hypothesis that the more time is spent on indexing, the less time is spent on re-use is based on the assumption that indexing with ontologies takes longer than with keywords because a fragment is annotated on 5 attributes instead of 1. The variables serve to test if indexing with ontologies takes longer than with keywords, and locate where costs are made.

The variables are:

- Average indexing time per fragment attribute [Keyword, Representational Type, Topic, Description Type, Knowledge Type, Instructional Role]
- Average indexing time per fragment of medium type [picture, sound, text, video]
Time is measured in seconds, logged in the annotation tool that was used (described in Section 6.3.1.3). Fragments were shown to subjects 6 times. The possible effect of the fixed order of attributes could be that fragments are displayed long when indexing the first attributes, and shorter when indexing the last attributes. To prevent a bias in indexing time, annotations were added while a fragment was being displayed, and the time that annotations were made was logged. Keyword indexing time was measured from activating the text field where keywords were typed until an annotation was completed. Ontology indexing time was measured from expanding an ontology (by clicking on a top concept) until an annotation was completed. All times are averaged across subjects and fragments. Indexing time per attribute is the average time to annotate one fragment on one attribute with 0 to 5 terms. Keyword indexing time is the average time to annotate one fragment with keywords. Ontology indexing time is the average time to annotate one fragment with ontologies summed over 5 attributes. Indexing time per medium type is the average time to annotate a fragment of one medium type on one attribute with 0 to 5 terms. The following operational definitions are used:

- Number of subjects, 1...n
- Number of fragments, 1...16
- Number of fragments of medium type m, 1...4

Average indexing time per attribute = total time spent by all subjects to annotate 16 fragments on attribute a / number of subjects * number of fragments.

Average indexing time per medium type = total time spent by all subjects to annotate 4 fragments of medium type m / number of subjects * number of fragments of medium type m. Because 0 to 5 terms are used per attribute, indexing time is normalized. The number of terms used per attribute is counted, and the time to add one term to an attribute is calculated.

6.3.1.2 Subjects

37 3rd year Social Science Informatics students from the University of Amsterdam (different from those who attended earlier experiments) were asked to participate.

6.3.1.3 Materials

The computers that were used had the same hardware and software configuration. Given the current state of affairs in hardware and technology, there is no reason to assume that the speed of computers influenced time.

A Prolog database was used, filled with 830 fragments (275 pictures, 69 sounds, 93 videos and 393 text fragments). Text and picture fragments were gathered from different websites. Video fragments were created from a number of videotapes. Sound fragments were recorded human voices (generally a few sentences) or gorilla sounds retrieved from the Internet. For the indexing experiment, a subset of 200 fragments was used. To prevent a bias caused by the size of a fragment, of each medium type 50 fragments of approximately the same size were selected. Most text fragments were small, a few lines of text, a paragraph or at most half a page. The average playtime of a video fragment was 25 seconds, and the average playtime of a sound fragment was 10 seconds.

An annotation tool was developed to annotate the fragments (similar to the tool in Figure 4-2).
and record the time of each annotation. In addition the tool records which fragments were presented to a subject; the order in which they were presented; the terms that were used; and the number of terms per attribute (0-5). The tool was created in SWI-Prolog and XPCE and Windows Media Player was used to display sound and video fragments. Before the experiment was carried out a pilot was conducted in which 7 subjects conducted the indexing task. This led to a technical adaptation. Instead of storing the fragments on a server they were stored on a local cd-rom to prevent delay in video playtime. Two screenshots of the annotation tool are shown in Figure 6-3 and Figure 6-4. The first screenshot shows a picture fragment that is annotated with keywords, the second shows a text fragment that is annotated with Topic concepts.

![Figure 6-3 Annotation tool, an annotation of a picture on the keyword attribute.](image)

![Figure 6-4 Annotation tool, annotation of a text on the topic attribute.](image)
The buttons “Previous fragment” and “Next fragment” are used for browsing through the fragments and assigning values to for example the Topic attribute of each of the 16 fragments. By clicking “Next fragment”, the annotation field is cleared and the next fragment is displayed. Previously made annotations can be changed using the “Previous fragment” button. Terms are transferred using the “<-” and “->” buttons. The attribute values are saved with the “Apply” button. Subsequently the next attribute (and next fragment) is presented.

6.3.1.4 Procedure

The test was carried out in a non-fixed time of approximately one hour on average. First the task and the tool were explained, then the subjects conducted the indexing task and filled in the questionnaire. Subjects were explained that they should be as specific as possible but that they did not necessarily have to use leaf-concepts only, and that a keyword could be composed of more than one term, as in “scary gorilla”. By means of support the subjects could look into a reference work with definitions and examples of concepts. The questionnaire contained a set of multiple-choice questions about the subject’s prior knowledge and a set of evaluative questions about the ontologies.

6.3.2 Results

6.3.2.1 Prior knowledge

Before presenting the results it is mentioned that subjects’ prior knowledge was found sufficient to conduct the indexing task. Of the subjects 66% found they had enough everyday knowledge of gorillas and approximately the same percentage subjects reported they had sufficient instructional experience in general (90% had experience with giving Microsoft Powerpoint presentations and 38% had teaching experience). All subjects used search engines on the Internet.

6.3.2.2 Data preparation

The raw data in XML files were automatically transformed into Excel files. From a dataset of 3552 data points (37 subjects * 16 fragments * 6 annotations) a small number of outliers were removed. To correct for a learning effect, the data of the first two fragments (all 6 attributes) were removed from the dataset. A learning effect was caused by familiarization with the tool and the ontologies, as can be seen in Figure 6-5. It shows the average indexing time for each attribute from the first to the 16th fragment.
Indexing the first two fragments took long, especially for the keyword and topic attribute. Per term, the keyword attribute took least time, and Topic most. Figure 6-5 also shows the long time it took to annotate the 16th fragment on the last attribute (Instructional Role). These data were removed too, because many subjects who finished the indexing task went on with the questionnaire without pressing the “Apply” button, resulting in relatively long indexing times. Furthermore indexing times of 0 seconds in which no annotation was made were removed. These are considered outliers as they represent a deliberate attitude of rushing though the task and cannot be considered a serious attempt to index a fragment. For the same reason the data of two subjects were removed. Because some subjects were interrupted, indexing times per term longer than 90 seconds (for Topic longer than 120 seconds) are considered outliers and were removed. The removal of outliers resulted in a dataset of 2875 data points, which was used for analysis.

6.3.2.3 Analysis of results
Table 6-1 shows the average indexing time in seconds per fragment per attribute. The second column from the left shows the sum of the average ontology attribute times.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Keyword</th>
<th>Sum of ontology attributes</th>
<th>Representational Type</th>
<th>Topic</th>
<th>Description Type</th>
<th>Knowledge Type</th>
<th>Instructional Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture</td>
<td>32</td>
<td>91</td>
<td>15</td>
<td>38</td>
<td>13</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Sound</td>
<td>41</td>
<td>89</td>
<td>8</td>
<td>43</td>
<td>12</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Text</td>
<td>42</td>
<td>97</td>
<td>10</td>
<td>44</td>
<td>15</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Video</td>
<td>46</td>
<td>108</td>
<td>14</td>
<td>49</td>
<td>14</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Average</td>
<td>40</td>
<td>96</td>
<td>12</td>
<td>44</td>
<td>14</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 6-1 Average indexing time (sec.) per fragment attribute.
On average it took 40 seconds to annotate a fragment with keywords and 96 seconds to annotate a fragment with ontologies. Not surprisingly, the assumption that indexing with ontologies takes longer than indexing with keywords is correct. Pictures required least time and videos most time, but with a T-test no significant differences between medium types are found. Much time was spent on domain indexes (Keywords and Topics) while Representational Type took least time. However, different numbers of terms were used per attribute. Long indexing times may be due to a large number of terms used per attribute.

Table 6-2 shows the average number of terms used per attribute. The second column from the left shows the sum of the average numbers of concepts used in the ontology attributes.

<table>
<thead>
<tr>
<th></th>
<th>Keyword</th>
<th>Sum of ontology attributes</th>
<th>Representational Type</th>
<th>Topic</th>
<th>Description Type</th>
<th>Knowledge Type</th>
<th>Instructional Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture</td>
<td>2.9</td>
<td>6.9</td>
<td>1.5</td>
<td>1.9</td>
<td>1.3</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Sound</td>
<td>2.9</td>
<td>7.1</td>
<td>1.2</td>
<td>2.2</td>
<td>1.4</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Text</td>
<td>3.1</td>
<td>7.2</td>
<td>1.2</td>
<td>2.4</td>
<td>1.3</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Video</td>
<td>3.3</td>
<td>8.4</td>
<td>1.6</td>
<td>2.8</td>
<td>1.5</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Average</td>
<td>3.0</td>
<td>7.4</td>
<td>1.3</td>
<td>2.3</td>
<td>1.4</td>
<td>1.1</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Table 6-2 Average number of terms per fragment attribute.

On average 3 keywords and 7.4 concepts were used to annotate a fragment. Most terms were used for videos. Many Topic concepts were used, and not many Knowledge Type concepts. The ontologies differ in size. The domain ontology (with Topic concepts) is by far the most extensive. In order to see how much time was used per term, indexing time was normalized by dividing the indexing time per attribute by the number of terms used per attribute.

Table 6-3 shows the average indexing time per fragment attribute per term. The second column from the left shows the average of the ontology attributes.
On average it took 13.1 seconds to annotate a fragment with one keyword, and 13 seconds to annotate a fragment with one concept. The different kind of indexing tasks (thinking up a keyword or selecting a concept) require about as much time. However determining a Topic took relatively long, indicating that indexing with an extensive concept hierarchy requires much time. Beside the size of an ontology, the nature of the attribute is of influence on indexing time: Representational Type and Description Type required not much time while Knowledge Type, which is an abstract aspect, required longer.

### 6.3.3 Conclusions about indexing time

Extensive indexing with ontologies took more than twice as long as keyword indexing. More than twice as many terms were used. Per term, there is not much difference: to specify one keyword took about as long as to select one concept from an ontology, except for Topic concepts. To specify one Topic concept took approximately 50% longer than keywords, likely because the domain ontology contains most concepts and requires navigation through the hierarchy. Time may be gained by reducing the mental effort that is required to browse a large concept hierarchy, for example by providing indexers with a search function for concepts in an ontology.

### 6.4 Measuring re-use time

A second experiment was designed to measure the time it takes to re-use a fragment, based on keywords and ontologies. Additionally, to locate more specifically where time was gained or lost, the time subjects spent on retrieving fragments (searching and selecting fragments) and composing lesson material was measured.

#### 6.4.1 Method

Subjects had the task to create an educational Web page from a database with over 800 multimedia fragments about gorillas (described in Section 6.3.1.3). The fragments in the database...
were indexed with keywords and ontologies by one or sometimes two people knowledgeable about the domain. Search and retrieval was performed in two conditions: a keyword and an ontology condition. In order to provide subjects with some guidance in topics to address, they had to choose one of three gorilla-related themes. Pedagogical and editing guidelines were provided to help the relatively inexperienced subjects decide when the lesson material was finished. The number of fragments to be used was set to 15 to be able to compare the results. The following subsections describe the design of the experiment in more detail: the experimental conditions, variables, subjects, materials and procedure.

6.4.1.1 Independent variables and experimental conditions
Subjects’ prior knowledge and the theme subjects choose can vary. Subjects were assigned to a keyword and an ontology condition using a matching procedure based on two factors that could be of influence on the re-use task: familiarity with search engines and experience with composing pedagogical or presentation material such as a Microsoft Powerpoint presentation. In each condition subjects had to select a set of fragments from a database using a retrieval tool. In the keyword condition, search and retrieval was based on keywords only. Subjects had to think up one or more terms, type them in, submit the query and browse the retrieved fragments. In the ontology condition subjects searched based on the same set of ontologies as used in the indexing time experiment. Subjects could select a concept for each of the attributes (Representational Type, Topic, Description Type, Knowledge Type and Instructional Role), submit the query and browse the retrieved fragments. In both conditions subjects then had to create lesson material from the selected set of fragments using the same tools and guidelines.

6.4.1.2 Dependent variables
The hypothesis that the more time is spent on indexing, the less time is spent on re-use is based on the theoretic assumption that re-use based on ontologies takes less time than keyword-based re-use, because a search from different perspectives yields more precise results, and the task-related indexes support lesson material composition. The variables serve to test if re-use based on ontologies takes shorter than keyword-based re-use, and to locate where time is gained or lost. The dependent variables are:

- Average re-use time
  - Average retrieval time
  - Average search time
  - Average selection time
  - Average composition time

Time was measured in seconds, by logging user actions (for example button clicks) in the retrieval tool that was used (described in Section 6.3.1.3), and by logging active windows of the three tools that were used (either the retrieval tool, Microsoft Word or Windows Media Player).

All times are averaged across subjects. Re-use time is the time to re-use 15 fragments, from the moment a subject is logged into the retrieval tool, until the lesson material is finished (the document is saved and Microsoft Word closed). Re-use time is the sum of retrieval time and
lesson material composition time. Retrieval time is the time spent in the retrieval tool. Retrieval time is the sum of the time spent on searching and selecting fragments. Search time is the time spent on generating queries, from thinking up keywords or browsing concepts in an ontology, until a query is submitted. Selection time is the time spent on selecting fragments, from browsing a list of retrieved fragments, until a basket with 15 selected fragments is exported to Microsoft Word. Composition time is the time spent on editing lesson material, represented by the time spent in Microsoft Word. In order to relate indexing time to re-use time, re-use time per fragment, the average time to re-use one fragment, is calculated.

In addition, to obtain a detailed view of the time spent on searching, the number of queries, terms and retrieved fragments is counted. To analyze the time spent on selecting fragments in more detail, the number of displayed fragments during retrieval time is counted, and the time spent on displaying fragments during retrieval time is measured. The time that the Windows Media Player window was active is an indicator of the time spent on displaying fragments during the entire task, including composition of lesson material. The following operational definitions are used:

- Number of subjects, \( 1 \ldots n \)
- Number of re-used fragments = 15

Average re-use time = total time spent by all subjects to re-use 15 fragments / number of subjects.
Average retrieval time = total time spent by all subjects to retrieve 15 relevant fragments / number of subjects.
Average search time = total time spent by all subjects to search 15 relevant fragments / number of subjects.
Average selection time = total time spent by all subjects to select 15 relevant fragments / number of subjects.
Average composition time = total time spent by all subjects to compose lesson material from 15 fragments / number of subjects.
Average re-use time per fragment = total time spent by all subjects to re-use 15 fragments / number of subjects \(*\) number of re-used fragments.
Average number of queries = total number of queries submitted by all subjects during search time / number of subjects.
Average number of terms = total number of terms used by all subjects in queries submitted during search time / number of subjects.
Average display time = total time spent by all subjects fragments are displayed during retrieval time / number of subjects.
Average number of displayed fragments = total number of displayed fragments by all subjects during retrieval time / number of subjects.

6.4.1.3 Subjects
30 First year Information Science students participated in the experiment.

6.4.1.4 Materials
Three tools were used: a retrieval tool to search and retrieve fragments, Microsoft Word to
edit lesson material and Windows Media Player to display sound and video fragments. The Prolog database with 830 fragments (275 pictures, 69 sounds, 93 videos and 393 texts) was used (described in Section 6.3.1.3).

A keyword and an ontology retrieval tool were developed (similar to the tools described in Chapter 5) for searching, browsing, viewing fragments and assembling fragments for the lesson material. Both tools contain three browsers. In the browser at the top one can generate and submit a query. In the middle browser one can browse through the list of retrieved fragments and select one, and in the browser at the bottom a selected text or picture fragment can be displayed (sound and video are displayed in Windows Media Player), and fragments can be added to or deleted from a basket at the bottom right. The menu option “Export basket” in the “File” menu is used for exporting the basket with fragments to Microsoft Word. Figure 6-6 shows the keyword retrieval tool.

![Keyword retrieval tool](image)

**Figure 6-6 Keyword retrieval tool.**

In the keyword retrieval tool, subjects search for fragments simply by typing in one or more keywords. In the ontology retrieval tool (see Figure 6-7), subjects search for fragments by generating a query from at most 5 terms (one per attribute).
Fragments were indexed with a flat list of keywords based on the full text or, in the case of non-textual media, based on an elaborate but also flat list of domain-related keywords as IR techniques as yet provide no other solution. Instead of using a fixed list of keywords, the keywords were added quite freely. The matching technique used in the keyword retrieval tool for text fragments is full text search including inflections (Baayen, Piepenbroek & Gulikers, 1995), and conjunctive (AND) queries. Queries with more than one term match first on all terms, than on all minus 1, etc. (coordination level or quorum searching). Since the database contains relatively few fragments, terms are not weighed. Because a general query (for example “gorilla”) results in a very long list of hits, a maximum of 50 fragments is presented to the subject. To prevent a bias caused by for example alphabetic presentation order of search results, these are presented in random order.

The matching technique applied in the ontology retrieval tool is based on a mapping between terms in a query and attribute values of a fragment. A conjunctive query can be composed of a maximum of 5 search terms (one term per attribute). Slots that are not filled are treated as wildcards.

Microsoft Word was used for composing lesson material. The 15 fragments are displayed in a Word document in the order in which they were exported from the retrieval tool. The order of fragments could be changed. Pictures, sounds and videos are displayed as links. The names of the links could be changed and the text fragments could be edited.

6.4.1.5 Procedure
The experiment was carried out in a non-fixed time of approximately 1.30 hour on average.
The task and tools were explained first, then subjects conducted the re-use task. The subjects were explained that the task comprised choosing a theme, collecting 15 fragments relevant for that theme in a basket, exporting the basket to Microsoft Word, and composing lesson material from the selected fragments. To compose lesson material, subjects had to place the fragments in an order suitable for instruction and if necessary edit some of the text. Subjects should not spend time enhancing layout, for example formatting text, but instead concentrate on the content.

The context in which the task had to be carried out was to deliver the content of a Web page for an educational site for children aged 12-16 called Klokhuis, a TV-program for children familiar to the subjects. Subjects had to choose one of the following themes:

- In cooperation with WNF, Klokhuis wants to develop a Web page that teaches children about threats to the gorilla and what can be done to save the gorilla.
- Klokhuis wants to develop a Web page for kids (and their parents) who go on gorilla safari. The page should teach them how to behave near the gorilla.
- Klokhuis wants to develop a Web page in cooperation with Artis (a zoo in Amsterdam) that teaches kids who are interested in becoming a zookeeper how to take care of the gorilla.

The database contained enough fragments that were either directly or indirectly relevant for the themes. The subjects were furthermore told that Klokhuis values the didactical quality of the Web pages, and that for this reason some guidelines had to be followed, shown below.

- Try to get the reader's attention
- Present a learning goal
- Introduce the reader to the topic
- Present the content
- Possibly elaborate on the content
- Provide illustrations and/or examples
- Enable the reader to test his/her knowledge

Before the experiment was carried out a pilot was conducted, which led to a sharper explanation of editing guidelines with examples.

6.4.2 Results

6.4.2.1 Data preparation

The raw data (text files) contain 9546 data points (3,704 in the keyword condition and 5,842 in the ontology condition). These data were automatically transformed into Excel files. All times are averaged across subjects.

Similar to the indexing task, a learning effect was caused by familiarization with the retrieval tools and the ontologies. Figure 6-8 shows the average time of the first, second and remaining search actions, logged in the retrieval tools. (The search actions in the ontology tool are represented by sum of the times spent on browsing each of the ontologies).
The first and second search action times were removed from the dataset (360 data points). No individual outliers were identified. Although times longer than 4 minutes occurred, it was not clear from the log files that these were caused by subjects that were interrupted. There was some difference in the quality of the lesson material, which may to some extent have been the result of the time subjects worked on it. Re-use times varied from 20 minutes to over 90 minutes, and it is assumed that this has had some influence on the quality of the lesson material. Therefore a correction for the quality of the lesson material was carried out. The quality was assessed by two coders. The criteria were 1) topical relevance, 2) completeness and order of fragments, (for example title followed by advance organizer, questions included etc.) 3) variation in medium and way of representation (for instance large text-files are less suitable for a Web page), and 4) overall attractiveness (for example repeatedly using the same subtitle is not appealing). 10% of a subject’s time was added if the lesson material was judged to be of low quality. If the lesson material was judged to be of high quality 10% of a subject’s time was subtracted. The coders disagreed on 3 out of 30 pages, indicating high inter-coder reliability. After discussion consensus was reached, 7 Web pages were judged to be of low quality, 15 medium and 8 high.

6.4.2.2 Analysis of results
Table 6-4 shows the average time in seconds to re-use 15 fragments. Re-use time is the sum of the average retrieval time and the average composition time, and retrieval time is the sum of the average search time and the average selection time.

![Graph showing average search action times](image)

**Figure 6-8 Average of first, second and remaining search action times.**

<table>
<thead>
<tr>
<th></th>
<th>Average re-use time</th>
<th>Average retrieval time</th>
<th>Average search time</th>
<th>Average selection time</th>
<th>Average composition time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword</td>
<td>3343</td>
<td>2394</td>
<td>544</td>
<td>1850</td>
<td>949</td>
</tr>
<tr>
<td>Ontology</td>
<td>3389</td>
<td>2756</td>
<td>1106</td>
<td>1649</td>
<td>633</td>
</tr>
</tbody>
</table>
Table 6-4: Average re-use time (sec.): retrieval (search, selection) and composition.

It took about 55 minutes to re-use 15 fragments. Overall it took only slightly longer to re-use fragments based on ontologies than it took based on keywords. However the assumption that re-use based on ontologies takes less time than keyword-based re-use turns out to be incorrect.

In a keyword setting much time was spent on selecting fragments and on lesson material composition. In an ontology setting time was invested in queries, and less time was spent on selecting fragments and on lesson material composition. These profiles of re-use in the keyword and ontology condition are visualized in Figure 6-9.

In the ontology condition about 1/3 of the search time was spent on browsing the ontologies before a concept was selected. The difference in search time between the two conditions is significant, as shows from a one way ANOVA test (Mk=544, Mo=1106, F=8.48, df=29, P<.05). To compare search time in more detail, the number of queries is counted and the number terms used in queries is counted. In addition the number of retrieved fragments is counted, and the number of retrieved fragments per query is calculated.

Table 6-5 shows the average number of queries and terms used in order to select 15 relevant fragments. The 6 rightmost columns show the average number of terms per attribute (percentages between brackets), and the average number of terms per query.
Many queries were submitted in the ontology condition. The difference with the keyword condition is significant (Mk=27, Mo=55, F=12.88, df=29, P<.05). A factor that could have contributed to the low number of queries in the keyword condition is that some subjects applied a “retrieve all” strategy. This means that one or two general queries were submitted, and fragments were selected based on browsing the long list of hits (showing from the longer selection time in the keyword condition, see Figure 6-9).

A query in the ontology condition often consisted of a Topic (43%), Representational Type (28%) and Instructional Role (16%). Knowledge Type may not have been used intensively because of its degree of abstraction. Possibly this attribute would be used more intensively by experienced instructional designers. The reason that Description Type was not used much may be that the Topic and Description Type ontology are conceptually the closest related, hence the Description Type ontology may not have added enough value to the Topic ontology, which was used the most by far.

Based on these queries, different numbers of fragments were retrieved. Table 6-6 shows the average number of fragments that were retrieved in order to select 15 relevant fragments.

<table>
<thead>
<tr>
<th></th>
<th>Average number of retrieved fragments</th>
<th>Picture</th>
<th>Sound</th>
<th>Text</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keyword</strong></td>
<td>1198</td>
<td>490</td>
<td>57</td>
<td>606</td>
<td>45</td>
</tr>
<tr>
<td><strong>Ontology</strong></td>
<td>1368</td>
<td>386</td>
<td>68</td>
<td>739</td>
<td>175</td>
</tr>
</tbody>
</table>

Table 6-6 Average number of retrieved fragments.

About half of the retrieved fragments were texts, probably because the database contained more texts than other medium types. No significant differences between conditions are found, except for the number of retrieved videos. In the ontology condition more videos were retrieved (Mk=45, Mo=175, F=11.29, df=29, P<.05), possibly because subjects could
explicitly search for videos by specifying medium type in a query. In order to see how many fragments were retrieved per query, the number of retrieved fragments is divided by the number of queries. Table 6-7 shows the average number of retrieved fragments per query.

<table>
<thead>
<tr>
<th></th>
<th>Average number of retrieved fragments per query</th>
<th>Picture</th>
<th>Sound</th>
<th>Text</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword</td>
<td>76</td>
<td>31</td>
<td>4</td>
<td>38</td>
<td>3</td>
</tr>
<tr>
<td>Ontology</td>
<td>31</td>
<td>9</td>
<td>2</td>
<td>16</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 6-7 Average number of retrieved fragments per query.

In the ontology condition more queries were submitted (see Table 6-5) and more fragments were retrieved (see Table 6-6), but the number of retrieved fragments per query was much lower than in the keyword condition (see Table 6-7), which indicates that search results per query were more precise. Consequently less time was necessary for composition, which shows from the re-use profile in Figure 6-9.

To compare the selection time (see re-use profiles in Figure 6-9) in more detail, the number of displayed fragments is counted. Table 6-8 shows the average number of fragments that were displayed during retrieval in order to select 15 relevant fragments.

<table>
<thead>
<tr>
<th></th>
<th>Average number of displayed fragments</th>
<th>Picture</th>
<th>Sound</th>
<th>Text</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword</td>
<td>186</td>
<td>52</td>
<td>14</td>
<td>110</td>
<td>11</td>
</tr>
<tr>
<td>Ontology</td>
<td>175</td>
<td>54</td>
<td>5</td>
<td>94</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 6-8 Average number of displayed fragments.

While fewer fragments were retrieved in the keyword condition (see Table 6-6), more fragments were displayed. (The differences in numbers of displayed sound and video are significant, sound: Mk=14, Mo=5, t=2.58, df=28, P<.05; and video: Mk=11, Mo=22, t=2.77, df=28, P<.05).

The average time fragments were displayed during retrieval in the keyword condition was 1742 seconds (about 29 minutes), and in the ontology condition 1243 seconds (about 21 minutes). The average time a single fragment was displayed in the keyword condition was 10 seconds, and in the ontology condition 7 seconds, (the difference is significant, F=13.68, df=29, P<.05). Possibly because subjects had less information about a fragment in the keyword condition, more fragments were displayed and fragments were displayed for a longer time than in the
ontology condition, which shows from the longer selection time in Figure 6-9. While it can be expected that sound and video fragments be displayed for some time in the retrieval phase in order to select 15 relevant fragments from the database, much time was spent displaying fragments during composition of lesson material as well. During the entire re-use task, including the composition phase, the Windows Media Player window was active in the keyword condition for 431 seconds, in the ontology condition for 635 seconds (the difference is significant, $F=4.43, df=29, P<.05$). Of the 635 seconds that the Media Player window was active in the ontology condition, 285 seconds during retrieval on displaying sound (50 seconds) and video fragments (235 seconds), which means that the remaining 350 seconds were spent during composition. In the ontology condition, more than half the time spent watching videos (and some listening to sound) occurred during the lesson material composition phase. The long display time during composition was possibly necessary to sequence the fragments, however it suggests that videos might have been watched (again) for recreational purposes. With this it should be noted that sound can be listened to without the Media Player window being active. Of the 431 seconds that the Media Player window was active in the keyword condition, 325 seconds were spent during retrieval (171 + 154 seconds sound and video display time). The remaining time spent during composition would be 106 seconds, however this may in fact have been longer because sound was listened to while the Media Player window was not active. In short, fragments were viewed for a longer time during composition than during retrieval in the ontology condition. Oppositely, in the keyword condition fragments were viewed longer during retrieval than during composition. In general this indicates that the use of metadata supports the relevance judgment from a retrieved set of fragments. In the keyword condition a lack of metadata made it necessary to view the fragments in order to judge their relevance. Subjects in the ontology condition based their relevance judgments on the metadata. After a set of fragments was selected and transferred to Microsoft Word, the fragments were viewed (again) in more detail.

6.4.3 Conclusions about re-use time
Overall it took about as long to re-use fragments based on ontologies as it did based on keywords. However in a keyword setting much time was spent on selecting fragments and composing lesson material, while in an ontology setting time was foremost spent on generating queries. The profile of re-use based on keywords was to generate a few queries, in a short time. Much time was spent on selecting fragments, because many fragments were retrieved per query. Consequently many fragments were displayed, mostly texts and pictures (which were retrieved more than sounds and videos). Because the list of retrieved fragments was generally long, and there was no extensive metadata to support relevance judgments, selection of fragments took long. Imprecise search results even had an effect on the time that was necessary to compose lesson material. Relatively much time was spent on composing lesson material. On the contrary, the profile of re-use based on ontologies was to spend much time searching, and less time selecting fragments. Many queries were submitted, but not many fragments were retrieved per query. Because of these precise search results, relatively little time was spent on selecting fragments, of which most was spent on watching videos (which were retrieved much). Not much time was spent on composing lesson material. Of the time spent on composing lesson material, about half of the time was spent on watching the selected
videos again. Thus, overall there is not much difference in time between a keyword and ontology-based approach, but the different times necessary to carry out separate steps (search, select, compose) indicates that there is a difference between the nature of a keyword and an ontology-based re-use task.

6.5 Indexing and re-use time related: cost-benefit

The goal of these experiments was to provide insight in the cost-benefit relation of manual indexing and re-use of small learning objects or fragments. The first two questions were:

- How much time does it take to index a fragment manually, with keywords and with ontologies?
- How much time does it take to re-use a fragment based on keywords and on ontologies?

To index a fragment manually with keywords took 40 seconds, and with ontologies 96 seconds. To re-use a fragment based on keywords took 3 minutes and 43 seconds (re-use time 3343 sec. / 15 re-used fragments = 223 sec.), and based on ontologies 3 minutes and 46 seconds (re-use time 3389 sec. / 15 re-used fragments = 226 sec.).

Table 6-9 shows an overview of the results. The total time is the sum of the average indexing time per fragment and average re-use time per fragment.

<table>
<thead>
<tr>
<th></th>
<th>Total time</th>
<th>Average indexing time per fragment</th>
<th>Average re-use time per fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword</td>
<td>263</td>
<td>40</td>
<td>223</td>
</tr>
<tr>
<td>Ontology</td>
<td>322</td>
<td>96</td>
<td>226</td>
</tr>
</tbody>
</table>

Table 6-9: Total time and average indexing and re-use time (sec.) per fragment.

It was hypothesized that the more time is spent on indexing, the less time is spent on re-use, based on the assumptions that ontology indexing takes long but re-use takes a short time. The hypothesis that the more time is spent on indexing, the less on re-use, does not hold. The assumption that keyword indexing is quicker than with ontologies is correct, but the assumption that re-use with ontologies is quicker than with keywords appears to be wrong. It is and will remain more time consuming to apply a knowledge-rich indexing strategy.

The third question was:

- How often does a fragment have to be re-used to compensate for extensive indexing with ontologies?

Strictly seen, this question is answered with: “extensive indexing can not be compensated.”
sensitivity analysis is performed to refine this answer.

The estimation of keyword indexing time could in reality hardly be any lower (unless done automatically). Keyword indexing time represents the minimal annotation time because subjects may have been thinking up keywords for some time before they started typing. First they viewed a fragment, then they thought of some keywords, then they started typing, which is the moment from which keyword indexing time was measured. The ontology indexing time does include this “thinking” time. When indexing with ontologies, subjects immediately opened an ontology to see which concepts are in it, in order to relate the concepts to the fragment they were viewing. Besides, despite instructions almost half of the subjects indexed with keywords in their native language (Dutch) instead of English. The keywords the subjects used were diverse and the quality of the annotations was not assessed. It might well be that re-using the fragments had taken longer if the keyword indexes as applied by the subjects were used for retrieval. This was not the case; fragments were retrieved based on full text and elaborate lists of keywords.

The estimation of keyword retrieval time, especially search time, is also rather low. Similar to indexing, search time excludes the time used to think about keywords for a query. Subjects may have been thinking up keywords for some time before they started typing, which is the moment from which keyword search time was measured. In the ontology condition this thinking time was included. To generate a query subjects immediately opened a hierarchy of terms, and then started thinking which concepts they would use. Ontology search time was measured from the moment a hierarchy of terms was browsed, hence includes thinking time. In the sensitivity analysis, the indexing time remains the same but the re-use time is refined by excluding the thinking time of a search action in the ontology condition, similar to the keyword measurement. In the ontology condition the time between the moment subjects opened a hierarchy of terms, and the moment the first concept is selected from that hierarchy, is subtracted from search time.

Table 6-10 shows the same overview as in Table 6-9, but search time in the ontology condition is measured from the moment a first concept is selected, until a query is submitted. Total time is the sum of the average indexing time per fragment and the average re-use time per fragment. When search time is refined, re-using fragments based on ontologies does save time.

<table>
<thead>
<tr>
<th></th>
<th>Total time</th>
<th>Average indexing time per fragment</th>
<th>Average re-use time per fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword</td>
<td>263</td>
<td>40</td>
<td>223</td>
</tr>
<tr>
<td>Ontology</td>
<td>308</td>
<td>96</td>
<td>212</td>
</tr>
</tbody>
</table>

Table 6-10 Total, indexing and re-use time (sec.) per fragment, without “thinking time” of a search action in the ontology condition.

The formula to calculate the number of times a fragment has to be re-used in order to make up for extra indexing time is shown below. (A linear relation between re-use time and
number of times a fragment is re-used is assumed).

\[ X = \frac{(KITf - OITf)}{(ORTf - KRTf)} \]

The extra time necessary to annotate material with ontologies is divided by the time saved during re-use with ontologies. Put in different words: a one time effort invested in indexing, counterbalances X times the time saved during re-use.

Based on this formula and the results in Table 6-10, a fragment needs to be re-used 5 times in order to regain the time invested in indexing. This is an outcome of research conducted in a controlled setting. A number of factors have to be taken into account when generalizations are made on the basis of this research, these are discussed in section 6.6.

The last question was: Where can time be saved in the indexing and re-use process? In indexing as well as re-use, costs may be cut by reducing the mental effort that is necessary to handle a large indexing vocabulary. The costs of manual indexing were high where large indexing vocabularies were used. Supporting navigation in long indexing vocabularies, for example by providing a search functionality for concepts in an ontology (for example Schreiber, Dubbeldam, Wielmaker & Wielinga, 2001) would save time in indexing as well as searching. The costs of re-use were high where time was invested in generating queries, and where time was lost if search results were not precise, which implied that long lists of hits had to be browsed in order to find relevant fragments. Avoiding long and diverse lists of retrieved fragments would save time selecting fragments.

6.6 Conclusions and generalizations

This chapter provides some insight in the costs and benefits of a keyword and an ontology-based approach to indexing and re-using document fragments. Manual indexing and re-use by non-professionals based on keywords is generally a little quicker than based on ontologies. The keyword indexing time estimated could in reality not be much lower, even when conducted by domain experts. Experts would probably use fixed lists to describe the material, usually with more than 5 keywords, at different levels of abstraction. This is different from the experimental setting in which subjects added keywords that came to mind quite freely. Besides the indexes would have to be of a certain quality. It was shown that keyword indexing with terms that come to mind is not intrinsically easier than indexing with fixed structured lists of concepts. To add one keyword took about as long as to add one concept to a fragment. Professional keyword indexers may benefit by using a structured is-a hierarchy, allowing annotating material at different levels of abstraction.

Retrieval and composition based on keywords might in real life settings take longer than estimated, because the retrieval technique based on keywords is normally applied in large heterogeneous sets of material, which enlarges the effect of imprecise search results found in this experiment. It is realistic to expect that a repository contains more objects about different topics than was used in this research. Under such circumstances costs of re-use can increase, nevertheless it is concluded that a keyword-based method is generally least expensive, in the first place because of low indexing costs.

Regarding the estimated ontology indexing time, professional cataloguers would probably index material in a shorter time than estimated because they would be more familiar with metadata in their field of expertise than our subjects. However, in practice not many people
are experienced in using for example the LOM metadata model to index material, so the estimation of ontology indexing time may as well be realistic. People's growing experience with metadata models and the developments in support of manual indexing will in the future speed up the indexing process (Najjar, Duval, Ternier, Neven, 2003; Neven, Duval, Ternier, Cardinaels & Vandepitte, 2003).

Retrieval and composition based on ontologies might in reality be done in a shorter time than estimated, when carried out by an experienced instructional designer. Usually people who create lesson material are more experienced than the subjects participating in the experiment in which re-use time was measured. For example, a teacher creating lesson material on a regular basis would be familiar with the metadata, hence would need less time to generate queries, and in addition would probably not spend so much time watching videos after relevant material has been retrieved. The developments in this area, for example resulting in templates for pieces of lesson material, will make it easier (also for less experienced people) to create lesson material in the future.

In general a knowledge-rich approach makes the focus of the retriever shift from the content to the metadata. The ontologies are used to generate queries, and to support the relevance judgment from a set of retrieved fragments. The content was only viewed in more detail after a set of relevant fragments has been selected and transferred to an autgoring tool.

The results of the experiments in this chapter show that a keyword-based approach has the advantage of very low start-up costs in terms of time. The costs of a knowledge-rich index are not regained on a very short term, but are worth the investment if precise search results play an important role. Because both methods have advantages in terms of time, and because people have different preferences, an approach that allows for both keywords and ontology-based retrieval methods is recommended.