Faceted Classification for Public Administration

Luca Rosati
(Tidysoft)
luca.rosati@tidysoft.com

Maria Elisabetta Lai
(CSI Piemonte)
mariaelisabetta.lai@csi.it

Claudio Gnoli
(Università di Pavia)
gnoli@aib.it

Introduction

Many people are quite dissatisfied with existing online search systems, especially since they don't give the feeling of "browsing the shelves" that one gets in a library (Flamenco Group 2002).

In particular, recent usability tests with users ([Postai 2004]) show that people cannot find all the information they need in Italian **public administration sites**, even if the information to retrieve is simple, like a telephone number or an address.

The usual full-text searches have these disadvantages:

- once found an item, there is no way of understanding its position in the taxonomy and its relationship with adjacent or related items (from this, a sort of spatial uneasiness);
- the items which are found, though containing one or more occurrences of the searched terms, are not necessarily relevant to the subject those terms should belong to (background noise);
- the search engine is unable to retrieve many relevant documents dealing with the subject wanted, as they don't contain any occurrence of the terms used for searching.

The Flamenco project and faceted classification

Between the end of the 1990s and the beginning of 2000s, a research group at Berkeley University designed a search engine just in order to resolve these problems. Flamenco (this is the name of the project) is conceived to explore complex collections by a flexible user interface – its name is an acronym for *FLexible information Access using MEtadata in Novel COmbinations*. The Flamenco project solves the search problem allowing a sort of information filtering, that is to say a kind of **access to information between searching and browsing** [Rosati 2004b].

The basis of this approach is the adoption of a particular system for classifying the collected items: faceted classification (also called analytico-synthetic classification or multidimensional classification) [Vickery 1960; Gnoli 2004].

In this way, a user is able:

- to choose whether to look for an information by direct search or by browsing a directory
- to refine or to expand the search results
- to explore items which are semantically related those which have been found
- to choose at any moment alternative paths or points of view.

Faceted classification vs. traditional classification

In traditional classification systems (also called hierarchical-enumerative systems), every element is classified under one and only one class. It has a unique position inside a hierarchical and very deep schema, and it can be found through a unique path: *father category > child category*. Linnaeus's system, Dewey Decimal Classification and the Library of Congress Classification are examples of hierarchical-enumerative systems.

The peculiarity of these systems is that all possible classes must be specified (enumerated) before the indexing process (this is the reason why they are defined enumerative classification).

Schemes they based on are very rigid and conservative, because they are structurally

closed, and require a centralized management organization. Those schemes do not allow the classifier to insert new classes during the indexing process.

An example of traditional classification, applied to the Italian wines, could be:

- Region (Piedmont, Tuscany, Sicily etc.)
    - Colour (white, red, rosy)
        - Vine
            - etc.

In contrast, faceted systems avoid the need of listing all the possible classes of the taxonomy, and allow instead to create the classes **on the fly** by combining elementary elements (facets and foci). Besides, they allow to associate an item to several categories or parameters at one time, each category representing an aspect or face of the item itself [Marino 2004].

Thanks to these features, faceted classification results really **flexible and extensible**, allowing everyone to create his own classes as he needs. Such systems are especially useful in managing a large amount of items, as they are able to satisfy a wide range of needs. Coming back to the previous example, a faceted list of wines could be as follows:

- Region
    - Piedmont
    - Tuscany
    - etc.
- Colour
    - White
    - Red
    - etc.
- Typology
    - Raisin wines
    - Sparkling
    - etc.
- Vine
    - Cabernet
    - Merlot
    - etc.

Looking for a wine, a user can choose among several query parameters; he can alternatively combine these parameters in order to restrict the range of the results. Sacco [2000] describe this kind of process as a "zoom" operation. For instance, a user can select only wines having characteristics like: Region: Sicily; Colour: red; Typology: raisin wines; results will be in a limited number.

Moreover, the user can further refine his query by adding more parameters, for instance: Vine: Aglianico or Plates: Red chicken.

Leading facets to their full potential: the CSI Piemonte case

Although *facet, faceted* have become very common terms in the information architecture field, their application is often far from its original meaning. The attribute *faceted*, indeed, is used in a large variety of meanings, and is often referred loosely to the availability of search by different keys [La Barre 2004]. The full theory of faceted classification, as it has been developed by Ranganathan and the Classification Research Group (CRG), including rules for citation order and notation, is less widespread as a background for website organization; remarkable exceptions are offered by projects including library staff, such as Hibrowse [Pollitt *et al*. 1994], FATKS [Slavic 2002], and FACET [Binding & Tudhope 2004].

So, Tidysoft thought to apply faceted classification theory to the eGovernement field fully respecting the original library theory, in order to take advantage of the full potentiality of faceted classification, and to obtain the maximum benefits. We have applied such theory to classification of community services in Piedmont (North-Western Italy) public administration on behalf of *CSI Piemonte* (Consorzio Sistemi Informativi Piemonte) – an organism who cares the eGov projects of Piedmont public administration.

Our proposal is inspired by the Flamenco project on one side, and by Classification Research Group (CRG) approach on the other one.

The facets for public administration community services

For chosing the facets to be used, we looked at the CRG theory [Vickery 1960]. An aspect often forgot in the Web, indeed, is that both Ranganathan and the CRG tuned up a **general schema** for faceted classification, which every schema can refers to. In such way, in a faceted classification project one has not to re-build the schema every time from scratch, but has a guide for builing its main categories (i.e. facets). CRG postulated 11-13 general categories:

- Things/entities
- Kinds
- Parts
- Properties
- Materials
- Processes
- Operations
- Patients
- Products
- By-products
- Agents
- Space
- Time.

Analysis of a first corpus of e-government topics allowed us to define **eight facets,** appearing to be suitable for classifying e-government resources:

- Services and practices (Entities)
- Life events (Kinds)
- Access mode (Properties)
- Stages (Processes)
- Citizens and companies (Patients)
- Public institutions (Agents)
- Geographic departments (Space)
- Dates (Time).

In this way, we have not rejected the classification criterion by "Life Events" (this criterion is suggested by Italian Innovation & Technology Ministry), but we have integrated it as one of the facets of our system. As a matter of fact, the problem with indexing by "Life Events" does not stand in the criterion itself, but in its mono-dimensional.

Foci have then been listed within each facet, on the basis of the literary warrant of existing topics and of a sample of web pages to be indexed. A helpful sequence has been established for them, and a notation has been assigned to each focus. We chose to use an expressive notation, as it is more suitable for information processing in a digital environment [Slavic & Cordeiro 2004]. We considered adopting the same literal facet indicators as FATKS, but as our facets are less than 10, we decided instead to use numerals to express facets and letters to express foci.

Facets plus search engine

How can facets be coupled with a search engine? After all, facets are essentially a classification system. Full-text search works on a set of 'flat' items lacking any semantic information. Faceted classification adds a **semantic layer** and semantic relationships to the collection items: a layer which in turn can be exploited in searching.

Rather than performing a trivial search of occurrences, the search engine will compare the searched terms with the values (i.e. *foci*) of the various facets, and will extract only (or first) the items associated with such values. In this way, the search works on the basis of a semantic classification of items in the collection, not simply on the basis of occurrence of keywords. Another important advantage is that, once a result is selected, the system also allows to specify (so to speak) its semantic position, namely its coordinates in the classification grid. So, the user has a clearer perception of the collection architecture in which he is moving, and can surf more easily in it.

Facets plus search engine plus controlled vocabulary

What happens if the searched terms don't coincide with the labels used for facets and foci? Here a controlled vocabulary plays its role.

Such a tool allows to associate a series of synonyms to the labels used for the facets and their foci. So, the query can be performed both on the labels and on the synonyms. If the searched term doesn't appear among the facets

or the foci but among the synonyms, the system will work in this way:

- Search engine discovers that the term doesn't appear between the facets or foci, but in the synonyms field
- from the synonym it goes up to the corresponding focus (for instance, focus C3 of facet C)
- it extracts all the items related to the focus (C3).

For instance: I am looking for information about "*illness grant*": I type "*illness grant*"

- the word "grant" is recorded as a synonym of focus "indemnity", belonging to facet "Services and practices"
- the system extracts all the items related to focus "indemnity".

References

Binding C., Tudhope D. 2004. "Integrating faceted structure into the search process", in: *Knowledge organization and the global information society*, Proceedings 8th ISKO conference, London, 13-16 July 2004, I.C. McIlwaine ed. Würzburg: Ergon, p. 67-72.

Dextre Clarke S. G. 2002. "Planning controlled vocabularies for the UK public sector", in: *Challenges in knowledge representation and organization for the 21st century*, Proceedings 7th international ISKO conference, Granada, 10-13 July 2002, M.J. López-Huertas, F.J. Muñoz-Fernández, eds. Würzburg: Ergon, p. 142-148.

Flamenco Group 2002. How to Build a Flamenco instance <http://bailando.sims.berkeley.edu/flamenco/howtobuild/howtobuild.html>.

Gnoli, C. 2004. *Classificazione a faccette*, AIB, Roma 2004.

La Barre, K. 2004. "Adventures in faceted classification: a brave new world or a world of confusion?", in: *Knowledge organization and the global society*, proceedings 8th international ISKO conference, London, 2004 July 13-16, I. McIlwaine ed. Würzburg: Ergon, p. 79-84.

Marino, V. 2004. "Classificazioni per il Web: i vantaggi dell'adozione di schemi a faccette", *AIB-Web,* <http://www.aib.it/aib/contr/marino1.htm>.

Pollitt, A. S., Ellis, G. P. & Smith, M. P. 1994. "HIBROWSE for bibliographic databases". *J. Information sci.*, 20, 6, p. 413-426.

Postai, S. 2004. Facilitare la reperibilità dell'informazione: la classificazione a faccette, Internet Pro, Ottobre, 2004 <http://www.tecnichenuove.com/Extra/Riviste/Internet.pro/INPro0401-054-ArchFaccette.pdf>

Rosati, L. 2002. "Per un accesso multidimensionale all'informazione: o della classificazione a faccette", *AifIA* <http://aifia.org/it/articoli/000204.html>.

Rosati, L. 2004a. "Classificazione a faccette", <http://lucarosati.it/risorse/classificazione_a_faccette>

Rosati, L. 2004b. "Mac OSX Tiger a faccette?" <http://www.trovabile.org/mac_os_x_tiger_a_faccette>

Sacco, G. M. 2000. "Conventional taxonomies vs. dynamic taxonomies" <http://www.di.unito.it/~sacco/sacco-dyntax.pdf>.

Slavic A. 2002-. *FATKS: Facet Analytical Theory in managing Knowledge Structures for humanities* <http://www.ucl.ac.uk/fatks>

Slavic A. & Cordeiro M. I. 2004. "Core requirements for automation of analytico-synthetic classifications" In: *Knowledge organization and the global society*, proceedings 8th international ISKO

conference, London, 2004 July 13-16, I. McIlwaine ed. Würzburg: Ergon, p. 187-192.

Vickery B. 1961. *Faceted classification: a guide to the construction and use of special schemes*. London: Aslib, 1960.