

The Functionality and Flexibility of Traditional Classification Schemes Applied to a Content Management System (CMS): *Facets, DDC, JITA**

Susanna Dal Porto* and Andrew Marchitelli**

*The University of Pisa, Department VI, Lungarno Pacinotti, 43/44 56126 Pisa, Italy,
susanna.dalporto@jus.unipi.it

** CILEA Team AePIC, <http://www.cilea.it>, <http://www.aepic.it>, Piazza G. Marconi, 10,
I-00100 Rome, Italy, marchitelli@cilea.it

Susanna Dal Porto was born in 1976 in Lucca (Tuscany). She lives in Pisa. She is working for the University of Pisa Library System as a manager of Librarian Dept. at Institutions Business and Market "A. Cerrai" Library. Nevertheless she is still involved in studying and researching activities about databases, e-journals and fundraising activity. She also is following information retrieval and classification. She worked in a big Tuscany public library, focusing on the Regional Multicultural Pole creation. She collaborated at the research project CiBit: Inter-university Centre for Italian Electronic Library, in a research unit headed by Prof. Alberto Petrucciani (University of Pisa). She graduated in 2003 in Library Science, majoring in the study of Historical Bibliography, in Pisa University: marks obtained 110/110 cum laude.



Andrea Marchitelli was born in 1975 and has been working in libraries since 1999. Actually he is working at CILEA (a non-profit consortium of Italian universities) since June 2006, and he is part of AePIC team, that provides innovative solutions for electronic publishing and digital libraries, employing open source and OAI-PMH compliant software. He is interested also in information retrieval and classification and collaborates with AIB (Italian Library Association) and AIDA (Italian Association for Advanced Documentation). His personal home page is: <http://www.biblioatipici.it/andrea>.



* An earlier version of this paper was presented at the 2nd ISKO Italy-UniMIB meeting, Milan, June 9 2006 <<http://www.iskoi.org/doc/milano06.htm>>.

Dal Porto, Susanna and Andrea Marchitelli. **The functionality and flexibility of traditional classification schemes applied to a Content Management System (CMS): Facets, DDC, JITA.** *Knowledge Organization*, 33(1) 35-44. 12 refs.

Abstract: Different classification schemes may be used for web indexing. The authors analyze three weblogs (Biblioatipici, Letture and Andrea Marchitelli's blog) to demonstrate that different contents may be classified using the appropriate scheme. Biblioatipici is a weblog about temporary workers in libraries and Italian documentation centres indexed with a faceted scheme, home made by authors. Letture, a diary about reading and books, is indexed by DDC. Marchitelli's blog about digital libraries and open access, is indexed by JITA, the scheme used for indexing e-prints in some different open archives. The three applications are presented starting from the most complex scheme (i.e. the faceted one) to the simple but least functional (JITA), passing through a traditional bibliographic classification scheme, the *Dewey Decimal Classification*. The analysis demonstrates that different web contents can efficiently be classified with different schemes. In particular, with the facets one (Biblioatipici) the indexer is able to generate the needed classes in the indexing phase. Moreover with a small number of facets and foci one

can obtain an exponential number of classes. Finally, one of the advantages for the user is that the faceted scheme allows multiple accesses on the basis of different information requirements, in addition to being coherent and intuitive. *DDC* scheme can be more suitable than the first web pages or blogs performing a sort of digital library. In this case in fact the *DDC* scheme constitutes a consolidated classification standard, widespread in a huge quantity of libraries and the use of a different scheme might get confusion. *JITA* scheme, finally, can find a huge applicability in web pages (or blogs) in LIS field showing several advantages: it's very simple and essentially pragmatic, intuitive and coherent. It's a conservative scheme because it is structurally closed, in fact it does not allow the classifier to insert new LCS's categories. If, from a certain point of view, this is a disadvantage, vice versa this feature is really useful because you get a complete matching between classes of all the web pages implemented with it.

1. Introduction

Library classification schemes (LCSs) have become increasingly available in electronic form and undergone many enhancements that make them attractive for Web knowledge organization. In fact, library professionals have been quite successful in applying library classification to Internet-based information services in a number of projects, both small and large. Yet, many opportunities remain for improving general knowledge organization tools and using them in new ways. The object of this project is to illustrate three different LCSs as applied to three different weblogs, highlighting their functionalities, characteristics and limits, and explaining why one scheme was chosen over another. The three weblogs taken into account have a layout that is similar to all blogs: that is, a series of posts can be seen on the home page, organised from the most recent to the oldest. These posts, after a fixed period of time on the home page, are transferred to an archiving page via an archiving mechanism. Each post has a publication date, a title and a "permalink," i.e. its own specific URL where it is always possible to trace the document. Moreover, a weblog offers various possibilities of searching for the posts. The most common are full text searching, and browsing by month of publication or by title. However it is a very simple and uniform structure, without any changes. What will change is the particular content and length of each article.

For example, one may observe the information architecture of Biblio(a)tipici <<http://www.biblioatipici.it/>>, which has a home page structured in four zones:

- the header
- the two central columns
- the footer

The header contains the identifying elements of the blog: the title and a short tagline. Moreover, all the elements that facilitate access to the posts, in particu-

lar the navigation bar, are concentrated in this area of maximum visibility. Beneath the header, the column to the left (the width of the column occupies two-thirds of the screen), contains the last seven posts in inverse chronological order. Over and above a permalink (a stable URL identifying the individual articles) each post has three links. The first is used to add comments through a form; the second for the printable version and the last is referring to facets used to classify posts (we shall come back to this later).

The right column contains some services linked to blog life: an easy search form, an access point through a month's calendar, the latest list issues for community growth, a recruitment examinations list and expiring job advertisements. Finally the link-of-the-week picked out from the directory and highlighted for a few days. The footer, in addition to the link to Andrea Marchitelli's blog (the creator of Biblio(a)tipici), indicates the reference standards (XHTML and CSS 2) and the group's adhesion to CABI. (The Biblioteca Nazionale Marciana di Venezia launched the CABI project, beginning with the Campaign for the Accessibility of the Libraries <<http://marciana.venezia.sbn.it/CABI/>>, which seeks to involve all culture "providers" and, naturally, the entire Public Administration, to provide information on and arouse interest in the problem of accessibility.)

These three weblogs are realised with a CMS (Content Management System). Right from the outset, usability and accessibility were targets to be attained. The use of a CMS allowed people to become involved in the editing, even though they did not have an in-depth knowledge of HTML, yet still maintaining a good standard of page design and compilation. In fact a CMS fully manages all the complex site contents, starting from the contents collection from external sources, dbase management, going down to storage process and final publication. Publication is a central aspect of a CMS: it is based on the initial creation of templates, i.e. models that define the texts layout, getting the content from the dbase.

The CMSs allow creation and management of contents separately from their publication, through pre-defined or customized templates. This is a great opportunity for content managers because they become completely autonomous. The blogs web pages are in XHTML code, with a full separation between content and graphic, using two different style sheets, one of which is used for the construction of a printable version of the individual posts. The code is periodically verified with the W3C validator, accessible at <<http://validator.w3.org>>. During the several site restylings, accessibility was also tested with an appropriate software able to simulate the graphic output of the different browsers. *MovableType* <<http://www.sixapart.com/movabletype/>>, version 3.2, is the CMS used for the three blogs. It comes with an intuitive web interface and many new functionalities, such as the trackback mechanism. This allows to keep track of citation swaps between blogs: if the post of one blog cites the post of another one, the latter is advised of the citation, (cf. <http://en.wikipedia.org/wiki/Trackback>). We use this tool to link other blogs, or for “internal” citation, i.e. one that cross-references two posts of the same blog. An example of this is at: <http://www.biblioatipici.it/2005/11/nuova_lista_di.shtml>.

Finally, a fundamental characteristic of *MovableType* (like other CMSs) is that it provides an appropriate module for categories management (the word “category” is used here in the same meaning as that in *MT System*) on n levels, i.e. hierarchically organised, with the possibility of selecting more than one category for each entry (or post in our case). This functionality was already provided in the previous versions of this software, whereas the possibility of easily creating an infinite number of sub-categories was only implemented with the 3.1 version. At that stage the drop-down menu used to assign one or more category to each entry was no longer capable of managing these powerful options, so in the 3.2 version the tree menu was chosen.

The CMSs functionalities, described above, are generally used in blogs to assign one or more categories or sub-categories to each post, depending on its content, in such a way as to gather posts by subject-matter and to offer a semantic access key to the posts archive. In the blogs we are going to examine, it was decided to apply three LCSs developed and used in the world of libraries to this hierarchical categorical structure, typical of all the CMSs.

The experimental use of LCSs on the web has become gradually more frequent, bearing in mind that

the semantic classification theory has a long history within the library and information science community (Mai, 2004). Moreover, the classifications developed in librarianship’s long history, even though they have renounced pure scientific method, show a greater level of coherence. A coherent classification scheme respects the rule of the single fundamentum divisionis, i.e. every node of the classification produces mutually exclusive LCS’s categories, as compared to the home-grown schemes of the Web (Gnoli et al., 2006). In addition the structure of the LCSs that we will see, marry very well with any system based on a database. The choice of applying three different LCSs to the same content structure allows one to highlight the advantages of using one scheme rather than another.

The three applications are presented starting from the most complex scheme, which, however, has the greatest potential (i.e. the scheme based on the principle of facets), to the simplest but least functional (i.e. the *JITA*), passing through a traditional bibliographic classification scheme: the *Dewey Decimal Classification*.

2. Faceted Classification Scheme

Biblioatipici, <<http://www.biblioatipici.it/>> is a weblog dedicated to the world of temporary workers in libraries and Italian documentation centres. It is maintained by librarians for other librarians and aspiring librarians, for the purpose of sharing knowledge on common themes pertaining to work and the profession. It came into being in April 2004 and it currently has about three hundred access “hits” per day. *Biblioatipici* is a collaborative weblog. It is not easy to classify it because it is a review weblog, containing announcements, a lot of comments, a few stories, and project reports. There are presently about five hundred posts on the blog, a number that is becoming substantial, requiring retrieval techniques. A “search” function is available on the home page that searches in the full text of the posts. In addition, as can be viewed from the archive page at <<http://www.biblioatipici.it/archives.shtml>>, the posts can be scrolled by:

- Popularity (one finds the twenty most recent posts with their popularity index, linked to the number of comments received);
- Category;
- Month of publication; and,
- Alphabetical list by title, of all the posts

1.1 FCS Categories

Scrolling by category is definitely the most flexible and functional search mechanism. A project, started in the summer of 2005 and still under fine tuning, is linked to the application of a LCS to the CMS's categories, based on facet analysis (Vickery, 1960). The choice of using this type of LCS, instead of a hierarchical-enumerative one, is motivated by different reasons. In the first instance, specialised contents are indexed and the facets work particularly well for a set of homogenous objects. Secondly, *Biblioatipici* is a collaborative blog and the faceted scheme allows whoever is carrying out the indexing to provide multiple access to the various posts. This multidimensional subject approach satisfies mental models and access strategies that vary substantially amongst themselves. Finally, this type of blog is fated to grow and to change continuously, including possible revision or integration of initial classification criteria. A faceted scheme is the best option because it is flexible and can be graduated.

In applying this scheme, the following methodology was used (Gnoli, 2004): in the first phase, at the level of the conceptual plan, the semantic content of each post was broken down into simple parts (*isolates*). Next, repeating the same analysis for more than one post, it was seen that all the isolates tended to become grouped into a fixed number of similar *facets*, which were then identified through inductive means. In this manner, it was easy to apply a Faceted Classification Scheme (FCS) to the *Biblioatipici* blog, because the blog already had several posts when we started. Therefore the categories and the facets emerged from the reading-matter itself, from the bottom up, from the act of classifying. Conversely, the application of a hierarchical-enumerative classification scheme like the *DDC* would have been possible even with only one post in archive (the classification comes from the top).

Seven fundamental FCS categories were identified in *Biblioatipici*: numerically more than those of Ranganathan and fewer than the thirteen worked out by the Classification Research Group (CRG). Ranganathan's fundamental categories are expressed by the formula PMEST (Personality, Matter, Energy, Space, Time) and were reviewed in the light of the CRG's studies (Foskett, 1996). The CRG work, in fact, was concentrated on identifying more specific categories compared to those of Ranganathan, just to facilitate the exercise of analysing the categories. On the other hand, the full range of categories in the

CRG's formulation is probably only used in technological subjects, or in any event in particularly complex topics (Broughton 2004). In this blog the scheme was simplified, because we are dealing with a site that collects information, materials, and announcements that are deliberately selected.

The following standard FCS categories were selected (see schedule at <<http://www.biblioatipici.it/archives.shtml#cat>>):

- **Entities:** the first category in the CRG scheme equates to the most simple use (because it makes up only a part) of the "Personality" category of Ranganathan and relates to the main interest or the object of any discipline. In this blog it is used for "Libraries," "Librarians," "*Biblioatipici*'s friends," and "*Biblioatipici*," without subclasses. The material category from the CRG scheme, equivalent to the category 'Matter' in the Colon Classification, was not included in this scheme because it has little significance in this context, where we talk about immaterial concepts rather than objects.
- **Activities:** corresponding to Ranganathan's Energy has two equivalents in the CRG scheme: 'Processes' and 'Operations'. The CRG scheme in fact distinguishes between processes that are intrinsic, spontaneous actions, and operations, i.e. actions caused by an external agent. In the FCS the distinction is not maintained because only operations are included. One can find processes particularly in physical and natural sciences.
- **Products:** present in the CRG scheme, products are the outcome or results of processes or operations on entities; usually they consist of physical products. Even though this category is to a large extent limited to the area of technology, and is generally absent in the humanistic and social disciplines, it finds a place in this blog because there are a lot of objects that are the result of activity.
- **Instruments** (see below).
- **Agents:** the means by which the operations are carried out. Agents can be differentiated into "Persons" and "Instruments," and, at a complex level, can be represented by institutions. The two categories of agents can occur together; e.g. in this scheme, the "Temporary workers" (M1, Person Agent) can confront their experiences using a blog (L1, Instrument Agent). The "Person" and "Instrument" agents are separated, unlike in the CRG scheme that provides a single category.
- **Space:** any type of political, physiographic or spatial dimension (e.g.: the USA, mountainous, inte-

rior). In this blog this category is very useful because, for instance, it is possible to collect recruitment examinations for given geographic areas (southern Italy, northern Italy, etc.).

- **Time:** any historical, chronological or temporal characteristic (e.g.: medieval, permanent, nocturnal). This category was only slightly developed (it contains only one post), because this blog is still quite new, and because a chronological search mechanism is available through the *chronological archive*, where the posts are subdivided by month depending on publication date. Of course, this does not mean that a post issued in March relates to an examination in March, but it is still a good approximate search key.

Each facet contains the foci that have been identified. For example the category “Agent” has, as foci: “Trade union,” “Private Entities,” “Public Entities,” “Professional Associations,” and, “Atypicals”. It becomes evident that the foci are mutually exclusive, with respect to each other. In fact, there is no semantic overlapping. Inside certain facets subclasses were identified. For example: the facet “Professional Associations” has the subclasses: ‘IAML-Italia,’ ‘AIDA,’ and ‘AIB.’ In these instances the facet acts as a grouping class and thus does not contain posts or contains only a few because, in the classification of a post, one always tries to use the most specific class.

Each post was classified with one or more facets, depending on the complexity of the subject-matter. Look at this post, for example <http://www.biblioatipici.it/2006/05/concorso_alluni_18.shtml> for which it was possible to identify the following FCS categories: ‘Activities,’ ‘Products,’ ‘Agents,’ and ‘Space.’ Thus in the classification string we will have the following facets: *H11: Recruitment examinations that have not yet lapsed::J94: Open-ended contracts::M58: Universities::S51: Northern Italy.*

In the classification string of each post, a citation order must be respected, i.e. an order of precedence amongst the facets. This order respects Ranganathan's rule of “decreasing concreteness;” the last facets in the citation order are less significant for subject specification (Ranganathan 1967). The order of facets in this blog respects the standard citation order that places the fundamental FCS categories in the following sequence: ‘Entities,’ ‘Activities,’ ‘Products,’ ‘Instruments,’ ‘Agents,’ ‘Space,’ ‘Time.’ Take the following post as an example: <http://www.biblioatipici.it/2006/03/nuovo_disegno_d.shtml>. From a subject point of view, it can be broken down into four iso-

lates relating to the following FCS categories: ‘Entities,’ ‘Agents,’ and ‘Space.’ Consequently, in the classification string, the facet order will be the following: *B6: Librarians::M1: Atypical::M59: Local authorities::S58: the Islands.*

As a mid phase of scheme application, an expressive notation was identified, able to get an expressive order useful for browsing. In fact, for getting automatically standard citation order in the classification strings, a mixed notation of letters and numbers was used. This allows the hierarchical structuring of the classes and the application of a systematic sorting criterion. The FCS categories are thus identified with a letter and the first and second level foci with letters and numbers. Letters in ascending alphabetical order were assigned to the FCS categories (and consequently the facets), because the system allows in this case only ascending alphabetical order. In the post analysed above, the classification string respects the standard classification order because the category “Entities” that must precede the category “Agents” has the letter “B” as an equivalent notation that comes before the letter “M.”

For a faceted LCS it is necessary to provide a schedule in which an inversion principle is applied. This means that when all the archive post facets are listed, the generic entity precedes the specific. Thus in the schedule of the blog, <<http://www.biblioatipici.it/archives.shtml#cat>> the facets are listed or tabulated in reverse order compared to the citation order.

Unfortunately, the CMS allows inverse alphabetical order only via tags linked to the preparation of the archives and thus only in the inverted schedule, not in the classification string of the individual objects. This is so because the inverted schedule is a product of the archive of the blog and uses tags linked to the archive templates (see the online software manual, at <http://www.sixapart.com/movabletype/docs/3.2/a_template_tag_reference/subcategory/>), whereas the classification string, not being an archive product, is linked to the post and thus uses different types of tags, <http://www.sixapart.com/movabletype/docs/3.2/a_template_tag_reference/entry/>. With the archive tag there is the possibility of changing the sorting, choosing the inverted order (<MTSubCategories sort_order="descend">, instead with the tags of the posts this is not possible and one must of necessity adopt an ascending order, <http://www.sixapart.com/movabletype/docs/3.2/a_template_tag_reference/subcategory/#entry-6090>).

Mainly for this reason the notation is studied so that it is directed at the inside of the string. In this way the less concrete standard FCS categories (Time, Space) have the highest letters (T: Time; S: Space), rather than the opposite situation, which would be logical and recommended in faceted schemes. A useful example is represented by FATKS (Slavic 2002), developed by Aida Slavić under the supervision of Vanda Broughton at the University College of London. Aim of this project is the creation of a FCS for humanities, with expressive notation (the notation will represent the hierarchy of the scheme). If one clicks on the facet “islands” in the inverted schedule, for example, a link opens up and all the posts that have the facet “islands” in the classification string are displayed. The order of these posts is alphabetical by title whereas, strictly speaking, it should be an order that follows the inversion principle, i.e. a descending alphabetical sorting of the facets. First the post whose classification string commences with the facet “L,” then those with the facet “H,” and finally that with the facet “B.” This is a limitation imposed by this CMS. Another limitation of the system is that it allows searching for posts by category with the ‘browse’ function only (browsing the news by subject) and not with the search function as well. This means for example, that it is not possible to use two or more facets as search characteristics together.

Finally, another possibility offered by the organisation of the blog in FCS categories is that of generating specific RSS feeds by subject-matter, in addition to the general feeds. A feed is a list of news from a site in XML language; an RSS feed is made available to a supplier of information content and is ready to be used by others. To use feeds, an aggregator programme is required, i.e. software that can follow the updates of more than one RSS feed and can display it to the user. The user indicates directly to the aggregator the RSS feeds that he intends to follow. In *Biblioatipici* the feeds are updated at each new post and can also contain the comments relative to the posts). All the RSS feeds available can be reached starting from the grouping URL: <http://www.biblioatipici.it/2005/01/feed_rss_di_bib.shtml>.

One can say that the semantic homogeneity of the posts contained in this blog makes the choice of this classification particularly pertinent compared, for example to the *DDC*. In the classification schedule, in fact, it is not necessary to foresee all possible concept combinations as opposed to traditional enumerative classifications. It is sufficient to list the iso-

lates that can appear in each facet and the rules to combine them. There are many advantages to the faceted application, both for site managers and for the user. To give but a few examples, one can say that the indexer, with the use of facets is able to generate the necessary classes in the indexing phase. With a small number of facets and foci one can obtain an exponential number of classes. Finally, one of the advantages for the user is that the faceted scheme allows multiple accesses on the basis of different information requirements, in addition to being coherent and intuitive.

3. Dewey Decimal Classification

AtipicheLecture, on line at <<http://www.biblioatipici.it/lecture/>>, is a weblog developed in December 2005 and closely linked to the *Biblioatipici* blog. It contains the reading-matter of the *Biblioatipici* group, but would like to be open to the collaboration of persons outside this group. A type of “diary of collective reading-matter,” each post corresponds to the reporting of a book that can relate “atypical” themes to the working world. This can be in the LIS field, but it can also simply be Italian or foreign fiction, a sociology or history book or, in short it can relate to any discipline. At present the blog contains seventy posts and it is assuming an increasingly stable aspect, with typical library features: the description of the books is drawn up according to the *International Standard Bibliographic Description for Monographic Publications (ISBD(M))*, every post is classified with the *Dewey Decimal Classification (DDC)*, and there there is a link active for every post, which locates the document in the Italian Union Catalogue, <<http://opac.sbn.it/cgi-bin/IccuForm.pl?form=WebFrame>>.

When the blog was created, in the *Biblioatipici* mailing list there was a pleasant debate about setting up the site according to library criteria. There were a couple of thoughts. On the one hand, the feeling was to be less of a “librarian” and to personalise the posts with original stories, comments, quotations, cross-references from one book to another, in a very free and tainted context. On the other hand, the conviction was to have a weblog as a gym for shaping the site architecture according to library methods. The mailing lists archives of January 2006, where it’s possible to follow the debate, are at <http://lists.biblioatipici.it/mailman/private/ml/2006-January/thread.html>. In support of the first point of view there was also another line of reasoning, i.e., that if

the entry of posts were to become too complicated this would discourage many from contributing. The solution was that “the views are not mutually exclusive” - that is to say [in the words of Livia Castelli, a member of the Biblioatipici group]:

What’s more tainting than a catalogue getting the opportunity to fully allocate a book, instead of continuing to speak about it without perhaps having ever read it? Why should we view “personalisation” as something that excludes our trade tools and the organisation of knowledge, which we use daily? Why we shouldn’t hope these powerful and useful tools become widespread [...]? And more generally: why should our knowledge, intuitions and professional ideas be isolated in a watertight compartment when we become readers, bloggers, or anything else?

These are the characteristics of each post:

- Every library announcement is accompanied by the book cover and the related ISBD card;
- It follows the plot of the book, the quotation, or a “suggestion;”
- Thanks to the use of keywords it is possible to index the following bibliographic data: author, publisher, series. In this way one can immediately view all the titles in the blog relating to the same author, publisher and series;
- It is possible to locate the document in the SBN national union catalogue. In fact, by clicking on the link “Locate title in the SBN” set up for each post, one activates a query on the SBN Index that automatically sets the title of the post (= the title of the book) as the search key for the title. In some cases the result is affected by noise, in that it is a search function by title only and is not capable of being cross-referenced. A very clear example is at <<http://www.biblioatipici.it/lettere/2006/01/piattaforma.shtml>>, where, clicking on “Localizza titolo in SBN” the results include records about rigs, not only about the novel by Michel Houellebecq;
- By indicating the ISBN amongst the keywords in the following form ‘ISBN=XXXXXXXXXXXX’ (without spaces amongst the groups of figures) one can automatically activate a search to buy the book in certain selected on-line bookshops (BOL, IBS, DEAstore);
- One can leave a comment for every post; and,

- the “reader” (the person who inserted the post) is included amongst the indexed data and thus one can view all the books read by a reader.

2.1 The Dewey Decimal Classification

The choice of using the *DDC* to classify the posts arises from the features of the blog (Kepner, 2002). *Atipichelettore* is a blog of library announcements relating to the widest possible subject fields, thus each post corresponds to a book, with the virtual feeling of being in front of the shelves of a multi-discipline library. In this context the *DDC* constitutes a consolidated classification standard and the adoption of a different scheme might have risked confusion, because this is a blog managed by librarians for other librarians.

By assigning a classification string to each post (=book), it is possible to retrieve all the books of the same subject in the same class and those of similar subjects in the classes that follow, exactly like an open shelf. For example, look at the archive of class 853.914 (Italian fiction, 1945-) <http://www.biblioatipici.it/lettere/800/850/853914/>. In addition to the functionality that the *DDC* offers, if applied to a CMS it overcomes the physicality of a shelf. In fact, it is possible to browse from one class in the class immediately higher, i.e. less specific. For example, beginning with a post classified as 21st century Italian fiction, with a simple link it is possible to view all posts classified under Italian fiction of all periods (without specifying the time). It is also possible to browse the classified catalogue <<http://www.biblioatipici.it/lettere/cdd.shtml>>. In this catalogue one can view the Dewey classes used to classify the posts, represented by a notation, or a caption (the verbal equivalent of a notation), and the links to the posts. The notation of the *DDC* has substantially mnemonic features (Foskett, 1996): it is based on 10 numbers; the entire knowledge is intended as the unit, divided into 10 main classes, each of which has 10 divisions and each of these has ten sections and so forth, by way of successive expansions, indicated with numbers alone. The numbers are read as decimals, as if preceded by 0 and the decimal point: thus not as full numbers, but figure for figure (thus 813.54 will precede 813.6). One is clearly dealing with an expressive notation because the longer symbols correspond to the more specific classes.

One could say that this classified catalogue, as it was devised, with the assistance of the captions, constitutes a type of Dewey browser (Vizine-Goetz,

2006) in the sense that it allows users to use the *DDC* like a retrieval tool (even if users are not very skilful in *DDC*). From the schedule it is apparent that the *DDC*, being a hierarchical-enumerative classification, provides nearly all the possible concepts and thus, compared to the faceted classification, which is analytico-synthetic, expands more because it is added to indicate very specific subjects. To do this, the *DDC* uses many hierarchical levels, exploiting to the utmost the infinitely divisible characteristics of the CMS categories. On the other hand, the faceted classification uses in general few hierarchical levels, in fact in the *Biblioatipici* blog we found three in all.

In this blog, on the other hand, even if there are still only a few posts and the *DDC* is not expanded at the moment, we are already at the 4th hierarchical level. For example the class 000 in which we have 020 son of 000, one 025 son of 020, 025.3 son of 025 and so forth. The classified catalogue of the blog is suitably housed by the CMS inside a tree structure, in which there are mandatory hierarchies and paths. Browsing takes the form of a step by step path, from the main category to the one below and so forth. Notwithstanding the efficient adaptation of *DDC* for this type of blog, the Web applications that were made are somewhat unsatisfactory when compared to the functionality of the faceted scheme. A few of the disadvantages of adopting the *DDC* (Gnoli et al., 2006), as shown in this site, are:

- The presence of an increased number of hierarchical levels hinders the immediacy of the browsing. The number of options in a menu, if too many, influences their usability and leads to information overload;
- The scheme is not easily graduated: it adapts with difficulty to the addition of components or to partial changes in its structure; and,
- It is structurally closed and conservative; it does not allow the classifier to insert new LCS categories. Only the compiler can modify the classification, drawing up an official revision sheet or publishing a successive edition of the scheme.

4. *JITA* Scheme

Andrea Marchitelli's blog, at <http://www.biblioatipici.it/andrea/>, is a personal blog. It contains posts that fall within LIS: professional meetings announcements, latest news in the Italian and overseas library scene, publications of scientific in-

terest, as well as all the information tied to the professional activities of the author of the blog. Actually this blog does not contain many posts (106, at 2006-07-30) and does not have a particularly broad cast (200 visits per month).

The retrieval techniques enabled in this blog are: the "search" functionality scanning the full text of the posts; scrolling by month of publication; scrolling by category, and finally tags. Considering the purpose of this blog, the LCS adopted was the *JITA Scheme* – acronym of the names of the authors of the scheme, José Manuel Barrueco Cruz, Imma Subirats Coll, Thomas Krichel, und Antonella De Robbio (De Robbio & Subirats-Coll 2005). It is a scheme in the LIS field that was created to classify the documents of *E-LIS* (E-prints in Library and Information Science, <http://eprints.rclis.org>), the international open-access archive for library and information science. It is also used by the CNR Research Area Library of Bologna, <http://biblio-eprints.bo.cnr.it/view/subjects/>. The subjects are systematically listed in the schedule, at <http://www.biblioatipici.it/andrea/jita.shtml>, highlighting relationships. It is a very simple LCS, a combination and reworking of the *NewsAgentTopic Classification Scheme* (kept by Mike Keen at Aberystwyth, UK, until March 31, 1998) and the *RIS* classification scheme (*Review of Information Science*), now no longer in use, originally conceived by Dagobert Soergel (University of Maryland). It is hierarchical on two levels (the second level was only opened in 2005) and today contains over 120 items. The documents can be classified with one or more items.

The *JITA* scheme, <http://eprints.rclis.org/jita.html> is divided into 12 blocks (A-L) created around 3 thematic areas, not explicit:

- Theoretical and general: includes theoretical and general aspects of library and information science, and the use of information and information sociology;
- User, executive and managerial functions at an intermediate level (including socio-economic and legal questions): here one finds works directed at users, the diffusion of literacy skills and reading; libraries and information repositories; publishing and legal themes, including copyright and the management of the legal issues, management for projects and industry, profession and education.
- Objects, pragmatic themes and technical questions at a specific level: relates to sources, supports and information channels; information processing for

information services, the technical services in library, archives and museums, information technologies and "library science technology."

The notation is made up of the Latin alphabet, in upper case characters and makes the order underlying the LCS immediately evident, even to a non-expert user. The aim of this scheme is essentially pragmatic, i.e. to present, through a browsing mode, in display and retrieve, the documents subdivided into broad LCS's categories without going into the details (there are only two hierarchical levels). Even if it is a very simple scheme, it falls within the hierarchical-enumerative or mono-dimensional LCSs, because the relationships between the subjects in the schedule are hierarchical and because the scheme provides the enumeration of all the possible simple and compound subjects.

There are many disadvantages to this scheme. First it was devised to classify documents in *E-LIS*, thus it is not easily adapted outside this field. Secondly it is a conservative scheme because it is structurally closed, in fact it does not allow the classifier to insert new categories. Finally periodical updates by the editorial committee, such as are common for the *DDC*, are not foreseen. Nonetheless, given its simplicity and the subject sector represented, it finds efficient application in this blog, also taking into account that it is a sufficiently coherent scheme.

JITA scheme is fully applied here, without any modification or limitation, but only few classes of the scheme are used, at the moment (less than 50%). In this blog, the archive of every category of post is dynamically linked to its counterpart in *E-LIS*, eg., see at <<http://www.biblioatipici.it/andrea/i/id/index.shtml>>. In addition, the latest papers published in *E-LIS* are listed and can be linked, on the home page, thanks to the use of the RSS feeds, made available in the *E-LIS* archive, at <http://eprints.rclis.org/last.xml>. Finally, another *subject indexing* functionality is active in this blog, which allows one to identify for each post on the blog one or more tags. The table of used tags is at: <<http://www.biblioatipici.it/andrea/tags.shtml>>. The tags constitute the subject/topic of the post, expressed with keywords, thus in the more generic and less hierarchical sense compared to the classes. Through "social tagging" (Hammond et al., 2005) services, (like *Technorati* a search engine, <<http://www.technorati.com/>>), the posts of different blogs are aggregated using tags. Thus by clicking on a certain tag of the blog one launches a search in *Technorati*

and finds the blogs that cite the same tag and which thus have similar topics. All the tags identified are found on the same level: a hierarchy does not exist (tag-fathers and tag-sons); the only distinctive factor is their popularity: certain tags are used more frequently to describe an item, others more rarely.

5. Conclusion

We can finally say that among the LCSs here applied to the web, *FCS* is the one that has more advantages. We are dealing with a highly flexible scheme, fully adaptable for more different kinds of blogs or elaborated web pages. The indexer, with the use of facets, is able to generate the needed classes in the indexing phase. Moreover, with a small number of facets and foci one can obtain an exponential number of classes. Finally, one of the advantages for the user is that the faceted scheme allows multiple accesses on the basis of different information requirements, in addition to being coherent and intuitive.

In spite of these features that make *FCS* the best, in some blogs or web pages the *DDC* scheme can be more suitable. This is especially true for web pages or blogs performing as digital libraries. In this case, the *DDC* scheme constitutes a consolidated classification standard, which is widespread in a huge number of libraries. Last of all we cannot forget the disciplinary LCSs like the *JITA* scheme. This scheme can be applicable in web pages (or blogs) in the LIS field because of its simplicity. It is essentially pragmatic, intuitive and coherent. It is considered a conservative scheme because it is structurally closed, and does not allow the classifier to insert new categories. Although from a certain point of view this can be considered a disadvantage, in reality it provides a complete matching between classes of all the web pages implemented with it.

References

- Broughton, V. 2004. *Essential classification*. London: Facet Publications.
- De Robbio, A., Subirats-Coll, I. 2005. *E-LIS : an international open archive towards building open digital libraries*. *HEP libraries webzine* 11 <http://library.cern.ch/HEPLW/11/papers/1/>.
- Foskett, A.C. 1996. *The subject approach to information*, 5th ed. London: Library Association.
- Gnoli, C. 2004. *Classificazione a faccette*. Roma: AIB.
- Gnoli, C., Marino, V., Rosati, L. 2006. *Organizzare la conoscenza*. Milano: Tecniche nuove.

- Hammond, T., Hannay, T., Lund, B., Scott, J. 2005. Social bookmarking tools (I): a general review. *D-lib magazine* 11 (4) <http://www.dlib.org/dlib/april05/hammond/04hammond.html>.
- Kepner, L.T. 2002. Workable solutions with Dewey classification. *Technicalities* 22n4: 3, 9, 12-13.
- Mai, J.E. 2004. Classification of the web: challenges and inquiries. *Knowledge organization* 31: 92-97.
- Ranganathan, S.R., 1967. *Prolegomena to library classification*. 3. ed. - Bangalore : Sarada Ranganathan endowment for library science, 1967.

- Slavic, A. 2002. *Facet analytical theory in managing knowledge structure for humanities*. University College London <http://www.ucl.ac.uk/fatks/>.
- Vickery, B.C. 1960. *Faceted classification: a guide to the construction and use of special schemes*. London: ASLIB.
- Vizine-Goetz, D., Hickey, T. 2006. Getting visual with DeweyBrowser. *Nextspace* n. 1: 18-19.

Scheme	<i>Biblioatipici</i> schema	<i>DDC</i>	<i>JITA</i>
Name of the blog	<i>Biblioatipici</i>	Letture	Andrea Marchitelli's blog
Number of post	495	85	112
Home made	YES	NO	NO
	Faceted	Hierarchical	Hierarchical
Number of classes / FOCI	61	67	135 (*)
URL of the scheme	http://www.biblioatipici.it/archives.shtml#cat	http://www.biblioatipici.it/lettture/cdd.shtml	http://www.biblioatipici.it/andrea/jita.shtml
Exemplum of classification string	<i>H11: Recruitment examinations that have not yet lapsed::J94: Open-ended contracts::M58: Universities::S51: Northern Italy.</i>	<i>346.0482 (Diritto della proprietà. Diritto d'autore - Copyright)</i>	<i>LJ (Software)</i>
Also used by	- - -	BUBL Information service < http://bubl.ac.uk/ > Canadian information by subject < http://www.collectionscanada.ca/caninfo/esub.htm >	E-LIS < http://eprints.rclis.org/view/subjects/ >

(*) Full scheme classes number. We use less than 50 classes, at the moment.