ISKO. CAPÍTULO ESPAÑOL. CONGRESO (7º: 2005: BARCELONA) 549-556

FACETED CLASSIFICATION FOR COMMUNITY SERVICES USING CRG STANDARD CATEGORIES

Luca ROSATI,[1] luca@lucarosati.it
Claudio GNOLI,[2] gnoli@aib.it
Maria Elisabetta LAI,[3] mariaelisabetta.lai@csi.it

1. Università per stranieri di Perugia (Perugia, Italy)
2. Università di Pavia (Pavia, Italy)
3. CSI Piemonte (Torino, Italy)

Abstract

Faceted classification is increasingly considered as a tool in non-traditional knowledge organization contexts, such as knowledge management and information architecture. These new applications, however, are rarely based on the full theory of facet analysis as developed by Ranganathan and the CRG (Classification Research Group). In this paper we report the creation of a faceted scheme to organize Web resources concerning regional community services, including remote management of cases (e-government). Community services are an important domain of application for knowledge organization systems and for usability models, as their target users may come from any social category and have any degree of computer skills. Determination of the facets in the scheme was based on a corpus of subjects, according to the methods and the standard categories recommended by the CRG. The Italian enumerative scheme of «life events» for e-government was incorporated as the Kinds category. Verbal headings were integrated with a thesaurus. An expressive notation was created, suitable for use in a digital database. The scheme was stored into an open-source content management system (*Drupal*), and an interface was designed for browsing and searching it in the Web environment. Usability issues are discussed, such as which part of the scheme to show in the main form, whether and how to show the whole underlying structure, notation, etc. Faceted classification meets the requirements of usability in that it does not force people to learn and navigate a single hierarchical tree; on the other hand it is necessary to get used t the classification structure, as the hierarchical model still dominates most computer interfaces.

Keywords

Faceted classification, search interfaces, community services, e-government, usability, Classification Research Group

1 Introduction

In recent years, faceted classification (Vickery 1960; Gnoli 2004) is frequently being mentioned as a tool for knowledge management, information architecture, and the organization of Internet resources: websites such as *Flamenco*,[4] *InformeDesign*,[5] and *Wine.com*[6] have become popular examples (Rosati 2003). Peter van Dijck has created a markup language for faceted data, XFML.[7] A *Faceted Classification* discussion mailing list[8] has also been active and viable since December 2002, with contributions mostly from knowledge managers and webmasters, but also from LIS students.

In all these contexts, however, the term «faceted classification» is used mostly in a loose sense, e.g. referring to the availability of alternative search fields for different features of the recorded items. Although the name of S.R. Ranganathan is sometimes mentioned, unfortunately the LIS origins of faceted classification and its full techniques, such as citation order, inverted schedules, notation, etc., are poorly known among computer people; reference to original LIS works is rare, as they are not easily available without a bibliographic search outside the usual Internet sources. This situation is producing a remarkable amount of confusion in the use of faceted classification concepts, as was shown by an examination of the archives of the aforementioned discussion group mentioned above (La Barre 2004).

These problems do not imply that a good integration between LIS theory and digital applications cannot be attained. Outstanding examples of this are projects involving both library and computer people, such as Hibrowse, FACET, and FATKS. Hibrowse (Pollitt 1994) applied facet analysis to the search and display of bibliographic records classified with the Dewey Decimal Classification (though not a truly faceted scheme), and more recently evolved into a search software produced by View-Based Systems.[9] *FACET*[10] (Binding 2004) exploits the faceted structure of the *Art and Architecture Thesaurus* to guide users in formulating their queries. *FATKS*[11] (Broughton 2002) aims at building a faceted classification to be used for indexing digital resources in the humanities. Interestingly, all these projects come from the United Kingdom, where most faceted classification theory has been developed, and the last one also involves members of the historical CRG (Classification Research Group).

An important field for application of knowledge organization systems is the public administration. As the use of the Internet spreads, public administration websites are expected to include not only information

4. <bailando.sims.berkeley.edu/flamenco>
5. <www.informedesign.umn.edu>
6. <www.wine.com>
7. <xfml.org>
8. <www.poorbuthappy.com/fcd>
9. <www.view-based-systems.com>
10. <www.comp.glam.ac.uk/~FACET>
11. <www.ucl.ac.uk/fatks>

on government departments, but also forms to be filled in order to begin a case or even to formalise it online (*e-government*). Since these services are to be used by all the community, accessibility and usability issues play a particularly important role in them: interfaces must be as understandable and easy as possible for many kinds of people, having different skills and needs.

Therefore, we believe that e-government is a challenging opportunity for applying the unexploited potential of knowledge organization tools such as faceted classification. The sophisticated techniques available from library science should be fully employed, but in the meantime their products should be easy to understand and use for everyone. The reputation of faceted classification of being a complex technique for specialists only could be confuted by paying attention to the development of usable interfaces, without losing its structural richness.

Some knowledge organization systems have been produced already for application in e-government, like the *UK Government Category List* (Dextre-Clarke 2002). The Italian government has published a list of «life event» classes to be used for indexing community information (Italia 2003). This is a simple hierarchical classification trying to represent the basic needs of users (birth, education, job, health, etc.); however, as it is, it cannot be used for searching in more flexible ways.

2 Construction of the scheme

CSI-Piemonte, Consortium for Information Systems, was founded with the aim of promoting the modernization of local administration in Piedmont (North-Western Italy) by using the most advanced information and IT-based tools to create information services and system. In 2004 CSI Piemonte started a project for collecting links to community services available on local government websites, including real e-government services, request forms to be printed and completed, and simple information pages. A web directory should gather all the links, and people would be encouraged to start from it as a gateway to available information and services on the matters and places they are interested in, and then go to the local websites and use them.

To organize the links to those resources, it was decided to create a faceted scheme. A working team was formed including the authors of the present paper, in order to gather skills in web interfaces, information architecture, and facet analysis and to apply all them to the purposes of the project. The team worked through both e-mail and meetings in Milan and Turin in order to produce the new scheme.

As the main structuring principle we chose the standard categories for faceted classification formulated by the CRG (Vickery 1960), namely Objects, Kinds, Parts, Properties, Materials, Processes, Operations, Products, Byproducts, Patients, Agents, Space, Time. An alternative would have been Ranganathan's «PMEST» categories, which are structurally equivalent, albeit less analytical. A first corpus of subjects, listed in na-

tional official documents concerning e-government or occurring in local websites was analyzed in order to recognize their facets and a first list of classes within each facet (*foci*). According to facet analysis theory, subjects should tend to cluster into facets, and each facet should be assigned to one of the standard categories.

We found eight facets that appeared to be suitable for the classification of community resources. They are reported below, each one with the category to which it has been assigned:

services and cases [Objects]
life aspects [Kinds]
access mode [Properties]
steps [Operations]
citizens and companies [Patients]
public institutions [Agents]
geographic departments [Space]
dates [Time]

The main question in stating them was which facet to take as the Objects one, as all the other facets would have depended on it. We considered taking life aspects as the Objects so to follow the national recommendation (we did not call them «events» to avoid confusion with their categorical function). However, we found that in this case it would have been very problematic to establish the appropriate syntactic relations between each facet and such Objects. Therefore, services and cases were chosen instead as the reference Objects of this domain, while life aspects were treated as a Kinds facet: e.g. a birth certificate is a kind of paper, birth being a life aspect. In this way the official enumerative scheme of life events was incorporated into the new faceted scheme.

Subclasses have were then listed within each facet on the basis of the literary warrant of the considered corpus of subjects. A helpful sequence (RANGANATHAN 1967) was established within each facet according to criteria of logic, of temporal sequence in life events or in case stages to be followed, of geographical position, of hierarchy between institutions, etc.

Finally, each class was assigned its notation. In our digital context notation is intended to be a tool for appropriately managing the hierarchy of classes (*chain*), the sequence of facets (*citation order*), and the place of classes within facets (*array*). Hence we chose to use an expressive notation, which is suitable for information processing in digital environment (SLAVIÆ 2004). We considered adopting as facet indicators the same letters as *FATKS* (B = Places, C = Times, K = Operations, etc.), but as we have less than 10 facets, we eventually decided instead to use numerals as facet indicators and letters as foci.

To form class numbers, notation from the occurring facets and foci can be combined, according to the inversion principle, or in other words facets coming last in the schedule will be cited first. As an example, the subject «refund request for medical expenses by a parent» would be represented by r8v6b5pi, namely «Service/case: refund, Life aspect: health, Step: re-

quest, Person/company: parents». When many subjects are listed, e.g. in a browsable menu, notation will automatically produce a helpful sequence between them according to the principles of faceted classification.

Verbal headings are intended to be searchable by users. To this purpose, they have been integrated with a thesaurus of synonyms for both facets and foci. These have been chosen using two methods:

1. by screening log files of the internal search engines in the websites of Comune di Torino[12] and Sistema Piemonte,[13] in order to ascertain the most popular search terms;
2. by selecting them from scratch, using digital and printed dictionaries: the CD-ROM *Dizionario italiano Sabatini-Coletti*, and the *Lessico di frequenza dell'italiano parlato* edited by Tullio De Mauro. The latter appeared to be especially useful, as it deals with the use of spoken language. Indeed, we considered that the terms used to search community services will often be the same used in everyday speaking.

3 Structure of the database

For the application development, we opted, together with CSI Piemonte, for an ad-hoc database, following the recommendations of the Flamenco project (Flamenco 2004). In so doing we should obtain greater control over the system features and be able to correct or modify them even while it is up and running.

This stage, still in progress, was preceded by the creation of a prototype to begin testing the functionality and usability of the system. To accelerate implementation we based the prototype on some modules of Drupal,[14] an open source CMS (content management system) oriented to semantic applications. The core of this CMS is a system of information categorization based on vocabularies, called the Taxonomy Module. The Taxonomy Module's main features are listed below:

— each content item can be related to one or more vocabulary entries at one time (multidimensionality);
— any number of hierarchically dependent levels can be recorded for each entry;
— a set of synonyms and related entries can be recorded for each entry.

We transposed the faceted scheme into Drupal in the following way:

— a first-level entry was created in the vocabulary for each facet;
— foci were represented by «sons» of first-level entries;
— «see also» relations to entries of other facets were created for some entries.

12. <www.comune.torino.it>
13. <www.sistemapiemonte.it>
14. <drupal.org>

4 Interface issues

> Hierarchies are not natural concepts for most people when it comes to storing and retrieving arbitrary information. […] In the current desktop metaphor used by most window systems, you can nest folder within folder ad infinitum. It's no wonder most computer neophytes get confused when confronted with this paradigm. […] Computer science gives us hierarchical structures as tools to solve the very real problems of managing massive quantities of data. But when this implementation model is reflected in the manifest model presented to users, they get confused because it conflicts with their mental model of storage systems. […] In other words, rather than forcing users to navigate deep, complex tree structures, give them tools to bring appropriate information to them. (Cooper 2003, p. 156).

The above study, and others (Rosati 2004), champion the higher cognitive ergonomy of faceted schemes as compared to enumerative ones. On the other hand, decades of computer science have accustomed us to the latter. Therefore, in order to avoid confusing users, some tricks are needed when presenting them with faceted systems. As Boscarol (2005) rightly points out, design of the information architecture is not the same as design of navigation paths: the latter do have to reflect the logic of the system in some way, but not necessarily to imitate it strictly. In fact, in many cases it is appropriate that navigation paths reflect architecture only partially, or reveal its logic only in a gradual way.

In particular, one of the major risks of a faceted system is information overcrowding: showing all facets and foci together at one time may generate a cognitive overload to the user, who may get confused. Perhaps this is one problem of interfaces which dynamically show the content of all facets in parallel frames, as in some Hibrowse displays. If there are many facets, it may be better to show only the options (facets and foci) which are supposed to be most popular, leaving the whole picture to a subsequent detailed display. If, instead, there are few facets, it may be useful to show all the facets together with the most popular foci for each of them, and leave all the remaining options to a detailed display («more… », «view all options/subcategories… », etc.). The latter option appeared to be the appropriate one for our scheme.

While the Drupal prototype cannot fully and correctly leverage all the features of expressive notation, this could be done in the final homemade interface. Notation could be used as a coded field, or even a record identifier, in order to get records automatically sorted in the correct way. This does not necessarily imply that notation must be displayed in every search mode: verbal headings will be understood more immediately by most users, while codes could be shown in advanced display as additional information, allowing users who pay more attention to understand the logic and operation of the faceted scheme.

5 Discussion

Albeit the result of a specific need, our scheme for community services could serve as a useful reference for the development of other knowledge organization tools in the same domain. Schemes for e-government, indeed, are increasingly discussed both at national and international level (Peristeras 2004). In this perspective, we could convert it into an RDF ontology suitable to be used for semantic tasks across the Internet. Our experience shows that facet analysis provides a significant enhancement to the organization of this domain as compared with the enumerative scheme of life events. Indexing a larger corpus of resources, as well as performing tests with users, would make it possible to improve details of the classification scheme and the interface.

Using existing software not specifically designed for faceted schemes, while good in semantic tasks, obviously implied some limitations. The full development of the home-made database should allow a better integration between the structure of the faceted scheme and the user interface. Unfortunately, few examples are yet available of search tools joining usability with facets. As we now have both these conceptual tools, making them interplay efficiently in actual implementations should be our goal.

References

(Binding 2004) Binding, Ceri; Tudhope, Douglas. «Integrating faceted structure into the search process». In: McIlwaine, Ia C. ed. *Knowledge organization and the global information society: proceedings 8th ISKO conference, London, 13-16 July 2004* (Würzburg: Ergon, 2004), p. 67-72.

(Boscarol 2005) Boscarol, Maurizio. «Organizzazione link» [electronic resource]. *Usabilità e architettura dell'informazione*, (22 February 2005). <http://forum.diodati.org/messaggi.asp?f=5&id=875>. [cited 28 February 2005].

(Broughton 2002) Broughton, Vanda. «Faceted classification as a basis for knowledge organization in a digital environment; the Bliss Bibliographic Classification as a model for vocabulary management and the creation of multidimensional knowledge structures». In: López-Huertas, Maria José; Muñoz-Fernández, Francisco J. eds. *Challenges in knowledge representation and organization for the 21st century: proceedings 7th international ISKO conference, Granada, 10-13 July 2002* (Würzburg: Ergon, 2002), p. 135-142. Available online: <http://www.ucl.ac.uk/fatks/ paper2.htm>. [cited 28 February 2005].

(Cooper 2003) Cooper, Alan; Reimann, Robert M. *About face 2.0: the essentials of interaction design*. Indianapolis: Wiley, 2003.

(Dextre-Clarke 2002) Dextre-Clarke, Stella G. «Planning controlled vocabularies for the UK public sector». In: López-Huertas, Maria José; Muñoz-Fernández, Francisco J. eds. *Challenges in knowledge representation and organization for the 21st century: proceedings 7th international ISKO conference, Granada, 10-13 July 2002* (Würzburg: Ergon, 2002), p. 142-148.

(Flamenco 2004) Flamenco group. *Flamenco technical details* [electronic resource], 2004. <http://bailando.sims.berkeley.edu/flamenco-technical.html>. [cited 28 February 2005].

(GNOLI 2004) GNOLI, Claudio. *Classificazione a faccette*. Roma: AIB, 2004.

(ITALIA 2003) ITALIA. PRESIDENZA DEL CONSIGLIO DEI MINISTRI. DIPARTIMENTO PER L'INNOVAZIONE E LE TECNOLOGIE. *Front office e servizi di e-government per cittadini e imprese* [electronic resource]. <http://www.innovazione.gov.it/ita/intervento/normativa/allegati/avviso_allegato1.pdf>, [2003?]. [cited 28 February 2005].

(LA BARRE 2004). LA BARRE, Kathryn. «Adventures in faceted classification: a brave new world or a world of confusion?». In: MCILWAINE, Ia C. ed. *Knowledge organization and the global information society*: *proceedings 8th ISKO conference, London, 13-16 July 2004* (Würzburg: Ergon, 2004), p. 79-84.

(PERISTERAS 2004) PERISTERAS, Vassilios; TARABANIS K. «The Governance Enterprise Architecture (GEA) object model». In: WIMMER Maria A. ed. *Knowledge management in electronic government: proceedings KMGov 2004: 5th IFIP international working conference, Krems, May 27-29 2004*. *Lecture notes in computer science*, 3035 (Springer, 2004), p. 101-110.

(POLLITT 1994) POLLITT, A. Steven; ELLIS, Geoffrey P.; SMITH, Martin P. «HIBROWSE for bibliographic databases». *Journal of information science*, v. 20, n. 6 (1994), p. 413-426.

(RANGANATHAN 1967) RANGANATHAN, Shiyali Ramamrita. *Prolegomena to library classification*. 3rd ed. Bangalore: Sarada Ranganathan Endowment for Library Science, 1967.

(ROSATI 2003). ROSATI, Luca. «Per un accesso multidimensionale all'informazione: o della classificazione a faccette» [electronic resource]. *AifIA*. <http://aifia.org/it/articoli/000204.html>. [cited 28 February 2005].

(ROSATI 2004) ROSATI, Luca. «Le faccette, il web semantico e l'ornitorinco» [electronic resource]. *Trovabile*. <http://www.trovabile.org/faccette_web_semantico_ornitorinco>. [cited 28 February 2005].

(SLAVIÆ 2004) SLAVIÆ, Aida; CORDEIRO, Maria Inês. «Core requirements for automation of analytico-synthetic classifications». In: MCILWAINE, Ia C. ed. *Knowledge organization and the global information society*: *proceedings 8th ISKO conference, London, 13-16 July 2004* (Würzburg: Ergon, 2004), p. 187-192.

(VICKERY 1960). VICKERY, Brian C. *Faceted classification: a* guide to the construction and use of special schemes. London: Aslib, 1960.