

## Personalizing Semantic Information Networks

A. MKADMI<sup>1</sup>, E. REYES<sup>1-2</sup>, N. BOUHAÏ<sup>1</sup>

<sup>1</sup>Laboratoire PARAGRAPHÉ, Université Paris8, 2, rue de la Liberté 93526 - SAINT-DENIS cedex 02

<sup>2</sup>Departamento de Comunicación, Tecnológico de Monterrey Campus Toluca, Eduardo Monroy 2000, 50110 Toluca, México

[abderrazak.mkadmi@univ-paris8.fr](mailto:abderrazak.mkadmi@univ-paris8.fr) ; [everardo.reyes@itesm.mx](mailto:everardo.reyes@itesm.mx) ; [nasreddine.bouhai@univ-paris8.fr](mailto:nasreddine.bouhai@univ-paris8.fr)

### Abstract

*In our present days, documents on the web are comprehensible both by humans and machines. This has been possible with the development of XML as a language for data structuring and exchanging. XML separates structure from content in documents and defines metadata in more precisely ways. This article deals with the problem of creating and recovering information spaces from the Web, which commonly are not enough structured and, thus, not well exploited. We introduce a description of the HyWebMap system as a tool for profiting the flexibility and extensibility of XML when working with structures and semantics.*

### Keywords

*Structured documents, information spaces, XML language, XML Schemas, Web semantics, XML databases.*

### 1. Introduction

As a result of norms and standards appeared these last years, we start to have a large amount of texts and multimedia that represent documents by its content and its logical structure. These types of documents are richer than the ordinary ones because they allow us to attach logical information, relationships between information elements, metadata and links to external information sources to the net content. We have seen that there are just a few tools able to handle the new representations and exploiting its richness.

Considering the previous frame, and from an university perspective, members of Laboratoire PARAGRAPHÉ have developed HyWebMap <sup>1</sup>

that could be defined as a system for creating, navigating and structuring personalized information spaces <sup>2</sup>. By the use of HyWebMap we would like to come up with three questions, the first one is linked to common problems of hypertext systems (e.g. disorientation and cognitive excess), the second one refers to the hypertextual writing, and the last one, treated on this article, has a link between research and information exchange. Our goal is double, in one hand, to confer a better structure to documents on the Web, in order to provide a practical solution for ever-strictly needs in the field of research, navigation and web documents, and, in the other hand, to exploit XML technology for evolving into a new web, where documents are to be more comprehensible for humans and for machines.

### 2. What is HyWebMap?

HyWebMap is a system for structuring and navigating documents through the web. It was created in a context of developing didactical hypermedia systems. It is used in several domains, from education and web monitoring to distance learning. Its constitution has five levels:

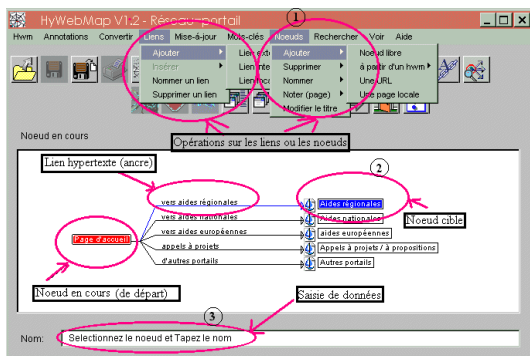
- 1- a monitoring agent for virtual networks, which allows users to have updated their document links,
- 2- a tool for getting nodes, which permits the user to create his own navigation course and renaming the nodes and links of a document,
- 3- a tool for assisting the navigation, which offers an advanced research by dates, tree structure, history, etc.,

<sup>1</sup> For more information about HyWebMap (Hypertext Web Mapping), refer to: <http://h2ptm.univ-paris8.fr/hywebmap/docs/>

<sup>2</sup> A personalized navigation space is an information space where we can modify the navigation course, rename the links between documents, annotate the documents and eventually attach descriptors or keywords.

- 4- a tool for generating virtual networks, created or downloaded in HTML format, which facilitates its view in most popular web browsers,
- 5- and finally, a tool for personalizing documents using comments, annotations and personalized links.

HyWebMap includes a graphical interface [figure 1] where different functions can be accessed from a group of elements distributed in the software.

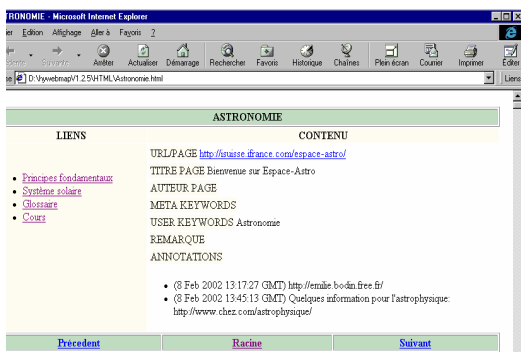


**Figure 1. Main HyWebMap Interface [Bouahi02]**

The system's functions respond to a bidimensional perspective [Papy01]:

- A reader dimension by the behavior of agents, in charge of searching the asked information and gathering it into the user's informational space;
- An author dimension by all of the operations that users may apply to the collected information for example: adding new links and nodes, attributing keywords and comments to the information nodes.

Figure 2 shows the HTML version generated from HyWebMap, created or downloaded by the user.



**Figure 2. Interface de génération d'une version HTML à partir d'un réseau HyWebMap**

### 3. Structuring Information: Network documents

Beyond the notions of Virtual Personnel Documents (VPD), that “could be considered as an assembly of elements (also known as fragments) associated to filter, organization and constraint set mechanisms... respecting a user model and narrative principles”. [Garlatti99] or Pedagogical Objects [Bourdoncle00] we are witnesses of an all-new type of document called “network documents”, a kind of type addressed at the convergence of structured documents and network hypertexts [Clement95]. Internet is the best illustrative example of them.

These documents are the result of constructing knowledge spaces on the web [Bouahi02]. We have to consider them as atomic entities—the same way we do with HTML pages- but the difference is that they are made out of several fragments. Each fragment cannot pretend to have its own autonomy, instead, all together give a coherent and semantic structure to documents where they are attached. Finally we are assisting to a new reality associated to the existence of the Net.

It is about to evidence a type of document in phase of documentary practices, unknown until now since the web has become a place for valid and pertinent information researches, as remarks [Dkaki99] “... it is useless to say that the web has become a highly-potential information source with strategic applications. Just remember the great methods for managing information that are insufficient for an optimal exploitation, essentially related to information access needs...”.

In the practice, this constant has been undeniable overpassed by the proliferation of filtering systems [Ben Hamadou01] with tendency to separate good grains and by unedited searching engines such as NeuroWeb [Lelu99] whose informational matter comes directly from the web. At his point, specific practices inherent to reading electronic information [Morizio99] and to its “cognitive apprehension” by the users [Hudon01] were derived. [Ferret01] says that practices of searching information are not going towards a research of “LA”, but towards a constitution of convergent documentary references that are being put in evidence by some monitoring tools and portals.

### 3.1. From “network documents” to “semantic information spaces”

Having network documents more semantic has implied several phases: structure, portability, exchange and access to documents. Briefly, the evolution goes from a syntax-based web, where only a human being is capable to use the information, to a semantic web [Berners-Lee01] where the information has an explicit meaning that allows machines to get a more active role for treating, describing and structuring information, leaving behind localization, transferring and presenting tools.

This evolution is supported on three concepts:

1. A richer representation of information, more structured and more rigorous, an essential condition for programs to work on the content,
2. a data model for validating all information networks created or downloaded,
3. metadata, a well-formalized external description of main information.

### 3.2. A logical structure for XML documents?

In the measure that XML is used as a language for exchanging data, it seems important to have a document model for understanding exchanged data. To create such model there are two recommendations available: the DTD and the XML schemas. The first one has been used for a long time and has been recognized in all applicative domains. The second one is more recent and has the purpose to overwhelm the deficiencies of DTDs, it is related to data types, use of language and handling name spaces [Mkadmi03].

We have chosen the XML schemas to develop our model, according to the following reasons:

- XML schemas use XML language, which facilitates data manipulation with the same tools and methods of XML documents,
- XML schemas contain a large number of data types such as Booleans, natural numbers, time intervals, etc. Additionally, it is possible to create new types by adding constraints on an existing type,
- XML schemas handle name spaces, which avoid confusion between elements coming from different sources,

Our document model is useful to define coherence among documents, which may be used by any informatics application that defines nothing but the mentioned model. It is about a grammar that allows checking out the conformity of XML documents. This grammar contains the tag names that we can use in an XML document, and its possible implications. It includes the data types contained in these elements, too. All of these advantages are important for saving time and money, as well as the reliability of collaborative works.

This XML schema allows to validate the HyWebMap networks created or downloaded, what is helpful for assuring documents homogeneity and, at the same time, for collaborative working based on a predefined structure. The HyWebMap network schema has two parts:

- The first one defines network metadata (title, author, description, categories, creation date, modification date and network reading)
- The second one, presents network content (the nodes). Each node has a title, an author, an URL, keywords, comments and annotation space.

```
<HYWEBMAP>
<HEADER>
<IDENTIFIANT>1</IDENTIFIANT>
<TITRE>XML & Travail collaboratif </TITRE>
<DESCRIPTION>Ce réseau représente une
étude sur les SRCIs utilisant
XML</DESCRIPTION>
<AUTEUR>M. Langlois </AUTEUR>
<CATEGORIE>structuration des
données</CATEGORIE>
<DATES CREATION="2004-01-17"
MODIFICATION="2004-02-10"
LECTURE="2004-02-10"/>
</HEADER>
</HYWEBMAP>
```

Figure 3. Network metadata

### 3.3. Information network metadata

Metadata is commonly defined as data about data. Precisely, it is about a structured information assemble describing a resource. The metadata elements that we have defined are [figure 3]:

- Identifier: a network identifier automatically endorsed by the system,
- Title: network title,
- Description: an abstract describing the network,
- Author: network author,
- Category: network general matter,



automatique de résumés. – in CIDE 2001, Toulouse, 23-25 October 2001.

**[Berners-Lee 01]** BERNERS-LEE T., HENDLER J., LASSILA O. - The Semantic Web. – in :*Scientific American*, May 2001, p35-43.

**[Bouhaï02]** BOUHAÏ, N. – Lire, Réécrire et partager le savoir sur le Web : problèmes et solutions, Ph.D. Thesis in Information Sciences, Université Paris8, 2002

**[Bourdoncle00]** BOURDONCLE F., BERTIN P. - Recherche d'aiguilles dans une botte de liens, La recherche, n°328 February 2000.

**[Clement95]** CLEMENT J. – Du texte à l'hypertexte : vers une épistémologie de la discursivité hypertextuelle. – in: « Hypertextes et hypermédias : Réalisations, outils, méthodes », Hermès, Paris, 1995.

**[Dkaki99]** DKAKI T. – Collecte, prétraitement et traitement des informations issues du Web dans un environnement coopératif. – in « Solaris », n°5, Presses Universitaires de Rennes, 1999.

**[Ferret01]** FERRET O., GRAU B., HURAUT-PLANTET M., ILLOUZ G., JACQUEMIN C. – Comment trouver « LA » réponse. – in « 3<sup>ème</sup> congrès du chapitre français de l'ISKO, 5-6 July 2001, Paris, pp. 159-168.

**[Garlatti99]** GARLATTI S., IKSAI S. - Documents virtuels personnalisables pour des systèmes d'informations en ligne, IHM99, Montpellier, 1999.

**[Hudon01]** HUDON M. – Structuration du savoir et organisation des collections dans les répertoires du Web. – in : BBF, Paris, T. 46, n°1, 2001, pp. 57-62.

**[Lelu99]** LELU A., HALLAB M., RHISSASSI H., BOUYAHI S., BOUHAÏ N., HE H., QI C., SALEH I. – Projet NeuroWeb : un moteur de recherche multilingue et cartographique. – in : H2PTM'99, 23-24 September 1999, Paris.

**[Mkadmi03]** MKADMI A., BOUHAÏ N., LANGLOIS M. – Partage des modèles XML: une solution pour les échanges électroniques professionnels, Conférence « Journées Francophones de la Toile » (JFT'2003), Tours, 30 June, 1 & 2 juillet 2003, École Polytechnique de l'Université de Tours.

**[Morizio99]** MORIZIO C. – Zapper, chercher, lire des documents électroniques. – in : BBF, T44, n°5, 1999, pp.48-51.