Prototype of KOS in Web 2.0: Contributions to Innovation Activities

Manel Ankoud, Manel Hmimida, Gabriella Salzono
DICEN, CNAM
Paris, France
{manel.ankoud, manel.hmimida}@cnam.fr

***ABSTRACT*:** *This paper has two main objectives: The first objective is to present the contributions of Knowledge organization System (KOS), Hypertagging, to explain business activities and especially to bring out weaknesses and opportunities for innovation. Hypertagging is qualified as a multidimensional and distributed KOS, adopts the perspective of participatory documentary classification. The second objective is to explain different metadata designs, planned to be used in Hypertagging. There are different choices which are used to manage the lifecycles of documents and elements classification (views, dimensions, tags). So, enriched metadata are needed to improve innovation: they can detect the emergence of new concepts and practices. They also reinforce the integrity of, the critical process of enterprise environments Web 2.0, the process of document indexing composed from multiple heterogeneous systems of production, storage and dissemination of information.*

Keywords: Scientific and Technical Observation, Knowledge Organization, Metadata and Interoperability, Temporal Data, Innovation, Information Gathering

Received: 18 February 2012, Revised 9 April 2012, Accepted 14 April 2012

1. Introduction

In an information environment, the recent activities require access to increasing volumes of data, produced and managed by heterogeneous digital ecosystem. Under the impetus of Web 2.0 within enterprises, these systems include records produced, exchanged, shared among members of collectives of various sizes (teams, departments, directorates). To capitalize and focus on these new opportunities and at the same time manage this growing complexity, the Knowledge Organization Systems (KOS) "*include all types of diagrams to organize information and promote knowledge management*" [1].

As part of a project ANR-Miipa Doc "*Methods and Integrated Services for Institutional and Participatory Content Faceted Classification Documentaries Complex*", our work aims to develop KOS-based facets, for large and small enterprises [2]. The prototype of this system, Hypertagging, is qualified as a multidimensional and distributed KOS with a perspective of participatory documentary classification.

In this context, the objectives of this paper are:

- To present the contributions of KOS business, Hypertagging (HT), to intelligence activities in a Web 2.0 environment
- To explain design elements of a future version of the current prototype of Hypertagging, to generate knowledge on the dynamics and changes in classification practices, indexing and retrieval of documents within an enterprise information system.

The second objective is strongly correlated to the first, because the information covered is essential to the emergence of «weak signals» and identifies opportunities for innovation. We focus here on choices for informational support within Hypertagging developments managed documents and classification elements (views, dimensions, tags, described below).

Managing these changes is based on rich metadata for the concept of "*life cycle*", stored in a data base. This metadata will generate, throughout the indexing process shared, historical, and consistent correlations in time between views, dimensions and tags. They will answer some questions:

- What are the classification element introduced/deleted over time and what are their periods of validity?
- Since when classification elements began to be shared among users? Which membership rate?
- How has the process of organizing documents in time?

The plan of this paper is as follows. We present a succinct Hypertagging, informational and technological aspects, as well as its position in Information Systems in Web 2.0 (section 2). We indicate Hypertagging advantages to watch activities, discussing various aspects concerning the phases of sleep, waking and types of information collected(section 3).Management changes within Hypertagging, treated by associating documents and elements of a classification concept of life cycle, is shown in (section 4). Finally the paper ends with conclusions and research perspectives (section 5).

2. Description of Hypertagging

The main functions of this prototype are storage (naming, classification), treatment (research) and communication (sharing).The functions of capture and creation of documents resources are not included in Hypertagging, there documents are either created by individual tools or collective, accessible workstations as Office suite, CMS (Content Management System) ... or acquired from external systems through web browsers.

2.1 Conceptual and functional components

Structuring elements of Hypertagging are conceptual and functional graphs, which represent a GUI with which the user can classify a document by simply "*clicks*". These elements can be Views, dimensions and tags:

Views allow classifying individually or collectively dimensions. This classification is based on mental users' classifications in different contexts. "*Product*" and "*Clients*" are examples of views in a commercial enterprise. Each view contains one or more dimensions. These are similar to classes (in the sense of classification) and contain multiple values. These values are the tags, represented as icons. A dimension can appear in multiple views. It can be multivalued (the user has set for it several tags) or single-valued (limited to a single tag). Compared to the collaborative aspects, a dimension can be individual or shared. It can contribute to the naming of a document (in the case where the user chooses that this dimension name appears in the naming of the document). The description of a document is based on the tags, the values of a dimension. Tags are injected as document metadata.

Metadata is a fundamental component of the conceptual document classification [3]. They are among the main concepts of the prototype Hypertagging, necessary to support the search functions of classified documents. There are two solutions to associate metadata to resources: internally and externally methods.

The internal metadata are included in the resource file (with embedded data). This technique allows associating intrinsically metadata to the data and then transmitting them with. Even if you change your environment, you can search and find the information resource by its description. The metadata associated to resource materials are very heterogeneous. Indeed, except the Dublin Core standard, generic, adapted for all types of resources, metadata is often dependent on resource types and each standard defines its own metadata description. This heterogeneity can be seen for example between the Dublin Core, IPTC (International Press Telecommunications Council) Core for digital images and the MIME (Multipurpose Internet Mail Extensions) for email. It makes more difficult the metadata injection into a single application. The external metadata, selected by the Hypertagging prototype, can remedy this problem. Hypertagging metadata is stored in a database in two formats: relational for documents classified and XML for Hypertagging elements such as (view, size, Tag).

Hypertagging is based on three main functional components. Components Saver and Finder are dedicated respectively to the function of naming and classifying (Saver) and also to the function of research (Finder). These components require a component of metadata management. For a future version of Hypertagging, we describe the principles of metadata specification in paragraph 4. The architecture of Hypettagging is based services. These services provide a communication between different components and the database of the system and manage metadata. For reasons of speed and reuse of requests data query, Hypertagging is based on REST (Representational State Transfer) [4] type architecture. In this architecture, each Hypertagging element is identified by a URI.

2.2 Hypertagging in the Enterprise IS in the Web 2.0 environment

The information system (IS) generally refers to "*an organized set of resources including hardware, software, personnel, data, procedures ... to acquire, process, store, transmit information* (*as data, text, images, sounds, ...*) *within and between organizations* " [5].

"*Personnel "* include users, also the designers and administrators of IS. The IS can be understood on many levels, schematized by: individual (desktop), group (eg in a service or department), organizational (for all company) and interorganizational (extended enterprise to its partners, for example). These levels are increasingly connected: in the data (individual data integrated in the composition of collective data can then be released outside) or tools (messaging, databases, workflows, search engines).

2.2.1 Hypertagging Positioning

The current prototype of Hypertagging integrates with Windows file system and shared spaces to store files, such as managed using software like Dropbox.

Hypertagging is located primarily at the individual (desktop) and collective (eg within a department or department of business). At these levels, the most common tools of production and exchange of documents and information office tools, messaging, blogs, FAQs and forums, wikis and shared workspaces. The information mobilized at these levels is mostly unstructured or poorly structured, with a predominance of volatile information [6]. The information more structured and "*sustainable*", based on rules rather standardized and centralized, is generally mobilized at organizational and inter-organizational. Collaborative tools which are best suited to these levels include content management systems and EDM (Electronic Document Management), workflow and enterprise portals, databases, the PDM (Product Data Management) systems and data warehouses. Hypertagging adopts two nonfunctional requirements: usability, in terms of ergonomics and consideration of perspective of business users, interoperability, particularly in terms of coexistence and exchanges with other IS components, such as CMSs (Joomla, SPIP, ...), the GELS (SharePoint, Alfresco ...) and data servers.

2.2.2 Web 2.0 Environment

Several characteristics of our prototype resemble those of Web 2.0 environments. These environments use the web as platform [7], which migrates to applications for creating, processing, editing content, facilitating processes shared and collaborative.

According to [8], "*Web 2.0 is under development and execution of web applications that try to combine the flexibility of web technologies 'clients' classic* (*XHTML / CSS / Javascript, Flash, Java applets ...*) *with the benefits that were previously the preserve of rich clients* (*complex interfaces, access to local resources of the client, optimized network connections*)."

Hypertagging design is based on the UGC (User Generated Content). It's one of basic principle of Web 2.0 (Read-Write Web). This principle is based on services offered by sharing platforms for digital content and "*social networks*", whether generalists (Facebook, Twitter) or specialized, such as YouTube and Dailymotion video, Delicious or Diigo for favorites management.

Its open infrastructure based on the use of Web services and the generation of data flow, increases the interactions between heterogeneous applications of organizing content. Facilitating interactions between users and applications, Hypertagging contributes to interoperability and development of collaborative work focuses on the organization of knowledge. The main goal is to continue in the future this opening, to allow interactions of Hypertagging with tools of various types: mail, Web content syndication (based on RSS feeds) or bookmarking.

3. Contributions of Hypertagging Activities in Entreprise Innovation

In 1996, D. Rouach defined the innovation as "*the art to identify, collect, process, store information and relevant signals*

(*weak*, strong) that will irrigate the company at all levels of profitability, help guide the future (technological, commercial, ...) and also to protect the present and the future face of attacks from the competition.

Innovation is done in the law and respect rules of conduct. In other words, it allows the transfer of knowledge from outside to inside the company, following strict operating rules. "[9]. Using recent research on the life of the watchfulness of Web 2.0 [10], we analyze aspects of the innovation on Web 2.0 and we discuss the contributions of Hypertagging.

3.1 For which stages of innovation?
The innovation is organized as an iterative process, each iteration containing several phases. Following [10], the main tools of Web 2.0 may be involved in one or more of these phases:

• Definition of research needs: Social bookmarking

• Collect of information: Blogs, CMS, RSS, social bookmarking, social networks.

• Analysis and treatment of information: Mashups (Web applications that combine content from different sites) .

• Information broadcasting: Blogs, Wiki, CMS, RSS, Mashups, Wid gets.

Hypertagging contributions to these different phases can be summarized as follows:

• The definition of classification elements (views, dimensions, tags) and the collaborative indexing, Hypertagging contributes to phase 1 '*definition of research needs*' and phase 2 ' *collect of information* '

• Integration of documents from heterogeneous environments within a single environment facilitates phase 3, ' *analysis and treatment of information* '

• Functionality for sharing documents and filing plans among collective and individual users support phase 4, '*information broadcasting* '.

3.2 For what types of innovation?
Hypertagging targets a large majority of players in the company, who perform activities of innovation for their own work, including classifying and indexing documents in "*private.*"

Beyond these individual objectives, Hypertagging aims also "*collective*" purposes and covers "*professionals*" activities: identify emerging issues, for the detection of "*weak signals"*, the emergence of new practices and new ways of collective representation of knowledge, connecting people with similar interests.

In terms of information and communication, Hypertagging allow users to work according to local and / or shared documents and interact with two levels:

• Individual, representing their own responsibilities, procedures, steps of filing and retrieval,

• Shared, representing an area of more comprehensive information, providing this level to their own contributions.

In fact, analysis of activities realized with Hypertagging participates in the emergence of tacit knowledge.

3.3 For what type of information?
Hypertagging is integrated in technological environments which increase every day and generating information. In these environments, the innovator identifies sources relevant to their research. These may belong to databases of patents (as Espacenet, http://www.espacenet.com), databases of scientific articles (such as Science Direct, http://sciencedirect.com), to collections of gray literature (research reports, theses, as http://www.inist.fr). This can be a part also to market research. These sources can be analyzed with bibliometric tools specific to the innovation and generate new documents.With the help of Hypertagging, the innovator may classify certain resources on shared drives, by associating them with metadata, and broadcasting them to collaborators. He can also be used to identify other Hypertagging.

3.4 Which contributions?
Supporting collaborative activities of actors in a Web2.0 environment, Hypertagging is able to support innovation processes in

companies. According to [11], "*an innovation process is the set of actions needed to turn a new idea into an actual implementation with commercial potential. This is an informational process, in which knowledge is acquired, interpreted and used.*". Schematically, an innovation process goes through different "*stages*":

• Generation of ideas and potential innovations,

• Choice and validation of ideas to fruition,

• Development,

• Industrialization,

• Placing on the market.

Hypertagging is able to support the innovation process by defining, for example one or more views in correspondence of each stage of the project, with dimensions and associated tags. Thus, for example:

• A view "*regulatory intelligence*" can understand the dimensions "*France*", "*European Union*", "*World* ", in which we will define such tags "*Energy*", "*Transport* ", "*Health*"

• A view "*sleep multifaceted* " may include the dimensions: "*Market potential*," "*Competition*", in which we will define such tags "*Customer*", "*Product* ".

As Hypertagging adopts a bottom-up approach to naming and classifying documents, it helps to bring out the emergence and consolidation of views, dimensions, tags and thus the generation of new ideas. For example, analysis of tags associated with a technology innovation, can bring forth new manufacturing processes, as well as analysis of tags associated with a regulatory monitoring, to measure the impacts and opportunities associated with the latest regulations (eg on transport of hazardous materials). The elements of classification will be more representative of a group they will be shared among multiple users.

Management of change is essential to support intelligence activities, including support of the innovation process. The following paragraphs describe the informational and technical choices to meet these changing needs within Hypertagging.

4. Informational Choice for the Evolution Management

To detect elements of innovation from individuals and the collective, the metadata should describe the structural elements of what we call a "*system Hypertagging*", a term combining in one hand the documents managed by Hypertagging and in the other hand views, dimensions and tags Hypertagging. The current prototype of Hypertagging does not support these changes:

• On the documents, there has been only two dates: the first relates to the creation or last modification, the second concerns the classification of the document.

• Concerning the classification of Hypertagging elements, no usable information to generate historical is stored.

An initial conceptual development of this prototype, proposed in [12], consists in associating a URI and ranges of validity to the Hypertagging classification elements. The first is used to manage the persistence of these elements, while the second are used to calculate views or elements which are valid in absolute (individual) or group (temporal contexts).

To achieve these objectives, we introduced the concept of life cycle, combined with elements of Hypertagging classification and to the documents managed by Hypertagging.

4.1 Classes model for the evolutions management

Figure 1 shows a class diagram of this system meets the needs of evolutions. This diagram includes classes: Document and HTElement. This element specializes in classes, View, Dimension and Tag, connected by aggregations:

• A view is set of dimensions (possibly empty),

• A dimension is composed from tags (possibly empty).

Conversely,

• A tag belongs to one and only one dimension

• A dimension cannot be associated with a specific view, or be associated with one or more views.

In the following, we described the structure of MetadataDoc and MetadataEltHT,two subclasses specialized from Metadata class.

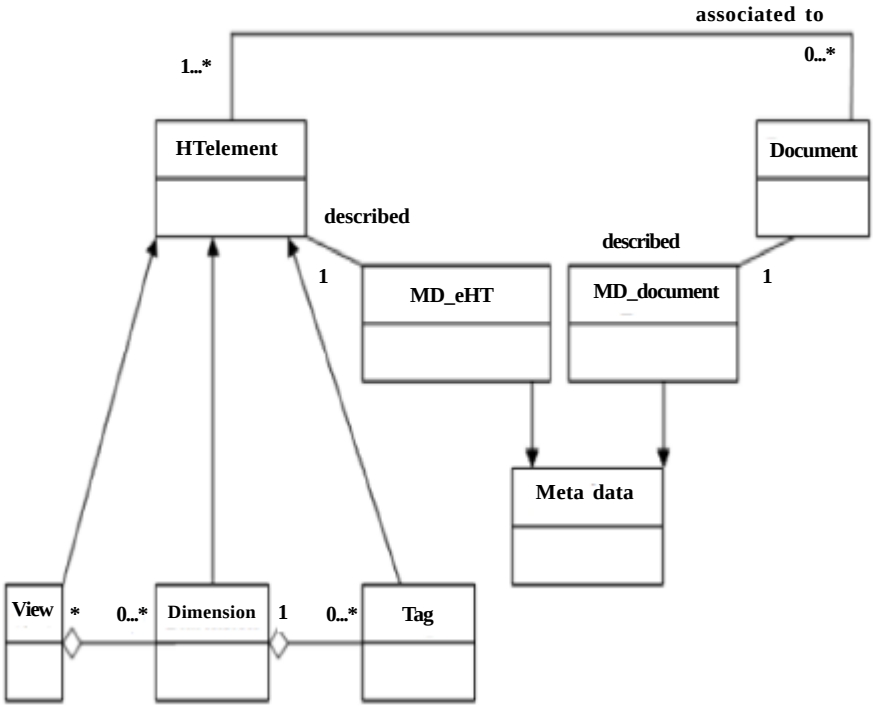

Figure 1. Class model for document classification and elements of Hypertagging

4.2 Representation of the life cycles of documents and Hypertagging

Metadata evolutions are based on the concept of "*life cycle*". This allows tracing the records management as Hypertagging elements, recording successive actions carried out on them, according to their type (eg creation, classification, change or deletion of documents, creation / editing elements classification, add / remove tags from dimensions ...). They are summarized in Figure 2.

These components are respectively:

- For documents: the URL, title, a set of keywords and, for each action on the document (creation, classification, delete, modify, add comment), the date on which this action is executed, the (or) author (s) of action and the name of the action.

- For Hypertagging elements: the URL, the name and type of the element, and, for each action on the item (creation, deletion, modification), the date on which this action is executed, author of the action and the name of the action.

The contributions of these informational metadata are multiple. They can:

- Determine the temporal contexts of validity of documents and Hypertagging elements (views, dimensions, tags). This will make it possible to detect configuration changes to the "*regulatory intelligence*", by the appearance, in the dimension "*Europe*" tag associated with the environment (such as "*Inspire Directive*") or financing ("*debt management* ")

- Evaluate the collaborative aspects of the system Hypertagging, analyzing the contributions of actors in the creation, sharing and broadcasting of materials and elements of classification, by detecting the "*leaders*" of these activities.

4.3 Support of evolution and document integrity

The approach proposed in this Section is based on distributed databases research.In [13] an approach is proposed to anage the

MD_document	MD_eHT
URL_doc	**URL_eHT**
title_doc	**name_eHT**
{[*keyword_doc*]}	type_eHT
[life_cycle_doc]	[*life_cycle_eHT*]
...	...

with

life_cycle_doc := ([create_doc], [classify_doc], [update_doc]}, [remove_doc], {[comment_doc]})

create_doc : = ({[create_*date_doc*]}, {[author_doc]})

%action_doc := ([*%action_date_doc*], [%action_author_doc], [%action])

Where%action := create | classify | update | remove | comment

and

life_cycle_eHT :=([create_eHT], {[add_eHT]}, {[update_eHT]}, [remove_eHT])

%action_eHT := ([*%action_date_eHT*], [%action_author_eHT], [%action_description])

Where%action := create | add | remove | update

Figure 2. Resources metadata

Notations: Bold, mandatory metadata, italic, metadata recommended.
The formalism {.}, [.] And (.,.,.) respectively indicate:
the repetition of an element, an optional element and a structure made up of several elements

temporal consistency of discrete objects moving in real-time systems, such as the value of stocks of financial products subject to rapid changes in time.

They introduce the notion of consistency and temporal objects (whose value changes over time). Time consistency of such an object has two aspects: the absolute consistency and relative consistency.

The first is associated with the objects individually, while the second relates to a group of objects, when read together. These concepts, which are based on intervals of validity, permit the computation of dynamic data views consistent.

The recommended changes, concerning the life cycle of documents and Hypertagging elements, extend this work to the needs of document integrity in collaborative environments. Life cycles combine temporal information to activities (grading, update, delete ...). They allow building rich historical developments of the classified documents and filing plans operated by users.

5. Conclusions and outlook

This article proposal presented the contributions to the activities of the day, especially for innovation provided by Hypertagging. Hypertagging is a prototype application allowing management of Knowledge Organization Systems heterogeneous, from the bottom up, multifaceted and distributed.

Has been described as design elements for a later version of the prototype, to take into account changes Hypertagging managed documents and items used for classification (Views, Dimensions, Tags). Managing these changes is essential to intelligence activities, based on the concepts of life cycles of documents and items of Hypertagging classification. It was clarified how the life cycles can be represented in the rich metadata.

We plan to continue this research by analyzing in detail the management of the development of collaborative indexing process in heterogeneous environments. For this we wish to build on scenarios related to the day before experimentation, in very strong trend fields, such as technological or regulatory monitoring.

References

[1] Hodge, G. (2000). Systems of Knowledge Organization for Digital libraries. Beyond traditional authority files. Washington, DC: the Council on Library and Information Resources.

[2] Zacklad, M. (2011). Miipa-Doc : Gestion de l'hétérogénéité des classifications documentaires en entreprise, Conférence Hypertextes et hypermédias". H²PTM, Université Paul Verlaine-Metz, p.12-14.

[3] Calderan, L., Hidoine, B., Millet, J., (coord.). (2008). Métadonnées: Mutations et perspectives, INRIA seminar , Dijon.

[4] Fielding, R. (2000). Architectural styles and the design of network-based software architectures, Dissertation for the degree of Doctor of Philosophy in Information and Computer Science, chapter 5.

[5] Reix, R. (2004). Systèmes d'information et management des organisations, Ed. Vuibert.

[6] Cigref. (2009). Communication Collaborative Unifiee, p.14

[7] O'Reilly, T. (2005). What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software.

[8] Pompidor, P. (2009). Évolutions récentes des technologies web du côté client. Le Web 2.0 au service des applications riches. Techniques de l'Ingénieur, Base Documentaire : Technologies logicielles Architectures des systèmes. Reference H7440.

[9] Rouach, D. (1996). La veille technologique et l'intelligence économique. Paris: PUF Coll. Que sais-je ?.

[10] Leitzelman, M. (2009). Etat de l'art et tendances sur le marché de la veille et de l'Intelligence compétitive, Rapport de Recherche, Projet ISICIL, Information Semantic Integration through Communities of Intelligence online.

[11] IMIST. (2010). Le rôle de la veille dans le processus d'innovation, La lettre de l'IMIST.

[12] Salzano, G., Ankoud, M., HmimidaA, M., Zacklad, M. (2011). Gestion des évolutions dans un SOC d'entreprise, multidimensionnel et distribué, 29' INFORSID, Lille.

[13] Kao, B., Kam-Yiu, L., Brad, A., Reynold , C., Tony, L. (2003). Maintaining Temporal Consistency of Discrete Objects in Soft Real- Time Database Systems, *IEEE Trans. Computers* 52 (3) 373-389.