Tags for Citizens: Integrating Top-Down and Bottom-Up Classification in the Turin Municipality Website*

Franco Carcillo[1] and Luca Rosati[2]

[1] Turin Municipality website & assistant professor at University of Turin,
Via Meucci, 4, 10121 Torino, Italy
`carcillo@comune.torino.it`
[2] Information architect & assistant professor at University for Foreigners of Perugia,
via XX settembre 25, 06124 Perugia, Italy
`luca@lucarosati.it`

Abstract. Tags for citizens project aims to empower the citizens' experience within the Turin municipality website (http://www.comune.torino.it) integrating a standard top-down taxonomy with a bottom-up classification by tags. The top-down taxonomy has been conceived following the UK Integrated Public Sector Vocabulary (IPSV 2006 – an ISO 2788 fully compliant classification scheme) and empirically refined by usability tests with users and by log files monitoring. The bottom-up classification works as a social tagging system. The latter it is not simply added to the former, but completely integrated to it, in order to obtain a coherent system.

Keywords: eGovernment, Metadata, Information Architecture, Social Classification, Folksonomy, IPSV (Integrated Public Sector Vocabulary).

1 Introduction: The Best of Both Worlds

The collaborative tagging (or folksonomy) may be not only an alternative to the canonycal classification models born in librarianship field, but it may works as a valid complement of conventional taxonomies or facets. In such a way we can correct the intrinsic limits of each classification approach and enforce their benefits. Considering the wide range of mental models an e-gov site has to satisfy, this may take strong advantages by a "blended" approach.

So, the purpose of the project is to show how the flat keywords space of user-generated tags can be effectively mixed with a richer poly-hierarchical classification scheme to improve the system information architecture.

Besides enforcing order on the flat space of keywords, the blend of tags and taxonomy is able to empower the *information scent* [1] and the *berrypicking* [2] capabilities of the system. Every information architecture project refers to two different information axes [3]:

- a **vertical** (or paradigmatic) axis, i.e. the hierarchical relationship that each item of a system engages with the others.

---

* Though the article is the result of a common strength, Franco Carcillo wrote the paragraphs 1-2, Luca Rosati the paragraphs 3-4.

- a **horizontal** (or syntagmatic) axis, i.e. the semantic, contiguity relationship that each item engages with the others.

In our case, the combination of tags and taxonomy allows for better management of both these axes:

- from the vertical or paradigmatic point of view, when a user is going to associate a keyword to a category of the taxonomy (in order to tag a resource), the system suggests similar tags or hierarchy of tags pertaining to the same category
- from the horizontal or syntagmatic point of view, at the same time, the system will allow the user to see all the other tags belonging to the same category.
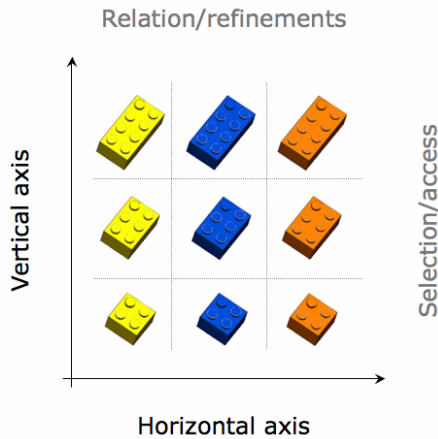

Fig. 1. The two axes of information architecture

Other main features will be:

- tags suggestion (by Ajax) during the tag assignation phase or during the keyword/tag insertion in the search engine
- hybrid search both by tags and keywords
- usage opportunity of compound tags or keywords
- tagging history.

2   Taxonomies and Folksonomies: Pros and Cons

My guess is that we have a folk theory of categorization itself. It says that things come in well-defined kinds, that the kinds are characterized by shared properties, and that there is one right taxonomy of the kinds.

It is easier to show what is wrong with a scientific theory than with a folk theory. A folk theory defines common sense itself. When the folk theory and the technical theory converge, it gets even tougher to see where that theory gets in the way-or even that it is a theory at all (Lakoff) [4].

Taxonomies are top-down classification systems designed by specialised staff and thus fairly centralised and accurate. They resemble the form of classification Lakoff calls "scientific theory", reflecting Aristotelian-type logic.[1]

On the other hand, folksonomies are bottom-up classification systems, created by users themselves by adding keywords. As they are distributed, these systems are not too accurate, but better reflect a user's viewpoint. They come closer to the classification model Lakoff calls "folk theory" - in this case comparable to the Eleanor Rosch prototype theory (more than to Aristotle).

However, in philosophical terms, no classification system can call itself completely scientific or completely empirical (folk). Any classification, even one starting out as scientific, tends to become folk, based on the empirical needs its use inevitably requires of it.

> So why do we sometimes appear in practice prototypical in our classilications, even if in principal we are Aristotelian? For two main reasons: because each classification system is tied to a particular set of coding practices; and because classification systems in general (we are not making this as an ex cathedra pronouncement) rellect the conflicting, contradictory motives of the sociotechnical situations that gave rise to them. [...]
>
> In addition to this inheritance, there is a practical Occam's razor. When doctors come to code causes of death they are frequently faced with a set of difficult judgments (which may require an autopsy and further diagnostic work). They can simply go for the easiest way, by using a generalized "other" category. They can then get back to dealing with their live patients [...]. So the classical beauty of the Aristotelian classification gives way to a fuzzier classification system that shares in practice key features with common sense prototype classifications – heterogeneous objects linked by metaphor or analogy. [...]
>
> There is no great divide between folk and scientilic classifications. Below, we discuss one particular fault line between the two: a fracture that is constantly being redelined and changing its nature as the plate of lived experience is subducted under the crust of scientific knowledge (Bowker & Star) [5].

All these considerations have led us to the idea of combining top-down and bottom-up classifications in order to both gain (so to say) the best of two worlds and – above all – to try to get a better picture, through folk classification, of citizens' mental models and correct the limits of top-down taxonomy.

Taxonomy and folksonomy can be seen as alternative, complementary tools and not opposing ones. Due to such individual peculiarities, each can be used in specific contexts with specific objectives.

2.1 Taxonomies Anatomy

Taxonomies are knowledge organisation and management systems made up of terms set out in a hierarchical tree-like structure. They exploit hierarchical-enumerative classification systems, from which they inevitably inherit qualities and defects.

---

[1] An Aristotelian classification works based on binary properties that the object being classified has or has not. Hence, at each classification level, you choose a group of these properties to decide on whether an object belongs or does not belong to a single class: e.g. a polygon belongs to the rectangle class if it has four sides, these sides are at right-angles and if the opposite sides are the same length.

Tree-like systems permit precision and order when classifying elements within a group, favouring a 'known-item seeking' approach. These same characteristics are also their limits.

- As they are top-down (creation and upkeep by a few experts or an institution), taxonomies are fairly centralised systems where any changes to be made can only be made from above.
- The strict precision of tree-like models often makes for excessive system rigidity and closure. Introducing new categories at any point in the system often requires changes being made to other parts and can only be done by whoever created the system.
- As they branch out deeply, tree systems tend to hide the more popular classes (those used by users to a greater extent) and 'push' them downwards.
- The hierarchical structure is highly suited to known-item seeking strategies (used by users who know what they are looking for), while it does not suit exploratory-seeking strategies (users who do not know what they are looking for, those browsing or non-expert users). Moreover, they do not help the berry-picking model type (see Par. 1) so typical of the web.

2.2   Folksonomies Anatomy

The strength of these systems lies not in their precision but their wide-scale popular use, i.e. people. Instead of making an effort to build classification systems suiting the users' mental model, you simply enable the users themselves – from below – to 'create' common, shared mental models, which emerge spontaneously (through a collaborative process and natural selection).

As a matter a fact, folksonomies should be called **distributed classification** systems. Folksonomies can also be used for analysis: to study the mental models of entire communities through the way their members describe (classify) a group of collected items. They can be used as support when processing categories, choosing favourite terms or setting up controlled vocabularies [6].

Critics of folksonomy mention the following main defects:

- Folksonomies are not precise either in terms of categorising or language. The risk is that of having many language variables for the same concept thus loss of control over vocabulary and its synonyms. It is also true that recent tag systems foresee a 'suggestion-based' system displaying tags linked to the one chosen. Moreover, to criticise this position we can turn to the Lakoff quote found in Par. 2: a classification system is always somewhat hybrid as its end is practical (what it is used for) and this inevitably mixes a logical-scientific base with strictly popular, empirical criteria.
- Tags used in folksonomies have no hierarchical structure whatsoever. There is no division into categories and sub-categories, whether in the hierarchical-enumerative sense or analytical-synthetic. Tags are a flat space of keywords. This objection – though essentially true – does not highlight the most innovative aspect of the phenomenon, which is not so much the flat mass of tags but their aggregation by users themselves.

- Folksonomies have low findability and are therefore more suited to exploratory seeking than known-item seeking.

As far as the positive or interesting aspects of folksonomies are concerned, we can list these advantages as follow.

- Above all, not all folksonomy limits are defects. If on the one hand you lose precision, on the other this loss is offset by an important number of benefits in terms of suggestions as to the representation and use of certain concepts by a large number of users, and thus on navigators' mental models. In this sense, folksonomies can be considered not only as classification systems but as support tools for designing the classification itself.
- Folksonomies are more suited to finding information (discoverability) than targeted searches (findability). Referring to models seen, we could say that folksonomies favour the **berry-picking** search model type, and strongly integrate the idea of **serendipity**.
- Inexpensiveness: tag systems are extremely economical to set up (in terms of both time and money), as all classification work is carried out by users successively.

3   "Tags for Citizens" Project: Classification Systems Components

3.1   The Taxonomy Layer: IPSV (Integrated Public Sector Vocabulary)

We are running a taxonomy redesign of the Turin Municapality website following the British standard IPSV (Integrated Public Sector Vocabulary) [7]. The Integrated Public Sector Vocabulary (IPSV) is a controlled vocabulary, also known as an 'encoding scheme' for populating the Subject element of the e-Government Metadata Standard (e-GMS). It should be used with all the electronic resources produced in the UK public sector, so that citizens may access them more easily. IPSV complies with ISO 2788 and BS 8723, the International and British Standards for monolingual thesauri.

The IPSV was developed with the backing of the Department for Communities and Local Government (DCLG) (formerly the Office of the Deputy Prime Minister) and the e-Government Unit (e-GU) of the Cabinet Office.

We have structured the redesign process into three steps:

- content inventory (mapping of the first two navigation levels of Turin Municipality website (www.comune.torino.it)
- comparison between the actual taxonomy, the IPSV model, and a sample of IPSV application (the UK Public Services website, www.direct.gov.uk) in order to extract the new top-down information architecture of Turin Municipality website
- a successive deeper comparison between the labels of the two systems in order to design a more coherent labelling system.

The matching results are shown in Table 1 below (we have not enough space to explain the taxonomy redesign procedure in detail).

Table 1. Comparison between the actual Turin Municipality website taxonomy, the IPSV model, and the UK Public Services website, www.direct.gov.uk

Turin website	IPSV	Direct.gov.uk
Ambiente e verde [Environment and green]	Environment	Environment and greener living
Casa e territorio [Home and territory]	Housing	Home and community
Commercio e impresa [Commerce and business]	Business and industry	Information for businesses [in the footer menu]
Cultura, sport e tempo libero [Culture, sport and leisure time]	Leisure and culture	Young people > Leisure
Formazione e Scuola [Education, training and school]	Education and skills	Education and learning
Lavoro e Orientamento [Work and orientation]	Employment, jobs and careers	Employment
Sanità e Servizi Sociali [Healthcare and social services]	Health, well-being and care	Health and well-being
Sicurezza ed emergenze [Safety and emergencies]	Public order, justice and rights > Security	Emergencies	In: Newsroom > Public safety [utilities menu]
Soldi e risparmio [Money and savings]	In: Economics and finance > Personal finance	Money, tax and benefits
Tasse e agevolazioni fiscali [Taxes and tax benefits]	In: Economics and finance > Tax	Money, tax and benefits
Trasporti e viabilità [Transport and roads]	Transport and infrastructure	Travel and transport
Jolly Item (e.g. Special: Olympic Games 2006)	--	--

3.2 The Folksonomy Layer

The folksonomy layer works as a speed classification/movement skin overlapping the taxonomy layer.

This system has two main aims.

- Allow users to save pages that interest them from the City of Turin website in a reserved area so as not to have to search for them each time (labelling them through tags). As such, tags become access shortcuts for frequent users.

- Furthermore, allow users who have not saved their "own pages" to use tags created by other users as complementary browsing tools to the taxonomy and search motor.

From the home page, the user can use the tags both as main tools to access content or to refine a search started through the standard taxonomy or search engine. User tags will be saved in a personal area on the website - with each user able to decide to what extent to make this information public: completely, partially or not at all.

These two forms of use correspond to two levels of interactivity:

- one more active, the former.
- one more passive, the latter.

The first form of use is for users who are more inclined to take part actively or used to using similar systems (e.g. Del.icio.us). The second is for those who, though more conservative, can be stimulated to change their behaviour by taking advantage of work done by other users.

3.2.1 Save and Find Functions

The system is displayed to users through two key functions in two boxes on the page (normally placed in the right-hand column) - see Fig. 2.

The Save function allows users to save the page in a personal bookmark (tagging it) to find it more easily later on. No registration will initially be required. Users will only need to tag the page and then will be subsequently requested to complete a brief registration. Alternatively, they can use the Delicious account (if they already have one) or another similar system.

The Find function allows users to find pages similar to, or linked to, the current one, through tag clouds. The system will visualise the tag/s (if available) other users used to tag the page and any other tags linked semantically to the first ones (e.g. "Whoever used this tag also used...", or "Whoever consulted this page also read..."). This function will also allow users to search for pages on a specific topic through the use of tags.
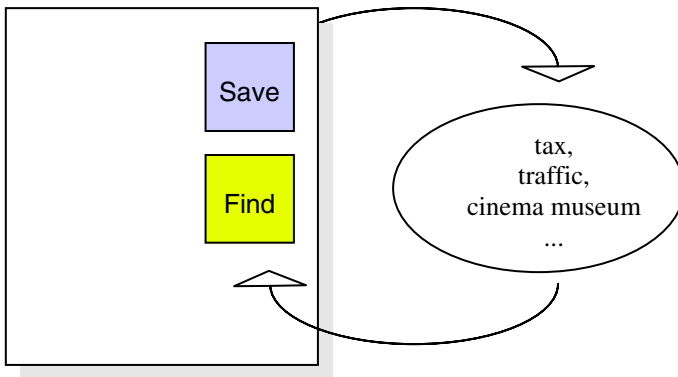
Fig. 2. Main functions of the folksonomy layer

4 Putting All Together

We said the folksonomy layer is not just added to the taxonomy one, but integrated with it. The synthesis happens in the tagging phase: during the tag inserting process. The system asks users to link tags to the site category taxonomy creating a link between the two classification layers. One tag can be associated to several categories at the same time.

Integration has further advantages in terms of system functionality.

- The user has two complementary access systems available: through taxonomy and tags.
- There can be interchange between the two systems: moving from taxonomy to tag and vice-versa, without interruption.

In greater detail, this enables:

- Showing (all) other tags belonging to a certain category when the user associates a tag to that category.
- Avoiding the tag flatness problem (tags on a single indistinct level), allowing macro-category grouping.
- Helping to solve (at least partly) the problem of synonyms.
- Monitoring the most frequently used tags and their correlation with categories to correct or integrate the taxonomy (the classification system created from above).
- Progressively absorbing into a meta-data or controlled vocabulary system tags going over a certain threshold of use (as happens in the BBC website).

> *BBC staff have suspected for many years that metadata could be the solution to guiding the audience through the site, but it has not been simple to find the right approach. [...]*
> *Maintenance costs and responsiveness [of the controlled vocabularies] were still a problem. A compromise solution (known as the metadata threshold) allows for free-text tagging that is absorbed into formal controlled vocabularies when enough content is tagged with that term. The solution aims to combine cheap and responsive tagging with unambiguous aggregation power. So far it has been very successful at slashing overheads. The controlled vocabulary, semi-automatic suggestions and metadata threshold were still coupled to the CMS, but the development of an application programming interface (API) should resolve the issue of tagging content being produced in different systems (Loasby) [8].*

References

1. Chi, E.H., Pirolli, P., Chen, K., Pitkow, J.: Using Information Scent to Model User Information Needs and Actions on the Web. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM Press, New York (2001), http://www2.parc.com/istl/projects/uir/publications/items/UIR-2001-07-Chi-CHI2001-InfoScentModel.pdf
2. Bates, M.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface, UCLA (1989) http://www.gseis.ucla.edu/faculty/bates/berrypicking.html

3. Rosati, L.: I due assi dell'architettura dell'informazione [= The Two Axis of Information Architecture], Trovabile (2006) http://trovabile.org/articoli/assi-architettura-informazione
4. Lakoff, G.: Women, Fire and Dangerous Things: What Categories Reveal About Mind, p. 121. University of Chicago Press, Chicago (1987)
5. Bowker, G.C., Star, S.L.: Sorting Things Out: Classification and Its Consequences, pp. 64–67. MIT Press, Cambridge (1999)
6. Quintarelli, E.: Folksonomies: Power to the People, paper presented at the ISKO Italy-UniMIB meeting, ISKO Italia (2005) http://www.iskoi.org/doc/folksonomies.htm
7. UK Cabinet Office eGovernment Unit: IPSV-Integrated Public Sector Vocabulary, (2006) http://www.esd.org.uk/standards/ipsv/
8. Loasby, K.: Changing Approaches to Metadata at bbc.co.uk: From Chaos to Control and Then Letting Go Again, in ASIS&T Bulletin (October/November 2006) http://www.asis.org/Bulletin/Oct-06/loasby.html