TAMPERE UNIVERSITY OF TECHNOLOGY
Degree Programme in Information Technology

JAAKKO SALONEN

KNOWLEDGE MANAGEMENT FOR VISUALISING SOCIAL ONLINE SERVICES

Master of Science Thesis

Examiners: Senior researcher Ossi Nykänen (TUT)
Professor Seppo Pohjolainen (TUT)

Examiners and topic approved in the Information Technology
Department Council meeting on 12 February 2007

PREFACE

This thesis work was done at the Hypermedia Laboratory at Tampere University of Technology between June 2006 and August 2007, as part of my work in European Union's Network of Excellence project, OPAALS.

Even though the work on this thesis has taken a lot of my time and efforts, it certainly has been very rewarding. Yet, it is very unfortunate that only a fraction of the knowledge and wisdom I have gained during this process has ended up to the actual work as written words and illustrations.

There are a lot of people who have aided me during this task. First of all, I wish to thank my examiners. Secondly, my most grateful thank yous to all the nice people at Hypermedia Laboratory for both directly helping me out and creating truly inspirational and motivating working environment. My special thanks go to my collegues Anne-Maritta Tervakari and Jukka Huhtamäki, who – despite not being formally involved with my thesis work – assisted me in considerably many aspects of the work. Also, for proof-reading my work twice during the process, I wish to thank my good friend Tutta Kesti.

Finally, I wish to thank my family and my beloved friends - for being there for me.

Tampere, August 2007

Jaakko Salonen
Opiskelijankatu 4 D 685
33720 Tampere

CONTENTS

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Degree Programme in Information Technology
SALONEN, JAAKKO: Knowledge management for visualising social online services
Master of Science Thesis, 69 pages, 4 appendix pages
September 2007
Major: Hypermedia
Examiners: Senior Researcher Ossi Nykänen and Professor Seppo Pohjolainen
Keywords: Web-based services, Online Communities, Knowledge Management, Visualisation, Social Software, Wikis, Blogs

Typically in information systems, the technical infrastructure is built upon fixed structures that are determined by the used tools rather than by the community of its users. Community-managed systems, especially digital ecosystems, should, however, be able to control not only the evolution of the community, but also the evolution of structures and tools.

In this Master's thesis, we study knowledge and its management in web-based services as a guideliner for creation of community-managed systems and application of visualisation methods. We believe that web-based services, especially those supported by social software tools, are pioneers in building systems for online communities and community-based content building. Our research hypothesis is that by studying knowledge and its management in web-based services, we can identify key practises of knowledge management and point out potential areas for application of visualisation technigues.

For the study, we chose 19 of current web-based services, selected according to researcher's own experience and openness of communities. Content and metadata in the services where studied as an expert evaluation. According to found characteristics of content and metadata management, we classified the communities into four groups, after which one community from each group was selected for scenario-based analysis. Basing on this analysis, we classified the practises in knowledge management and visualisation.

As a result of the study, we identified four essential knowledge areas in web-based systems, and potential areas of applying methods of visualisation. Two of the knowledge areas are more focused on content management, while the other two are more emphasised on community management. As such, these results may be utilised to guideline the application of visualisation methods for community-managed systems.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Tietotekniikan koulutusohjelma
SALONEN, JAAKKO: Tiedonhallinta sosiaalisten verkkopalveluiden visualisoimiseksi
Diplomityö, 69 sivua, 4 liitesivua
Syyskuu 2007
Pääaine: Hypermedia
Tarkastajat: erikoistutkija Ossi Nykänen ja professori Seppo Pohjolainen
Avainsanat: verkkopalvelut, verkkoyhteisöt, tietämyksen hallinta, visualisointi, sosiaalinen media, wikit, blogit

Tietämyksen hallinnan tekninen infrastruktuuri rakentuu tietojärjestelmissä tyypillisesti järjestelmän määräämien kiinteiden rakenteiden varaan. Digitaalisten ekosysteemien ja muiden yhteisöllisesti hallittujen järjestelmien tulisi kuitenkin kyetä hallitsemaan paitsi yhteisön, niin myös sisällön ja sen rakenteiden evoluutiota.

Työ tavoitteena on tutkia tietämystä ja sen hallintaa nykyisissä avoimissa verkkopalveluissa, reunaehtona yhteisöllisesti hallittujen järjestelmien ymmärryksessä sekä visualisoinnin menetelmien soveltamisessa näissä järjestelmissä. Tutkimuksen kohdentaminen avoimiin erityisesti sosiaalisen median välineiden tukemiin verkkopalveluihin perustui siihen olettamukseen, että nämä palvelut ovat pioneereja yhteisöllisesti hallittujen järjestelmien ja verkkoyhteisöiden luomisessa. Oletuksena on, että tutkimalla tietämyksen ja sen hallinnan piirteitä näissä palveluissa, voitaisiin tunnistaa yhteisöllisen tiedonhallinnan keskeisiä piirteitä sekä potentiaalisia sovelluskohteita visualisoinnin eri menetelmille.

Tutkimusotokseen valittiin 19 verkkopalvelua tutkijan oman kokemuksen ja palveluiden avoimuuden perusteella. Palveluiden sisältöä ja metatietoa tutkittiin asiantuntija-arviointina. Palvelut luokiteltiin sisällön ja metatiedon hallinnan piirteiden perusteella neljään ryhmään. Yksi palvelu jokaisesta ryhmästä valittiin tarkempaan analyysiin, jossa hyödynnettiin skenaarioita - eräänlaisia tyypillisiä käyttötarinoita. Tämän analyysin perusteella työssä eriteltiin palvelujen käyttämiä tietämyksen hallinnan ja visualisoinnin keinoja.

Työn tuloksena tunnistettiin verkkopalveluista neljä keskeistä tietämyksen aluetta sekä näihin liittyviä potentiaalisia sovelluskohteita visualisoinnin eri menetelmille. Tietämyksen alueista kaksi painottuu sisällönhallinnan tarpeisiin, kun taas toiset kaksi vastaavat paremmin yhteisön hallinnan erityispiirteisiin. Kartoitusta voidaan hyödyntää erityisesti tuotaessa visualisoinnin eri menetelmiä osaksi yhteisöllisesti hallittuja järjestelmiä.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
CMS	Content Management System
DCMES	Dublin Core Metadata Element Set
DE	Digital Ecosystem
FOAF	Friend of a Friend
JHS	Julkisen Hallinnon Suosituksia
ICT	Information and Communications Technology
OPAALS	Open Philosophies for Associative Autopoietic Digital Ecosystems
OKS	Open Knowledge Space
RDF	Resource Description Framework
URL	Uniform Resource Locator
WYSIWYG	What You See Is What You Get

1. INTRODUCTION

In this chapter, we introduce background for the work, define key concepts and give an overview of the work.

1.1. Background

The world-wide web is more than just a unidirectional, publishing medium. Web-based services offer a wide variety of facilities for users of the web, and online communities have emerged from groups of people carrying out discussions, communicating and interacting over the web.

O'Reilly describes that the second generation of these online communities and web-based services, or the web 2.0, has revolutionised the way people use the web. Caused by the shift from Internet as a medium to Internet as a platform, web 2.0 has changed the way of how online software is seen. In web 2.0, software and services are not specified and created by precious few, but rather created by engaging communities of users in the design [Ore05].

Users of web 2.0 software are treated rather as co-developers than just as end-users or service consumers (cf. [Ore05]). The kinds of systems that utilise this pattern of "perpetual beta", are one step further of giving the community of users the control over system's management. In community-managed systems - from which web 2.0 services are not very far from - the creation of technical infrastructure is built upon structures that are completely determined by the community of the users and not dictated by the technical contributors.

Web 2.0 leaves much for hoping in regard of social inclusion[1] and web accessibility[2]. As Zajicek depicts, web 2.0 services are often built for "young socially integrated

1 Social inclusion as reducing inequalities between the least advantaged groups and communities and the rest of society

2 Web accessibility in the meaning that people with disabilities can use the Web [Sha06]

people who own their own laptop and live in a world of readily available radio LAN and fast access broadband". Meanwhile, the inaccessible rest will become increasingly excluded from mainstream life. As such, web 2.0 improves infrastructure for those who already enjoy great adaptability, while disregarding the rest. [Zaj07].

Additionally, while many services from the web 2.0 breed can be used freely and without direct cost, they do not always give users the freedom to control their data. In these services it is not unusual that data - along with digital identity, communities and services - are thereof easily bound to the providers of these services.

While web 2.0 might the best support the needs of technologically able young people, the approach of digital ecosystems take the concept of community management more inclusively. According to Nachira and Louarn, digital ecosystem can be seen as "a pervasive ICT infrastructure with a particular architecture and framework, which is collectively defined and built following the specific requirements defined through a multi-stakeholder participative process, and which exhibit some characteristics of the natural ecosystems". When digital ecosystems in the web are considered, the focus is on web as a platform for knowledge and services. [Nac06].

In European Union, the digital ecosystems research is carried out by a network of excellence project, OPAALS[3]. Funded by the Sixth Framework Programme, the project aims to build a sustainable, interdisciplinary research community for digital ecosystems. From the research point of view, OPAALS aims to establish a theoretical foundation for digital ecosystems research and - along with it - build a digital ecosystem around the participants. [OPA06].

In the context of OPAALS, digital ecosystems are seen as a novel approach for the catalysis of sustainable regional development driven by small and medium-sized enterprises (SMEs). The claim is that archiving sustainable digital business ecosystems of SMEs and software components require in-depth knowledge on collaborative processes and ICTs (Information and Communications Technologies) that underpin the continuous creation, formalisation, and sharing of knowledge in the form of business models, software infrastructure for e-Business transactions, and new formal and semi-formal languages. [OPA06].

The creation of a digital knowledge ecosystem in OPAALS is approached from the concept of open knowledge space (OKS) that is developed as a seed for the creation of a distributed, self-organising knowledge work environments. As both the efforts for formalising the definition of such space and the building of it are ongoing, there is not much we can define straight away. [OPA06]. The term *open knowledge space (OKS)* describes the way how knowledge is viewed in OPAALS's digital ecosystems research.

3 OPAALS as an acronym of "Open Philosophies for Associative Autopoietic digitaL Ecosystems"

1.2. Scope of the work

In this master's thesis, we study knowledge and its management in web-based services for application of visualisation methods. We see that even though currently available web-based services do not readily offer the infrastructure for digital ecosystems, they are pioneers in building systems for online communities and community-based content building.

Our research hypothesis is that by studying knowledge and its management in web-based services, we can identify key practises of knowledge management and point out potential areas for application of visualisation techniques. The application of different visualisation methods in practise, however, is outside the scope of this work, as our interest lies more on understanding how and where the methods would be applicable.

1.3. Definitions

1.3.1. Web-based services and online communities

Web-based services are online services that offer services for end-users of the web. The definition should be distincted from more technical terminology such as web services or web APIs (Application Programming Interface), which are often used in the context of machine-to-machine interaction over the web.

In this context, *service* refers to mediation of immaterial commodities. According to Grönroos, *services* are more or less immaterial set of acts that are produced and consumed more or less concurrently. The consumer, i.e. the users of the service, is seen also to have a role in the production process. As such we may see services as subjective experiences from the consumer point of view. [Grö90].

An online community consist of a group of people that primarily or initially communicate or interact by using collaborative or social software tools, which in this case consist of web-based services. The concept has much resemblance with that of virtual communities, presented by Rheingold. According to his definition, these ”...[virtual] communities are social aggregations that emerge from the Net when enough people carry on public discussions long enough, with sufficient human feeling...” [Rhe95].

1.3.2. Collaboration

Collaboration is about working together, joining intellectual efforts resulting in a shared base of knowledge. It involves both producing collaborative knowledge and communication about how the work is done. Hence, collaborative knowledge is such knowledge that both has been produced and is managed collaboratively. In the view presented by Awa, Ghaziri and Hassan, we could state that collaboration is a key

process supporting knowledge cultivation [Awa04].

All online communities are collaborative at some level, as characteristics of collaboration arise from continuous social interaction. Yet, the depth of the collaboration is much dependent on the available tools and technical infrastructure. It could be well said that quality and quantity of available social and collaborative tools dictate much of the nature of collaborative work and social interaction.

1.3.3. Visualisation

One of the challenges in community-managed systems, and digital ecosystems in particular, is how the knowledge can be managed and visualised in distributed, yet collaborative and socially aware environments. As in community-managed systems, digital information does not restrict only to building of the digital content but also to that of the community, practises that enhance insight to both aspects are much needed.

Visualisation is seen as a practise in which data is represented in order to give better insight to the underlying knowledge. We see that visualisation potentially complements the intensions of knowledge management in general in the cultivation of organisational knowledge. Successfully applied practises of visualisation could well intensify organisational knowledge creation and therefore promote innovation.

1.4. Overview

In Chapter 2 we review theories of knowledge to the extent that is relevant to our study. We also take a look into different paradigm according to which knowledge is interpreted in interactive systems.

In Chapter 3 we deepen our understanding on visualisation and study of it as a process in which data is transformed into visual representations, and ultimately interpreted by human beings.

In Chapter 4 we present the research methods and the research sample used in the study.

In Chapter 5 we perform knowledge analysis on our research sample, present the results and conclusions. According to the analysis, we select four web-based services for more detailed study. Several scenarios are presented, familiarising the reader with the selected services.

In Chapter 6 we perform qualitative analysis of the scenarios presented in chapter 5 and study how they reflect knowledge management and visualisation practises in the web-based services.

In Chapter 7 we summarise the work and present our conclusions.

2. KNOWLEDGE IN THE WEB

In this chapter we review theories of knowledge to the extent that is relevant to our study. We also take a look into different paradigm according to which knowledge is interpreted in interactive systems.

2.1. Knowledge

The classification of different types of knowledge is more of a philosophical issue rather than something we are easily able to do explicitly. There are numerous different terms for defining information in detail, of which data, information and knowledge are perhaps most used. These terms, however, are remarkably ambiguous and easily mixed up.

The classical definition of *knowledge* is justified true belief[4]. The definition implies that these "objects of knowledge" are beliefs that are both justified and true. The types of warrants that justify a belief could be discussed, but are outside of the scope of this work[5]. The definition is far from perfect in terms of logic, but is generally a definition that Western philosophers have agreed on (See for instance [Non95, 21]).

It should be noticed that in different domains, there are diverging definitions of knowledge. As we have to study theories from different domains, we must deal with same terms possessing divergent meanings. Thus, for disambiguating concepts between different domains, we express the context along the term. For instance when we refer to knowledge in organisational theories, we talk about organisational knowledge.

As an integrating approach between different domain, we define the concepts of knowledge, information and data, according to Lowe and Hall as follows [Low99]:

4 The view of knowledge as justified true beliefs is seen to originate from Plato dialogue Thaetetus, where Socrates considers different theories of what knowledge might be.

5 For more about different types of knowledge, see for example Pecorino's writing [Pec00] on the matter

- **Knowledge** is the base of personal information which is integrated in a fashion which allows it to be used in further interpretation and analysis of data.
- **Information** is the interpretation of data within a context set by *a priori* knowledge and the current environment.
- **Data** are the artifacts which exist as a vehicle for conveying information.

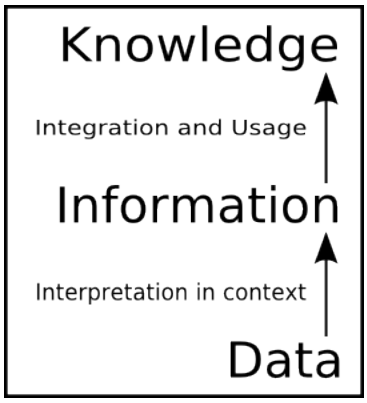


Figure 2.2: *Data, information, knowledge and their relations [Low99]*

Let us clarify these definitions by making an example. Let us consider that we read a line of text from a book. The interpretation of the data, the black ink on the paper, relies on a priori knowledge of the reader, e.g. information that has been acquired earlier. This knowledge includes understanding meanings of the symbol, words and linguistic rules. According to this understanding, different kind of knowledge may be acquired from the same piece of text with different, preceeding knowledge.

Our definitions for data, information and knowledge are also in coherence with the classical definition of knowledge as imposed by constructivism[6]: contextual interpretation invokes personal justification and hence implied knowledge becomes true personally. In this way, the issues of truthfulness rely on personal, subjective understanding of reality.

2.1.1. Dimensions of human knowledge

On the matter of human knowledge, Michael Polanyi has written that "we can know more than we can tell" ([Pol67, p. 4] according to [Smi03]). According to this statement, not all knowledge possessed by humans, is shaped in a way that can be readily communicated.

Polanyi has separated two different types of human knowledge: tacit and explicit knowledge. *Tacit knowledge* is personal, context-specific, and therefore hard to formalise and communicate. *Explicit knowledge (*or "codified" knowledge), refers to knowledge that is transmittable in formal, systematic language. ([Pol66] according to [Non95]).

We see that the separation of tacit and explicit knowledge could be described as the difference between practise and theory. Where tacit knowledge is gained through experience, explicit knowledge is the theory or the rationality behind things that are gradually formed in interaction with the environment. Nonaka and Takeuchi see that

6 "Constructivism refers to the idea that learners construct knowledge for themselves---each learner individually (and socially) constructs meanings---as he or she learns" [Hei91].

Polanyi's separation of knowledge types contends human beings acquiring knowledge by actively creating and organising their own experiences. Thus, the knowledge that can be expressed in words and numbers, represent only a part of the entire body of [human] knowledge. [Non95].

While the idea of tacit knowledge is presented by Polanyi in philosophical context, Nonaka and Takeuchi see that it can be expanded on a more practical direction. According to them, tacit knowledge can be seen to include both cognitive and technical elements. Cognitive elements center on mental models, such as schemata, paradigms, perspectives, beliefs, and viewpoints that help individuals to perceive and define their world. Technical elements of tacit knowledge, on the other hand, include concrete know-how, crafts and skills. [Non95].

2.1.2. Creating and communicating human knowledge

Based on Polanyi's dual nature of human knowledge, Nonaka's and Takeuchi's model of knowledge conversion assumes that human knowledge is created and expanded through social interaction between the two types of knowledge. According to the two types of human knowledge, Nonaka and Takeuchi define four different modes of knowledge conversion (as illustrated also in figure 2.3) [Non95]:

- **Socialisation.** The process in which tacit knowledge is directly mediated between [human] individuals. The process is largely supported through personal, direct interaction.
- **Externalisation.** A process in which tacit knowledge is articulated into explicit knowledge. This is done by giving tacit knowledge in the shape of metaphors, analogies, concepts, hypotheses, or models.
- **Combination.** The process of systemising concepts into a knowledge system. Individuals exchange and combine knowledge through such media as documents, meetings, telephone conversations, or computerised communication networks.
- **Internalisation.** A process in which explicit knowledge is translated to tacit knowledge. At the level of individuals, this could be interpreted as accumulating the personal know-how in the sense that can be then directly utilised or socialised.

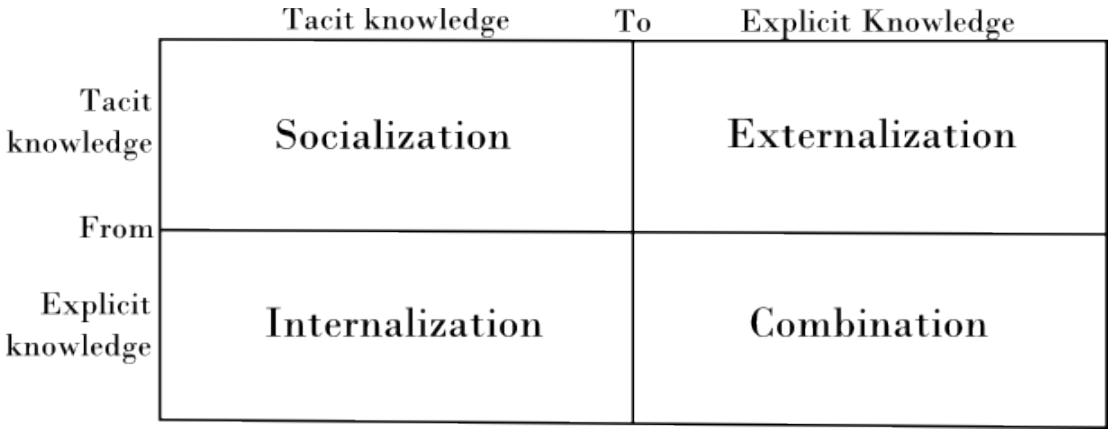


Figure 2.3: Four modes of knowledge conversion [Non95]

Nonaka and Takeuchi see that this conversion occurs socially between people rather than is confined within an individual. Under favourable circumstances the knowledge conversion process becomes circular and repeating. By enabling the complete four-staged process of knowledge conversion, an organisation is seen to leverage its members' accumulation of tacit and explicit knowledge, but will also foster the creation of new knowledge. [Non95].

2.1.3. Can computers possess knowledge?

Knowledge building comes natural to humans, but machines do not possess this ability by nature. When studying knowledge in interactive systems, the question that naturally rises is: can computers also possess knowledge?

Computers are able to compute. In computational processes, data is simply processed and transformed in mathematically just processes. If a computer would operate on the level of knowledge, it should have ability to understand the information beyond mechanical processing. (See especially [Rus03]).

In the field of artificial intelligence (AI), it is believed that intelligent entities, that think as intelligently as humans, could be built. Assumptions of *weak AI* state that machines can be programmed to act intelligently, but may not actually think. According to assumptions of *strong* AI, it has been predicted that machines could be able to actually think, instead of simply emulating the process of thinking [Rus03].

Whichever hypothesis is taken, it is clear that computers are able to handle information, e.g. interpret and analyse data they possess, if programmed to do so. Interpretation of data is based on how computers are programmed to understand data. The description of the format in which data is encoded into these internal structures is called data *schemata* (pl. for *schema*).

With sophisticated programming, computers could be programmed even in a fashion that enables them to handle knowledge, e.g. integrate and use information and act

intelligently according to what is known. The ability of computer programs for being able to make decisions and act on it without being reprogrammed, is called *reasoning. Semantics*, e.g. programmatically defined rules of how information should be understood, govern the process of reasoning. Thus, how computers are able to process *knowledge* depends on how knowledge is represented to them. (Adapted from [Rus03]).

Our main concern on computers is on how computers could be able to assist human users as *intelligent agents*[7]. By assisting humans on knowledge-level tasks, like information integration and usage, computers should be able to handle tasks in which the completion would otherwise require human attention.

2.1.4. Ontologies

A knowledge agent builds its conception of the reality on some a priori knowledge. According to Gruber, this kind of personal context could be called a conceptualisation: an abstract view that gives us a view of the world for some purpose. We see that every knowledge agent is committed to some conceptualisation: it would be impossible to make any interpretations of the existance without such specification. Thus, what "exists", is exactly what can be represented. [Gru93].

In order to share our conceptions of the existing, we need an explicit specification of knowledge conceptualisation, an *ontology,* for knowledge representation (See for example [Gru93]). By sharing such specification and committing on them, different knowledge agents will able to communicate on the level of knowledge.

As a knowledge management practise, ontologies offer technical framework for describing arbitrary information semantics as an explicit specification. Typically, ontology-committed information is modelled as entities and relations between them. Thus, practically any knowledge can be specified according to a select ontology language.

The entities and their relations must be well defined in order to execute semantic queries to a specified set of data. In practise, this means that the underlying knowledge repository must commit on a well-defined, formal ontology description language such as the Web Ontology Language[8] (OWL).

Ontologies are often defined and populated by expert users which hides dealing with the inconsistencies and uncertainty in information from end-users (For example [Hyv05]). When access to information authoring is given to a large base of users, variations in understanding and use of the ontology concepts will occur, as not all of the end-users are likely experts of the used domain and it's ontologies.

7 An intelligent agent is something that is able to act on behalf of others
8 http://www.w3.org/TR/owl-features/

2.1.5. Challenges in ontologies

Ontologies, however, should not been seen just as a technical framework for describing information. Hepp presents that building ontologies is, by its nature, "a social process constrained by technical, social, economic, and legal bottlenecks". Drawing on his conclusions, there are four main obstacles that will be faced on ontology related processes [Hep07]:

- **Conceptual dynamics.** As ontologies rise from the user needs, may the involved communities of practise define new concepts or find current ones obsolete. The need for new concepts rises as time passes and current culture become history.
- **Economic incentive.** Ontology building requires resources like human labor. Especially if building is not integrated to the other knowledge work, economic feasibility is something that should be taken into account.
- **Ontology perspicuity.** Using an ontology for knowledge modelling is about trusting its creator and committing to it. A good ontology is something that users are willing to commit to and therefore something that can be understood.
- **Intellectual property rights.** Intellectual property rights apply also to standards, controlled vocabularies and other existing systems of classification. Exporting industrial standards to ontologies will most certainly require legal agreements.

Thus what we gain from the richer level of semantics, formalisation and query abilities, are results of successfully created knowledge community with technical understanding or such alike tools for ontology management. Basically this requires that the community should be able to commit to and perhaps to even build ontologies as the needs arise.

2.2. Knowledge in organisations

Usually in organisational theories, knowledge is defined as actionable information, e.g. information that can be used to pursue certain tasks (See [Awa04, pp. 40-41]). According to the definition of knowledge we assumed from Lowe and Hall, we see that organisational knowledge is focused on the subsection of all available knowledge that is relevant in the context of the organisation in focus.

Awad and Ghaziri present *knowledge management (KM)* as a newly emerging, interdisciplinary business model that has knowledge within the framework of an organisation as its focus. The practise of knowledge management involves people, technology and organisational processes in overlapping parts (as in figure 2.3). The goal of KM accordingly, is to present a balanced view of how computer technology captures, distributes, and shares knowledge in the organisation by linking human experts and

documented knowledge in an integrated KM system [Awa04].

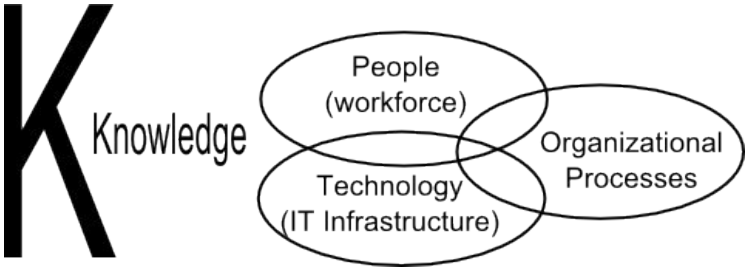


Figure 2.3: *Overlapping human, organisation, and technological factors of knowledge management (adapted from Awa et al. [Awa04])*

Like Nonaka and Takeuchi have also stated, Awad and Ghaziri see that people are the driving, creation force of [organisational] knowledge. They suggest that up to 95% of all knowledge is tacit by its nature, while only 5% can be described explicitly, on paper, in documents and in databases. If we rely on these estimated percentages, majority of knowledge is mediated in processes of knowledge socialisation [Awa04].

In KM studies, all processes in organisation are viewed as knowledge processes. These processes form the a cycle of knowledge management. This cycle of knowledge management in organisation includes knowledge creation, collection and capturing, organisation, refinement, dissemination and maintenance. [Awa04].

2.2.1. Knowledge processes in organisations

A knowledge management system (KM System) is a set of different tools, organisational processes and guidelines that aim at enabling organisational knowledge creation and management. According to Awad and Ghaziri, knowledge organisations consist mainly of people and technology, tied together by different organisational processes. With interfacing of people and technology splitting the knowledge organisation, we can see that there are two levels: informal and formal level. One possible depiction of this kind of a a knowledge management system is illustrated in figure 2.1. (Partly adapted from [Awa04]).

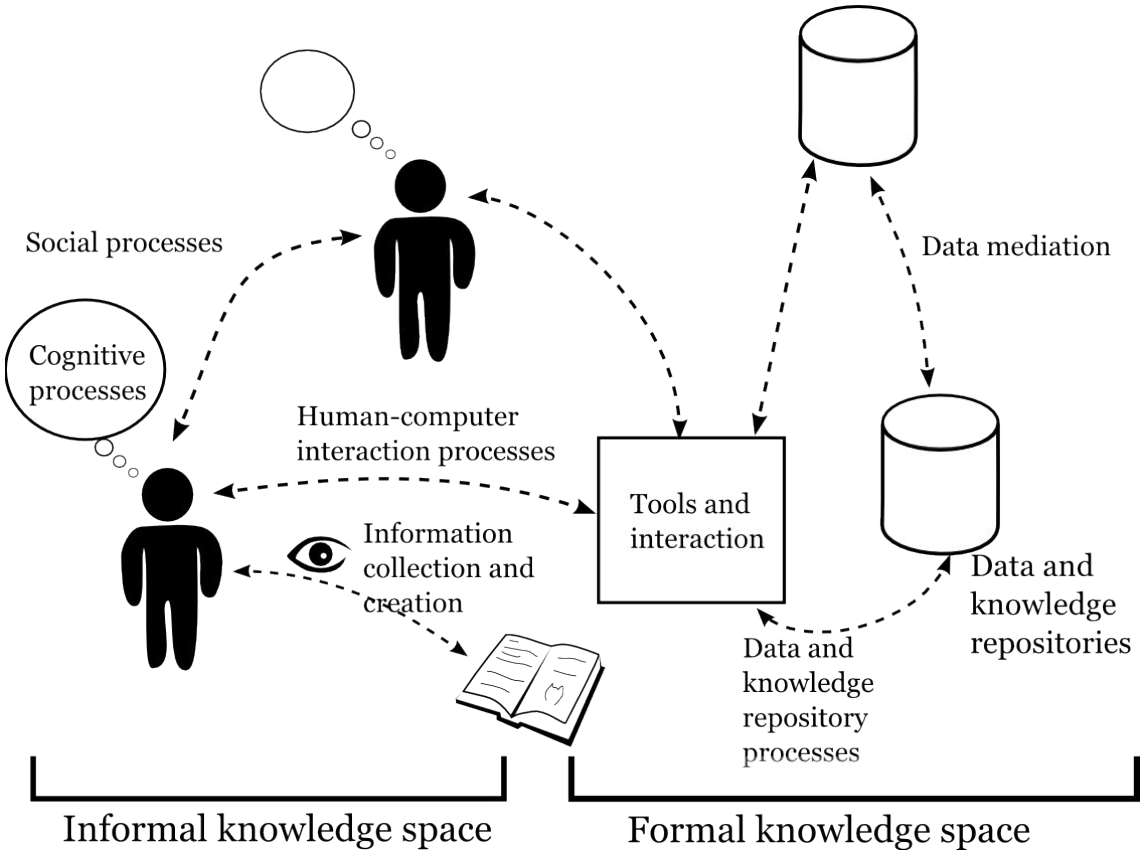


Figure 2.1: *Knowledge management processes in different levels of KM systems*

We can see that knowledge processes in KM systems revolve around four main stages (Adapted from [Awa04] and [Sta01]):

- **Creation and import.** The contents needs to be created or converted so that they fit the conventions of the organisation.
- **Capture.** Knowledge items have to be captured in order to determine their importance and how they mesh with the organisation's vocabulary conventions.
- **Retrieval and access.** This step satisfies the searches and queries for knowledge by the knowledge worker[9].
- **Use**. The knowledge worker will not only recall knowledge items, but will process them for further use.

Informal level of knowledge organisations (as illustrated on the left in figure 2.1) is focused on informal and social processes. The interaction between individuals mainly occurs as direct social interaction, which is usually some form of face-to-face communication, but may also be a technically assisted process like a phone call or video conference. The [organisational] knowledge is informal, but at the borderline formalised

9 Knowledge worker, a term coined by Drucker in 1960s, is used to highlight the importance of knowledge in modern business organisations (See [Awa04, 20])

semi-formally with some sort of tools such as pen and paper.

Formal level of knowledge organisations (as illustrated on the right in figure 2.1) includes the tools and technology that enable access to formal, persistently data repositories. Different kinds of technologies can be used to implement repositories, such as databases or flat-file systems. As the level of formality in data may vary, different levels of data structure granularity occur in these backends. Formal level of knowledge organisations optionally includes the access to computer-assisted knowledge-processing, like computer agents.

2.3. Document management

Many interactive systems, especially different content management systems (CMS), draw heavily on the metaphor of documents in the way they manage knowledge. *Document management* is the practise of managing information in various digital data containers that resemble physical documents. *A document management system* is an application that facilitates different tasks of content management, including content authoring and content browsing. In practises of document management, the actual content in the document, e.g. the physical or virtual container of data, is separated into *data* and *metadata*. Whereas *data* in documents is considered as the main content of interest, *metadata* is description about the data itself.

The separation of these two content types is made on basis of data needs. Data in documents contain the actual information of interest, whereas metadata is used to facilitate the finding, sharing and management of information held in the content of the documents (See for example [Tay03] or [DCMI06]). In JHS[10] 143 specification, different metadata properties are categorised according to their use cases as follows [JHS143]:

- **Identification.** Defines the identity that helps to locate and identify documents.
- **Content description.** Describes the content of the document that helps to classify and search the content. It is recommended to use coherent description conventions such as known vocabularies.
- **Access conditions.** Defines the legislative and other access conditions that might limit the availability and distribution. Elements that govern archiving are also included in the category of access conditions.
- **Context.** Describes the context of the document. Context is seen to be important because it contributes to comprehensibility and interpretability of information. It is a crucial requirement for information documentation as it links people, places and events among others to the documents.

10 As an abbreviation of 'Julkisen Hallinnon Suosituksia'

- **Life-cycle management.** Elements of the category make the up the life trail of the document and track the history. Governing the metadata of document life-cycle may be also important because of legislative purposes.

Various standards for organising the metadata in documents exist. Especially Dublin Core Metadata Element Set (DCMES), has been widely adopted by American, European and International standardisation organisations. According to DCMES specification, the Dublin Core Metadata element set is a framework for describing information about "anything that has identity". In total, the set contains 15 elements that aim to describe an entity of information and supplementary several occurrences of elements [DCES06].

DCMES encourages use of known encoding schemes and controlled vocabularies. In practise, this means that element values are picked among known, well-defined sets of keywords. According to Dublin Core's guideline this gives rise to "dramatic improvements in search results because computers are good at matching words character by character, but weak at understanding the way people refer to one concept using different words, i.e. synonyms" [DCUG05].

2.3.1. Challenges of document management

Even though the paradigms of documents or digital assets do not dictate how content is structured to the documents, it assumes that knowledge is managed in document entities enriched with metadata. Well-defined schemas may be useful from computational point of view, but may be challenging from users' point of view as understanding standard sets of metadata and schema definitions may call for expertise on the domain of informatics. This not only leaves a gap between users and creators of the schemata, but may also restrict the ways how metadata may be used to manage documents.

Among many, the criticism of metadata has been laid out by Cory Doctorow, who has presented seven "insurmountable obstacles" that exist in the way of reliable and thus useful metadata. We see that these points generally reflect two challenges of metadata (Adapted from [Doc01]):

- **People may have alterior motives and beliefs.** Not all share the complete meaning of the schema, nor have time or the interest to focus on authoring complete and descriptive metadata.
- **Schemas are limited and easily biased**. Even though schemas aim to be complete, they easily bias towards something and are limited to certain uses. Readily built schemas also facilitate the phenomenon "what you measure is what you get".

It could be well said that metadata in practise is not only about managing metadata. It is

always about managing the metadata authoring process and the users, both authors and end-users. With systems that rely completely on document management paradigm, the users do not entirely control the ways of how content is managed.

2.4. Social awareness

One of the reasons why interactive systems might lack the support for more intricate social interaction, is that the design rationale is based on the conception that collaboration and social interaction take place completely outside the digital infrastructure. Accordingly, it is not uncommon that the role of interactive systems in organisational knowledge management is seen as supporting, not as enabling in, the whole organisational system.

2.4.1. Social translucence

According to Erickson and Kellogg, "we humans as social creatures are immersed in a sea of social information". It is well worth noticing that many of our every day decisions are made according to social hints. Erickson's and Kellogg's claim is that also the production and use of knowledge are "deeply entwined with social phenomena". When designing system of communication and collaboration, Erickson and Kellogg state that this side of knowledge is frequently neglected. They see that digital systems are still commonly completely opaque to social information, whereas need for such does not cease to exist. [Eri00].

Even though information systems should make social phenomena visible, without some degree of transparency, Erickson and Kellogg see that "it is difficult to conduct a long-running, productive conversation through the digital medium". [Eri00]. This is confirmed for example by Franco et al.: "People who communicate through computers do not receive the nonverbal cues that can enhance meaning or soften harsh words." [Fra95]. Thus, digital systems that do not acknowledge the social aspects of knowledge management, are not sufficient for truly digital knowledge work.

Erickson and Kellogg list three features that are important in social processes [Eri00]:

- **Visibility**. Humans are more attuned to movement, human faces and figures than to printed signs or documents. Also information systems should encourage this kind of visibility of human actors.
- **Awareness.** Awareness if about being aware of each other. Perhaps most importantly it is about user knowing to whom his or her actions have an impact on. The element of awareness helps to attenuate socially undesirable actions by bringing culture's social rules into play.
- **Accountability**. Vague origin of statements may contribute to insecurity and

therefore there should be a way to cultivate liability: "I know that you know that I know and therefore I will be held accountable for my actions". Flames and taunting behaviour could be interpreted as symptoms of a system with insufficient framework for accountability.

Complete transparency of social information is, however, discouraged by Erickson and Kellogg. They state that social processes should be more *translucent* by nature than completely transparent. According to them, translucency is necessary because "there is a vital tension between privacy and visibility" [Eri00]. It is easy to make an example about this in ordinary social context: when talking to a trusted good friend, the characteristics of conversation may be entirely different from talk aimed to a large, public audience.

2.4.2. Social software

The tools that enable collaborative work are usually called *collaborative software* or *social software*. Whether the software is seen more as “social" or "collaborative", is a matter of applying software for a use case than an issue of the software itself. We could say that in comparison to "ordinary" software, social software tools depend "more on social convention than on software features to facilitate interaction and collaboration" [Sen04].

In order to better understand what types of collaborative or social tools are available, we make a classification of them (See for instance [Wik07a]). The complete taxonomy based on this distiction is illustrated in figure 2.2. The two different main types are synchronous and asynchronous tools.

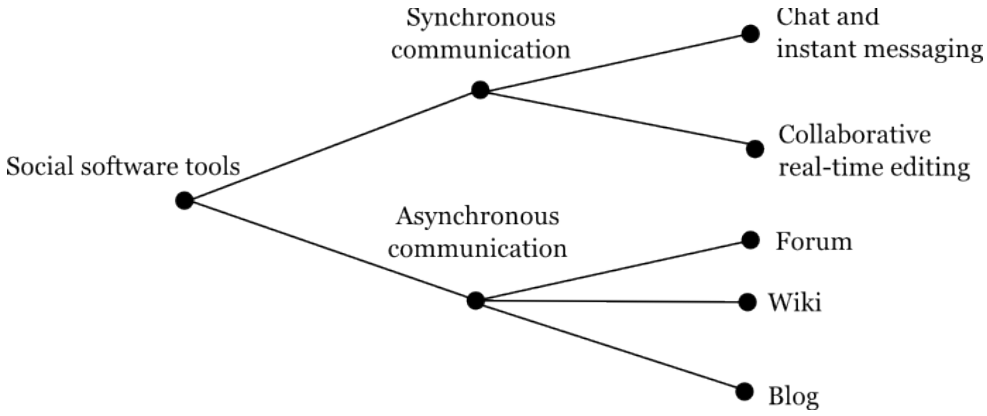


Figure 2.2: A taxonomy for social software tools (Adapted from [Wik07a])

Synchronous communication tools include tools of instant messaging, text chat and collaborative real-time editing [Wik07a]. In these occasions, the interaction is "in situ", leading to fairly informal and instant communication. The context built by the presense can be fairly well utilised in the communication. The most common types of synchronous tools can be classified and described by two main types: chat and instant

messaging and collaborative real-time editing.

Chat and instant messaging are a form of communication where short messages are instantly mediated between two people or a group of people. Term chat and instant messaging can be used synomously, while typically instant messaging typically occurs between two people whereas chatting usually takes place in a chatroom, containing larger group of participants.

Collaborative real-time editing, is likewise a synchronous form of communication, but with the distiction that participants share a workspace they can collaborate in. The key characteristic in such tools is that participants may not only make changes at the same content, but they also are able to perceive the made changes, leading to greater social awareness.

Asynchronous communication tools consists of tools that do not require immediate presence for collaboration. This kind of software includes internet forums, blogs (or weblogs) and wikis [Wik07a]. Whereas synchronous tools could utilise the richness of context, have asynchronous forms of communication more formalised procedures of communication. The most common types can be classified into three categories: forums, blogs and wikis.

Internet forums (or simply forums) work as facilities for holding discussions and posting user generated content, forming a social place for an online community. User contributions are organised into chronological threads of discussions which are further organised into topics or "sub-forums". Access to content is restricted to user's own postings. [Wik07b].

A blog (or weblog) is a website where contributions are made in entries and displayed usually in reverse chronological order [Wik07c]. Visitors of a blog may contribute by leaving comments, but usually blog softwares support cross-blog references technically with a peer-to-peer communication technologies, such as TrackBack (See [Tro02]).

Wikis are socially managed online databases that have little technical restrictions for content management. Unlike in blogs or forums, users may typically edit all content, leading to cumulative contributions rather than separate, consecutive edits. (See for example [Leu01]). Content in many wikis is not very highly structured, but there are also wiki tools for creating very structured and well-defined content.

2.5. Digital ecosystems

In the paradigm of *digital ecosystems* (DEs), systems are viewed as organisational systems with self-organising digital infrastructure. with self-organising digital infrastructure. The paradigm excavates the fundamental principles from biological ecosystems aiming to build distributed, self-organising environments that do not require

centralised management [Nac05].

A digital ecosystem (DE) is a self-organising digital infrastructure for networked organisations. Nachira has defined the infrastructure as follows:

> "The digital ecosystems infrastructure is a pervasive 'digital environment' which is populated by 'digital components' which evolve and adapt to local conditions thanks to the re-combination and evolution of its 'digital components'. 'Digital components' could be: software components, applications, services, knowledge, business processes and models, training modules, contractual applications, services, knowledge, business processes and models, training modules, contractual frameworks, law." [Nac05]

A digital ecosystem has decentralised infrastructure and thus, should not require any centralised management. While centralised infrastructures are typically based on server-client model in which few servers provide services for numerous clients, in digital ecosystems any single participating entity may function as a service provider or as a service consumer. Thus, a digital ecosystem is a network formed of independent organisations (cells), which communicate through the network.

Decentralisation enables the participants to slowly develop new practises, e.g. *evolve*, in regions of the ecosystem while maintaining existing connections with current infrastructure. In centralised systems, the evolution of structures and services in interactive systems is handled by renewing complete infrastructure at once, for instance as software or database schema updates. These all-at-once evolutionary changes usually require notable devotion from technical contributors.

Like living organisms, digital ecosystems should be able to adapt: grow or shrink according to needs. Digital ecosystems should not be able to not only evolve content, structures and services, but also be able to add or remove new independent organisations. Thus, infrastructures for digital ecosystems should be able to adapt new systems to the evolutionary infrastructure.

When digital ecosystems in the web are considered, the focus is on web as a platform for knowledge and services. Each web-based service could be interpreted as an environment for a digital web ecosystem, supporting the needs of its users. On a large scope, we could also see the web as a digital service ecosystem, consisting of web-based services as digital agents, which we do not, however, consider as the focus in this work.

3. UNDERSTANDING VISUALISATION

In this chapter we deepen our understanding on visualisation and study of it as a process in which data is transformed into visual representations, and ultimately interpreted by human beings.

3.1. Why visualisation matters?

According to Ware, information visualisation has five potential benefits: it helps to comphrehend huge amounts of data, to perceive emergent properties of the system, to evaluate correctness of the data, to facilitate understanding of both large- and small-scale features and finally, to faciliate hypothesis formation. [War04].

Visualisation also potentially complements the intentions of knowledge management: the process in which organisational knowledge is cultivated. Successfully applied practises of visualisation could be well a key factor in itensifying organisational knowledge creation and therefore promote innovation.

3.2. Defining visualisation

In the literature, there seems to be several conceptions of what visualisation is about. WordNet defines *visualisation* as "a mental image that is similar to visual perception" and as synonym of *visual image* which may be also "a percept that arises from the eyes; an image in the visual system" [Wor07a].

While visualisation on the other hand is seen to involve formation of a "visual" mental image, it meanwhile should arise "from the eyes". In the regard that the very basic concept of visualisation has not an easily agreeable foundation, it is not surprising how diversely it has been interpreted in different domains.

3.2.1. Scientific visualisation

Scientific visualisation, as phrased by McCormick et al., is the "mechanism by which humans perceive, interpret, use and communicate visual information". Scientific visualisation is seen as an alternative for mere numbers, as means to provoke insights and to communicate about them with others. We see that the underlying idea in scientific visualisation is the conversion of numerical data into visual representations. Thus, input according to such methods of visualisation, consists of numerical data. [McC87].

Foley and Ribarsky, on the other hand, see that scientific data visualisation is not necessarily restricted visual data representations. Disputably, they see that visualisations can be any kinds of perceivable data bindings. Accordingly, a visualisation can be not only visual, but also aural, tactile, as long as the binding is perceivable. Where the visual bindings are created by assigning data into visual properties like color, shape, size, orientation and texture, may for instance auditory data bindings map data onto pitch or spatial localisation. [Fol94].

Robertson sees that the whole concept of [scientific] visualisation is underlain by the idea that an observer can build a mental model as a result of the data representation process. While it is important to define how visual representations of data are created, it should also be seen that cognitive processing with mental models is the key factor when interpretation of visualisations are concerned. [Rob91].

3.3. Information visualisation

According to Ng, *information visualisation* is "the use of computer-supported interactive visual representations of abstract data to aid understanding and analysis". Ng also states that in many works on visualisation including his, the visual representations are graphical with text strings mainly serving the purposes of content labelling and annotation. [Ng00].

Ng also points out that in constrast to scientific visualisations, information visualisation does not necessarily deal with measurable, physical phenomena. Instead, information visualisation generally deals with abstract data that does not have intentional spatial components. Examples of such abstract data are vast, including financial data, documents and hypermedia and include much additional freedom for choosing effective visual representations. [Ibid.].

The lack of spatial information is a challenge, as in information visualisation, there is no obvious starting point. On the other hand, the freedom for choosing all spatial properties offers the opportunity for pre-visualisation data mining. By *data mining*, we refer to the act of extracting (or "mining") knowledge from large amounts of data [Han01].

Data mining based knowledge discovery is in general based either on descriptive or predictive methods. While *descriptive mining tasks* characterise the general properties of data, *predictive mining* perform inference in order to make predictions. The kind of patterns that can be mined in such tasks, include data characterisation and discrimination, association analysis, classification and prediction, cluster analysis, outlier analysis and evolution analysis. We see that any kinds of data mining tasks can take place before or during the information visualisation process [Han01].

3.4. Process model for visualisation

Regardless of how the data is obtained or how the sensory bindings of data are made, it seems to make a lot of sense to understand visualisation as a data representation process. We can see that there are two distinguishable strategies for creating visualisations as a result of data representation process: information visualisation and data mining. These different methods of data processing should not be seen as exclusionary, but rather as completementary to each other. While *information visualisation*[11] draws primarily on the *deductive properties* of data, the strategies of *data mining* enrich raw data in terms of *induced properties*.

Ware has defined a four-staged model for the process of visualising information (as seen in figure 3.5). The four stages of this process are 1) the collection and storage of data 2) preprocessing and transformation 3) rendering of visual data representation and 4) perceptual and cognitive processing of the visualisation. [War04].

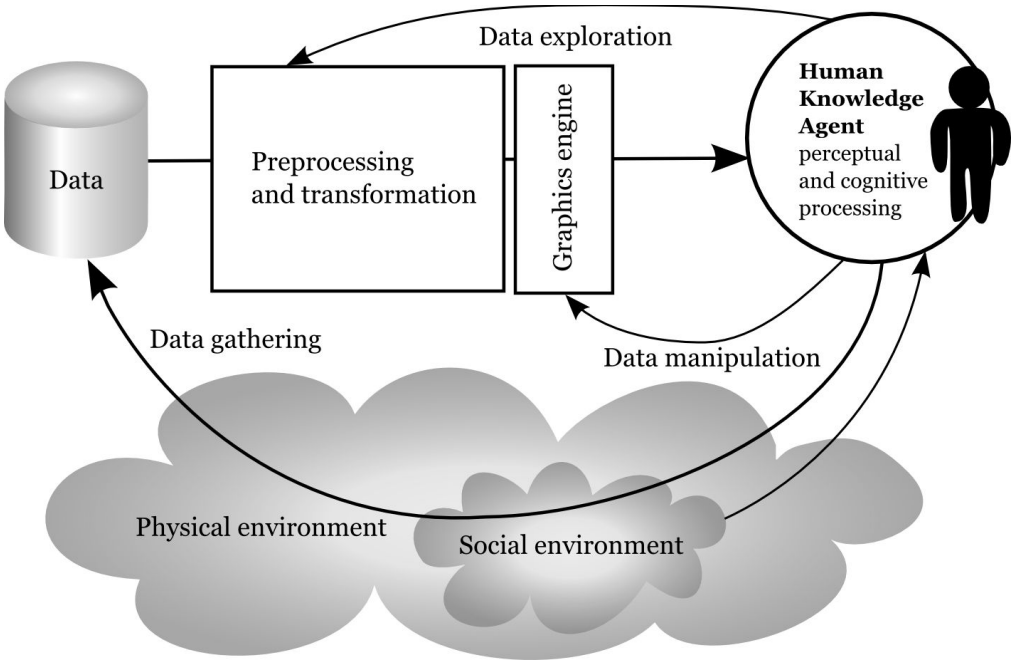


Figure 3.5: *Process model for visualisation (Adapted from [War04])*

11 And also scientific visualisation thereof

Data gathering is the subprocess in which the data is collected and stored. According to Ware data is gathered by a human (actor) in the physical and social environment. Ware defines that physical environment provides the source of the data, whereas social environment influences the way how data is interpreted and what is collected. Ware sees that the stage of data gathering includes all data analysis, the human information analyst chooses to execute. After the data is presented to the visualisation system, automated information processing will take over the control. [War04].

Secondly, preprocessing and transformation is the stage in which the gathered data is processed into human interpretable format [War04]. In this stage, data is preprocessed and transformed into structures that include visual properties. On the other hand, data that is already presentable in visual format, may not require any preprocessing or transformations at all.

In the third stage, visual presentation perceivable by a human knowledge agent, is created. The display hardware and a set of graphics algorithms produce an image on the screen or if the extended definition of visualisation as arbitrary data representation is assumed, any sensory feedback may be produced by the visualisation hardware. [War04].

In the fourth stage, human information analyst processes the produced visualisation visually and cognitively [War04]. It should be acknowledged that equal sensory data will yield different interpretations by various analysts.

3.5. Notes on data gathering

Simply gathering raw data is not sufficient in order to enable visualisation in an understandable sense: if the method of data gathering does not provide sufficient means for capturing the semantics of the data, they must be entered explicitly by the knowledge agent in order to preserve the original context.

3.5.1. Human-entered versus automatically collected data

One classification for the data gathered from the socio-physical environment is the separation of computer collected and user-entered data.

Computer collected data can be automatically gathered by the computer. Current date and time, application status and other computer contextual data are already available for the computer. This is mostly the can of data that can be harnessed in providing the kinds of mediated visualisations, as also audio and video feeds essentially originate from digital sources.

Human entered data is everything that users enter to the computer by using different interfaces, provided. In some sense they are also computer collected data, but with a

degree of interpretation involved. For example, a computer can collect the keypresses, but when they are combined and understood as written words, they present data entered by the user. User entered data usually also involves some interaction with the system so that entering user feeds to the system usually is a process rather than an event.

3.5.2. Types of data

We may see that data has a certain structure that aims to conceptualise knowledge. Again, one possibility to classify data is presented by Ware. He suggests that there are two fundamental forms of data [War04]:

- **Entities**[12] are generally the objects of our interests. Entities can be for example peoples or documents. Also groups of entities can be considered to be of a single entity, if convenient.
- **Relationships** relate entities to each other. There can be many kinds of these relations, for example an apple might be related to a human so that an apple is owned by one such entity.

In addition to relationships, entities may also have *attributes:* abstract properties strongly bound to entities themself. Ware sees that using an attribute to model data is approriate when it is a property of an entity and cannot be thought of independently. Especially numerically-valued types are better described as attribute values rather than entity to entity relationships. In Ware's model, the attributes are classified into three different kinds [War04]:

- **Category attributes**. Serves the function of labelling. There is generally no sense of order in which different categories can be placed into.
- **Integer data attributes**. The use of number in different purposes such as to order things or express a scale. Integers can be used also to generate ordered sequences.
- **Real-number data attributes**. in the meaning of intervals and ratio scales.

According to Ware, many visualisation techniques are capable of conveying all categorial, integer and real-number data attributes. This is also the perspective taking in the entity modelling: we do not focus on the comprehensive classification of different data types, but instead on how such data types can actually be used in different visualisations. [War04].

Practises of document management are keen to make a difference between the actual data, (the primary content of interest) and metadata (data supporting content discovery). From the viewpoint of visualisations design, the representational problems are the same

12 Or as Chen has nicely put it: "Entity is a 'thing' which can be distinctly identified" [Che76].

for both, and hence, no distinction between data and metadata is made when visualisation is concerned (See for example [War04]). We see that both data and metadata form entities, relationships and attributes, regardless of the primary content management tasks.

3.5.3. Evoluationary aspect to data

In the entity-relationship-model, we should keep in mind that in time, there will be evolution of all entities, relations and attributes. From this aspect, it is reasonable to say that we may require visualisations for both the actual structure and relations of the concepts and the underlying process in which the entities inspection evolve. According to this evoluationary viewpoint, visualisations may be divided into two groups: structure and process visualisations.

Structure visualisations illustrate entities and relations between them, as according to our presented visualisation ontology. Structural visualisations may include attributes and numeric data encoded. This structural information can be encoded using space, color, shape or any other channel of encoding.

Process visualisations represent the evolution of entities, their relationships and attributes within the scope of inspection. In accordance to our visualisation ontology, this means representating the evolution of entities, relationships and attributes associated with them. Typically these processes of information evolution are visualised with animations, but may also be conveyed using any other method of encoding.

Separating process and structure visualisation aims to highlight the fact that content authoring tools should not always be limited to capturing the most recent revision of data. The need of perceiving history and evolution of information are, in author's opinion, fairly easily forgotten when designing interactive system.

3.6. Presentation versus representation

The visualisation process model we have presented, does not imply, whether data is presented in its original context. By original context, we mean the overall settings that the visualisation was produced in and designed for. According to this criterion, visualisations can be thus classified into mediated visualisations and emergent visualisations.

Mediated visualisations generally aim to reproduce visualisation as in its original, ”real-world”, context. Such visualisations are video clips, images, digital representations of books and any of other type that try to resemble the sensory stimuli in a preserved context. Mediated visualisations in particular aim to preserve the original context.

Emergent visualisations utilise different computational methods that transform the

underlying data by the means of calculations that do not preserve, but rather generate new sensory stimuli. Such emergent visualisations can be further classified into different types by the origin method used for newly discovering the data: statistical visualisations and deduced visualisations.

Statistical visualisations utilise statistical calculus in order to generate new insight to data. Unlike in emergent visualisations, statistical visualisations may be used to predict data and thus, to discover trends. Statistical visualisations may as well enable extraction of the kind of tacit or implicit features that are not yet formalised.

Deduced visualisations are logically derived from the underlying data. Therefore a selection of logical calculi are utilized such as tuple-, predicate- or proposional calculus to name a few. From the view of knowledge, such calculi can be seen to emulate human knowledge usage and integration in different levels of sophistication[13]. In general, deductive calculus simply manifests knowledge by reformalising what is already known via a set of well-defined proportions.

3.7. Visualisation as an interface to data

An aspect worth noticing is that we do not make any difference between the end-user interfaces or the other possible visualisation clients: they all provide different kinds of view to the same base of data. Of course different kinds of user interfaces may have varying data manipulative features. From this basis we acknowledge that from the user interface point of view, the data-driven visualisations can be divided into two groups: interactive visualisations and static visualisations.

Interactive visualisations are such presentations of the data that provide users with both a representation of the contained data and means to directly manipulate it. In this sense such visualisations work also as a user interface. To put it the other way around, any graphical user interface is a kind of interactive visualisation of underlying information.

Fixed visualisations provide users only with a one-way view to the data. They might allow manipulation of different, contextual parameters or navigate the data, but do not provide edit access to the actual data source. Common examples of such visualisations could be different graphs that use numeric data sources or other kinds of derived presentations of data. Some types of fixed visualisations are by nature such that intuitive means for manipulating the representation are not even possible to implement.

13 Basically all deductions rely on the computational models of the domain calculus, and hence are limited to its expressive power. A commonly known domain are relational databases that are based on tuple calculus. Also different artifical agents and recommendation systems usually represent information based on deduction.

3.8. Perceptual and cognitive processing

Cognition is what goes on in our heads when we carry out our everyday activities. With *cognitive processing*, we refer to human or humanlike information processing including the process and results of recognising, interpreting, judging and reasoning. According to our view, cognitive processing is based on *internal representations*, e.g. hypothetical internal cognitive symbols that represent external reality. These internal representations form the base of personal information. (Adapted from [Pre94] and [Wik07c]).

A dual coding theory presented by Allan Paivio [Pai06], suggests that human cognition involves the activity of two distinct subsystems. The other of these subsystems is very specialised in dealing directly with language, whereas the other is more of a nonverbal system, specialised in dealing with non-linguistic objects. According to this theory, internal representations can be associated to either one of these two subsystems. [Pai06].

What the dual coding theory states is that both of these subsystems are involved in the representation of tactile information, even language. There are associative, referential connections between these two subsystems that together explain the different responses to verbal and nonverbal stimuli. Even though the importance of individual subsystems are emphasized in different tasks, the associative interplay between these two subsystems is what Paivio sees cognition is made of. [Pai06].

*Perception i*s the process in which external, sensory information is acquisitioned, interpreted, selected and organised, which also may result in modification of internal representations. Saariluoma sees that immediate stimuli of adequate intensity in different sensory receptors are required in order to produce an actual perception [Saa04]. There are several theories for how the actual process of perception works. To evaluate them in detail, we roughly split them into two categories that are *constructivist* and *ecological theories of perception*.

3.8.1. Constructivist view to perception

Constructivist theories of perception, according to Preece et al., see perception as an active process, where world is constructed gradually from perceived information from the environment and from previous knowledge. [Pre94, p. 76].

One such constructivistic theory of perception is presented by Ulric Neisser [Nei82]. He sees that perception is a fundamental cognitive process that other processes are reliant on. According to his model, perception is an ongoing, cyclic process. From such approach, our internal representations are the factors that direct our perception. The information that the perceiver extracts from the sensory data will therefore be interpreted in the context of such priori understading. The new, perceived information will further modify the existing internal representations which again ultimately continue to direct our next circulation of the perpectual process. [Nei82].

3.8.2. Ecological view to perception

Preece et al. explain the ecological approach on the basis of Gibson's, Michael's and Carello's view. According to their view, ecological approach sees perception as a direct process in which information simply is detected, rather than constructed. The primary concern is on how perception is used to deal with continuous events over time. Where constructivists view perception as a mean to recognise objects, the ecologists propose that we actively explore the objects in our environments. By seeing, smelling, listening to, tasting and touching the environment we carry out our activities. [Pre94].

An important concept in ecological approach is affordances. Preece states that *affordances* are what we perceive as the behaviour of a system, object or an event. The statement is generally about how our actions are assumed to and do affect, not on how we reason on the level of schemas. Naturally, affordances that are more obvious, are easier to exploit. Ambiguous or contradictive affordance signals make interaction with the environment more harder as perhaps was intented. [Pre94].

We can interpret that ecological view to perception suggests that the affordances should be matched with the operational goals: the perceptual representation should express the desired functionality of the system as accurately as possible.

4. DESCRIPTION OF RESEARCH METHODS AND RESEARCH SAMPLE

In this chapter we present the research methods and the research sample used in the study.

4.1. Description of research methods

The subject of our research consists of a set of web-based services. The research is split into two parts: knowledge analysis and scenario study. The knowledge analysis will be performed to all services included in our sample. As a result of the analysis, we classify services into groups. From each group of services, one web-based services was selected for more detailed, scenario-based study.

4.1.1. Knowledge analysis

In knowledge analysis, we aim to capture in general the overall quality and quantity of knowledge, according to content available in the web-based service. The knowledge analysis is split into two parts: *concept analysis* and a more detailed *metadata analysis*.

Data from web-based services for the knowledge analysis was collected by actually using the web-based services and thus, by joining and taking part of the community of users. This data gathering method can be seen as resembling a simplified form of ethnographic research[14], as the web-based services are used in situ. This kind of experimental approach to data gathering is seen important, as – include traditional applications – online communities that form around web-based services influence greatly the way the services are used.

14 For more on Ethnography, see for instance [Pre94]

4.1.1.1. Concept analysis

Concept analysis aims to understand how the knowledge can be seen to consist of different concepts and their relations, forming a domain of the knowledge. As we do not aim to understand the ontological domain in great detail, we focus on analysing five key qualities of the content framework: primary concept, number of its occurences, supporting concepts, intended use and content management type.

Primary concept is seen to be the concept that is most heavily weighted and most noticeable. In some cases identifying a primary concept is not obvious. The key heuristics for identifying such concepts are that the quantity of such concepts is most extensive or significant and without such concept the service would not have a reason to exist. The number of primary concept's occurences, is used to give us a rough estimate of the service's knowledge base size.

Supporting concepts are the kind of concepts that can be seen to have an important, possibly supporting role in the creation of the knowledge in the service. Such concepts are typically strongly related to primary concepts and together with primary concepts form a remarkable amount of the knowledge captured in the service. Leaving out secondary concepts would typically greatly reduce the meaning of the service.

Intended use is the (primary) use scenario that the service itself states to be most focal. As that kind of advertising is important also in the case when an arbitrary user seeks to learn about the service, it is also used as one criteria of the heuristics. Intended use will be used as a seed for finding out customary service use cases.

Content management type (with valid values of 'social' and 'technical') defines whether the practises of content management are 1) enforced socially by community of users or 2) more or less dictated by technical structures which the users can not influence. The evaluation of content management type is based on expert evaluator's opinion.

4.1.1.2. Metadata analysis

Metadata analysis aims to discover what kind of metadata are used in managing the content. We choose Dublin Core metadata element set to measure metadata since it is not only a very well recognised standard for cross-domain information resource description, but is also - by far - the most popular extension of RDF metadata model[15].

The heuristics began by selecting a sample of three occurences of primary concepts in the authored content. For each property in metadata element set, the samples are used to analyse the type of a property occurence. All properties were classified to have one of three possible values: implicit, explicit and unknown.

15 Most popular, publicly used extensions at least according to http://ebiquity.umbc.edu/blogger/100-most-common-rdf-namespaces/

Implicit property values are typically of the kind that require use of heuristics involving also some degree of guessing. Implicit properties can typically "read between the lines" and hence, are hard to extract without any knowledge on the context.

Explicit property values can be easily extracted from the data without little or no a priori knowledge on the context. An explicit value may contain also the schema for the property, such as reference to the corresponding Dublin Core element name. Naturally, when a value is explicitly available it will also be counted as implicit.

Even though many properties exist, there are cases when we can not even widly guess the a type of a value property. It is evident that the difference between implicit and unknown values may not be obvious: the interpretation largely depends on the evaluating expert's opinion.

As different occurances of primary concept may have different metadata granularity, the definite classification relies on expert evaluator's consideration. As a rule of thumb, if two of the three occurences of a concept indicate that such value exist, it will be seen as prevalent. If only one of the picked values referred to a type, a half weighting was used with discretion.

4.1.1.3. Measuring content quality and quantity

From all services, we calculate four key figures that characterise the properties of primary concepts in these services. The calculated properties are 1) the number of explicit and 2) implicit properties, 3) difference of explicit and implicit properties and 4) certainty, all according to Dublin Core.

As it has been established earlier, an explicit value is also implicit. However, as an element can be available only implicitly, it might be interesting to see, if such difference is very common. The difference is hence calculated directly, as substraction of implicit and explicit concept counts. Hence

$$Difference = Count_{explicit} - Count_{implicit}$$

Difference value is used to reflect the quality of available metadata. We see that difference value should reflect the formality of the data provided. The more negative the difference value gets, the more imprecisely the values of metadata properties have been defined, and vice versa: higher difference value indicates more formally defined metadata.

As there is a maximum number of 15 elements that all can score from 0, 0.5 or 1 points for both implicit and explicit values, we know that total scoring per service can vary from 0 to 30 points. Therefore we can derive the total coverage of Dublin Core, per score, so that:

$$Coverage = \frac{Count_{implicit} + Count_{explicit}}{Max_{implicit} + Max_{explicit}} = \frac{1}{30}(count_{implicit} + count_{explicit})$$

We see that coverage value should well reflect, how extensively Dublin Core elements have been used in the content. The smaller the coverage value is, the less of Dublin Core metadata elements have been used to describe content. If we assume that Dublin Core is a good description framework for any information resource description, then coverage value should give us an estimate of how well resources have been assigned with metadata.

4.1.2. Scenario analysis

Scenarios are informal narrative descriptions of human activities that allow us to discuss the challenges and possibilities of knowledge management and visualisation in presented web-based services (See for instance [Pre94]). For building the scenarios, we attempt to understand what kind of activities the members typically perform. Our motivation for using scenarios lies on understanding how the software tools, provided by end-user web services, have been intended to be used.

4.2. Research sample

For purposes of our research, we selected 19 of current web-based services, around which online communities have – more or less – emerged. Our focus is not on communities, but on the ways of action communities of users have adopted for the web-based services.

As the author has most knowledge on web-based services he has used himself, the core of the services was selected to those the author himself uses frequently. The core consists of six services: 1) Wikipedia, an online encyclopedia, 2) IRC-galleria, a social networking service, 3) Google Search, 4) YouTube video service and 5) del.icio.us social bookmarking service and 6) Last.fm, an internet radio and an online music database.

This set of services was then refined with popular or otherwise well referenced services, in order to study the types of online communication the author was not readily familiar with. According to the taxonomy of social software tools presented in chapter two, we added examples of popular or unique services utilising these tools, according to following categorised rationale:

- **Chat and instant messaging**: Suomi24 chat, Habbo Hotel
- **Collaborative authoring and wikis:** Discogs, 43 Things, Psychology Wiki, Ontoworld Wiki, WikiHow, Citeseer

- **Personal web publishing and blogs:** Technorati, 43 Things, Flickr
- **Read-only database interfaces:** PIKI online library, MuseumFinland

It is acknowledged that the used approach does not provide us with an extensive web-based service study. As the web-based services that enable the growth of online communities are hand-picked by the author, there will be a vast amount of different types of web-based services that will not be covered by the study. As the purpose of this sampling is to determine guidelines for the actual qualitative study, we see the coverage of the sample should be sufficient.

Table 4.1: Overview to the research sample of online services

#	Service name	Language	Rationale
1	Wikipedia	Int. English / Finnish	Perhaps the most popular wiki-based service
2	IRC-galleria	Finnish	The most of popular social networking service in Finland
3	Google Search	Finnish	Typical (and popular) Internet search service
4	YouTube	Int. English	Popular Internet video service
5	Del.icio.us	Int. English	Popular service
6	Last.fm	Int. English	Social approach to music metadata creation
7	Discogs	Int. English	Collaborative authoring
8	43 Things	Int. English	Social approach to goal management
9	Flickr	Int. English	Popular service, aspects of metadata authoring
10	Technorati	Int. English	Interesting mechanism of metadata aggregation and annotation
11	Psychology Wiki	Int. English	Collaborative authoring of specialist topics
12	Ontoworld Wiki	Int. English	Collaborative authoring of specialist topics
13	WikiHow	Int. English	Collaborative authoring for a long tail of specialists
14	Citeseer	Int. English	Collaborative authoring of academic metadata
15	PIKI online library	Finnish	Classical interface to library database searching
16	MySpace	Int. English	Very popular social networking service
17	Habbo Hotel	Finnish	Popular social networking service / chat with elements of gaming
18	MuseumFinland	Finnish	Interesting approach to content aggregation and searching
19	Suomi24 chat	Finnish	An example of a typical online chat service

5. KNOWLEDGE ANALYSIS AND SCENARIO-BUILDING

In this chapter, we perform knowledge analysis on our research sample, present the results and conclusions. According to the analysis, we select four web-based services for more detailed study. Several scenarios are presented, familiarising the reader with the selected services.

5.1. Knowledge analysis on research sample

5.1.1. Results

Content in total of 19 web-based services was evaluated using the heuristics defined above. The collection of key results from the analysis are presented in table 4.2.

As for the metadata analysis, we found it very difficult to make a clear distinction between implicit and explicit properties. Quite usually, if a property was found to exists, it was found to be defined explicitly. Thus, we can say that this context plays a little role in capturing of Dublin Core metadata.

Even though the implicit and explicit values scattered quite well to the possible values, most of the services had a very similar total certainty for primary concepts. 80 percent of primary concepts had metadata coverage between 57 and 88 percent. This suggest that there are radical differences between how various services handle Dublin Core metadata type of properties.

In figure 5.1 we see the distribution of web-based services on a 2-dimensional space, with coverage and difference as orthogonal axes. As we recall, larger difference illustrates higher numbers of implicit than explicit properties. Therefore implicitly richer elements are on bottom. Coverage distributes services along the x-axis, where services of highest coverage are on the right-hand side.

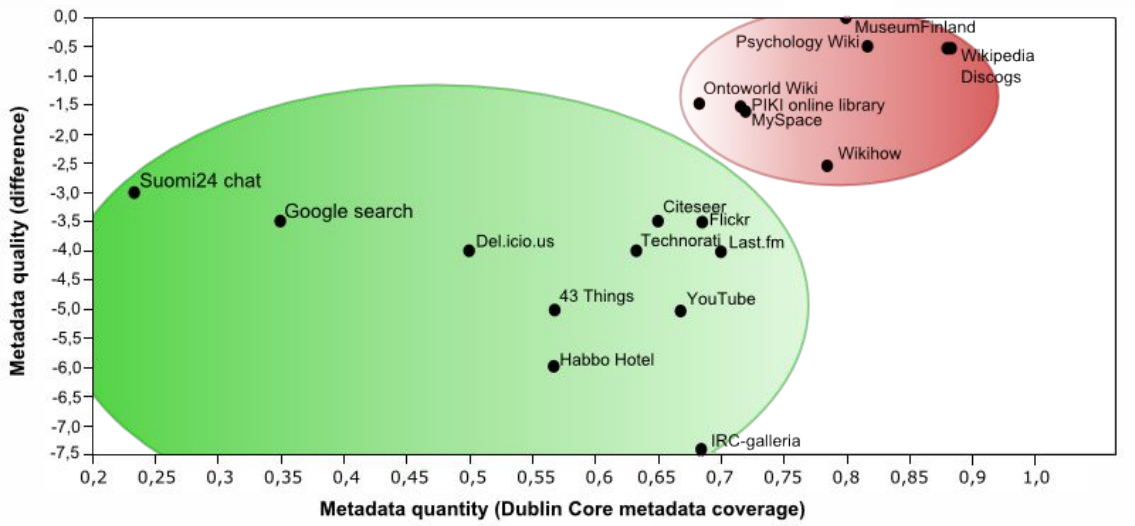


Figure 5.1: *Studied online services distributed by quality and quantity of Dublin Core metadata*

Difference and coverage give much more variation in distribution of services. Some of the services are really close in the terms of our metadata analysis. Directly from the figure, we can see that the set of services consisting of MuseumFinland, PIKI verkkokirjasto, Discogs, MySpace and all the wiki-based services form a set with both high Dublin Core coverage that is explicitly rich.

The result of the services with lower coverage of Dublin Core and greater number of implicit properties consist of the rest: Last.fm, Flickr, Citeseer, YouTube, Technorati, Del.cio.us, Google search, Suomi24 chat, Habbo Hotel and IRC-galleria. The common denominator is that they are very rich in total Dublin Core coverage and have relatively high implicit coverage. These services tend to extend their metadata coverage with implicitly available information.

5.1.2. Conclusions

On the basis of this, we present that there are two diverging orientations: *document orientation* and *data-domain orientation*.

Document oriented services focus on document based content management, offering a document management platform for diverging needs. They serve rich and definite metadata and usually offer good document management capabilities, such as content versioning and access control.

Data-domain oriented services focus on the structural requirements, unique to the domain of concern. Metadata in data oriented services does not conform generally very well to the Dublin Core elements. Metadata in data-domain oriented services "may be there", but sometimes only implicitly, via understanding of the context. The data, however, seems to capture the specific knowledge needs of the domain fairly well.

Table 5.1: Overall results of knowledge analysis

Service	Primary concept	Dublin Core metadata coverage	Supporting concepts	Content management
Discogs	release	88 %	Song, artist, label	technical
Psychology Wiki	article	82 %	category, user	social
Wikipedia	article	80 %	category, user	social
MuseumFinland	historical artefact	78 %	artefact type, material, producer, place, time, user and place of use, place of use, usage situation, collection	technical
Wikihow	article	72 %	category, user	social
MySpace	user	72 %	comment, photo, video, calendar entry, blog message, message, forum post, forum category, event	technical
PIKI web library	library resource	72 %	-	technical
Last.fm	played song	70 %	user, artist, song, album, label	technical
Ontoworld Wiki	article	68 %	category, user, attribute, predicate	social
IRC-galleria	user profile	68 %	photo, channel, community, comment, diary message	social
YouTube	video clip	67 %	user, message, channel, group, category	technical
Citeseer	citation	65 %	document, acknowledgement	technical
Flickr	photo	63 %	user, collection, tag, group	social
Technorati	post	63 %	blog, tag, user	technical
43 things	things	57 %	people, place, entry, photo	social
Habbo Hotel	user	57 %	chat message, room	social
Del.icio.us	bookmark	50 %	user, tag, bundle	technical
Google Search	search result	35 %	-	technical
Suomi24 chat	chat message	23 %	user, nickname, chatroom	social

The two presented orientation, however, do not explain all the differences in content management. Especially the level of abstraction in concepts vary fairly greatly: wiki-based services contain usually only articles and users, whereas document oriented services like MuseumFinland and Discogs define fairly intricately the valid structure for content. The same seems to apply for data-domain oriented services: Last.fm is very specific about structure of data, whereas Del.icio.us can be very genericly used.

In a way, the level of abstraction in content can be seen as an issue of how the content is managed. Services with more abstract and generic content management framework, delegate tasks of content management from technical tools to the community its of users, We see that this orientation of content management can be used to split the content management practises into two type: socially managed content and technically managed content.

Drawing on the selected two properties, content management type and Dublin Core metadata coverage, we have classified our services into four different types according to their content and metadata management practises:

- **Socially managed content and metadata**. Both content and metadata are managed by the community, rather than by the technical framework. Participation is an important aspect as most of the content is community created. Content management practises favor convention over configuration leading to little or no available metadata.
- **Technically managed metadata, social content.** Structure of actual content is socially managed, yet content is document based. Content in documents is mostly defined socially by the community. Used framework for document management provides automation for metadata productions and documents are typically rich of it.
- **Technically managed content with social metadata.** Actual content is managed in domain-spefic structures, yet management of metadata is more social than technical. These kinds of services are typically rich of community-created metadata that is usually used to create different kinds of social navigation systems.
- **Technically managed content and metadata.** Structure of both content and metadata is technically enforced. Content may be created collaboratively, but has typically low context and high commitment to domain ontologies.

These categories will be used as the basis for selecting the services for content authoring analysis. Next chapter, we shall select a service from each of these categories and create scenarios of content browsing and authoring accordingly.

5.2. Scenario-building for IRC-galleria

With more than 390,000 users, IRC-galleria is the largest web-based online community (service) in Finland. When originally founded in late 2000, the service was built to be a photo gallery for Finnish users of Internet Relay Chat (IRC). The community, however has notably outgrown from photo sharing IRC users. Nowadays it is actually a fairly significant medium with some estimates stating almost 25 percent of Finnish people using it weekly[16]. Especially teenagers are found to be keen users, with estimatedly 65 percent of people aged between 15 and 17 visiting daily. [17].

IRC-galleria is maintained by a Finnish company Dynamoid Oy[18]. The company has also introduced several, relocalised versions of the original Finnish gallery for other regions. In this study, however, we inspect only the original, Finnish IRC-galleria the author himself is most familiar with.

5.2.1. Intended use

It is not very clearly stated, what is the ultimate purpose for a user to use IRC-galleria. The service's own online tour presents rather mechanisticly the key functionalities of the online service[19]:

- Introduce yourself in your profile page,
- Comment other users and their pictures,
- Organise your photos into albums,
- Search for a friend [user] with the tool provided,
- Join and create communities,
- Write your own web diary,
- View your "guestbook" to find out who has visited your page.

The functionalities provided do not directly suggest any well-defined use cases, however quite clearly they underpin different forms of social interaction between users of the service. Profile enables sharing of personal information along with the photos and diary entries. The channels and nickname integrate IRC-galleria to the interaction in actual IRC, where the communities form new associations between users within the service. Commenting further enables discussing via photos.

5.2.2. Scenario 1: daily socialisation in IRC-galleria

One of the key purposes of IRC-galleria and many other user oriented and social

16 http://irc-galleria.net/tour.php
17 http://irc-galleria.net/mediakortti.php
18 http://www.dynamoid.com/yritys
19 http://irc-galleria.net/tour.php?page=intro

networking services is the ability to navigate in social networks and find out information about it's actors. In the use case the user is performing a routine sign in to the service and checking the service for changes.

Story 1

18-year-old Lilli is an average user of IRC-galleria. Living with her parents, she is going to high school for the last year. She spends in average 75 minutes per day using IRC-galleria. An important thing in her life are her friends which she has many. Four out of five are also users of IRC-galleria. (Profile adapted from [Dyn07]).

One of the key motives for Lilli to use IRC-galleria is to keep in touch with her friends. Again, she opens the frontpage of the site and signs in to the service. The logged in view of front page is pretty similar, however small orange number before the first menu item indicates that she has received comments. She clicks the number ending up to the picture to which the comment is associated to.

The new comment is from one of Lilli's friends, commenting on the new picture she has put to the gallery. Reading it, Lilli clicks on her friends profile, scrolls down the page and writes a brief answer under the picture page that is behind the link.

Answering her friend, Lilli ends up browsing her pictures. Casually she checks her friends gallery thumbnails for new pictures, clicks some of them to see new comments and looks for changes in the profile. Her friend is also writing a diary which Lilli also checks for new entries.

Lilli repeats the same pattern for some other users that are in some way associated with her. Especially users sharing communities with her are potentially interesting.

5.3. Scenario-building for Wikipedia

Wikipedia is a multi-lingual, web-based free content encyclopedia project. The encyclopedia contained by wikipedia is written collaboratively by volunteers, and in the wiki-spirit, most articles can be changed by anyone with a web access. "The free encyclopedia that anyone can edit", as the frontpage tagline of english Wikipedia states[20].

As formentioned, the Wikipedia encyclopedia is provided in multiple languages, according to Wikipedia itself in 171 "active" language editions with more than 100

20 http://en.wikipedia.org/w/index.php?title=Main_Page&oldid=89893301 [referenced 26.11.2006]

articles each. There are, however, significant differences between the different editions of Wikipedias. At the moment of writing, only the largest 12 editions contain more than 100,000 articles. The largest edition is the English edition: it contains three times the articles of the second largest, German version. In this case study our focus will be on the English edition of the Wikipedia – the ancestor of them all. [21].

Wikipedia is funded and service is provided by Wikimedia Foundation Inc. It is a non-profit organisation, based on Florida, USA. The goal of the foundation is to develop and maintain open content, wiki-based projects and to provide all contents of projects to the public free of charge. Other services founded by Wikimedia include Wiktionary, a wiki thesaurus, Wikiquote, an encyclopedia of quations and Wikisource, a free library and collection of original texts. All these projects are mostly dependent on donations, therefore it might be well justified to say that the main driving-factors of Wikimedia projects are not commercial. [22].

Even though the base of all the different projects in various languages are quite rich, the underlying infrastructure is quite similar in all different locales and projects, as especially all editions of Wikipedia rely on MediaWiki. In the matter of fact, the history goes vice versa: MediaWiki has grown out from Wikipedia into a free, open source software wiki.

5.3.1. Intended use

Wikipedia is a free encyclopedia. Encyclopedia, defined by Wikipedia itself, means a comprehensive, written compilation of knowledge containing information from all branches of knowledge[23]. It has been desribed to be an compendium: a concise yet comprehensive body of knowledge that may summarise larger work[24].

It is not easy to say, what kind of information should be put on Wikipedia. The open nature of the encyclopedia makes it yet even harder as new readers are potentially new authors. The orientation, however, can be seen so that Wikipedia should summarise priori human knowledge, like traditional encyclopedia, not to make up new information. According to an official Wikipedia policy, only arguments, ideas or other information that are already published elsewhere, should be included to Wikipedia.

In this sense, Wikipedia's main use scenario would be to find facts about publicly known information. In a certain sense, it can be seen as an index to human knowledge: a comprehensive yet compact presentations of all available knowledge in different branches. Naturally our first use case is a user seeking for encyclopedic knowledge.

21 http://en.wikipedia.org/w/index.php?title=Wikipedia&oldid=90070521 [referenced 26.11.2006]
22 http://en.wikipedia.org/w/index.php?title=Wikimedia_Foundation&oldid=88900034 [referenced 26.11.2006]
23 http://en.wikipedia.org/w/index.php?title=Encyclopedia&oldid=90164560 [referenced 26.11.2006]
24 http://en.wikipedia.org/w/index.php?title=Compendium&oldid=89206904 [referenced 26.11.2006]

5.3.2. Scenario 2: Searching for encyclopedic information

As we recall from chapter 2 of our study and how the knowledge process, discovering encyclopedic information from Wikipedia largely corresponds to knowledge access and retrieval. A presentation of the captured knowledge is retrieved, on basis of user searches and queries.

We assume that as the user does searches or queries, he has some needs for using them. It might not be an accurate subject that we are interested in, as it is not always that we know exactly what we are looking for. It could also simply be an interest subject to intuition. In terms of formalisation, we could say that a user could either have a specific subject of interest or perhaps something more intuitive that would guide user to search for additional knowledge. Therefore we shall split the scenario into two stories: the first story, where user knows exactly what to look for and the latter one where the user is generally interested in the subject.

Story 2

John has been tasked in school to make a brief presentation about a sports hobby he has just recently began: Shotokan Karate. He already knows many facts about the subject, but he would like to make some further confirmations in two questions:

Question 1: who is the founder of Shotokan Karate?
Question 2: what are the different belt ranks in Shotokan Karate?

John starts his search from Wikipedia's main page. It is full of different kinds of content, but as he is familiar with search engines like Google, he types in Shotokan Karate to the search box and pushes Results.

The results page state that 292 results have been included. John starts looking at the results in the presented order and even all the results include terms Shotokan and Karate, it is only the 9th result that has been titled Shotokan Karate.

The link entitled Shotokan Karate leads to an article in Wikipedia named Shotokan. After the title, there is a short text that gives some clues about the content in general. The short summary is preceeded by a table of contents, including 11 headings with 2 subheadings. In addition to the textual content, there are two pictures on the side of the page, including a black and white photo of a man, and latter one with two karatekas in a competition.

The answer to the first question is found from the first clause, of the page: as John thought, Shotokan school of karate was developed by master Gichin Funakoshi. The summary part does not, however, mention anything about belt

ranks and therefore John continues to the article's table of contents. The 8th item in the list is a link entitle *Rank System*. Clicking the link brings the focus of his browser window to the part of the article covering the topic.

What John finds out is that the answer is not straightforward. The article lists many ranking systems. This makes him realise there can actually be many of them, in addition to what he uses. This makes him a bit confused: why are there so many of them and what is the one I am using?

The list organisation makes it a bit easier for John to to compare different ranking systems: every item of the list is organised in to a new line and denoted with a gray square. What he finds out is that none of the lists actually matches his own ranking system. This makes John a bit suspicous: *is this article actually comprehensive? Which one of the ranking system is actually correct?*

The founding information forces John to rethink his question: the different belt rankings are dependent on what school of Shotokan you are in. **An answer to question two** could be that there are different belt ranks, but in general, the ranking starts from white as novice through different colors and turns finally black. John makes himself a clever rule of thumb: the darker the color, the better the know-how.

5.3.3. Scenario 3: Authoring Wikipedia

Wikipedia states that it is an open dictionary that "anybody can edit"[25]. Because it is completely written by the community of it's users, authoring is an important part of Wikipedia's functionality. Let us introduce Jane, a beginning semantic web researcher:

Story 3

Jane is making a paper about folksonomies, but does not know very well what is already written about the subject for wider audience. She goes into Wikipedia looking for what is available there. Wikipedia proves to contain a short article about the matter, a stub. From Jane's point of view, the information in Wikipedia is incomplete and partially incorrect. She decides to contribute her thoughts by making an edit.

Actual edit view reveals underlying wiki-syntax: the article is not editable in it's original, WYSIWYG-format[26]. Jane happens to be somewhat familiar with this notation, and making a few plain text additions to the article is not a big deal for

25 http://en.wikipedia.org/wiki/Main_Page [referenced 3.1.2007]
26 WYSIWYG is a commonly used acronym for 'What You See Is What You Get'

her. Jane verifies her changes by using the provided preview function and then hits "save changes". The changes are now instantly visible in the article's own page. Jane admires the freshly fixed content for some time and finds herself pleased with it.

However, after a few hours when he comes back to check the content, it has already been changed. Most of the content is as Jane has left them, but some of her text has been modified by someone. Jane clicks history to figure out the actual changes. Comparison tool on history view shows changes as parallel tables.

It turns out that Jane and another user disagree in used terminology. Jane sees that this is certainly a matter that should be discussed further: the version system does not itself provide easy means for argumentative discussions.

Jane unfolds discussion page from the article and writes a short note that will be hopefully read. After some hours, another user has participated the discussion. Yet, it takes some days of collaboration on the discussion page, before Jane and other users have ended up with a satisfactory result.

5.4. Scenario-building for MuseumFinland

MuseumFinland is both a semantic portal "for publishing heterogenous museum collections on the Semantic Web" and a research pilot in Finnish Semantic Web research. The virtual exhibition in MuseumFinland[27] consists of more than 4,000 cultural artifacts and 260 historical sites in Finland. What is of our interest is that the kind of cultural collections that are used as content for the portal are semanticly rich. This was also confirmed in our heuristic study: MuseumFinland had relatively the most explicit and almost the widest coverage of Dublin Core metadata. [Hyv05].

5.4.1. Scenario 4: Searching for historical artefacts via facets

For end-users, MuseumFinland is almost all about exploring the repository of historical artefacts. We do not know to what extent the service is actually used, given its research-oriented background. Nevertheless, the search capabilities in the service are clearly deviant from basic keyword search and hence are worth of investigating by a scenario.

Story 4

Juuso is a historian who is working on a historical toy exhibition. Looking for potential exhibits, he browses around MuseumFinland.

27 Online at http://www.museosuomi.fi/

The frontpage of MuseumFinland[28] provides Juuso with access to numerous aspects of the underlying data. There is a total of nine different "main categories" with different subcategories that Juuso can choose to browse: artifact type, material, creator, location of creation, time of creation, user, location of usage, situation of usage and collection [Hyv05, p. 18]. Juuso finds out that artifact type includes type selection of "toys" and clicks the link on it.

The next view[29] already includes a set of 200 items classified into several subtypes of toys such as "toy animals", "balls", "rattles" and so on. It seems that the results yet do not match Juuso's preferences. As the main categories can be still used to refine the search, Juuso clicks open a complete set of selection from time of creation and selects the 19th century.

The resulting view consist of only five artefacts[30]. Juuso gathers from the results that the matched items are actually precisely matched as "toys from the 19th century". The results provided Juuso with the knowledge he was looking for.

5.5. Scenario-building for Del.icio.us

Del.icio.us is a web-based service for keeping a collection of favourites. The general idea is that rather than storing your Internet bookmarks to your local computer, you can use the service instead. The service further provides users with capabilities to organise and share their favourites of any kind. These favorites could be in practise for instance collections of research materials, podcasts, vacation trip activities or simply links [31]. The service also includes mechanisms for easy sharing of these favourites for individual users or groups.

5.5.1. Scenario 5: Social bookmarking

The social bookmarking aspect of del.icio.us is something that makes it unique. Our scenario for this service reflects this view:

Story 5

Jane is keeping record of web sites that contain interesting technologies related to her research issues. For keeping track of different online resources, she has an own del.icio.us account. She already has hundreds of links stored in her account.

28 http://www.museosuomi.fi/
29 http://www.museosuomi.fi/?l=fi&m=0&n=%2500%2512&g=c%2500%2512
30 http://www.museosuomi.fi/?l=fi&m=0&c=%2500%2512&n=%2504%2501%2502&g=c%2504%2501%2502
31 Http://del.icio.us/about/ [referenced 14.2.2007]

Again while browsing the web, she stumbles upon a page containing documentation for a web standard. The content looks promising and Jane decides to add the page as a del.icio.us favourite.

Jane begins posting the resource by clicking del.icio.us bookmarklet that is provided to help with adding new content to the service. Along with the page URL[32] address, she has three fields – description, notes and tags – that she may utilise in organising the link to her account.

URL and description fields are precompleted by the service which Jane complements. She also adds some notes to the third provided field in the fashion that helps her to write out her current thoughts.

To associate the new resource to what Jane already has added to her account, she uses tags. The service readily recommends tags used by other users and shows the tags Jane has already used herself. Additionally she can of course also make up completely new ones. Tags provide Jane also a possibility to channel new resources to her friends using del.icio.us.

Adding new resources is complemented by clicking save. The new resource appears to be integrated to all the various visualisations of her account content, helping Jane to complete various tasks with her technology related issues.

32 An abbrevation of 'Uniform Resource Locator'

6. SCENARIO ANALYSIS

In this chapter, we perform qualitative analysis of the scenarios presented in chapter 5 and study how they reflect knowledge management and visualisation practises in the web-based services.

6.1. IRC-galleria

Our heuristic analysis on IRC-galleria showed that even though the service is relatively rich of metadata, it is relatively implicit by it's nature. We could perhaps better depict the information in the service as indicative rather than categoric. Especially understanding the discusions that emerge from the streams of comments, is quite deeply embodied to the talkers. Thus, interpreting meanings of messages out of that context would not very likely give any further insight.

In IRC-galleria, sign in was required in order to build the architecture that gives as much support to social translucence as it does. This can clearly be seen as a challenge if an architecture of interaction is built in more distributed fashion or reaching over a set of services. Infrastructure for such hypertextual network of messages that are distributed over services, can be found from the network of blog messages (See for instance [Maj06]).

From the author's point of view, IRC-galleria is an example of a large online community, where interaction is more imporant than formal validity of the messages. It might actually be the low level of formal requirements that helps to feed the community with new content, keeping it alive.

6.1.1. Knowledge management in scenario 1

In the presented scenario, Lilli performed a daily routine in which she checks the gallery service for possible changes. When signed in, the front page provided her with information about new comments in her gallery pictures. The more pictures she has in

the gallery, the more crucial this information is to her. Otherwise she could not easily keep on record with the recent changes.

A new comment triggers Lilli to visit another users profile but also to make a responding comment to her friend's gallery. She could have also left a comment in her own picture, keeping the dialogue in context, but as the reply was more personal than about the photo context, the decision was made that in mind. She also knows that putting the comment there will trigger the new comments alert, thus making it more certain that her friend sees it.

What is special in such communication is that the actual conversation will be distributed to the attending users' profiles. Lilli and her fellow conversationist might not have any problems understanding the dialogue, however users viewing individual profiles could have problems understanding the complete conversation. Perhaps intentional, such organisation could be a key factor in keeping the comments themed: the only shared feature of the comments is the gallery photo they are associated to.

This kind of associative commenting or annotating is also found from many other socially rich services. Similar patterns of communication emerge also in Last.fm's shoutboxes, Flickr's photo commenting or YouTube's comments and responses, just to name but a few.

Commenting enables hypertextual navigation in the virtually endless network of comments. This is clearly a completely new model for communication in comparison to oral or written forms of speech, which also most online tools, chats and forums clearly try to model.

IRC-galleria seems to support fairly well the use of comments in this fashion of distributed conversations, when the discussion is being held one on one. The service, however, does not provide users with any additional tools for compiling the discussion to a chronological thread or another view containing more inclusive aggregation of comments.

Another feature of commenting in IRC-galleria is that the distribution of comments is limited to the scope of the service. Friends of Lilli who are not users of the service, are unable to participate or even see the perpetual discussion. When such medium becomes a significant mediator of online social interaction, it is unsurprising that more new users will adopt it. This could perhaps explain some of the popularity of IRC-galleria and similar social networking sites.

Evidently when all social interaction is capsulated to a single service such as IRC-galleria, the service lock-in should not give rise to any major problems. However, combining ongoing discussions from different social networking oriented services without open architecture might prove troublesome. These boundaries set by the

services further converge the interaction in service-centered fashion.

6.1.2. Social networks and social translucence

Clearly, in IRC-galleria, the different social processes (visiting profiles, commenting pictures, reading and writing blog and many other activites perceivable by others) dictate much of the service. One of the most popular contexts for understanding social processes is the concept of social networks. According to Finin et al., social networks are explicit representations of the relationships between individuals and groups in a community. The links between inviduals may define arbitrary types of relationships such as professional or family relations to name just a few. [Fin05].

Web-based services that have been built upon structures of social networks began appearing in 2002 and since have become among most popular web-based applications (see for example [Fin05]). Examples of most popular, such services include MySpace, Friendster, LinkedIn, Orkut and many more. Our research sample also includes services that are not primarily social tools, but benefit even greatly from the availability of networked social structures. Different kinds of social navigation is utilised in services like 43Things, Del.icio.us, Flickr, Last.fm, LinkedIn, Technorati and many more[33].

Let us next consider, how the aspects of social translucence, as presented in chapter 2, have been taken into account. In IRC-galleria, it is not possible to leave anonymous comments. All the comments left are signed with link to the writing user's own profile. This helps social tranclucence giving rise to visibility – exposing the commenting users to the receiver of the feedback. Further, all users have sign in status in their profile pages. If they are signed in, it is clearly visible to other users. This raises awareness when making a comment: "I know that you are on the other side" [Eri00].

Guestlist, if enabled, keeps exact track of registered users visiting user's profile. It is important that the use of guestlist relies on mutual agreement: when users choose to use the list, they also choose to share their own visiting status. In the framework of social translucence this is as much of encouraging visibility as well as accountability. As well as Lilli sees the visits to her profile, she also knows that a visit to another user's page will leave a mark.

6.1.3. Visualising IRC-galleria and socially aware systems

What we see as visualisations in IRC-galleria, largely depends on how loosely we interpret different features as visualisations. The ones picked here represent what author sees as fundamental to support the centric use cases of the service. Bundled into three categories, the visualisations can be divided into four main categories:

33 For a more comprehensive list, see for instance Wikipedia's article on social networking websites at http://en.wikipedia.org/wiki/List_of_social_networking_websites

- **Comment and diary entry streams**. Just as diary entries, the comments under user pictures are ordered by their posting dates. The order is generally fixed to the date of appearance, forming the backbone for interaction.
- **Activity indicators**. For a signed in user the frontpage lists briefly some of the new content that might be relevant for the user. Also a small orange number in parentheses on the top of the interface hints user of new comments in his or her pictures. Also user status indicators appear before user names consistently all around the service. Online indicator is a small blue dot whereas users celebrating their birthday have small crown-like icons nearby their nicknames.
- **Thumbnail overviews**. A table-layout of user pictures give overview to user galleries.
- **User peek windows**. Leaving mouse cursor over user's nickname will make a peek window to users information to appear, containing a thumbnail picture and some textual facts about user, including time since last appearance.

We could predict that also perhaps up to four different visualisations could be potentially beneficial: emerging discussions, social trends, community activity and related communities and users visualisations.

*Emerging discussions visualisations***.** As discussions are already taking place in the picture comments, a visualisation of the hypertextual discourse that arises might be useful. The visualisations could answer to questions like "what kind of dialogue have I had with another user?" and "What have I said in general?". The granularity of the textual content does not suffice for making a specific map of arguments, but nor is the IRC-galleria potentially the main forum of discussions requiring such capabilities.

Social trends visualisation. The service is full of different community user forums and yet it does not provide any further insight to past or current communities and trends that emerge from them. Currently the service leaves user with little or no information on history or trends. Some kinds of social trend visualisations could be utilised to discover past and current trends for example in community memberships, topics of discussion and so on.

Community activity visualisations. Clearly, IRC-galleria was not originally meant to be a social forum. However, considering it's current status, we could say that does not sufficiently give insight to group related activity. Such visualisations closely related to personal activity indicators could be used to better keep up with what's going on in a community.

Related communities and users visualisations. Along with the user profiles, the different communities in IRC-galleria form a network of actors, loosely related by their own

interests, community memberships and activity. This information could be refined giving end users access to information about their own position in the social space of IRC-galleria. On its simplest, community pages could include a list of closely related communities.

6.2. Wikipedia

Wikipedia is a typical wiki tool that has knowledge management focus on articles as documents. Very similar services as Wikipedia from our study are Psychology Wiki, Wikihow and Ontoworld Wiki, of which all fundementally rely on the MediaWiki application that originates from Wikipedia. From the document management aspect also Discogs is very near to Wikipedia, however in comparison with some missing version control capabilities.

The tools in these services serve well the use cases where users focus on single articles. Getting insight to both larger scale and smaller scale artefacts, such as sentences of texts or groups of articles, is much more troublesome. Also situations that require cross-article work are not easy to visualise nor manage explicitly.

We could say that Wikipedia supports well management and visualisation of knowledge on the granularity level of articles. Moving the focus to finer parts of the content, such as actual sentences or tables, some problems are introduced. For example in our first use scenario, Wikipedia did not provide viable means for making comparisons or synthesis of different belt ranks in Shotokan, let alone a comparison of different martial arts.

Some of these problems in Wikipedia and wiki's in general have been acknowledged. For example, the Semantic extension to MediaWiki provides means for making structuring and linking also to parts of articles. Relying on the Resource Description Framework, this extension is seen as enabling technology for Wikipedia to transform it to "a resource of semantic statements, hitherto unknown regarding size, scope, openness and internationalisation". [Völ06].

The idea of improving wiki semantics by combining the power of the Semantic Web and Wiki tools can be seen as one of the emerging phenomena during 2006. It seems that adding semantics to wiki content could potentially improve directly search and navigation capabilities, but also enable non-technical users to contribute to the Semantic Web. [34].

6.2.1. Knowledge management in scenario 2

In this story, our imaginary friend, John started his explorations to Wikipedia by using the provided search tool. Even though the main page did not provide him with much

34 The first workshop on Semantic Wikis was held in June, 2006. For more information, see http://www.semwiki.org/ [visited 3.1.2007]

insight to the subject of his study, it provided him with different possible views to the content. But why did the main page have the content it did?

Some of the content is clearly there because they are related to current events. Sections "In the news" and "On this day" are clearly connecting today's important aspects to the actual content. Featured articles and featured pictures are also present, however why they are seen so important, remains behind understanding of why such practises exist. Also the main page is full of smaller links put on the sides of the web page when viewed with a visual browser. They had numerous links to other editions of Wikipedia and different projects of Wikimedia.

It it not perhaps a single piece of information that has an impact on John, but the page as a whole that contributes to John's image of Wikipedia. There are at least three things, however, that can be perceived from the frontpage:

- **Content language.** It is quite self-explainatory that John understands English. However, it was just the main page that proved that to be true for John: as main page is in English, John perhaps quite implicitly assumes that all the content would be in the same language.
- **Subject coverage**. The beginning of the page states that there are over 1,500,00 articles available at the moment. Naturally it is impossible to put them all to the main page: even listing bare titles would give too much information. However, making some summary would be helpful in determining what subjects have been covered. By using broader terms of the articles, we may condense the content to fewer terms that describe it in a more abstract level. In Wikipedia's classification scheme, there are seven topmost concepts that capsulate narrower topics. For John this information is relevant if he is suspicious whether the knowledge he is looking could exist in the service.
- **Topical coverage.** Several views to the content have been picked in the emphasis of current moment. The views link content of Wikipedia to current events with dozens of hyperlinks. This might be relevant to John if he ponders whether the content is up to date.

Continuing the study of affordances, when John uses the search engine provided by Wikipedia, it is mentioned that John is familiar with other search engines like Google. The study of how the user interface of the search engine should be done, is more of a user interface issue. However, the results give some insight to the content.

The results are organised into a list with *title* and *relevance* specified as percentage values. The order of the list is that the most relevant item is placed topmost, whereas subjects of lower importance are listed later or are available on further pages. Again, we may see that the results screen makes an effort for making it clear to John, what is the

coverage of the page.

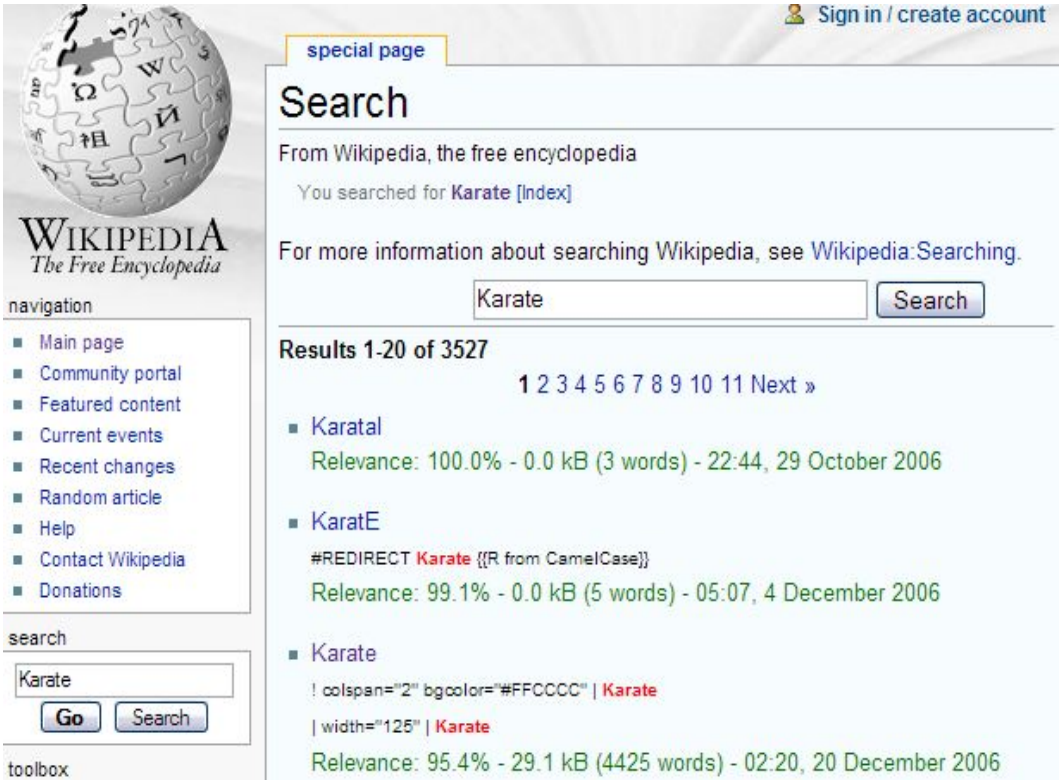


Figure 6.1: *Search results in Wikipedia*

The results are presented in a structure that reminds an *ordered list*. Therefore the first good match should not leave any larger doubts about if there are further entries that are better. The primitive structure that has order may be seen as a crucial element for efficient search in this case. As the order assumes descending relevance, first good match should not leave John any doubts about if there are any better results later on the list.

Moving on to the Karate article page, John finds it troublesome that the found information about Karate belt ranking system does not match his own notions. Understanding the concept of parallel ranking systems is totally dependant on the quality of the article written about Karate. The article encoded the values into multiple, separate lists that somewhat helped John to internalize the differences, however for making such comparison, *a comparison table* might have proven to be more intuitive.

Still, we may see that the real problem is that what John found out, did not match his understanding of what he had discovered. John assumed that the belt ranking system in karate is universally unambigous. In reality, however, there is deviation at least in different regions and different trades of Karate.

In some sense, we may see that this problem rises from the lack of social and contextual information in Wikipedia. As English is an internationally spoken language, the users

and editors of English Wikipedia could be globally very distributed. As there is no regional coverage in the articles that synthesize information, it is truly difficult to keep track on actual regional or user coverage of the articles. If John at least knew where the editors of the article are from, he perhaps could have figured out that the article did not match very well his specific, even very regionally unique needs for information.

The pattern of action in scenario one resembles that of searching. Wikipedia seems to be well suited for finding and managing specific, factual information that rarely changes. Regional or otherwise subjective variations may introduce problems: it might be difficult to make a comprehensive synthesis about a very ambiguous matter in a format of article.

6.2.2. Knowledge management in scenario 3

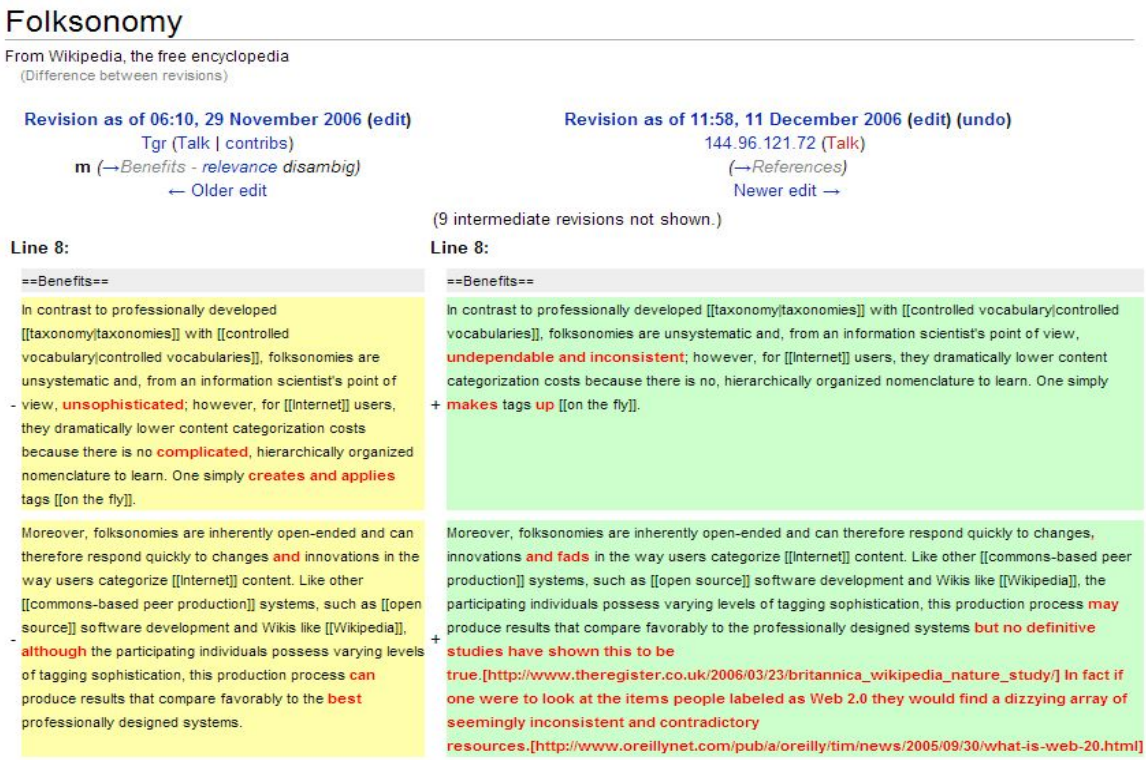
Jane began authoring Wikipedia by editing the existing article from the edit view. She supposed that the version of the article she found is up to date, otherwise she would have continued browsing for more recent information. Whatever Jane's motive for editing Wikipedia were, the edit operation was fairly quick and straightforward. Freshly updated article provided her with instant gratification: the results were instantly available to all visitors on the site.

The version comparison tool provided Jane with quick access to preceeding changes. It did not, however, suffice to complete the collaboration that turned out to need some argumentive discussion, which was needed in order to resolve all conflicts beyond technical data differences.

Perhaps a tricky part in the process was the comparison between Jane's and preceeding edits. Wikipedia actually keeps all articles under version control, meaning all the cascading revisions of an article are kept in history. Provided comparison tool helps users to track made changes between arbitrary two revisions.

As demonstrated in figure 6.2, the compared versions are shown side by side laid out with a table. The article is split into pieces per paragraph and changes in text are formatted with bolded, red font. The view is fairly simple to use, if the user can view the content in the assumed, broad tabular format. This is something that John would have found useful earlier when comparing different belt ranks.

Derived from our visualisation classifications, the comparison view is a deduced visualisation. From version history that is saved by Wikipedia, it is possible to accurately reason how cumulate changes have affected the content. The comparison view is more about facts, than something new emerging by use of a certain method. The interpretation of the view is very much dependent on the visual layout of the page. Visual layout enables faster insight to the changes that have been made to the article.


Folksonomy

From Wikipedia, the free encyclopedia

(Difference between revisions)

Revision as of 06:10, 29 November 2006 (edit) Tgr (Talk | contribs) m (→Benefits - relevance disambig) ← Older edit	Revision as of 11:58, 11 December 2006 (edit) (undo) 144.96.121.72 (Talk) (→References) Newer edit →
(9 intermediate revisions not shown.)	
Line 8:	Line 8:
==Benefits==	==Benefits==
In contrast to professionally developed [[taxonomy|taxonomies]] with [[controlled vocabulary|controlled vocabularies]], folksonomies are unsystematic and, from an information scientist's point of view, unsophisticated; however, for [[Internet]] users, they dramatically lower content categorization costs because there is no complicated, hierarchically organized nomenclature to learn. One simply creates and applies tags [[on the fly]].	In contrast to professionally developed [[taxonomy|taxonomies]] with [[controlled vocabulary|controlled vocabularies]], folksonomies are unsystematic and, from an information scientist's point of view, undependable and inconsistent; however, for [[Internet]] users, they dramatically lower content categorization costs because there is no, hierarchically organized nomenclature to learn. One simply makes tags up [[on the fly]].
Moreover, folksonomies are inherently open-ended and can therefore respond quickly to changes and innovations in the way users categorize [[Internet]] content. Like other [[commons-based peer production]] systems, such as [[open source]] software development and Wikis like [[Wikipedia]], although the participating individuals possess varying levels of tagging sophistication, this production process can produce results that compare favorably to the best professionally designed systems.	Moreover, folksonomies are inherently open-ended and can therefore respond quickly to changes, innovations and fads in the way users categorize [[Internet]] content. Like other [[commons-based peer production]] systems, such as [[open source]] software development and Wikis like [[Wikipedia]], the participating individuals possess varying levels of tagging sophistication, this production process may produce results that compare favorably to the professionally designed systems but no definitive studies have shown this to be true.[http://www.theregister.co.uk/2006/03/23/britannica_wikipedia_nature_study/] In fact if one were to look at the items people labeled as Web 2.0 they would find a dizzying array of seemingly inconsistent and contradictory resources.[http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html]

Figure 6.2: *Wikipedia's version comparison tool*

The sequential order of different versions is apparent in all the views of the history tool. The links provided on the top of the page make it easy to browse and compare newer and older version of the same resource. In a way, these operations provide access to the lifestream of an article.

In Jane's case the comparison tool proves to be useful. However, it has some limitations for more demanding tasks. The basic element that is versioned, is an article. There is no tools for tracking, for example a specific paragraph or a specific sentence. This is quite fine in Wikipedia's article, as the service itself is quite document oriented. However, if it would be required to track specific sentences, both the editing tool and versioning should have fine *granularity*. That is: finer basic elements for versioning.

Moving on from version comparison, the actual discussion that took place in the article's discussion page, was not directly linked to any part of the text. Still, the context was enough accurate for the collaborators to continue. Collaboration in Wikipedia is perhaps easy because there are no strict rules that govern modification of the wiki content. However, what Jane would have discovered, is that there are many best-practises that guide the content production process[35].

As we may see, the knowledge management in Wikipedia is largely focused on articles as documents. This means that there are not very good tools provided for tracking

35 For detailed information on such guidelines, see Wikipedia's Policies and guidelines, online at http://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines

changes that span over multiple articles. For instance if an article is renamed or document text is distributed by moving it to parts of several articles, it is very hard to keep in track of the changes with the tool provided. This brings another restriction of the model visible for us: as the finest item of knowledge that has identity and support document management, are articles, applying use cases of knowledge management to finer parts of the content is more difficult.

6.2.3. Collaboration, reputation and "the wiki way"

Jane probably did not began the collaboration a process in mind. It just emerged along the way. Initially, her motive was just to contribute a thing she knew as a fact. What ultimately motivates Jane to contribute and collaborate and kept her involved, is a good question. Perhaps it is the instant grantification that she gets from correcting a fact or adding her own notions about things she feels she is good at.

We can see that the engagement to open-content collaboration happened step by step. Jane could stop contributing after any step, but she still chose to continue until resolving the conflict. It is worth noticing that the effort Jane had to made, was quite minimal. No registrations, confirmations or publishing were required in order to collaborate. In general terms, Jane had only to do the creative effort in the collaboration, everything else such as document versioning was automatically taken care of.

Making everything "just a click away" is actually presented as a key criterion for usefulness of wikis by Leuf and Cunningham in The Wiki Way. They see that simple navigational and collaborative model feed ease of writing and spark collaboration. In a way, the ideology is about not making collaborative processes formal. The trust is gradually built in collaborative processes rather than in mechanical verifications, hence emphasizing the importance of the community of the wiki users. [Leu01].

The aspect this leads us to, is also noticed by Bruckman and Forte. They present that the incitive system in Wikipedia resembles that of the scientific community: credit helps Jane to promote her social status. Jane could also utilise the popularity of Wikipedia to contribute legible facts about her scientific studies in encyclopedic format. [For05].

6.2.4. Visualising Wikipedia and document driven systems

We have already taken a look into the version history tool that clearly provides end users with particularly visual and quite efficient version comparison view giving insight to the document change history. That included, we ended up with four visualisations:

- **Overview to content with text snippets.** The frontpage of Wikipedia provides multiple windows presenting some content from various aspects, such as "today's featured article", "in the news", "did you know..." and "on this day". These text snippets effectively function as a looking glass into the content

available in Wikipedia. Once in a while the frontpage is updated by the community, allowing even frequent visitors to discover new things about the content.

- **Search results and article version history streams**. Both search results and article versions are shown as visually rather primitive lists of items. The only way a user can control the listing is selecting a page.
- **Version comparison view.** The version comparison tool provides insight to actual changes in the article content with a side-by-side view of different article versions. The user may browse through various article versions to see how differences incrementally change the content.

The already available visualisations suffice for the very basic user needs. There are, however, several content-based visualisations that could potentially help with the current model of content authoring.

Richer category browsing tools and visualisations. Article categories are fairly strongly used to build a more abstract, conceptual backbone for the content in Wikipedia. Yet, the categories can only be browsed with rather simple, generated category indices. Other relationships between categories, wether emergent or explicitly defined, can not be easily perceived nor utilised. Currently getting even a list of articles that belong to two separate categories is at least difficult, let alone a process easily automated.

Related articles and topics visualisations. Finding out related articles and related topics could be made more flexible for instance with associatively or statistically generated relations between articles. Such weaker associations would perhaps help both readers and co-authors even better to discover topics emerging around article documents.

Spatial, editorial and conceptual coverage visualisation. As we pointed out in our scenarios, Wikipedia does not very well make it especially clear who are the actual contributors and in what is the general scope of an article. Understanding the spatial, editorial and conceptual coverage of an article is currently not easy to track. Another issue is that current categories do not very well point out the amount of information that has been gathered around various topics.

Visualisation of inter-article information transactions and knowledge evolution. Current article versioning facilitates only understanding the sequential versions of an artifcle. Further, the version history and comparison tool does not provide easy means for getting insight to the complete, procedural history of the articles. Visualisations for better insight to articles that have been changed, moved, deleted or integrated as parts of other articles, do not exist at all. Different visualisations that would potentially help to understand the procedural nature of information in Wikipedia might be helpful in facilitating understanding of collaboration in various scales.

As noted, the suggested new content visualisations focus more on enabling better insight to the content than on making Wikipedia more suitable as a place of socialisation. For latter we have already gathered experiences from IRC-galleria, however, it is worth questioning whether encouraging members of Wikipedia community to social interaction would be even necessary let alone desired.

6.3. MuseumFinland

MuseumFinland is the richest service of all compared in the means of available formal, ontology-committed information. As such, it shows us a glimpse of what can be done with ontology-commited system. The end-user interface of MuseumFinland does not provide any tools for content authoring and thus, it would be better described as an online front-end or a web-based service to museum resources rather than as an online community. In the scope of our research sample only the Ontology Wiki actually provided some community control to ontology building. It, however, as a wiki relied heavily on document management practises. It seems that at least in the scope of our research sample, collaboratively authored ontology-systems are yet to come.

6.3.1. Knowledge management in scenario 4

The presented use case where Juuso searches for historical information, resembles our scenario 2 where John was searching Wikipedia for information about karate. Perhaps the greatest difference between these two cases is that where Wikipedia provided John with only keyword search, Juuso had the possibility to search and browse keywords in context.

The search engine in MuseumFinland is what Hyvönen et al. call a semantic view-based search engine. A sample of search results of the engine is shown in figure 6.3. It is based on the paradigm of multi-facet search. In the figure 6.3, the applied facets are shown on the top-right. The novelty of the engine lies in the use of ontologies that are the "main categories" (on the left) or the conceptual models that form the backbone of the data repository. [Hyv05].

The faceted approach offered Juuso multiple advantages over traditional keyword based approach. First of all, the flexible ontology-based model enabled instant and extensive overview to all available information in the repository (Compare [Hyv05, p. 18]). The kind of overview also available in Wikipedia's frontpage should, however, not be taken for granted. In, Wikipedia frontpage is maintained manually by the users, whereas MuseumFinland generates it automatically without user intervention.

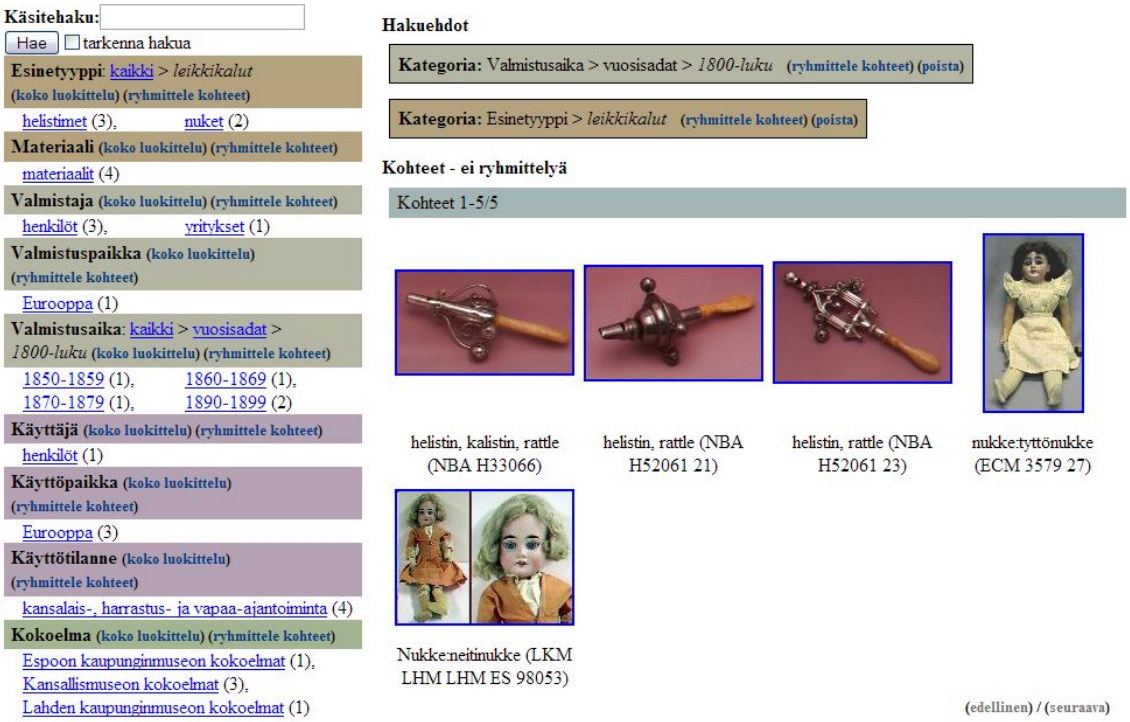


Figure 6.3: *Search results view from MuseumFinland's faceted search*

The ontology approach not only reduces the amount of manual work but also enables all kinds of arbitrary views to the repositorial information. In the presented scenario, the hierarchies can be used to guide Juuso to formulate the accurate query "in terms of appropriate concepts" [Hyv05, p. 18]. The hierarchies may be used in cumulative fashion, to gradually build the list of desired results by using facets as result limiters. A small but important aspect is also that the number of hits for each category can be calculated beforehand. This further helps user to guide the search towards a more suitable sample.

6.3.2. Visualising MuseumFinland and ontology-committed systems

From the author's point of view, MuseumFinland should be seen as a working pilot of what semantic technologies could possibly enable in web-based services. The flexible faceted search gives just one view to how the information could be visualised: the technology behind MuseumFinland is potentially even more flexible. In this spirit, we shall list visualisations that could be in theory implemented over MuseumFinland:

- **Faceted content exploration**. Despite the rather unexciting user interface for the faceted search, we can see that the search gives us more than just an ability for keyword matching. The facets give fairly much creative freedom for end-users to explore the content by limiting the results with the different ontological facets. More user-centered visualisations of this facet exploration would most

likely enable the end-users to make arbitrary roaming to the content and iteratively found what they are looking for – even when information needs are fairly blurry.

- **Dynamic content overview visualisations**. Ontology-enabled framework could easily generate dynamic overviews to content. The thumbnails of content could be automatically generated from basis of date last changed or any such criteria or combinations of them.

What is clearly left outside the scope of traditional domain ontology approach, is the ability to utilise weak associations or emergent trends in content. For example, making navigation of loosely related content objects in MuseumFinland may prove difficult with the current approach alone. It is still clearly required to utilise some sort of non-straightforward ontology based deductive queries, such as statistical means, to extract the data needed for such implications.

6.4. Del.icio.us

Del.icio.us provides a practical and lightweight framework for describing, organising and browsing personal collection of resources. Generally anything that has an identity can be added to del.icio.us, however the service works the best for resource available directly online.

Similar to many other services in close proximity at our metadata study, such as Flickr, Technorati, Last.fm and YouTube, also del.icio.us uses tags for organising content. This style of organising content, often referred to as folksonomy (in constract to taxonomy) does not set any artificial restrictions for harnessing users own personal preferences for information organisation. The folksonomy rising from the tags further feeds this process with communal support in the use of tags (Compare to [Ore05]).

6.4.1. Knowledge management in scenario 5

Del.icio.us gives surprisingly much support for Jane as she posts new resource to her account. The service does not only pick some relevant metadata available from the resource page, but also uses content available via other users that have added the same resource. Besides, the service offers several ”helpers” for picking right tags to organise the resource with: these helpers include a tag recommendator, your tags, your network and popular tags (See figure 6.4 below).

Even though user may freely enter whatever he or she wishes to the tags field, there is no denying whether the recommended tags would affect the outcome. The tags function as affordances, helping users to appereciate already existing tags in the service. This social guidance further maintains the coherence of the tags: without any guidance there would be likely less shared use of the tags.

Figure 6.4: *A view for posting a new resource in del.icio.us, displaying two tag helpers*

As we already mentioned, a notable feature in del.icio.us is that anything that has an identity, can be posted to the service. Unlike many traditional web based services, del.icio.us not only enables description of in-service resources, but also makes it possible to describe sets of resource over the service domain. If similar distributed identity architecture would be used in services like IRC-galleria, it would enable ability for social networking regardless of service provider. In the level of information integration, such architectures could be implemented in the top of FOAF[36] (Friend of a Friend) ontology.

6.4.2. Folksonomies

Unlike in many web-based services, in del.icio.us the vocabulary for metadata is almost entirely controlled by the community. If we call controlled vocabularies the top-down approach of metadata usage, the bottom-up approach would be that users may use their own vocabulary to describe content. Vander Wal sees that whereas controlled vocabularies contribute directly to building of shared metadata, tagging content with own vocabulary contributes to personal information retrieval. A key point is that "the act of tagging is done by the person consuming the information". Thus, *folksonomy* is the result of the free tagging in social environment, a vocabulary emerging from the community, not from committing to a controlled, third party created vocabulary. [Van05] (See also [Gru05a]).

We could say that folksonomies aim at understanding the user's own conception of reality, rather than force or commit user on any arbitrary vocabulary. In this sense, folksonomy-based systems not only contribute to metadata, but also to the building of

36 http://www.foaf-project.org/

user models, used for example to adapt content and navigation systems [Bru01].

6.4.3. Visualising del.icio.us and folksonomic systems

Most of the visualisations in del.icio.us aim at giving users better understanding to the resource stores of the system. The influence of tagging and the emerging folksonomy as the main organising system is clearly visible also in the visualisations. We identified five differing visualisations from the service:

- **Tags as search facets**. Users may browse metadata of resources stored to their account by browsing the different tags. Also combinations of tags can be chosen, resulting in boolean intersection of the described resources.
- **Related and recommended tags**. Using some unknown methods of data mining, the service automatically discovers related tags that are associated with the selected tag view, but also finds recommended tags when adding new resources.
- **Tag bundles**. By default, all tag associations are more of statistical than explicit. Users, however, may choose to define *bundles*; explicitly defined collections of tags.
- **Tag clouds (figure 6.5)**. Two illustrate popularity of tags and see what subjects a user personally has covered with his or her tags, del.icio.us offers a tag cloud. The tag cloud gives also an impression of what the social trends in the service currently are.
- **Resource thumbnails**. For most popular web pages, the service generates a thumbnail image of the page content.

Clearly, the tags along with the different visualisations offer versatile system for organising and perceiving the content of the service. Especially as tagging encourages associative use of what user knows, it is a tempting method for informally yet efficiently organising content.

Figure 6.5: *Tag cloud from del.icio.us*

Yet, the service does not support associating tags to concepts outside the service, such as tags in other services or concepts in formally defined ontologies. The loose organisional value of tags stays as such in del.icio.us as there is no support for stating explicitly relations between the organisational system used in del.icio.us and possibly elsewhere. The more tagging as organisational system gains popularity in the web, the more important it potentially is to be able to build cross-service aggregations of the descriptive metadata.

Further, the support for discovering social trends is fairly limited in del.icio.us. The tag cloud (in figure 6.5) gives a fairly good overview, but does not scale well for understanding recent trends. A prototype of similar trends discovery visualisation is already implemented with another resource organiser service: Digg[37].

37 http://labs.digg.com/stack

7. SUMMARY AND CONCLUSIONS

When interactive systems design migrate towards community-managed structures, becomes knowledge management even more demanding. As data is enriched with information relevant to community-based dynamics, the task of presenting users with information relevant to their needs, becomes even more challenging.

In this thesis, we studied knowledge and knowledge management in 19 of current web-based services. In a two-phased study, we began by overally studying how knowledge in these services could be described according to combination of content and metadata analysis.

After initial knowledge analysis, we classified the services into four different types according. From all four groups of services, we selected one service for more scenario study. For each of the selected services, we created usage scenarios according to which we analysed the knowledge-level interaction processes in the services.

According to scenario study, we found out that the typical tasks in these selected services represent different areas of knowledge with varying requirements for both practises of knowledge management and visualisation. According to our findings, we will name the four areas of knowledge as *documents*, *ontologies*, *folksonomies* and *people and discussions*. In chapter 6 we also highlighted different visualisation needs, applicable to each of the presented areas. In figure 7.1, we have integrated the four knowledge areas and their typical tools into a four-field table with corresponding visualisation needs.

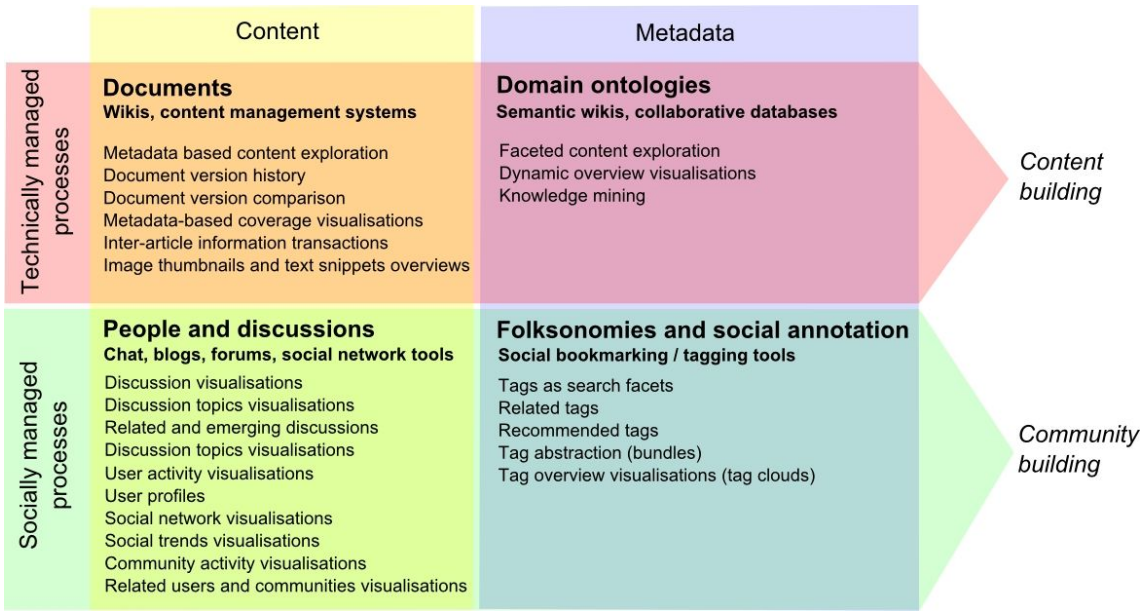


Figure 7.1: *Four essential areas of knowledge and related management and visualisation needs in social online services*

These four areas of knowledge can be described as follows:

- **Documents.** Even though the content and structure of documents can be managed socially, as in wikis, have the content management practises been committed to the document management metaphor. This commitment provides automation for metadata production and typical management tasks like document versioning.
- **Ontologies.** Content in ontology-committed systems may be created collaboratively, as in semantic wikis, but the used concepts draw on ontologies that formally specify their meaning. Methods of knowledge-mining can be applied to ontology-committed information. Collaborative databases[38] can be used for community-managed ontologies.
- **Folksonomies and social annotation**. Actual content is managed in domain-specific structures, yet users may annotate content socially. Folksonomic systems not only add to social navigation, but also contribute to the building of user models via personal tags.
- **People and discussions**. Knowledge is managed in discussions that take place between members of a community. As discussion is an important form of participation and community building, are discursive tools generally good for community building.

38 E.g. databases that have shared access not only to database content but also to its schema

We see that these four knowledge management tasks in web-based services have two diverging orientations: while documents and domain ontologies are more focused on content building practises, serve discussions and social tagging more the building of an online community. Tools of content building may also be used for tasks of community building, but they do not serve this task very efficiently. And vice versa: community-building tools such as chat and blogs, may be used for content building, but they lack the technical automation that otherwise benefits the processes of content management.

When documents or domain ontologies are in focus, we see that Dublin Core vocabulary is fairly extensively used. However, when people, discussions or any other social information is being described, Dublin Core offers little help for describing such resources. This may imply that Dublin Core element set provides poor capabilities for the the description of social information resources. Thus, for the task of describing social resources, vocabularies like FOAF [FOAF07] and TagOntology [Gru05b] might prove beneficial.

Besides the four categorial knowledge classification, we have presented related, current and potential areas of visualisation application. All the four areas of knowledge pose distinct challenges: needs for the community building knowledge generally focus on leveraging social information, whereas challenges in content building knowledge lie on supporting understanding of the content and the underlying production process. The little attention given to visualisation challenges with domain ontologies is a deliberate choice; all knowledge domains possess their unique challenges that we think should not be appreciated individually, rather than are commonly shared among all communities.

Our study still leaves open, how different techniques of visualisation should be applied to practise, in order to support the internalising of knowledge of presented categorial types. Even though many of current visualisation research efforts concentrate on data-driven visualisations research, we see that there is clearly a risen need for more efficient community-based visualisations design. Especially combining approaches of data mining and information visualisation depict interesting possibilities for enhancing community-building practises in interactive systems.

BIBLIOGRAPHY

[Awa04] Awa, Elias M., Ghaziri, Hassan M. *Knowledge management (international ed.)*. 2004. Pearson Education. ISBN 0-13-122784-X.

[Bru01] Brusilovsky, P. *Adaptive Hypermedia*. 2001. User Modeling and User-adapted Interaction issue 11, pages 87-110. Kluwer Academic Publishers.

[Che76] Chen, P. P. *The Entity-Relationship Model – Toward a Unified View of Data*. 1976. ACM Transaction on Database Systems, Vol. 1, No. 1., March 1976, Pages 9-36.

[DCUG05] Hilmann, D. *Dublin Core Usage Guide*. 2005. [online] http://dublincore.org/documents/usageguide/ [accessed 7.12.2006].

[DCMI06] Dublin Core Metadata Initiative. *Homepage*. 2006. [online] http://dublincore.org/ [accessed 7.12.2006]

[DCES06] Dublin Core Metadata Initiative. *Dublin Core Metadata Element* Set, Version 1.1: Reference Description. 2006. [online] http://dublincore.org/documents/dces/ [accessed 7.12.2006].

[Doc01] Doctorow, C. *Metacrap: Putting the torch to seven straw-men of the meta-utopia*. 2001. [online] http://www.well.com/~doctorow/metacrap.htm [accessed 7.12.2006].

[Dyn07] Dynamoid Oy. *Mainostus IRC-galleriassa – IRC-galleria – Kävijäprofiili*. 2007. Dynamoid Oy corporate website. [online] http://www.dynamoid.com/mainostus/kavijaprofiili/ [accessed 9.1.2007].

[Eri00] Erickson, T., Kellogg, *W. Social Translucence: An Approach to Designing Systems that Support Social Processes*. 2000. ACM Transactions on Computer-Human Interaction, Vol. 7, No. 1, March 2000, Pages 59-83.

[For05] Forte, A., Bruckman, A. *Why do people write for Wikipedia? Incentives to contribute to open-content publishing*. 2005. GROUP 05 workshop: Sustaining community: The role and design of incentive mechanisms in online systems. Sanibel Island, Fl.

[Fra95] Franco, V. Piirto, R. Hu, H.-Y. Lewenstein, B.V. Underwood, R. Vidal, N. K. *Anatomy of a flame: conflict and community building on the Internet*. 1995. Technology and Society Magazine, IEEE Volume 14, Issue 2, 1995, p. 12 – 21.

[FOAF07] Brickley, D., Miller, L. *FOAF vocabulary specification 0.9*. 2007. [online] http://xmlns.com/foaf/spec/20070524.html. [referenced 16.8.2007]

[Fol94] J. Foley and B. Ribarsky. *Next-generation Data Visualization Tools in Scientific Visualization*. 1994. Advances and Challenges, Ed: L. Rosenblum, R.A. Earnshaw, J. Encarnacao, H. Hagen, A. Kaufman, S. Klimenko, G. Nielson, F. Post, D. Thalmann. Academic Press.

[Gru93] Gruber, T. *A Translation Approach to Portable Ontology Specifications*. 1992 (revised 1993). Knowledge System Laboratory Technical Report KSL 92-71. Stanford University.

[Gru05a] Gruber, T. *Ontology of Folksonomy: A Mash-up of Apples and Oranges*. 2005. [online] http://tomgruber.org/writing/ontology-of-folksonomy.htm [accessed 5.9.2006].

[Gru05b] Gruber, T. *TagOntology – a way to agree on the semantics of tagging data*. 2005. Presentation to Tag Camp, www.tagcamp.org, October 29, 2005. [online] http://tomgruber.org/writing/tagontology.htm [referenced 16.8.2007].

[Han01] Han, J., Kamber, M. *Data Mining: Concepts and Techniques*. 2001. Morgan Kaufmann. ISBN 1558604898.

[Hei91] Hei, G. *Constructivist Learning Theory*. 1991. [online] http://www.exploratorium.edu/ifi/resources/constructivistlearning.html [accessed 12.10.2006].

[Hep07] Hepp, M. *Possible Ontologies – How Reality Constrains the Development of Relevant Ontologies*. 2007. IEEE Internet Computing, vol. 11, no. 1, pp. 90-96, Jan/Feb, 2007.

[Hyv05] Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S. *MuseumFinland – Finnish Museums on the Semantic Web*. 2005. Preprint submitted to Elsevier Science. [online] http://www.seco.tkk.fi/publications/2005/hyvonen-makela-et-al-museumfinland-finnish-2005.pdf [accessed 15.1.2007].

[JHS143] Julkisen hallinnon tietohallinnon neuvottelukunta (Juhta). *Asiakirjojen kuvailun ja hallinnan metatiedot*. JHS-suositus (Finnish government recommendation). 2004. [online] http://www.jhs-suositukset.fi/intermin/hankkeet/jhs/home.nsf/pages/C61E1E20B7AC1CF0C2256F85004BB16B [accessed 7.12.2006]

[Leu01] Leuf, B., Cunningham, W. *The Wiki Way – Quick Collaboration on the Web*. 2001. Addison-Wesley. ISBN 0-20-171499-X.

[Low99] Lowe, D., Hall, W. *Hypermedia & the Web – An Engineering Approach*. 1999. John Wiley & Sons Ltd, Chichester, England. ISBN 0-471-98312-8.

[Maj06] Majava, J. J. *Suomalaisten weblogien verkosto keskustelevana julkisuutena (Finnish weblog network as discoursive public*. 2006. Helsingin Yliopisto, Valtiotieteen tiedekunta. Sosiologian pro gradu -tutkielma.

[McC87] McCormick, B. H., DeFaranti, T. A., Brown, M. D. *Visualization in scientific computing – a synopsis*. 1987. IEEE Computer Graphics and Applications, vol. 7, issue 7, pp. 61-70.

[Nac05] Nachira, F. *What is a Digital Ecosystem? Essence of the Digital Ecosystems in 2 pages*. 2005. On-line paper. [online] http://www.digital-ecosystems.org/doc/de-summary-0205.pdf [accessed 13.2.2007]

[Nac06] Nachira, F., Louarn, M.Le. *What is an European Digital Ecosystem? Policy Priorities and Goals*. 2006. Bruxelles. [online] http://www.digital-ecosystems.org/doc/fp7-de-shortintro.pdf [accessed 17.8.2007]

[Nei82] Neisser, U. *Kognitio ja todellisuus [Cognition and Reality]*. Weilin+Göös. 1982.

[Non95] Nonaka, I., Takeuchi, H. *The knowledge-creating company*. Oxford University Press. 1995. ISBN 0-19-509269-4.

[Ore05] O'Reilly, T. *What Is Web 2.0 – Design Patterns and Business Models for the Next Generation of Software*. O'Reilly Media, Inc. 2005. [online] http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [accessed 16.11.2006].

[Opa06] *"Description of Work" for project Open Philosophies for Associative Autopoietic Digital Ecosystems (OPAALS)*. 2006. **Internal project documentation, not publicly available**.

[Pai06] Paivio, Allan. *Dual Coding Theory and Education*. University of Western Ontaria. Draft chapter for the conference on "Pathways to Literacy Achievement for High Poverty Children". 2006.

[Pec00] Pecorino, P. A. *Introduction to Philosophy – an Online Textbook*. 2000. Queensborough Community College, CUNY. Available: http://www2.sunysuffolk.edu/pecorip/SCCCWEB/ETEXTS/INTRO_TEXT/default.htm [accessed 26.10.2006].

[Pre94] Preece, J., Rogers, Y., Sharp, H., Beyon, D., Holland, S., Carey, T. *Human-computer Interaction*. 1994. Addision-Wesley, England. (775 p.). ISBN 0-201-62769-8.

[Pol66] Polanyi, M. *The Tacit Dimension.* 1966. Routledge & Kegan Paul, London.

[Pol67] Polanyi, M. *The Tacit Dimension*. 1967. Anchor books, New York.

[Rhe95] Rheingold, H. *The virtual community: finding connection in a computerized world*. 1995. Minerva. Electronic version of the book. [online] http://www.rheingold.com/vc/book/ [accessed 25.9.2006].

[Rob91] Robertson, P.K. *A Methodology for Choosing Data Representations*. 1991. IEEE Computer Graphics and Applications, Vol. 11, No. 3, May 1991, pp. 56-68.

[Rus03] Russell, S. J., Norvig, P. *Artificial Intelligence: A Modern Approach.* 2nd edition. 2003. Prentice Hall. ISBN 0-13-080302-2.

[Saa04] Saariluoma, P. *Käyttäjäpsykologia – Ihmisen ja koneen vuorovaikutuksen uusi ajattelutapa*. 1. painos. 2004. WSOY. ISBN 951-0-29184-6.

[Sen04] Seng, James. *Social Software*. James Seng's Wiki. 2004. [online] http://james.seng.cc/wiki/wiki.cgi?action=browse&id=Social_Software&revision=97 [accessed 27.4.2007].

[Sha06] Shawn, H. et al. *Introduction to Web Accessibility*. 2006. Web Accessibility Initiative (WAI). [online] http://www.w3.org/WAI/intro/accessibility.php [referenced 12.7.2007].

[Smi03] Smith, M. K. *Michael Polanyi and Tacit Knowledge*. 2003. The encyclopedia of informal education. [online] http://www.infed.org/thinkers/polanyi.htm [referenced 23.5.2007].

[Sta01] Staab, S., Studer, R. *Knowledge Processes and Ontologies*. 2001. IEEE Intelligent systems, January/February 2001, pp. 26-34.

[Tay03] Taylor, C. *An Introduction to Metadata*. University of Queensland Library. 2003. [online] http://www.library.uq.edu.au/iad/ctmeta4.html [accessed 7.12.2006].

[Tro02] Trott, B., Trott, M. *TrackBack Technical Specification*. 2002. [online] http://www.movabletype.org/docs/mttrackback.html

[accessed 2.5.2007]

[Van05] Vander Wal, T. *Folksonomy Definition and Wikipedia*. 2005. [online] http://www.vanderwal.net/random/entrysel.php?blog=1750 [accessed 2.5.2007].

[Völ06] Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H. *Semantic Wikipedia*. International 15th International World Wide Web Conference Proceedings. 2006. [online] http://www2006.org/programme/item.php?id=4039 [accessed 20.12.2006].

[War04] Ware, C. *Information visualization – perception for design*. 2nd ed. Morgan Kaufmann Publishers. 2004. ISBN 1-55860-819-2.

[Wik07a] Wikipedia. *Social software --- Wikipedia, The Free Encyclopedia.* 2007. [online] http://en.wikipedia.org/w/index.php?title=Social_software&oldid=125992173 [accessed 27.4.2007]

[Wik07b] Wikipedia. *Internet forum --- Wikipedia, The Free Encyclopedia.* 2007. [online] http://en.wikipedia.org/w/index.php?title=Internet_forum&oldid=127489076 [accessed 2.5.2007]

[Wik07c] Wikipedia. *Blog --- Wikipedia, The Free Encyclopedia*. 2007. [online] http://en.wikipedia.org/w/index.php?title=Blog&oldid=127607182 [accessed 2.5.2007]

[Wor07a] WordNet® 3.0. *visualisation*. 2007. Dictionary.com, Princeton University. [online] http://dictionary.reference.com/browse/visualisation [accessed 4.7.2007]

[Zaj07] Zajicek, M. *Web 2.0: Hype or Happiness?*. 2007. ACM International Conference Proceedings Services; Vol. 225, pages 35-39.

APPENDICES

A ANALYSIS SHEET FOR WIKIPEDIA

Service	Wikipedia (English)
Homepage	-
Year founded	-
Primary concept	Encyclopedic article
Primary concept occurences	> 100000
Supporting concepts	article, category
Knowledge Domain	-
Intented use	Open, collaboratively written encyclopedia for discovering and authoring encyclopedic information
Source(s)	http://en.wikipedia.org/wiki/Wikipedia

Metadata analysis based on Dublin Core

Category	Element label	Implicit	Explicit	Sample values / rationale
1. Locating	*1.1. Identifier*	**1**	**1**	http://en.wikipedia.org/wiki/Duck http://en.wikipedia.org/wiki/Andrea_Bacci http://en.wikipedia.org/wiki/Kitchen_Stories
2. Content description	*2.1. Title*	**1**	**1**	”Duck”, ”Andrea Bacci”, ”Kitchen Stories”
	2.2. Subject(s)	**1**	**1**	”Living people”, ”Italian racecar drivers”, ”Norwegian films” / Wikipedia categories
	2.3. Description	**½**	**½**	Occassional table of contents
	2.4. Language(s)	**1**	**1**	/ XML-lang attribute, context
	2.5. Coverage	**½**	**-**	/ No systematic, spacial, temporal or jurisdictive coverage explication practises
	2.6. Source(s)	**1**	**½**	/ Best practise for referencing, not enforced
	2.7. Type	**1**	**1**	/ Encyclopedic categorisation, ”all articles are articles”
3. Access conditions	*3.1. Format*	**-**	**1**	/ All article content is HTML, information about format perhaps not well visible to end-user
	3.2. Rights	**1**	**1**	/ Link to copyright, textual explanation
4. Context	*4.1. Creator*	**½**	**-**	/ Version history reveals first creator of an article, however, it might not tell much about the content itself
	4.2. Contributor	**1**	**1**	/ Version history reveals all contributing users
	4.3. Publisher	**1**	**1**	/ All articles have notion ”From Wikipedia, the free encyclopedia”
	4.4. Relation	**1**	**1**	/ There are occasionally links under sections ”See also” and ”External links”
5. Life-cycle management	*5.1. Date(s)*	**1**	**1**	- Last modified, version history with all dates of modifications
Total		**12,5**	**12**	
Maximum		**15**	**15**	

B ANALYSIS SHEET FOR IRC-GALLERIA

Service	IRC-galleria
Homepage	http://irc-galleria.net/
Year founded	2000
Primary concept	User profile
Primary concept occurences	>35000
Supporting concepts	User, photo, channel, community, user
Knowledge Domain	IRC users / Finnish people
Intented use	Socialize by sharing personal information, photos, diary entries
Source(s)	http://irc-galleria.net/parentsinfo.php

Metadata analysis based on Dublin Core

Category	Element label	Implicit	Explicit	Sample values / rationale
1. Locating	*1.1. Identifier*	**X**	**X**	/ Unique nicknames for all users
2. Content description	*2.1. Title*	**X**	**X**	/ Nickname + user entered description
	2.2. Subject(s)	**X**	**-**	/ List of channels, communities, user pictures and comments
	2.3. Description	**X**	**-**	/ List of channels, communities, user pictures and comments
	2.4. Language(s)	**X**	**X**	/ XML lang attribute
	2.5. Coverage	**X**	**½**	/ Location city per user, image descriptions may give some clue of spatial coverage
	2.6. Source(s)	**X**	**½**	/ Not explicit, however, according to use policies and context some source(s) could be derived. User comments are sourced
	2.7. Type	**X**	**-**	/ List of channels, communities, user pictures and comments
3. Access conditions	*3.1. Format*	**X**	**X**	/ All user profiles are about same format
	3.2. Rights	**X**	**-**	/ Every page have a reminder about copyright law, not easy heuristics for the rights are available
4. Context	*4.1. Creator*	**X**	**½**	/ All user content is created by those who have access to respective user accounts. In a way putting creator explicitly is not even relevant
	4.2. Contributor	**½**	**-**	
	4.3. Publisher	**X**	**½**	/ IRC-galleria publishes all user profiles, understanding this may perhaps not be easily made a heuristic
	4.4. Relation	**X**	**½**	/ Only comments provide direct links to other users, however, a paid version advertises a friend tool
5. Life-cycle management	*5.1. Date(s)*	**½**	**-**	/ Explicit dates about any modifications are not generally widely available. Only easily available event is "last logged in" + when commented
Total		**14**	**6,5**	
Maximum		**15**	**15**	

C ANALYSIS SHEET FOR MUSEUMFINLAND

Service	Museosuomi
Homepage	http://www.museosuomi.fi/
Year founded	2004
Primary concept	Historical artefacts
Primary concept occurences	> 2000
Supporting concepts	Artefact type, material, producer, producing place, producing time, artefact user, use place, use situation, collection
Knowledge Domain	Museums / history
Intented use	Museum/cultural data discovery
Source(s)	http://www.seco.tkk.fi/applications/museumfinland/

Metadata analysis based on Dublin Core

Category	Element label	Implicit	Explicit	Sample values / rationale
1. Locating	*1.1. Identifier*	**X**	**X**	/ Collection item numbers
2. Content description	*2.1. Title*	**X**	**X**	/ All artefacts have titles
	2.2. Subject(s)	**X**	**X**	/ Per artefact keywords
	2.3. Description	**X**	**X**	/ Usually many images plus textual summary
	2.4. Language(s)	**½**	**½**	/ Language is globally Finnish
	2.5. Coverage	**X**	**X**	/ Geographical, jurisdictive and temporary coverages per item
	2.6. Source(s)	**X**	**X**	/ Museum responsible + collection that artefact is in
	2.7. Type	**X**	**X**	/ Resource type + type schema
3. Access conditions	*3.1. Format*	**X**	**X**	/ Size,
	3.2. Rights	**-**	**-**	/ No per-item rights available
4. Context	*4.1. Creator*	**X**	**X**	/ Usually present
	4.2. Contributor	**-**	**-**	/ No contributors listed
	4.3. Publisher	**X**	**X**	/ Museum responsible + collection that artefact is in
	4.4. Relation	**X**	**X**	/ Artefacts that are related by equal properties
5. Life-cycle management	*5.1. Date(s)*	**½**	**½**	/ Timespan for artefact creation is specified, no further dates available
Total		**12**	**12**	
Maximum		**15**	**15**	

D ANALYSIS SHEET FOR DEL.ICIO.US

Service	Del.icio.us	
Homepage	http://del.icio.us/	
Year founded	2003	
Primary concept	Bookmarks	
Primary concept occurences	>	1000000
Supporting concepts	User, tag, bundle, neighbour	
Knowledge Domain	Bookmarks	
Intented use	Keep and mange own links, share them and discover new things	
Source(s)	http://del.icio.us/about/	

Metadata analysis based on Dublin Core

Category	Element label	Implicit	Explicit	Sample values / rationale
1. Locating	*1.1. Identifier*	**X**	½	/ All bookmarks have ID's, but per user identifiers are not easy to discover
2. Content description	*2.1. Title*	**X**	**X**	
	2.2. Subject(s)	**X**	½	/ Tagging may give some insight
	2.3. Description	½	½	/ Optional notes field
	2.4. Language(s)	**X**	**-**	/ From context, service assumes English
	2.5. Coverage	½	**-**	/ Tagging may give some insight
	2.6. Source(s)	½	**-**	/ Tagging may give some insight
	2.7. Type	½	**-**	/ Tagging may give some insight
3. Access conditions	*3.1. Format*	½	½	/ Most common Internet media types are recognised, such as sound has an icon
	3.2. Rights	**-**	**-**	/Usually unavailable, generally may be unimportant
4. Context	*4.1. Creator*	**X**	**X**	/ User owns the links him/herself
	4.2. Contributor	½	½	/ Other users who have added the same link
	4.3. Publisher	**-**	**-**	/ Generally unimportant or unavailable
	4.4. Relation	**X**	½	/ Other people who have saved this link, related links by tags, related links by user
5. Life-cycle management	*5.1. Date(s)*	½	½	/ Date added in format ”added x days ago”
Total		**9,5**	**5,5**	
Maximum		**15**	**15**	