UDC in subject gateways: experiment or opportunity?

Aida Slavic
UDC Consortium Advisory Board
aida@acorweb.net

ABSTRACT: The paper gives a short overview of the history of use of UDC in Internet subject gateways (SGs) with an English interface, from 1993 to 2006. There were in total, nine quality controlled SGs that were functional for shorter or longer periods of time. Their typology and functionality is described. Quality SGs have evolved and the role of classification has changed accordingly from supporting subject organization on the interface and automatic categorization of resources, towards supporting a semantic linking, control and vocabulary mapping between different indexing systems in subject hubs and federated SGs. In this period, many SGs ceased to exist and little information remains available regarding their status. SGs currently using UDC, for some part of their resource organization, do not use a UDC subject hierarchy at the interface and its role in resource indexing has become more difficult to observe. Since 2000, UDC has become more prevalent in East European SGs, portals and hubs, which are outside the scope of this research. This paper is an attempt to provide a record on this particular application of UDC and to offer some consideration of the changes in requirements when it comes to the use of library classification in resource discovery.

1 *Introduction*

In information retrieval, the importance of controlled vocabularies with a hierarchical structure, such as classifications and thesauri, is well known. Their most obvious purpose is in supporting subject browsing and enabling visualisation of subject access through hierarchical presentations of subject areas, thus enabling more interactive information retrieval. They also allow selection and contextualisation, i.e. disambiguation of terms in the process of searching, and can be used to improve recall through enabling search expansion based on semantic linking. The value of classification has also been recognized as significant in the digital and networked environment (Koch et al., 1997; Soergel, 1999; Hodge, 2000; Currier & Wake, 2001; McGuinness, 2002 etc.). Classification becomes even more relevant when it comes to providing access to subjects in a multilingual environment in which widely used library classifications have important advantages. This stems from their use of notation as language-independent indexing terms but also because such schemes have been translated into various languages. It is, therefore, not strange that readily available and known library classifications started to be used in supporting resource discovery on the Internet early in the 1990s.

The Web, as an information environment, differs from the controlled setting of a traditional information retrieval system for which bibliographic classifications were originally developed and in which they have been used for a long time. It may be important to see and understand how and to what extent a classification is actually used to support subject access. For instance, the way it is presented to users; whether this involves browsing or searching or both and which editions or what part(s) of the vocabularies are used etc. Such information may be instructive to both future users and scheme owners alike and may be useful in adapting and improving the way traditional library classifications are developed, maintained and distributed. This paper discusses the evolution and the change of use of one such traditional knowledge organization tool in the Web environment but some observation may be relevant for other general and special classification schemes.

Universal Decimal Classification has been used in document indexing worldwide since its first edition in 1904-1907 and it has long been considered an international *de facto* standard in indexing. Initially developed based on the Dewey Decimal Classification (DDC), UDC was re-designed to embrace a modern analytico-synthetic structure, which allows unlimited

combination of terms in indexing and retrieval and supports detailed indexing. As a result, UDC can be used, similar to other library classifications, for simple shelf arrangement (to any arbitrary level of specificity/complexity) but is often chosen as a tool by special libraries and bibliographic services for its strength in detailed indexing. Over the years, and under the ownership of the then owner, the International Federation for Documentation (FID), UDC has been developed to consist of up to 200,000 classes in its full editions. In 1992, the ownership of UDC was transferred from the FID to the UDC Consortium (UDCC), a non-profit organization of publishers based in The Hague. Since 1993, a standard version (in English) containing around 67,000 classes has been regularly maintained, updated, and annually distributed as a file (UDC Master Reference File). Further information regarding UDC structure, history, products and recent literature can be found at the UDCC Website (http://www.udcc.org).

Recent research confirmed that UDC is used in libraries and information centres in 124 countries. In 34 (mainly in Europe, Asia and Africa) it is the main classification and its schedules can be found translated into 39 languages (Slavic, 2006; 2004). It is important to note that UDC has never been widely used in English speaking countries, which, incidentally, were those most involved and influential in the development of the Web. Introduced some time after DDC (and LCC) had already been adopted by libraries in the United States, Canada, the United Kingdom and Australia, UDC has been used in these countries mainly in special libraries and bibliographic services for which its power in indexing and rich vocabulary offer necessary advantages over other library classifications. Today, there are UDC users in 43 English-speaking countries; however, the number of libraries and information centres using UDC is fairly low. Hence, the fact that the UDC first appeared in subject gateways with an English interface as early as it did is a point of interest for this research. In Europe, UDC is the main classification system in 22 countries where it is used across library networks. Apart from Spain and Portugal these are mainly central and eastern European countries that, since 2000, have been quickly catching-up with the development of Web based information services.

2 Research background

SGs[1] emerged as a response to the challenges imposed on resource discovery at the very earliest phase of the Web and became more prevalent towards the mid-nineties. The majority were developed by libraries, academic, educational or related institutions and only a few SGs were developed by commercial providers. The majority of services from this period were developed in the UK, USA, Australia, the Nordic countries, the Netherlands and Germany.

SGs offer a selection of resources with respect to quality, stability and authority, applying a set of quality measuring criteria and supporting systematic resource discovery (Koch, 2000, 2000a; Vizine-Goetz, 1998). They may be designed with a target audience in mind or be created for a certain domain, community of users or field of knowledge. From the management point of view, apart from programs for automatic resource harvesting and occasionally automatic indexing, SGs usually employ manual labour to test and improve quality. One of the common features of their management is that they base resource organization and presentation on metadata rather than on the resources themselves.[2]

Koch (2000, 2000a) distinguishes two levels of service: *subject gateways* which consist of a list of links with minimal descriptions and shallow subject structure and *quality-controlled subject gateways*, which are subject services based on comprehensive metadata descriptions and resource quality control. Quality-controlled SGs have a well-defined approach to collection management and development. They select resources based on an established

policy and quality evaluation criteria: expected durability of the resource (lifetime), its importance for the user community, the amount of information, size etc.. [3]

According to Koch (2000a), the typology of SGs may be based on coverage (related to the subject, geography, language of the resources or type of resource), collaboration with other gateways (shared selection for a common service, record exchange, translation and other co-ordination efforts, cross-browsing, cross-searching, mirroring remote service), and co-operation (co-operation in one subject area, close national co-operation in different subject areas, co-operation of an integrated national service, co-operation between subject gateways and regional or national digital libraries). Because of this, interoperability is one of the key principles in the development of SGs and consequently this translates into interoperability of subject vocabularies used in the organization of resources: vocabulary selection, their mapping and complementing of one another.

Under the pressure of increased labour in indexing, insufficient funding and the struggle to maintain users' interest, quality SGs evolved. The longevity of some of the best gateways has been linked to the deployment of new technological solutions,[4] inclusion of new user-orientated services and federation of individual SGs into subject orientated hubs and portals. Collaboration between SGs and sharing of technological solutions became a logical trend, crucial for their survival, while their involvement in national e-learning programmes seems to represent an important rôle for their future (Huxley & Joyce, 2004). Automatic metadata harvesting and automatic indexing and classification have become fundamental in resource discovery and the adoption of these developments is important for quality SGs (cf. Dempsey et al. 2004). In 2000, the number of *SG*s was still said to be growing, and existing entities were constantly changing in scope, coverage, content, methods and features (Koch, 2000a).

At the time, Dempsey (2000) suggested that the overall trend was towards brokerage services based on distributed communication components. He also pointed out that the future of these services would depend on their ability to move to some sustainable funding source and that it was not evident whether they were going to retain their institutional identity or whether they were going to be absorbed into a national learning services, professional portal services, etc. In July 2006, for instance, eight of the U.K. hubs (*Altis*, *Artifact*, *BIOME*, *EEVL*, *GEsource*, *Humbul*, *PSIgate* and *SOSIG*) were re-launched through a single service now called *Intute* which continues to offer a "free national service enabling lecturers, researchers and students to discover and access quality Internet resources" (see http://www.intute.ac.uk/development/).

Collaboration and federation of subject gateways, portals and hubs, aiming to provide a better and more sustainable subject orientated service to a wider audience, has been the obvious trend after 2000. International collaborative information services have become an important part of this new phase (cf. Koch, Neuroth & Day, 2003; Stoklasová, Baliková & Celbová, 2003). This development coincides with a wider presence of SGs and portals in non-English speaking parts of the world. Without digressing into more technological issues of cross domain / cross language resource discovery, suffice it to say that this context puts greater stress on the interoperability of the infrastructure supporting SG services (cf. Serre, 2004; Dempsey et al. 2004). In this context, language-independent controlled vocabularies such as classifications seem to continue to play an important rôle.

3 Research objective and method

This overview of quality SGs with an interface in English, that used or still use UDC, is an attempt to record a small but potentially very important niche of UDC application. The way classification is used in resource discovery may indicate if there are any new and important requirements with respect to the classification scheme that need to be accommodated in the future.

Information on SGs was collected through literature, mainly articles, conference papers and project reports. In 1995 I compiled the first list of directories using UDC that was used when teaching classification and which was maintained until 1999. From 1999 to 2005 my observations continued while maintaining a list for the UDC Consortium website. As far as it was possible to establish through Internet searching and literature research, I had reason to believe that the list included all such services with an English language interface. Part of the information that was important for this research was collected through searching and browsing services and checking the level of UDC use. The focus was on the way the classification was presented on the interface (notation, caption), if it was visible or hidden, if it was present in metadata and if so, were these displayed on the interface. The majority of information regarding the technical side of UDC implementations, and especially when UDC was used hidden 'behind the system', had to be found in the literature. It was fortunate that many of these applications were part of research projects, which were concerned with research dissemination through publications and conference reports.

The focus on gateways with an English interface was influenced by two facts. The first is the fact that UDC is used in a large number of countries; hence different languages and scripts are an obvious problem. Although it was possible on many occasions to search and discover these services, the language and script barrier limited the possibility in testing the functionality and observing the way the classification is used in browsing (especially if notation is not present). Finding, accessing and using the literature on subject gateway projects in various national languages and scripts is also difficult. For instance, if only Europe is taken into account, the UDC is used in 41 countries, which means that one has to deal with over twenty languages and three different scripts. The second reason why the services with English interfaces were considered sufficiently interesting for this research was their early introduction, which provided a longer period of time for recording the changes and evolution of services and their attitudes towards classification.

This research looks into the type of services, their lifespan, and the way they use UDC. The objective is to record the changes in the use of classification as noticeable on the SG's interface or reported in the literature, from their creation to 2006. It is outside the scope of this paper to assess the general influence that library classifications, in this case UDC, have had on resource discovery on the Internet. The objective of this research was rather to look into what impact its use on the Internet may have on the classification itself. With respect to this, some implications on the scheme use and development are considered and some new possibilities for the UDC are suggested.

4 Subject gateways using UDC

UDC in SGs appeared to be linked to the following types of applications:

- manual classification of manually collected links on small to medium-size directories (from a few hundred to a few thousand resources), with or without the help of metadata (i.e. simple directories of links)
- manual classification of a large number of automatically harvested resources using harvesting and metadata creation tools and more advanced technology (quality controlled SGs)
- automatic harvesting and classification (quality controlled SGs)

Simple directories that used UDC for the basic organization of links when the Web was still in its infancy, were rather short-lived. They gradually disappeared when manual link collection and update, usually performed by individual enthusiasts, could not keep up with the growth of the Web. Very few readers will still remember directories such as *Informazioni Classificate per Discipline*, *Services in classified order* at the University of Wales, or the

Plambeck UDC directory. It is a quite different situation for institutionally based and maintained quality SGs, which although they may use UDC in a very limited way, stand as representatives of this specific kind of its application.

The majority of quality SGs using UDC (with an English interface) were developed in the U.K. as a part of the *Electronic Libraries Programme* (eLib). In 1999 some of these services joined the *Resource Discovery Network* (RDN) a free national gateway for the educational and research community - a network organization consisting of subject hubs such as SOSIG (social sciences hub), EMC (engineering, maths, computing hub) and Humbul (humanities hub).

The changes in SG services that occurred towards the end of the 1990s, affected the use of classification. With respect to SGs using UDC, the following changes were noticeable:

- closing of individual services caused by lack of funding
- migration from UDC to other classifications in order to meet the needs of a particular community of users
- disappearance of UDC based subject categories from the interface and their replacement with a more user orientated subject hierarchy
- shift towards use of UDC 'behind the system' as a pivot in managing various vocabularies or as a control in keyword assignment

There were in total nine SGs using UDC in the period 1993-2006 with an English language interface, two of which were based on automatic indexing/classification (Table I). All are freely available resource discovery services. Three of these are non-UK based SGs with a parallel interface in languages other than English. Only two of these gateways are special in subject coverage (PORT and SOSIG) and only two emerged after 2000 (OKO and PORT).

Table I: Overview of quality subject gateways using UDC 1993-2006

	Created	**Status 2006**	**Subject coverage**	**Interface language**	**Content indexing**
1. WAIS/WWW	1993	accessible but not updated	general	English	automatic
2. BUBL	1994	operational, no UDC 1996-	general and LIS	English	manual
3. OMNI	1995	operational, no UDC 1998-	medicine	English	manual
4. NISS	1995	not operational from 2003	general	English	manual
5. SOSIG	1994	**operational**	social sciences	English	manual
6. FVL	1996	**operational**	general	En., Fin., Swed.	manual
7. GERHARD	1997	not accessible from 2006	general	En., Fr., Germ.	automatic
8. OKO	2000	**operational**	general	Eng., Sloven.	manual
9. PORT	2000	**operational**	maritime information	English	manual

Bulletin Board for Libraries (BUBL) was the first quality subject gateway in the U.K. that used UDC, and it set an example for other services to follow. During the period 1990-2006, UDC was frequently mentioned in relation to the following services: Nordic WAIS/WWW, BUBL, FVL (now Science Linkhouse), OMNI, NISS, SOSIG, GERHARD, OKO - Slovenian catalogue of the Web resources and PORT. Some, such as BUBL and *Organising Medical Networked Information* (OMNI) migrated very early to other classification systems.[5]

Table II shows a summary overview of UDC use that can be observed for seven subject gateways using UDC (BUBL and OMNI not included). Four of these systems use very little of the UDC vocabulary (NISS, OKO, SOSIG, WAIS/WWW) and three do not use UDC

synthesised concepts. Two gateways, PORT and FVL, are probably using some of the recent editions of the UDC but it is not possible to confirm this assumption, as they are not displaying the UDC structure on the interface or UDC numbers in the metadata. However, they contain subjects that are pre-combined concepts in UDC.

Table II: The use of UDC

	Hierarchy display [6]	**Hierarchy level** [7]	**caption**	**number synthesis**	**shown in metadata**	**number search**	**search to browse**	**Use UDC MRF**
1.WAIS/WWW	YES	3	YES	NO	NO	NO	NO	NO
2. NISS	NO 2001>	---	NO 2001>	NO	YES	NO	NO	NO
3. SOSIG	NO 2000>	---	NO 2000>	NO	NO	NO	NO	NO
4. FVL	NO	---	NO	YES	NO	NO	YES	----
5. GERHARD	YES	no limit	YES	YES	NO	NO	YES	YES [8]
6. OKO	YES	1-4	YES	YES	NO	YES	YES	NO
7. PORT	NO	---	NO	YES	NO	NO	YES	---

All services that are based on manual content indexing are based on metadata although the metadata records are not always shown in the result display. The general trend is, also, not to display UDC numbers on the interface (only WAIS, which is not updated, and OKO do). In 2006 only three services displayed a UDC hierarchy and showed UDC captions compared to five before 2000. FVL, NISS, PORT and SOSIG have a gateway specific subject directory for browsing.

4.1 Subject gateways with manual content indexing

All five gateways that are described below can be considered medium scale quality information services. They use metadata and both indexing and metadata production are fully or partially manual.

NISS - National Information Services and System at http://www.niss.ac.uk (Bath University) has been maintaining professional information services for UK education since 1988. In the 1990s, its function was extended to government information, health and other sectors. The part of the NISS service using UDC was the Directory of Networked Resources (Lafford & Stone, 1997). As is usually the case, the NISS system was modelled on library practice i.e. it was based on a catalogue of Internet resources and used simplified UDC numbers to provide subject searching and browsing (Figure 1).

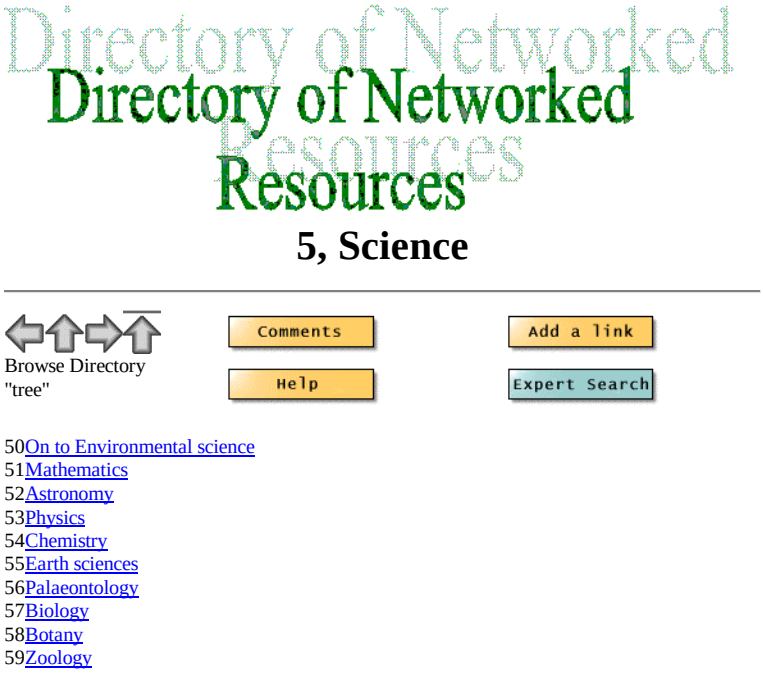


Figure 1: Subject directory in NISS in 1999

Up to 2003, the subject gateway was browsable by inverted 'subject tree' UDC menus, by UDC classmark, and by alphabetical subject heading.

Since 2003, NISS Campus has no longer been maintained. Its content was transferred to the HERO (Higher Education & & Research Opportunities) portal at http://www.hero.ac.uk (Lonsdale, 2003). HERO offers access to subject directories through a simple menu of eight broad subject areas such as arts and sport, humanities, applied sciences, language and literature etc. (Figure 2).

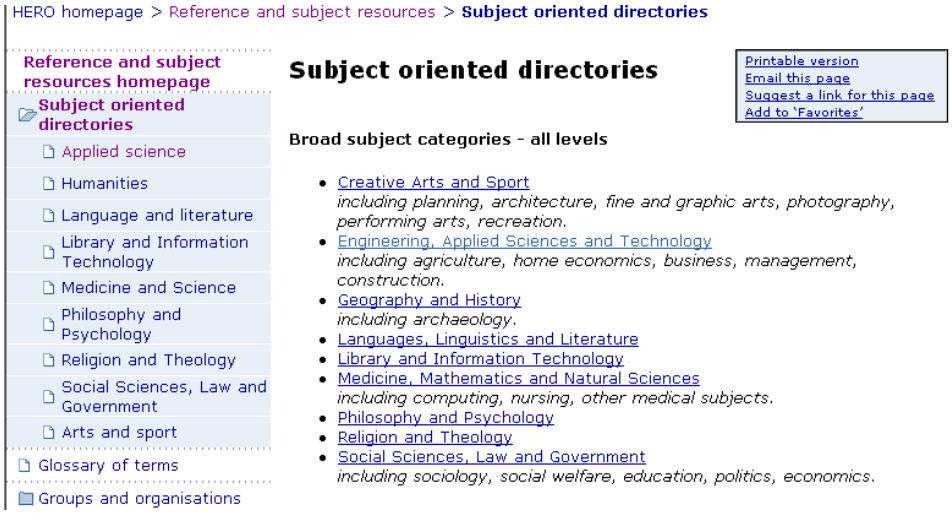


Figure 2: Subject approach directory in HERO in 2006

SOSIG - Social Science Information Gateway at http://www.sosig.ac.uk (Bristol University) is an Internet catalogue containing around 28,000 quality selected information resources for students, academics, researchers and practitioners in the social sciences, business and law (Huxley & Joyce, 2004). Launched in 1994 as an information service for

social sciences, it was re-launched as a social science, business and law hub with a new interface and additional features [9] as a part of the RDN in 2000 (Huxley, 2000; Huxley & Joyce, 2004). The service changed its interface again in 2006 when it was re-launched through *Intute*. SOSIG was one of the most important gateways used as a test bed for the *DESIRE* project and it has applied the project's automatic Web harvesting technology. Since then, SOSIG has continued exploring XML/RDF technology in the storage, query and interchange of controlled vocabularies such as HASSET and UDC (Hiom, 1998, 1998a).

Initially, SOSIG used a total of 161 UDC numbers from an obsolete scheme edition. A smaller selection of these classes was made available for browsing and the remainder was used for cataloguing. The detail of the numbers used to vary from being at the top of a hierarchy (e.g. Philosophy = 1) to being fairly specific (e.g. Environmental Issues = 551.588 or 658.562 = Quality management). Interdisciplinary resources or those treating more than one subject were given as many UDC numbers as necessary in a manner of post-coordinate indexing (Hiom, 1998). The UDC-linked browsing hierarchy was replaced in 2000 by a hierarchy of subjects that now consist of 17 headings and over 1, 000 sub-sections. UDC is still used in the process of indexing to control keyword assignments for 13 out of 16 main subject areas (SOSIG Guidelines for assigning keywords, 2004).

Catalogue OKO at http://www.zrc-sazu.si/oko/ is a Slovenian catalogue of Web resources hosted at *The Scientific Research Centre of Slovenian Academy of Sciences and Arts (SRC SASA)*. The catalogue covers a rather small selection of quality resources on the Web irrespective of the language, the criterion probably being their usefulness for the Slovenian academic and research community. In spite of its small size, it is included in this overview of the SGs because it has an interface in English. Also, with UDC numbers shown on the interface it illustrates the way SGs looked like in their earliest phase.
Resources in OKO are classified manually using UDC and can be searched and browsed in English and Slovenian (an alphabetical subject sequence and a UDC sequence are available for browsing). The UDC hierarchy contains 2-3 levels, while synthesised UDC numbers are searchable only as whole strings. The UDC edition used is from the 1980s (with the exception of *004 Computer science*). The total number of UDC classes used is around 700. A small increase in the number of resources indicates that not much effort has been put into its regular maintenance and update since it first appeared in 2000.

FVL - Finnish Virtual Library (in December 2004 changed its name to *Science Linkhouse*) at http://www.linkkitalo.fi/ is the result of the *Finnish Virtual Library Project* initiated in 1996. This is a quality subject gateway with general knowledge coverage. The languages on the interface are Finnish and English and it includes resources in Finnish and English (one database is in Swedish). The FVL cooperated with the *Renardus* (http://www.renardus.org) project and is now accessible through the *Renardus* service. The FVL i.e. Science Linkhouse combines its own classification with UDC, DDC and NLM classifications. UDC is used as an underlying structure for the extraction of subject classes for browsing in certain areas but the original UDC structure is not shown.

PORT - Maritime Information Gateway is the National Maritime Museum's (London, U.K.) subject gateway at http://www.port.nmm.ac.uk/about.html and provides catalogues of around 2500 Internet resources containing maritime information. The gateway is created in collaboration with ILRT (Institute for Learning and Research Technology at the University of Bristol) and is based on a model developed by other UK subject-based information gateways as part of the *eLib* programme. Librarians in the National Maritime Museum select resources, classify them manually using UDC and create metadata.

Resources are organized into twenty subject categories which, at first glance, have no resemblance to the UDC structure. The display of a full subject hierarchy (listed alphabetically by main heading), however, shows a structure that is obviously based on more complex UDC numbers. Each subject category can then be browsed by choosing from period,

form, place or some other helpful facet and it is obvious that this kind of restructuring and reorganization of display is achieved with the help of the UDC (Ashton & Robertson, 2000).

4.2 Subject gateways with automatic indexing

Automatic text processing and indexing is usually implemented on textual resources in a single language. When terms extracted from the documents are, however, mapped to a given classification vocabulary, resources may be categorised and a subject hierarchy accessed via any language in which the classification scheme is made available. There were a number of projects that explored the application of library classification for automatic indexing and categorization, between 1993 and 1999.[10] Research was largely focused on the ability of general library classifications to hierarchically group and organize resources into smaller subject-related "chunks".[11]
UDC was the first classification used for automatic indexing of Internet resources and the first one to support automatic indexing of Internet resources with a multilingual search interface. There were two quality subject gateways using UDC in the period 1993-2006 but in 2006 neither of these is still active. Since they still represent an important and relevant model in resource discovery with two very different levels of complexity and scope, they are hereby represented in more detail.

Nordic WAIS/WWW, at http://www.lub.lu.se/auto_new/UDC.html, is one of the first Internet information services to use classification and was certainly the first to apply automatic classification in 1993. The project was funded by *Nor Info* and involved the *Danish Technical Library* and *the Lund University Library*. The project goal was to apply a library system model on the Internet and to build a gateway for Wide Area Information Server (WAIS) databases. Its primary aims were to enable the automated processing of WAIS source files to produce classified indexes available over the World Wide Web, and to develop better gateways between WAIS and the Web. In its final report it was established that the project successfully demonstrated the possibility of network-based resource discovery and retrieval through a subject gateway based on library classification. (Ardö et al., 1994).

The project had three parts: demonstration of how an existing library system (ALIS) can be integrated in an open client/server architecture, simplifying the use of the WAIS database by establishing a Web interface and combining hypertext browsing with search options. The project used only 51 UDC numbers to classify approximately 700 WAIS database descriptions that were quite homogeneous in their content.

Vocabulary from the WAIS databases' description was indexed and compared to the UDC vocabulary using only the two top levels of division (there was no explanation as to whether or not it was only from the UDC caption or whether a richer index was created for the purpose). The classification number was automatically assigned to the source when a term match was established. Based on the final classification, the individual WAIS database was linked into the gopher WAIS UDC subject tree. The classification was also used to build a hypertext hierarchy corresponding to the UDC structure.

GERHARD - German Harvest Automated Retrieval and Directory, was available until 2006 at http://www.gerhard.de, and was a fully automatic indexing and classification system of academic resources on the German Web and developed as a research project at the University of Oldenburg. The project had two phases (1997 to 1998 and 2001-2002) and was funded by the *Deutsche Forschungsgemeinschaft*. *GERHARD* was a database-driven robot that collected academically relevant documents which were indexed using computer-linguistic and statistical methods and classified by UDC. The user interface was trilingual (German, English and French) and allowed the user to look for similar documents very easily through its tight integration of searching and browsing mechanisms. In the second phase of the project statistical methods were improved, the UDC vocabulary was extended and the database was

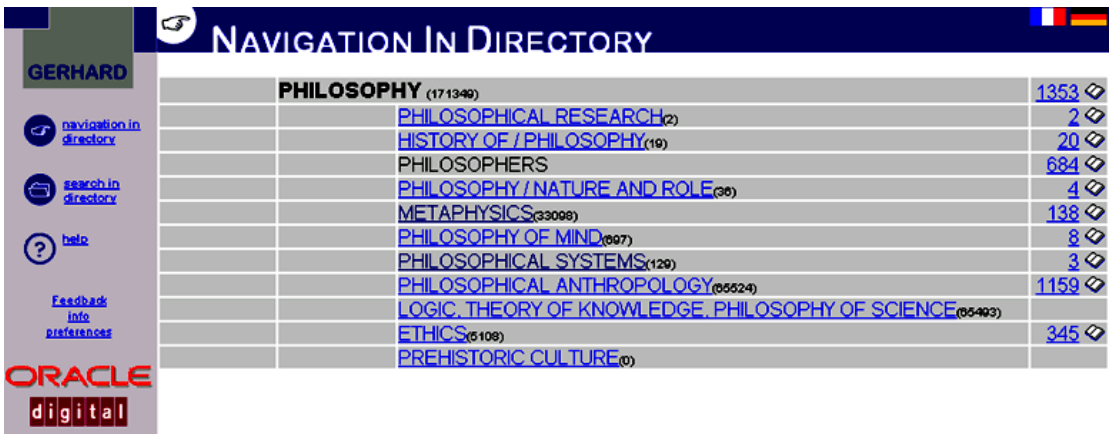
enriched with some manually classified resources. The development and architecture is explained in an article and conference papers (Wätjen, 1998; Möller et al., 1999; Möller et al. 1999a).

GERHARD's architecture consisted of a database-driven gatherer, fast automatic classification and an integrated searching and browsing service. The generated metadata and index of documents were held in a relational database (ORACLE) which contained almost 1,300,000 records. Automatic classification was based on the UDC authority files from the Eidgenössische Technische Hochschule (ETH) library system in Zürich (a system called ETHICS[12]). The ETHICS subject authority files support searching of compound and complex UDC numbers using indexing terms (in English, German and French) or numbers and it supports about 15 different relationships that could be established between UDC numbers. It is a reliable controlled vocabulary which consists of approximately 70,000 entries.

The first step in *GERHARD* was to extract a vocabulary called UDCZ-Lexicon from the ETHICS subject authority file. The creation of UDCZ-Lexicon was done automatically by conversion of 500,000 lines of text from ETHICS authority data to language expressions that normally occurred on Web pages. Conversion consisted of morphological analysis of each word in the UDC entry and its reduction to a stem. To each stem a variable was added in order to enable a specific form of words that may occur in the text. Once made, this lexicon was compiled to serve as a "recogniser" with the purpose of identifying regular expressions and providing the corresponding UDC notation.

German Web pages were harvested into the database and underwent text analysis to extract words that were matched to the UDCZ-lexicon and then assigned a notation. Notations were clustered and an occurrence frequency was weighed and statistically compared with the UDCZ database. Additionally, quality and specificity were checked using special algorithms until the average number of 14 per document was reduced to the best six-eight notations (secured by threshold applied). Finally, there was a statistical post-processing in which doublets of documents were weeded by matching the title of the document with titles already in the database. Titles are analysed and matched to exclusion lists produced to prevent classifying pages such as e.g. Web server statistics. The notation gathered from titles, headings and body text were matched and their concordance and precision was further checked through special algorithms.

GERHARD permitted both navigation and searching through a subject directory. Each subject class was hierarchically displayed and was represented with a short class caption. Moving through the hierarchy of categories was easy and user-friendly. Against each class there was a 'hyperlinked' number of documents available and users could browse the list of documents for each level of the hierarchy (Figure 3).



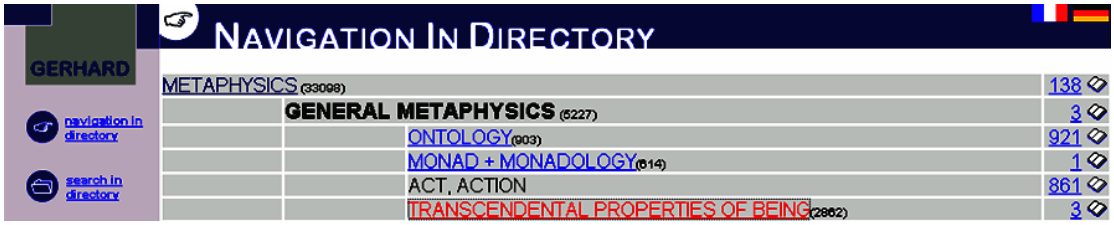


Figure 3: Two steps in directory browsing on GERHARD

Within each class documents were ranked according to accuracy. This was said to be achieved through the use of the complex UDCZ/SOIF (Summary Object Interchange Format) design (Möller et al., 1999). *GERHARD* was the most sophisticated application of UDC, and classification in general, for the purpose of the Internet resource discovery and while it was still accessible was the best interface example of the UDC browsing feature.

5 Summary and discussion

In spite of the fact that it was not a widely used scheme in English speaking countries, UDC seemed to be popular in resource discovery in the earlier phases of the development of the Web. There were several smaller manually compiled directories that gradually became unable to cope with the growth of the Internet, became obsolete and disappeared over time. These used a selection of top levels of UDC schedules from old printed editions. More advanced gateways using UDC appeared within the *eLib* project towards the mid-nineties. By 2000, quality SGs evolved into more complex services and their requirement towards controlled vocabularies changed in terms of interoperability, availability of mappings between systems and verbal access to subject categories. This evolution is best illustrated through the history of SOSIG (Huxley & Joyce, 2004).

It is indicative that NISS and SOSIG (and BUBL and OMNI before they changed to other classifications) used old and obsolete printed editions of the UDC. The reason for this may be either because they were unable to sustain the costs of a UDC MRF licence or this was thought to be unnecessary, considering the small count of UDC numbers they used. According to the statements from the DESIRE I project report in 1997, it transpires that the collaborators from these gateways did not have knowledge of the content of a standard UDC MRF file and its revised editions that preceded their report.[13] On the other hand, although it made no practical use of UDC MRF data, GERHARD purchased a UDC MRF licence for several years.

While it was easy to describe the level of UDC hierarchy applied in the early phases of general subject gateways (e.g. WAIS/WWW, NISS), this is not so with subject specific SGs (e.g. social sciences) which require more specific or pre-combined UDC concepts or services that use classification in the background. No matter how small the selection of UDC used in a SG, it still may contain numbers from the 6th or 7th level of a hierarchy.

No subject gateway (apart from OKO) allowed a search of UDC numbers, which was expected. By hiding the UDC notation and by altering the subject browsing structure and captions, SGs' interfaces have become more adjusted to the information needs of their users. If the reason for the use of library classification was, in the first place, to provide a browsing hierarchy it is then very indicative that subject orientated SGs have replaced the UDC browsing structure with a list of subject headings. With SOSIG this happened in 1999, as soon as it joined the RDN, as this consisted of six institutions that used different subject vocabularies (Huxley & Joyce, 2004). Subject headings in SOSIG and PORT, which have specific subject coverage, represent a shortcut to the subject areas of interest to the majority

of users, reduce the length of navigation, and make it easier for users to access the most pertinent subjects. In parallel to this, the use of UDC in content indexing 'behind the system' allows consistency in indexing and its independence from the interface browsing structure to which UDC terms are mapped. This means that the browsing categories on the interface itself can evolve and be adjusted easily and independently from subject metadata and can be adapted to the changes in users needs.

One can summarise by saying that the infrastructure behind browsing directories has significantly changed and grown in complexity and classifications have moved from their 'naive' presentation on the interface to metadata and authority files, in which they are used to control natural language terms. For this purpose a subject vocabulary may be managed as an authority tool and shared between metadata repositories. It can serve as a source of keywords or as a way of semantic control.

UDC's most straightforward use is in automatic classification. However, the quality of GERHARD's browsing interface, as one example of such an approach, is due to the fact that it was built on an unusually well developed library classification authority file which contained richer and better linked classification data than that available in the UDC MRF. UDC headings, from the ETH library authority file, are linked to a tri-lingual thesaurus, and contain codes for handling hierarchies that are great help in building a browsing interface (cf. Loth, 1996). Generally speaking, however, the size of the UDC MRF vocabulary (around 67,000 numbers) may represent a good vocabulary base for automatic classification. More so, since UDC captions are well adjusted to contain meaningful terms, as a result of its long history of use in online retrieval systems, as opposed to library classifications that have never been used in IR and usually contain context dependent and often meaningless captions.

It is important to note that the total number of SGs using UDC, with an English interface, decreased from 9 to 4. In addition, the two with automatic indexing (WAIS/WWW and GERHARD) are better described as successful experiments rather than real user orientated resource discovery services. The count of SGs using UDC, especially after 2000, would be, however, very different if the East European countries were taken into account. After this period, new portals and gateways using UDC appeared mainly in Central and Eastern Europe, especially in the Russian Federation (Stoklasová, 2003).[14]

6 Implications

The owners and users of UDC alike may be uncertain as to how to interpret this relatively short phase in a long history of UDC applications. It may be, therefore, instructive to examine the impact the use in resource discovery had on other classifications, DDC in particular. The research in improving classification data, enhancing the subject alphabetical index and testing and classification mapping to one another and to subject headings seemed to be encouraged by the owners of LCC and DDC.

A comparison between DDC and UDC, as reported by the *Renardus* project in Table III, clearly indicates how the previous research and testing of classification systems can lead to the 'enhancement of classification data' and can clearly influence and determine the future use of a particular scheme and selection of classification for cross-browsing/searching. As shown in Table III, Neuroth & Koch (2001) put their arguments for the choice of classification around the following requirements:

- online availability and tools
- global usage
- functionality of classification system
- updates (frequency and character)
- research/development efforts

- use in other gateways

In spite of a few weak points, they have put forward three important arguments why DDC was more suitable for *Renardus* than UDC and these are absolutely fundamental:

- research based development supported by its owner
- mapping to other vocabularies
- being used in many gateways

Table III: Comparison between DDC and UDC in selecting the classification

CRITERION	DDC/UDC COMPARISON (Neuroth & Koch, 2001)
ONLINE AVAILABILITY AND TOOLS	The DDC was adapted earlier and more quickly to usage in digital systems via the Internet. It is completely and easily available as "WebDewey" for all Web browsers and platforms. Furthermore the DDC is connected to large databases of real documents via the Web, incl. the CORC service, that allow the verification of potential subject content of classes and the correct usage of the classification.[15]
GLOBAL USAGE	The DDC has a much larger user community on a global level. DDC is more frequently chosen as the classification system in Internet resource discovery services. Some gateways already use DDC or provide a mapping to DDC.[16]
SUITABILITY OF THE CLASSIFICATION SYSTEM AND ITS FUNCTIONALITY	The UDC is probably more "modern" and has made faster progress towards a faceted structure. Its way of allowing the construction of new special classes by a straightforward relationship between existing ones is in the Renardus application context, however, a disadvantage. In addition, the UDC has 61 000 classes; three times larger than the DDC and might be far too detailed for Renardus' purpose.[17]
FREQUENCY AND CHARACTER OF THE UPDATES:	Both systems have large amounts of outdated captions and class structures. Both are slow with integration of external standard vocabularies.
RESEARCH AND METHODOLOGICAL DEVELOPMENT EFFORTS	With the rich, broad and long-term activities of the OCLC Office of Research Knowledge Organization Group and related Dublin Core metadata and XML/RDF work, the DDC has a decisive advantage when it comes to research based development potential and is future proof in digital library contexts. So the enhanced DDC contains intellectually and statistically mapped vocabularies like the LCSH which are extremely useful in classification and mapping work.
USE IN OTHER SGs	The main advantage of DDC here is the effort that has been made to adapt captions of the system to general end-users and, even more importantly to Renardus, to the real content of digital documents in today's Internet.

As the DDC's presence in subject gateways made a strong case for its use in *Renardus* (Table III), in return, the DDC presence in *Renardus* has influenced its further use in other national and international gateways that planned to link to *Renardus*. A recent report that included research on subject gateways in six countries, given by Stoklasová (2003), shows that the predominant classification in SGs in four of them: Australia, Norway, Sweden and the USA was DDC. According to her research, UDC was more frequently used in information services in the Czech Republic and Russia, which led the author to emphasize the importance of mapping different classification systems to DDC. For instance, the concordances between DDC and UDC (500 classes) were reported as being prepared for the Czech Uniform Information Gateway in order to make it interoperable with gateways using DDC (Stoklasová, Baliková & Celbová, 2003). Digital repositories and open archives (learning materials, electronic journals, research papers and theses) in Western Europe may also become biased towards DDC if this scheme is already used in national and international gateways, since the interoperability of subject access is an important requirement for their integration (cf. Koch, 2006). The wider use of UDC may, however, still be expected in central and eastern European countries in which it is a dominant classification system across library and information networks and which has been quickly catching up with Web developments since 2000.

7 Possibilities

The Web interface has been greatly improved by enabling browsing using expandable lists, interactive frames and navigation through hyperlinks that span windows and frames and allow the display and easier tracking of browsing steps. Following this, subject orientated taxonomic structures on gateways and portals show the trend to be combined with, or replaced by, a faceted organization of object/subject properties and attributes (Merholz, 2001; Devadason, 2003; La Barre, 2004; Ramshirish, 2004).

Discussion of the benefits faceted classifications may bring to resource discovery has been going on for some time. This is true whether one interprets faceted classification in a 'pure' sense as a system whose structure is based and built on fundamental concept categories; or in a pragmatic sense, i.e. the system that organizes vocabulary in mutually exclusive facets of concepts usually based on a specific practical purpose. Some papers describing the potential benefits a pure faceted library classification may have in resource discovery, deal with the subject on the theoretical plane mainly. The authors do not explain the role that the distinction between fundamental facets may have in the information seeking process in a real IR environment. Even more importantly, they do not suggest how the proposed library classification schemes, containing data that is not adjusted to be processed by programs, can support resource discovery services or any facet based, non-linear browsing on their interfaces (Ellis & Vasconselos, 1999, 2000; Broughton & Lane, 2000; Broughton, 2002). On the contrary, Web portal developers were primarily interested in the benefits of concept organization into practical and purposeful categories that their users/customers may easily combine or independently navigate. The practical applications and demonstrations, in this case, have often preceded the literature describing them. In addition, these applications led to the development of a format and tools for the management of faceted vocabularies for their use in a Web interface (Tzitzikas et al., 2002; Van Dijk, 2003).

UDC is an analytico-synthetic classification that can be related to both of the above interpretations of a faceted classification, although it is more often confused with purely faceted schemes. This is because some of the disciplines in UDC are, indeed, based and organized on the basis of fundamental facet categories. But what is more relevant for a 'faceted interface' is that the whole of the UDC vocabulary is organized into separate and clearly marked, independent parts specifically designed to enable concept combination in information retrieval. It is therefore easy to imagine an UDC supported interface in which one would be able to independently navigate and combine a facet of subject with a number of facets of common isolates e.g. place, time, materials, persons, ethnic grouping etc. The attention given to faceted as opposed taxonomical structures simply demonstrates that a new window of opportunity may be opening for UDC, providing its data is made more affordable and readily available in standardized vocabulary formats that support hierarchical and facet browsing.

The fundamental question still remains whether an *a priori* automatic classification is a good choice for resource discovery on the Internet. There are very few user studies measuring IR performance or assessing the information seeking behaviour on SGs with browsing based on a library classification but none has been conducted on the SGs described in this paper. There is, however, a report by Koch, Golub & Ardö (2006) on user browsing behaviour in a DDC based service in *Renardus*, which seems to be relevant. The authors have established the dominance of browsing over searching. However, their research was based solely on service log analysis and this raises questions. It is not clear how many of the browsing log sessions analysed belonged to international project members, associates or readers targeted through the project dissemination whose focus was primarily to check the browsing interface (which was singled out as *Renardus*' unique feature), and were not genuine information seeking users. Also, although 650 robots were recognized and excluded from the initial log data, it still remains uncertain whether all data analysed really represents human users' behaviour.

Another thing that may be questioned, without proper user feedback, is the authors' assumption that session lengths, or as they say "long and highly elaborate paths", is really an indicator of user satisfaction with the function of classification browsing.
Sparck Jones (2005) stresses the dangers of hidden and ill-grounded biases which cannot be avoided with any *a priori* given classification, in particular when combined with human indexing, are even greater in the context of the Internet. She points out that researchers (in the field of natural language processing) exploring how the words are actually used in large text corpora found that they reflect a "different and changing paradigm of cultural worlds". She allows that classifications developed by humans as they appear in ontologies may be justified in a special technical context that requires richer descriptive apparatus. When it comes to a general context, however, recent research in machine learning applied to web data shows "that it can recover surprisingly refined information" (Sparck Jones, 2005: 600). Lacking better solutions, the use of library classification in quality search gateways, both with manual and automatic indexing, seems still to be relevant in 2006, especially in a multilingual context.

8 Concluding remarks

In the evolution of subject gateways, from simple directories to subject orientated hub services, UDC was seen from its simple rôle as a tool for link organization on static web pages to a more sophisticated source of vocabulary for automatic classification and finally as a way to control and link vocabulary behind the system. As is the case with other library classifications, the presence of UDC in resource discovery services has been largely experimental in nature. This was, however, very important as it introduced the scheme to a new kind of user and thus contributed to being better known outside the library domain. The further use of UDC in resource discovery is equally important as this new application may test the scheme, as well as highlight and speed up the changes necessary to ensure its better development.

From this research, it transpires that there are some new approaches in the way UDC is used in the Web environment and these may be relevant for any library classification scheme. The most obvious is related to the nature of applications and their attitude towards human *vs.* automatic indexing, which indicates that any development and the support a classification scheme can offer for the latter, might be more relevant. In addition, while libraries recognize and use classifications primarily because of the purpose its notation has in collection organization, classification in the networked environment is employed to support vocabulary control and semantic relationships between subjects relevant for resource discovery, while notation may or may not be used or displayed at all. In relation to this, classification vocabulary, i.e. class captions, can be used as a source of natural language vocabulary for coordinate indexing or term matching in automatic classification or may even be extended or mapped to the terminology more pertinent for the service. Thus we can safely assume that the use of classification as a tool "behind a system" is probably that which is more relevant for subject access on the Internet. Also, unlike libraries, which tend to continue using a classification system once it is implemented, on the Internet we are witnessing a greater speed with which the changes in the use of vocabulary occur. This may happen under the influence of user feedback, service development or oscillation in funding and this may impose even greater stress to classification maintenance and development. Equally necessary to take into consideration are the differences in the requirements the implementers of Web applications may have with respect to the availability of a classification vocabulary (formats, language, mapping), freedom of adaptation/alteration/implementation, versioning, extraction, expansion, and more relaxed copyright. As we have seen in the example of GERHARD in the case of UDC, classification application on the Internet may be based on a subject authority file and not on a standard scheme. Last but not least - subject indexing expertise or investment into it, which was assumed in traditional bibliographic services, may not be available for Web-based implementations. This may mean that the choice of classification scheme for a

Web service may not be based on indexing or classification expertise and will rarely be motivated by an intrinsic quality of the scheme itself. Instead this is likely to be influenced by availability, familiarity, popularity, cost of the scheme and other circumstances that are normally influenced by scheme owners.

With respect to UDC specifically, and in order to facilitate its applications on the Web, some changes in scheme management and distribution, with respect to the networked environment, may be worth considering. These changes relate to two types of requirement pertinent to the majority of stakeholders: one relates to the quality of UDC data, the other to the conditions of use. The quality of UDC data will determine how easy it is to implement in resource discovery services and comprises of:

- export data format according to bibliographic and network standards
- machine processable semantic and structural linking
- controlled word access (i.e. subject-alphabetical index)
- availability of mapping to other systems
- availability of multilingual machine readable data

Classification schemes are not cheap to buy but are even more expensive to implement and use, which is why the UDC's future use, for instance, will depend on the following:

- clear information on UDC credentials as an international scheme
- affordability of UDC data i.e. adequate pricing with respect to the cost of implementation and running
- a transparent and logical copyright policy, better suited to the reality of the networked environment - concerning UDC use (rather than publishing) (cf. Stevens, 2005)
- easily accessible information on UDC, documentation and training

The Internet is a fast changing environment that evolves largely through short-lived experiments, testing and research, as is well exemplified with this overview of quality SGs. In spite of the fact that quality SGs, portals and hubs are very much alive, it is hard to predict what will be the future of UDC and any other library classification in this context. The above recommendations may, however, contribute towards increasing the opportunities of UDC in the Web environment. Research in the Web application of UDC in countries, in which it is a well known system and is traditionally used online in a more advanced way may produce more interesting results.

Notes

[1] Synonyms to subject gateway are 'subject-based information gateway' and 'subject service' and are sometimes used when other features are added to a gateway. The terms 'gateway' or 'portal' have a wider scope and are more general in meaning, while 'hub' stands for a model that offers several services, only a part of which can be a subject gateway (Koch, 2000a).

[2] Some of these services build a comprehensive catalogue of Internet resources based on a widely accepted metadata standard such as Dublin Core. Earlier services, such as NetFirst at The Online Computer Library Centre (OCLC), contained resource descriptions derived from a bibliographic description standard but extended with elements specific to electronic resources. Apart from descriptive information and summaries, a NetFirst catalogue record contains DDC classification numbers and LCSH.

[3] Kirriemuir (1999), for instance, identified 141 candidates where only 39 could be placed in the category of quality controlled. Koch (2000a) lists 50 subject gateways and quality-controlled subject gateways.

[4] This in particular related to automatic resource and metadata harvesting and indexing. In the U.K. for instance, the system support, i.e. a freely available and configurable kit for building information gateways, was developed by the ROADS (Resource Organisation and Discovery for Subject based services) project. The DESIRE project (DESIRE I -1996; DESIRE II -1997) - a development of a European Service for Information on Research and Education, for instance, was an international effort specifically aimed at the improvement and development of resource discovery services focusing on both strategic and technical issues. Among other things they explored the rôle of classification in knowledge organization on the Internet (www.desire.org), including automatic classification. One of the useful project outcomes is the DESIRE Information Gateways Handbook (http://www.carnet.hr/CUC/cuc2000/handbook/welcome.html).

[5] BUBL changed from UDC to DDC in 1996 when the service obtained permission to use DDC from OCLC. OMNI changed to the NLM classification in 1998 as it was more suitable for resources in the field of medicine.

[6] UDC based subject hierarchy shown on the interface (with or without UDC numbers).

[7] Approximate number of hierarchical levels of subjects available for UDC browsing.

[8] GERHARD paid for the UDC MRF licence and had access to the file, although it is not certain whether this was used at all.

[9] These include personalized alerting services Grapevine and MyAccount, which provide a selection of resources, conferences, courses and communication channels for selected areas.

[10] Automatic text indexing matched to a classification structure given in advance is also known as an *a priori* automatic classification as opposed to document clustering, according to a derived classification which is termed *a posteriori* (Serre, 2004; Sparck Jones, 2005).

[11] One of the research institutions that started to explore this field early in 1992 was NetLab - the Research and Development Department at Lund University Libraries (http://www.lub.lu.se/netlab/). Members of the research team participated in several important projects on automatic indexing of Web resources using classification including, Nordic WAIS/WWW project, DESIRE I, DESIRE II and *Renardus*. Also, the OCLC's Scorpion project, was one of the most important contributors to the field of automatic indexing and classification of Web resources. DESIRE II, for instance, demonstrated that the use of a library classification assists in achieving good results in automatic indexing with very simple weighting algorithms and simple heuristics (Ardö & Koch, 1999, 1999a; Ardö et al., 2000). The project's research, among other things, intended to explore the possibility of expanding and connecting the vocabulary between the Ei vocabulary, UDC and DDC (Koch & Vizine-Goetz, 1998). The outcomes and experiences from DESIRE II were taken further in the *Renardus* project (Koch, Neuroth & Day, 2001).

[12] The ETHICS Library Information Control System was developed and used in the ETH library from 1983-1999. In 1999 ETH migrated to the Aleph vendor system to support NEBIS (Netzwerk von Bibliotheken und Informationsstellen in der Schweiz) that consists of over 70 academic libraries (Pika, 2002).

[13] For instance, for UDC that is (since 1993) revised, updated and released annually there are statements such as: "...main weakness of the scheme is that it is out of date" or " ...SOSIG and NISS suggested that UDC is not updated often enough: some subjects are outdated, weak in environmental studies and developments studies" or "...does not have good vocabulary in all subject areas" or "...complex structure of the scheme is considered to be a problem [because] main tables of the classification can be combined with auxiliary tables.." or "NISS finds it too complicated to use composite classification codes and decimal notation does not reflect a true hierarchy" (Koch et al., 1997). These and other statements are not supported by any facts/examples and are not paralleled in description of other classifications in the same report. Hence, someone with no knowledge of library classifications may come to completely the wrong conclusions that: other library classifications are updated more frequently than UDC (i.e. have new editions more often than once a year), have good vocabulary in all areas of knowledge and do not contain obsolete terminology in certain areas. Or even more absurdly: that other library classifications do not contain auxiliaries which can

be combined with main schedules and that decimal point in other classifications does actually represent a true hierarchy. A similar approach in comparing classifications based on outdated or unverified information and partial knowledge of classification systems was later repeated in the HILT project report (cf. Currier & Wake, 2001: 10) and was copied even in some recent textbooks on classification (cf. Batley, 2005: 163).

[14] Some of these services may be using UDC for simple subject browsing such as EJOL - Electronic Journals Online Library at http://ejol.irb.hr/ (Stojanovski, 2006). More important examples are quality subject gateways using UDC for searching, browsing and vocabulary mapping, some of which can be found at the Russian Federal Education Portal (http://www.edu.ru) or the Czech UIG - The Uniform Information Gateway (http://www.jib.cz). Some European library and documentation services index Internet resources using UDC. A good example here is CADIST for physics (Centre d'Acquisition et de Diffusion de l'Information Scientifique et Technique) in Grenoble at http://web.ujf-grenoble.fr/BUS/Physique/Ressources_Internet_Physique.php. There are also services that use only a part of the UDC vocabulary, such as the ECLAS - European Commission Library Catalogue (http://europe.eu.int/eclas/), which includes Internet resources. This catalogue allows, in expert search mode, a combination of UDC area codes and ECLAS thesaurus.

[15] The fact that it is considered to be more rapidly adapted to the Internet through WebDewey may not be a valid argument as UDC-online was made available to users earlier than DDC (in June 2001), while the UDC database file has been distributed to users and can be easily purchased since 1993.

[16] One has to put a "much larger user community" at least in geographical context i.e. in English speaking and Western European countries. While DDC tends to be used widely (e.g. in public libraries) UDC is used in libraries with larger collections. For instance in Switzerland, where the majority of libraries use Dewey, the largest libraries are using UDC and hence the majority of the country's library holdings are classified by UDC.

[17] Every classification, including UDC, can be used to an arbitrary level of specificity and detail. UDC's ability to be used as a synthetic classification need not be exploited at all. For many purposes UDC can be used as a simple classification. In addition, even if subjects are synthesised in the process of indexing they may be treated as simple terms for the purpose of resource collocation and resource discovery.

References

ARDÖ, A.; KOCH, T. (1999) "Creation and automatic classification of a robot generated subject index", Digital Libraries '99, The Fourth ACM Conference on Digital Libraries, August 11-14, 1999 in Berkeley, California.
http://www.lub.lu.se/desire/poster.htm.

ARDÖ, A; KOCH, T. (1999a) "Automatic classification applied to full text Internet documents in a robot-generated subject index", 23rd International Online Information Meeting, London, 7-9 December 1999 : proceedings. Edited by B. McKenna. Oxford : Learned Information Europe Ltd., 1999, 239-246.

ARDÖ, A. et al. (1994) "Improving resource discovery and retrieval on the Internet : the Nordic WAIS/World Wide Web Project : summary report". NORDINFO project, 1994.
http://www.lib.lu.se/W4/summary.html.

ARDÖ, A. et al. (2000) "Browsing engineering resources on the Web : a general knowledge organization scheme (Dewey) vs. a special scheme (EI)", Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 385-390.

ASHTON, S.; ROBERTSON, S. (2000) "Re-casting our net : broadening information access at the National Maritime Museum", Museums and the Web 2000, April 16-19, 2000, Minneapolis, Minnesota, USA.
http://www.archimuse.com/mw2000/papers/robertson/robertson.html.

BATLEY, S. Classification in theory and practice. Oxford [etc.]: Chandos Publishing, 2005.

BROUGHTON, V. (2002) "Facet analytical theory as a basis for a knowledge organization tool in a subject portal", Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries: proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain). Eds. María J. López-Huertas with the assistance of Francisco J. Muñoz-Fernández. Würzburg: Ergon Verlag, 2002. (Advances in Knowledge Organization; Vol 8). 135-142

BROUGHTON, V.; LANE, H. (2000) "Classification schemes revisited : application to web indexing and searching", Internet searching and indexing : the subject approach. Edited by A. R. Thomas , J. R. Shearer, New York etc. : The Haworth Information Press, 2000, 143-155.

CALESS, T. W.; KIRK, D. B. (1967) "An application of UDC to machine searching", Journal of Documentation, 23 (3) 1967, 208-215

CURRIER, S.; WAKE, S. (2001) "Negotiating subject access : resource discovery on the Web", Library & Information Briefings, 97 2001, 2-14.

DEMPSEY, L. (2000) "The subject gateway : experiences and issues based on the emergence of the Resource Discovery Network", Online Information Review, 24 (1) 2000, 8-23.

DEMPSEY, L. et al. (2004) “Metadata switch: thinking about some metadata management and knowledge organization issues in the changing research and learning landscape.” Forthcoming in LITA guide to e-scholarship (working title), ed. Debra Shapiro.
http://www.oclc.org/research/publications/archive/2004/ dempsey-mslitaguide.pdf

DESIRE Information Gateways Handbook
http://www.carnet.hr/CUC/cuc2000/handbook/welcome.html

DEVADASON, F. J. (2003) "Faceted indexing application for organizing accessing Internet resources", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 147-159.

ELLIS, D.; VASCONCELOS, A. (1999) "Ranganathan and the Net : using facet analysis to search and organise the World Wide Web", *Aslib Proceedings*, 51 (1) 1999, 3-10.

ELLIS, D.; VASCONCELOS, A. (2000) "The relevance of facet analysis for World Wide Web subject organization and searching", *Internet searching and indexing : the subject approach*. Edited by A. R. Thomas, J. R. Shearer. New York etc. : The Haworth Information Press, 2000, 97-114.

HIOM, D. (1998) "The Social Science Information Gateway: putting theory into practice", Information Research, 4 (3) 1998.

HIOM, D. (1998a) "Mapping classification schemes", SOSIG, ILRT, University of Bristol, 1998. http://www.sosig.ac.uk/desire/class/mapping.html.

HODGE, G. (2000) *Systems of knowledge organization for digital libraries : beyond traditional authority files*. Council of Library and Information Resources (CLIR). http://www.clir.org/pubs/reports/pub91/contents.html.

HUXLEY, L. (2000) "New features of The Social Science Information Gateway", *Proceedings of the JANET User Support Workshop, 4th-6th July 2000*. Cambridge : University of Cambridge, 2000, 99-107. http://www.sosig.ac.uk/about_us/docs/sosig_paper.pdf.

HUXLEY, L. & JOYCE, A. (2004) "A social science gateway in a shifting digital world: shaping SOSIG for users' needs of the future", Online Information Review 28 (5) 2004, pp.328 – 337. http://www.ilrt.bris.ac.uk/publications/conf/DigLib2003/abstract.pdf

KIRRIEMUIR, J. (1999) "A brief survey of Quality Resource Discovery Systems : commissioned by the JISC-funded Resource Discovery Network Centre : final report". JISC RDN, September 1999. http://www.rdn.ac.uk/publications/studies/survey/.

KOCH, T. (2000) "Quality-controlled subject gateways : definitions, typologies, empirical overview", Online Information Review, 241, 2000. http://www.lub.lu.se/~traugott/OIR-SBIG.txt.

KOCH, T. (2000a) "Dublin Core Metadata Initiative in transition : report", DC 8, Ottawa, 3-6 October 2000. http://www.lub.lu.se/metadata/dc8-report.html.

KOCH, T. (2006) "Electronic thesis and dissertation services: semantic interoperability, subject access, multilinguality", E-Thesis Workshop, Amsterdam 2006-01-19/20. http://www.ukoln.ac.uk/ukoln/staff/t.koch/publ/e-thesis-200601.html

KOCH, T.; GOLUB, K.; ARDÖ, A. (2006) "Users browsing behaviour in a DDC-based Web service: a Log Analysis". *Cataloging & Classification Quarterly*, 42 3-4 (2006).

KOCH, T; NEUROTH, H.; DAY, M. (2003) "Renardus : cross-browsing European subject gateways via a common classification system (DDC)", Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 25-33.

KOCH, T.; VIZINE-GOETZ, D. (1998) "Automatic classification and content navigation support for web services : DESIRE II cooperates with OCLC". OCLC, Dublin, OH ; NetLab, Lund University Library Development Department, Sweden. 1998. http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003489.

KOCH, T. et al. (1997) "The role of classification schemes in Internet resource description and discovery : project final report". DESIRE project, 1997. http://www.lub.lu.se/desire/radar/reports/D3.2.3/.

LA BARRE, K. (2004) "Adventures in faceted classification : a brave new world or a world of confusion", Knowledge Organization and the Global Information Society : proceedings of the 8th ISKO Conference, London 13-16 July 2004. Edited by I. C. McIlwaine. Würzburg : Ergon Verlag, 2004. (Advances in knowledge organization 9), 79-84.

LAFFORD, A.; STONE, O. (1997) "NISS - National Information Services and Systems", Ariadne, 10 1997. http://www.ariadne.ac.uk/issue10/niss/.

LONSDALE, M. (2003) Global gateways: a guide to online knowledge networks. Australian Council for Educational Research, 2003. http://www.educationau.edu.au/jahia/webdav/site/myjahiasite/users/root/public/papers/global_gateways_v3.pdf

MCGUINNESS, D. L. (2002) "Ontologies come of age", *Spinning the semantic web : bringing the World Wide Web to its full potential*. Edited by D. Fensel et al. Cambridge, MA : MIT Press, 2002, 171-194. http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm.

MERHOLZ, P. (2001) "Innovation in classification", [personal blog archive] 2001. http://peterme.com/archives/00000063.html.

MÖLLER, G. et al. (1999) "Navigating the Web with GERHARD", presented at The Third European Conference on Research and Advanced Technology for Digital Libraries, Paris, France, 22-24 September 1999. http://www.ifi.unizh.ch/CL/carstens/ecdlarticle.pdf

MÖLLER, G. et al. (1999a) "Automatic classification of the World Wide Web using Universal Decimal Classification", 23rd International Online Information Meeting, London, 7-9 December 1999 : proceedings. Edited by B. McKenna. Oxford : Learned Information Europe, 1999, 231-237. http://www-is.offis.uni-oldenburg.de/~gemoe/veroeffentlichungen/online_99/article.pdf

NEUROTH, H.; KOCH, T. (2001) "Cross-browsing and cross-searching in a distributed network of subject gateways : architecture, data model and classification" ELAG 2001, Integrating Heterogeneous Resources, 25 Library Systems Seminar, Prague, 6-8 June 2001. http://www.stk.cz/elag2001/Papers/HeikeNeuroth/HeikeNeuroth.ppt.

PIKA, J. (2002) "Anwendung der UDK in NEBIS in der Schweiz : ein Ausblick", [talk given at the seminar] Jahrestagung der Gesellschaft für Klassifikation (GfKl), Universität Mannheim, 23. Juli 2002. http://www.ethbib.ethz.ch/pub/nebisUDK.pdf.

RAMSHIRISH, R. (2004) " Faceted classification & its advantages in current perspective" [presentation]. http://drtc.isibang.ac.in/~ram/classification.pdf.

SERRES, A. (2004) "Recherche d'information sur Internet : où en sommes-nous, où allons-nous?". Paris : CNDP, Savoirs CDI, June 2004. http://savoirscdi.cndp.fr/culturepro/actualisation/Serres/Serres.htm

SLAVIC, A. (2004) "UDC translations : a 2004 survey report and bibliography", *Extensions & Corrections to the UDC*, Vol. 26, pp. 58-80. Also available at http://dlist.sir.arizona.edu/649/.

SLAVIC, A. (2006) "Use of the Universal Decimal Classification: a worldwide survey " accepted for publication in *Journal of Documentation* (2006). Preprint available at http://dlist.sir.arizona.edu/

SOERGEL, D. (1999) "The rise of ontologies or the reinvention of classification", *Journal of the American Society for Information Science*, Vol. 50 No.12, pp. 1119-1120.

SOSIG Guidelines for assigning keywords : section editors workshop. (2004) SOSIG, ILRT, University of Bristol, June 2004 [internal document].

SPARCK JONES, K. (2005) "Revisiting classification for retrieval", Journal of Documentation, 61 5 (2005), 598-601.

STEVENS, A. (2005) "UDC licences - their value and purpose", *Extensions & Corrections to the UDC*, 27 (2005), 127-130.

STOJANOVSKI, J. (2006) "Wissenschaftliche Informationen aus einer Hand - Erfahrungen aus Kroatien", GMS Medizin - Bibliothek - Information, 6 1 (2006), Doc08. http://www.egms.de/pdf/journals/mbi/2006-6/mbi000026.pdf.

STOKLASOVÁ, B. (2003) "Short survey of subject gateways (SG) activity", IFLA [paper presented at] 69th IFLA General Conference and Council, World Library and Information Congress, 1-9 August 2003, Berlin.
http://www.ifla.org/IV/ifla69/papers/152e-Stoklasova.pdf.

STOKLASOVÁ, B.; BALIKOVÁ, M.; CELBOVÁ, L. (2003) "The relationship between subject gateways and national bibliographies in international context", IFLA [paper presented at] 69th IFLA General Conference and Council, World Library and Information Congress, 1-9 August 2003, Berlin.
http://www.ifla.org/IV/ifla69/papers/054e-Stoklasova_Balikova_Celbova.pdf.

TZITZIKAS et al. (2002) "Extended faceted taxonomies for Web catalogs", *ERCIM News*, 51 (2002). http://www.ercim.org/publication/Ercim_News/enw51/tzitzikas.html.

VAN DIJK, P. (2003) "eXchangeable Faceted Metadata Language - XFML Core . [Last changed in 2003]. http://www.xfml.org/spec/1.0.html.

VIZINE-GOETZ, D. (1998) "OCLC investigates using classification tools to organize Internet data", Visualizing subject access for 21st century information resources : [proceedings of 34th Annual Clinic on Library Applications of Data Processing, Champaign, 1997]. Edited by P. Atherton Cochrane, E. H. Johnson. Champaign, IL : Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 1998, 93-105.

WÄTJEN, H.-J. (1998) "GERHARD - automatisches Sammeln, Klassifizieren und Indexieren von wissenschaftlich relevanten Informationsressourcen im deutschen World Wide Web. B.I.T. online, (1998), 279-290.