Merging semantic and participative approaches for organising teachers' documents

Benjamin Huynh-Kim-Bang, Eric Dané, Monique Grandbastien

Merging semantic and participative approaches for organising teachers' documents

Benjamin Huynh-Kim-Bang*, Éric Dané**, Monique Grandbastien*

* Université Henri Poincaré / LORIA
54506 Vandoeuvre-lès-Nancy, France
{benjamin.huynh-kim-bang, monique.grandbastien}@loria.fr

** Université Paris Descartes
75270 Paris, France
eric.dane@univ-paris5.fr

Communities of teachers need tools to gather and organise educational resources scattered over the Web. As current generic search engines are insufficient to deal with contexts, these tools must rely on human annotations. First, we compare the Semantic Web and the Participative Web (2.0) approaches to annotate resources and we argue for the need to merge them. Secondly, we propose a social bookmarking tool merging the flexibility of tags with basic inference processes attached to ontologies. Thus we define the concept of *structurable tags* which are tags that can be linked through relations while keeping their spontaneous, on-the-fly and aggregative properties.

Introduction: Sharing documents in a community of teachers

Context. New issues around shared resources are raised by the increasing participation of users into the production and the organisation of these resources. Our research deals with communities of teachers on the Web and their process to share educational documents. Indeed, results provided by search engines do not seem to totally satisfy those teachers who use other tools to organise and search resources.

Presentation of the community. Our research focuses more particularly on a group of French teachers of mechanical engineering that have been exchanging for 10 years via a mailing-list[1] (more than 100 messages/month). In this community of practice [1] composed of 1800 members, we can observe exchanges of information, personal experiences and references to online resources (technical or pedagogical documents).

Scattered resources. A major aspect of this community is that resources used by teachers are highly scattered over French institutional websites (local or national) or a huge number of teachers' ones. One member in ten has a website where s/he provides his/her own resources.

Organising resources through a mailing-list. To cope with this scattering aspect, the community transformed the mailing-list into a tool to regroup and comment on resources: when a member asks for a resource on the mailing-list, other members answer him/her privately; then this member sends back to the mailing-list a summary of the received answers. Thus the mailing-list is not overloaded and broadcasts rich summary messages. Nevertheless, this system reaches limits as users cannot easily browse previous messages which are poorly archived.

Other tools to organise resources. To resolve this archive problem, the community has tried several solutions. The foundator tried to put manually the most interesting resources from the mailing-list to a website. Yet the updates

[1] http://listepgm.org/

were irregular. A Web forum was set up where small groups of members could discuss more specific subjects than those tackled on the mailing-list. Thus we observed that some threads of the forum were especially used to share links to resources. Finally, a few months ago, the community completed the mailing-list with a collaborative content management system (Joomla).

Issues of the community. Interviews processed with members of this community pointed out the need to enhance the efficiency of participation, which requires to complete the mailing-list with tools to organise scattered resources. Our work is a tentative to help communities of teachers by providing them with suitable sharing tools. If search engine based on automatic annotations are insufficient, then tools based on manual ones seem necessary.

State of the art: semantic vs. participative approaches

Limits of current search engines. Sharing documents requires finding them. Most of the current search engines (like Google's or Yahoo's ones) select documents by looking for terms then rank them by pertinence. So this approach reaches its limits with implicit information. Except for basic synonyms or plural forms they do not find a page if it does not include terms of the query. For example, they cannot determine if an educational document is an exercise or a course without the terms appearing in it. Among current solutions to resolve this difficulty are metadata added by manual annotations (other solutions are described in [3]). We distinguish two major methods based on a priori or a posteriori manual annotations.

A priori annotations and Semantic Web

The first method comes from paper-based publishing and is built on the annotation of documents by specialists according to a model produced by experts e.g. LOM. More generally, this trend can be situated in the Semantic Web's context where experts create ontologies, which are models of domains on which computers can perform reasoning [2]. According to [3] and [4], this approach includes at least two weaknesses. Firstly, the creation of a common model is a difficult task — even for experts — requiring an important negotiation phase. Secondly, the annotation process is costly because it must be operated by specialists who understand and are able to apply the pre-defined model. Consequently, this centralized approach can be difficult to manage as difficulties encountered in the application of LOM [5] and the numerous Application Profiles illustrate it.

A posteriori annotations and participative Web (2.0)

The second method has appeared with the mass participation of Internet users in producing and annotating resources. This participative Web is called "Web 2.0"[2] to highlight the shift from a "traditional" Web run by more rigorous rules for publishing. This trend is illustrated by the popularity of tools like wikis, blogs or tools for sharing videos, photos and bookmarks.

Social bookmarking and tags. Sharing bookmarks consists in annotating URLs and in displaying them to others on a public space. In order to make easier the annotation of bookmarks, the first social bookmarking tools like Del.icio.us[3] enabled users to annotate resources with freely chosen keywords called tags instead of using controlled vocabularies or taxonomies. Indeed each user owns his/her own tags that cannot be modified by anyone else. Thus s/he can use his/her own vocabulary without constraints. Moreover tagging does not separate the creation of a domain model and the annotation process based on it. The model is not defined a priori but it is built by each tag associated to a resource. So tagging has the advantage to allow many users to associate tags to a URL and concurrently to enrich its description. By aggregating individual tags, social bookmarking tools compensate for the lack of predefined structures by generating browsing interfaces based on statistical calculations (frequency, co-occurency, etc).

---

[2] O'Reilly, T. "What is Web 2.0" (2005): http://www.oreillynet.com/pub/a/oreilly/\\tim/news/2005/09/30/what-is-web-20.html

[3] http://del.icio.us

Limits of tags. Nevertheless, tagging reaches its limits in the fact it does not allow advanced structuring [6] [7]. First of all, tags leads to flat structures, letting few ways for users to organise their own tags. Moreover, tags' efficiency decreases because of problems of typography or synonymy. For example, a user has no way to express that "car", "cars" and "automobile" are a same concept for him/her.

Merging semantic and participative approaches

The rigidity of approaches based on controlled vocabularies (thesaurus, ontologies) realized by experts seems incompatible with the flexibility of tags built in a very distributive way. However it exists an important issue in combining these approaches in order to take advantage of their qualities and to compensate for their weaknesses (Table 1). This combination is supported by studies demonstrating the consensus reached through tags in huge communities [9]. Few authors describe this merging trend as "Semantic Web 2.0" [10] or "Web 3.0" [11].

	Semantic Web	Participative Web
Qualities	Powerful logical and computed reasoning	Low-cost metadata with diversity of point of view
Weaknesses	High-cost metadata Difficulty to obtain a single model	Difficulty to structure metadata Only basic clustering and co-occurencies around metadata

Table 1: Semantic vs. participative

Problem and hypotheses

Communities of teachers, like the one studied, need tools to share resources. Human annotations are necessary because of the need, among others, to determine the implicit context of a document. However, semantic and participative approaches provide complementary solutions which seem to be incompatible.

General hypothesis: It is possible to provide a tool merging the flexibility of the participative approach completed with few inference possibilities from the semantic approach.

Our proposition: We propose a model to build a social bookmarking tool allowing basic inferences on tags. This model relies on the concept of *structurable tags* which are tags that can be linked together by users.

Constraints: Web 2.0 approach is based on the participation of users allowed by ergonomic features. Our proposition must gingerly take into account this aspect.

Structurable tags

Based on an existing social bookmarking tool. In order to test our hypothesis, we add functionalities to Scuttle[4], a social bookmarking tool under free licence and which is a clone of the web application Del.icio.us. The Scuttle version extended by our developments is called SemanticScuttle and is documented and downloadable on Sourceforge[5].

Presentation of a social bookmarking tool. Like Del.icio.us, SemanticScuttle allows a logged user to add a bookmark composed of a URL, a title, a description and a list of tags. Once added, the new bookmark appears on the

[4] Scuttle (by Marcus Campbell): http://sourceforge.net/projects/scuttle/

[5] http://sourceforge.net/projects/semanticscuttle/

user page and on the home page listing the last added items. To browse among bookmarks, the tool offers on the left a list of bookmarks and on the right, the corresponding tags clouds listing tags according to frequency or recency criteria (Fig. 1).

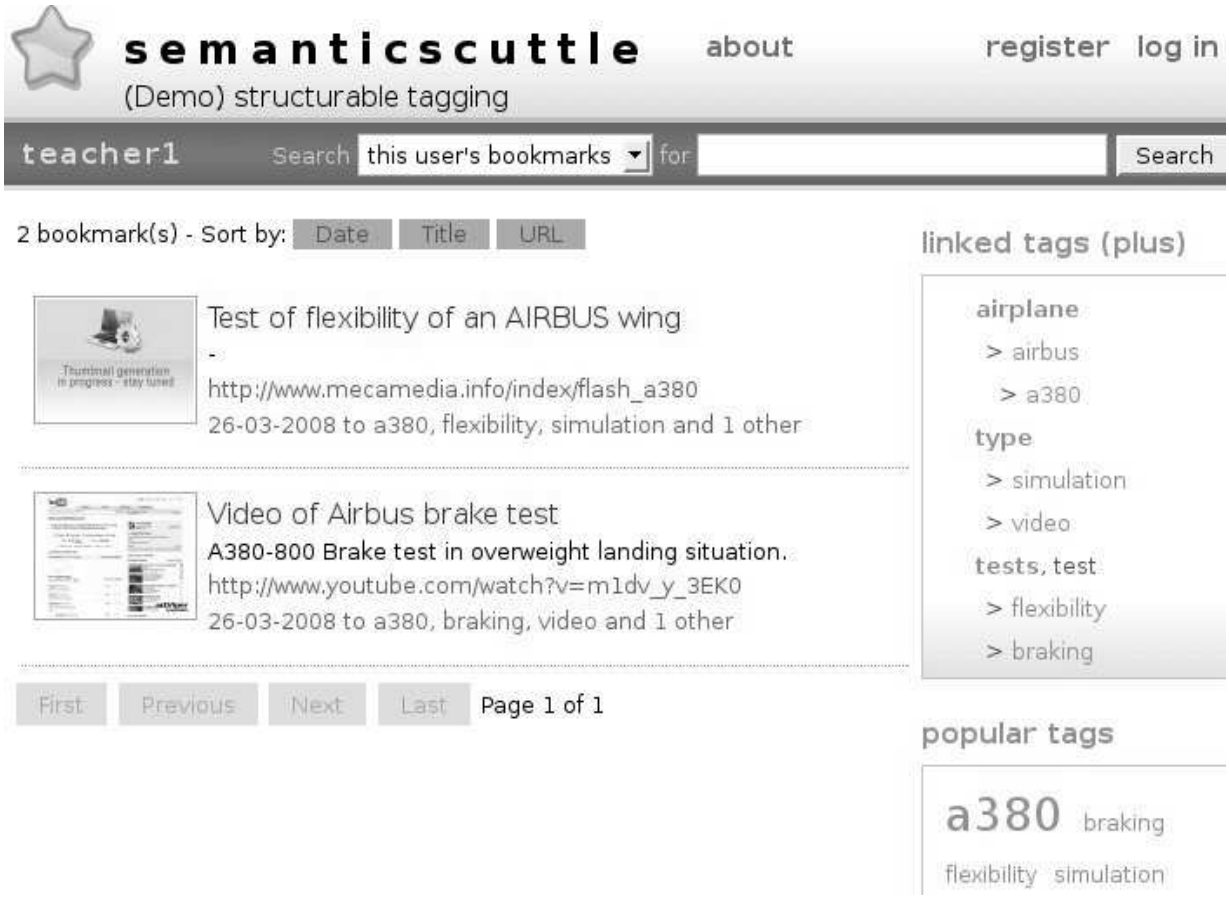

Figure 1: Bookmarks of "teacher1" (Screenshot from SemanticScuttle)

Relations between tags

In order to organise tags and to allow inferring mechanisms by the system, SemanticScuttle offers to users the possibility to create two types of relations between tags: the inclusion one symbolised by the character ">" and the synonymy one symbolized by "=".

Inclusion relation. Let's consider a webpage including a video of an AIRBUS A380 airplane showing a test for emergency braking. A user decides to create a bookmark for this page and fill the tags zone:

airplane>airbus>a380, test>braking, type>video

Thus the system interprets this succession of tags by the creation of oriented relations of inclusion between tags *airplane*, *airbus* and *a380*, between tags *test* and *braking*, and between *type* and *video*. Then the bookmark is displayed with simplified tags (*a380*, *braking*, *video*) which are the smallest according to the inclusion relation (see bookmarks at the bottom left of Fig. 1).

Then let's consider that the user adds a bookmark about a simulation illustrating the test of the flexibility of an A380's wing. The user describes the bookmark with:

test>flexibility, a380, type>simulation

Thanks to these inclusion relations, the system can infer on tags and make easier the search of bookmarks. Thus we created a basic inference engine which is able to complete a query for *test* into a query for bookmarks annotated by *test* but also *braking* or *flexibility* included into *test*.

Synonymy relation. Now let's suppose when s/he adds the second bookmark, the user does not remember if s/he previously used the tag *test* or the plural form *tests*. Thanks to the synonymy relation, the user can declare on-the-fly an equivalence between both tags by typing in the tag zone:

test=tests, test>flexibility, a380, type>simulation

The system interprets the firsts tags as the creation of a synonymy relation between *test* and *tests*. In the context of a search of bookmarks, the inference engine interprets a query on the tag *tests* as a query for bookmarks tagged with *test* and *tests* which are synonymous, but also *flexibility* and *braking* which are still included into *test*.

These relations between tags decrease the number of synonymous tags that must be typed by the user. Users can also find "interesting" bookmarks described with not queried tags thanks to inclusion.

As classic tagging (and on the contrary of most of the approaches based on ontologies), this system is error-tolerant. If a user creates an inclusion cycle (for example *airplane>airbus>airplane*), SemanticScuttle does not block the user by forcing him/her to immediately correct the relations. Each query on one of these tags will be simply completed by the other tag. If a user does not like a relation then s/he can modify it in an administration interface.

The "structured linked" on the right of Fig. 1 shows how the system displays the relations to the user. Included tags appear below their parent tags while synonym tags appear side by side like *test* and *tests*.

Multi-users context

Aggregation As the system cannot display all the tags, they have to be gathered and selected. One major benefit of structurable tags is their possibility of aggregation like classic tags. If two users used two tags with the same name, the system considers it as one tag and regroups the bookmarks under this one. There are no possible conflicts. The same effect is produced with structurable tags and their relations. If two users create tags with the same name but different relations, these relations are just gathered under the same tag. For example, if *tag1* includes *tag2* for one user and if *tag1* includes *tag3* for another user then the aggregation of these relations on the home page is *tag1* including *tag2* and *tag3*.

Visualisation Home pages of social bookmarking tools generally display selected tags as a view of the global users' activity. These highlighted tags have a huge influence on which tags will be used by taggers [7]. Scuttle, the tool on which are based our developments, is able to display tags according to classic criteria as the most recent or the most used ones. Nevertheless, structurable tags offer new possibilities like the next three ones already implemented:

- Number of descendants: tags are sorted according to the number of tags they include. For example, if *tag1>tag2>tag3* and *tag1>tag4* then tag1 include three tags. This mode focuses on items which cover a lot of other items.
- Length of branches: tags are sorted according to the length of their longest branches. For example, if *tag1>tag2>tag3* then tag1 has a branch whose length is two. In a tree, it corresponds to the height (or depth) of a node. This mode focuses on items which are the head of particularly detailed fields (on the difference with the first mode which could focus on vast field).
- Number of updates: tags are sorted according to the number of updates of its branches. Any times a tag A is added to or removed from a tag B then all the tags which include tag B increments one update. This mode

focuses on tags whose branches are often modified, which could mean that they are often re-worked and thus maybe particularly interesting to users.

These possibilities can be combined with each others or with classic ones. We need more experimentations to confirm their benefit but we already can say that they offer new perspectives for visualisation of tags.

Related works on tags and relations

Other works. Several projects currently explore how to make more flexible the use of ontologies. Few works allow users to create tags and then to link them to a common underlying ontology managed by a supervisor e.g. for semantic wiki [12] or semantic blog [8]. Other works allow users to complete an ontology by adding concepts and relations e.g. for the annotation of pictures [13] or the annotation of pedagogical documents [14] [15].

Comparison with other works. Our work is different from the previous ones in three points. Firstly, there is no supervisor with special rights allowing him to aggregate items. Any final user can add relations between tags. Thus our approach is more respectful of the tagging process where user is totally responsible of his/her actions.

Secondly, there is no underlying common model (like an ontology) with which every users must be consistent. SemanticScuttle accepts that users disagree with each others. Indeed, if users are forced to respect a too constraining frame, it will decrease the usability of the tool. The popularity of social bookmarking tools support this hypothesis. We understand the supervisor approach or the one based on a common model as an easy way to unify visions of a domain. However, we justify our choice of highest decentralization by the exploration of the possibility to obtain a partial common view by interactions and negotiations instead of a centralized control. If all users want to use the same tags they can do it, but they are not forced. If few users do not agree with tags from others, they can continue to keep their way to tag but others can try to convince them through discussions (see perspectives about collaborative spaces). We consider the success of the Wikipedia community as a proof of this concept of negotiation. Furthermore, if previous works about constraining tools can work in a constraint context like a company [8], there is no real clue that they can be applied in the context of an online community.

Thirdly, we brought a special attention to the ergonomic issue for the creation of structurable tags. Thus SemanticScuttle is the only tool we know which allows to create relations between tags on-the-fly thanks to special characters. Yet this characteristic seems essential to us in order to keep the flexibility in the use of tags.

Finally, all these works try to merge ontological and tagging approaches. Yet SemanticScuttle keeps tagging properties and adds ontological ones, while other works keep ontological properties like a single underlying model and add participative ones (Table 2).

	SemanticScuttle	Related works
Control	Any user can create tags and relations between them	One supervisor has advanced control on tags and relations
Model	No common model but an aggregative one tolerating inconsistency	One underlying common model or constraining one
Ergonomy	On-the-fly relations	Relations added through browsing in a tree
Conclusions	Based on tagging approach	Based on ontological approach

Table 2: Comparison with other works

Comparison with currently experimented applications. SemanticScuttle is situated in a *current* trend exploring the mix of human and computational search. We are waiting more results on the usage of Google co-op search[6], Swicki[7], Gnizr[8] or the Wikia search engine[9] (based on the Wikipedia principles) in order to compare them with SemanticScuttle ones.

Discussion and perspectives

Proof of concept and further experimentations. Structurable tags implemented in SemanticScuttle prove the technical possibility[10] to make inferences on tags while keeping their spontaneous and flexible aspect. We are currently experimenting SemanticScuttle into the community of teachers described in the introduction to verify usability of structured tags.

Perspectives about structurable tags. As described in this article, structurable tags offer promising possibilities to organise tags. Thus we are thinking of new types of relations between tags. For example, we foresee a "-" relation which means that two tags are linked for the user. For example, if a user creates the relation *vehicle-trademark*, then when this user will add a bookmark described by *bus* (which is included into *vehicle* in his/her structurable tags) then the system will suggest under the tags zone some tags included into *trademark* (as *bmw* or *ford*).

Perspectives about experimentations. We are particularly attentive to three experimentation points. First, we are attentive to the usability of the relations between tags and to the best views to manage them. We think about testing alternative visualisations by linking SemanticScuttle to visualization tools[11]. Secondly, we are attentive to the human interactions into the community around the tool. How will SemanticScuttle complete the mailing-list currently used to share resources? Will SemanticScuttle need an open system of moderation as in Wikipedia? Finally, we are attentive to the terms and relations used to organise resources. We wait for more data in order to compare structures of tags realized by teachers and taxonomies from French educational program.

Collaborative and progressive annotations into SemanticScuttle. Our objective is to transform a social bookmarking tool into a structuring one, where annotations can be easily added and then gradually structured and improved. Structurable tags described in this article are one aspect on this structuring process. We are also interested into the collaborative structuring process and negotiations between members of the community in order to reach a common vision on some parts of the domain. That's why we also added collaborative spaces into SemanticScuttle. Users can describe resources in an individual way (as in any social bookmarking tool) or they can describe resources on common wiki-based spaces in a collaborative way. Indeed we make the hypothesis that structuring process takes place at the frontier between individual and common spaces. More user observations will allow us to investigate this trend further.

References

1. Wenger, E.: Communities of practice: Learning as a social system. The Systems Thinker 9 (June 1998)
2. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American Magazine (May 2001)
3. Brooks, C., Bateman, S., Liu, W., McCalla, G., Greer, J., Gasevic, D., Eap, T., Richards, G., Hammouda, K., Shehata, S., Kamel, M., Karray, F., Jovanovic, J.: Issues and directions with educational metadata. In: Third Annual International Scientific Conference of the Learning Object Repository Research Network, Montreal (November 2006)
4. Downes, S.: Resource profiles. Journal of Interactive Media in Education 5 (2004)

---

6 Personalised search engine by Google: http://www.google.com/coop/
7 Search portal customised by a community: http://www.eurekster.com/
8 Custom bookmarking tool: http://code.google.com/p/gnizr/
9 Wikia Search engine: http://search.wikia.com
10 A demo is available at: http://festic.fr/semanticscuttle/demo/en/
11 Visualisations for Del.icio.us: http://www.solutionwatch.com/252/visualizing-delicious-roundup/

5. Friesen, N.: International lom survey: Report (July2004)
6. Guy, M., Tonkin, E.: Folksonomies: Tidying up tags? D-Lib Magazine 12 (2006)
7. Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M., Riedl, J.: tagging, communities, vocabulary, evolution. In: CSCW06, Alberta, Canada, ACM (2006) 181-190
8. Passant, A.: Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs: Theoretical background and corporate use-case. In: International Conference on Weblogs and Social Media, Boulder, Colorado, USA (March 2007)
9. Halpin, H., Robu, V., Shepherd ,H.: The complex dynamics of collaborative tagging. In: WWW07, Alberta, Canada, ACM (2007) 211-220
10. Greaves, M.: Semantic web 2.0. IEEE Intelligent Systems 22 (March 2007) 94-96
11. Mika, P.: Social networks and the Semantic Web. Springer, NewYork (2007)
12. Buffa, M., Gandon, F.: Sweetwiki: semantic web enabled technologies in wiki. In: International symposium on Wikis, Odense, Denmark, ACM (2006) 69-78
13. Trichet, F., Nizon, Y.: Une plateforme de développement d'espaces webs semantiques communautaires dediés au partage de ressources multimedia. In: Ingénierie des Connaissances (IC2007), Grenoble (2007)
14. Bateman, S., Brooks, C., McCalla, G.: Collaborative tagging approaches for ontological metadata in adaptive elearning systems. In: Fourth International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL2006) in conjunction with 2006 International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2006), Dublin, Ireland (June2006)
15. Al-Khalifa, H.S., Davis, H.C.: Replacing the monolithic lom: A folksonomic approach. In: Seventh IEEE International Conference on Advanced Learning Technologies, Nigata, Japan (July 2007) 665-669