**KATHOLIEKE UNIVERSITEIT LEUVEN**
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT COMPUTERWETENSCHAPPEN
AFDELING INFORMATICA
Celestijnenlaan 200 A — B-3001 Leuven

# INFORMATION VISUALISATION: FLEXIBLE ACCESS TO REUSABLE COMPONENTS

Promotoren :
Prof. Dr. ir. E. DUVAL
Prof. Dr. H. OLIVIÉ

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

**Joris KLERKX**

2008

**KATHOLIEKE UNIVERSITEIT LEUVEN**
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT COMPUTERWETENSCHAPPEN
AFDELING INFORMATICA
Celestijnenlaan 200 A — B-3001 Leuven

# INFORMATION VISUALISATION: FLEXIBLE ACCESS TO REUSABLE COMPONENTS

Jury :
Prof. Dr. ir. Dirk Vandermeulen, voorzitter
Prof. Dr. ir. E. Duval, promotor
Prof. Dr. H. Olivié, promotor
Prof. Dr. ir. Ph. Dutré
Prof. Dr. ir. K. De Vlaminck
Prof. Dr. M.-F. Moens
Prof. Dr. M. Specht (Open Universiteit, Nederland)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

**Joris KLERKX**

U.D.C. 681.3*H5, 681.3*H51

2008

D/2008/7515/121
ISBN 978-94-6018-012-5

# Preface

Over the last few years, numerous information sources have been deployed over the Internet, with large amounts of reusable objects and associated metadata. Examples include Wikipedia, social bookmarking tools like del.icio.us, learning object repositories like the Ariadne Knowledge Pool System, etc. Most of the tools for getting access to these resources are based on electronic forms that enable end users to define boolean combinations of search criteria. However, users find it often quite difficult to adequately define such queries. There is thus an obvious need to improve access to those repositories.

Information visualisation is the visual presentation of abstract information spaces and structures, in order to facilitate their rapid assimilation and understanding. This dissertation investigates whether methodologies and technologies for information visualisation can be a solution for the challenge above by answering the following questions:

- how can we effectively and efficiently represent a collection of reusable components by visualisation?
- how can we effectively and efficiently interact with visualisations of a collection of components?

The dissertation is organised as follows:

**Chapter 1** introduces the problem domain and presents challenges and issues impeding flexible, efficient and effective access to large collections of reusable components.

**Chapter 2** presents an extensible information visualisation software framework for reusable components. This framework is structured to support the visualisation pipeline or the computational process of converting a collection of reusable components into a visual form that users can interact with. Having this framework enables us to rapidly experiment with visualisation techniques at the one hand and with different kinds of data-sets, distributed over different data domains, on the other hand.

**Chapter 3** outlines a case study that was created to enable access to the Ariadne Knowledge Pool System (KPS) which is an infrastructure for educational learning objects, with a focus on enabling “Share & Reuse” of learning material for end users. A number of visualisation techniques for mapping a collection of learning objects into visual form are discussed. A prototype-application is created to enable access to the Ariadne KPS. This prototype has been evaluated to validate the used information visualisation techniques.

**Chapter 4** focusses on a collection of reusable components that has been created by disaggregating a collection of powerpoint presentations. A node-link graph is used to enable insight in the contents of this collection. A working prototype has again been evaluated to validate the used information visualisation techniques. One of the results of this evaluation was the idea to focus more on social aspects. This is what we investigate in chapter 5.

**Chapter 5** presents a case study to enable access to collections of social bookmarks and implicit networks between users. A prototype application is created which uses a cluster map visualisation for representing the collection of bookmarks. The evaluation sessions of this prototype introduced the idea of visualising the evolution of tags through time. This has been investigated in chapter 6

**Chapter 6** focusses on enabling access to a network of Learning Object Repositories and looks into visualising search history through time.

Finally, **Chapter 7** concludes this dissertation with a summary of contributions and an exploration of the potential it offers for future research.

# Acknowledgement

Writing a good acknowledgement is perhaps more difficult then writing the dissertation itself. The chance of forgetting someone is pretty high as there are too many people that deserve my gratitude. But let's give it at try..

First of all, I am grateful to my supervisors Erik Duval and Henk Olivé for their support and thoughtful comments through the years.

I would also like to thank the members of my jury Karel De Vlaminck, Philip Dutré, Marie-Francine Moens and Marcus Specht for their time and effort in evaluating this work, and Dirk Vandermeulen for chairing the committee.

Many thanks of course to all my former and current colleagues for creating an inspiring but also relaxing working environment.

I gratefully acknowledge the financial support of the K.U.Leuven research council through the BALO project and the Interdisciplinary Institute for Broadband Technology (IBBT) through the Acknowledge project.

And last but certainly not least, I would like to express my deepest gratitude to my parents, TNT, Dan, Sofie and many other relatives and friends for their continuous encouragement and support.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Over the last few years, numerous repositories have been deployed over the Internet that contain vast amounts of reusable objects with associated metadata. Examples of them are the Wikipedia, social bookmarking tools like del.icio.us, the Ariadne Knowledge Pool System (KPS), etc. Most of the tools for getting access to these resources are based on electronic forms that enable end users to compose boolean combinations of search criteria. However, these kind of queries are not flexible enough to meet all the demands [Marchionini, 2006a]. There is thus an obvious need to improve access to those repositories [Duval and Hodgins, 2003].

Information visualisation is the visual presentation of abstract information spaces and structures to facilitate their rapid assimilation and understanding [Andrews, 2006]. The goal of this dissertation is to investigate how we can improve access to repositories with the help of information visualisation techniques. Before going into the details of challenges and issues impeding this approach, we briefly introduce the field of information visualisation.

The chapter is organised as follows: Section 1.1 presents an overview of the different fields within visualisation research. Section 1.2 outlines the context of technology enhanced learning in which this work was carried out: learning objects research. The concept of Learning Object Repositories is described in section 1.2.1. Section 1.3 summarises issues and challenges that are tackled in this dissertation. An overview of all case studies that have been worked out, is provided in section 1.4. Finally, section 1.5 provides an overview of the subsequent chapters.

## 1.1 Information Visualisation

Information visualisation is a growing field that has emerged from research in human-computer interaction, computer science, graphics, visual design, psychology, and business methods. It is increasingly applied as a critical component in


Figure 1.1: Reconstituted Ptolemy's World Map - The British Library Harley MS 7182, ff 58v-59

scientific research, digital libraries, data mining, financial data analysis, market studies, manufacturing production control, and drug discovery [Shneiderman and Bederson, 2003].In essence, It is concerned with the use of interactive visual representations of abstract data to amplify cognition [Card et al., 1999].

The use of visualisation to present information is not new. It has been used in maps and drawings since a thousand years. One example is Ptolemy's world map (Figure 1.1) that has been created somewhere in the second century BC. Today, there exist a number of fields within the visualisation domain. The most important ones are the following:

- **Scientific visualisation** deals primarily with three-dimensional physical objects and processes such as blood flowing through heart valves. It focuses more on volumes and surfaces [Shneiderman and Bederson, 2003].

- **Information visualisation** concentrates on the visual presentation of vast amounts of abstract data, for which there is no inherent mapping to space and there may not be a natural physical reality [Ware, 2004]. One example of these components are the social relationships that may exist between people, but also attributes of people such as gender, race, home ownership, etc. The goal of information visualisation is to represent this abstract data in a dynamic way to facilitate human interaction for exploration and understanding.

- **Knowledge visualisation** is a field of research which examines our innate potential to effectively process visual representations in knowledge intense tasks. [Burkhard and Meier, 2005]. It examines the use of visual representations to improve the transfer and creation of knowledge between people.
- **Visual analytics** is an area of research and practice that aims for supporting analytical reasoning by interactive visual interfaces. It has been defined as "the science of analytical reasoning supported by the interactive visual interface" [Wong et al., 2006]. Visual Analytics is an inherently multi-disciplinary field ranging from cognitive psychology to database research with applications in medicine & biotechnology, business, security & risk management, and environment & climate research.
- **Visual communication** is the communication of ideas through the visual display of information [Barlex and Carrre, 1985]. It includes mostly two dimensional images like alphanumerics, art, signs, and electronic resources.

This dissertation focusses on techniques from the field of information visualisation. In the context of this dissertation, we consider information visualisation as technology, i.e. a collection of methods, techniques, and tools developed and applied to satisfy a need. Visualisation has to be effective and efficient: it has to do what it is supposed to do, and it should do this with a minimal amount of resources [Van Wijk, 2006].

## 1.2 Learning Objects Research

This section outlines the context of technology enhanced learning in which this work was carried out: learning objects research. According to the IEEE LTSC Learning Object Metadata standard, a learning object is "any entity, digital or non-digital, that may be used for learning, education or training" [IEEE, 2002]. Learning objects research generally falls into one of two categories [Wiley, 2007].

1. The traditional approach to using learning objects focuses on enabling the just-in-time automated assembly of carefully structured learning objects to create personalised educational experiences.
2. The permissive approach to learning objects focuses on making the reuse and localisation of all resources, regardless of their structure, as effective and efficient as possible.

This dissertation fits in the second approach of improving access to learning objects.

Learning objects are often described with extensive standardised metadata. The IEEE LTSC Learning Object Metadata (LOM) standard [IEEE, 2002] is the primary standard for the description of learning objects (LOs) [Wiley, 2007]. Relevant attributes of learning objects that are described include: title, author, owner, format, interaction style, etc. LOM enables sharing of descriptions of learning resources between resource discovery systems which should lead to a reduction in the cost of providing services based on high quality resource descriptions [Duval and Hodgins, 2003]. It will be used in this work to improve access to the learning objects.

Learning objects, together with metadata, can be are stored in Learning Object Repositories (LORs). The concept of LORs will be outlined in Section 1.2.1. Section 1.2.2 introduces the research that is performed within ARIADNE, the association in which the work of this thesis was carried out.

### 1.2.1 Learning Object Repositories

LOs can be stored in databases called learning object repositories. Different types and examples of Learning Object Repositories exist. We will focus on the ARIADNE Knowledge Pool System, which is a distributed repository of learning objects and associated metadata [Duval et al., 2001]. We will validate our work by improving access to this repository. A thorough review of the different types and examples of LORs can be found in [Verbert et al., 2008] and [Ternier, 2008].

Current research on learning object repositories focuses, amongst others, on learning object discovery [Ochoa and Duval, 2006b], [Orzechowski et al., 2007] and interoperability between repositories [Prause et al., 2007]. Interoperability research focuses on connecting and using learning objects located in heterogeneous and unaligned repositories. For instance, through the Simple Query Interface [ [Ternier et al., 2008], a protocol for searching repositories, the ARIADNE KPS, Merlot and EdNA Online are currently interconnected. Finally, research on discovery of learning objects is concerned with user proling for more accurate learning object discovery [Orzechowski et al., 2007], enhanced search mechanisms [Zimmermann et al., 2007], ranking and recommendation of learning objects [Ochoa and Duval, 2006a] and information visualisation techniques to enable flexible and efficient access to learning object repositories [Klerkx et al., 2004].

### 1.2.2 ARIADNE

ARIADNE (Association of Remote Instructional Authoring and Distribution Networks for Europe) [Duval et al., 2001] is an association whose aim is to promote share and reuse of learning objects. The core of the ARIADNE infrastructure is a distributed network of learning object repositories. Current research in ARIADNE focuses, amongst others, on:

- Access to distributed collections, through federated search based approaches or metadata harvesting;
- Content models, that define learning object components and how they can be aggregated, so as to enable reuse and repurposing;
- Information visualisation, as a means to enable analysis of the overall content of a large-scale repository;
- Metadata and how they enable flexible access to learning objects, with a specific focus on both automatic metadata generation and "attention metadata", that describe user interactions with content;
- Social information retrieval techniques for flexible access to large-scale collections of content.

The information visualisation research is presented in this dissertation. For this research, we built further on other ARIADNE research. For example, chapter 4 describes visual access to a repository that has been created in the context of the content models research [Verbert et al., 2008].

## 1.3 Issues and Challenges

Duval & Hodgins have listed a number of interrelated research issues in [Duval and Hodgins, 2003]. Those issues are important for enabling learning object reuse on a global scale. This section outlines the issue of flexible access to huge repositories and the challenges that have been addressed in this dissertation.

### 1.3.1 Flexible Access to Reusable Components

Applications can take advantage of the structured nature of the metadata while issuing queries on those repositories. In this way, they can select learning material that is most suited to the needs of end users like learners or teachers. For example, if a teacher searches for a slide presentation of 15 minutes on inheritance in object-oriented programming, with a target audience of university students, he or she can compose an advanced query that can be matched against the LOM descriptions that describe the objects.

Most of the tools at this moment to access LORs are based on an electronic form that enables end users to compose boolean combinations of search criteria or on a simple text box, that fails to take advantage of the structured nature of LOM. However, queries typed into search boxes are not effective enough to meet all the demands [Marchionini, 2006a]. End users should be able to quickly zoom in on those LO's that are relevant to them. They should be able to find those LOs without requiring him or her to go through a lengthy process to formulate complex search criteria, to evaluate some of the results, refine the search criteria, etc.

The **main challenge** that this dissertation addresses is therefore

- how to provide flexible, efficient and effective access to large collections of reusable components and enable insight in the contents of such a collection.

One possible paradigm that was proposed by Duval & Hodgins is information visualisation [Card et al., 1999]. This enables end users to manipulate controls over the metadata, zoom in on potentially more relevant LO's and continuously keep an overview of how additional search criteria restrict the remaining number of LO's. It is important to notice that this approach tries to empower end users and keep them in control of the selection process, rather than trying to transfer the focus of control to the Learning Management System. For example, the learning management system might process rules that match characteristics of the learner (style, intended learning outcome, already acquired knowledge and skills) against the LOM descriptions, in order to select the most appropriate LO's. In this case, the user is not in control of the selection process.

Information visualisation thus follows the exploratory search approach that is suggested by Human-computer information retrieval (HCIR) research. This is the study of information retrieval techniques that brings human intelligence into the search process [Marchionini, 2006c]. The goal of HCIR is to integrate human and system interaction so that people are continuously engaged with meaningful information, thus incorporating understanding with finding [Marchionini, 2006b].

A somewhat "other" approach to LO selection, relies on social recommending and collaborative filtering techniques where similarities between end users are exploited to suggest potentially relevant LO's [Goldberg et al., 1992].

In both approaches, LO's are identified in a more natural, less laborious way than by "filling in electronic forms". Other research in ARIADNE is concerned with this second approach. In this dissertation, we will focus on the first approach. We will investigate whether methodologies and technologies for information visualisation can be a solution for the main challenge above by answering the following questions:

- how can we effectively and efficiently represent a collection of reusable components by visualisation?

- how can we effectively and efficiently interact with visualisations of a collection of components?

### 1.3.2 Extensible Framework for Visualising Collections of Reusable Components

In the context of this research, we do not want to narrow down the scope of our research to enable flexible access to only one collection. We want to enable access to multiple collections with different kinds of information visualisation techniques to improve access to them. To be able to carry-out this work, we need a framework that supports the visualisation pipeline, which is the process of converting a collection of reusable components into a visual form that users can interact with. This visual form can then be used to explore those collections and gain insight in them.

The objective is therefore to design an infrastructure that allows us to rapidly experiment with visualisation techniques at the one hand and with different kinds of collections, distributed over different domains, on the other hand. This infrastructure should allow us to create visualisations that are efficient, effective, flexible and scalable. Many algorithms for visualisation techniques already exist. Because we do not want to reinvent the wheel, we need to find ways to integrate those algorithms in this infrastructure.

Note that although this framework is a big challenge, it only is a means to an end to carry-out our research to solve the main challenge that we discussed in the previous section.

## 1.4 Case Studies

For this dissertation, we adopted a case study methodology. This section will introduce the case studies that have been worked out during this dissertation. Each case study will be discussed in detail in the subsequent chapters. They investigate how to enable access to different kinds of repositories in different ways by making use of information visualisation techniques. A working prototype has been created for each study to be able to validate the used techniques with a number of evaluations.

### 1.4.1 Visualising a Learning Object Repository

Chapter 3 outlines a case study that has been created to enable access to the Ariadne Knowledge Pool System (KPS) (section 1.2.2). This case study investigates how to enable access to a collection of learning objects by starting from a visualisation of the topic of the objects in the repository. Therefore, a number of visualisation techniques for mapping the topics of a collection of learning objects into visual form are discussed.

A working prototype has been created and used to perform user evaluations which enabled us to validate the use of information visualisation techniques for improving access to learning object repositories. This prototype is used nowadays in ARIADNE, the ZoEp-project in the Netherlands and in the Aristomusic company. The contents of this chapter have been published in [Klerkx et al., 2004] and [Klerkx et al., 2005].

### 1.4.2 Visualising Actual Reuse of Reusable Components

Besides enabling access to a LOR by starting from a visualisation of the topic of the learning objects, relationships between those objects can also reveal relevant information for finding them. Chapter 4 therefore focusses on a case study for opening up a collection of reusable components by exploiting those relationships. This repository has been created by disaggregating a collection of powerpoint presentations. As an alternative to visualising the topic of those components, this case study will start from a visualisation of the inner relationships between the different components. A working prototype has been created to validate whether the visualisation and interaction techniques are efficient and effective to improve understanding of the material in the collection. This prototype has been used to help the research on content models that has been discussed in section 1.2.2. This chapter has been published in [Klerkx et al., 2006].

### 1.4.3 Visualising Social Bookmarks

Chapter 5 presents a case study that enables access to collections of social bookmarks. This time, we start from a visualisation of the implicit relationships between users, tags and content in Web2.0-style social information retrieval systems. Those relationships are not always clear in the traditional ways of accessing social bookmarks. This chapter will therefore focus on using information visualisation techniques

- that enable insight in those implicit relationships, and
- that may offer end users new ways to find content and information that could be of interest to them, but that would not have been found through explicit searches.

A prototype application has been created on top of the del.icio.us social bookmarking tool [delicious, 2008]. Del.icio.us is probably the most well-known social bookmarking tool, designed to store and share bookmarks on the web instead of saving them in the browser. We chose del.icio.us as a source of data as it is highly popular with many users, lots of data and it has a very easy API to access this data. Those characteristics make del.icio.us a very good reference point for other social bookmarking tools. This working prototype is currently online available for all del.icio.us users. An online survey has been performed to validate the information visualisation techniques that have been used to enable access to these kinds of collections.

This chapter has been accepted as such for publication in a special issue on "Social Information Retrieval for Technology Enhanced Learning" of the Journal of Digital Information (JoDI). Earlier versions of this chapter has been published in [Klerkx and Duval, 2007c].

### 1.4.4 Visualising a network of Learning Object Repositories

Chapter 6 focusses on enabling visual access to a network of Learning Object Repositories instead of only one LOR. GlobeMash is a mashup application that focuses on opening up the Global Learning Objects Brokered Exchange consortium [Globe, 2008], which aims to unlock the deep web of the learning repository networks that its members maintain. GlobeMash enables users to

- find learning objects within the GLOBE network of repositories,
- obtain a visual insight in the search history of users.

Early results of this work have been published in [Klerkx and Duval, 2007a] and [Klerkx and Duval, 2007b].

## 1.5 Outline

This dissertation describes

- an extensible visualisation framework,
- the designs and prototypes of tools that have been developed with this framework,
- existing information visualisation techniques that have been used within those designs and prototypes, and
- the evaluations that have been conducted.

This work enables us to validate the use of information visualisation techniques to provide flexible, efficient and effective access to large collections of reusable components and enable insight in the contents of such a collection. Earlier versions of the chapters have been published, in whole or in part: [Klerkx et al., 2004], [Klerkx et al., 2005], [Klerkx et al., 2006], [Klerkx and Duval, 2007c], [Klerkx and Duval, 2007a] and [Klerkx and Duval, 2007b].

The remaining chapters are organised as follows: Chapter 2 tackles the issue of an open infrastructure for visualising collections of reusable components and presents an extensible information visualisation software framework for reusable components. Points of interest are extensiveness, openness, scalabilty and performance.

Chapter 3 outlines the case study on visualising a learning object repository (Section 1.4.1).

Chapter 4 focusses on opening-up a collection of reusable components by visualising the relationships between the different components (Section 1.4.2).

Chapter 5 presents the case study that enables access to collection of social bookmarks (Section 1.4.3).

Chapter 6 focusses on enabling visual access to a network of Learning Object Repositories (Section 1.4.4).

Finally, chapter 7 concludes this dissertation with a summary of contributions, lessons learned, and an exploration of the potential it offers for future research.

# Chapter 2

# An Extensible Framework for Visualising Collections of Reusable Components

## 2.1 Introduction

The main challenge that this dissertation addresses is how we can provide flexible, efficient and effective access to large collections of reusable components with associated metadata and enable insight in the contents of such a collection. Whenever we use the term reusable component in this dissertation, we mean reusable content components like e.g. videos, social bookmarks, Wikipedia articles, etc. The hypothesis is that information visualisation techniques can be a solution for this challenge. The methodology to investigate this hypothesis is case study research.

To enable this research, we need an extensible framework for information visualisation techniques. The purpose is to have a test bench which enables fast experimentation with (i) new visualisation techniques on the one hand, and with (ii) collections of reusable components on the other hand. Those components may be distributed over different domains and described with any possible metadata standard.

The second challenge of our research is therefore to develop such infrastructure that allows us to create visualisations that are efficient, effective, flexible and scalable. This chapter presents this framework, that was designed and developed so that it would be open and extensible. The core functionality is to map collections of reusable components onto a visual form that can be used to amplify understanding and hence enable insight in the contents of those collections.

This chapter discusses the architectural design of the software framework we developed. We will start this chapter by outlining related work in section 2.2. The goal of architectural design is to create an architecture with a set of properties that form a superset of the system requirements [Fielding, 2000]. Therefore, we describe our framework requirements in section 2.3. These are deducted from the problem domain, because "Form follows function" [Fielding, 2000]. Section 2.4 presents the framework architecture that was designed to support those requirements. Section 2.5 outlines a number of implementation notes on the framework. The framework architecture has been evaluated by common quality attributes that provide the means for measuring the fitness and suitability of a system [Bass et al., 2006]. These are discussed in section 2.6. Finally, this chapter is concluded in section 2.7.

A number of case-studies that were created with this framework have been published in [Klerkx et al., 2005], [Klerkx et al., 2006], [Klerkx and Duval, 2007a] and [Klerkx and Duval, 2007c].

## 2.2 Related Work

Over the last years, a number of information visualisation projects and products have been created from scratch. Several applications have been developed for specific data structures and visualisations. A thorough overview of existing toolkits and libraries can be found in [Fekete, 2004]. All toolkits and libraries are developed for a specific purpose:

- For instance, the Prefuse toolkit [Heer et al., 2005], implemented in java, is more oriented towards data whose structure corresponds to that of a graph. Among them you can include networks, hierarchies, trees, etc.
- The Geovista studio is mainly oriented towards geoscientific analysis. However the toolkit can also be reused in other projects where the focus lies in some other domain [Takatsuka and Gahegan, 2002].
- Other toolkits focus more on one specific visualisation technique. The TreeMap Java Library [TreeMap, 2003] and the HCIL Treemap 4.0 toolkit [HCIL, 2008] both focus on visualisations of treemap algorithms. However the first one can visualise squarified cushion treemaps where the latter can visualise ordered and quantum treemaps [Bederson et al., 2002].
- The InfoVis Cyberinfrastructure is a central resource unit that provides access to a comprehensive set of software packages easing the exploration, modification, comparison, and extension of data mining and information visualisation algorithms. Its website is complemented with a series of learning modules about the different aspects of data mining and information visualisation, software, databases and the available computing resources [Shashikant et al., 2004].

Our goal is to have a test bench which enables fast experimentation with (i) new visualisation techniques on the one hand, and with (ii) collections of reusable components on the other hand. The overview in [Fekete, 2004] shows there exist a lot of libraries that consist of implemented visualisation techniques. We do not want to recreate these implementations but rather reuse them whenever possible. We want to be able to integrate a collection of reusable components once, and visualise them with multiple visualisation techniques without having to create mappings between the collection and every existing library. Our framework is therefore specifically oriented towards collections of reusable components with structured metadata that describes them.

## 2.3 Frameworks Requirements

The core functionality is to map collections of reusable components into a visual form. If we consider this framework as a collection of methods, techniques and tools developed to satisfy these needs, then standard measures apply [Van Wijk, 2006]: visualisation has to be *effective* and *efficient*. In other words: visualisation should do what it is supposed to do and has to do this using a minimal of resources. To achieve these goals, the software framework has to be open and extensible, scalable and performant. These requirements are detailed in the rest of this section.

### 2.3.1 Extensibility

Extensibility is defined as the ability to add functionality to a system and to support easy integration of new software components, diverse software packages, etc . [Bass et al., 2006]. Dynamic extensibility implies that functionality can be added without impacting the rest of the system [Fielding, 2000]. To meet our goal of having a software framework for information visualisation techniques, so that we can validate their effectiveness for different kinds of collections, we need to ensure that the design of the architecture is extensible on multiple grounds:

- First of all, easy integration of existing information visualisation techniques and third-party libraries should be possible. As there are more than hundred different visualisation algorithms, implemented in different libraries and different programming languages, the framework must have a means to quickly select a visualisation technique for a given data collection.

  Research on different visualisation techniques and algorithms is a continuous process. For example, the use of a tree-map visualisation was suggested in [Card et al., 1999] as a compact visualisation of directory tree structures. Numerous extensions and implementations of this idea have been created during the following years. We want to be able to use the different implementations of this technique and its extensions for every possible data collection, without having to write code for well-documented algorithms from scratch.

- Secondly, it should be easy to add new data collections, in various formats and structured according to various metadata schemes. In this way, we can easily create case studies of and validate whether the visualisation techniques can provide visual insight in these collections.

### 2.3.2 Scalability

Scalability refers to the ability of the architecture to support large numbers of components, or interactions between components, within an active configuration [Fielding, 2000]. As we want to enable effective and efficient access to collections with any size of reusable components, the software framework should be able to cope with very large numbers of objects. Dealing with such size is quite challenging in terms of memory and performance. The design of the architecture of the software framework needs to cope with this accordingly.

### 2.3.3 Performance

Performance is all about timing [Bass et al., 2006]. Events like interrupts, messages, requests from users, or the passage of time, occur and the system should react to them. Looking from an information visualisation perspective, the most important events are the ones coming from human interaction. We know from [Shneiderman and Bederson, 2003] that visualisation systems should achieve a rapid 100ms interaction loop if we want the user perceived performance to be any good. As we work with third-party libraries for information visualisation techniques, achieving this proved to be quite challenging since the performance of those libraries is not always in our hands. Memory performance is also important. When the framework is used for visualising a repository with 50.000 objects and more, the time that is needed to create the underlying data models for the framework should stay low. The reason for this is that visualisation of collections of reusable components onto a visual representation should be effective and efficient.

## 2.4 Framework Architecture

### 2.4.1 Introduction

As said before, the core functionality is to map collections of reusable components into a visual form. In other words, our framework should support the visualisation pipeline which is the computational process of converting a collection of reusable components into a visual form that users can interact with [Card et al., 1999]. We will explain the overall structure of the software framework in the following

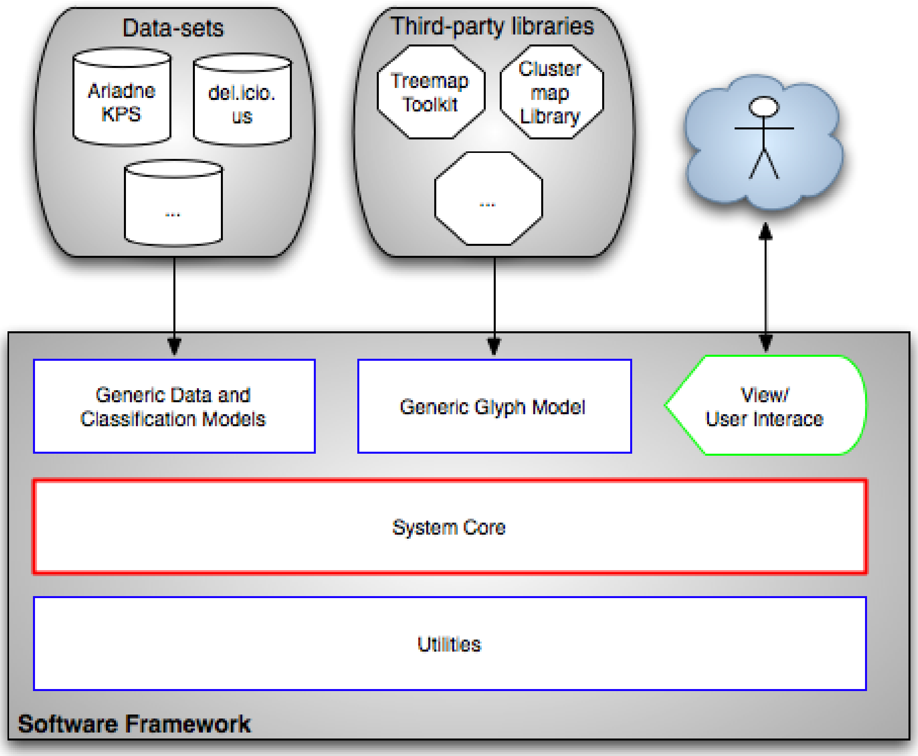



Figure 2.1: Structure of the extensible framework for visualising collections of reusable components.

sections. Section 2.4.7 explains how the different modules of the framework work together to support the visualisation pipeline. Figure 2.1 shows an overview of the framework architecture which consists of the following modules:

1. The generic data and classification models allow easy integration of collections of reusable components with associated structured metadata schemes. For example, the case study in chapter 3 shows how collections of learning objects can easily be visualised if they are described by the Learning Object Metadata (LOM) standard. Section 2.4.2 explains these models in detail.

2. Different libraries for both visualisation and interaction can be plugged into the framework by implementing a generic glyph model. A glyph is a graphical object designed to convey multiple data values [Ware, 2004]. Section 2.4.3 presents this generic glyph model.

3. The system core is responsible for a number of activities in the framework. First of all, it takes the created glyph model and renders it to the final visual representation in the user interface. Secondly, the user can interact with this visualisation and make changes by e.g. defining range queries to filter out objects that are not relevant or include others that are [Ahlberg and Shneiderman, 1994]. The system core module is responsible for making these changes happen as fast as possible because performance is an important system requirement. The system core is detailed in section 2.4.5.

4. A simple user interface is available in the framework that allows easy visualisation of collections of reusable components (see section 2.4.4).

5. The last module is a collection of utilities that can be used in all parts of the framework. These utilities consists of logging facilities for user actions, database connectors, xml-parsers for metadata standards like LOM, and so on. Another important aspect is the possibility to create indexes for the created models of the reusable components in order to optimise the visualisation and filtering process during run-time (see section 2.4.6).

The remainder of this section explains these modules and the reason we need them in our framework. Section 2.4.7 explains how these modules support the visualisation pipeline [Card et al., 1999].

### 2.4.2 Generic Data and Classification Model

#### Data model

Different collections of reusable objects are often (i) described with different metadata standards and (ii) stored in different ways like relational databases, collections of xml-files, etc.. For example, the ARIADNE Knowledge Pool System (KPS), that will be described in chapter 3, uses LOM to describe their learning objects. They are stored in both a relational database and a XML-database. Del.icio.us, a social bookmarking tool, keeps non-standardised metadata about their bookmarks and are online available on the Web. To be able to visualise all collections of reusable components, we need a general model for internally representing reusable components. This model contains a small general-purpose element set to be reused in many different contexts.

Our framework has grown incrementally while investigating the use of visualisation techniques in various case studies. The element set of our model, has been stipulated by our first case study to open up learning object repositories (chapter 3). The elements were selected by choosing those in the LTSC IEEE LOM [IEEE, 2002] standard that are most frequently used in searcher queries [Najjar et al., 2004]. These are title, description, authors and classification.

The framework is able to distinguish the different reusable objects. If an interesting object would be found by an end user, the framework is able to present the real object to the user. Therefore, technical elements have been added like identifier and location of the object. For the case study, described in chapter 4, we needed to be able to model the “is part of / has part” relationships between components. Therefore, relations were added to the set.

All these elements are also available in other standards like Dublin Core, MPEG-7, etc.. The authors of [Kules and Shneiderman, 2003] came to the same conclusion by looking at a review of National Center for Health Statistics (NCHS) and Center for Disease Control. On top of the elements proposed above, they add “Document Type” with possible values as “report”, “summary”,“press release”, etc.

The elements are summarized in Figure 2.2 where they are presented as attributes in the UML-diagram of the interface “ReusableComponent”. When a new dataset is added to the framework, the metadata that describes the objects have to be mapped to the general model by implementing the interface of “ReusableComponent.” It is not mandatory to map all elements. For example, in the del.icio.us case, there are no relationships between different social bookmarks available. This element will therefore not be mapped from del.icio.us to our general data model.

Reusable components from ARIADNE and del.icio.us are thus described with different metadata. Figure 2.2 shows that “AriadneLearningObject” and “DeliciousObject” both extend “ReusableComponent” with extra features that are not available in the general model. A visualisation of these collections will be described in chapters 3 and 5 of this dissertation.

### Classification model

In many cases, components are classified in a classification system. For instance, in the ACM digital library, we need to keep track of the relationships between objects and their ACM classification. Therefore, the framework is equipped with a model for representing a classification. A classification is represented as a number of taxonomic paths in the specific classification system. Each taxonomic path is an ordered list of taxons where each taxon in the path is a particular term within the taxonomy. This is summarised in figure 2.3. This way of classifying content is also used in LTSC IEEE LOM standard [IEEE, 2002].

IMS Vocabulary Definition Exchange (VDEX) allows the exchange and expression of simple machine-readable lists of human language terms, along with information that may assist a human in understanding the meaning of the various terms. For instance, VDEX allows hierarchical classification systems to be expressed using source/target value pairs in specified relationships. In retrospect, VDEX would have been a better alternative to model classification systems in our framework because it is a standardised specification and would have therefore made it more open and extensible.

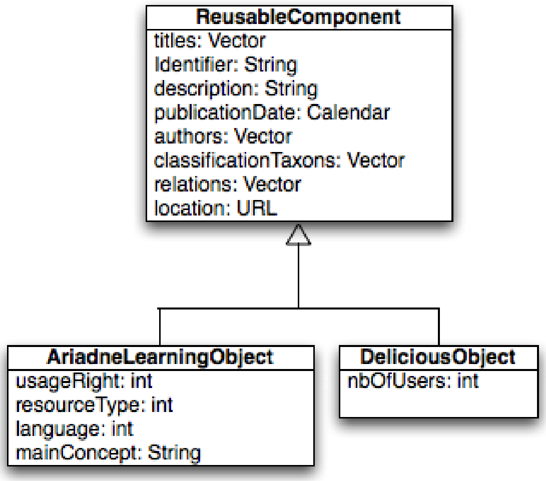



Figure 2.2: ReusableComponent interface and two examples of classes that implement it. The attributes of the ReusableComponent interface define the generic data model of the framework. Not all attributes are mandatory.

More recently, the W3C Semantic Web Deployment Working Group (SWDWG) published a second working draft of the "Simple Knowledge Organization System (SKOS)" specification. SKOS is an area of work developing specifications and standards to support the use of knowledge Organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web. We should consider this specification as an alternative to both our classification model and VDEX in the future.

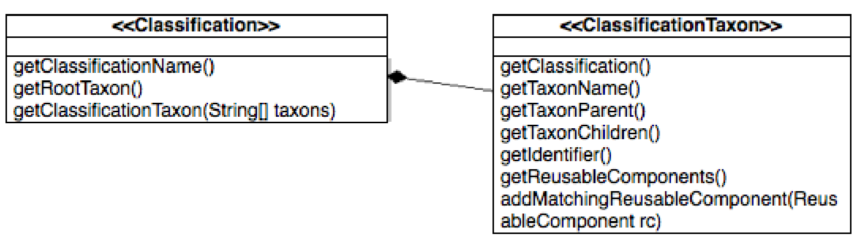



Figure 2.3: A classification is an aggregation of different taxonomic paths.

### 2.4.3 Generic Glyph Model

The glyph model is a model for the internal representation of a reusable component as a glyph. A glyph is a graphical object designed to convey multiple data values [Ware, 2004]. For instance if we would visualise a collection of music songs, we could represent each song visually as a glyph. The shape of the glyph might be used to represent the the genre, the size might be used to represent the length of the song, the color to represent the popularity, and so on. Every library for visualisation techniques has its own way of internally representing the glyphs that are to be rendered on screen.

- The “Infovis Toolkit” [Fekete, 2004], uses a table of columns as a unified data structure for representing the information about glyphs. Those columns contain objects of homogeneous types, such as integers and strings.
- The Aduna Clustermap library [Fluit et al., 2005] uses another approach and represents every glyph as simple java objects. Each glyph is an instantiated object of the class “DefaultObect” which provides only two attributes that are used to visualise a name and a location for the object in question.
- The Tree map Java Library [TreeMap, 2003] presents every glyph as a combination of three simple java objects “TMNode”, “NodeDraw” and “NodeSize”.
- etc.

Because every library has its own way to represent the glyphs, we have created our framework model for representing glyphs. Our model exists of two general java classes:

- CNode, i.e. classification node, represents a visual node of a specific taxon in our classification model.
- RCNode, i.e. reusable component node represents a visual reusable component in our data model.

To add new visualisation libraries, we have to create a mapping from our general glyph model to their internal representations. For instance, the mapping process of the “Infovis Toolkit” will have the responsibility to add an entry to the table of columns for each instantiated CNode and RCNode. Once that mapping is created, all techniques in those libraries are available for the visualisation of integrated data collections.

The advantage of using an intermediate model, is the fact that we only have to map a new collection into our model once. From that moment, it can be visualised with all techniques without having to know every detail about the different glyph models in the libraries. If we would not have chosen to work with internal models, we would have needed to create mappings from each data collection to each glyph model in all libraries. This would have been less extensible and more time-consuming. In other words, a one-to-many mapping is achieved instead of a many-to-many mapping between data collections and visualisation libraries.

### 2.4.4 Simple User Interface

We want to use the framework to quickly create new prototypes of applications to open up collections of reusable components with information visualisation techniques. In this way, we can investigate whether some visualisation techniques would work for a specific collection of objects. Therefore, we have added a simple default user interface that can be reused and customised to the requirements of the newly added case study. This interface consists of basically nothing more than a simple panel that can be used to quickly view a visualisation of a data collection.

### 2.4.5 System Core

The system core glues all modules together. The module consists of a number of different components. Each component is responsible for a number of tasks that are needed at runtime to support the visualisation pipeline of mapping collections of reusable components onto a visual form. Those components and their tasks are described below. Section 2.5 presents implementation notes for the most important tasks.

#### VisualisationMgr

This component is responsible for:

- creating the visual mapping between the data and classification models and the glyph model, and
- rendering and showing the visualisation in the simple user interface.

#### RegistryMgr

This component is responsible for:

- ensuring a global point of accessing resources that are needed at runtime. Resources are the instantiated data and classification models, but also all components of the system core that are described in this section.

#### ReusableComponentMgr

This component is responsible for:

- the creation of the internal data and classification models for a specific collection of reusable components,
- the creation of indexes to optimise the performance of search and filter processes,
- returning the correct reusable component when it is requested by one of the other components in the system core.

**QueryMgr**

The QueryMgr is responsible for

- issuing queries in the collection of the reusable components and retrieving the results in an efficient way.

**FilterMgr**

The FilterMgr is responsible for

- performing filtering operations. For example, if a user only wants to visualise material that has been created in a specific period of time, the FilterMgr will select only those components that match.

**LogMgr**

The FilterMgr is responsible for

- keeping logs of every event. An example of an event is a query being issued by the QueryMgr.

### 2.4.6 Utilities

This module consists of a number of utilities to facilitate the process of quickly creating prototypes of case studies. They can be used in all parts of the framework. The most important ones are listed below.

- **Logging facilities** are used to send logs to the Contextualized Attention Metadata (CAM) web service. CAM services capture the attention of a user on content in an application to support a feedback loop, enabling learning from the way people actually use technologies and tools.
- **Database connectors** are used to easily connect to a DBMS like mySql, oracle, to speed up the process of plugging new datasets into the framework.

- For the same reason, **xml-parsers** are available for e.g. LTSC IEEE LOM metadata. In this way a learning object repository (LOR) can quickly be visualised with the framework without much effort. Ongoing work on "Harmonization of Metadata Standards" in the context of the Prolearn project, analyses a number of existing metadata specifications in order to isolate reasons and issues behind harmonization problems of a plethora of metadata specifications (such as IEEE LOM, Dublin Core, METS, MODS, MPEG-7, etc.). This work could be helpful in the future for adapting those parsers so that they support other standards as well. This would make it easier to visualise collections of reusable components, described in other metadata schemes than LOM.

- **Index facilities** are necessary when visualising collections with many reusable components. The framework can harvest such a dataset, create the necessary intermediate data and classification models and put them in Lucene [Lucene, 2008] indexes. Those indexes are used to optimise the visualisation rendering process and the handling of actions. An example of such action is issuing a keyword query to the collection of reusable components.

- **IO-facilities** can be used to store information like e.g. the created intermediate data and classification models. This information can be restored afterwards so that it is not necessary to recreate those intermediate models each time the created application is used. On top of that, it can be used by end users to save information about reusable components that were possibly found during a visual exploration of the collection of reusable components with the created visualisation application.

### 2.4.7 The Visualisation Pipeline

The visualisation pipeline is the computational process of converting information into a visual form that users can interact with [Card et al., 1999]. The following chapters will present case studies that have been created with our framework. Every case study investigates how we can provide flexible, efficient and effective access to a specific collection of reusable components by using information visualisation techniques. We will therefore need to customise the visualisation pipeline, depending on the case study that we are working out.

The general involved steps of the visualisation pipeline can be seen in Figure 2.4. These steps are detailed below.

1. The first step is the **data analysis** step. Its purpose is to transform the information of the collection of reusable components into the generic data and classification models of the framework. Various data processing steps can be used in this step to manipulate the data as needed. For instance, in the

case of visualising a learning object repository (chapter 3), the "ReusableComponentMgr" harvests data from the ARIADNE repository using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) protocol [Sompel et al., 2004]. The returned data consists of learning object metadata (LOM) instances. Those instances are processed and a classification model is built, based on the classification information that may be returned within those LOM instances. After this step, objects of "ReusableComponents" are created and assigned to this classification model. Those objects are stored in Lucene indexes for optimising data retrieval.

2. The second step is the **visual mapping** step. Its purpose is to map the created data and classification models into visual form or in other words into the generic glyph model. For each object of "ReusableComponent", the "VisualisationMgr" creates the needed instances of CNode and RCNode. Whenever a visualisation library is integrated into the framework, a mapping is created from our glyph model to the underlying data models of that library. Based on the selected visualisation algorithm, the "VisualisationMgr" will use the correct mapping and create the necessary data models for the library of the visualisation algorithm.

3. Step three involves the **rendering** process. The glyph information that was created in the second step, is rendered on the screen. The "VisualisationMgr" instructs the underlying library to render its visualisation. This visualisation is shown to the user in the simple user interface. The data collection is then visually available for the user to explore and interpret.

4. Step four is the **filtering** step. It is responsible for the filters that have to be applied on the collection. For instance, if only a part of the collection of reusable components has to be visualised, a filter can be applied to achieve this goal. For instance, imagine an end user who is only interested in learning objects on "Exact Sciences". It makes no sense to visualise other material on screen because we would lose space for visualising the material the user is actually interested in.

5. Finally, users can **interact** with the user interface to alter the resulting visualisation by e.g. changing a dynamic query control, and make further interpretations of the visualised collection of reusable components. For example, in the simple user interface, a keyword query can be issued to the system. The "QueryMgr" will use the Lucene indexes to match the chosen keywords with the reusable components. The "FilterMgr", that is responsible for all filtering operations, will make sure that only those objects that match all current filters are visualised on screen.

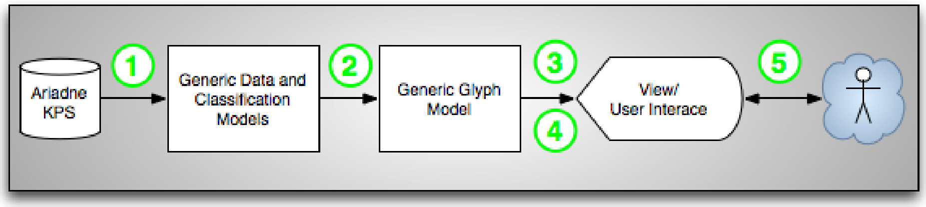



Figure 2.4: The visualisation Pipeline: Converting a collection of reusable components into interactive visual representation (adapted from [Card et al., 1999]).

## 2.5 Implementation Notes

### 2.5.1 Open source Java™Software Framework

The software framework is open source and implemented in Java™which produces a number of advantages:

- a number of visualisation techniques are already implemented in toolkits and libraries like Prefuse [Heer et al., 2005], the Infovis Toolkit [Fekete, 2004], Improvise [Weaver, 2004], Processing [Processing, 2008], JUNG [White et al., 2008], the visualisation Toolkit (VTK) [Schroeder et al., 1993], Piccolo [Bederson et al., 2004], etc. They all have their own implemented algorithms and are implemented in Java™which makes it easier to integrate them into the framework.

- platform independence is needed because our goal is to open-up collections of reusable components for every end user, working on every possible platform.

- easy integration on the Web by using Java Web Start Technology and by using Java Applets, helps the deployability of the created applications. For example, the visualisation application that has been created in Chapter 3, is online deployed as an applet in the ARIADNE web application for finding and sharing learning resources.

- availability of the Java 2D API. This is a set of classes for advanced 2D graphics and imaging, encompassing line art, text, and images in a single comprehensive model [Sun, 2008].

### 2.5.2 Software Design Patterns

A design pattern describes relatively complex protocols of interactions between objects as a single abstraction [Fielding, 2000] [Gamma et al., 1995]. The framework has a simple, developer-friendly API which uses well-known software design patterns to provide quality attributes like extensibility and modifiability (see section 2.6). The following design patterns are used:

- the Factory pattern [Cooper, 2000] in the interfaces for returning the right builders for (i) creating the internal data and classification models for a specified system, (ii) the underlying data models of the integrated libraries. For example, listing 2.1 illustrates the use of this pattern by showing how the "NodeBuilderFactory" looks up the correct builder class for building the underlying data models for the chosen visualisation algorithm.

- the Builder pattern [Cooper, 2000] in the classes for the construction of the (i) internal data and classification models for a specified system, (ii) the underlying data models of the integrated libraries. Those classes have detailed information on how to create those models. For example, listing 2.1 illustrates the use of this pattern by showing how the method "buildNodes()" in the "NodeBuilder" class gets overriden in the classes "ClusterNodeBuilder" and "TMNodeBuilder". The first one knows how to build the data models for the Aduna Cluster Map library and the latter one for the Treemap Library

- the Singleton pattern [Cooper, 2000] is for instance used for the RegistryMgr to ensure it only has one instance and to provide a global point of accessing it.

- the Facade pattern [Cooper, 2000] as unified interface between the simple user interface and the system core.

- the Observer-Observable pattern [Cooper, 2000] is used to define a one-to-many dependency between objects. For instance, if an end user issues a keyword query in the simple user interface, the facade pattern is used to change a keyword filter in the "FilterMgr". The "QueryMgr" is an observer of the FilterMgr and therefore gets notified of the new query. It will store matching results after which the "VisualisationMgr" will automatically get notified of this update. The visualisation in the user interface will therefore automatically be updated as well.

Listing 2.1: Factory and Builder pattern for creating the underlying models of the integrated visualisation models

```
public class NodeBuilderFactory {
  public NodeBuilder getBuilder(){
    try {
      Class nodeBuilderClass =
        Class.forName(RegistryMgr.getInstance().getSystemProperties().
                      getVisualisationTechnique());
```

```
        return (NodeBuilder) nodeBuilderClass.newInstance();
      } catch (Exception e) {
        ...
      }
   }

public class NodeBuilder {
    public abstract void buildNodes();
}

public class ClusterNodeBuilder extends NodeBuilder {
      public void buildNodes() {
        //will create DefaultObject of the Aduna Cluster Map Library
        ...
    }
}

public class TMNodeBuilder extends NodeBuilder {
      public void buildNodes() {
        //will create TMNodes  of the Treemap Library
        ...
    }
}
```

### 2.5.3 Open Source Search Engine

The QueryMgr that was described in section 2.4.5 makes use of the Lucene search engine [Lucene, 2008] to optimise text search within the collection of reusable components. By using this third-party open source search engine library, written in Java, we boost the performance and the scalability of the framework which are important framework requirements. Lucene is one of many open source search engines like ASPSeek, BBDBot, Datapark, ebhath, Eureka, Indri, ISearch, IXE, Lucene, Managing Gigabytes (MG), etc. We chose Lucene for the following features:

- scalable, high-performance indexing with small RAM requirements.
- powerful, accurate and efficient search algorithms.

The authors of [Middleton and Baeza-Yates, 2008] compared the main features of 17 search engines, as well as the performance during the indexing and retrieval tasks with different document collections and several types of queries. The results of their comparison show that Lucene performs well on keeping both the index size and the RAM usage small. These are important advantages in the area of both performance and scalability which are important requirements of our framework (see section 2.3). Lucene has been used in many domains and has been proven to be scalable. Our research group itself has also good experiences with Lucene in

ARIADNE where Lucene has been used to optimise search for many projects like MACE, MELT, Prolearn, Acknowledge, etc.

### 2.5.4 Overview of the Integrated Visualisation Toolkits

An overview of the integrated visualisation toolkits and libraries is presented in table 2.1. Some of the libraries presented in the table contain more information visualisation techniques than shown in the table. Only those information visualisation techniques that are currently used in the framework are taken into account. Those techniques will be explained in detail in the subsequent chapters. The convenience of the integration of those libraries mainly depended on the availability of documentation of the data models that are needed for the libraries.

Table 2.1: Overview Information visualisation Libraries

| Name | Year | Programming Language | Information visualisation Techniques Used |
|---|---|---|---|
| TreeMap Java Library | 2001 | java | Squarified Cushion treemap |
| HyperTree Java Library | 2001 | java | Hyperbolic Tree |
| GraphViz | 2003 | C | Twopi graph layout |
| The Infovis Toolkit | 2003 | java | Node-link diagram<br><br>Magical lens interaction<br>Execentric Labeling |
| Aduna ClusterMap Library | 2002 | java | Clustermap visualisation |
| Prefuse | 2005 | java | Radial Tree Layout |
| Piccolo | 2000 | java, C# | Zoomable User Interface |

### 2.5.5 Overview of the Integrated Collections of Reusable Components

A the moment, nine collections of reusable components are integrated in the framework. These are presented in table 2.2. Each collection has a different access mechanism. The metadata for describing the reusable components are structured according to various metadata schemes.

Table 2.2: Overview of the Data Collections: those with a (*) are presented in the case studies in the subsequent chapters.

| Name | Access Mechanism | Metadata Format | Context |
|---|---|---|---|
| Ariadne KPS (*) | Webservices (SQI, OAI-PMH) | LOM | Sharing & reusing of learning objects. |
| ALOCOM repository (*) | Webservices (SQI) | LOM | Disaggregating & repurposing learning objects. |
| ZoEP repository | MySQL database | Unstandardized metadata | Digital movies for high education in the Netherlands. |
| Aristo Music repository | Java Interface connectors | Unstandardized metadata | Quality Control of metadata describing musical parameters. |
| Attention.XML | Webservices | Attention.XML | Master thesis about capturing Attention.XML medata of WinAmp and IRC. |
| Del.icio.us (*) | HTTP, JSON | Unstandardized metadata | Social bookmarks in del.icio.us system. |
| Calibrate Repository (*) | MySQL | Unstandardized metadata | Personal bookmarking & tagging tool for digital learning resources. |
| ED-MEDIA | Flat XML files | Unstandardized metadata | Proceedings of the ED-MEDIA conference. |
| GLOBE (*) | Webservices (SQI) | LOM | Distributed network of learning object repositories across the world. |

## 2.6 Evaluation

The core functionality of our framework is to effectively and efficiently map collections of reusable components into a visual form. To achieve this, Section 2.3 presented a number of framework requirements. Our framework has been evaluated by quality attributes that provide the means for measuring the fitness and suitability of a system [Bass et al., 2006]. This section describes those attributes, the tactics that are used for their achievement and their evaluation results. The framework requirements are measured by these attributes as well. Section 2.3 will discuss advantages and disadvantages of the design of our framework.

### 2.6.1 Extensibility

Extensibility is defined as the ability to add functionality to a system and to support easy integration of new software components, diverse software packages, etc [Bass et al., 2006]. Dynamic extensibility implies that functionality can be added to a system without impacting the rest of the system [Fielding, 2000]. The following tactics are used for the achievement this quality attribute:

- The architecture is structured to support library extensibility and data collection extensibility by introducing generic models for glyphs, data and classifications (sections 2.4.2 and 2.4.3). In this way, a one-to-many mapping is achieved instead of a many-to-many mapping between data collections and visualisation libraries.
- The visualisation pipeline is supported by the framework architecture. Section 2.4.7 illustrates this.
- The framework has a simple, developer-friendly API where well-known software design patterns are used throughout the framework (section 2.5.2), providing a low entry barrier for developers who are aware of these patterns.
- The framework supports the LTSC IEEE LOM standard [IEEE, 2002] and the OAI-PMH protocol [Sompel et al., 2004]. Therefore, any collection of reusable components that supports LOM and allows the harvesting of their metadata with this protocol, can be visualised by changing a configuration file.

Table 2.1 gives an overview of the libraries that are currently integrated into the framework which shows the library extensibility. The case studies, summarised in Table 2.2 show data collection extensibility. Not all of these case studies have been created by us. Our framework has been extended with the prefuse toolkit in a master thesis student project where a visualisation has been created for providing insight in captured attention metadata. The ZoEP repository is a repository of digital movies for use in high education in the Netherlands. Our framework has been used to create a tree-map visualisation on top of this repository.

### 2.6.2 Performance

Performance is all about timing. Events like interrupts, messages, requests from users, or the passage of time) occur and the system should react to it [Bass et al., 2006]. The following tactics are used to achieve this quality attribute:

- Lucene indexes are used to optimise the performance of issuing queries and filtering large collections of reusable components (see section 2.5.3).
- Well-tested third-party libraries for visualisation techniques are used for rendering the visual form of the collection of reusable components (see section 2.1).
- Well-known software design patterns are used to bind all the modules of the architecture together (see section 2.5.2).

### 2.6.3 Scalability

Scalability refers to the ability of the architecture to support large numbers of components, or interactions between among components, within an active configuration [Fielding, 2000].The following tactics are used to achieve this quality attributes:

- Lucene indexes are used to optimise the performance of issuing queries and filtering when the scale of the collections of reusable components increases (see section 2.5.3).
- Well-tested third-party libraries for visualisation techniques are used for rendering the visual form of the collection of reusable components (see section 2.1).
- Information visualisation techniques are used for handling scale, such as excentric labelling and fisheye techniques. Those techniques will be explained in detail in Chapter 4.

### 2.6.4 Functionality

Functionality is defined as the ability of the system to do the work for which it was intended [Bass et al., 2006]. The main task of the framework is to be able to map collections of reusable components onto a visual form with the help of information visualisation techniques. This visual form can help to amplify cognition and therefore enable insight in the contents of those collections. Performing this task requires that many or most of the components of the system work in a coordinated manner to complete the job. Therefore, if the components have not been assigned the correct responsibilities or have not been endowed with the correct facilities for coordinating with other components, the system will be unable to perform the required functionality. Section 2.4.7 shows how the software framework supports the visualisation pipeline and therefore achieves the core functionality.

### 2.6.5 Configurability

Configurability refers to post-deployment modification of components, or configurations of components, so that they are capable of using a new service or data element type [Fielding, 2000]. The framework can be configured by the use of property files. Those property files are used by the factories that have to return the right builders for creating the internal data and classification models for a specified system (section 2.5.2).
Listing 2.2 illustrates this by showing how the ClassificationFactory looks up the correct builder class for building a model for a specific classification by reading the properties of the system. Those properties are stored in a property file.

Listing 2.2: Returning a builder for creating the classification model

```
public ClassificationBuilder getBuilder(String classificationName){
try {
  Class classificationBuilderClass =
     Class.forName(RegistryMgr.getInstance().getSystemProperties().
     getClassificationClassName());

  return (ClassificationBuilder)
     ClassificationBuilderClass.newInstance();
} catch (Exception e) {
//...
}
}
```

## 2.7 Conclusions

The created software framework that is presented in this chapter, has merely been created as a tool that allows us to investigate the main challenge of this dissertation:

- how can we provide flexible, efficient and effective access to large collections of reusable components and hence, enable insight in the contents of such a collection with the help of information visualisation techniques.

This framework

- enables fast experimentation with (i) new visualisation techniques on the one hand, and with (ii) collections of reusable components on the other hand. Those components may be distributed over different domains and described with any possible metadata standard.
- is structured to support the visualisation pipeline or the computational process of converting a collection of reusable components into a visual form that users can interact with (see Section 2.4.7).
- supports the LTSC IEEE LOM standard [IEEE, 2002] and the OAI-PMH protocol [Sompel et al., 2004]. Therefore, any collection of reusable components that supports LOM and allows the harvesting of their metadata with this protocol, can be visualised by changing a configuration file (see Section 2.4.6).
- has been used in several case studies to create visualisations of different collections of reusable components (see Table 2.2).
- has several integrated information visualisation libraries and toolkits (see Table 2.1).

- has been evaluated by quality attributes that provide the means for measuring the fitness and suitability of a system (see Section 2.6).

The goal of this chapter has been to provide understanding of the design of this framework. The subsequent chapters of this dissertation will describe a number of case studies that used this framework to validate the hypothesis of this dissertation that information visualisation techniques can provide flexible, efficient and effective access to large collections of reusable components.

# Chapter 3

# Visualising a Learning Object Repository

## 3.1 Introduction

The main challenge that this dissertation addresses is how to provide flexible, efficient and effective access to large collections of reusable objects and enable insight in the contents of such a collection. In the previous chapter, we have discussed the software framework for information visualisation techniques that we have developed with the purpose to have a test bench to enable this research. In this chapter, we use this framework to investigate whether methodologies and technologies for information visualisation can be a solution for the challenge above by answering the following questions:

- how can we effectively and efficiently represent a collection of learning objects?
- how can we effectively and efficiently interact with visualisations of learning objects?

According to the Learning Object Metadata (LOM) standard, a learning object is "any entity, digital or non-digital, that may be used for learning, education or training" [IEEE, 2002]. This definition allows for an extremely wide variety of granularities. This means that a learning object could be a picture of the Mona Lisa, a document on the Mona Lisa (that includes the picture), a course module on da Vinci, a complete course on art history, or even a 4 year master curriculum on western culture [Duval and Hodgins, 2003]. Learning objects are stored in Learning Object Repositories (LORs), together with extensive metadata that describe them. A few examples of such LORs are the Ariadne Knowledge

Pool System (KPS) [Duval et al., 2001], Merlot [McMartin, 2004], EdNA [Ivanova, 2004], EducaNext [Quemada et al., 2004], etc.

A crucial problem is how to produce and deliver quality content for online learning experiences. Educational institutes all over the world offer access to their learning objects. By improving share and reuse of learning objects, the cost of creating them declines. Finding the most useful learning objects for a given learning task is not easy [Ternier et al., 2008]. This chapter investigates this problem and presents a solution, based on an exploratory search application that uses information visualisation techniques. We focus on opening-up a collection of learning objects by starting from a visualisation of their topic. In the subsequent chapters, we will focus on opening-up collections by focusing on relationships between topics, social data, etc.

We will start this chapter in section 3.2 with a description how people get access LORs nowadays. Section 3.3 presents requirements for a visual exploratory search application that enables access to LORs. Section 3.4 looks at learning object metadata in more detail. Three visualisation techniques are presented in section 3.5. Section 3.6 discusses the prototype application that is currently used in Ariadne. Two evaluations that have been conducted to validate the uses techniques, are discussed in section 3.7. Implementation notes are presented in section 3.8. Section 3.9 shows related work and section 3.10 concludes this chapter.

The contents of this chapter have been published in [Klerkx et al., 2004] and [Klerkx et al., 2005]. The prototype that has been created in this chapter, is nowadays used in ARIADNE and the ZoEp-proejct in the Netherlands. A customised version is currently used in the ARISTO music company [AristoMusic, 2008] where it is used to control the quality of metadata of songs.

## 3.2 Current Situation

### 3.2.1 Problem Statement

Learning objects (LOs) in repositories like the Ariadne KPS, are usually described by metadata schemes that conform to LOM, Dublin Core, MPEG, etc. Applications can take advantage of the structured nature of the metadata while performing queries on those repositories. In this way, they can select learning material that is most suited to the needs of end users like learners or teachers. An example of a learning object and its associated metadata is shown in Figure 3.1. Section 3.4 will elaborate further on the structure of LOM.

There are many information retrieval models [Manning et al., 2008]. A typical and traditional setting of an information retrieval task involves a user, a document collection, and an IR system [Kraaij, 2004]. Imagine for instance teachers who want to use relevant resources from the Ariadne KPS in one of their courses. They first express their information need as simple or advanced queries by filling out


Figure 3.1: An example of a Learning Object and its associated LOM metadata (From [Ternier et al., 2008])

electronic forms. These enable them to compose boolean combinations of search criteria. However, queries typed into search boxes are not effective enough to meet all the demands [Marchionini, 2006a]. The Ariadne KPS processes this query and returns a ranked list of learning material to the user. This list of documents is the result of a matching process, that compares each LO to the query. Teachers can evaluate the results in this list and, if necessary, reformulate the query to filter out some results or include some more. This process of formulating queries and evaluating the results can be lengthy and is rather time-consuming and user-unfriendly [Duval and Hodgins, 2003]. Another approach is to use directory or category search by browsing the learning objects by subject. However, directory-based search does not seem to offer increased relevance over keyword-based search and also takes longer [Dennis et al., 2002]. A third approach, dynamic taxonomies or faceted search, focuses on user-centered conceptual exploration and is gaining acceptance the last years [Sacco, 2000]. Faceted search adds the option to filter search results on predefined context aspects. This works well if users really know what they are searching for, but we think that it does not help enough to provide a thorough overview of the contents of the complete LOR.

### 3.2.2 Exploratory Search

In the traditional IR setting, users start their search for material with an empty list of results which the IR system tries to fill by matching the queries with the documents [Kraaij, 2004]. Human-computer information retrieval (HCIR) has emerged as the study of information retrieval techniques that brings human intelligence into the search process [Marchionini, 2006c]. The goal of HCIR is to integrate human and system interaction so that people are continuously engaged with meaningful information, thus incorporating understanding with finding [Marchionini, 2006b]. Marchionini suggests that one possible way of reaching this goal is exploratory search which engages people fully in the search process and puts them in continuous control. Therefore, researchers are devising highly interactive user interfaces [Marchionini, 2006a], [Ahlberg and Shneiderman, 1994] and [Shneiderman and Plaisant, 2004]. Those interfaces support the "Visual Information-Seeking Mantra": overview first, zoom and filter, then details on demand [Shneiderman, 1996]. This is a basic visual design guideline that has been rediscovered in many existing projects. Our application, to enable access to LOs, supports this alternative approach of exploratory search instead of the traditional access that most LORs support, by using information visualisation techniques.

A basic design cycle has the following ingredients: set up requirements, invent number of solutions, match the solutions against the requirements and finally select the best one [Van Wijk, 2006]. Therefore, we will start with setting up the requirements of our approach in the next section.

## 3.3 Requirements

The main requirement of our approach is that it should support exploratory search by using information visualisation techniques. Seven user tasks are described in [Shneiderman, 1996] which are at a high level of abstraction and which should be supported in the design of graphical user interfaces for exploratory search. Applied to the context of finding material in LORs, these tasks are the following:

- **Overview**: gain an overview of the entire contents of a LOR.
- **Zoom**: zoom in on LOs of interest.
- **Filter**: filter out uninteresting LOs.
- **Details-on-demand**: select a LO or a set of LOs and get details on them.
- **Relate**: gain an overview of possibly existing relationships among LOs.
- **History**: keep a history of user actions.
- **Extract**: allow extraction of sub-collections of LOs that can be used in e.g a course.

Exploratory search applications typically follow the "Visual Information-Seeking Mantra": overview first, zoom and filter, then details on demand. This mantra and the tasks above, form the basic requirements of our approach to access LORs.

## 3.4 Data

Visualisation should show the data: it should reveal what the data is about [Tufte, 2001]. In order to know which metadata elements may be used for enabling visual exploratory search, we must first analyse the structure of the Learning Object Metadata (LOM) that describes the objects. This section starts in 3.4.1 with an overview of the data type taxonomy for visualisation environments, that was first described in [Shneiderman, 1996]. Section 3.4.2 shows how the metadata elements of LOM fit into this taxonomy.

### 3.4.1 Data Type Taxonomy

A taxonomy of seven data types for visualisation environments is summarised below, together with a number of examples. For each data type, visualisation techniques and visualisation tasks have been designed ( [Shneiderman, 1996]). These may be used as design guidelines when building advanced visualisation applications.

- **1-dimensional** data: sequential lists, often text based. Examples are program listings, text documents, document search results, etc.
- **2-dimensional** data: planar or map data. Examples are Geographic Information Systems (GIS), building blueprints, etc.
- **3-dimensional** data: real-world objects. Examples are molecules, the human body, etc.
- **Temporal** data: timeline data. Examples are project management, video data, etc.
- **Multi-dimensional** data: data with more than three dimensions. Examples are a movie database, a personnel database of company, etc.
- **Tree** data: collections of items where each item has a unique parent but may have many siblings. Examples are the hierarchical ACM classification [Inc., 2008], directory tree structures, etc.
- **Network** data: nodes and links, where a node represents a data item, and a link represents a relationship between two nodes. Examples are links between web pages on the World Wide Web, citation analysis, etc.

This taxonomy of data types is well adopted [OLIVE, 1997] and therefore a good choice for deciding how the metadata elements of LOM can be visualised effectively.

### 3.4.2 Learning Object Metadata

The IEEE LTSC Learning Object Metadata (LOM) standard [IEEE, 2002] is the primary standard for the description of learning objects [Wiley, 2007]. Relevant attributes of learning objects to be described include: title, author, owner, format, interaction style, etc. The elements of LOM are organized into nine categories:

1. **General**: description of the learning object as a whole.
2. **Life Cycle**: history and current state of the learning object.
3. **Meta-Metadata**: information about the metadata instance.
4. **Technical**: technical requirements and characteristics.
5. **Educational**: educational and pedagogical characteristics.
6. **Rights**: intellectual property rights and conditions of use.
7. **Relation**: the relationship with other learning objects.
8. **Annotation**: comments on the educational use of the learning object.
9. **Classication**: relation to a particular classication system.

A goal of LOM is to enable sharing of descriptions of learning resources between resource discovery systems which should lead to a reduction in the cost of providing services based on high quality resource descriptions [Duval and Hodgins, 2003]. Creating a visualisation application starts by looking at the data [Card et al., 1999]: LOM contains many elements [Cardinaels, 2007] and therefore we have to make a choice which elements are most interesting to use for reaching our goal.

An application profile is a scheme that consists of LOM elements which is optimised for a particular local application [Duval et al., 2006]. The Ariadne application profile consists of a number of mandatory LOM elements. Those elements were thoroughly selected by a group of researchers within the Ariadne foundation. This is a European Association open to the world, for knowledge sharing and reuse. The mandatory elements have been found most interesting to effectively describe learning objects to allow users to share and reuse them. We therefore use this subset of LOM elements in our visualisation prototype to enable users to effectively find learning material and possibly reuse it.

The discussed subset is summarised below, together with the data taxonomy type of the previous section. For every data type in the taxonomy, a number of tasks and visualisation techniques are proposed in [Shneiderman, 1996] that can be useful when designing a visual exploratory search application. These techniques will be the fundament in the remainder of this chapter for choosing which visualisation techniques to employ. For each element in the subset, we have indicated the data type of the collection of different values for that specific element. The reason for this, is that we want to open up a collection of LOs that are all described by metadata instances. If we would for instance want to issue a query, based on the title of the objects, we will need to process all values of the "title"-element in all metadata instances of the repository. All the values of this element together can be seen as an ordered list of titles and therefore this element is of the 1-dimensional data type.

- Identifier: a globally unique identifier that identifies the LO.
  - 1-dimensional: e.g ordered list of identifiers, representing the search results of a query
- Title: name given to the learning object.
  - 1-dimensional: e.g ordered list of titles, representing the search results of a query.
- Language: the primary human language(s) within the object.
  - 1-dimensional: e.g list of languages, representing the different languages in a collection of LOs.
- Description[1]: a textual description of the content of the LO.
  - 1-dimensional: e.g ordered list of descriptions, representing all learning material within a course that consists of a number of LOs.
- Aggregation Level: granularity of the LO, e.g. course, lesson, etc.
  - tree data: e.g collection of LOs hierarchically divided by their aggregation level.
- Date: creation date of the LO.
  - temporal data: e.g timeline of a collection of LOs based on their creation date.
- Author(s)[2]: person(s) who created the LO.
  - Multi-dimensional data: e.g collection of authors of LOs, described with names, addresses, institutions, etc.

[1] This element is actually not mandatory in the Ariadne application profile, but if the element is not filled in, a combination of other elements can be used to function as a description of the object. For instance, a learning object that contains a picture of the Mona Lisa, could be described by its title and its hierarchical classification

[2] In the LOM specification, authors are captured as a combination of multiple elements under the parent "Contribute" (LOM element 2.3)

- Classification: category describing where the LO falls within particular classification system. A classification is represented as a number of taxonomic paths in the specific classification system. Each taxonomic path is an ordered list of taxons where each taxon in the path is a particular term within the taxonomy. An example of this will be presented in section 3.5.1.
  - tree data: e.g collection of LOs divided by a hierarchical classification scheme.
  - network data: e.g collection of LOs divided by a non-hierarchical classification scheme
- Rights: category describing the intellectual property rights and conditions of use for LOs.
  - multidimensional data: e.g collection of the rights of LOs, described with cost, rights, licenses, etc.

The Ariadne Knowledge Pool System is the source of data for our visualisation experiment that is described in this chapter. How we get access to this repository will be discussed in section 3.8.

## 3.5 Visual Overview of a LOR

The "Visual Information-Seeking Mantra" dictates: overview first, zoom and filter, then details on demand. This section focuses on the first step and will discuss a number of solutions for visually presenting an overview of the complete contents of a LOR. The second and third step are discussed in the following sections.

Students and teachers usually search for learning material on a particular subject or topic ( [Najjar, 2008a], [France et al., 1999]), by searching in classification or directory structures like e.g. the Google Directory [GoogleDirectory, 2008]. The classification is therefore a useful starting point when creating the visual overview of the repository. Clustering techniques can be used to make it easier to browse the resources in a LOR and to present them in an overview [Zamir and Etzioni, 1998]. Therefore, we will first describe the LOM classification element which we will use to cluster the LOs in the overview.

### 3.5.1 Starting Point: LOM Classification

Learning Object Metadata (LOM) contain a classification of the subject of the learning object. For example, the classification of an LO that visualises the algorithm of the "Towers of Hanoi" [Buneman and Levy, 1980] could be :

- Exact, Natural and Engineering sciences;

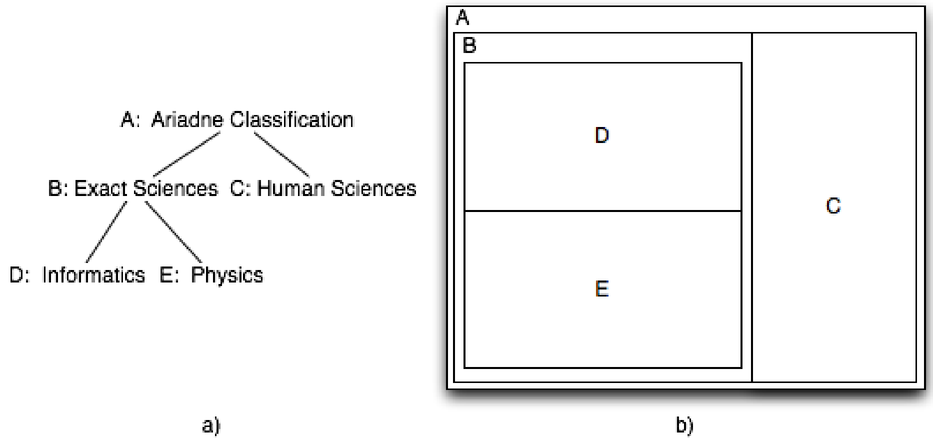



Figure 3.2: Example of the use of the tree-map algorithm. a) represents a small subset of the Ariadne Classification. b) is the mapping of the small subset in a) onto a tree-map.

- Informatics/Information Processing;
- General/Sundry;
- Recursion.

In the list of mandatory metadata elements in the Ariadne application profile of LOM, the classification element is of the tree data type. The tree data type represents a collection of items where each item has a link to one parent item (except the root) [Shneiderman, 1996]. We can thus consider the contents of a repository as a hierarchically structured classification of LOM instances. The basic tasks for tree data types are mostly related to structural properties of the tree. For example, find out how many levels there are in the classification tree. In the example above, the answer would be 4. Another typical task would be to see how many children an item has, or in other words, how many learning objects are classified with a certain classification. In the following sections, we will compare three well-known visualisation techniques for the tree data type that have been applied to the Ariadne KPS. These three techniques are squarified cushion tree-maps, hyperbolic trees and Venn-diagrams.

### 3.5.2 Squarified Cushion Tree-map

A tree-map is a visualisation of hierarchical structure that makes 100% use of the available display space. It maps the complete LOM classification hierarchy

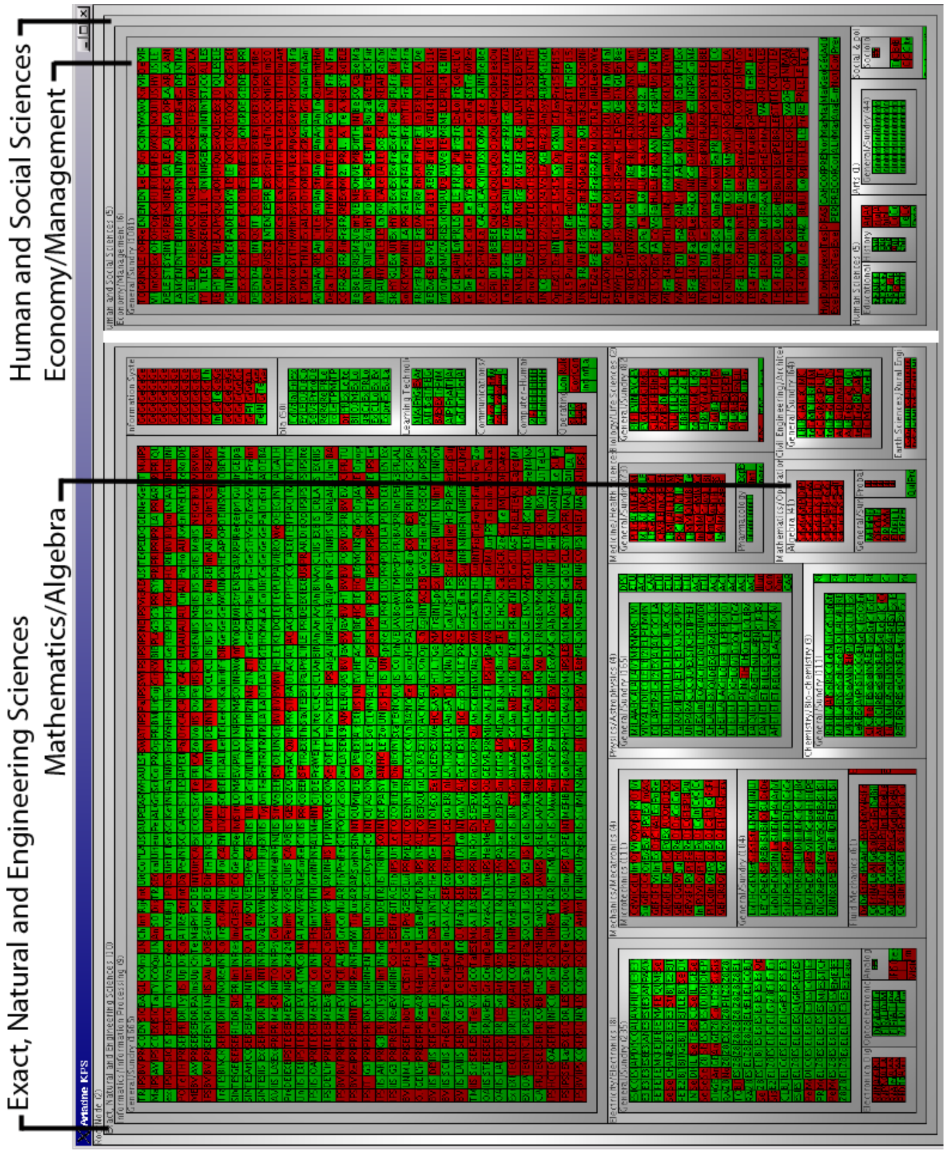



Figure 3.3: Squarified Cushion Tree-map of the Ariadne KPS. The left part consists of learning material about "Exact, Natural and Engineering Sciences", the right part about "Human and social sciences".

onto a rectangular region in a space-filling manner [Shneiderman and Johnson, 1991]. This is shown in Figure 3.2 where a small subset of the classification is mapped onto a tree-map. The size of a rectangle is proportional to the size of the corresponding nodes. In the figure, “C” gets less space allocated than “B”. The size of the rectangle represents the amount of LOs that have been classified by the respective classification taxon. Therefore, it means that there are more LOs in Ariadne, that have been classified by “Exact Sciences” than “Human Sciences”. Choosing this criteria for the size attribute is again that users usually search on topic. By seeing the size of a classification rectangle, they can immediately see if there are many materials of their interest available for them or not.

Several partitioning algorithms have been designed for tree-maps [Kobsa, 2004]. We chose the squarified tree-map algorithm that tries to make the nodes as square as possible, because it is hard for users to compare rectangles with different aspect-ratios. Cushion tree-maps inherit the elegance of standard tree-maps: compact, space-filling displays of hierarchical information, based on recursive subdivision of a rectangular image space. Intuitive shading is used to provide insight in the hierarchical structure [Wijk and van de Wetering, 1999].

The result of using this technique for creating a visual overview is shown in Figure 3.3. Visualising a document collection involves mapping data attributes to visual features like shape, size, orientation, etc. [Ware, 2004]. Parent rectangles represent a single taxon in a taxonomic path of the Ariadne classification scheme. Leaf rectangles represent an individual learning object within the KPS. The size of the taxon rectangles represent the amount of LOs that have been classified by the respective classification taxon. The size of the LO rectangles are all the same because users could otherwise get the false impression that a bigger LO is better in quality than a smaller LO but we did not have any real data available to support this impression. However, there is ongoing research to indicate relevant indicators of quality for learning [Duval, 2006] that may change this lack of quality data.

Colour is an important and frequently-used feature for information encoding [Healey, 1996a]. Well chosen colours allow for rapid and accurate identification of individual data elements by users [Healey, 1996b] [Few, 2004]. The “rights” element of LOM is colour encoded in Figure 3.3. Choosing colours for this element is rather straight-forward by looking at historical facts: “red” means “stop” and “green” means “go”. This has been known by almost everyone in the world since the convention on road signs and signals in 1968 [UNECE, 1968]. Therefore, red rectangles represent LOs that cannot be used without contacting the owner. The green ones represent objects that can be freely used. The choice of the “rights” element is illustrative and can be changed by any other metadata element with a fixed number of values. For instance, we could also choose to the language element to encode the colours like we will suggest in the next section.

Users should be able to interactively explore the contents of the LOR (section 3.2.2) so we enable them to zoom in on classification taxons of their interest by clicking on the corresponding rectangles. Tooltips, a standard way to visualise extra information, are provided with basic information about the LOs: title, author and identifier. This information can again be used to explore the available learning material.

**Observations**

Visualisation takes advantage of the immense power, bandwidth, and pattern recognition capabilities of the human visual system. It enables analysts to see large amounts of data in a single display, and to discover patterns, trends, and outliers within the data. Quite a number of patterns can be derived from looking at this visualisation of the Ariadne KPS:

- There are many more LOs on "Exact, Natural and Engineering Sciences", than there are on "Human and Social Sciences". One can deduct this by looking at the size of the corresponding rectangles (Figure 3.3). One level deeper in the classification hierarchy, we can see that half the material belongs to "Informatics/Information Processing". If a student with a background in computer science starts exploring this repository, he can thus immediately see that there is quite a good coverage of material of his interest in the LOR. We may also learn from this, that the authors that share material in the Ariadne KPS, are more exact sciences oriented than human sciences. This is for example interesting for researchers who want to find out if there is a correlation between the background of authors and their willingness of sharing their material [Gynn and Acker, 2003].

- The LOs within the "Exact, Natural and Engineering Sciences" classification are more scattered over sub-topics than the ones on "Human and Social Sciences". In "Human and Social Sciences", almost all LOs belong to the "Economy/Management" classification.

- 100% of the LOs of the topic of "Mathematics/Algebra" have restricted access. Further observation points out that all 64 objects are created by the same author.

- About 70% of the objects on "Human and Social Sciences" have restricted access while this is the case for about 50% of the objects on "Exact, Natural and Engineering Sciences". This insight could be used by researchers to investigate the hypothesis that users in exact sciences are more eager to share and reuse material than users from human sciences.

- In Figure 3.4, colour encoding is used for the representation of the languages of the LOs, instead of the "rights" element of the LOs before. Quite notable is that all objects on the topic of "Fluid Mechanics" are in magenta which

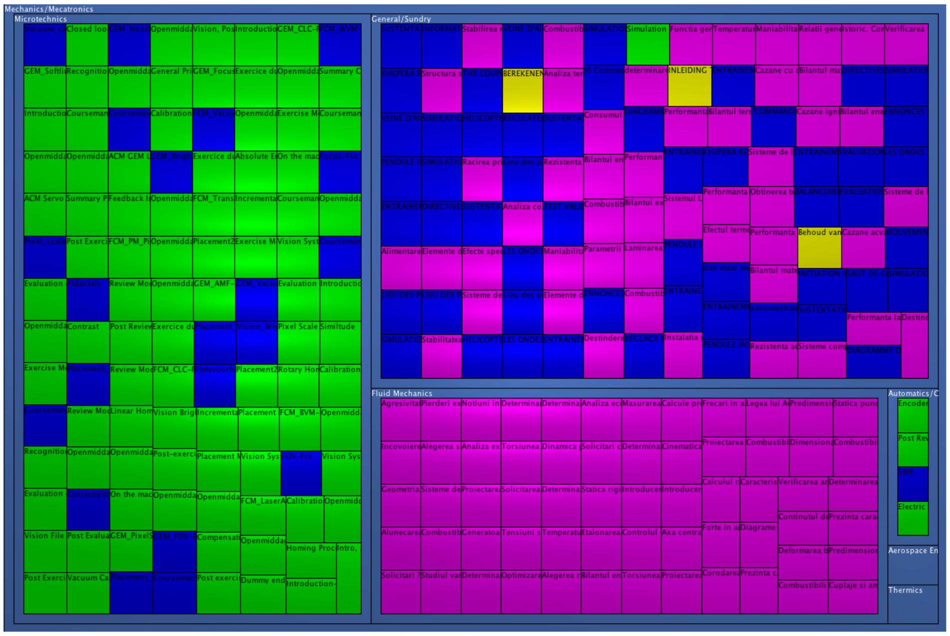



Figure 3.4: Tree-map of the "Mechanics/Mecatronics" classification where the colours represent the language of the objects.

represents the Romanian language. Further observation points out that all 62 LOs on this topic are created by the same author.

We can conclude from these observations that the tree-map visualisation can be used as a valuable presentation of an overview of a LOR. It enables users to compare nodes and sub-trees even at varying depth in the tree, and help them to spot patterns and exceptions. Therefore, this techniques is a valuable starting point in the exploration process on a particular subject or topic ( [Najjar, 2008a], [France et al., 1999]), by searching in classification or directory structures like e.g. the Google Directory [GoogleDirectory, 2008].

### 3.5.3 Hyperbolic Tree

The second visualisation that we investigated for creating a visual overview of LORs is the hyperbolic tree. This is a focus+context technique based on hyperbolic geometry [Moise, 1974] for visualising and manipulating large hierarchies [Lamping and Rao, 1996]. This technique blends in detailed views of particular parts of the LOM classification, together with a view of the overall structure of the complete classification. Users can navigate through the hyperbolic tree by clicking on the nodes of interest. By doing so, the user brings the selected nodes into focus at the centre. Those nodes at the centre get more space allocated. Transitions are animated to help maintain perceptual continuity when moving from one place to another. The classification elements have again been used to cluster the learning objects in the hyperbolic tree of Figure 3.5. By clicking on a node that represents a learning object, the user can view its complete metadata in his web browser.

**Observations**

If we look at Figure 3.5, we can derive a number of patterns about the Ariadne KPS:

- The space that is allocated to "Human and Social Sciences" is about the same size than the node of "Exact, Natural and Engineering Sciences'. At first glance, this gives the impression that both sciences are well represented in the Ariadne KPS. However, this is not the case. This can be seen by looking at dark areas at the side of the tree, which represent high densities of LOs. There are more of those areas, and thus more LOs, connected to the node of exact sciences than to the node of human sciences.

- It is difficult to estimate the number of LOs in the dark areas. Therefore, it is difficult and often wrong to make a visual comparison of those. Consider for instance the area at the bottom which consists of LOs on the topic of general Informatics. The black area is rather small but in fact it contains 1802 LOs. The area that consists of LOs on the topic of general economy, is a lot larger but it only contains 1124 LOS.

- Using the classification element of LOM is not enough when searching for specific individual LOs in a hyperbolic tree. The number of internal nodes was too small and as a result there are too many leaf-nodes with the same parent-node. This is shown in Figure 3.6 where the user brought the node into focus that represents the topic of general Informatics with 1802 LOs. This could be solved by

  - only showing those LOs that match against a simple query that has to be performed first. However, the visualisation would not show a complete overview of the repository anymore. This is exactly the opposite of what we wanted to accomplish.
  - adding extra levels of metadata elements to further cluster the learning objects. These elements could be chosen on the basis of the empirical analysis of the actual use of metadata in the Ariadne KPS that was performed in [Najjar et al., 2003]. We could for instance choose those elements that are most frequently used while performing queries. This will be further discussed in section 3.5.5.

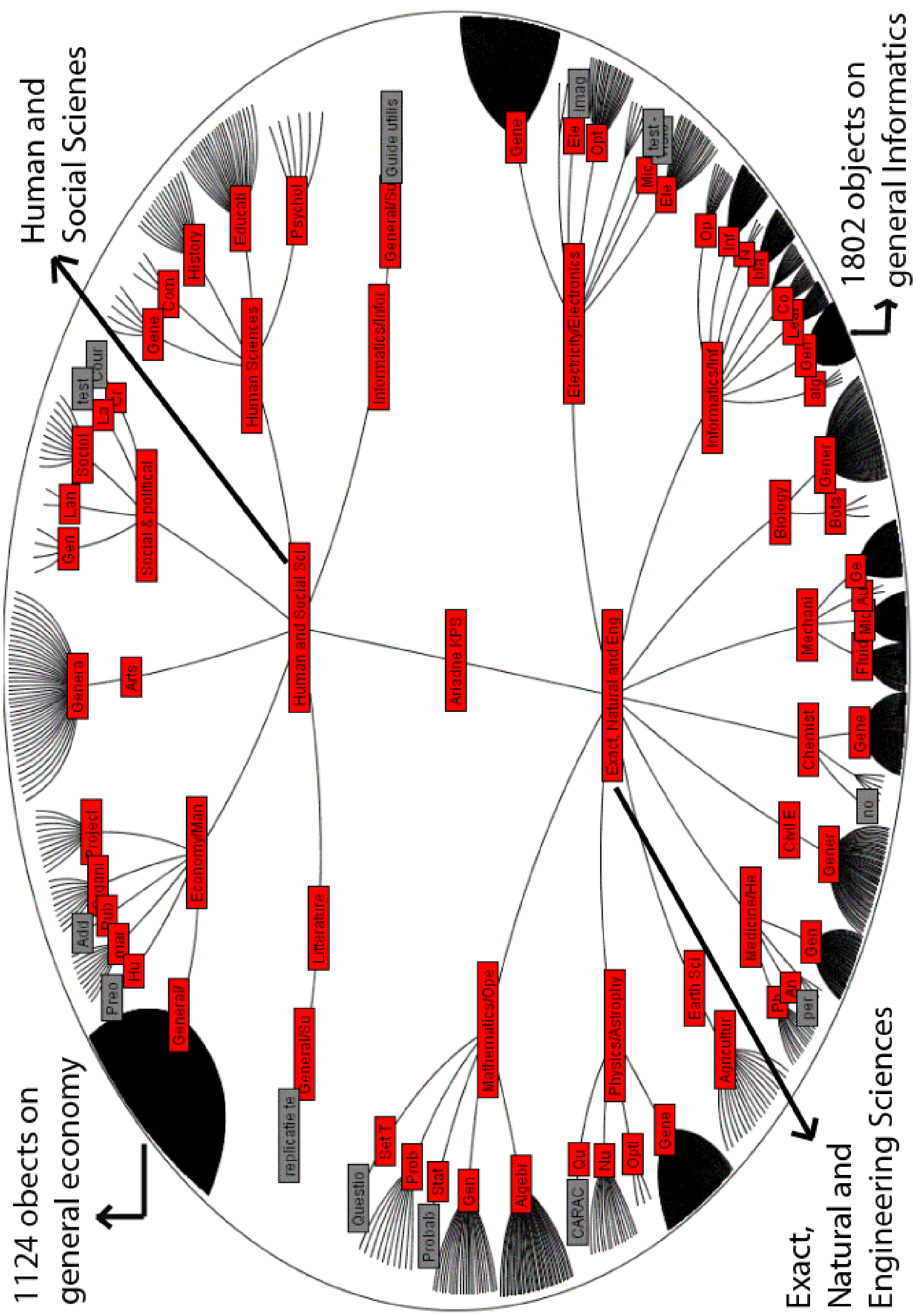



Figure 3.5: Hyperbolic tree of the learning objects in the Ariadne KPS

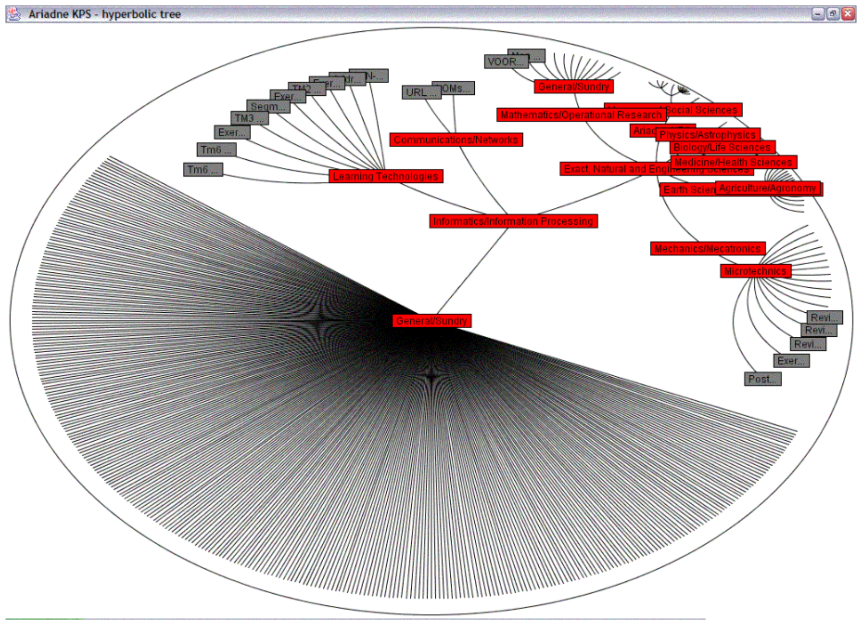



Figure 3.6: Hyperbolic tree: Too few internal nodes on one classification level. 1802 LOs are classified by the topic of general Informatics.

We can conclude from these observations that the hyperbolic tree visualisation can be used as a valuable presentation of an overview of a LOR. Section 3.5.5 will compare this visualisation with the tree-map visualisation to decide which of them has the most potential for enabling an effective overview of the LOR.

### 3.5.4 Venn Diagrams

A third alternative for the visual overview of the LOR is a Venn diagram visualisation. This is an illustration of the relationships between and among sets or groups of objects that share something in common [Venn, 1880], [Michard, 1982]. Grokker [Groxis, 2008] is a program that creates a graphical knowledge map. It makes use of non-intersecting Venn-diagrams to graphically present information. Navigating through the diagrams is possible by clicking on the set interest. Grokker will then zoom in on that particular set by using animated transitions. We have used this program to investigate this technique for visualising the LOM Classification of Ariadne. The result can be seen in Figure 3.7.

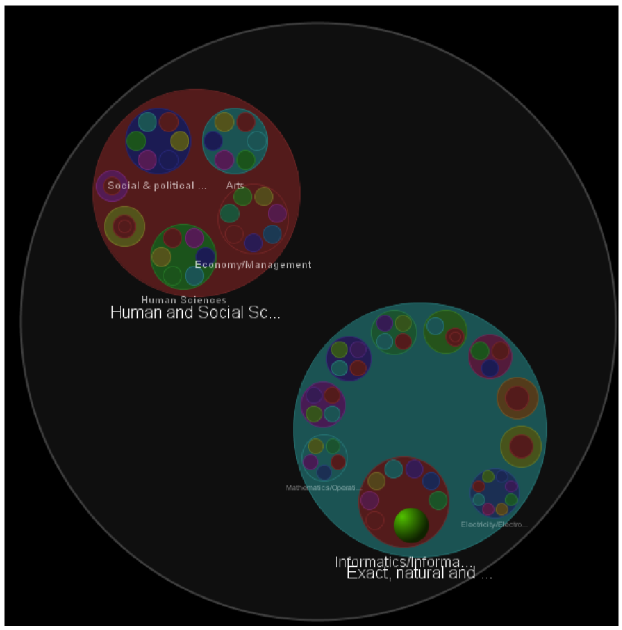



Figure 3.7: A Venn diagram of the Ariadne KPS

#### Observations

- Using the classification element of LOM is not enough when searching for specific individual LOs. There are too few inner nodes or in this case top-level sets: e.g the root node in the Ariadne classification has two children nodes: “Exact, Natural and Engineering Sciences” and “Human and Social Sciences”. This can be clearly seen in figure 3.8, where a lot of space is wasted. A solution for this, would be to e.g. draw half circles instead of complete circles. However, this would mean that some classification nodes would be represented by circles, and others by half circles. This geometric difference for representing the same type of item, i.e. classification nodes, could be confusing for users.

- Exploring the LOR can be done by clicking on the classification nodes of interest. A number of navigation steps are needed when navigating deeper into the classification. It is more difficult to keep an overview of the levels you already zoomed in. This can be seen in Figure 3.8 where a few navigation steps were undertaken to look at the learning material, classified with “Exact, natural and Engineering Sciences, Informatics/Information Processing, General/Sundry.” However, the only textual cue of the current zoom level, is “General/Sundy”. The parent nodes of this level should be remembered by the user who is currently exploring the LOR.

We can conclude from these observations that Venn diagrams have a number of disadvantages when presenting an overview of a LOR. The next section will compare all discussed visualisation techniques to decide which of them has the most potential for enabling an effective overview of the LOR.

### 3.5.5 Discussion

This section elaborates on a number of general observations that apply to all visualisation technique for opening-up the content of learning object repositories. They are summarised in Table 3.1.

#### Navigation

The visualisation technique should provide an overview of the entire LOR and allow users to navigate rapidly to any location in the LOR without getting lost.

- **Tree-map**: zoom in on any location in the LOR by clicking on the rectangles that represent the classifications nodes.

- **Hyperbolic tree**: browse through the tree by clicking on the nodes of interest. A number of navigation steps are needed when navigating deeper into the classification. However, intermediate navigation steps can be skipped

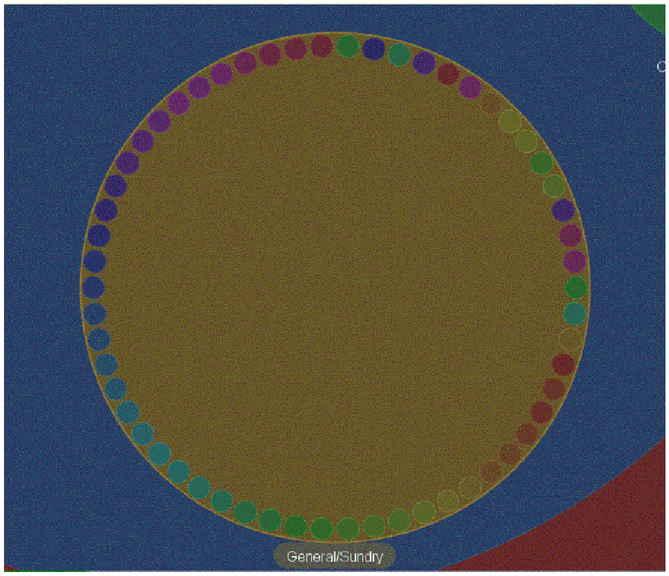



Figure 3.8: Learning Objects with classification: “Exact, natural and Engineering Sciences, Informatics/Information Processing, General/Sundry.”

by directly clicking on nodes that are positioned deeper in the classification. Because this visualisation is a focus+context technique, users will not get lost because a view of the overall structure of the complete classification is always visible.

- **Venn diagrams**: browse through the classification by clicking on the sets of interest. A number of navigation steps are needed when navigating deeper into the classification. However, intermediate navigation steps can be skipped by directly clicking on nodes that are positioned deeper in the classification.

**Space Usage**

LORs contain a vast amount of material that has to be visualised in the overview. Making maximum use of the available screen space is therefore an advantage.

- **Tree-map**: is a space-filling technique. Therefore, no space is lost in this kind of visualisation as it makes use of every available pixel.
- **Hyperbolic tree**: space will be dynamically reallocated based on the focus of the user. There will always be lost space between the classification nodes (cfr. Figure 3.5).

- **Venn diagrams**: depending on the situation, a large amount of space is wasted. For example, when using the LOM classification element as the starting point to cluster the results, almost 50% of the screen is lost. This can be clearly seen in Figure 3.7 where the main circles represent the nodes of human and exact sciences.

### Extra levels

The classification element has been used to cluster the contents of the repository in a hierarchical way. Adding extra deeper levels to this hierarchy should be easy without degrading the overview of the repository. It would for example be possible to add a level that represents the aggregation level of the LOs which defines the granularity of a LO and it could be one of the following: raw media element, course, lesson, etc. Users are then able to navigate to a specific topic and immediately know if there are complete courses on that topic or not. A disadvantage of this approach would be that those extra nodes would appear at multiple places in the hierarchy, which would complicate the structure of the classification hierarchy.

We could use most of the metadata elements that were described in section 3.4.1 as internal nodes of the hyperbolic tree. These elements could be chosen on the basis of the empirical analysis of the actual use of metadata in the Ariadne KPS that was performed in [Najjar et al., 2003]. We could for instance choose those elements that are most frequently used while performing queries. Another possibility is to bring human intelligence into the process [Marchionini, 2006c]: users should be able to select these metadata elements that seem most efficient at the moment of the query. Adaptive Hypermedia techniques could be used for the user so that they wont have to select it every time they access the LOR. Adaptation can assist the user in a navigational sense [Brusilovsky, 1996], which is particularly relevant in the case of visualising an overview of a LOR.

- **Tree-map**: easy to add extra levels to the hierarchy without losing orientation. However, the extra space that is needed to represent the extra levels, will not be available anymore for the objects itself. Zooming into a level deeper will solve this problem as more space comes available.

- **Hyperbolic Tree**: easy to add extra levels to the hierarchy. However, when too many metadata elements are added to the hyperbolic tree to cluster the LOs, the users overview degrades and it becomes easy to lose orientation. Users have to keep the top-levels of the tree in mind to make sense of the classification of the resources.

- **Venn diagram**: When completely zoomed in on a set of objects, it is difficult to keep an overview of the complete LOR. This is shown in Figure 3.8 where the user zoomed in on all learning objects with the classification:

Table 3.1: Overview advantages and disadvantages of the visualisation techniques for presenting an overview of the contents of LORs.

| Properties | Treemap | Hyperbolic Tree | Venn Diagram |
|---|---|---|---|
| Navigation | + | + | + |
| Space Usage | + | +/- | - |
| Extra Levels | + | - | - |
| Many LOs in one level | + | - | - |
| Colours | + | +/- | - |
| Visual Cues | +/- | +/- | +/- |
| Textual Information | + | + | + |
| User Studies | + | +/- | - |

"Exact, natural and Engineering Sciences, Informatics/Information Processing, General/Sundry.". However, only the textual cue "General/Sundry" is displayed. The parent nodes "Exact, natural and Engineering Sciences" and "Informatics/Information Processing" should be remembered by the user. Adding extra hierarchical cues could tell users which level they are in and how deep the structure is.

### Many LOs in one level

Many leaf-nodes within the same parent-node can occur. Consider for instance the case of the Ariadne KPS where 1802 resources share the same classification: "Exact, Natural and Engineering Sciences – Informatics/Information Processing – General/Sundry"[3].

- **Tree-map**: the size of a classification rectangle is proportional to the size of the corresponding nodes. Many LOs in one level would therefore be an advantage for general questions like "Does the LOR contain many material on biology or chemistry?" as the size of the classification node would enable users to quickly spot this information.

[3]Last checked: March 2008

- **Hyperbolic Tree**: cannot cope with this situation. An example of many LOs in one level is shown in Figure 3.6.
- **Venn diagram**: space is used insufficiently when there are many LOs in one level. This can be seen in figure 3.8.

### Visual Cues

A typical method of visualising a document collection involves mapping data attributes to visual features like shape, size, orientation, etc [Ware, 2004]. First of all, colours may be used for showing extra LOM metadata elements:

- **Tree-map**: easy to use.
- **Hyperbolic Tree**: easy to use. However, it may be difficult to distinguish colours of LOs when they are not in focus of the hyperbolic tree. This can be even worse when there are too many elements in one level of the tree where the objects cannot be drawn in the overview (Figure 3.6).
- **Venn diagram**: easy to use.

Secondly, other visual cues may be used like e.g using pictograms to show the mime-type of the LOs or country flags to show the language of the LOs. However, there is not much visual space available within each node in all three visualisation techniques, due to the vast amount of LOs that have to be presented in the visual overview of the repository. The visual space for each node is too small to add textual information like document titles, authors, publication dates, etc.. This can be solved a number of ways:

- The visualisation technique could be combined with extra dynamic query controls to manipulate the metadata. With these controls the user should be able to filter out many learning objects. Consider for instance the use of a data slider [Eick, 1994] to manipulate the metadata element “Language” that the user could use to filter all LOs but the ones in English. In this way the sizes of the different learning objects will increase and space will become available to add visual cues.
- Distortion techniques [Leung and Apperley, 1994] like fisheye-views [Furnas, 1999] could be used. These techniques magnify the area of interest while minimizing all other data, thus preserving global context. However, the author’s in [Shi et al., 2007] suggest that the multi-distortion technique is not always effective due to the high level of distractibility created by the animations.
- Magical lenses [Fishkin and Stone, 1995] could be used for manipulating metadata elements. Users would be able to move the magical lens over the

LOs in the visualisation. The LOs under the lens would then get the colour that represent the appropriate metadata element.

#### Textual Information

Textual information is needed for identification and extra information on the LOs like titles, authors, publication dates, etc. In all three visualisation techniques, this has been established by using a combination of text labels and tooltips.

#### User Studies

The study in [Kobsa, 2004] compared several information visualisation systems, among which the tree-map and the hyperbolic tree, for tree hierarchies in a between-subjects experiment. The aim of the experiment was to determine whether solving tasks in those system differs with respect to task completion times, accuracy and user satisfaction. Tree-map turned out to arguably the best visualisation system overall.

The authors of [Wang et al., 2006] performed a user study to find the effectiveness of tree visualisation methods in aiding knowledge discovery. They focussed on answering simple search or identification questions. The study shows that tree-maps are able to improve user's perception of some patterns file directory structures which is just another tree structure like the hierarchical LOM classification structure.

A user study on the Venn-diagram approach has been performed in [Rivadeneira and Bederson, 2003]. Most important result of that study has been that hierarchical cues are needed for users to tell them which level they are in and how deep the structure is. Without these cues, using this technique is not a good choice for presenting a visual overview of the repository as users quickly lose orientation.

Tree-maps are very effective when size is the most important feature to be displayed [Bruls et al., 2000]. This is the case in presenting the overview of the LOR as the size of the classification nodes enables users to quickly spot information on the available topics within the repository.

The results of these user studies are added as an evaluation property in Table 3.1, which summarises different properties that have been used to decide which visualisation algorithm to use as the starting point in our visualisation design. The next section presents a prototype application that has been created to validate the used visualisation techniques. Section 3.7 presents results of a user evaluation that was set up for this purpose.

## 3.6 Prototype

Based on the information in Table 3.1 , we chose the tree-map visualisation as the starting point in our application to enable exploratory search within LORs. How-

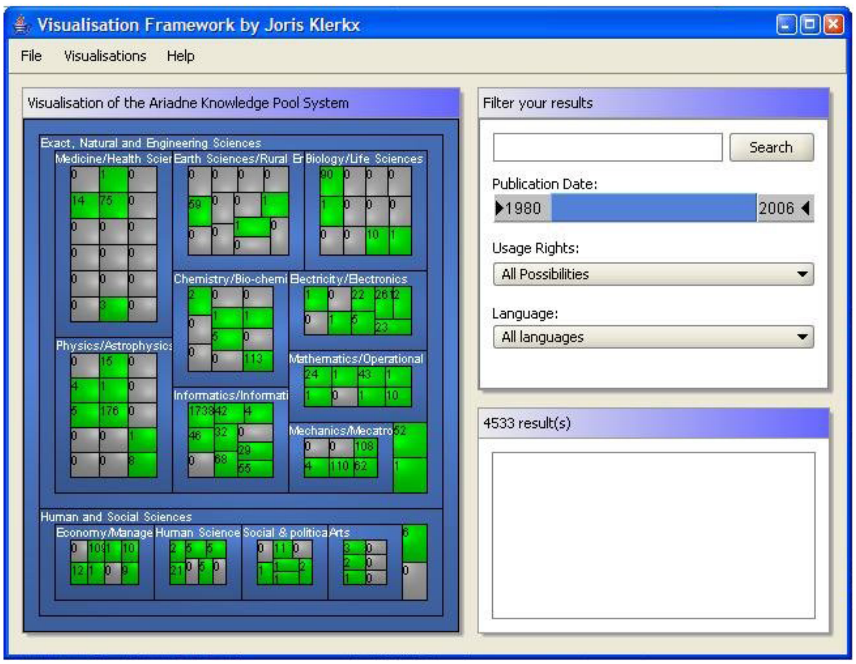



Figure 3.9: Prototype mockup for visual access to a LOR.

ever, the tree-map visualisation should be combined with methods to manipulate dynamic query controls over the metadata elements that are used to cluster the learning objects. This would enhance the usefulness of the visualisation [Kobsa, 2004] and will be described in the section 3.6.2.

Rather than creating a very advanced user interface for our prototype, we chose to create a very basic user interface with these 3 areas:

- a **visualisation** panel where the tree-map is shown that represents an overview of the learning resources in the Ariadne KPS,
- a **dynamic query control** panel where dynamic query controls are located for creating filters on the LOM metadata elements of the LOs, and
- a panel where textual **result** information is shown.

All these areas are co-ordinated which means that if one area is changed, the others are adapted as well. This basic interface is sufficient for the moment to do testing and validation of our visualisation design. However, It may be adapted afterwards if more advanced functionality would be needed. Figure 3.9 shows a screenshot of this interface. The following paragraphs explain the different areas in detail.

### 3.6.1 Visualisation Panel

In contrast to Figure 3.3, the prototype does not show the learning objects from the start. Only the amounts of learning objects that match a specific classification taxonomic path, are visualised (Figure 3.9) by displaying this number within the corresponding rectangles. We say that a LO matches a specific visual classification taxonomic path if two conditions are fulfilled:

1. The classification taxonomic path of the visualisation is the same as the one that can be found in the metadata of the resource.
2. The learning object matches the filter that is created with the dynamic query control panel.

Classification taxonpaths that do have matching LOs are green coloured. The green color was chosen because we wanted the matching resources to be strikingly visible. The green color is hard-wired into the human brain as a primary [Ware, 2004] and was therefore chosen for this purpose.

The approach of showing only numbers of matching LOs instead of the LOs itself, has been chosen as it is less influenced by dramatic increases of LOs in LORs. Therefore this approach is more scalable. However, users can zoom in on classification taxonpaths by clicking on the rectangles after which individual LOs are once again presented as in Figure 3.3.

### 3.6.2 Dynamic Query Control Panel

The dynamic query control pane consists of

- a search text box,
- a range slider for the publication date,
- a drop box for the usage rights of the LO, and
- a drop box for the language of the LO.

This set of metadata controls was chosen on the basis of the empirical analysis of the most frequently used metadata elements in searcher queries [Najjar et al., 2004]. With these controls, the user can quickly change the number of matching learning objects. Changing one of the controls invokes a dynamic update of the visualisation. The numbers of the matching LOs are updated and if some classification taxonpaths do not have anymore matching objects, the green colour is faded out. The result panel is updated as well and shows the total number of results and textual information about the individual matching objects.

### 3.6.3 Textual Result Information Panel

This panel shows detailed metadata about the LOs that are either selected in the tree-map visualisation, or the ones that match the filter, created by the dynamic query control panel. The metadata cover

- the title,
- the description,
- the author(s),
- the main concept in the classification,
- the location where the object itself can be found, and
- the location where the complete LOM metadata instance can be found.

As said before, users usually search on topic (see section 3.5). The elements above, allow users to get acquainted with the contents and thus, the topic of the learning object. If more information is needed, users can view the complete metadata instance that describes the learning object in the ARIADNE search application [Duval et al., 2001].

## 3.7 Discussion

In the previous section, we have presented a prototype application of our visualisation design to access learning object repositories. Effective visualisation guides users toward relevant information [Spoerri, 2004]. By means of describing a typical usage scenario in section 3.7.1, we will show how users of our prototype can effectively gain insight in the contents of a LOR by using the visual information-seeking mantra. Section 3.7.2 discusses an expert evaluation of our approach. Section 3.7.3 presents a user study with a complete prototype version of our visualisation that has been created to validate the used information visualisation techniques.

### 3.7.1 Typical Usage Scenario

Consider a biology teacher who wants to find free material in English about “trees”. The teacher can start by typing in “tree” in the search text box. The visualisation updates itself and only shows those classification paths that contain LOs that match against “tree”. The total number of results gets updated as well to 49 results which are also listed in the textual information panel. The teacher can either decide to go through the complete list or to filter some more results by changing the drop box for the usage rights to “free” and for the language to “English”. In the latter case, the visualisation updates itself and the total number of results drops to 40. Changing the range slider of the publication date from 2004 until now, results in 25 results that belong to the classification “Exact, Natural and Engineering Science; Earth Sciences/Rural Engineering; Agriculture, Agronomy”. Zooming in on that classification will make the teacher see all the individual LOs that match (Figure 3.10). If a LO is interesting enough, the full metadata instance can be viewed in the web browser. The web page of the full metadata instance is always the web page of the LOR itself and therefore up-to-date. The teacher may also decide to export a list of all the matching resources for future reference. In this scenario, the user is continuously engaged with meaningful information by integrating human and system interaction.

This example shows the use of the visual information-seeking mantra: overview first, zoom and filter, then details on demand. This mantra has been one of de main requirements of our design (section 3.3).

### 3.7.2 Evaluation 1

Seven user tasks have been described which should be supported in the design of an application that supports exploratory search (Section 3.3). This section presents the perception of an expert user how the prototype supports these tasks. There is a general acceptance that the feedback provided by such an evaluation method, is valid and useful [Nielsen, 1992b] for finding major usability issues in an early stage of development.

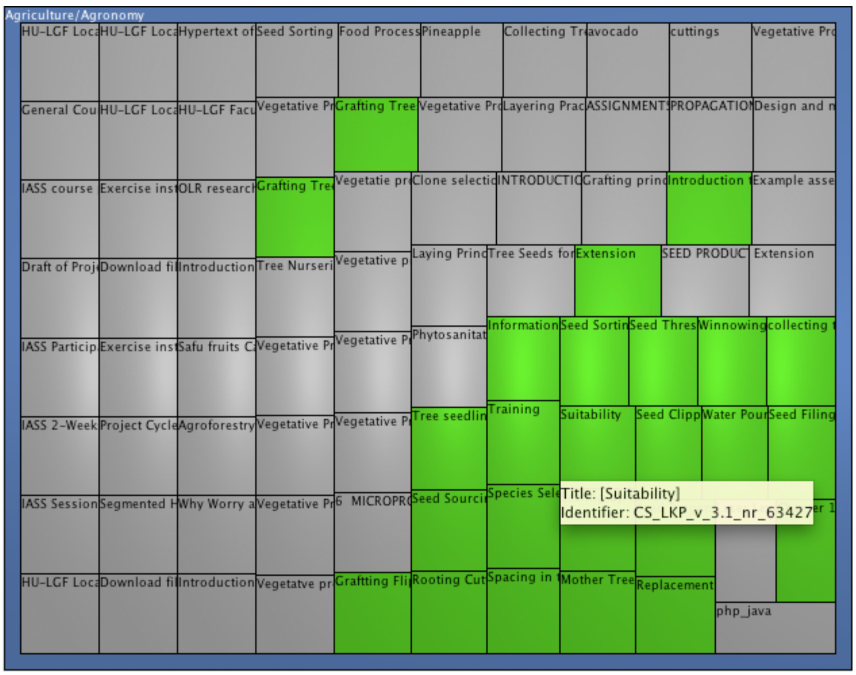



Figure 3.10: Matching Learning Objects in a Typical Usage Scenario(see 3.7.1).

- **Overview**: gain an overview of the entire contents of a LOR.
  - Interactive tree-map visualisation for representing the LOR (section 3.5) allows understanding of the contents of a LOR by being able to visually browse through the hierarchical classification of the learning objects in the repository.
- **Zoom**: zoom in on LOs of interest.
  - Zoom in on both taxonomic paths of the classification and individual LOs by clicking on elements in the tree-map (section 3.5.5).
  - Zoom in on individual LOs by creating filters in the dynamic query panel (3.6.2).
  - Find information of those LOs in (i) the result information panel and (ii) the visualisation panel (section 3.6.1);
- **Filter**: filter out uninteresting LOs.
  - Change the filters in the dynamic query panel to exclude uninteresting LOs from (i) the result information panel and (ii) the visualisation panel (section 3.6.2).
  - Filter out LOs by zooming on specific classification taxonomic paths and thus filtering out LOs on other topics (section 3.5.5).
  - Doing so, uninteresting LOs are excluded in (i) the result information panel and (ii) the visualisation panel (section 3.6.1).
- **Details-on-demand**: select a LO or a set of LOs and get details on them when needed.
  - While browsing through the interactive tree-map visualisation, information on specific LOs can be seen by (i) textual labels and (ii) the displayed information in the result information panel when one clicks on a node in the tree-map (Sections 3.6.1 and 3.6.3).
  - The metadata that is presented in the result information contains the location of the complete LOM metadata instance. This can be used to view all available details on the LO (Section 3.6.3).
- **Relate**: see possibly existing relationships among LOs.
  - Relationships between the topics of the LOs can be easily seen in the tree-map visualisation where all nodes in one classification node represent LOs on the same topic 3.5).

  - Creating filters in the dynamic query panel allows for seeing patterns in the visualisation. For example, if a keyword query is issued, LOs that match the same filter criteria are related to each other. They are not faded out in the visualisation and are presented in the result information panel (Sections 3.6.2 and 3.6.3).

- **History**: keep a history of user actions.

  - Logging facilities were not available at the time of the expert review of this prototype. However, these are added to our extensible framework for collections of reusable components for the case study that is described in Chapter 5. They are now also available for the prototype that is presented for this case study.

- **Extract**: allow extraction of sub-collections of LOs that can be used in e.g a course.

  - The result information panel shows created sub-collections of LOs. However, the individual results can only be downloaded one at the time from the repository. However, the implementation of downloading multiple LOs mainly depends on the constraints of the LOR. Many repositories only allow the harvesting of metadata. Harvesting content is often not allowed. This feature does not influence the use of information visualisation techniques for accessing LORs. Therefore, it has been moved to future work.

### 3.7.3 Evaluation 2

This section presents a user study that has been conducted to validate our design that enables effective and efficient visual access to a collection of learning objects.

#### Study Description

Our study compared two systems for getting access to the ARIADNE Knowledge Pool System. The first system was the ARIADNE search tool or "SILO". The second system was our prototype. Each session involved one participant, who performed the same set of tasks on both systems. There were 10 participants in this study. They were chosen to be representative of the intended user community, including research assistants, teachers and students. Each session involved one participant.

Each session started with an introduction of the context of the study. Besides that, we gave participants an introduction of tree-maps as it takes users about 15 minutes to get acquainted with them [Wang et al., 2006]. After these introductions, subjects were asked to complete a set of tasks with both tools.

The tasks for both tools were the same and were carefully selected to be a representative task for exploring the repository, such as the task "Find out the amount of objects on the topic of "java", that were created between 1990 and 1998." All tasks were chosen to have a similar difficulty level. Appendix A provides an overview of all tasks.

Two groups of 5 users were composed. Group A first performed the tasks with the ARIADNE search tool, and then with our visualisation tool. Group B first started with our visualisation tool and then with the ARIADNE search tool. All tasks were independent of each other.

#### Data Collection

During the sessions, subjects were asked to use the "think aloud" method, which is a method for user testing where the users verbalise their thoughts while using the system that is being tested [Nielsen, 1992a]. In this way, those users reveal their view on the system and possibly, their misconceptions. Participants were also asked to complete a questionnaire after the tasks. This questionnaire was adopted from a usability of the ARIADNE search tool in earlier work [Najjar et al., 2005].

#### Measurements

The following characteristics were measured for the experiment:

- **Task time** represents the time needed to finish each task. Comparing timings between the two tools allows us to investigate whether the use of the visualisation techniques can lead to savings in time. We know that time can be influenced by other factors, but this characteristic allows us to get a first indication of improvements for time-on-task.

- **Task accuracy** measures if users were able to complete the task at hands. It is calculated as the percentage of participants performing successfully [Rubin, 1994].

- **Satisfaction** of the users was assessed through a questionnaire filled in by each participant after finishing tasks. To measure the subjects agreement to those statements, we used the Likert scaling method [Likert, 1932] which is a popular scaling method. All items were rated on a 1-to-5 rating scale: strongly disagree - disagree - undecided - agree - strongly agree.

#### Results

- **Task Time**. Table 3.2 shows the average time participants spent on all tasks. At first sight, the difference between the two systems is rather limited: on average, 1,61 minutes were spent to complete a task with the ARIADNE

Table 3.2: Task Time and Task Accuracy

| TASKS SILO | Task Accuracy (%) | Mean Time (minutes) | Standard Deviation (minutes) |
|---|---|---|---|
| T1 | 100 | 1,57 | 0,24 |
| T2 | 100 | 2,28 | 0,15 |
| T3 | 100 | 1,21 | 0,16 |
| T4 | 60 | 1,31 | 0,12 |
| T5 | 100 | 1,74 | 0,32 |
| T6 | 60 | 1,9 | 0,11 |
| T7 | 60 | 1,27 | 0,16 |
| ALL TASKS | 82,86 | 1,61 | 0,18 |
| | | | |
| TASKS INFOVIS | Task Accuracy (%) | Mean Time (minutes) | Standard Deviation (minutes) |
| T1 | 60 | 2,58 | 0,24 |
| T2 | 100 | 1,57 | 0,36 |
| T3 | 100 | 1,59 | 0,38 |
| T4 | 100 | 0,79 | 0,12 |
| T5 | 80 | 1,14 | 0,08 |
| T6 | 100 | 1,56 | 0,36 |
| T7 | 60 | 2,19 | 0,14 |
| ALL TASKS | 85,71 | 1,63 | 0,24 |

Table 3.3: Satisfaction

| Statement | Mean | Standard Deviation |
|---|---|---|
| Ease of use visual tool | 3,56 | 1,01 |
| Performance visual tool | 3,78 | 0,97 |
| Performance trad. tool | 3,44 | 0,53 |
| In control of the search process | 3,89 | 0,78 |
| Clarity tree-map visualisation | 3,78 | 0,67 |
| Result lists easy to read | 3,44 | 0,53 |
| Would use a visual tool in future | 3,00 | 0,71 |
| Confidence correctness visual tool | 3,89 | 0,60 |
| Confidence correctness trad. tool | 3,44 | 0,53 |

search tool and 1,63 minutes with the visualisation tool. Solving the first task T1 with the visualisation tool took significantly more time (2,58 minutes) than the following tasks for all users. This can be explained as the fact that users need time to get acquainted to the tree-map visualisation.

Task T2, T4 and T5 were tasks that measured how quickly participants could find out statistics about the repository. For example, T4 was about finding out how much “English” material was available on “Informatics” in the repository. Solving those tasks with the visualisation tool took less time than with the ARIADNE search tool. This indicates that our visualisation design helps users to keep a good overview of the amount of matching results, when changing filters with the dynamic query control panel (section 3.6.2). However, we are aware that we need to include more participants to be sure of this.

Task T7 measured if users could easily find the difference in numbers between free and non-free material on the topic of ”human sciences”. Solving this with the visualisation tool (2,19 minutes) took longer than solving it with the traditional tool (1,27 minutes). We think this is related to the absence of a search box for classifications in the tree-map visualisation. Participants in the visualisation tool started to look for the correct classification node in the tree-map visualisation zoom. This took quite a while for most of them which explains the mean task time.

- **Task Accuracy**. Table 3.2 shows the percentage of users that completed a task correctly for both the visual and the traditional tool. The difference between the two systems is again rather limited: 82,86% for the traditional tool versus 85,71% for the visual tool. A task was considered completed incorrectly if (i) a wrong answer was given to a task and (ii) if the answer was not found in the most efficient way possible.

  For instance, T5 asked to find the amount of documents on “java” that were classified by “Informatics”. In the visualisation tool, two users issued a keyword query with the term “java” and started to browse through the result list while counting the ones with the correct classification. However, the most efficient way would have been to use the overview visualisation to zoom in on this classification. This would have filtered out all other results automatically. Therefore, T5 has a task accuracy of 80%.

  We can conclude that the visualisation tool is at least as effective as the traditional tool. However, there is no significant difference to conclude that one way is better than the other one.

- **Satisfaction**. Table 3.3 presents the responses of participants to questions concerning the overall use of the visualisation tool. The mean for the level of ease-of-use was higher than 3, meaning that the majority of participants found the visualisation application ease to use. The level of the clarity of the tree-map was also perceived as rather good (mean 3,78) which was better than expected as only 3 out of 10 participants used a tree-map visualisation before. Performance of both the traditional and visualisation tool was perceived as fast (resp. 3,44 and 3,78). Overall, we can conclude that users were satisfied by the use of information visualisation techniques for accessing the ARIADNE repository.

#### Findings and Recommendations

In this section, findings and recommendations of the participants are discussed.

- **Lack of Search Filter for Classifications**. Some participants remarked that including a search box where classifications could be filled in, is required for successful exploring of the tree-map visualisation. For instance, if a user is interested in material on history, he needs to be able to type "history" in e.g. a search box, after which the classification "history" should light up in the visualisation. This would save a lot of time.

- **Ordering of classification nodes**. The classification nodes are sometimes displayed in the visualisation at different locations. For example, if the user resizes the application window, the squarified tree-map algorithm recalculates the tree-map based on the available space. This makes it sometimes difficult for users to get a grip on what is visualised at which place. Other tree-map algorithms could be investigated to solve this. For example, the "slice and dice" tree-map algorithm is known to be stable in the ordering of nodes, but it has very high aspect ratios in turn which degrades the overview of the tree-map.

- **Zooming**. For the moment, zooming out can only be done by clicking with the right mouse-button on the classification node that a users wants to zoom out from. This was far from clear for the majority of the participants. We could solve this by adding buttons to zoom out or by showing context menus in the visualisation.

- **Rating**. If users look at results in detail, they should be able to indicate a rating of the object that should be available to all other users as well. This rating can be used as an interesting extra dynamic query control. To include this in the future, we can build further on other ARIADNE research.

- **Navigational cues**. Extra navigational cues can be used to inform users which elements have been visited already. This would save time while exploring the repository.

## 3.8 Implementation Notes

A Learning Object Repository (LOR) is a repository that manages learning objects and/or their metadata. The reference model for LORs in [Ternier, 2008] thus contains a metadata store and a content store. The metadata store is a functional component that is responsible for storing and managing metadata. Our visual exploratory prototype application has been created with the help of our framework (see Chapter 2) to enable access to this metadata store. Some implementation notes on this integration are listed below.

- The KPS-Client of the Ariadne system [Ternier and Duval, 2003] has been integrated in our framework. This is a java-library that uses the Ariadne Web Services (AWS) to send queries to the Ariadne system. Results are sent back in LOM-format that validate against the Ariadne Application Profile.
- AWS have been generalised and standardised [Ternier, 2008] to enable interoperability for searching in and publishing to LORs. Since this standardisation, the use of the KPS-client has been deprecated and a new connection to the Ariadne KPS had to be created by implementing the Simple Query Interface (SQI).
- More recently, the Ariadne KPS supports harvesting of LOs by using the OAI-PMH protocol [Sompel et al., 2004]. Our application has been extended to support OAI-PMH so that it is now possible to not only visualise the Ariadne KPS but every LOR that supports both harvesting wtih OAI-PMH and LOM.
- The starting point of our application is a visual overview of the KPS where all resources are represented. For creating this overview, we should issue queries for each taxon in the Ariadne classification that result in the identifiers of the LOs that are classified with it. Issuing these queries and parsing the results at runtime, takes too much time for end users. Therefore, this work had to be done beforehand.
- When something is changed in the dynamic query panel, each LO has to be compared to the query. The visualisation should respond within a 100 ms interaction loop if we want the user perceived performance to be any good [Shneiderman and Bederson, 2003]. As a consequence, we cannot use to issue queries to the KPS at runtime because of the amount of queries that are needed and the possible network delays and interruptions that can happen. Index facilities have been added as a solution. The framework can harvest all LOs and put them in a Lucene [Lucene, 2008] index. Those indexes are used at runtime to optimise the visualisation rendering process and the handling of user actions.

- The squarified cushion tree-map visualisation of the Ariadne KPS, has been created by integrating a tree-map library [TreeMap, 2003] into our framework.
- The hyperbolic tree algorithm has been created by integrating the hypertree-library [Hypertree, 2003].
- The Grokker application [Groxis, 2008] has been used to create the Venn diagram visualisation of thet Ariadne KPS. This application has not been integrated into the framework as it was not available as open source. Therefore, we first made an overview manually to be able to evaluate the visualisation.

## 3.9 Related Work

A few examples of Learning Objects Repositores are the Ariadne Knowledge Pool System (KPS) [Duval et al., 2001], Merlot [McMartin, 2004], EdNA [Ivanova, 2004], EducaNext [Quemada et al., 2004], etc. All of them provide a web application where users can share and reuse learning material.

The Baltimore Learning Community used a simplified starfield display that had a discrete grid to find educational resources [Shneiderman and Bederson, 2003]. This worked pretty well for in their experiment for a rather small database of 826 resources. However, this approach has not been evaluated on a large scale.

The authors of [Brusilovsky and Su, 2002] present an adaptive visualisation component of a distributed web-based adaptive educational system. This component focused on the visualisation of expression execution during program execution. It adapts the level of details in the visualisation to the level of student knowledge about these constructs.

The tree-map algorithm has been applied and evaluated in many applications [Shneiderman and Johnson, 1991], [Wijk and van de Wetering, 1999], [Bederson et al., 2002], [Wang et al., 2006], [Bruls et al., 2000] and [Shi et al., 2007]. This research has been taken into account for selecting tree-maps as overview visualisation (see Section 3.5.5) when visualising the contents of a learning object repository.

## 3.10 Conclusions

This chapter has presented the use of information visualisation techniques

- that offers users new ways to find content and information that could be of interest to them in a learning object repository (LOR).

The contributions of this chapter are the answers to the following question:

- how can we effectively and efficiently provide visual access to a collection of learning objects?

To answer this question, we created a tightly-coupled, visual exploratory search prototype that uses a tree map visualisation for representing learning objects (Section 3.6.1). Our prototype supports the "visual information-seeking mantra": overview first, zoom and filter, then details on demand.

Real data has been visualised by coupling our prototype to the ARIADNE Knowledge Pool System (Section 3.4). Two evaluations have been conducted to measure the effectiveness and the efficiency of our design (Section 3.7). From those evaluations, we have learned that the use of information visualisation techniques proves a useful alternative to more traditional ways of accessing LORs. Other possible learning object repositories can be used in the future.

The framework that we have created and presented in chapter 2 has been adapted while creating the prototype for the techniques above. More concretely, we integrated:

- a library with tree-map visualisation algorithms,
- a library with hyperbolic tree visualisation algorithms,
- support for the OAI-PMH protocol [Sompel et al., 2004] so we can easily visualise other LORs if they support LOM and OAI-PMH

These libraries are therefore available for all case studies that are developed with our framework.

The previous chapter described the framework that has been used to create prototype applications for the case studies in this dissertation. This chapter described how we enabled visual access to a collection of learning objects, based on their topic. The prototype that has been presented in this chapter, is publicly available at [Klerkx, 2008b] and can be freely used to explore the ARIADNE Knowledge Pool System.

# Chapter 4

# Visualising Actual Reuse of Reusable Components

## 4.1 Introduction

The previous chapter described how methodologies and technologies for information visualisation can be used to answer the following questions:

- how can we effectively and efficiently support visual access to a collection of learning objects?
- how can we effectively and efficiently interact with visualisations of learning objects?

As users typically search on topic, our visualisation techniques were based on the LOM classification of the learning objects. This chapter investigates how we can answer the questions above when there is no LOM classification available. In this case, we will use the relationships between the different components as a starting point to enable visual access to this collection. If this way of enabling access proves effective, it could be used as a complementary alternative to the proposed solutions from the previous chapter. To validate the use of our visualisation techniques for the challenge above, a prototype will be presented that has been created to enable access to a Learning Object Repository (LOR) with a large amount of small LO components. This repository has been created in the context of the research in ARIADNE on content models, that define learning object components and how they can be aggregated, so as to enable reuse and repurposing [Verbert et al., 2008].

Learning objects (LOs) are often very coarse-grained and difficult to reuse due to the fact that they are stored in a final presentation form. This static representation is not suitable for flexible content reuse, as the components cannot be easily accessed. In many cases, paragraphs, images or diagrams are assembled manually by copy and paste actions. In order to support this process in a more methodical way, the ALOCOM framework was developed to decompose composite learning objects, and make those components available for more flexible content reuse [Verbert et al., 2005] [Verbert et al., 2004]. This process results in a Learning Object Repository (LOR) with a large amount of small LO components. This LOR requires advanced support for

- gaining insight in the actual reuse of the different components,
- searching and finding relevant components.

This chapter is structured as follows: in the next section, we offer background information on ALOCOM, its objectives and the problems with current access techniques. A number of requirements for opening the view on the ALOCOM repository, are discussed in section 4.3. A visual overview of the repository is presented in section 4.4. This overview is used to develop, in order to validate the chosen visualisation techniques. This prototype is presented in section 4.5, and discussed and evaluated in section 4.6. Implementation notes are presented in section 4.7 and we conclude the chapter in section 4.8.

This chapter has been published in [Klerkx et al., 2006].

## 4.2 ALOCOM

The ALOCOM framework supports two processes: the disaggregation of learning objects into their components (text fragments, images, definitions, diagrams, tables, examples, audio and video sequenceq) as well as the automatic assembly of these components in real-world authoring tools. Those processes are explained in the next section. Section 4.2.2 presents the ALOCOM repository that contains all the disaggregated components. Understanding the contents of this repository and their reuse, can be a tedious task. This problem is discussed in section 4.2.3.

### 4.2.1 Disaggregation / Aggregation Process

- In the disaggregation process, a composite learning object is decomposed into its components. In the case of a slide presentation, the presentation is disaggregated into the individual slides and each slide is further decomposed into its images, diagrams, tables, text fragments, definitions, references, etc.. Metadata is added automatically to each learning object component and the component is stored in a repository [Cardinaels et al., 2005] [Verbert et al., 2005]. This process will be described in more detail in the next section.

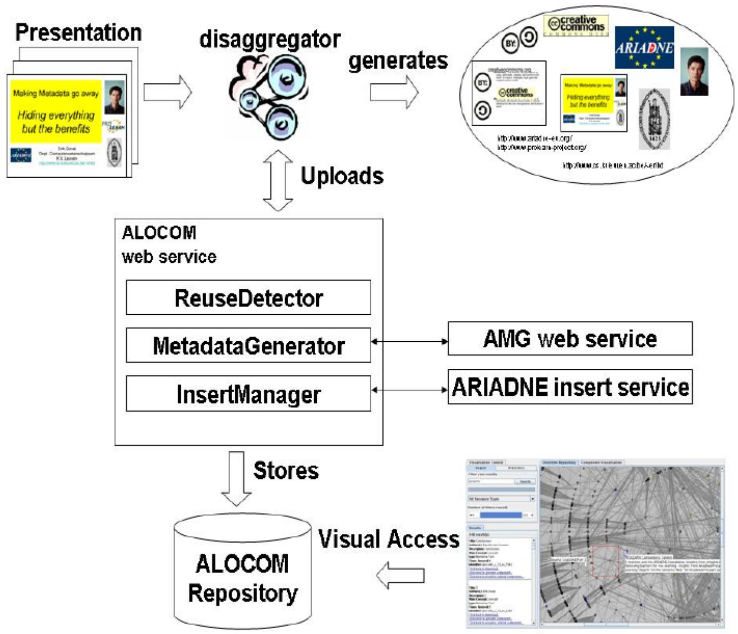



Figure 4.1: The ALOCOM framework.

- The aggregation process searches for components in the learning object repository and adds them to the learning object. Since end users prefer to use authoring environments they are familiar with, this functionality has to be integrated in ordinary authoring tools. Therefore, we have developed a component search plug-in for MS PowerPoint and MS Word in ARIADNE. A user can specify the type of component he or she is interested in (e.g. slide, image, text fragment, table, diagram), as well as keywords that best describe the component. All components that satisfy the specified search criteria are shown and the author can easily incorporate them into the learning object he or she is working on. For more details on the aggregation process, we refer to [Verbert et al., 2008].

### 4.2.2 The ALOCOM Repository

Disaggregating learning objects is a complex task that does not only involve extracting components from existing learning objects. Similarity measures are needed to detect reuse at the component level, and to avoid (near-) duplicates in the repository. For instance, it is necessary to detect if an image in presentation X is the same image that has been used in presentation Y. Furthermore, metadata need to be automatically added to each individual component, taking into account information from the original learning object to which the component belonged. Relationships between components need to be stored in order to keep track of reuse.

Our prototype supports decomposition for MS PowerPoint presentations (Figure 4.1). In a first step, a presentation is parsed in order to retrieve its content and structure. A .Net dis-aggregation client has been developed that extracts this information. Furthermore, the presentation is physically decomposed into its components. Each slide is stored in the MS PowerPoint format, images are extracted and text fragments, tables and diagrams are represented in a corresponding XML format.

In the next step, each individual component is uploaded to a disaggregation web service, which provides the following functionalities:

1. Reuse is detected using simple metrics that compute similarities between learning object components. The cosine similarity measure is used to detect overlaps in text fragments, hash functions are used to detect identical images, and a combination of these techniques is used for detecting similar slides, presentations, tables and diagrams [Verbert et al., 2008].

2. Metadata is generated for each component using the Automatic Metadata Generation framework (AMG) [Cardinaels et al., 2005].

3. Learning object components are stored in the ALOCOM repository using the ARIADNE insert service [Ternier and Duval, 2003]; an identifier is generated, isPartOf and hasPart relationships between components are stored and the identifier is returned to the disaggregation client.

In our case study, the ALOCOM repository was filled with 48286 components that were extracted from 653 presentations[1]. These components include 14113 slides, 5768 images, 198 tables, 26 diagrams and 27543 text fragments.

### 4.2.3 Problem Statement

The reuse is calculated as follows for each presentation: the number of reused components in the presentation is divided by the total number of components in that presentation. For instance, a presentation that contains 10 components, among which 2 images that are reused in other presentations, will have a 0.2 reuse value. The average reuse in this complete ALOCOM repository is 0.22. However, this number does not say anything about what is really going on in the repository. It does not provide answers to the following questions:

[1] Last checked: September 2007

- What is the distribution pattern of reuse in the repository? Are there few components that are reused many times which leads to the average reuse of 0.22, or are all components reused e.g one time?
- Which components are frequently reused? Why are they more popular than others?
- How are components put together in aggregate content?
- Why did reuse take place for some components?
- Which characteristics of the components contribute in making them reusable?
- etc.

These are the kind of questions that researchers have tried to answer for more than two decades [Collis, 1995], [Strijker, 2004] and [Verbert et al., 2008]. The goal of the case-study that is described in this chapter, is not to give answers to the questions above, but rather provide those researchers above a means to an end and improve their understanding.

The isPartOf and hasPart relationships between the components in the repository are of the network type (see chapter 3 - section 3.4.1). The following section provides detailed information about our use of information visualisation techniques to provide the means to improve understanding of both

- the structure of these relationships or the reuse within the repository, and
- the contents of the repository itself.

## 4.3 Requirements

To make the problem statement more concrete, we brainstormed together with a number of experts in the Technology Enhanced Learning (TEL) research field, to map it to a number of requirements for a tool that can be used as a means to solve it. This kind of tool should offer the following:

- An overview that shows researchers in a glimpse if there is much reuse in the repository or not.
- An overview of all objects in the repository, their reuse and how they are put together in aggregate content. This should enable the researcher to see patterns that are more difficult or impossible to see in statistical data like “the average reuse in the repository is 0.2”.

- Access to interesting objects by having the ability to interactively filter out objects of interest. It should e.g. be easy to find objects on a specific topic that have been reused, to be able to compare them and to find out why some objects are more reused than others.
- The possibility to easily see the structure of presentations and the interrelationships with other presentations.
- The possibility to cluster similar objects together and to save them for further reference and study. This is needed because discovery of information and patterns is seldom an instantaneous event. It often requires multiple times of studying and manipulating the data from different angles [Plaisant, 2004].
- The possibility to watch a preview of objects while exploring the repository. It should be easy to actually see a component in the repository instead of being forced to download the component first and view it with a separate application. Otherwise, a lot of time would be waisted by the end user.
- The possibility to compare two components; e.g compare two slides that have been reused a specified amount of times to see if there is any resemblance between those slides.
- etc.

We can conclude that creating a visual overview is very important because it enables researchers to improve their understanding of the relationships between the components in the repository. In a glimpse of the overview visualisation, researchers should be able to find interesting patterns in the underlying data. The next section will discuss two options for creating this overview.

## 4.4 Visual Overview of ALOCOM

In this section, we will elaborate on how we provide our target audience, researchers of "reuse" of content, the means to explore the content and the relationships between the components. More specifically, we make use of information visualisation techniques to create an exploratory search prototype to access the ALOCOM Repository. This prototype supports the Visual Information-Seeking Mantra "Overview first, zoom and filter, then details-on-demand" because this design guideline has been rediscovered in many existing projects (see chapter 3 - section 3.2.2).

### 4.4.1 Node-Link Graph

The isPartOf and hasPart relationships between the components in the repository are of the network type (see chapter 3 - section 3.4.1). Visualising this type of

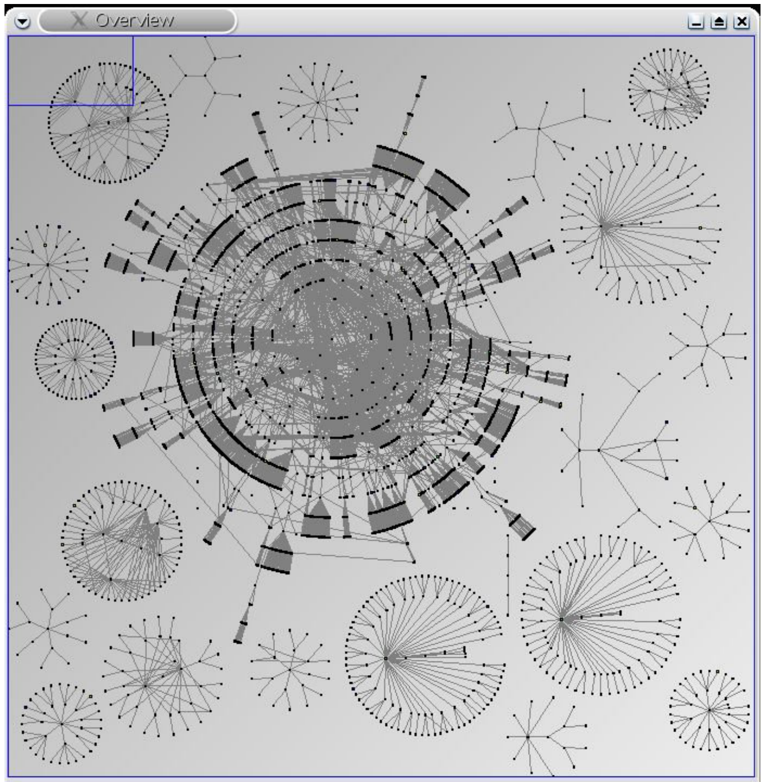



Figure 4.2: Node-Link visualisation of the ALOCOM repository.

| | image 1 | text fragment 1 | slide 1 | slide 2 |
|---|---|---|---|---|
| slide 1 | x | | | |
| slide 2 | x | x | | |
| presentation 1 | x | | x | |
| presentation 2 | x | x | | x |

Figure 4.3: Matrix presentation of 2 simple presentations with one slide.

data is already an old art but it is still an important research topic because of the complexity of relationships and user tasks [Shneiderman, 1996]. Representations of network data typically include the following:

1. Node and link diagrams, in all possible forms, shapes and different layouts; e.g 2D diagrams in circular layout (Figure4.3), 3D diagrams [Ware and Franck, 1994], etc.

2. A square matrix of the components where each component is assigned to one row and column. A link between two nodes "x" and "y" is represented by a value at the (x, y) and at the (y, x) position. Figure 4.3 shows a very simple example for two presentations, each consisting of one slide. Slide 1 contains only one image and belongs to presentation 1. Slide 2 contains a text fragment and an image and belongs to presentation 2. This figure also shows that image 1 is reused two times within this repository of two presentations.

From these options that have proven to work well in numerous projects, we chose the first one and created a node-link diagram visualisation to present an overview of the ALOCOM repository. We didn't pick the second option because of the following disadvantages:

- It is more difficult to grasp the meaning of the relationships when scaling up the amount of components that are visualised. Interpreting overview images like Figure 4.4 takes more cognitive overhead from the user.

- It is more difficult to see possible patterns within the data. Imagine a presentation where each separate slide is reused in other presentations. This is quite hard to find out in a matrix visualisation, but it is interesting information for a researcher because once found, this presentation can be compared with others that are not reused that often. Figure 4.9 shows that it is quite obvious to see this in a node and link visualisation.

### 4.4.2 A Node-Link Graph of ALOCOM

The overview of the repository is presented as a node-link graph, or a discrete structure, consisting of vertices and edges, where the vertices correspond to the objects and the edges correspond to the relations between the objects. Graph drawing is the task of the design, analysis, implementation, and evaluation of algorithms for automatically generating graph layouts that are easy to read and understand [Jünger and Mutzel, 2003]. Many layout algorithms and many state-of-the-art graph drawing software tools exist for many applications. For a thorough overview, we refer to [Nishizeki and Rahman, 2004] and [Jünger and Mutzel, 2003].

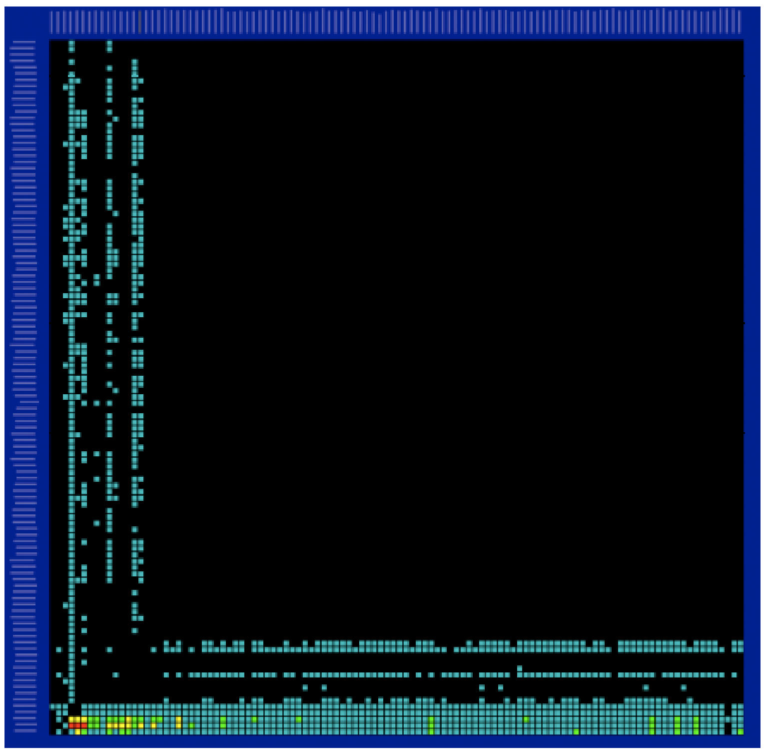


Figure 4.4: More complex matrix presentation with more connections between nodes [Becker et al., 1995]

The node-link diagram of the ALOCOM repository consists of a “twopi” drawing layout [Wills, 1999]. This algorithm tries to arrange all elements of the graph in circular or radial arcs, as can be clearly seen in Figure 4.2. The vertices in the graph represent the different components in the repository. Colours are used to represent the granularity of the components. The edges represent a “hasPart / isPartOf’ relation between the different components in the repository. An important criterion in many applications, is a small number of edge crossings in the node-link graph. Different kinds of layout algorithms are invented to achieve a minimum of crossings [Jünger and Mutzel, 2003]. However, one of the most important requirements in the case of the overview for the ALOCOM repository, is the possibility to see if there is a lot of reuse in the repository or not. If there would be little or no edge crossings at all, it would mean that there would be little or no reuse because the layout algorithm would be able to separate the nodes from each other. Figure 4.5 shows a number of presentations with little edge crossings and therefore little reuse of their disaggregated components. Figure 4.2 shows many crossings which means that there is much reuse in the repository. This static display of the overview is therefore a good choice for researchers that want to know in a glance if this repository is interesting for further study or not.

However, the static display can be cluttered with large amounts of data. Some applications solve these problems by aggregating or tresholding data. For instance, only show those components that have been reused a specified amount of times. Hence, not every component would be visualised and it would not be a complete overview anymore. This would make it impossible to assess the amount of reuse in the total repository. Imagine for instance a repository of 50.000+ components where only 10 slides are reused 10 times. With a specified threshold of 10, the user would only see a maximum of 100 vertices in that graph, but he does not have an idea what this means because of the lack of the amount of all components within the repository. On top of that, it is impossible to use this overview to compare those components that are reused, with the ones that are not because they are nowhere to be found. To solve the clutter-problem, we use direct-manipulation techniques that permit the graphs to continue to reveal relationships in the context of much more data. These techniques are discussed in the next section.

## 4.5 Prototype

The node-link diagram is the starting point for the exploration of the ALOCOM repository. This visualisation has been combined with methods to manipulate dynamic query controls over the metadata elements that are used to cluster the learning objects. This enhances the usefulness of the visualisation [Kobsa, 2004] for researchers.

The following section will elaborate on the user interface of the created prototype and all the available functionality that is supported in our prototype ap-

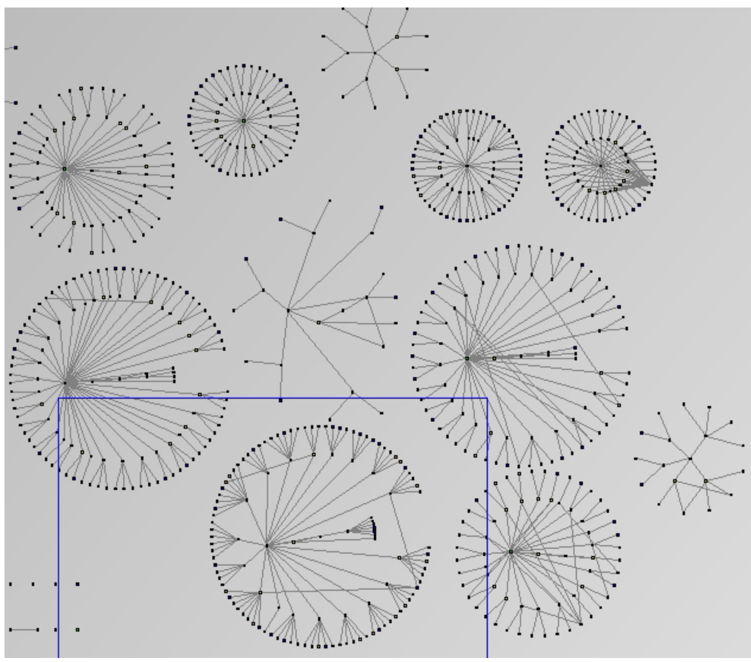


Figure 4.5: Node-link visualisation of a number of presentations with little reuse of components between the presentations

plication. Section 4.6.1 discusses how the defined problems (section 4.2.3) can be solved with this prototype, by means of describing typical usage scenarios.

Rather than creating a very advanced user interface for our prototype, we chose to reuse the very basic user interface that we have also used in chapter 3 with these 4 co-ordinated areas:

- A **visualisation** panel with the node-link diagram, presenting the “hasPart / isPartOf” relationships between the components;
- A **dynamic query control** panel where dynamic query controls are located for creating filters on the LOM metadata elements of the different components; and
- A textual **result** information panel;
- A **preview** panel, which displays a selected component.

The combination of these areas proved to be effective in the case of visualising a Learning Object Repository (chapter 3). This interface is therefore a good starting point but it may be adapted afterwards if more advanced functionality is needed. Figure 4.6 shows a screenshot of this interface. The following paragraphs explain the different areas in detail. All these areas are co-ordinated which means that if one area is changed, the others are adapted as well.

### 4.5.1 Visualisation Panel

The node-link graph within this panel shows the “hasPart / isPartOf’ relations between the different components in the repository. Colors are used to encode the resourcestype of the components. For example, the green color is used to encode slides or presentations, yellow to encode an individual slide, orange for figures, etc. In total, there are six resource types and thus six different colors. Users should be able to distinguish them while exploring the visualisation, according to [Miller, 1956] who showed that 7 ± 2 is the maximum number of colors that humans can work with. However, not all humans can distinguish the same set of colors [Bergman and Johnson, 1995]. Therefore, we made sure that users can change the default set of colors that are used to represent the different resource types.

Interaction with our node-link visualisation is necessary because

- We integrate human and system interaction so that people are continuously engaged with meaningful information [Marchionini, 2006b]
- The static overview display of the repository can be cluttered. The use of direct-manipulation techniques is a means to solve this (section 4.4.2).

Interaction is therefore supported in a number of ways:

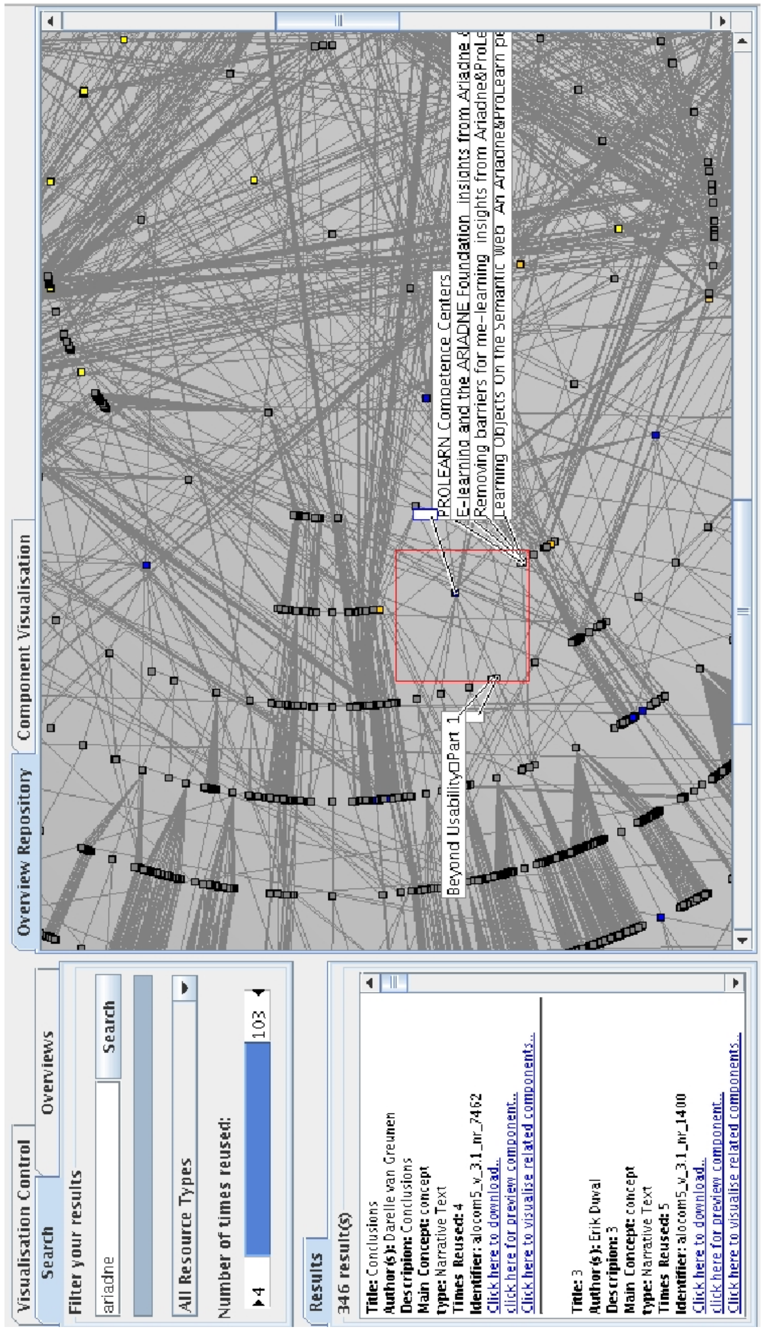



Figure 4.6: Visual access to the ALOCOM repository.

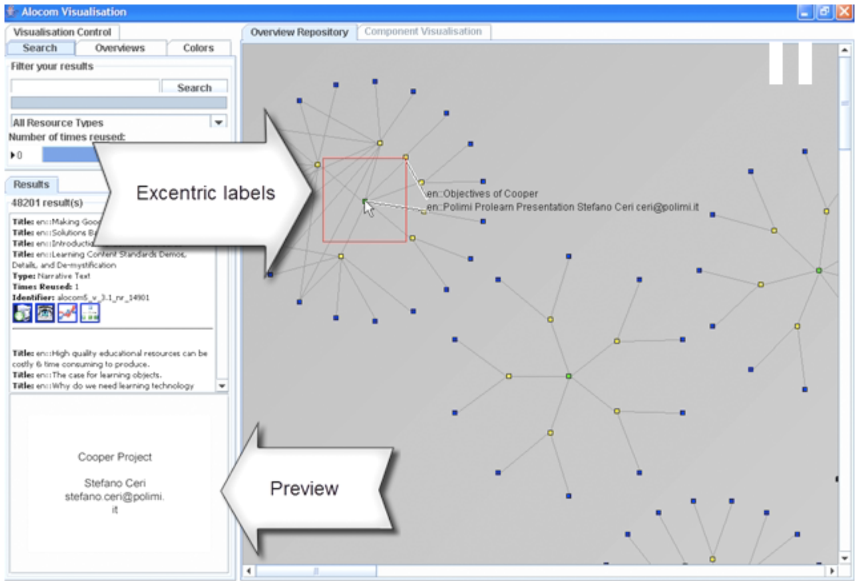



Figure 4.7: Excentric Labeling and Preview panel

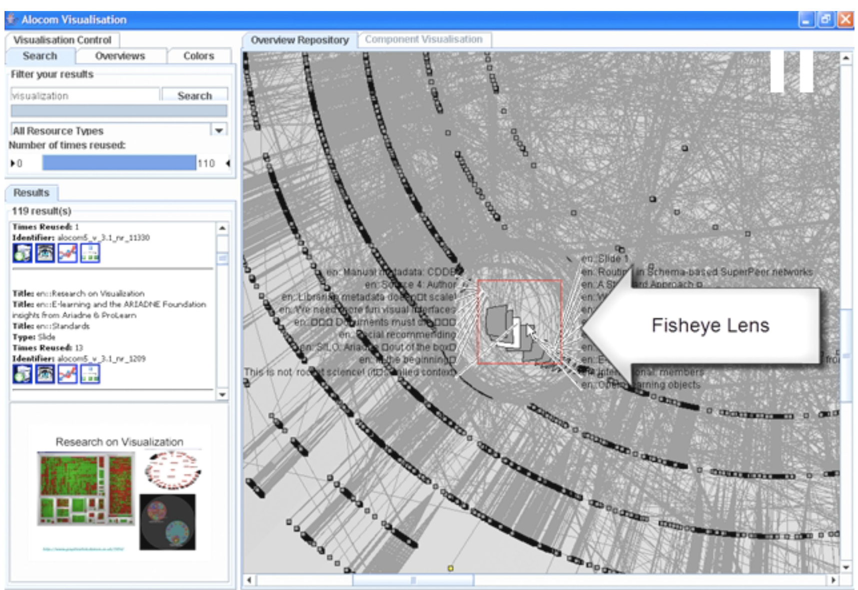



Figure 4.8: Fisheye View

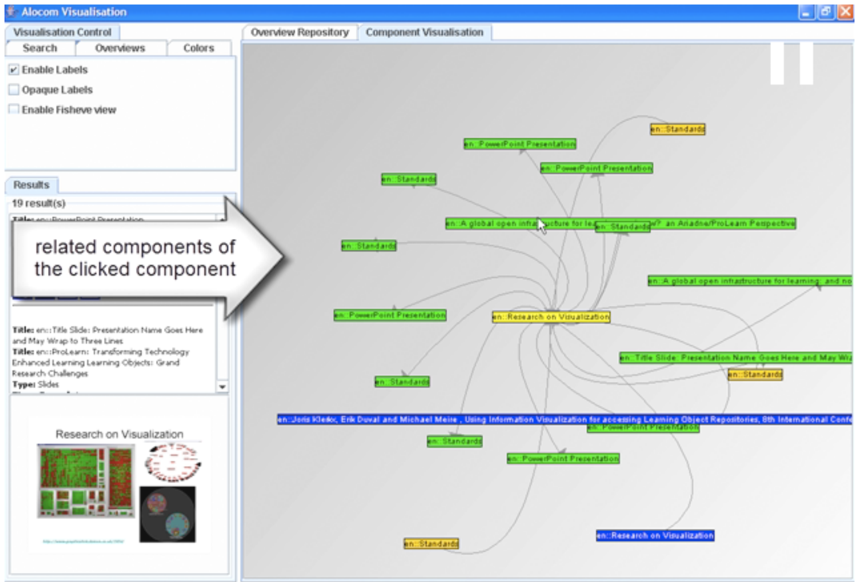



Figure 4.9: Detailed view of a slide with an image and its reuse in different presentations.

1. First of all, one can **pan** through the visualisation by scrolling through it, or by using the overview-control. This shows a zoomed-out view of the visualisation and a rectangle indicating the part that is currently visible. This rectangle can be moved to pan the visualisation. Panning allows navigation through large information spaces via direct manipulation. Exploring the repository by panning through the node-link visualisation is therefore an effective way of understanding the relationships between all components due to the natural spatial abilities of people that increases user's intuitive access to information [Bederson et al., 1996]. Besides that, it allows to easily find interesting patterns of reuse.

2. Secondly, the twopi layout that we have used to create a node-link overview of ALOCOM, tries to place related nodes near each other. However, this makes it more difficult to distinguish between those nodes. To solve this, we have combined two information visualisation techniques:

   - The first one is the excentric labeling technique, which labels a neighbourhood of components that are located around the cursor [Fekete and Plaisant, 1999]. Figure 4.7 shows a textual label for each node that is situated within the rectangle that is drawn around the cursor.

- The second one is the fisheye view [Furnas, 1999] which is a distortion technique that presents the data at the focal point of the cursor much larger than the other data around that point. Figure 4.8 presents an example of this.

3. Although the methods above help users to explore the ALOCOM repository quite effectively, they are not always sufficient enough to limit the clutter in the overview visualisation. We know from research that users often explore information by following the visual information-seeking mantra “Overview first, zoom and filter, then details-on-demand” [Shneiderman, 1996]. As we explained in section 4.5.1, users may discover patterns in the overview visualisation. Once they have identified these patterns, they should be able to zoom in on them. Therefore, we implemented the functionality that users can click on specific nodes in the graph visualisation. This will lead the system to visualise only the clicked node and its related components in a new visualisation. This new visualisation will have far less or even no clutter at all, so that it will be easier to understand the reuse of the clicked component in other components and in the composition of the clicked component itself. Figure 4.9 presents an example of this.

Indirectly, the user can also interact with the visualisation by using the dynamic query controls that are described in the next section.

### 4.5.2 Dynamic Query Controls

With the filter controls, one can find different components of interest in an interactive and flexible way. There are a number of filter methods available to the user. With these methods, the user can:

- enter keywords.
- select resource types from a list that includes slides, slide, diagram, figure, table and narrative text.
- select an interval that represents a number of times that a component was reused in the repository.

By combining these filter methods, an end user can easily find figures about trees that are 5 times reused. The filter methods all work together with an AND-operator. Components that do not match the filter criteria, are faded out in the visualisation. Components that do match, are highlighted. Moreover, metadata on the matching components are shown in the results information part of the interface.

### 4.5.3 Textual Result Information Panel

This panel shows detailed metadata about the components that are either selected in the visualisation, or the ones that match the filter, created by the dynamic query control panel. The metadata cover

- the title,
- the description,
- the author(s),
- the main concept,
- the location where the object itself can be found.

These chosen characteristics allow the user to get basic knowledge about the components. The user is then able to select a result and either (i) show it in the visualisation where related components can be shown, (ii) see a complete metadata instance about it, (iii) see a preview of the selected component, and (iv) download and possibly re-use it.

### 4.5.4 Preview Panel

One requirement (section 4.3) of this prototype, is the ability to easily and quickly compare components. For example, researchers would typically compare two slides that have been reused a specified amount of times to see if there is any resemblance between those slides. The preview panel enables a user to actually see the component he is interested and can also be used to compare multiple components (Figure 4.7). A preview of a component is immediately shown when a user is in progress of exploring the visualisation if he hovers the cursor over a component in the visualisation. This is not sufficient when a user is evaluating the components in the results list. Therefore he can also choose to view a preview by clicking on a button in the result information panel.

## 4.6 Discussion

In the previous section, we presented a prototype application that we created as a means to improve the understanding of reuse patterns within the ALOCOM repository. The hallmark of an effective visualisation is that it guides users toward relevant information [Spoerri, 2004]. By means of describing a typical usage scenario in section 4.6.1, we will show how users of our prototype can effectively gain insight in those patterns. Section 4.6.2 discusses an expert evaluation of our approach.

### 4.6.1 Typical Usage Scenario

Imagine a researcher that wants to investigate the ALOCOM repository and wants an answer to the questions in section 4.2.3. He can use the designed prototype to find them in the following way.

**Is there reuse in this repository?**

The node-link visualisation shows if this is the case or not in one glance. Figure 4.2 shows a lot of reuse because of the presence of the big circle in the middle.

**What is the distribution pattern of reuse in the repository?**

Researchers want to know if there are few components that are reused many times, or rather if all components are reused e.g one time. The node-link visualisation immediately gives an impression about those patterns. The big circle in Figure 4.2 with many components displays substantial reuse between the different components in this circle. The clutter of edges between the components means that there is a lot of reuse. On the sides of the graph, there is less clutter. Those components are not as much reused as the ones in the big circle. A researcher may therefore decide to further investigate the components in the big circle and see why they are more reused than the others.

**Which components are frequently reused?**

One can use the node-link visualisation as we described above, but the dynamic query controls can be used for this as well. One can use them to select an interval that represents a number of times that a component was reused in the repository. If we would change this interval to [10, 20], it means that only those components are shown in the visualisation panel that are reused at least 10 times and at most 20 times. All other components are faded out in the visualisation. We use tightly-coupled, co-ordinate views of the data so besides the visualisation area, the textual result information panel is adapted as well. It will only show those components that match the selected interval. Tight coupling eliminates the strict distinction between query and result displays. Both views on the data displays can be used to filter the data and easily find the components that are frequently reused.

**Why are some components more popular or more reused than others?**

Once the components are found that are reused many times with the help of either the node-link visualisation or the dynamic query panel, one can investigate the reasons by

- using the preview panel to actually see those components and study them. For instance, if most of the slides components that are frequently reused are composed of limited text fragments, one might conclude that this is a good criterion for reusability.

- using the textual result information panel to study the metadata that describes those components. For instance, if a majority of the slides that are reused have a specific topic, one might think that this topic is highly popular, and therefore that reuse of components depends on popularity.

- using the node-link visualisation. For instance, if a majority of all the components that match the specified interval [10, 20] in the dynamic query control panel, have an orange color, it means that a majority of frequently reused components in ALOCOM are images. For example, we found a figure in the ALOCOM repository about the Globe Consortium [Globe, 2008]: this figure has been reused in 10 different slides by different authors.

### How are components put together in aggregate content?

The node-link visualisation can be used for investigating how components are put together in aggregate content. Both the overview visualisation and the detailed visualisation van be used to investigate this. The detailed visualisation only shows

- a selected component,
- the components in which selected component is reused, and
- the components that are reused in the selected component.

For example, figure 4.9 shows a slide that has been reused in 16 different presentations. This slides consists of 3 figures (orange), and two text-fragments (blue).

### Which characteristics of the components contribute in making it reusable?

The textual information panel can be used to study the metadata of the components. The preview panel can be used to actually see the components and manually deduct their characteristics. For instance, there is a text-fragment about the Simple Query Interface [Simon et al., 2005] that is reused 110 times in 25 different slideshows in the centre of the big circle. When looking at this fragment, we can see that it exists out of one single word: “SQI”. One might conclude from this, that the length of text fragments should be short in order to be reusable. This corresponds with the “Reusability Paradox”: To make learning objects maximally reusable, learning objects should contain as little context as possible [Wiley, 2002] [Verbert et al., 2008].

### 4.6.2 Evaluation

Rapid-prototype developing allows the developers of an application to present examples of their concepts for early visualisation, verification, iteration, and optimisation [Luqi, 1989]. Our application was developed following this principle and we used the early development of the application to conduct some evaluation of it at the same time. The target audience of this prototype are researchers in the field of reusability of content. Although this is a very narrow scope of users, this work was needed in the context of our research group as we are interested in reusability of learning material [Verbert et al., 2008].

Because of the scope of the target audience, organising large evaluation sessions with users is quite difficult because of the number of users that are needed. Therefore, we followed the recommendations in [Andrews, 2006] to involve a limited number of highly experienced users in the early stage of development to get relevant feedback to better capture the exploration and analysis tasks needed. A fairly small number of test users ($4 \pm 1$) is enough to find most usability problems [Nielsen, 1992c]. For this reason, we chose to involve 4 expert users in the TEL community in an evaluation of our approach. First of all, a video demo was presented to those expert to give them a glance of the available exploration ways in our prototype. After this demo, we asked the subjects to explore the ALOCOM repository to see if they:

- understood the possible ways of exploration and found them effective and efficient,
- were able to find interesting reuse patterns, and
- had general usability remarks on the tool.

During their exploration, they were asked to use the "think aloud" method, which is a method for user testing where the users verbalise their thoughts while using the system that is being tested [Nielsen, 1992a]. In this way, those users reveal their view on the system and possibly, their misconceptions. From these sessions, we learned quite a number of interesting items:

1. The visualisation of a repository of fine-grained components is a promising and effective alternative to get insight in the components and their relations. The node-link visualisation is a clear visualisation that shows the relationships between the different components. However, removing the clutter with e.g the dynamic query panel is something that the users should get used to. In other words, the time to learn to solve that task is quite high. This would be a big issue if the target audience would be typical end users. However, our target audience exists of expert users, and therefore this learning time is considered to be an issue of little importance.

2. Especially the combination of the visualisation and the preview panel were found interesting ways of exploring the repository and discovering reuse patterns.

3. Not only insight in the reuse of components can be obtained from the visualisation, but also possible similar regions of interests between different authors that do not know each other can be obtained when for instance an author notices that he or she uses the same sources of components as another author. With this information, an overlapping social network could also be visualised, based on the reuse of different components. This is quite an interesting angle for future work. In chapter 5, information visualisation techniques are used to provide understanding of the relationships between users, tags and bookmarks in social bookmarking tools. The results of that case study could be used here in the future.

4. We should validate if the system will still have a high performance when we scale up the number of components in the repository. In our current prototype with ± 50.000 objects, panning while activating the fisheye view (section 4.5.1) was sometimes slow when the experts were trying to explore the repository by using this. This will be further discussed in section 4.7.

5. We should ask authors of uploaded presentations to assess the automatic decomposition of their learning objects. Results of this study might indicate to what extent the automatic decomposition approaches a decomposition that is performed manually. This work has been done in the context of [Verbert et al., 2008]. This remark however, has little to do with our approach of visual access to the ALOCOM repository but we added it to this list for completeness.

A number of problems were also found which could not immediately be supported by our visualisation prototype. We discuss them in the next section on future work, and propose solutions for them.

### 4.6.3 Future Work

- **Calculate statistics about the reuse in learning objects.** This seems obvious, but this functionality has not been available in the framework. Researchers want to be able to select a number of interesting components and the system should be able to calculate statistics like the reuse value.

- **Show social network of users.** The researchers suggested that it would be interesting to visualise a social network of authors of presentations. For example, if author A would reuse large amounts of components that were created by author B, author A and B would be connected in the social network.

- **Evolution of reuse through time.** It is not possible to see how a specified component is reused through time. This information is important because two components could have the same reuse value, e.g these slides are reused in 10 presentations. But one of them could have been reused 10 times in 1 week, while the other one could have been reused 10 times in 2 years. A visualisation of growing bubbles where each bubble represents a component could e.g be used for this. This visualisation is used and described in chapter 6.

- **Expand the dynamic filter controls**. For now, users can select an interval of the amount of times that a component is reused, and search for components that are matching this interval. However, more interesting is to be able to create multiple intervals like: this component has been reused in at least 2 slides that belong to different presentations because it would enable the users to create more powerful filters that can be used to discover new patterns in the data. One way of doing this, is to use parallel coordinates [Inselberg, 1985]. Figure 4.10 shows a parallel coordinates presentation of 4 components. The selected component 2 is reused in 8 presentations and 10 slides. Per axis, different intervals can be created by changing the scales on the Y-axis.

- **Generalise the target user group.** For this chapter, the target audience was of narrow scope, namely researchers who are interested in the reuse of content. One idea would be to expand this audience to e.g managers of research groups, both in the academic world and in companies. If each member of the team of a manager would save his given presentations within ALOCOM, the manager could use the visualisation to see how integrated the work of his staff is. E.g if there are no big circles in the node-link visualisation, it means that everyone is doing work on its own, with no regard for the work of others. This is most probably not what the manager wants.

## 4.7 Implementation Notes

### 4.7.1 Graph Drawing

The graph drawing itself is performed by using the INRIAs Infovis Toolkit [Fekete, 2004], an interactive open source toolkit, written in java, which aims to ease the development of visualisation applications and components. In this toolkit, a software bridge in Java$^{TM}$was implemented to the GraphViz [Ellson et al., 2003] toolkit, which essentially is a set of graph drawing algorithms and programs written in C. By plugging the infovis toolkit in our framework, we are automatically able to

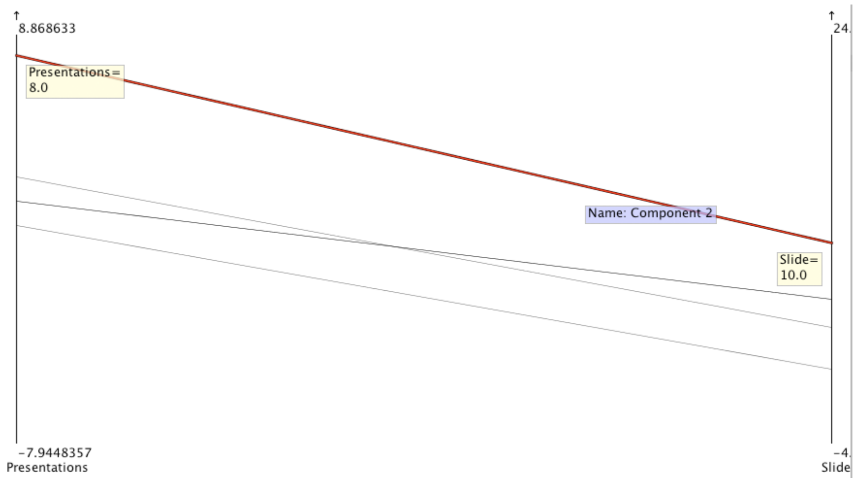



Figure 4.10: Parallel coordinates presentation of 4 components. The selected component 2 is reused in 8 presentations and 10 slides.

reuse this bridge and therefore all the graph drawing algorithms that are available in the GraphVis toolkit without having to create our own bridge. This also allows us to use those algorithms in combination with other information visualisation libraries that we have included in our framework. For example, Piccolo is a toolkit that supports the development of 2D structured graphics programs, in general, and Zoomable User Interfaces (ZUIs), in particular. A ZUI is an interface that presents a huge canvas of information on a traditional computer display by letting the user smoothly zoom in, to get more detailed information, and zoom out for an overview [Bederson et al., 2004]. We have integrated piccolo in our framework to show the open nature of our framework. We can know combine both toolkits and use the graph layout algorithms of GraphVis to create a zoomable user interface with Piccolo.

## 4.8 Conclusion

This chapter reported on interactive visualisation techniques that aims to enable visual access to a large repository of small reusable content components that were created by disaggregating legacy content. The goal has been to provide researchers who are interested in reuse of content, with a means to improve their understanding of reuse patterns within the repository.

The contributions of this chapter are the answers to the following questions:

- how can we effectively and efficiently support visual access to a collection of small learning objects, based on the relationships between those components? For this question, we have created a node-link graph visualisation that represent all components in one overview (Section 4.4)

- how can we effectively and efficiently interact with visualisations of learning objects? For this question, we have integrated a number of interaction techniques like dynamic query controls, excentric labeling, fisheye views and previews (Section 4.5)

An expert evaluation has been conducted (Section 4.6.2). This has shown us that users considered the visualisation techniques for accessing a repository of fine-grained components effective.

The framework that we have created and presented in chapter 2 has been adapted while creating a prototype for the techniques above. More concretely, another library, INRIA's infovis toolkit, has been integrated into the framework with new information visualisation techniques. Those are now available for all case studies that are created with the framework.

The previous chapter has described how methodologies and technologies for information visualisation can be used to effectively and efficiently support visual access to a collection of learning objects, based on their topic that is defined in the LOM classification. This chapter described how we can enable visual access, when there is no LOM classification available for the learning objects. In this case, we will use the "has Part / is Part Of" relationships between the objects, as a starting point. Because the expert suggested that it would be interesting to focus on social aspects of the data, the next chapter will discuss how we can support visual access to social bookmarks.

# Chapter 5

# Visualising Social Bookmarks

## 5.1 Introduction

The main challenge that this dissertation addresses is how to provide flexible, efficient and effective access to large collections of reusable components and enable insight in the contents of such a collection.

The ability to store or bookmark web pages by describing them by tags or terms, has been one of the most important features of browsers since the beginning of the Web [Hammond et al., 2005]. Social bookmarking tools became possible when the process of keeping bookmarks migrated from the browser to the Web.

Tagging is the process of assigning meaning to online items, such as web pages, images, videos, etc. by labelling them with personalised keywords that are shared among users [Guy and Tonkin, 2006]. The purpose of web-based social bookmarking tools is to tag the content of other users, mainly for the benefit of the tagger, although the bookmarks and tags are generally public, and users can establish networking opportunities [Hammond et al., 2005].

Social networks are an important factor for finding and spreading information because a big part of learning is social [Wenger, 1996]. For example, talking with your neighbour about how to build a garden house, may provide new insights about this process. Wenger introduced the idea of the "community of practice" which means that people can satisfy their need for information more efficiently if they are embedded in a community with similar interests and problems [Wenger, 1998]. "Weak ties" are relationships between people that don't know each other very well. These weak ties are important to allow information to spread from one closely-knit community to another [Granovetter, 1983].

Implicit relationships between users, tags and content are therefore very useful in Web2.0-style social information retrieval systems. However, those relationships are not always clear in the traditional ways of accessing social bookmarks. Our research therefore focusses on using information visualisation techniques

- that enable insight to analysts in those implicit relationships, and
- that may offer end users new ways to find content and information that could be of interest to them but that would not have been found through explicit searches.

We will start this chapter in section 5.2 with an overview of existing social bookmark tools and the traditional ways of accessing the information within those systems. Section 5.3 looks at the data of two existing social bookmarking systems in more detail. A number of requirements for opening the view on social bookmark spaces, are discussed in section 5.4. Section 5.5 elaborates on a prototype application with a cluster map visualisation for exploring social bookmarks. This prototype has been used to explain some typical usage scenarios in section 5.6. Implementation notes are presented in section 5.7. Section 5.8 discusses the prototype and presents evaluations of the methodologies. We conclude this work in section 5.9.

This chapter has been accepted as such for publication in a special issue on "Social Information Retrieval for Technology Enhanced Learning" of the Journal of Digital Information (JoDI). Earlier versions of this chapter has been published in [Klerkx and Duval, 2007c].

## 5.2 Social Bookmark Tools

There are quite a few social bookmarking initiatives, like CiteULike [CiteULike, 2008], Connotea [Connotea, 2008], del.icio.us [delicious, 2008], Furl [Furl, 2008], BlogMarks [BlogMarks, 2008], etc. A thorough general review can be found in [Hammond et al., 2005]. A number of ways are usually available for browsing personal bookmarks:

1. First of all, one can browse through scrollable pages of bookmarks.
2. Secondly, one can perform a "tag query" of the collection by clicking on a tag [Millen et al., 2007].
3. Exploratory search activities are also often supported in social bookmark applications. Some examples are browsing bookmarks by time or by popularity like on del.icio.us.

Tagging is an unstructured bottom-up approach of classifying content, in contrast to a top-down structured approach based on taxonomies, thesauri or ontologies [Weinberger, 2007]. The semantic structures that result from a tagging approach are often referred to as a folksonomy. Tags generally produce a flat namespace, rather than the hierarchical structures that taxonomies or other formal classification systems provide [Hammond et al., 2005]. However, there can also be rich implicit structures between tags, bookmarks and users that are not immediately clear. For instance, there is a clear relationship between two separate tags that have been used to classify the same bookmarks. Two users that use exactly the same tags for different bookmarks is another example. This means that they are interested in the same topics, and therefore they might be interested in the bookmarks of the other.

On top of that, some researchers claim that social bookmarking applications become harder to navigate as the amount of tags increases an they thus become less meaningful [Chi and Mytkowicz, 2007]. The reason for their claim is that e.g. del.icio.us is moving closer and closer to the proverbial "needle in a haystack" where any single tag references too many documents to be considered useful.

An important goal of our research is thus to provide understanding in bookmarks, tags, users and the relationships between them. In related work, "tag clouds" are often used to visualise the tag structures of one or more users. Tag Clouds are visual presentations of a set of words, typically a set of tags, in which attributes of the text such as text, weight or color can be used to represent features (e.g. frequency) of the associated terms [Halvey and Keane, 2007]. When a user clicks on a tag, the user obtains an ordered list of tag-described resources, as well a list of related tags [Hassan-Montero and Herrero-Solana, 2006]. Usually, the tags are displayed in alphabetical order.

Many other visualisations of tag, bookmark and network structures were created for del.icio.us and others. Notable such visualisations include HubLog [HubLog, 2008] that enables graphical browsing of del.icio.us tags in a mind mapping way. Extisp.icio.us Text [Extisp.icio.us, 2008] provides a random textual scattering of user tags, sized according to the number of times that they have been used. Revealicious [Revealicious, 2008] is a set of 2D graphic visualisations that enables the user to browse, search and select tags and bookmarks. Vizster [Heer and Boyd, 2005] is a tool that is designed for visualising the online social network Friendster, as a browseable network of social relations. Vizster is very useful for sociological research but does not take tag structures into account.

All these initiatives are based on either visualising the tag structures or the community structures but they do not take both structures into account for the visualisation at the same time, to make implicit community and tag structures more apparent for e.g. analysts that are studying folksonomies. The remainder of this paper investigates how to use information visualisation techniques in such ways to make those structures more apparent and to make bookmarks accessible.

Two social bookmark tools were chosen as the basis of our visualisation design: del.icio.us and the CALIBRATE portal. These tools are described in the next section, together with the rationale for choosing them.

## 5.3 Data

### 5.3.1 del.icio.us

Del.icio.us [delicious, 2008] is probably the most well-known social bookmarking tool, designed to store and share bookmarks on the web instead of saving them in the browser. We chose del.icio.us as a source of data as it is highly popular with many users, lots of data and with a very easy API to access this data. Those characteristics makes del.icio.us a very good reference point: if the visualisation techniques prove to be effective and efficient for exploring del.icio.us, they might be valuable for other social bookmark tools as well.

A number of steps are involved in the process of collecting del.icio.us data:

1. We start from one or more del.icio.us user names and collect their bookmarks, and the tags of those users that describe those bookmarks.

2. Del.icio.us enables users to create a network of other users, so that they can access the bookmarks of these users. In the second step, we collect the tags and bookmarks for all the users that are in the network of the users identified in step 1.

3. Del.icio.us also enables users to be a fan of other users. User A "is a fan of" user B if user A has added user B to his network. This relation is not necessarily bi-directional. In this third step, the same data is again collected for all the fans of the chosen user(s) in step 1.

This process of collecting data takes place at runtime but can also be performed beforehand. However, the del.icio.us API does have some limits:

- Only those bookmarks, tags and users in the network can be used that are explicitly made public by the user that added them.

- The del.icio.us API does not allow to submit a web page and query and retrieve information on the users that tagged that particular web page. If this would be allowed, we could for instance take a number of random webpages, ask which users tagged them, and afterwards visualise the implicit social structures that possibly exist between those users. What can be retrieved for specific web pages is the number of users that tagged the web page and added it to their collection. We could use this number to visualise the popularity of a web page if we consider a higher number as more popular.

- The usage throttling and abuse monitoring software at the del.icio.us website limits the amount of queries that can be performed in order to prevent the harvesting of all the data of del.icio.us [del.icio.us, 2008]. On top of that, a maximum of a hundred bookmarks per user can be retrieved in one query.

Summarised, the available information characteristics on bookmarks are:

- the URL, e.g “www.google.com”,
- tags that describe the web page, e.g “search, search engine, web”,
- title of the bookmark, e.g. “Google search engine”, and
- number of users that tagged the bookmarks, e.g “saved by 22974 people”.

The characteristics of a particular user are:

- the del.icio.us username, e.g “jkofmsk”
- the users in his network, and
- fans of the given user.

### 5.3.2 CALIBRATE portal

Applying social tagging and bookmarking to an educational setting, CALIBRATE users are teachers who create a social bookmark and add tags on learning material, for their own use and to share them with their students or other teachers. This work is done in the context of the CALIBRATE project [CALIBRATE, 2008], which makes K-12 digital learning resources available to 78 schools in Hungary, Austria, Estonia, Czech Republic, Lithuania and Poland in different curriculum areas. Schools can access material in different languages through a portal that is connected to a federation of learning resource repositories (GLOBE) [Globe, 2008] in the pilot countries. Within this portal, a personal bookmarking and tagging tool has been available to all users since the beginning of 2007. The data of these users, their tags and their bookmarked learning resources are taken into account for our visualisation where it is used to visualise the implicit structures between users, tags and learning resources.

Learning material is described by Learning Object Metadata (IEEE LOM [IEEE, 2002]. A goal of LOM is to enable sharing of descriptions of learning resources between resource discovery systems, which should lead to a reduction in the cost of providing services based on high quality resource descriptions [Duval and Hodgins, 2003]. The availability of LOM works to our advantage because it contains more metadata than the available metadata in traditional social bookmark systems like del.icio.us. This metadata can be taken into account while visualising them to further enable insight in the collection. Compared to the features that were available in del.icio.us, we use the following extra elements that describe the resources and are important in the exploration process:

- authors; e.g. while searching for material on the semantic web, a learner might prefer material created by Tim Berners-Lee, rather than an unknown author.

- language; e.g material in Spanish is not interesting for non-Spanish speaking learners.
- unique identifier; e.g. important for actually retrieving the material if it is interesting.
- creation date; e.g. material might be out-of-date and therefore less interesting than newer created material.

On top of this, we have access to information about the users in CALIBRATE, like the languages they speak and their country of residence. This information will for instance be used to filter out material that cannot be understood by the user.

The information from CALIBRATE and del.icio.us is summarised in Table 5.1. The data itself for this analysis came from a period of ten months, January 1 to October 31 2007 and includes real data. There were 142 teachers, hereafter referred as users, who made 1022 posts of favourite learning resources. These bookmarks covered 682 different learning resources or items. There were 1029 multilingual tags recorded in the system, some of which were reused by users.

## 5.4 Requirements

Our research therefore focusses on using information visualisation techniques

- that provides insight to analysts in the implicit relationships between bookmarks, tags and users;
- that may offer end users new ways to find content and information that could be of interest to them but that would not have been found through explicit searches.

This section discusses a number of requirements for applications that are designed with the goal of opening up the view on social bookmark spaces. A prototype application that uses information visualisation techniques, has been designed with these requirements in mind.

1. Hypertext makes a dynamic organisation of information, i.e. nodes with content, possible through links or connections. Navigation between the information is possible by following those links. Many rich hypertext systems are criticised since its users may suffer from the "lost-in-hyperspace" syndrome which means that they cannot identify where they are, or that they cannot remember what they have covered already [Theng and Thimbleby, 1998]. Much work has been done to address this problem, like providing breadcrumbs that show the trail of links leading to the current page [Blustein et al., 2005]. A number of visualisations of hypermedia systems have

Table 5.1: Overview data features in del.icio.us and CALIBRATE portal

| Object | Feature | del.icio.us | CALIBRATE Portal |
|---|---|---|---|
| Resource | authors | ✘ | ✔ |
| | language | ✘ | ✔ |
| | ID | ✘ | ✔ |
| | timestamp | ✘ | ✔ |
| Bookmark | user ID | ✔ | ✔ |
| | tag(s) ID | ✔ | ✔ |
| | resource ID | ✔ | ✔ |
| | timestamp | ✔ | ✔ |
| User | ID | ✔ | ✔ |
| | country | ✘ | ✔ |
| | language(s) | ✘ | ✔ |
| | repository Information | NA | ✔ |
| Tag | ID | ✔ | ✔ |
| | language | ✔ | ✔ |
| | timestamp | ✘ | ✔ |

been developed to address this problem by using a novel paradigm instead of the hypertext paradigm [Mukherjea, 1999]. Most of them have been based on the node and link graph diagram metaphor like the "Navigational view builder " [Mukherjea and Foley, 1995] where web pages are represented as nodes.

Social bookmarking tools like del.icio.us consider both users and bookmarks as pieces of content. Tags are presented as links between those users and bookmarks and can therefore be used to navigate through the social bookmark space. Applications with the goal of opening up the view on social bookmark spaces should be careful not to create the same lost-in-hypertext syndrome that can be found in traditional hypertext systems. In our design, we use a novel paradigm, called a cluster map visualisation, for browsing and exploring social bookmarks. This paradigm and the reason for choosing it, will be explained in section 5.5.2.

2. Secondly, we want the visualisation system to be able to automatically identify implicit tag and community structures. Such a structure is formed when two or more tags or users describe or share common bookmarks. After the identification of these structures, they should be visualised for the users, so that these users can become aware of them. We think that this can lead to the discovery of interesting material for those users, that could not have been found through explicit searches.

3. We consider visualisation as a technology. This means that visualisation has to be effective and efficient or that it should do what it is supposed to do by using a minimal amount of resources [Van Wijk, 2006]. Information exploration should be a joyous experience, but many commentators talk of information overload and anxiety [Wurman, 1989]. A key challenge for information visualisation researchers, often suggested by the community, is to make their systems usable by common computer users [Mukherjea, 1999]. Visualisation applications do not always fit in the normal workflow of users. If we therefore want that our visualisation design is useful, efficient and effective, users should be able to explore the social bookmarks in a playful manner in a fun and engaging space.

4. Human-computer information retrieval (HCIR) has emerged in research as the study of information retrieval techniques that brings human intelligence into the search process [Marchionini, 2006c]. To achieve this, users should be able to interactively create a selection of bookmarks and get details on them when needed [Shneiderman, 1996].

## 5.5 Social Bookmark Visualisation

Figure 5.1 shows that our visualisation consists of three parts:

1. a selection widget that presents lists of the users and tags in the currently loaded data (section 5.5.1),

2. a cluster map visualisation (section 5.5.2), and

3. a filter pane with integrated results list (section 5.5.3).

These panes are synchronised with each other. We discuss these panes in detail in the next paragraphs. Typical usage scenarios will be described in section 5.6 where we focus on exploring and discovering new bookmarks, by identifying and visualising those implicit structures that are e.g. formed when two or more tags or users describe or share common bookmarks.

### 5.5.1 Selection Widget

Users of social bookmarking tools most frequently browse the bookmark space by other people and by tags [Millen and Feinberg, 2006]. Providing easy access to the people and tag space is therefore important as we want to create a visualisation that is both effective and efficient while exploring the social bookmark space. End users of e.g. del.icio.us know how to find interesting items within lists. Therefore, the selection widget presents overview lists of both users and tags, that are currently loaded in the system. Alternatives to list presentations are e.g. tag clouds [Halvey and Keane, 2007]. However, our main goal is not to find interesting tags on itself but to provide understanding of relationships between users, tags and bookmarks. A simple list presentation is sufficient as a starting point for selecting interesting tags and users for this goal.

Figure 5.2(a) shows that each item is presented as a node in the list. A node consist of the particular tag or user and the number of bookmarks that are described with that tag or user. In CALIBRATE, more information is available about the resources. Therefore, Figure 5.2(b) not only shows the tags and users but also countries and languages. Each node also contains a check-box that can be used to create a selection of items that should be added to the visualisation. To optimise browsing of tags, users, countries and languages, a filter widget has been built that is visible on top of the selection widget and that can be used to find items faster. Figure 5.2(a) shows for instance only those items in the lists that start with “web” because the word “web” has been typed in the widget.

In our design, we have chosen to follow the philosophy of “start with what you know, then grow” [Heer et al., 2005]. This means that, by default, nothing is visualised in the cluster map but the bookmarks of the current user. The user can thus start from his own bookmarks, and afterwards use checkboxes of

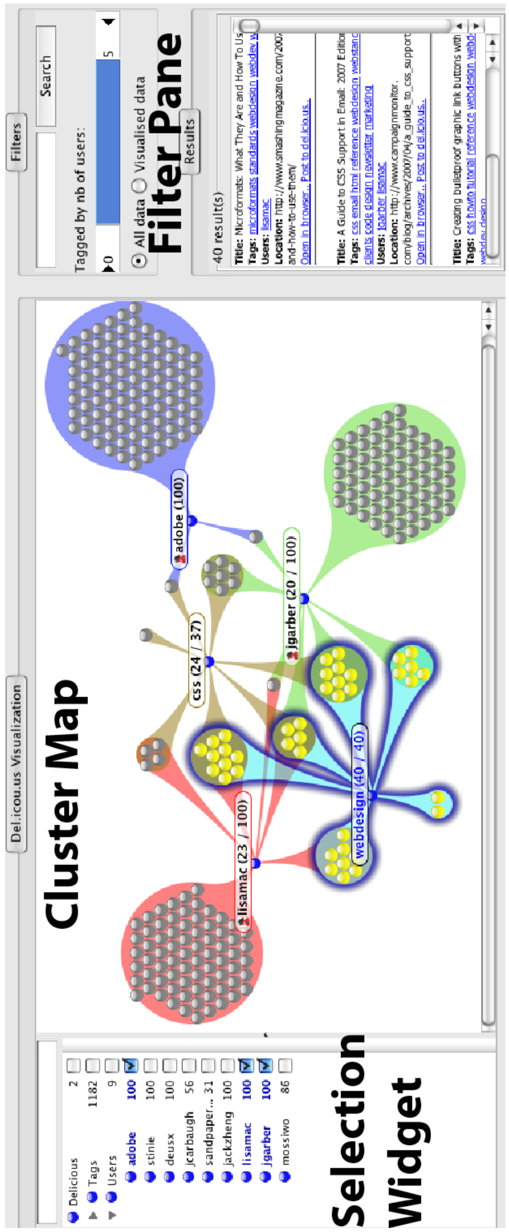
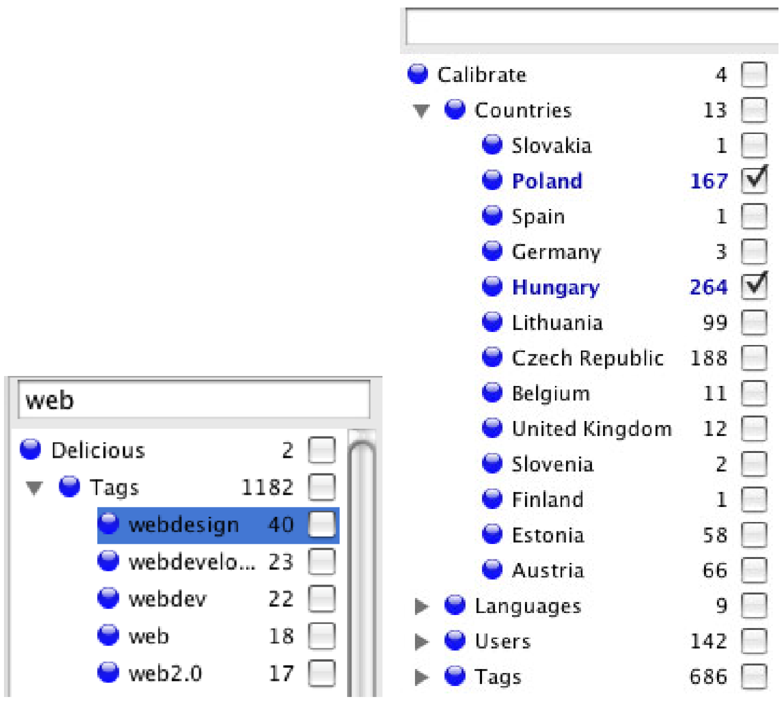



Figure 5.1: Cluster Map visualisation




(a) del.icio.us data (b) CALIBRATE data

Figure 5.2: A view of the selection widget. When a user clicks a checkbox, the item is added to the cluster map visualisation. (a) shows the lists for del.icio.us, (b) for the CALIBRATE portal.

either tags, users, languages or countries, to include the corresponding items in the visualisation. In this way, the initial visualisation carries less perceptual and computational burden to start with, because the user does not need to figure out complex structures that could be visual if more information was added to the visualisation. An example of this can be seen in Figure 5.8 which shows a rather complex cluster map of 7 users and one tag "java". When a user would see this visualisation from the start, he would need a some time to make sense of the patterns.

The same philosophy is also adopted in related research like Vizster which is a visualisation system for playful end-user exploration and navigation of large-scale online social networks [Heer et al., 2005]. HubLog [HubLog, 2008]) enables graphical browsing of del.icio.us tags in a mind mapping way and starts from one tag that a user should fill in. The del.icio.us network explorer [explorer, 2008] starts with one user and enables the exploration of that users' network.

### 5.5.2 Cluster Map Visualisation

Visualisation should show the data: it should reveal what the data is about [Tufte, 2001]. Our visualisation of the social bookmark space should therefore be expressive. Clear visualisation of the triple users, items and tags should be supported to create an effective navigational and exploration tool. Many visualisations have been developed for web resources (see [Fluit et al., 2005] and [Dodge and Kitchin, 2001] for an extensive overview). Especially in the context of social bookmark tools, a number of graph visualisations have been created (see section 5.2). Another metaphor that has frequently been used for accessing social bookmark tools is the landscape metaphore like the "Islands of Music" where tags of music songs are represented as islands on a map [Pampalk, 2006]. Similar tags are located closer to each other.

A cluster map [Fluit et al., 2005] is a technique that is very expressive in visualising classes and items that belong to those classes. One can also consider items like tags, users, language and countries as classes where bookmarks belong to one or more of those classes. Those classes and relationships between them need to be easy to detect. A cluster map makes use Venn diagrams which makes it easy to detect those relationships between those classes [Venn, 1880]. Therefore, we decided to customise a cluster map for visualising social bookmark spaces with the goal of validating the use of this technique for exploring those spaces.

A cluster map visualises the objects of a number of selected classes. Those classes and their relationships are easy to detect. It is immediately apparent which items belong to one or multiple classes, which classes overlap and which do not. A bookmark is represented as a small circle in the visualisation. Each bookmark belongs to the collection of one or more users. In Figure 5.3, two users are shown, each with a hundred bookmarks in their collection. A small icon of a human, followed by a user name identifies a user. Those two users have six bookmarks in

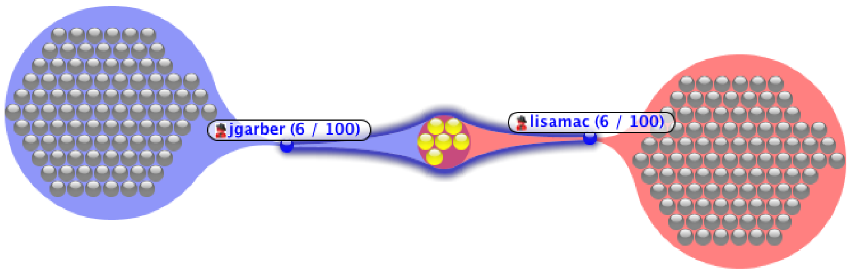



Figure 5.3: Cluster Map showing 2 users with 100 bookmarks and 6 of them in common. An icon followed by a username identifies a user.

common which is represented by the label “6 / 100”. This is represented in the visualisation by the smaller common cluster of bookmarks in the middle. By using the selection widget that has been described in the previous paragraph, users can select which users and tags are drawn on the cluster map.

Bookmarks can be clustered by the users that have them in their collection, or by the tags that describe them. This can be seen in Figure 5.4, where the tag “email” is shown. Users “lisamac” and “jgarber” have one bookmark in common which is tagged by “email”. The user “jgarber” has 3 extra bookmarks tagged with “email” that are not in the collection of the other user.

An important feature of social bookmarking tools is the popularity of bookmarks. For instance, if a web page on the topic of “java” is added by 10.000 people, and another web page on the same topic is only added by a hundred people, one might trust the first web page more than the latter. The rapid and accurate identification of popular bookmarks is therefore an important requirement in the design of our visualisation. Color is an important and frequently-used feature for information encoding [Healey, 1996a] because well chosen colors allow for rapid and accurate identification of individual data elements by users [Healey, 1996b] [Few, 2004]. In Figure 5.4, two main colors are used to represent bookmarks  yellow and grey. A color-saturation sequence is used for the yellow color to identify the popularity of the bookmarks among the users. This means that a bookmark that is presented with a brighter yellow color, is more popular among users.

Such a set of selected bookmarks can be created by clicking on a user, a tag, a country or a language in the selection widget, or by performing a keyword query in a search box. A selection of bookmarks is color-encoded in the visualisation:

- grey bookmarks are items that were filtered out by the users because they were less interesting, and
- yellow bookmarks are items that are in their interest.

The yellow bookmarks in Figure 5.4, belong to a selection of bookmarks that

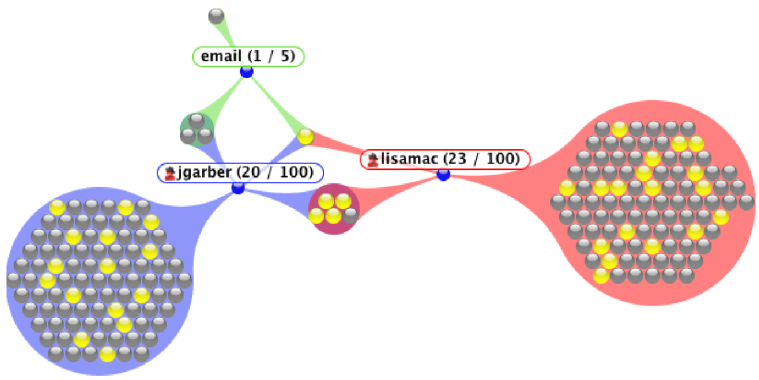



Figure 5.4: Cluster Map visualisation, showing 200 bookmarks with 2 users and 1 tag “email”. 20 bookmarks of user “jgarber” and 23 bookmarks of users “lisamac” are about “webdesign”. The selection of bookmarks on webdesign has been created by using the filter pane (section 5.6(a)).

were tagged with the tag “webdesign”. The yellow bookmark between “lisamac” and “email” therefore belongs to the collection of the users “lisamac’ and “jgarber” and is described by the tags “email” and “webdesign”. This selection was created by clicking on this tag in the list of tags in the selection widget, which can be seen in Figure 5.2(a) where this tag is highlighted.

Looking at CALIBRATE data, we can visualise bookmarks belonging to countries. Figure 5.5 shows that people from Poland and Hungary tagged respectively 167 and 264 learning resources in total of which there are 29 in common.

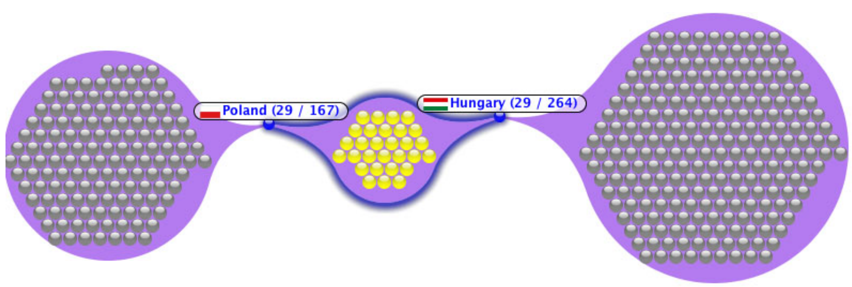



Figure 5.5: A view of shared learning resources: of all the 167 learning resources that were bookmarked and tagged by users from Poland, 29 are shared with the 264 learning resources that were tagged by users from Hungary.

### 5.5.3 Filter Pane with Integrated Result List

In the previous paragraph, we already mentioned that users should be able to interactively create a selection of bookmarks and get details on them when needed [Shneiderman, 1996]. In this way, they can zoom in on potentially relevant bookmarks and continuously keep an overview of how the additional search criteria restrict the remaining number of bookmarks. Therefore we have integrated a filter pane with a number of controls that can be used to filter out less interesting bookmarks or add more interesting ones. Figure 5.6(a) shows controls that can currently be used:

1. The first one is a search box where a typical keyword query like "javascript" can be performed.

2. The second one is a data visualisation slider [Eick, 1994] that can be used to indicate an interval of the number of people that tagged a bookmark.

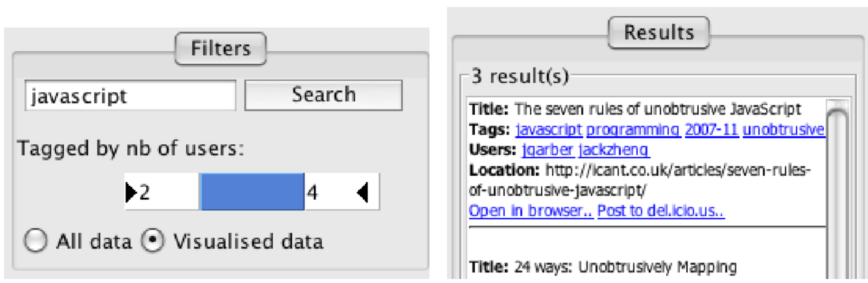
These combined controls allow users to quickly find popular bookmarks on a particular topic. This is a basic use-case that is available in most social bookmarking tools like del.icio.us and Dogear [Millen et al., 2006]. Other controls could be added to the filter pane as well. In the case of CALIBRATE, we could e.g. choose to add a drop box for the language of the resources. Users would then be able to filter out languages they do not master. The results of the combination of these filters are shown in two places:

1. **the clustermap visualisation**, where selected resources that do match the criteria in the filter pane are represented as yellow circles. The resources in Figure 5.5 that are represented in yellow match the criteria in the search pane (Figure 5.6(a)) where the user added the keyword "javascript" to include only results on this topic, and also changed the slider to include only the objects that were tagged by at least 2 users and at the most 4 users.

   The complexity of the visualisation can be high when many tags are added with a lot of overlaps between those tags. An important aspect of using the filter controls, is therefore, that they can be used to reduce this complexity.

2. **the result list**(Figure 5.6(b)), that shows detailed metadata about the bookmarks that are either selected in the cluster map, or bookmarks that match the search terms when a query was performed. In the case of the data from del.icio.us, the metadata cover the title, the location, user(s) that added the bookmark and tag(s) that describe the bookmark. On top of those, the metadata covers the language and the country of the users in the case of the CALIBRATE portal. A user can interact with the detailed information by clicking on e.g. a tag that describes a bookmark. If this tag is not already drawn on the cluster map, the visualisation automatically updates

itself and the tag classification is shown, together with the possible relationships between this tag and currently visualised bookmarks. The results should preferably be ordered by relevance because users always want to find the most relevant items. Therefore, we would e.g. need relevance feedback from users. However, capturing and incorporating this feedback has been of less priority because our main goal has been to visualise the relationships between users, tags and bookmarks. The results are therefore ordered alphabetically in the result list for the time being. Ordering by relevance has been moved to future work.




(a) Filter Pane (b) Result view

Figure 5.6: (a) shows the filter pane with a search box, a slider for the number of users that tagged a bookmark and radio buttons for indicating on which data set to search. (b) shows the result view with detailed metadata about bookmarks.

## 5.6 Exploring Social Bookmarks

We mentioned in the introduction that the use of information visualisation techniques offers end users new ways to find content and information that could be of interest to them. On top of that, it is important that these techniques enable insight in the implicit relationships as they are not always clear in the traditional social bookmark tools (section 5.4); insight that could enable them to find interesting material that would not have been found through explicit searches. For example, if a user finds a relationship with another user, he can find material that has been tagged by that user that is interesting to him. As an alternative to exploring social bookmarks on the del.icio.us and CALIBRATE portal, we offer a novel access paradigm that tries to enable a user to explore social bookmarks in a playful manner in a fun and engaging space by creating a cluster map visualisation. The cluster map visualisation enables users to identify tag and community structures. The following sections will explain how users can become aware of them by interpreting the visualisation.

### 5.6.1 Exploration ways

Users can explore the social bookmark space by using the cluster map visualisation in a number of ways.

1. First of all, the basic exploratory functionality on the del.icio.us website is supported because this is the basic functionality that typical end users of social bookmark tools expect to be able to do: One

   - can look at one's own bookmarks,
   - see which tags are used to tag a web page, either by the user himself or by others in his network, and
   - see all other users in one's own network that have saved the same bookmark. Note that this means that a user will not see del.icio.us users that added the same bookmarks but are not in the same network. The reason for this is the usage throttling and abuse monitoring software at the del.icio.us website (section 5.3.1).

   We chose to let a user start from an egocentric point-of-view, with a visualisation of one's own bookmarks, much along the lines of the philosophy of start with what you know, then grow [Heer et al., 2005]. Users will see those bookmarks from the start, which will give them a more comfortable feeling because they are already familiar with those bookmarks and tags. They can then select a number of bookmarks in the visualisation or perform a keyword query in the filter pane, after which detailed metadata about the resulting bookmarks is shown in the result list. Only those bookmarks that are also currently visualised, are added to the result list because both the visualisation and the result list are continuously synchronised with each other or tightly coupled. This is needed to have comprehensible and consistent affordances to guide the end users [Ahlberg and Shneiderman, 1994]. The metadata of the results contains all the tags that describe the bookmarks, i.e. not only the tags of the user itself. The metadata also contains all the users that added the specific bookmarks. The user can click on those tags and users, and by doing this add them to the cluster map, where the layout of the bookmarks updates itself to represent the new sub-clusters, like in the example of Figure 5.4 where the tag "email" is added.

2. The first way of exploring is not sufficient for efficiently exploring social bookmarks. Users should also be able to explore all tags that are in the data to find interesting bookmarks. A second way of exploring is therefore

by browsing the selection widget to find those tags. This can be compared with sifting through a list or cloud of tags on the del.icio.us website. Users can order the tag list alphabetically or by the number of bookmarks they describe. When ordering them in the latter way, one can see which tags are used the most. By adding tags to the visualisation, the corresponding bookmarks, possibly tagged by different users, can be explored in the cluster map. Moreover, tags often denote concepts, so relationships between those concepts may also be discovered. An example of this, is shown in Figure 5.7 where users can see the relationships between the concepts "programming", "dev" which stands for development, and "java". However, it's the responsibility of the users to interprete the visualised relationships. "Coffee" is for example related with the coffee brand "java" but "coffee" is not related with the concepts of "programming" and "dev".

3. Highly interactive applications support human-computer information retrieval (HCIR) [Marchionini, 2006c]. Users should for instance be able to easily see all metadata, associated tags and users of a bookmark in the visualisation. This can easily be done by clicking upon the bookmark in the visualisation after which this information is either shown in the result list.

4. When a user finds an interesting bookmark, he should be able to keep it for further reference [Shneiderman, 1996]. Implementing and maintaining a system with user information and the history of his actions is not in the scope of our design because we can use the del.icio.us system for this purpose. Users can easily post newly found bookmarks to del.icio.us by clicking upon them in the visualisation. For the same reason, users that are found to have the same interests, can also be easily added to the del.icio.us network of the current user by clicking on the corresponding nodes in the visualisation. This is only possible when the user that is currently exploring the social bookmark space, has a user account on del.icio.us.

5. Only a fraction of data is loaded from the start as we follow the philosophy of "start with what you know, then grow" (section 5.5.1). To enable further exploration, one can click the nodes in the visualisation that represent users after which the network and the fan data of those users are loaded into the system so that they become available for exploring. The reason that we do not add the data of all users that saved a bookmark automatically from the start is twofold:

   - the data is loaded incrementally which reduces the volume of data that has to be retrieved [Chaudhuri and Dayal, 1997].
   - the usage throttling and abuse monitoring software at the del.icio.us website, that we discussed in section 5.3.1, prevents us of adding this

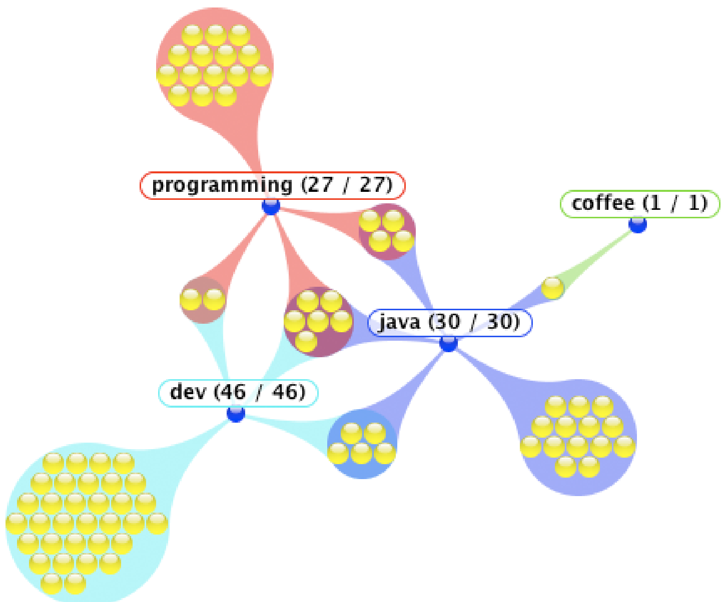



Figure 5.7: Cluster Map visualisation, showing the relationships between a number of concepts, denoted by tags.

information from the start. In the case of CALIBRATE, we do have this information and the data is therefore immediately loaded.

### 5.6.2 Example 1

The following relationships exist between the users in the underlying del.icio.us data:

- “jkofmsk” has two users in his network: “cougare” and “mmeire”.
- “cougare” has two users in his network: “phOng” and “vuorikari”.
- “phOng” on his turn has two users in his network: “pieterjelle” and “woumpousse”.

While exploring the bookmarks and tags of these users, one might want to find out if there are implicit relationships between the non-related users of them. One can easily find tags that are frequently used by ordering the list of tags by popularity. Figure 5.8 shows the visualisation of the users above and the tag “java” that is mostly used in this data. Showing this tag leads to the uncovering of an implicit community of users that are not originally in the same network of users, and thus probably are not acquainted. However, they all seem to be interested in the concept of the “Java” programming language. By adding those other users to one’s own network, one can create a sort of community-of-practice [Wenger, 1998] with a similar interest: programming in java. Establishing this community may prove useful in the future while searching for new material on this topic. This kind of visualisation therefore offers the opportunity to find content and information that could be of interest.

This example also shows that semantic closeness results in geometric closeness. If two items share many objects, they are semantically close [Fluit et al., 2005]. The user “vuorikari” did not tag a bookmark with “java”. Therefore this user is drawn further away from the item that represents this tag “java”.

### 5.6.3 Example 2

Examining the social bookmark space of the CALIBRATE portal with the cluster map visualisation allows users to get an impression of the resources that are “used” across country and language barriers. Some researchers in ARIADNE, are interested in learning resources that “travel well”. In the context of this work, we may define a learning resource as “travels well” if it satisfies one or more of the following criteria:

- an item that is bookmarked by a user is of a different language than user’s mother tongue,
- an item has tags in (a) different language(s) than that of the item language,
- an item is from a different country than where the user is from.

Travel well resources may be interesting for e.g repository owners or researchers, in seeing what kind of resources users from other countries bookmark. An example of this is shown in Figure 5.9 that gives an overview of the Hungarian learning resources that travel well: from 226 Hungarian resources, that are represented by the green color:

1. 5 are shared by users from both Poland and Estonia,
2. 15 are shared by Estonian users that are not shared by Polish users,
3. 20 are shared by Polish users that are not shared by Estonian users, and
4. 1 is shared by German users.

These hungarian resources that “travel well”, mainly cover an advanced music education area, which contains hardly any explanatory text, but mostly images and animations, which could explain why they are popular in the other countries as well.

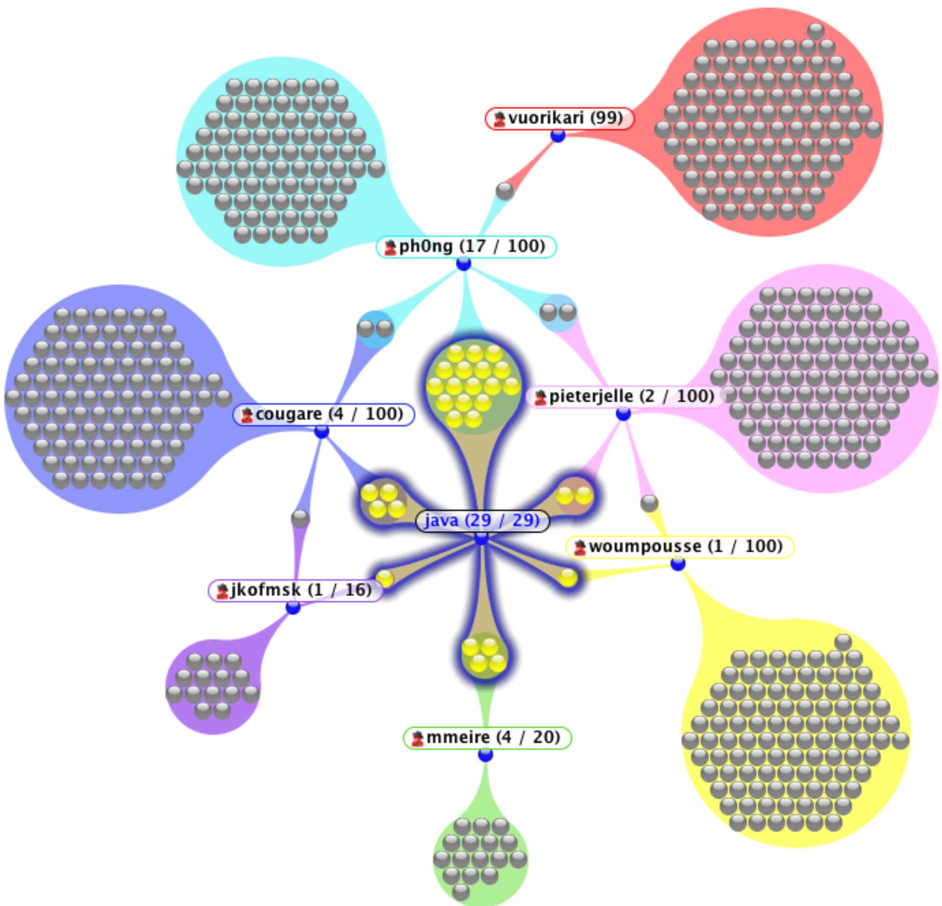



Figure 5.8: A Cluster Map visualisation showing a network of users, formed by common bookmarks that are all tagged by "java".

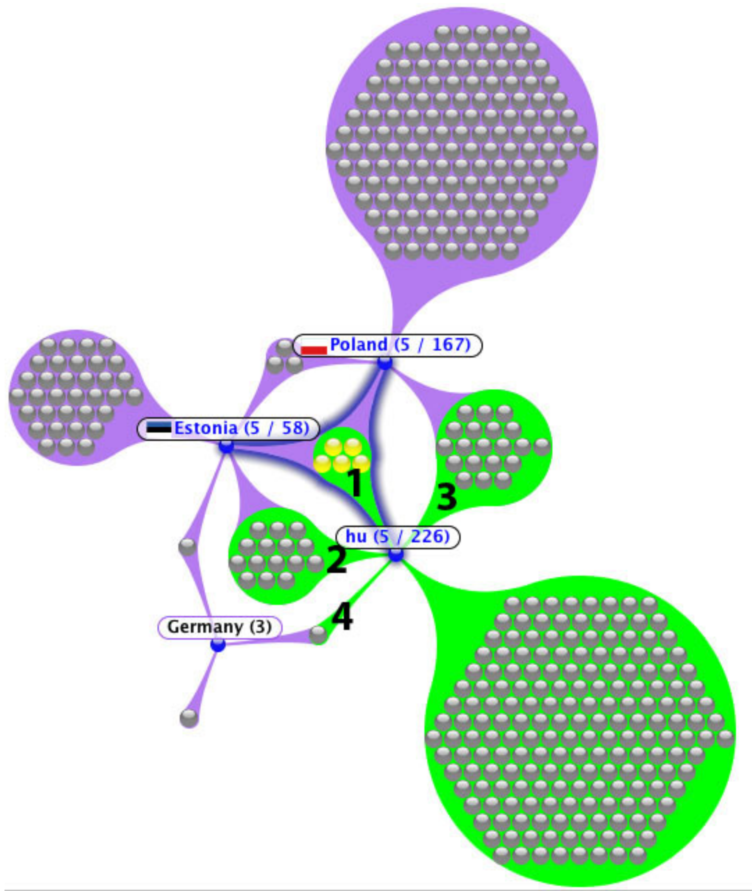



Figure 5.9: A view of Hungarian learning resources that travel well in both Estonia as in Poland.

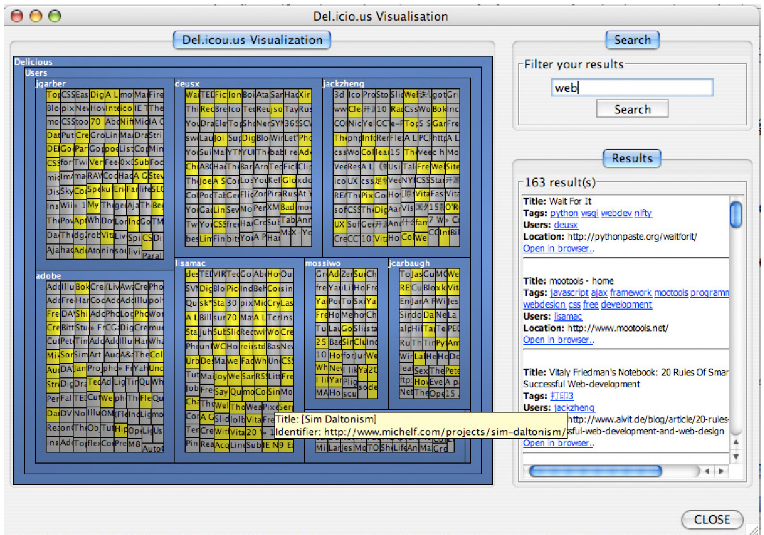



Figure 5.10: Tree Map visualisation: blue rectangles represent the users, yellow rectangles represent the bookmarks that match the selection made by the keyword web in the search box and grey rectangles are bookmarks that do not match.

## 5.7 Implementation Notes

Our prototype for visually searching and analysing del.icio.us and CALIBRATE social bookmarks is created with our open and extensible information visualisation framework that has been described in chapter 2. With this framework, it is easy to choose other visualisations for representing the bookmarks. We may for instance visualise the bookmarks with a tree-map where the bookmarks are classified per user in a flexible and efficient manner, or with any of the other techniques that our framework provides. Such a tree-map may be beneficial by providing an extra overview when exploring bookmarks by visualising all of them at once. Performing a keyword query in the filter pane on the topic "web" gives users an impression of the distribution of resources on that topic over all users. This can be seen in Figure 5.10. It makes sense to further explore those alternatives in subsequent future work.

For creating visualisation of the del.icio.us bookmarks, we plugged the Aduna Cluster Map [Aduna, 2008] software library into our framework. This library contains functionality for creating visualisations of collections of classified objects. By integrating this library in our framework, we don't have to reinvent the wheel and implement cluster map visualisations from scratch. On top of that, this library becomes available for other case studies developed with our framework as well and therefore it demonstrates the open and extensible nature of our framework.

The Contextualized Attention Metadata(CAM) framework [Wolpers et al., 2006], has been integrated into our information visualisation framework for this case-study. CAM can be used to capture the attention users spend on content in an application [Najjar, 2008b]. We gather user data of the prototype to evaluate the effectiveness and usefulness of our design, and adapt the design where needed. For instance, if we would notice from those logs that users never use the filter widget (section 5.5.1), this could mean that the functionality is:

- not clear to the users and should be explained by e.g adding a help-function,
- not found by the users in the user interface, so a better place needs to be found for it,
- not found useful by the users so we would need to find a new way of filtering items, or no filtering widget at all.

## 5.8 Discussion

We strongly believe in rapid prototyping development, where software systems are delivered incrementally and requirements analysis continues throughout the process, interleaved with implementation and evolution [Luqi, 1989]. It enables us to uncover problems like usability issues in the early stage of development. During development, we have involved a number of users to sollicit feedback. The results are described in the next paragraph. A user study with a rather complete prototype version of our visualisation has been conducted afterwards to validate the used techniques. These results are discussed in section 5.8.2.

### 5.8.1 Evaluation 1

A fairly small number of test users ($4 \pm 1$) is enough to find most usability problems [Nielsen, 1992c]. Four users, both del.icio.us users and repository owners in the CALIBRATE project, were involved to locate problems during the design of our visualisation. We found out a number of things while using the "think aloud" method during interviews with those users. The "think aloud" method is a method for user testing where the users verbalise their thoughts while using the system that is being tested [Nielsen, 1992a]. In this way, those users reveal their view on the system and possibly, their misconceptions. We learned the following issues during those sessions:

- **Portal Integration**. Our third design requirement (section 5.4) stated that we want our visualisation to be useful, efficient and effective. Among other things, this means that our visualisation needs to be tightly coupled with the normal workflow of users. A browser application is normally used to access del.icio.us and CALIBRATE for exploring the bookmark space, but our visualisation is not directly integrated in those portals. If users find interesting bookmarks with our visualisation, it should be easy for them to save and tag those bookmarks within e.g. del.icio.us. Therefore, we integrated the functionality to directly post newly found bookmarks to del.icio.us but this is not enough. People want to be able to export a number of bookmarks at once in order to save time. Adding users to one's own network on del.icio.us is a requested functionality that has recently been integrated.

- **Zooming**. This functionality could help some users while exploring the social bookmark space to keep a good overview of the visualisation. This functionality will be added to the general framework (see section 5.7) so that it will be available for this application but also for other case studies that have been created with it. It still needs to be further investigated if this would actually help in keeping a good overview. On top of that, we need to make sure that this does not cause the "lost-in-hyperspace" syndrome in the visualisation that we want to avoid (section 5.4).

- **Implementation issues**. There were some issues with the visualisation prototype. For instance, we found out that some users in the educational CALIBRATE portal bookmarked some resources and tagged them by concatenating keywords like for instance "water, pollution, water pollution", which has been used as one tag. Either our prototype or the CALIBRATE portal needs to be adapted to be able to cope with this.

- **Learning Curve**. We noticed during the think-aloud sessions that the users did not directly use all available functionality because they either didn't know how to use it, or did not know it was there. Some users therefore suggested to include a help-function or create a small video-demo to keep this learning curve as small as possible.

- **Complexity**. The more tags and users are added, the more complex a cluster map visualisation can become with overlaps between social bookmarks and different categories. This would certainly be a problem if the cluster map would be a static image. However, the user is given continuously control over the items that are currently visualised and therefore the user can decide to remove items from the visualisation at any point during the exploration of the social bookmark space.

- **Timeline**. Visualisation of social bookmarks through time would enable users to learn about the evolution of their bookmarks, tags and network. This would reflect how users' interests and focus develop over time. One idea to add time to our visualisation of social bookmark space is to represent time by moving graphics. One existing example of this is Gapminder where the authors use animated scattergrams to show statistics on world health [Rosling and Rosling, 2006].

### 5.8.2 Evaluation 2

The second evaluation that has been performed has been a subjective evaluation by using a web application: subjects were asked to fill out an online post-experimental questionnaire. First of all, they had the possibility to watch a short demo that explained the purpose of our prototype. After viewing this movie, subjects were asked to explore their del.icio.us bookmark space with our prototype. The questions in the survey were stipulated in such way to ensure that the users needed to perform a number of tasks. The goals of this evaluation were threefold: (i) to assess the effectiveness of our approach to enable exploration of social bookmarks by visualising the relationships between users, tags and bookmarks; (ii) to assess the subjective acceptance of a visualisation tool for the purpose of exploring social bookmarks; (iii) to find out possible usability issues of the prototype.

#### Survey Statements

In total, there were 15 statements:

- 11 statements to measure the effectiveness of visually exploring social bookmarks, and the subjective acceptance of this approach. To measure the subjects agreement to those statements, we used the Likert scaling method [Likert, 1932] which is a popular scaling method. All items were rated on a 1-to-5 rating scale: strongly disagree - disagree - undecided - agree - strongly agree.
- 4 open questions where users could fill in their opinion, suggestions, and usability issues.

#### Findings

Table 5.2 presents the responses of all subjects to the statements above. The mean for the level of expressiveness was higher than 4 with a standard deviation of 0.52, meaning that all subjects found the cluster map visualisation expressive for effectively exploring social bookmarks.

Finding items by using the search box has a value of 4.44 with a standard deviation of 0.72, which means that all users found this a valuable feature for exploring their bookmarks. We believe that this is related to the fact that people are used to using keyword queries for searching bookmarks in their own contexts.

Sorting tags by frequency or name with a mean of 2.55 was perceived as not easy. From the general remarks that we received as answers to our open questions, we learned that users thought that it would be a valuable and interesting feature while exploring but it should be made more clear in the interface by e.g specific buttons for this functionality.

Except for the previous statement, all others that measured the criterion of effective exploration have values higher than 3. We can therefore assess that our design has been found effective. The statement that measured the experienced efficiency of exploring the social bookmark space, has a value of 4.11 with a standard deviation of 0.92. This means that users found our design fast and efficient. All statements that measured the subjective acceptance of our visual design have values over 3.22.

We can therefore conclude this section that users considered the use of information visualisation techniques an effective and efficient means to explore social bookmark spaces.

**Recommendations**

From the open questions that we asked users we received a number of recommendations for further enhancement of the design. These are summarised below:

- The items move in the visualisation, when e.g adding new items to the visualisation. The reason for this, is that the visualisation algorithm tries to recalculate the layout of the cluster map whenever new items are added. The advantage of this is that users always see a visualisation with the most optimal layout for the current elements in the visualisation. However, this makes it sometimes difficult for users to get a grip on what is visualised because the continuity is lost for them. One way to solve this, would be to only recalculate the layout when the users chooses so. This is of course something that has to be further investigated.

- Adding an initial ranking on the tags in the tag lists so that users will not have to bother to do that.

- Adding "clear"-buttons to the search box and the filter widget of the tag lists (section 5.5.3 and 5.5.1).

- The difference between users and tags should become more clear in the visualisation. For now, users are identified by an icon and a username but we could for instance use different geometric shapes for users and tags.

## 5.9 Conclusions

This chapter presented the use of information visualisation techniques

- that enables insight to analysts in the implicit relationships between users, tags and bookmarks; and

Table 5.2: Effective Exploration and Subjective acceptance.

| Criteria | Statement | Mean | Standard deviation |
|---|---|---|---|
| Effective Exploration | Find the amount of bookmarks of a user is easy | 4.22 | 0.97 |
| | Sorting tags by frequency or name is useful and easy | 2.55 | 1.01 |
| | Adding data of users to the visualisation is easy | 3.66 | 1.00 |
| | Finding items with search box is useful | 4.44 | 0.72 |
| | Expressiveness: Relationships between bookmarks, tags and users are easy to detect | 4.44 | 0.52 |
| | Filter controls are effective in exploration process | 3.33 | 1.00 |
| | Finding popular bookmarks in the network of users, is easy | 3.55 | 0.88 |
| | Result lists easy to read | 3.77 | 0.44 |
| Subjective acceptance | Ease of use in general was easy | 3.89 | 0.33 |
| | Exploring the social Bookmark Space, is fast and efficient | 4.11 | 0.92 |
| | Users can always stay in control of the exploration process | 3.22 | 0.97 |
| | Seeing relationships between tags and users, is useful | 3.22 | 0.97 |
| | Would use the tool to explore social bookmarks more often | 3.44 | 0.88 |

- that offers end users new ways to find content and information that could be of interest to them but that would not have been found through explicit searches.

The contributions of this chapter is the answer to the following question:

- how can we effectively and efficiently provide visual access to a collection of social bookmarks?

To answer this question, we have created a tightly-coupled visualisation prototype that uses a cluster map visualisation for representing bookmarks, users and tags (Section 5.5). This prototype uses real data from del.icio.us and CALIBRATE. We chose del.icio.us as a source of data as it is highly popular with many users, lots of data and with a very easy API to access this data. Those characteristics makes del.icio.us a very good reference point (section 5.3.1) for other social bookmark tools. Two evaluations have been conducted to measure the effectiveness and the efficiency of our design (Section 5.8). From those evaluations, we learned that the use of information visualisation techniques prove useful in opening up the view on social bookmark spaces. Other possible data sources that can be used in the future, are folksonomies like Flickr [Flickr, 2008] and YouTube [YouTube, 2008], where respectively photographs and videos are described by tags.

The framework that we have created and presented in chapter 2 has been adapted while creating the prototype for the techniques above. More concretely, a library for visualising cluster maps has been plugged into the framework. On top of that, the Contextualized Attention Metadata(CAM) framework (Section 5.7), has been integrated and is therefore available for all case studies that are developed with our framework.

The previous chapter described how we enabled visual access to a collection of reusable objects, based on the “has Part / is Part Of” relationships between those objects. This chapter described how we opened up a collection, based on the implicit relationships between users, tags. The prototype that has been presented in this chapter, is publicly available at [Klerkx, 2008a] and can be freely used to explore the social bookmarks of the del.icio.us system.

# Chapter 6

# Visualising a Network of Learning Object Repositories

## 6.1 Introduction

Chapter 3 presented the use of information visualisation techniques to improve access to a single learning object repository (LOR). For this purpose, we have created a tightly-coupled, visual exploratory search prototype that uses a tree map visualisation for accessing learning objects in the ARIADNE Knowledge Pool System [Duval et al., 2001]. The prototype makes use of the LOM classification that describes the topic of the learning objects.

This chapter presents an additional case study for enabling visual access to a network of Learning Object Repositories instead of only one LOR. The Global Learning Objects Brokered Exchange (GLOBE) consortium [Globe, 2008] aims to unlock the deep web of the learning repository networks that its members maintain. However, the objectives of this chapter differ from the ones of the other chapters, where we start from a collection of reusable components and investigate the use of information visualisation techniques as a solution for enabling effective and efficient access to them. Prototypes are created with our extensible framework (chapter 2) but the framework is merely seen as a means to be able to validate the use of those techniques. The created applications have always been deployed as Java applications, either by using Java webstart or Java applets.

In this chapter, we only briefly explore the possibilities of visualising networks of LORs, because our primary objective of this case study has been the practical creation and integration of a mashup application with our framework. We wanted to know the practical implications on the architecture of the framework. A mashup is a web application that uses services from more than one source to create a new service. Services used in mashups are typically sourced from a third party via a public interface or API.

As a consequence, we offer a solution to enabling access to network of LORs, but this solution has not been evaluated by end users. However, we do believe that this solution can serve as an interesting opportunity for future work.

We will start this chapter in section 6.2 with a number of requirements for a visual exploratory search application that enables access to a network of LORs. Section 6.3 presents the prototype application that has been created with our framework. Implementation notes with the implications of creating a mashup application with our framework are outlined in section 6.4. Section 6.5 shows related work and section 6.6 concludes this chapter.

Results of this work have been published in [Klerkx and Duval, 2007a] and [Klerkx and Duval, 2007b].

## 6.2 Requirements

Each member in GLOBE has implemented a web interface for querying the GLOBE repositories. They all have the following features in common:

- Users can search in all repositories by filling out a simple query in an electronic form.
- Search results are always presented as a human readable abstract of the metadata that describes the learning object.
- Users can consult a result from a particular repository through the web interface of the repository where the result was found.

GLOBE members (Ariadne, Edna, Merlot, Nime & Edusource) do federated search in each others repositories. Federated search allows client applications to search in multiple repositories at once, taking advantage of the available metadata that these repositories maintain. The idea is to merge results from multiple repositories in the web application that end users are used to work with. Typically, only (i) the title, (ii) the repository the result originated from, and possibly (iii) a description of the result, are visualised in the result lists of these web applications.

### 6.2.1 Requirement 1

Not only can we opt to visualise the content of these repositories, but we can also look into the possibilities of visualising and modelling a landscape of interoperable LORs. These landscapes could provide users with helpful information like e.g. the physical location of a learning resource in the internet IP world. This could be used to get an idea how fast or slow it will take to get access to it in terms of bandwidth. A different kind of landscape could give information of the geographical location of the repository. By knowing this geographical location, it is possible to get an idea how easy it would be to directly contact the creator of a learning resource and possibly cooperate with him or her.

Visualising a landscape of networked repositories could also provide insight in the actual content of the federated repositories. As some repositories are used within a context, it makes sense to visualise this context. The EdMedia digital library e.g. will only store papers on educational multimedia, hypermedia & telecommunications while a learning object repository that acts as a storage layer for a Learning Management System, will probably cover more heterogeneous content. We can imagine that the user would like to see this context. We can also imagine that a user wants to see which repositories contain large numbers of learning resources on a specific topic like economy, and more specifically those repositories that would actually contain components that have been repurposed for a number of times within a context that matches the one of the user.

### 6.2.2 Requirement 2

Saving the history of a user is a functionality that resulted from usability research of user interfaces to access libraries [Dorner and Curtis, 2003]. Furthermore, this can be used to enable support systems for administration decisions. It can provide insight into the users behaviour thus enabling management to recognise needs and potential savings (e.g. cheaper learning material) earlier [Wolpers et al., 2006].

Our design should therefore include a history visualisation that provides understanding of the evolving patterns and focus of the users interest.

## 6.3 Prototype

Each repository uses another classification scheme their learning objects so that we cannot simply reuse the information visualisation techniques of chapter 3. There is also no information available on the relationships between the learning objects so that we can also not use the techniques that were presented in chapter 4. This section will therefore present GlobeMash, a mashup application that focuses on enabling effective access to the network of learning object repositories. By enabling users to

1. find learning objects within GLOBE by presenting the results on a geographical map, and
2. obtain a visual insight in the search history of users,

The following section focusses on the first challenge and section 6.3.2 on the second one.

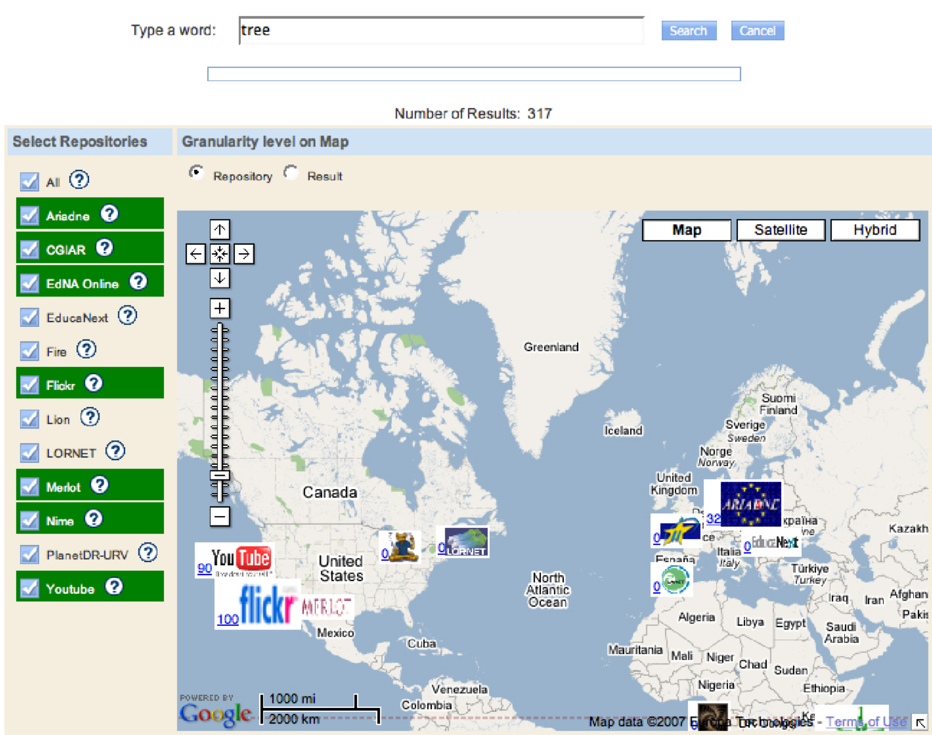



Figure 6.1: GlobeMash: A Mashup Application to Search within GLOBE.

### 6.3.1 Visualisation of Search Results

All existing interfaces in GLOBE present results of a keyword query as a list of objects, with a small number of metadata fields. These fields typically include the title, a description, the author(s) and the repository where the result was found. In figures 1 and 2, the reader can see that we integrated both such a list presentation, and on top of that, a googlemap with geographic information. In the following subsections, we discuss all functionalities of representing results in GlobeMash. Besides those presentation methods, a user can always look at the full LOM metadata instance that was returned by the Globe network, which is not always the case in existing interfaces.

#### Geographic Information of the Repository

When a user selects this level of granularity, all repositories that are searched on, are displayed on the geographical map as a logo at the geographic location of the headquarters of the organisation that maintains it. This can be seen in Figure 6.1. If the user issues a keyword query, a number is shown to the left of the icon that presents the number of search results that are returned. The size of the icon grows proportionally to the number of results as a clear visual cue. This way of presenting the results information, provides a clear visualisation of the relation between the repositories and the results. We can for instance imagine that a user wants to see which repositories contain large numbers of learning resources on a specific topic like economy. The user can click on the repository logo of his choice to only see the results of that repository in the list below the map (Figure 6.2), so the map can also be used as a filtering mechanism.

#### Geographic Information of the Individual Results

When a user selects this level, all results are indicated on the map. To be able to do so, we need a mapping from the metadata of the result to a location. We try to automatically deduce this location from the metadata of the result.

1. First of all, if present, we map the address of the institution of the author(s) through an open geolocation database.

2. If the metadata lacks the address of the institution, we look at the email address of the author(s) and map the domain in this address to an IP address. We can then again map this IP address to a location. Clearly, we cannot always do this. For instance, if the email address of the author(s) is not within the domain of his institution, but rather from a domain like "gmail.com" or "hotmail.com", the resulting location of the result will not be correct.

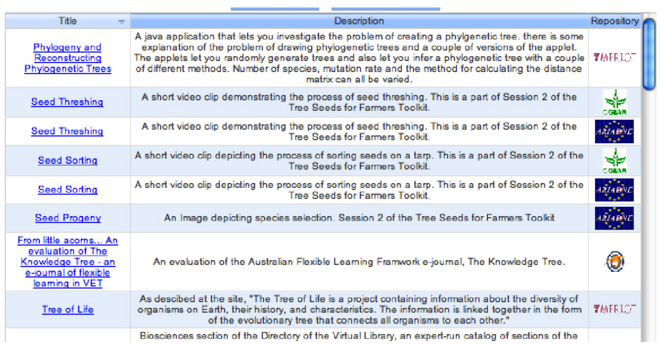



Figure 6.2: Result list presentation of GlobeMash. Results are sorted on the description field.

3. Always available in the returned metadata is the location of the description of the result on the web in the form of a handle. This location is mapped to a geolocation in case that all previous ways fail.

With all results displayed on the map, one can get an impression about the distribution of learning objects on specific topics over the world.

#### List Presentation of the Results

A table of results is always shown when the user performed a query. The columns indicate the title, a description, the author(s) and the repository where the result was found. All the results can be sorted on every column as research proofs that this is a usability requirement for searching libraries and databases [Dorner and Curtis, 2003]. This functionality is not available in the existing interfaces to find material in GLOBE. When the user selects a repository in the map, the table filters out the results of the other repositories so the list is synchronised with the geographical map.

Another usability requirement, is that a user should be able to merge results from different repositories [Dorner and Curtis, 2003] in one list. Therefore, we added functionality to enable users to create their own list with results of their interests that were returned from the network of repositories. The user can click on a result and look at the object itself without having to leave the current webpage. Results from multiple queries to the GLOBE network can be added to this list.

### 6.3.2 History Visualisation

In order to create a feedback loop that enables learning from the way people actually use the search interface within GLOBE, it is essential to track the behaviour of users with the web application. It is needed to answer questions like

- Which keywords are used how many times and could therefore be interpreted as representative keywords for a time period?
- Which keywords are used that have many or no results?
- Which keywords have results in multiple repositories?
- Which keywords or topics have the attention of a particular user over time?
- What is the behavior and the focus of the community within GLOBE?
- How many results does a user look at in detail after performing a query?
- When did a user perform a query and which keywords did he or she use?
- What is the evolution of search keywords over time?
- etc.

These questions are relevant for the end users and researchers that are interested to get an insight in the evolving patterns and focus of the users interest, or to use the users search history for e.g. recommendation purposes or to establish social networking opportunities. Every search event in GlobeMash is therefore logged in Contextualized Attention Metadata (CAM) [Najjar, 2008b]. A visualisation of this data is created to enable users to get an insight in the questions above in a flexible, efficient and browser-based way.

Our visualisation is made up of two synchronised parts; an extended tag cloud and a timeline that indicates the different events that are tracked by the Attention Metadata and the current interval of time that is visualised in the tag cloud. These parts can be seen in Figures 6.3 and 6.4. In the next paragraphs, we describe these parts in detail.

### Extended Tag-Cloud

Tag Clouds are visual presentations of a set of words, typically a set of tags, in which attributes of the text such as text, weight or color can be used to represent features (e.g. frequency) of the associated terms [Halvey and Keane, 2007]. When a user clicks on a tag, the user obtains an ordered list of tag-described resources, as well a list of related tags [Hassan-Montero and Herrero-Solana, 2006]. Usually, the tags are displayed in alphabetical order.

In our visualisation, every search keyword that is used in GLOBE is represented as a tag in the Tag Cloud. If we would use an alphabetical order to place the keywords, users would not be able to get an insight in questions like

- Which keywords are used that have many or no results?

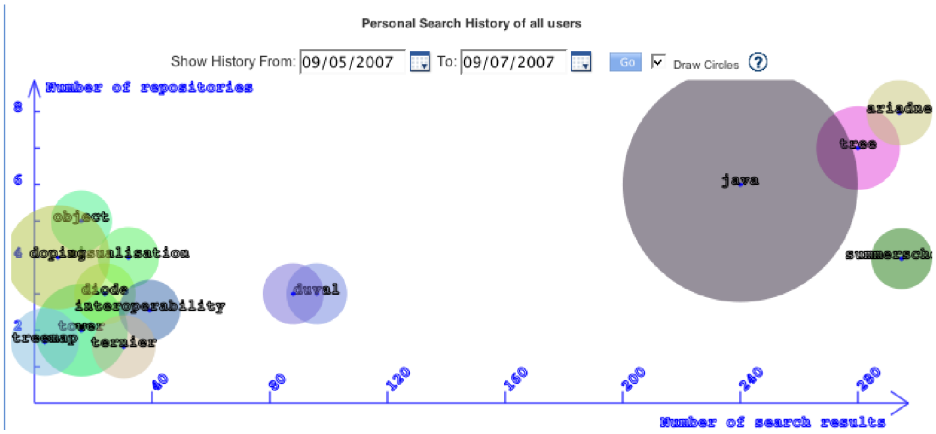



Figure 6.3: A Tag Cloud visualisation of Users Search history

- Which keywords have results in multiple repositories?

Therefore, tags are not ordered alphabetically, but in a coordinate system where the X-axis represents the average number of search results and the Y-axis the number of repositories where results were found, both in relation to the performed keyword query. The scale of the x-axis and the y-axis are automatically computed from the data that has to be drawn in the visualisation.

On top of that, the visualisation offers the functionality to draw circles around a tag or keyword. The radius of the circle represents the frequency of the keyword used in GlobeMash over time. The color encoding of these circles can be random or it can be used to encode the number of different users that actually used the keyword query, depending on the selection of the user. Another possibility to encode colors, would be the number of results that were downloaded and used by e.g. a teacher in his course.

### Timeline

The tag cloud visualisation gives an overview of the keyword used in queries in the federated network of repositories. However, it does not show when a keyword is used. Therefore we integrated a timeline that we synchronised with the tag-cloud visualisation. On the timeline, all keyword queries are displayed at the exact time and date that they were issued. The timeline consists of three synchronised bands with different time scales: daily, monthly and yearly. The band with the daily scale shows the actual keywords. The other bands are meant to give an overview when a search event happened in GlobeMash.

In this way, users can get an idea of the topics he was interested in over time.

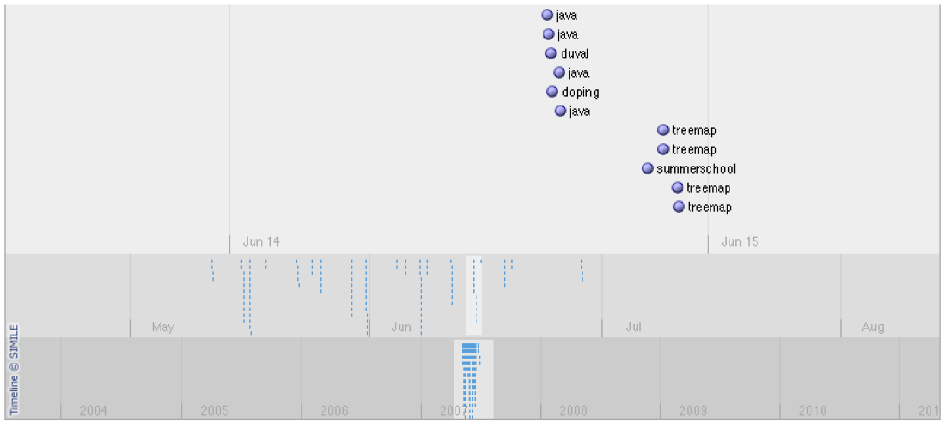



Figure 6.4: A Timeline visualisation of users' search history

If a user scrolls the timeline to a specific date, he can choose to only show the keywords from a start date to the chosen date on the visualisation. This can be used to filter out keywords that were used in a time period that does not have the interest of the user. If a user clicks on a keyword in the tag-cloud visualisation, the timeline will smoothly scroll to the last timestamp that the keyword was used.

#### Interaction

We synchronised the tag cloud and the timeline by using JavaScript. An end user can explore the visualisation and therefore get insight in the questions of section 6.3.2 by either performing actions on the tag cloud or on the timeline. If a user clicks on a tag in our tag cloud, the timeline scrolls automatically to the latest timestamp that the keyword was used in a query. If a user clicks on a keyword in the timeline, a pop up is displayed with detailed information

- when the keyword was used,
- how many results there were, and
- how many results were looked at in detail.

If a user double-clicks on a keyword in the timeline, the timeline serves as a means to select an interval of time that should be visualised on the tag cloud. The tag cloud will filter out all keywords that were not performed in that specified interval.

### 6.3.3 Findings and Recommendations

- A thorough evaluation has not been conducted in the scope of this dissertation and has been moved to possible future work. A usability test however, would allow us to evaluate GlobeMash against a number of usability criteria that were defined in [Dorner and Curtis, 2003]. The results can be used to add new features or leave out the ones that are not useful. Other visualisation techniques should also be investigated in more detail.

- Users should be able to easily see the history of himself, but also the history of all other users. Therefore, if a user does not login, the history of all users of the last time period is visualised without giving details on which query belongs to which user. If the user does login, he can see his own history. With the current software, we are already able to do this.

- To make GlobeMash a social tool, a user should be able to publish his extended tag-cloud to other users like people can do in the social bookmarking tool del.icio.us [delicious, 2008]. There, they can publish their tag-cloud that is based on their bookmarks.

- Obviously, if many queries were performed, it is possible that the tag cloud gets cluttered as it is possible that some tags overlap. In this section, we describe a number of countermeasures to avoid this problem.

    1. A first and rather obvious way to avoid clutter, is to give users the opportunity to select a small enough interval of time. A small enough interval of time means a smaller number of keywords or tags, and therefore there is less chance of overlapping keywords. This can be done by using the timeline or by filling in a small electronic form with two dates that define the interval.

    2. Secondly, the user is now able to select an area of interest in the tag cloud, after which the tag cloud is zoomed in at that area. Zooming in, has the advantage that the coordinate system can be rescaled and therefore overlapping keywords can be pulled away from each other. This can be seen in Figure 6.5.

    3. A last way of reducing clutter is the notion of only showing interesting data [Dubinko et al., 2007]: only show the tags during a particular period of time that are most representative for that time period. This method is not yet implemented in GlobeMash, but is certainly a worthwhile idea for future work.

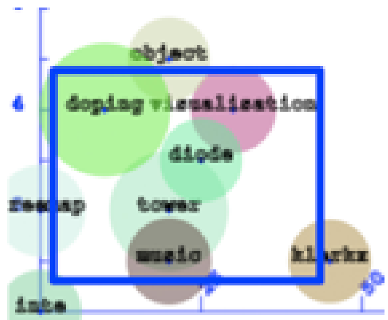



Figure 6.5: Selection of an Area of Interest

## 6.4 Implementation Notes

### 6.4.1 Objectives

Our objective was to design and implement a mashup web application that

- is modular and integrated with our extensible framework for visualising collections of reusable components,
- uses standardised data formats (LTSC IEEE LOM [IEEE, 2002], XML [XML, 2006], JSON [JSON, 2006], SVG [SVG, 2003] and CAM [Wolpers et al., 2006]), and
- uses standardised protocols (HTTP, SOAP).

This ensures that different parts of GlobeMash can also be used in other contexts than GLOBE. The history visualisation can e.g. be used in other tools that collect attention metadata.

### 6.4.2 Visualisation of Search Results

To create GlobeMash, we integrated the Dojo toolkit into our framework. This is an open source DHTML toolkit written in JavaScript. It allows developers to easily build dynamic capabilities into web pages [Dojo, 2006].

The federated search layer of GLOBE is accessed when a user issues a keyword query. We integrated this by adding a web service to our framework that uses the CEN/ISSS Simple Query Interface (SQI) [Ternier, 2008] for querying the repositories within GLOBE. Whenever a query is issued in GlobeMash, the following steps are performed:

- An AJAX request is sent from the browser client to the web service where our framework is deployed.

- The request is parsed, and a query is issued to the federated search layer by using SQI. The results from this layer are formatted in the IEEE LTSC LOM. The utilities in our framework (Chapter 2) convert these LOM results into objects of “ReusableComponent”.
- In the next step, those components are mapped into the general glyph model. For visualising the components on the the geographical map, we need the geographical locations of the results. Those are found by performing three look-up requests in GeoIP [GeoIP, 2007], NetGeo [Moore et al., 2000] and Hostip [hostIP, 2007]. By performing multiple requests, it is more likely to get a valid location as the three geolocation databases together have a better coverage of the world.
- The system core will then send those results in JSON format to the client of the user where those objects are added to the geographical map and the list presentation. The map is rendered using the Google Maps API [GoogleMaps, 2007].

### 6.4.3 History Visualisation

The history visualisation consists of two parts. The upper part is a Scalable Vector Graphics [SVG, 2003] drawing that is rendered by the Dojo framework in the client browser. The timeline visualisation part is created using the SIMILE project and is a DHTML-based AJAX widget for visualising time-based events [SIMILE, 2003]. We synchronised the SVG visualisation and the timeline widget by using the javascript event system.

### 6.4.4 Attention Metadata

All events in GlobeMash are logged in the Contextualized Attention Metadata (CAM) specification. They are transferred in JSON format from the browser client to a second web services that has been added to the framework. This web service will first map the JSON events to CAM and afterwards, it sends the created attention metadata to CAM services. This task is solved by using the software, that we created to support CAM while developing a prototype for enabling visual access to social bookmarks spaces (see section 2.4.2).

### 6.4.5 Framework Implications

Creating the mashup (web) application with our framework involved the following:

- a web service has been added to the framework that creates the necessary data and glyph models.

- a second web service has been added to create attention metadata and send it to CAM services.
- a dynamic web page has been created with the help of the DOJO toolkit that uses AJAX requests to communicate with those web services, and that visualises the results in the browser.
- adding utilities for creating and parsing of data in JSON format.

## 6.5 Related Work

All existing interfaces in GLOBE present results of a keyword query as a list of objects, with a small number of metadata fields. Our work distinguishes itself from those interfaces

- by displaying the repositories on a geographical map,
- by adding a visualisation of the history of the keyword queries of the users, and
- by adding functionalities that improve the usability of search.

For the history visualisation, a tag cloud is used. This is a rather new interface paradigm that quickly gained popularity in Flickr [Flickr, 2008] where it was necessary to find a way to represent vast amounts of information. Variations on tag clouds are presented in [Bielenberg and Zacher, 2006] where a circular form is used for a tag cloud. In [Dubinko et al., 2007] a visualisation of tags over time is presented for the Flickr system.

Finding the ideal tag cloud is a hot topic of interest, which is proved in [Hoffman, 2006] where the author starts a search for the Perfect Tag Cloud. The history of the searches of users has been the interest of research for a long time. One example is Google Zeitgeist or its recent successor Google Hot Trends [GoogleTrends, 2006] that shows interesting patterns and queries of users in a list presentation.

## 6.6 Conclusion

The main objective of the described work in this chapter, has been the practical creation and integration of a mashup application with our framework. The contributions of this chapter are the following:

- A solution is provided to enable visual access to a distributed network of learning object repositories. This solution has not been evaluated but we believe that this solution can serve as an interesting opportunity for future work.

- Our framework is extended to enable the creation of web applications that uses information visualisation techniques.

In chapter 3, we created an open and extensible information visualisation framework, which we used to access the Ariadne Knowledge Pool System in Chapter 3. This chapter broadened our scope by moving from one LOR towards opening up a collection of repositories, namely GLOBE. The created web application distinguishes itself from existing interfaces by displaying the repositories on a geographical map, by adding a visualisation of the history of the keyword queries of the users and by adding functionalities that improve the usability of search by e.g. enabling users to merge results from multiple repositories. The following chapter will present general conclusions about the work in this dissertation.

# Chapter 7

# Conclusion

In this dissertation, we have investigated how we can provide flexible, efficient and effective access to large collections of reusable components and hence, enable insight in the contents of such a collection with the help of information visualisation techniques. This work is an engineering dissertation. Therefore we have focused more on the technical aspects of using information visualisation techniques and less on the more psychology aspects of users that interact with those techniques. This chapter concludes this dissertation with a summary of contributions, and an exploration of the the potential it offers for future research.

## 7.1 Extensible Framework for Visualising Collections of Reusable Components

The created software framework that has been presented in chapter 2,

- enables fast experimentation with (i) different visualisation techniques on the one hand, and with (ii) multiple collections of reusable components on the other hand. Those components may be distributed over different domains and described with any possible metadata standard.
- is structured to support the visualisation pipeline or the computational process of converting a collection of reusable components into a visual form that users can interact with (see Section 2.4.7).
- supports the LTSC IEEE LOM standard [IEEE, 2002] and the OAI-PMH protocol [Sompel et al., 2004]. Therefore, any collection of reusable components that supports LOM and allows the harvesting of their metadata with this protocol, can be visualised by changing a configuration file (see Section 2.4.6).

- makes use of the Lucene search engine [Lucene, 2008] to optimise text search within the collection of reusable components. By using this third-party open source search engine library, written in Java, we boost the performance and the scalability of the framework which are important framework requirements.
- has been used in several case studies to create visualisations of different collections of reusable components (see Table 2.2).
- has several integrated information visualisation libraries and toolkits (see Table 2.1) that can be used for future research and case studies.
- has been evaluated by quality attributes that provide the means for measuring the fitness and suitability of a system (see Section 2.6).

A number of case studies have been worked out by using this framework. Those case-studies have been used to validate the hypothesis of this dissertation that information visualisation techniques can provide flexible, efficient and effective access to large collections of reusable components. The following sections will elaborate on the different contributions of each of them.

## 7.2 Flexible Access to Reusable Components

To show that information visualisation techniques can help to get flexible access to collections or reusable components, we adopted a case study methodology. Each case study in the next sections shows how we used well-chosen visualisation techniques to enable access to different collections of reusable components.

### 7.2.1 Visualising a Learning Object Repository

The case study in chapter 3, has presented the use of information visualisation techniques

- that offers users new ways to find content and information in a learning object repository (LOR) that could be of interest to them.

The contributions of this case study is the answer to the following question:

- how can we effectively and efficiently provide visual access to a collection of learning objects?

To answer this question, we have created a tightly-coupled, visual exploratory search prototype that uses a tree map visualisation for representing learning objects (Section 3.6.1). Our prototype supports the "visual information-seeking mantra": overview first, zoom and filter, then details on demand.

Real data has been visualised by coupling our prototype to the ARIADNE Knowledge Pool System (Section 3.4). Two evaluations have been conducted to measure the effectiveness and the efficiency of our design (Section 3.7). From those evaluations, we have learned that the use of information visualisation techniques prove a useful alternative to more traditional ways of accessing LORs. However, a number of recommendations have been found that need to be taken into consideration like adding navigational cues. Other possible learning object repositories can be easily visualised in the future if they support LOM and OAI-PMH.

Our software framework has been adapted while creating the prototype for the techniques above. More concretely, we have integrated:

- a library with tree-map visualisation algorithms,
- a library with hyperbolic tree visualisation algorithms,
- support for the OAI-PMH protocol [Sompel et al., 2004] so we can easily visualise other LORs if they support LOM and OAI-PMH

These libraries are therefore available for all case studies that are developed with our framework.

The prototype that has been presented in this chapter, is publicly available at [Klerkx, 2008b] and can be freely used to explore the ARIADNE Knowledge Pool System. This prototype is nowadays used in ARIADNE and the ZoEp-proejct in the Netherlands. A customised version is currently used in the ARISTO music company [AristoMusic, 2008] where it is used to control the quality of metadata of songs.

### 7.2.2 Visualising Actual Reuse of Reusable Components

Chapter 4 reported on an interactive visualisation techniques that aims to enable visual access to a large repository of small reusable content components that were created by disaggregating legacy content. The goal has been to provide researchers who are interested in reuse of content, with a means to improve their understanding of reuse patterns within the repository.

The contributions of this chapter are the answers to the following questions:

- how can we effectively and efficiently support visual access to a collection of small learning objects, based on the relationships between those components? For this question, we created a node-link graph visualisation that represent all components in one overview (section 4.4)
- how can we effectively and efficiently interact with visualisations of learning objects? For this question, we integrated a number of interaction techniques like dynamic query controls, excentric labeling, fisheye views and previews (section 4.5).

An expert evaluation has been conducted (Section 4.6.2). This has shown us that users considered the visualisation techniques for accessing a repository of fine-grained components effective.

Our software framework that we have created and presented in chapter 2 has been adapted while creating the prototype for the techniques above. More concretely, we have integrated:

- INRIA's infovis toolkit, which is an interactive open source toolkit with a number of visualisation techniques like excentric labeling (Section 4.5.1). On top of this, it contains an integrated software bridge in Java to the GraphViz [Ellson et al., 2003] toolkit, which essentially is a set of graph drawing algorithms and programs written in C.

- the Lucene search engine [Lucene, 2008], which we have used to optimise text search within the collection of reusable components. By using this third-party open source search engine library, written in Java, we boost the performance and the scalability of the framework.

These libraries are available for all case studies that are developed with our framework.

The framework that we have created and presented in chapter 2 has been adapted while creating a prototype for the techniques above. More concretely, another library, has been integrated into the framework with new information visualisation techniques. Those are now available for all case studies that are created with the framework.

The previous chapter has described how methodologies and technologies for information visualisation can be used to effectively and efficiently support visual access to a collection of learning objects, based on their topic that is defined in the LOM classification. This chapter described how we can enable visual access, when there is no LOM classification available for the learning objects. In this case, we have used the "hasPart / isPartOf" relationships between the objects, as a starting point. The next chapter will discuss how we can support visual access to a network of learning object repositories.

### 7.2.3 Visualising Social Bookmarks

Chapter 5 has presented the use of information visualisation techniques

- that enables insight to analysts in the implicit relationships between users, tags and bookmarks,

- that offers end users new ways to find content and information that could be of interest to them but that would not have been found through explicit searches.

The contributions of this chapter are the answers to the following question:

- how can we effectively and efficiently provide visual access to a collection of social bookmarks?

To answer this question, we have created a tightly-coupled visualisation prototype that uses a cluster map visualisation for representing bookmarks, users and tags (Section 5.5). This prototype uses real data from del.icio.us and CALIBRATE. We have chosen del.icio.us as a source of data because it is highly popular with many users, lots of data and with a very easy API to access this data. Those characteristics make del.icio.us a very good reference point (section 5.3.1) for other social bookmark tools. Two evaluations have been conducted to measure the effectiveness and the efficiency of our design (Section 5.8). From those evaluations, we have learned that the use of information visualisation techniques proves useful in opening up the view on social bookmark spaces. Other possible data sources that can be used in the future, are folksonomies like Flickr [Flickr, 2008] and YouTube [YouTube, 2008], where respectively photographs and videos are described by tags.

The framework that we have created and presented in chapter 2 has been adapted while creating the prototype for the techniques above. More concretely, we have integrated

- a library for visualising cluster maps has been plugged into the framework, and
- the Contextualized Attention Metadata(CAM) framework (Section 5.7), which makes it possible to capture the attention users spend on content in all case studies that are developed with our framework.

### 7.2.4 Visualising a network of Learning Object Repositories

The main objective of the described work in this chapter, has been the practical creation and integration of a mashup application with our framework. The contributions of this chapter are the following:

- A solution is provided to enable visual access to a distributed network of learning object repositories. This solution has not been evaluated but we believe that this solution can serve as an interesting opportunity for future work.
- Our framework is extended to enable the creation of web applications that uses information visualisation techniques.

In chapter 3, we created an open and extensible information visualisation framework, which we have used to access the Ariadne Knowledge Pool System in Chapter 3. This chapter has broadened our scope by moving from opening up a single LOR towards a collection of repositories, namely GLOBE. The created web application distinguishes itself from existing interfaces by

- displaying the repositories on a geographical map,
- by adding a visualisation of the history of the keyword queries of the users, and
- by adding functionalities that improve the usability of search by e.g. enabling users to merge results from multiple repositories.

## 7.3 Further Research Topics

For each case study that has been presented in this dissertation, a number of recommendations have been described. Besides these recommendations, there are also several more general avenues and issues for future work:

- Framework validation: Our framework has been used in several case studies and has several integrated third-party libraries which already shows its extensibility. However, integrating even more libraries and more information visualisation techniques will further validate the fitness of our framework for its intended purpose. By creating more case studies with our framework, we will be able to further validate and polish the architectural design.
- Framework translation: For the time being, our framework has been developed in Java™. It would be interesting to reimplement it in other technologies like .NET or Adobe Flex/AIR/Flash. These implementations would enable comparative research between those technologies for enabling access to reusable components with the help of information visualisation techniques.
- Visualisation for the masses: This is a highly popular information visualisation topic at the time of writing. The barrier between ordinary end users and implementing visualisations of their data should be low. Creating of new case studies in our framework requires knowledge about programming in Java™and therefore, the barrier is not low. Creating an high-level interface for normal end users would therefore be an interesting future direction of the research.
- Quality Measures: Some attempts have been made in research to find quantiable quality measures for visualisation and interaction algorithms. Finding such measures would enable usability studies to evaluate the fitness of the proposed visualisation algorithms for its problems. Therefore, it would also simplify the development of proposed solutions.

- Scalability: A number of case studies have been developed with lots of reusable components that have to be visualised. Some tactics were followed to achieve the support of these numbers of components like (i) the integration of Lucene indexes to optimise search, and (ii) the use of well-tested third-party libraries for rendering those components on screen. In ARIADNE, we are involved in the Global Learning Objects Brokered Exchange consortium which aims to unlock the deep web of the learning repository networks that its members maintain. Millions of reusable components will be become available for visualisation in this context. Creating case studies for such numbers is quite an interesting and challenging direction of the research.

- Visualisation of the reusable components: In this dissertation, we use metadata that describe the reusable components to visualise and enable flexible access to them. However, investigating how we can visualise the inner contents of such components, and how to make this clear to end users, may provide added value while searching for such components. Imagine for instance a novel. We could integrate the research of the authors in [Keim and Oelke, 2007] and create a visual fingerprint of the novel that enables users to gain a deeper understanding of the book in one glance.

# Chapter 8

# Informatie Visualisatie: Ontsluiten van Herbruikbare Componenten

## 8.1 Inleiding

De afgelopen jaren zijn er talrijke informatiebronnen verschenen op het internet, met grote hoeveelheden herbruikbare materialen en bijhorende metadata. Voorbeelden hiervan zijn Wikipedia, sociale bookmarking tools zoals del.icio.us, collecties met leermateriaal zoals de Ariadne Knowledge Pool System, enz. De meeste van de instrumenten voor het verkrijgen van toegang tot die middelen, zijn gebaseerd op elektronische formulieren waarmee eindgebruikers booleaanse combinaties van zoekcriteria kunnen formuleren. Eindgebruikers vinden het echter vaak moeilijk om dergelijke combinaties adequaat te definiëren [Marchionini, 2006a]. Er is dus een duidelijke behoefte om de toegang tot deze repositories te verbeteren.

informatie visualisatie is een groeiend onderzoeksgebied waarbij een visuele representatie van abstracte informatie omgevingen gebruikt wordt, voor het versnellen van hun assimilatie en begrip [Andrews, 2006]. Dit proefschrift onderzoekt hoe we betere toegang kunnen voorzien tot zulke informatiebronnen met de hulp van informatie visualisatie technieken. In dit werk worden vier case studies beschreven waarbij telkens een werkende proof-of-concept werd geïmplementeerd voor de validering van deze technieken.

We starten dit hoofdstuk in 8.1.1 met een beschrijving van de context van dit proefschrift. Secties 8.1.2 en 8.1.3 behandelen de onderzoeksvragen van dit werk. Secties 8.2 tot en met 8.5 beschrijven de verschillende case studies die uitgewerkt werden. Sectie 8.6 sluit dit hoofdstuk af met enkele vertrekpunten voor toekomstig werk.

### 8.1.1 Leermateriaal

Dit gedeelte schetst de context waarin dit werk werd uitgevoerd: "Learning Object Research'. Een "learning object" of een leerobject is volgens de definitie "any entity, digital or non-digital, that may be used for learning, education or training" [IEEE, 2002]. Leerobjecten worden vaak beschreven met uitgebreide gestandaardiseerde metadata. IEEE LTSC Learning Object Metadata (LOM) is de belangrijkste standaard voor de beschrijving van leerobjecten (LOs) [Wiley, 2007]. Deze standaard bevat kenmerken over leerobjecten zoals de titel, de auteur, de eigenaar, het formaat, de interactie stijl, enz. LOM maakt onder andere het delen van zulke beschrijvingen mogelijk, zodat eindgebruikers zoals leraars en studenten makkelijker toegang kunnen krijgen tot het benodigde leermateriaal.

LOs kunnen worden opgeslagen in databases of "Learning Object Repositories". Er bestaan verschillende voorbeelden van Learning Object Repositories zoals EDNA, MERLOT, NIME, KERIS, LACLO, etc. Wij hebben ons echter geconcentreerd op het ARIADNE Knowledge Pool System. Dit is een gedistribueerde repository van leermateriaal en bijhorende metadata [Duval et al., 2001]. Een grondig overzicht van andere voorbeelden van LORs zijn opgenomen in [Verbert et al., 2008] en [Ternier, 2008].

### 8.1.2 Flexibele Toegang tot Herbruikbare Materialen

Duval & Hodgins hebben een aantal samenhangende onderzoekthema's vooropgesteld [Duval and Hodgins, 2003]. Deze punten zijn belangrijk voor het verkrijgen van hergebruik van LOs op een wereldwijde schaal. Dit gedeelte schetst het probleem van flexibele toegang verkrijgen tot LORs met enorme hoeveelheden leermateriaal.

Wanneer een applicatie een query stuurt naar een LOR, worden de resultaten teruggestuurd in LOM formaat. Applicaties kunnen profiteren van het gestructureerde karakter van de metadata en op basis van deze metadata het juiste leermateriaal kiezen dat is aangepast aan de behoeften van de eindgebruikers. Bijvoorbeeld, als een leraar op zoek gaat naar een diapresentatie van 15 minuten over overerving in object-georiënteerd programmeren, met een doelgroep van universitaire studenten, dan kan hij of zij een geavanceerde query opstellen. Deze query kan worden vergeleken met de LOM beschrijvingen van de objecten.

De meeste applicaties om toegang te krijgen tot LORs zijn op dit moment gebaseerd op een elektronisch formulier dat de eindgebruikers in staat stelt om booleaanse combinaties van zoekcriteria te formuleren. Een andere manier is het invullen van een eenvoudig tekstvak, maar die manier profiteert niet van het gestructureerde karakter van LOM. Queries die op deze manier geformuleerd worden, zijn vaak niet doeltreffend genoeg om te voldoen aan alle eisen [Marchionini, 2006a]. Eindgebruikers moeten snel kunnen inzoomen op de juiste leerobjecten die voor hen relevant zijn. De gebruikers moeten dit materiaal kunnen vinden zonder dat ze het langdurig proces moeten doorlopen van het formuleren van complexe zoekcriteria, het evalueren van enkele van de resultaten, het verfijnen van de zoekcriteria, enz..

Een mogelijk alternatief paradigma voor toegang te krijgen tot LORs is informatie visualisatie [Card et al., 1999]. Hierbij kunnen eindgebruikers door het manipuleren van controles op de metadata, inzoomen op potentieel relevantere LO's. Tijdens dit proces, hebben ze continu een overzicht van hoe aanvullende zoekcriteria het resterende aantal LO's beperken. Human-computer information retrieval (HCIR) is de studie van information retrieval technieken die menselijke intelligentie in het zoekproces brengt [Marchionini, 2006c]. Het doel van HCIR is het integreren van de mens en systeem interactie, zodat de eindgebruikers tijdens het zoekproces voortdurend voorzien worden van zinvolle informatie, waardoor de integratie van begrip & vinden verbeterd wordt [Marchionini, 2006b]. Binnen HCIR, noemt men dit proces exploratief zoeken.

In dit proefschrift onderzoeken we dus het gebruik van informatie visualisatie technieken voor het aanbieden van flexibele en efficiënte, exploratieve toegang tot repositories met herbruikbare materialen.

### 8.1.3 Een Uitbreidbaar Software Raamwerk voor het Visualiseren van Herbruikbare Materialen

In het kader van dit onderzoek willen we geen beperking van de reikwijdte van ons onderzoek. We willen dus niet enkel flexibele toegang aanbieden tot slechts één verzameling, maar we willen in staat zijn om toegang tot meerdere collecties aan te bieden. Bovendien willen we verschillende en bestaande soorten informatie visualisatie technieken gebruiken ter verbetering van de toegang tot zulke collecties.

Een onderzoeksdoel is dan ook het ontwerp en de implementatie van een software raamwerk dat het mogelijk maakt om snel te experimenteren met visualisatie technieken aan de ene kant en met verschillende soorten collecties, verdeeld over verschillende domeinen, aan de andere kant. Deze infrastructuur moet ons in staat stellen om visualisaties te creëren die efficiënt, effectief, flexibel en schaalbaar zijn. Veel algoritmen voor visualisatie technieken bestaan al. We willen uiteraard niet het wiel opnieuw uitvinden, en daarom willen we een ontwerp opstellen voor de integratie van deze algoritmen in deze infrastructuur.

Merk op dat, hoewel het uitwerken van zulk een raamwerk een grote uitdaging is, dit slechts een middel is voor het uitvoeren van het onderzoek dat werd besproken in de vorige paragraaf. Voor dit raamwerk hebben we gekozen voor een architectuur die in de lijn ligt van "The Visualisation Pipeline" [Card et al., 1999]. Dit is het proces van het omzetten van informatie in een visuele vorm waarmee de gebruikers kunnen interageren. Deze architectuur wordt besproken in sectie 2.4.7.

Belangrijke aspecten van deze architectuur zijn uitbreidbaarheid, schaalbaarheid en efficiëntie.

Voor dit proefschrift opteren we voor een "case study" methode. In de volgende secties wordt elke studie achtereenvolgens besproken. Zij bespreken hoe we toegang voorzien tot verschillende soorten van repositories door gebruik te maken van informatie visualisatie technieken. Elke studie toont de uitbreidbaarheid van ons raamwerk aan. De eerste studie die we bespreken, is het visualiseren van een LOR.

## 8.2 Visualiseren van een LOR

Een voorbeeld van een leerobject met geassocieerde LOM metadata, wordt getoond in figuur 3.1. Typische eindgebruikers zoals leraren en studenten, zoeken leermateriaal op basis van het onderwerp van het leermateriaal [Najjar, 2008a], [France et al., 1999]. Dit doen ze dikwijls door te zoeken in classificatie of directory structuren zoals bijvoorbeeld de Google Directory [GoogleDirectory, 2008]. Het classificatie-element dat is opgenomen in LOM, is daarom een goed startpunt voor het creëren van een visueel overzicht van de repository. Een voorbeeld van een classificatie van een leerobject over de "Torens van Hanoi" kan zijn:

- Exacte Wetenschappen;
- Informatica;
- Algoritment;
- Recursie.

### 8.2.1 Visueel Overzicht van een LOR

Shneiderman heeft een taxonomie opgesteld van 7 data types voor visualisatie omgevingen [Shneiderman, 1996]. Voor elk van deze types zijn er visualisatie technieken en taken ontworpen. De classificatie structuur in LOM is hierarchisch gestructureerde "tree-data". Voor het visualiseren hiervan, hebben we drie technieken onderzocht:

- **Squarified Cushion Tree-map**; Een visualisatie van een hierarchische structuur die 100% gebruik maakt van de beschikbare ruimte op het scherm. Het maakt een volledige mapping van de hierarchische LOM classificatie naar een rechthoekig gebied op een ruimte-vullende manier [Shneiderman and Johnson, 1991]. Dit wordt getoond in Figuur 3.2 waar een kleine deelverzameling van de classificatie wordt afgebeeld op een tree-map. De groote van een rechthoek dat een bepaalde classificatie voorstelt, is recht evenredig met de grootte van de overeenstemmende leerobjecten die geclassificeerd werden met deze classificatie.

- **Hyperbolic Tree**; Een focus+context visualisatie techniek die gebaseerd is op een hyperbolische geometrie [Moise, 1974] voor het visualiseren en manipuleren van grote hierarchieën [Lamping and Rao, 1996]. Deze techniek biedt een continu overzicht van de classificatie structuur, gecombineerd met een gedetailleerd zicht op een gedeelte van de classificatie. Binnen deze visualisatie kan de gebruiker navigeren door te klikken in de visualisatie. Een voorbeeld kan men zien in figuur 3.5.
- **Venn Diagram**; Een visualisatie techniek die de onderliggende relaties visualiseert tussen groepen objecten [Venn, 1880], [Michard, 1982]. Een voorbeeld van de classificatiestructuur als een Venn diagram kan men zien in 3.7.

Na een vergelijkend onderzoek tussen bovenstaande technieken, werd er gekozen om het overzicht van de LOR te creëren met een tree-map visualisatie. Bij dit onderzoek hebben we gekeken naar navigatie, gebruikerstudies, nuttig ruimte- en kleurgebruik en de mogelijkheid om (i) extra niveaus, (ii) visuele hints en (iii) tekstuele informatie toe te voegen. Een gedetailleerde beschrijving van dit onderzoek hebben we opgenomen in sectie 3.5.5.

### 8.2.2 Ontwerp

Voor het valideren van de gebruikte visualisatie technieken, werd een werkend prototype gebouwd bovenop het Ariadne Knowledge Pool Systeem. Dit protoype bevat een

- **visualisatie** paneel waar de tree-map getoond wordt die een overzicht geeft van het leermateriaal in ARIADNE (sectie 3.6.1),
- **dynamisch query controle** paneel waar filters gemaakt kunnen worden op de verschillende elementen in de LOM metadata (sectie 3.6.2), en
- een paneel waar tekstuele informatie over de **resultaten** getoond wordt zoals de titel, de locatie van het object, de auteur, enz. (sectie 3.6.3.)

Al deze panelen werken op een gecoördineerde manier wat wil zeggen dat als 1 paneel aangepast wordt, de andere panelen eveneens aangepast worden. Deze basis gebruikersinterface voldeed aan onze behoeften voor het valideren van het visualisatie ontwerp. Figuur 3.9 toont deze interface.

### 8.2.3 Besluit

Twee evaluaties werden uitgevoerd aan de hand van bovenstaand prototype. De eerste evaluatie was een expert evaluatie voor te valideren of 7 typische gebruikerstaken ondersteund zijn in het prototype. Deze werden gedefinieerd bij de vereisten in sectie 3.3. De tweede evaluatie was een gebruikersevaluatie waarbij een traditioneel zoeksysteem in ARIADNE vergeleken werd met het prototype. Tijdens deze evaluatie werd gebruik gemaakt van de "think aloud"-methode voor het capteren van de gedachten van de gebruikers [Nielsen, 1992a]. 3 karakteristieken werden gemeten tijdens het experiment: (i) tijd om een taak uit te voeren, (ii) juistheid van een taak en (iii) gebruikersvoldoening. Dit laatste werd gemeten door gebruik te maken van de "Likert Scale" methode [Likert, 1932]. Voor een compleet overzicht over de resultaten, verwijzen we naar sectie 3.7.3. Belangrijk is dat deze studie duidelijk maakte dat ons ontwerp, dat gebruik maakt van informatie visualisatie technieken, effectief en efficient kan werken voor het ontsluiten van een repository met leermateriaal. Voor de implementatie van het prototype werden er een aantal bibliotheken toegevoegd aan het uitbreidbaar raamwerk. Een opsomming van deze bibliotheken kan gevonden worden in sectie 3.8.

In dit geval probeerden we een repository van leermateriaal te ontsluiten op basis van het onderwerp van het leermateriaal. In de volgende sectie bespreken we hoe we een repository kunnen visualiseren op basis van de "hasPart / isPartOf" relaties die kunnen bestaan tussen verschillende leerobjecten.

## 8.3 Visualiseren van Hergebruik van Materialen

ALOCOM [Verbert et al., 2005] en [Verbert et al., 2004] is een raamwerk voor (i) het automatisch disaggregeren van leermateriaal in kleine componenten enerzijds en (ii) het aggrereren van kleine componenten in grotere leerobjecten anderzijds. De "has part/is part of" relaties die bestaan tussen deze componenten zijn een interessant startpunt voor het visueel ontsluiten van een repository, die al deze componenten bevat.

De ALOCOM repository bevatte, op het moment van deze studie, 48286 componenten die geëxtraheerd werden uit 653 presentations[1]. Deze componenten bestonden uit 14113 slides, 5768 afbeeldingen, 198 tabellen, 26 diagrammen en 27543 tekst fragmenten. Een belangrijk aspect van ALOCOM is hergebruik van leermateriaal. Hergebruik in een repository van verschillende presentaties wordt gemeten als volgt: het is gelijk aan het aantal componenten die hergebruikt werden in een presentatie, gedeeld door het totaal aan componenten in die presentatie. Als een presentatie bijvoorbeeld 10 componenten bevat, waarvan 2 afbeeldingen herbruikt worden in andere presentaties, is de hergebruik waarde van die presentatie gelijk aan 0.2. Het gemiddelde hergebruik in de volledige ALOCOM repository is 0.22. Dit cijfer zegt echter niet veel over wat echt gaande is in de repository. Het geeft geen antwoorden op een aantal vragen zoals:

- Wat is de distributie van het hergebruik-patroon in de repository? Zijn er componenten die veel herbruikt worden zodat dit tot een gemiddeld herge-

[1] Laatste update: September 2007

bruikt leidt van 0.22, of worden alle componenten bijvoorbeeld 1 keer hergebruikt?

- Welke componenten worden dikwijls hergebruikt? Waarom zijn sommige componenten populairder dan andere?
- Hoe worden componten samengeplaatst in geaggregeerde content?
- Welke karakteristieken dragen bij om componenten herbruikbaar te maken?
- enz.

Onderzoekers zoeken vooral antwoorden op dit soort vragen [Collis, 1995], [Strijker, 2004] en [Verbert et al., 2008]. Het doel van deze studie is niet zozeer een antwoord geven op bovenstaande vragen, maar eerder om een middel aan te bieden aan de onderzoekers waarmee zij het antwoord kunnen vinden.

### 8.3.1 Ontwerp

De "hasPart/isPartOf" relaties tussen de componenten in de repository zijn van het netwerk type (sectie 3.4.1). Het visualiseren van dit type is een oude kunst maar is opnieuw een belangrijk onderzoeksonderwerp door de complexiteit van de relaties en de gebruikerstaken [Shneiderman, 1996]. Na een vergelijkend onderzoek tussen een aantal mogelijke visualisatie manieren, hebben we gekozen voor een node-link diagram voor het presenteren van een visueel overzicht van de ALOCOM repository. Dit kan men zien in Figuur 4.2 waar een "twopi" layout [Wills, 1999] gebruikt wordt. Alle componenten worden hier geplaatst in een cirkel. Deze layout laat ons toe om in één oogopslag te zien of er veel of weinig hergebruik is in de repository. Veel overlappingen wilt immers zeggen, veel hergebruik. Als er heel veel overlappingen zijn, wordt deze overzichtsfiguur echter onoverzichtelijk. Daarom hebben we een aantal interactie-technieken toegevoegd aan deze applicatie zoals "fisheye" lenzen [Furnas, 1999], "excentric labelling" [Fekete and Plaisant, 1999], en het toelaten om enkel die componenten te tonen die de gebruiker interesseren.

Voor het valideren van al deze technieken werd opnieuw een prototype gebouwd met dezelfde eenvoudige gebruikersinterface als het prototype voor het ontsluiten van een LOR. Het dynamisch query controle paneel (sectie 8.2) werd dus opnieuw toegevoegd waarmee gebruikers filters kunnen instellen over de verschillende LOM metadata velden. Bovendien werd er een "Preview"-paneel toegevoegd waarmee de gebruikers gemakkelijk en snel componenten kunnen vergelijken. Dit paneel toont volledige slides, afbeeldingen en individuele tekstfragmenten. Een afbeelding van het volledig prototype kan teruggevonden worden in Figuur 4.6.

### 8.3.2 Besluit

Het prototype werd gebruikt voor het evalueren en valideren van de gebruikte technieken. Vanwege de omvang van de doelgroep is het organiseren van evaluaties op grote schaal moeilijk. Daarom hebben we de aanbevelingen gevolgd in [Andrews, 2006] en zijn we vertrokken van een beperkt aantal zeer ervaren gebruikers in een vroege fase van de ontwikkeling. Hierdoor krijg je relevante feedback over de gebruikte exploratie- en analysemethoden die beschikbaar zijn in het prototype. Een klein aantal gebruikers in de test ($4 \pm 1$) is voldoende om de meeste bruikbaarheidsproblemen eruit te halen [Nielsen, 1992c]. Om deze reden hebben wij ervoor gekozen om 4 deskundige gebruikers in de TEL gemeenschap te betrekken in onze evaluatie. Een uitgebreid overzicht over de resultaten van deze evaluatie hebben we opgenomen in sectie 4.6.2. Het belangrijkste resultaat was dat de experten vonden dat het gebruik van de visualisatie technieken effectief werkte. Uiteraard waren er een aantal aanbevelingen voor het verbeteren van het prototype. Het software raamwerk werd voor dit prototype uitgebreid met een aantal bibliotheken voor het connecteren van het raamwerk met ALOCOM, het creëren van een twopi grafe layout, enz. Een volledig overzicht van de implicaties op het raamwerk kan teruggevonden worden in sectie 4.7.

Uit de evaluatie bleek onder andere dat de gebruikers erg geïnteresseerd waren in een visualisatie van het sociale aspect tussen auteurs in ALOCOM. In de volgende sectie bespreken we daarom de studie die we verricht hebben voor het visualiseren van sociale bookmarks of favorieten.

## 8.4 Visualiseren van Sociale Bookmarks

De mogelijkheid om "bookmarks" of favoriete webpagina's op te slaan, samen een aantal "tags", is één van de belangrijkste kenmerken van browsers sinds het begin van het Web [Hammond et al., 2005]. Tagging is het proces van het toekennen van betekenis aan items die online opgeslaan zijn, zoals webpagina's, afbeeldingen, video's, enz., door middel van etikettering met gepersonaliseerde zoekwoorden die worden verdeeld onder de gebruikers. Het doel van web-gebaseerde sociale bookmarking tools is het taggen van content van andere gebruikers, voornamelijk ten behoeve van de tagger, hoewel de bookmarks en tags publiek zijn waardoor gebruikers de mogelijkheid krijgen om te netwerken. Er bestaan dus impliciete relaties tussen gebruikers, tags en content items. Deze relaties zijn echter niet altijd duidelijk in de traditionele ontsluiting van sociale bookmarks. Ons onderzoek focust daarom op het gebruik van informatie visualisatie technieken:

- waarmee analysten inzicht kunnen krijgen in zulke impliciete relaties, en
- die eindgebruikers nieuwe manieren kunnen aanbieden voor het vinden van content dat hen interesseert maar dat ze niet gevonden zouden hebben door

gebruik te maken van expliciete zoekacties.

Van een heel aantal bestaande sociale bookmarking tools is del.icio.us wellicht één van de meest bekende [delicious, 2008]. We hebben del.icio.us gekozen als bron voor onze data omdat (i) het zeer populair is bij vele gebruikers, (ii) het zeer veel data bevat en (iii) dat het een erg gemakkelijke API heeft om zijn data te ontsluiten. Als we kunnen aantonen dat de visualisatie technieken effectief en efficiënt kunnen werken voor het exploreren van del.icio.us, hebben we daardoor reden te geloven dat ze ook waardevol kunnen zijn voor andere sociale bookmarking tools.

### 8.4.1 Ontwerp

Qua ontwerp hebben we in deze visualisatie gekozen voor een "cluster map" visualisatie [Fluit et al., 2005]; dit is een zeer expressieve techniek voor het visualiseren van klassen, zoals tags, en objecten, zoals bookmarks, die tot deze klassen behoren. In dit geval hebben we eveneens een werkend prototype gecreëerd van ons ontwerp voor het valideren van de gebruikte technieken. Het ontwerp wijkt in dit geval af van de vorige ontwerpen. Waar we in de vorige studies vertrokken van een volledig visueel overzicht van de repository, en erna inzoomden op details ervan, volgen we deze keer een andere filosofie: "start with what you know, then grow" [Heer et al., 2005]. Dit betekent dat er bij de start niets gevisualiseerd wordt buiten de bookmarks en de tags van de gebruiker zelf. De gebruiker heeft volledige controle over welke tags, bookmarks en andere gebruikers hij of zij toevoegt aan de visualisatie.

Het prototype, getoond in figuur 5.1 bestaat opnieuw uit een eenvoudige gebruikersinterface met 3 onderdelen:

1. een selectie widget dat een lijst voorstelt met gebruikers en tags die beschikbaar zijn in de data (sectie 5.5.1),
2. een cluster map visualisatie paneel (sectie 5.5.2), en
3. een filter paneel met een hierin geïntegreerde lijst van resultaten (sectie 5.5.3).

Al deze onderdelen werken opnieuw op een gesynchroniseerde manier zodat als de gebruiker iets verandert op eender welk paneel, de gegevens op alle andere panelen eveneens aangepast worden.

### 8.4.2 Besluit

Wij geloven sterk in software ontwikkeling aan de hand van "rapid prototyping", waar software systemen stapsgewijs worden opgeleverd en de vereisten analyse gedurende het gehele proces wordt bijgeschaafd [Luqi, 1989]. Het stelt ons in staat problemen zoals bruikbaarheid te ontmaskeren in een vroege fase van ontwikkeling. We hebben dan ook in een vroege fase een aantal interviews gedaan met eindgebruikers voor het verkrijgen van feedback (section 5.8.1). Achteraf hebben we een gebruikersstudie uitgevoerd met het ontwikkelde prototype voor het valideren van de gebruikte technieken. De volledige resultaten van deze studie zijn opgenomen in sectie 5.8.2. Het belangrijkste besluit dat we uit deze studie kunnen trekken is dat gebruikers het gebruik van informatie visualisatie technieken een effectief en efficiënt middel vonden voor het exploreren van social bookmarks. Uiteraard waren er ook een heel aantal aanbevelingen voor het verbeteren van het ontwerp. Eén hiervan was het visualiseren van tags doorheen de tijd. Dit zou de interesse en focus reflecteren van de gebruikers doorheen de tijd. Dit idee hebben we daarom uitgewerkt in de case studie die we bespreken in de volgende sectie.

## 8.5 Visualiseren van een Netwerk van LORs

Deze case studie biedt een visuele toegang tot een netwerk van LORs in plaats van toegang tot slechts een enkele LOR 8.2. Het GLOBE ("Global Learning Objects Brokered Exchange") consortium [Globe, 2008] is gericht op de ontgrendeling van het "deep web" van de netwerken van LORs die haar leden onderhouden. De doelstellingen van dit hoofdstuk wijken echter af van die van de andere hoofdstukken. Daar gingen we telkens uit van een verzameling herbruikbare componenten die we wilden ontsluiten met het gebruik van informatie visualisatie technieken als alternatief voor de meer traditionele ontsluiting van deze verzamelingen. Prototypes werden telkens ontwikkeld met het uitbreidbaar software raamwerk (hoofdstuk 2), maar het raamwerk zelf zien we slechts als een hulpmiddel om het gebruik van deze technieken te kunnen valideren. De gemaakte applicaties zijn altijd online beschikbaar gesteld als Java-toepassingen, hetzij door gebruik te maken van Javawebstart of Java-applets.

In dit geval exploreren we slechts kort de mogelijkheden voor het visualiseren van netwerken van LORs, omdat de primaire doelstelling van deze case studie de praktische opzet en integratie was van een mashup applicatie met ons uitbreidbaar software raamwerk. We wilden weten wat de praktische gevolgen zijn voor de architectuur van het raamwerk. Een mashup is een webapplicatie die gebruik maakt van de services van meerdere bronnen voor het maken van een nieuwe service. Services die gebruikt worden in mash-ups zijn doorgaans afkomstig van een derde partij via een publieke interface of API. Als gevolg daarvan bieden we een oplossing aan voor het ontsluiten van een netwerk van LORs, maar deze oplossing werd niet geëvalueerd door eindgebruikers. Deze oplossing kan echter dienen als een interessant vertrekpunt voor toekomstig werk.

### 8.5.1 Ontwerp

UIt de vereisten analyse kwamen 2 topics naar voren (sectie 6.2). Gebruikers moeten in staat zijn om

1. leermateriaal te vinden in GLOBE.
2. een inzicht te krijgen in de zoekgeschiedenis van gebruikers. Dit laat de gebruikers toe om vragen te beantwoorden over GLOBE zoals:
    - Welke sleutelwoorden worden hoeveel keer gebruikt in welke periode en kunnen daarom geïnterpreteerd worden als representatief voor een bepaalde periode?
    - Welke sleutelwoorden hebben veel of weinig resultaten?
    - Welke sleutelwoorden hebben resultaten in meerdere repositories?
    - enz.

Voor de eerste vereiste hebben we de resultaten van een zoekactie gevisualiseerd op een geografische kaart die gesynchroniseerd werd met een lijst presentatie. Dit kan men zien in figuur 6.1. Voor de tweede vereiste hebben we een uitgebreide tag cloud ontworpen die gesynchroniseerd werkt met een tijdlijn. Deze tag cloud hebben we opgenomen in figuur 6.3 waar men kan zien dat de zoektermen afgebeeld worden in een assenstelsel waarbij de X-as het aantal resultaten voorstelt en de Y-as het aantal repositories waarin resultaten gevonden worden.

### 8.5.2 Besluit

Ons doel, waarin we geslaagd zijn, was het ontwerpen en creëren van een mashup web applicatie die

- geïntegreerd is met ons uitbreidbaar software raamwerk voor het visualiseren van verzamelingen herbruikbare componenten,
- gestandardiseerde data formaten gebruikt (LTSC IEEE LOM [IEEE, 2002], XML [XML, 2006], JSON [JSON, 2006], SVG [SVG, 2003] en CAM [Wolpers et al., 2006]), en
- gestandardiseerde protocollen gebruikt (HTTP, SOAP).

De implicaties voor het creëren van een software raamwerk zijn vooral een aantal toegevoegde web services. Een volledig overzicht van deze webservices werden beschreven in sectie 6.4. Ons ontwerp voor het ontsluiten van een netwerk van LORS, werd niet geëvalueerd maar is een interessant startpunt voor toekomstig werk.

## 8.6 Toekomstig Werk

Al de uitgewerkte case studies tonen aan dat het gebruik van informatie visualisatie technieken voor het ontsluiten van verzamelingen herbruikbare componenten effectief en efficiënt kan werken en dat het inzicht kan bieden aan gebruikers in de inhoud van zulke repositories. Uiteraard gaat het hier om een ingenieursthesis waardoor de nadruk in dit werk meer lag op het ontwerp van een oplossing en de praktische implementatie ervan in het raamwerk. Allerlei integraties van verschillende bibliotheken, protocollen en standaarden kwamen daarbij aan bod. De nadruk lag daardoor minder op de meer humane, psychologische aspecten waarom het gebruik van informatie visualisatie technieken meer of sneller inzicht kan bieden dan alternatieve technieken.

We beëindigen dit hoofdstuk met een aantal interessante topics voor toekomstig werk:

- Aanbevelingen uit evaluaties; Uit alle evaluaties die uitgevoerd werden, zijn er telkens een aantal aanbevelingen geformuleerd die de voorgestelde oplossingen kunnen verbeteren. Niet al deze aanbevelingen zijn reeds uitgewerkt. Dit is echter interessant toekomstig werk.
- Raamwerk validatie; Ons software raamwerk is gebruikt voor het uitwerken van verschillende prototypes en heeft verschillende geïntegreerde bibliotheken die door derden ontwikkeld zijn. Dit toont reeds de uitbreidbaarheid aan. Het integreren van nog meer bibliotheken, visualisatie technieken en de ontwikkeling van andere case studies kan echter de geschiktheid van het raamwerk voor zijn doel verder aantonen.
- Raamwerk vertaling; Op dit moment, is het raamwerk volledig geschreven in Java™. Het zou interessant zijn om een herimplementatie te maken in andere technologieën zoals .NET of Adobe FLEX/AIR/FLASH. Deze implementaties zouden het mogelijk maken om een vergelijkend onderzoek te starten tussen deze verschillende technologieën.
- "Visualisation for the masses"; Dit is een erg populair onderwerp op het moment van schrijven. Gewone eindgebruikers zouden gemakkelijk in staat moeten zijn om hun data te visualiseren. Met ons raamwerk is dit niet mogelijk aangezien een stevige achtergrond vereist is over het programmeren in Java™. Het creëren van een hoog-niveau gebruikersinterface is daardoor een interessante uitdaging.
- Kwaliteitsmetrieken; Er zijn een aantal pogingen gedaan naar het vinden van kwantificeerbare kwaliteitsmetrieken voor visualisatie en interactie algoritmen. Het vinden van zulke metrieken zou gebruikersstudies gemakkelijker mogelijk maken voor het evalueren van de geschiktheid van een visualisatie techniek voor een bepaald doel.

# Bibliography

[Aduna, 2008] Aduna (2008). The aduna clustermap library. http://www.aduna-software.org/projects/display/CLUSTERMAP/. Last retrieved: April 2008.

[Ahlberg and Shneiderman, 1994] Ahlberg, C. and Shneiderman, B. (1994). Visual information seeking: tight coupling of dynamic query filters with starfield displays. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 313–317, New York, NY, USA. ACM.

[Andrews, 2006] Andrews, K. (2006). Evaluating information visualisations. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pages 1 – 5.

[AristoMusic, 2008] AristoMusic (2008). Aristomusic. http://www.aristomusic.com/. last retrieved: April 2008.

[Barlex and Carrre, 1985] Barlex, D. and Carrre, C. (1985). *Visual Communication in Science*. Number 978-0521278607. Cambridge University Press.

[Bass et al., 2006] Bass, L., Clements, P., and Kazman, R. (2006). *Software Architecture in Practice*. Number ISBN 0321154959 in SEI Series in software engineering. Addison-Wesley Professional, second edition edition.

[Becker et al., 1995] Becker, R. A., Eick, S. G., and Wilks, A. R. (1995). Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics*, 1(1):16–28.

[Bederson et al., 2004] Bederson, B. B., Grosjean, J., and Meyer, J. (2004). Toolkit design for interactive structured graphics. *IEEE Trans. Softw. Eng.*, 30(8):535–546.

[Bederson et al., 1996] Bederson, B. B., Hollan, J. D., Perlin, K., Meyer, J., Bacon, D., and Furnas, G. W. (1996). Pad++: A zoomable graphical sketchpad for exploring alternate interface physics. *Journal of Visual Languages and Computing*, 7(1):3–32.

[Bederson et al., 2002] Bederson, B. B., Shneiderman, B., and Wattenberg, M. (2002). Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *ACM Trans. Graph.*, 21(4):833–854.

[Bergman and Johnson, 1995] Bergman, E. and Johnson, E. (1995). Toward accessible human-computer interaction. *Advances in human-computer interaction (vol. 5)*, 5:pp. 87–113.

[Bielenberg and Zacher, 2006] Bielenberg, K. and Zacher, M. (2006). Groups in social software: Utilizing tagging to integrate individal contexts for social navigation. Master's thesis, Universitat Bremen.

[BlogMarks, 2008] BlogMarks (2008). Blogmarks. http://www.blogmarks.net. last retrieved: April 2008.

[Blustein et al., 2005] Blustein, J., Ahmed, I., and Instone, K. (2005). An evaluation of look-ahead breadcrumbs for the www. In *HYPERTEXT '05: Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*, pages 202–204, New York, NY, USA. ACM.

[Bruls et al., 2000] Bruls, D. M., Huizing, C., and van Wijk, J. (2000). Squarified treemaps. In *Proceedings of the joint Eurographics and IEEE TVCG Symposium on Visualization*, pages 33–42. Springer.

[Brusilovsky, 1996] Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Model. User-Adapt. Interact.*, 6(2-3):87–129.

[Brusilovsky and Su, 2002] Brusilovsky, P. and Su, H.-D. (2002). Adaptive visualization component of a distributed web-based adaptive educational system. In *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems (ITS'2002), Biarritz, France.* Lecture Notes in Computer Science, Vol. 2363, pp. 229-238, Springer-Verlag, Berlin.

[Buneman and Levy, 1980] Buneman, P. and Levy, L. (1980). The Towers of Hanoi Problem. *Information Processing Letters*, 10(4/5):243–244.

[Burkhard and Meier, 2005] Burkhard, R. A. and Meier, M. (2005). Tube map visualization: Evaluation of a novel knowledge visualization application for the transfer of knowledge in long-term projects. *Journal of Universal Computer Science*, 11(4):473–494.

[CALIBRATE, 2008] CALIBRATE (2008). The calibrate project. http://calibrate.eun.org. last retrieved: april 2008.

[Card et al., 1999] Card, S. K., Mackinlay, J. D., and Shneiderman, B. (1999). *Using vision to think.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[Cardinaels, 2007] Cardinaels, K. (2007). *A Dynamic Learning Object Life Cycle and its Implications for Automatic Metadata Generation.* PhD thesis, Katholieke Universiteit Leuven - Faculteit Ingenieurswetenschappen, Belgium.

[Cardinaels et al., 2005] Cardinaels, K., Meire, M., and Duval, E. (2005). Automating metadata generation: the simple indexing interface. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 548–556, New York, NY, USA. ACM.

[Chaudhuri and Dayal, 1997] Chaudhuri, S. and Dayal, U. (1997). An overview of data warehousing and olap technology. *SIGMOD Rec.*, 26(1):65–74.

[Chi and Mytkowicz, 2007] Chi, E. and Mytkowicz, T. (2007). Understanding navigability of social tagging systems. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI'07)*, pages (San Jose, USA, April, 2007), New York, NY. ACM Press.

[CiteULike, 2008] CiteULike (2008). Citeulike. http://www.citeulike.org. last retrieved: April 2008.

[Collis, 1995] Collis, B. (1995). The evolution of educational software productivity. In *Educational media and technology yearbook*, volume 21, pages 76–97. CO: Libararies Unlimited, Englewood.

[Connotea, 2008] Connotea (2008). Connotea. http://www.connotea.org/. last retrieved: April 2008.

[Cooper, 2000] Cooper, J. W. (2000). *Java design patterns: a tutorial.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[del.icio.us, 2008] del.icio.us (2008). The del.icio.us abuse monitor. http://del.icio.us/help/json/url. last retrieved: April 2008.

[delicious, 2008] delicious (2008). The del.icio.us social bookmarking tool. http://del.icio.us. last retrieved: April 2008.

[Dennis et al., 2002] Dennis, S., Bruza, P., and McArthur, R. (2002). Web searching: A process-oriented experimental study of three interactive search paradigms. *Journal of the American Society for Information Science and Technology*, 53(2):120–133.

[Dodge and Kitchin, 2001] Dodge, M. and Kitchin, R. (2001). *Atlas of Cyberspace.* Number ISBN 978-0201745757. Addison-Wesley Professional.

[Dojo, 2006] Dojo (2006). The dojo toolkit. http://dojotoolkit.org. last retrieved: April 2008.

[Dorner and Curtis, 2003] Dorner, D. and Curtis, A. (2003). A comparative review of common user interface software products for libraries. *Victoria University of Wellington, Wellington, available at: www. natlib. govt. nz/-files/CCI_Report_Final. pdf.*

[Dubinko et al., 2007] Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., and Tomkins, A. (2007). Visualizing tags over time. *ACM Trans. Web*, 1(2):7.

[Duval, 2006] Duval, E. (2006). *Handbook on Quality and Standardisation in E-Learning*, chapter LearnRank: Towards a real quality measure for learning, pages 457–463. Springer Berlin Heidelberg. ISBN: 978-3-540-32787-5.

[Duval et al., 2001] Duval, E., Forte, E., Cardinaels, K., Verhoeven, B., Durm, R. V., Hendrikx, K., Forte, M. W., Ebel, N., Macowicz, M., Warkentyne, K., and Haenni, F. (2001). The ariadne knowledge pool system. *Commun. ACM*, 44(5):72–78.

[Duval and Hodgins, 2003] Duval, E. and Hodgins, W. (2003). A lom research agenda. In Hencsey, G., White, B., Chen, Y., Kovacs, L., and Lawrence, S., editors, *Proceedings of the 12th International World Wide Web Conference*, pages 659–667, New York. ACM Press.

[Duval et al., 2006] Duval, E., Smith, N., and Coillie, M. V. (2006). Application profiles for learning. In *ICALT '06: Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies*, pages 242–246, Washington, DC, USA. IEEE Computer Society.

[Eick, 1994] Eick, S. G. (1994). Data visualization sliders. In *UIST '94: Proceedings of the 7th annual ACM symposium on User interface software and technology*, pages 119–120, New York, NY, USA. ACM.

[Ellson et al., 2003] Ellson, J., Gansner, E., Koutsofios, E., North, S., and Woodhull, G. (2003). *Graph Drawing Software*, chapter Graphviz and dynagraph static and dynamic graph drawing tools, pages 127–148. Springer-Verlag.

[explorer, 2008] explorer, N. (2008). The del.icio.us network explorer. http://www.twoantennas.com/projects/delicious-network-explorer/. last retrieved: April 2008.

[Extisp.icio.us, 2008] Extisp.icio.us (2008). Extisp.icio.us text. http://kevan.org/extispicious. last retrieved: April 2008.

[Fekete, 2004] Fekete, J.-D. (2004). The infovis toolkit. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*, pages 167–174, Washington, DC, USA. IEEE Computer Society.

[Fekete and Plaisant, 1999] Fekete, J.-D. and Plaisant, C. (1999). Excentric labeling: dynamic neighborhood labeling for data visualization. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 512–519, New York, NY, USA. ACM.

[Few, 2004] Few, S. (2004). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Number ISBN 0970601999. Analytics Press.

[Fielding, 2000] Fielding, R. T. (2000). *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, University of California, Irvine.

[Fishkin and Stone, 1995] Fishkin, K. and Stone, M. C. (1995). Enhanced dynamic queries via movable filters. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–420, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

[Flickr, 2008] Flickr (2008). Flickr. http://flickr.com/. last retrieved: April 2008.

[Fluit et al., 2005] Fluit, C., Sabou, M., and van Harmelen, F. (2005). Ontology-based information visualisation: Towards semantic web applications. In Geroimenko, V., editor, *Visualising the Semantic Web (2nd edition)*, pages 36–48. Springer Verlag.

[France et al., 1999] France, R. K., Nowell, L. T., Fox, E. A., Saad, R. A., and Zhao, J. (1999). Use and usability in a digital library search system.

[Furl, 2008] Furl (2008). Furl. http://www.furl.net. last retrieved: April 2008.

[Furnas, 1999] Furnas, G. W. (1999). The fisheye view: a new look at structured files. *Readings in information visualization: using vision to think*, pages pp. 312–330.

[Gamma et al., 1995] Gamma, E., Helm, R., Johnson, R., and Vlissides, J. (1995). *Design patterns: elements of reusable object-oriented software*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

[GeoIP, 2007] GeoIP (2007). Geoip technology. http://www.maxmind.com/app/ip-locate. last retrieved: April 2008.

[Globe, 2008] Globe (2008). The globe consortium. http://globe-info.org/globe/go. last retrieved: April 2008.

[Goldberg et al., 1992] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Commun. ACM*, 35(12):61–70.

[GoogleDirectory, 2008] GoogleDirectory (2008). The google directory. http://www.google.com/dirhp?hl=en. last retrieved: April 2008.

[GoogleMaps, 2007] GoogleMaps (2007). Google maps api. http://www.google.com/apis/maps/. last retrieved: April 2008.

[GoogleTrends, 2006] GoogleTrends (2006). Google trends. http://www.google.com/trends. last retrieved: April 2008.

[Granovetter, 1983] Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233.

[Groxis, 2008] Groxis (2008). Grokker. http://www.grokker.com. last retrieved: April 2008.

[Guy and Tonkin, 2006] Guy, M. and Tonkin, E. (2006). Folksonomies: Tydying up tags? *D-Lib Magazine*, 12(1).

[Gynn and Acker, 2003] Gynn, C. M. and Acker, S. R., editors (2003). *Course Management Systems: Trapped Content Silos or Sharing Platforms?* The Ohio State University.

[Halvey and Keane, 2007] Halvey, M. J. and Keane, M. T. (2007). An assessment of tag presentation techniques. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 1313–1314, New York, NY, USA. ACM. ISBN: 978-1-59593-654-7.

[Hammond et al., 2005] Hammond, T., Hannay, T., Lund, B., and Scott, J. (2005). Social bookmarking tools (i). *D-Lib Magazine*, 11(4). ISSN 1082-9873.

[Hassan-Montero and Herrero-Solana, 2006] Hassan-Montero, Y. and Herrero-Solana, V. (2006). Improving tag-clouds as visual information retrieval interfaces. In *Proceedings of International Conference on Multidisciplinary Information Sciences and Technologies (InScit2006)*. ISBN: 84-611-3104-5.

[HCIL, 2008] HCIL (2008). Treemap 4.1.1 toolkit. http://www.cs.umd.edu/hcil/treemap/. last retrieved: April 2008.

[Healey, 1996a] Healey, C. G. (1996a). Choosing effective colours for data visualization. In *VIS '96: Proceedings of the 7th conference on Visualization '96*, pages 263–270, Los Alamitos, CA, USA. IEEE Computer Society Press.

[Healey, 1996b] Healey, C. G. (1996b). *Effective visualization of large multidimensional datasets*. PhD thesis, The University of British Columbia (Canada). Adviser-Kellog S. Booth and Adviser-James T. Enns.

[Heer and Boyd, 2005] Heer, J. and Boyd, D. (2005). Vizster: Visualizing online social networks. In *INFOVIS '05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, page 5, Washington, DC, USA. IEEE Computer Society. ISBN ISSN:1522-404x , 0-7803-9464-x.

[Heer et al., 2005] Heer, J., Card, S. K., and Landay, J. A. (2005). prefuse: a toolkit for interactive information visualization. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430, New York, NY, USA. ACM.

[Hoffman, 2006] Hoffman, K. (2006). In search of ... the perfect tag cloud. http://dotnetaddict.dotnetdevelopersjournal.com/tw.htm. white paper: http://dotnetaddict.dotnetdevelopersjournal.com/tw.htm.

[hostIP, 2007] hostIP (2007). hostip geolocation database. http://www.hostip.info. last retrieved: April 2008.

[HubLog, 2008] HubLog (2008). Hublog. http://hublog.hubmed.org/archives/001049.html. last retrieved: April 2008.

[Hypertree, 2003] Hypertree (2003). A hypertree java library. http://hypertree.sourceforge.net. last retrieved: April 2008.

[IEEE, 2002] IEEE (2002). *IEEE Standard for Learning Object Metadata*. The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, USA.

[Inc., 2008] Inc., A. (2008). The 1998 acm computing classification system. http://www.acm.org/class/1998/. last retrieved: April 2008.

[Inselberg, 1985] Inselberg, A. (1985). The plane with parallel coordinates. *The Visual Computer*, 1(4):69–91.

[Ivanova, 2004] Ivanova, N. (2004). EdNA Online Shared Information Services: Case studies. In *Proceedings of the 21st ASCILITE Conference*, pages 455–457, Perth, Western Australia.

[JSON, 2006] JSON (2006). Rfc4627: Javascript object notation (json). last retrieved: April 2008.

[Jünger and Mutzel, 2003] Jünger, M. and Mutzel, P. (2003). *Graph Drawing Software*. Number ISBN 3-540-00881-0. Springer-Verlag.

[Keim and Oelke, 2007] Keim, D. A. and Oelke, D. (2007). Literature fingerprinting: A new method for visual literary analysis. In Ribarsky, W. and Dill, J., editors, *Proceedings of IEEE symposium on Visual Analytics Science and Technology (VAST)*, pages 115–122.

[Klerkx, 2008a] Klerkx, J. (2008a). Delivis. http://www.cs.kuleuven.ac.be/ hmdb/infovis/delicious/del.icou.ustion.html. last retrieved: April 2008.

[Klerkx, 2008b] Klerkx, J. (2008b). Visual access to the ariadne kps. http://www.cs.kuleuven.ac.be/ hmdb/infovis/. last retrieved: April 2008.

[Klerkx and Duval, 2007a] Klerkx, J. and Duval, E. (2007a). Globemash: A mashup for accessing globe. In Tochtermann, K. and Maurer, H., editors, *Proceedings of 7th International Conference on Knowledge Management*, pages 392–399, Graz, Austria.

[Klerkx and Duval, 2007b] Klerkx, J. and Duval, E. (2007b). A synchronized tag cloud and timeline visualization. In *Poster Compendium of IEEE Information Visualization Conference*, pages 86–88, Sacramento, CA, USA.

[Klerkx and Duval, 2007c] Klerkx, J. and Duval, E. (2007c). Visualizing social bookmarks. In *Proceedings of SIRTEL07 workshop, EC-TEL07*, Krete.

[Klerkx et al., 2004] Klerkx, J., Duval, E., and Meire, M. (2004). Using information visualization for accessing learning object repositories,. In Banissi, E., Borner, K., Chen, C., Dastbaz, M., Clapworthy, G., Faiola, A., Izquierdo, E., Maple, C., Roberts, J., Moore, C., Ursyn, A., and Zhang, J., editors, *Proceedings of Information Visualization (IV04)*, pages 465–470, London, United Kingdom.

[Klerkx et al., 2005] Klerkx, J., Meire, M., S.Ternier, Verbert, K., and Duval, E. (2005). Information visualisation: Towards an extensible framework for accessing learning object repositories. In Kommers, P. and Richards, G., editors, *Proceedings of World Conference on Educational Multimedia, Hypermedia & Telecommunications*, pages 4281–4287, Montreal, Canada.

[Klerkx et al., 2006] Klerkx, J., Verbert, K., and Duval, E. (2006). Visualizing reuse: more than meets the eye. In Tochtermann, K. and Maurer, H., editors, *Proceedings of 7th International Conference on Knowledge Management*, pages 389–396, Graz, Austria.

[Kobsa, 2004] Kobsa, A. (2004). User experiments with tree visualization systems. *IEEE Symposium on Information Visualization, 2004.*, pages 9–16.

[Kraaij, 2004] Kraaij, W. (2004). *Variations on Language Modeling for Information Retrieval*. PhD thesis, University of Twente.

[Kules and Shneiderman, 2003] Kules, B. and Shneiderman, B. (2003). Designing a metadata-driven visual information browser for federal statistics. In *dg.o '03: Proceedings of the 2003 annual national conference on Digital government research*, pages 1–6. Digital Government Research Center.

[Lamping and Rao, 1996] Lamping, J. and Rao, R. (1996). The hyperbolic browser: A focus+context technique for visualizing large hierarchies. *Journal of Visual Languages & Computing*, 7(1):33–55.

[Leung and Apperley, 1994] Leung, Y. K. and Apperley, M. D. (1994). A review and taxonomy of distortion-oriented presentation techniques. *ACM Trans. Comput.-Hum. Interact.*, 1(2):126–160.

[Likert, 1932] Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140):1–55.

[Lucene, 2008] Lucene (2008). Lucene homepage. http://lucene.apache.org/. last retrieved: April 2008.

[Luqi, 1989] Luqi (1989). Software evolution through rapid prototyping. *Computer*, 22(5):13–25. ISSN 0018-9162.

[Manning et al., 2008] Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

[Marchionini, 2006a] Marchionini, G. (2006a). Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41–46.

[Marchionini, 2006b] Marchionini, G. (2006b). Leveraging Context for Adaptive Multimedia Retrieval: A Matter of Control. *Adaptive Multimedia Retrieval: User, Context, and Feedback: Third International Workshop, AMR 2005, Glasgow, UK, July 28-29, 2005: Revised Selected Papers*.

[Marchionini, 2006c] Marchionini, G. (2006c). Toward human-computer information retrieval. http://www.asis.org/Bulletin/Jun-06/marchionini.html.

[McMartin, 2004] McMartin, F. (2004). Case study: MERLOT: A model of user involvement in digital library design and implementation. *Journal of Digital Information*, 5(3).

[Michard, 1982] Michard, A. (1982). A new database query language for non-professional users: Design principles and ergonomic evaluation. *Behavioral and Information Technology*, (1):279–288.

[Middleton and Baeza-Yates, 2008] Middleton, C. and Baeza-Yates, R. (2008). Technical report: A comparison of open source search engines. Technical report, Universitat Pompeu Fabra (Barcelona, Spain).

[Millen et al., 2007] Millen, D., Yang, M., Whittaker, S., and Feinberg, J. (2007). Social bookmarking and exploratory search. In Bannon, L., Wagner, I., Gutwin, C., Harper, R., and Schmidt, K., editors, *Proceedings of the Tenth European Conference on Computer Supported Cooperative Work*, ECSCW, pages 21–40, Limerick, Ireland. Springer.

[Millen and Feinberg, 2006] Millen, D. R. and Feinberg, J. (2006). Using social tagging to improve social navigation. In *Workshop on the Social Navigation and Community based Adaptation Technologies*.

[Millen et al., 2006] Millen, D. R., Feinberg, J., and Kerr, B. (2006). Dogear: Social bookmarking in the enterprise. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 111–120, New York, NY, USA. ACM.

[Miller, 1956] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97.

[Moise, 1974] Moise, E. E. (1974). *Elementary Geometry from an Advanced Standpoint*. Addison-Wesley.

[Moore et al., 2000] Moore, D., Periakaruppan, R., Donohoe, J., and Claffy, K. (2000). Where in the world is netgeo.caida.org? *CAIDA Publications*.

[Mukherjea, 1999] Mukherjea, S. (1999). Information visualization for hypermedia systems. *ACM Comput. Surv.*, page 6.

[Mukherjea and Foley, 1995] Mukherjea, S. and Foley, J. D. (1995). Visualizing the world-wide web with the navigational view builder. *Comput. Netw. ISDN Syst.*, 27(6):1075–1087.

[Najjar, 2008a] Najjar, J. (2008a). *Actual use of metadata*. PhD thesis, Katholieke Universiteit Leuven.

[Najjar, 2008b] Najjar, J. (2008b). *Tracking Contextualized Attention Metadata*, pages 93–117. Faculteit Ingenieurswetenschappen, Katholieke Universiteit Leuven.

[Najjar et al., 2004] Najjar, J., Klerkx, J., Ternier, S., Verbert, K., Meire, M., and Duval, E. (2004). Usability evaluation of learning object indexation: the ariadne experience. In Remenyi, D., editor, *Proceedings of the 3rd european conference on e-learning*, pages 281–290.

[Najjar et al., 2005] Najjar, J., Klerkx, J., Vuorikari, R., and Duval, E. (2005). Finding appropriate learning objects: An empirical evaluation. *Lecture Notes in Computer Science*, pages 323–335.

[Najjar et al., 2003] Najjar, J., Ternier, S., and Duval, E. (2003). The actual use of metadata in ARIADNE: an empirical analysis. In *Proceedings of the 3rd Annual ARIADNE Conference*, pages 1–6. ARIADNE Foundation.

[Nielsen, 1992a] Nielsen, J. (1992a). Evaluating the thinking-aloud technique for use by computer scientists. *Advances in human-computer interaction*, 3:69–82. ISBN: 0-89391-751-6.

[Nielsen, 1992b] Nielsen, J. (1992b). Finding usability problems through heuristic evaluation. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 373–380, New York, NY, USA. ACM.

[Nielsen, 1992c] Nielsen, J. (1992c). The usability engineering life cycle. *Computer*, 25(3):12–22.

[Nishizeki and Rahman, 2004] Nishizeki, T. and Rahman, M. S. (2004). *Planar Graph Drawing*, volume 12 of *Lecture Notes Series on Computing*. World Scientific Publishing CO. Pte. Ltd.

[Ochoa and Duval, 2006a] Ochoa, X. and Duval, E. (2006a). Relevance ranking metrics for learning objects. In *Proceedings of the 1st international workshop on Contextualized Attention Metadata*, pages pp. 9–16, New York. ACM Press.

[Ochoa and Duval, 2006b] Ochoa, X. and Duval, E. (2006b). Use of contextualized attention metadata for ranking and recommending learning objects. In *CAMA '06: Proceedings of the 1st international workshop on Contextualized attention metadata: collecting, managing and exploiting of rich usage information*, pages 9–16, New York, NY, USA. ACM.

[OLIVE, 1997] OLIVE (1997). On-line library of information visualization environments. http://www.otal.umd.edu/Olive/. Last retrieved: April 2008.

[Orzechowski et al., 2007] Orzechowski, T., Ernst, S., and Dziech, A. (2007). Profiled Search Methods for e-Learning Systems. *First International Workshop on Learning Object Discovery & Exchange*.

[Pampalk, 2006] Pampalk, E. (2006). *Computational Models of Music Similarity and their Application in Music Information Retrieval*. PhD thesis, Vienna University of Technology, Vienna, Austria.

[Plaisant, 2004] Plaisant, C. (2004). The challenge of information visualization evaluation. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 109–116, New York, NY, USA. ACM.

[Prause et al., 2007] Prause, C. R., Ternier, S., de Jong, T., Apelt, S., Scholten, M., Eisenhauer, M., Vandeputte, B., Specht, M., and Duval, E. (2007). Unifying learning object repositories in mace. *Proceedings of the First International Workshop on Learning Object Discovery & Exchange (LODE'07)*.

[Processing, 2008] Processing (2008). Processing. http://processing.org/. last retrieved: April 2008.

[Quemada et al., 2004] Quemada, J., Huecas, G., de Miguel, T., Salvachùa, J., Fernandez, B., Simon, B., Maillet, K., and Lai-Cong, E. (2004). Educanext: a framework for sharing live educational resources with isabel. In *WWW Alt. '04: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pages 11–18, New York, NY, USA. ACM.

[Revealicious, 2008] Revealicious (2008). Revealicious. http://www.ivy.fr/revealicious/. last retrieved: April 2008.

[Rivadeneira and Bederson, 2003] Rivadeneira, W. and Bederson, B. (2003). A study of search result clustering interfaces: Comparing textual and zoomable user interfaces. *Studies*, 21:5.

[Rosling and Rosling, 2006] Rosling, H. and Rosling, O. (2006). *New software brings statistics beyond the eye*, pages 522–530. Organisation for Economic Co-operation and Development.

[Rubin, 1994] Rubin, J. (1994). *Handbook of Usability Testing*. Number ISBN: 0-471-59403-2. Wiley Technical Communication Library.

[Sacco, 2000] Sacco, G. M. (2000). Dynamic taxonomies: A model for large information bases. *IEEE Trans. on Knowl. and Data Eng.*, 12(3):468–479.

[Schroeder et al., 1993] Schroeder, W., Martin, K., and Lorensen, B. (1993). *The Visualization Toolkit, Third Edition*. Kitware Inc.

[Shashikant et al., 2004] Shashikant, P., Mane, K., and Börner, K. (2004). A Toolkit for Large Scale Network Analysis. SLIS SLISWP-04-02, Indiana University.

[Shi et al., 2007] Shi, K., Irani, P., and Li, P. C. (2007). Facilitating visual queries in the treemap using distortion techniques. In *HCI (8)*, volume 8, pages pp. 345–353.

[Shneiderman, 1996] Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *VL '96: Proceedings of the 1996 IEEE Symposium on Visual Languages*, page 336, Washington, DC, USA. IEEE Computer Society.

[Shneiderman and Bederson, 2003] Shneiderman, B. and Bederson, B. B. (2003). *The Craft of Information Visualization: Readings and Reflections*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[Shneiderman and Johnson, 1991] Shneiderman, B. and Johnson, B. (1991). Treemaps: A space filling approach to the visualization of hierarchical information structuress. In *Proceedings of Proceedings of IEEE Information Visualization '91*, pages 175–282. UM Computer Science Department; CS-TR-2657.

[Shneiderman and Plaisant, 2004] Shneiderman, B. and Plaisant, C. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley.

[SIMILE, 2003] SIMILE (2003). The simile project. http://simile.mit.edu/wiki/SIMILE:about. last retrieved: April 2008.

[Simon et al., 2005] Simon, B., Massart, D., van Assche, F., Ternier, S., Duval, E., Brantner, S., Olmedilla, D., and Miklos, Z. (2005). A Simple Query Interface for Interoperable Learning Repositories. In *Proceedings of the 1st Workshop on Interoperability of Web-based Educational Systems*, pages 11–18.

[Sompel et al., 2004] Sompel, H. V. D., Nelson, M. L., Lagoze, C., and Warner, S. (2004). Resource harvesting within the oai-pmh framework. *D-Lib Magazine*, vol. 10(nb. 12).

[Spoerri, 2004] Spoerri, A. (2004). Coordinated views and tight coupling to support meta searching. *Coordinated and Multiple Views in Exploratory Visualization, 2004. Proceedings. Second International Conference on*, pages 39–48.

[Strijker, 2004] Strijker, A. (2004). *Reuse of Learning Objects in Context: Human and Technical Aspects*. PhD thesis, Universiteit Twente.

[Sun, 2008] Sun (2008). Enhanced graphics and imaging for the java platform. http://java.sun.com/products/java-media/2D/whitepaper.html.

[SVG, 2003] SVG (2003). W3c scalable vector graphics (svg) format. last retrieved: April 2008.

[Takatsuka and Gahegan, 2002] Takatsuka, M. and Gahegan, M. (2002). Geovista studio: A codeless visual programming environment for geoscientific data analysis and visualization. *The Journal of Computers & Geosciences*, 28(10):pp. 1131–1144.

[Ternier, 2008] Ternier, S. (2008). *Standards based Interoperability for Searching in and Publishing to Learning Object Repositories*. PhD thesis, Katholieke Universiteit Leuven - Faculteit Ingenieurswetenschappen, Belgium.

[Ternier and Duval, 2003] Ternier, S. and Duval, E. (2003). Web services for the ARIADNE knowledge pool system. In *Proceedings of the 3rd Annual ARIADNE Conference*, pages 1–9. ARIADNE Foundation.

[Ternier et al., 2008] Ternier, S., Massart, D., Campi, A., Guinea, S., Ceri, S., and Duval, E. (2008). Interoperability for searching learning object repositories, the prolearn query language. *D-Lib Magazine*, 14(1/2).

[Theng and Thimbleby, 1998] Theng, Y. L. and Thimbleby, H. (1998). Addressing design and usability issues in hypertext and on the world wide web by re-examining the "lost in hyperspace" problem. *JUCS*, 4(11):839–855.

[TreeMap, 2003] TreeMap (2003). Treemap java library. http://treemap.sourceforge.net/. last retrieved: April 2008.

[Tufte, 2001] Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press.

[UNECE, 1968] UNECE (1968). Convention on road signs and signals.

[Van Wijk, 2006] Van Wijk, J. (2006). Views on visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):421–433.

[Venn, 1880] Venn, J. (1880). On the diagrammatic and mechanical representation of propositions and reasonings. *Philosophical Magazine and Journal of Science S. 5.*, 9(59).

[Verbert et al., 2005] Verbert, K., Jovanovic, J., Gasevic, D., and Duval, E. (2005). Repurposing learning object components. In *Proceedings of WOSE05: OTM 2005 Workshop on Ontologies, Semantics and E-Learning*, Lecture Notes in Computer Science, pages pp 1169–1178. Springer. ISSN0302-9743 (Print) 1611-3349 (Online).

[Verbert et al., 2004] Verbert, K., Klerkx, J., Meire, M., Najjar, J., and Duval, E. (2004). Towards a global component architecture for learning objects: An ontology based approach. In Meersman, R., Tari, Z., and Corsaro, A., editors, *On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops*, volume Volume 3292/2004 of *Lecture Notes in Computer Science*, pages 713–722, Agia Napa, Cyprus. Springer Berlin / Heidelberg.

[Verbert et al., 2008] Verbert, K., Ochoa, X., and Duval, E. (2008). The alocom framework: Towards scalable content reuse. *Journal of Digital Information*, 9(26). URL: http://journals.tdl.org/jodi/article/view/225/199.

[Wang et al., 2006] Wang, Y., Teoh, S. T., and Ma, K.-L. (2006). Evaluating the effectiveness of tree visualization systems for knowledge discovery. In *Proceedings of Eurographics Visualization Symposium*, pages 67–74.

[Ware, 2004] Ware, C. (2004). *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[Ware and Franck, 1994] Ware, C. and Franck, G. (1994). Viewing a graph in a virtual reality display is three times as goodas a 2D diagram. *Visual Languages, 1994. Proceedings., IEEE Symposium on*, pages 182–183.

[Weaver, 2004] Weaver, C. (2004). Building highly-coordinated visualizations in improvise. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*, pages 159–166, Washington, DC, USA. IEEE Computer Society.

[Weinberger, 2007] Weinberger, D. (2007). *Everything Is Miscellaneous: The Power of the New Digital Disorder*. Times Books.

[Wenger, 1996] Wenger, E. (1996). Communities of practice: The social fabric of a learning organization. *HealthCare Forum Journal*, 39(4):20–26.

[Wenger, 1998] Wenger, E. (1998). *Communities of Practice: learning, meaning, and identity*. PhD thesis, Cambridge University Press, New York.

[White et al., 2008] White, S., JO'Madadhain, Fisher, D., and Boey, Y. (2008). Jung: Java universal network/graph framework. http://jung.sourceforge.net/.

[Wijk and van de Wetering, 1999] Wijk, J. J. V. and van de Wetering, H. (1999). Cushion treemaps: Visualization of hierarchical information. In *INFOVIS '99: Proceedings of the 1999 IEEE Symposium on Information Visualization*, page 73, Washington, DC, USA. IEEE Computer Society.

[Wiley, 2002] Wiley, D. (2002). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. *The Instructional Use of Learning Objects*, 3(23).

[Wiley, 2007] Wiley, D. A. (2007). *The Learning Objects Literature*, pages 345–353.

[Wills, 1999] Wills, G. (1999). Nicheworks - interactive visualization of very large graphs. *Journal of Computational and Graphical Statistics*, 8(2):pp. 190–212.

[Wolpers et al., 2006] Wolpers, M., Gunnar, M., Najjar, J., and Duval, E. (2006). Attention metadata in knowledge and learning management. *Proc. 6th International Conference on Knowledge Management (I-Know 2006), Journal of Universal Computer Science (ISSN 0948-695x)*.

[Wong et al., 2006] Wong, P. C., Rose, S. J., George Chin, J., Frincke, D. A., May, R., Posse, C., Sanfilippo, A., and Thomas, J. (2006). Walking the path: a new journey to explore and discover through visual analytics. *Information Visualization*, 5(4):237–249.

[XML, 2006] XML (2006). W3c extensible markup language (xml). last retrieved: April 2008.

[YouTube, 2008] YouTube (2008). Youtube. www.youtube.com. last retrieved: April 2008.

[Zamir and Etzioni, 1998] Zamir, O. and Etzioni, O. (1998). Web document clustering: a feasibility demonstration. In *SIGIR '98: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 46–54, New York, NY, USA. ACM.

[Zimmermann et al., 2007] Zimmermann, B., Meyer, M., Rensing, C., and Steinmetz, R. (2007). Improving retrieval of reusable learning resources by estimating adaptation effort. In *Proceedings of the First International Workshop on Learning Object Discovery & Exchange (LODE-2007)*, pages 46–53. CEUR Workshop Proceedings.

# Appendix A: Overview Evaluation Tasks: Flexible Access to a LOR

1. How many objects are classified with human and social sciences, human sciences, history?
2. How many objects can you find that were created between 1990 and 1998 on java?
3. Find the document Tree Seeds for Farmers introduction and find out the following information on it:
   - the author
   - classification & concepts
4. Find the number of available material in English on Informatics/Information Processing.
5. How many documents exist on java that are classified with Informatics/Information processing. How many documents exist on java that are classified differently?
6. Find a document on the Farmers Toolkit that is freely available for everybody. Find out the document format of this result.
7. Find out if the amount of freely available material in the classification category of human & social sciences is bigger or smaller than the amount of non-fee material in that classification category.