

# Design of a Content Management System Based On the Use of Metadata

Kalonji Kalala H<sup>1\*</sup>, Mbuyi Mukendi E<sup>1</sup>, Mulopo Tona D<sup>1</sup>, Kabengele Mpunga Y<sup>1</sup> Amisi Fataki C<sup>1</sup>, Galekwa Napifongo R<sup>1</sup>,

<sup>(1)</sup> Department of Mathematics and Computer Science/ University of Kinshasa, DR Congo

## ABSTRACT

In this paper we present the design of a content manager (CMS) based on the use of metadata. This article will help us to understand the development of software for content management and the various constraints and resources used for this kind of development. This design will allow us to see how the development of a practical application on the data in the web can be done, and to take maximum advantage of the benefits of the use of metadata in content management: indexing, integration, organization, storage and retrieval of objects (documents).

**Keywords:** archiving xml, content management, indexing, metadata, website

## I. INTRODUCTION:

A content manager represents a complete system for the storage, processing and information security. New horizons open to the use of computer data, revealing a new technological environment where some problems can be solved. All actions are performed from a web browser and require no special technical skill. Dedicated to a website, intranet or extranet, the content management tool allow to administer the site interactively.

On the other hand, the Content Manager System (CMS) are described as repositories which can also include the creation, sequencing and content aggregation tools, with a goal of simplifying the creation and administration online content. [8] From the architecture view, a content management system consists of a front office and a back office or administrative part.

The main feature of a tool such as a Content Manager System (CMS) is to separate the contents from the container. Thus, it is possible to update independently of each other. In this work, we will design a CMS.

The rest of the paper is organized as follows: Section I: The management of content, we will present an overview of content management, Section II: Metadata and Section III: Development of a CMS.

## II. THE MANAGEMENT OF CONTENT

### 1.1 A content

The content is all information and knowledge available to the visitor to a website. More generally, the content is a set of documents, multimedia or various data components. The internet technology

provides this content online or allows downloading it. The content can take many forms, structured with a database or unstructured such as documents. Web data can be classified as data content, data structures, data from the user profile and usage data. Web data is a collection that describes the use of web resources. [4 -5 -6].

### 1.2 Static site

Managing static pages were problematic when the number of pages on the site became important in number, difficult to maintain relationships, tedious capture content for common pages, and difficult reorganization of the site during the evolution of the architecture requiring computer knowledge and programming skills [5-6]

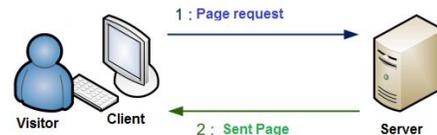


Fig.1: Treatment of a static page from the server

### 1.3 Dynamic Site

A web page is an HTML document. The web pages constituting a web site may be arranged in width and in depth, or the combination of the two modes, ie, indicating the structure or the website topology [9]. The revolution was to organize sites through the use of databases that link to pages written in scripting languages such as PHP, ASP, JSP, PERL ... Dynamic websites have more advantages than static sites that their contents are generated based on input from the user or requests or system. And before any display, entries (or applications) are considered to be optimal responses. The image below shows how the demand of a page is first treated prior to display in the client browser. [4-5-6]

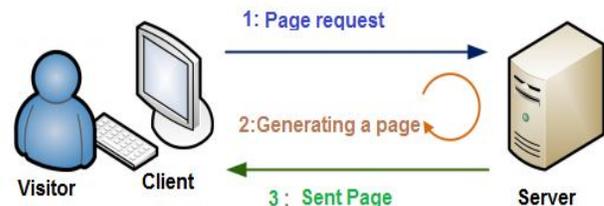


Fig.2: Treatment of a dynamic page

Also called tools of “Content Web Management (CMS)”, these solutions allowed a real separation of the contents and container, in other words “to separate the page layout (presentation) and the documents, the texts, images... (Contents).

The advantages of these solutions are: A preserved graphic unit; An automatic generation of the menus of navigation and bonds; A management of the structured and not structured contents (documentary database); A separation of the roles of each speaker (authors, validator, developer, administrators...); A decentralization and devolution of the update of the contents; Possibility of organizing, of planning and of automating the production of contents [4-5]

The image below shows the contribution of the content management system to manage content with separation of presentation and content

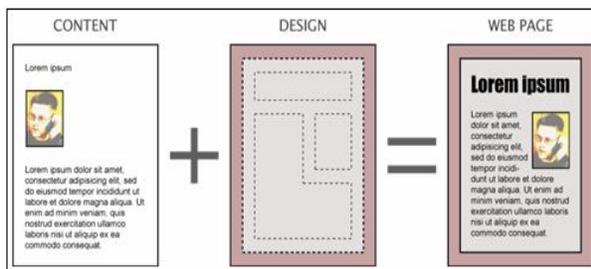


Fig. 3: Separation of presentation and content

To logically structure the contents of the documents so as to separate the content and the form, XML (Meta languages specialized markup) appears as the most appropriate way to describe both the metadata and the structure of documents in order to ensure the sustainability of the electronic documentation. Companies use the Web as a means of disseminating information, and XML has become the de facto standard for representing and exchanging data over the Web. Information available over the web is extremely diverse and different levels of quality (eg, accuracy, etc...) and security [1]

## 2. Issues a CMS

### 2.1 Reduction of costs

The same content can then be displayed in various forms and presentations. CMS management tools use what is called "templates" that are shaped skeleton for the presentation of content that prevents companies to rework the integration of HTML pages. In particular, in the case of multi-language websites, we see the need to use a CMS tool is essential.

### 2.2 Reducing the time of publication

The publishing process can be decomposed into two steps: selection of content and formatting. The selection of content often use tags or filters that allow you to choose the information to be

presented, the formatting is based on templates that define how content should be presented in the pages and what are the indications presentations (typography, alignment ...).

### 2.3 Increase the added value of information

Whatever the type of site (institutional, E-commerce, portal), content is the most effective generator of internet traffic. The added value of a site depends on the relevance and quality of information found there, freshness and ease of access to content.

### 2.4 Formalism and compliance business processes

Because of the separation of content and container, the CMS allows compliance of business processes in the organization where it is established. It is based on circuits performing tasks or validate information before publication. These circuits are commonly called "Workflow" and are based on a set of rules and actions.

### 2.5 Decentralization updates content without technical knowledge

Each contributor and approver can edit and validate the part of the site which it is responsible. There are no geographical or temporal limitations. Access is done with a simple web browser requiring no special technical skill.

## 3. The main functions of a CMS

### 3.1 The collaboration and version control

CMS tool will allow more people to participate in the creation of content in a common goal. A complete management of rights is necessary for several stakeholders can share the content update. Version control keeps the old versions so you can return to an earlier stage.

### 3.2 Planning and content validation

The supply of content is decentralized, ie delegated to a large number of participants. So it is important to maintain control over the content. Content validation then follows a process called "workflow".

### 3.3 User management and access control on the content

A content management software must support a number of users with different projects. The government allocates rights. Rights management allows to define the roles of each user, whether editor, validator or administrator on the site.

### 3.4 Integration of external data

A content management system must be integrated into the information system of the organization. It must integrate: One or more external databases, a directory, an XML-based management of large volume of data.

## **4 Architecture of a CMS**

The architecture of a CMS usually has three types of components: The presentation features (or Human / Computer Interface); Treatments and data access

### **4.1 Content optimization policy with a cache**

The purpose of the cache systems is to avoid recalculating the same content multiple times.

This applies for example: An html page, loading the configuration settings of the site, information on the current visitor, the result of a search in the database

## **III. METADATA**

Metadata is data used to define or describe other data and can be used in the management of the data they describe. Metadata provides systematic, understandable and deductive information on the content, the structure, the relationships, the representation and the context of use of a particular data set. [2, 3].

Metadata can appoint, designate, describe or categorize the various resources in a "readable and understandable manner" by machines such as humans to make them identifiable, usable and sharable. They therefore make the management of document databases (identification, storage, indexing, search ...) possible.

### **1. Importance of metadata**

Facilitate the search for information: Describe the contents and relationships between files of the resource; better reference the resource; classify content according to degree of difficulty or target audience.

### **Facilitate interoperability**

Share information to maintain and enrich the collective memory training, exchanging information among computer systems, among sites and among institutions.

### **Facilitate management and archiving**

Inform on the documents life cycle; manage collections of resources, manage electronic records.

### **Sustainability**

Minimize redevelopment resources despite changes in technological environments.

### **Interoperability**

Enable the use of a resource from a management system to another, from one institution to another with a different technological environment.

### **Collaboration**

Exchange, syndication, sharing, co-production, enrichment of educational resources.

## **Reusability / adaptability**

Recycling educational resources for different purposes, in different environments and in different contexts, etc...

Metadata is an addition to the semantic content of published documents. They are increasingly important for efficient and effective research, as well as filtering and content management. The creator of each document posted on the Web is best placed to add the information constituting the metadata. The producer or the site editor is concerned, in particular, to ensure compliance with metadata standards through harmonization, the semantic and syntactic vocabularies between the standards. [2, 3, 13]

## **2. Types of metadata**

Metadata are classified according to their utilities. There are several types of metadata. There are descriptive metadata, management, conservation and technical metadata.

Depending on the goals, we can also have other types of metadata, for illustration, the metadata may change during the life cycle of the resource and others that do not change.

There are also internal metadata to a resource and others that are external to a resource. As they are external to a resource, they may be stored in a database.

[Pereira, F. et al., 2008] Pereira, F. et al. all provide a list of types of metadata: metadata about the content, identification metadata and localization of content, interaction metadata with the user, metadata of content management, metadata about the context of the user. [12]

## **IV. APPLICATION**

This manager will aim to create, publish articles, which will be indexed to allow a good search content by keyword and then managed by an administrator so that the front office part appears as a portal, the data entered from the back office. Our content management system named "CMS BKG" is considered as a two parts site, one for anonymous users and back office part reserved for various persons involved in the management of information. For information management, we mean the creation, publication and validation of information, and user management by the software.

It should be noted that this article is an elaborate scientific work and all the complete architecture of the application and the source code and some comments will not be included in this article for reasons of compliance and space.

### **Development environment**

We used the following software to achieve our content manager:

- Php version 5
- Mysql Version 5



Fig.4: CMS front office

```
<?php
if(file_exists(ROOT.'/user/config.xml'))
{
    $config = new DOMdocument();
    $config->load(ROOT.'/user/config.xml');
    $default = $config->getElementsByTagName('default')->item(0)->nodeValue;
    $login = $config->getElementsByTagName('login')->item(0)->nodeValue;
    $password = $config->getElementsByTagName('password')->item(0)->nodeValue;
    $sitename = $config->getElementsByTagName('sitename')->item(0)->nodeValue;
}
else
{
    die('Le fichier \' Config.xml n\'existe pas !');
}
define('DEFAULT_PAGE', $default);
define('PASSWORD', $password);
define('LOGIN', $login);
define('SITENAME', $sitename);
global $id;
```

Fig.5: Source code of config.php file

- Include / initialisation.php: is considered like a controller, because it is the file that the vast majority of functions are done.

```
<?php
require_once ROOT.'/include/session.php';
require_once ROOT.'/include/url.php';
require_once ROOT.'/pages/index.php';
require_once ROOT.'/include/pages.php';
require_once ROOT.'/include/strings.php';
$session = Session::getInstance();
global $id;
$id = analyze_url();
$pages = Pages::getInstance();
$index = Index::getInstance();
if(!empty($_POST) and $_POST["inscript"]!=1)
{
    require_once ROOT.'/include/processing.php';
    data($_POST);
}
if (!empty($_POST) and $_POST["inscript"]==1)
{
    require_once ROOT.'/include/inscription_User.php';
    inscription_user($_POST);
}
if(!empty($id))
{
    $result = $pages->charger_page($id);
    if(!$result)
    {
        $pages->charger_page(DEFAULT_PAGE);
    }
}
else
```

```
{
    $pages->charger_page(DEFAULT_PAGE);
}
require_once ROOT.'/include/theme.php';
Theme::affiche_zone_theme();
Theme::charger_index();
```

Fig. 6: Code source du fichier initialisation.php

We present our various files that enabled the realization of our CMS



Fig. 7: Overview of the structure of the site

The file "param.php" contains the configuration parameters such as server name, the version of PHP used, the instance name (site folder), the registration ID to collect data from users connected.

```
<?php
session_start();
session_regenerate_id();

define('VERSION', '0.5');
define('PATH', '/bahambula');
define('DOMAINE', 'http://'. $_SERVER['SERVER_NAME']);
define('ROOT', $_SERVER['DOCUMENT_ROOT'].PATH);
define('NL', "\r\n");
?>
```

Fig. 8: source code of param.php file

The "index.php" file is our startup file, it allows you to load two configuration files with "config.php" and "initialisation.php" located in the include directory and "param.php" file.

```
<?php
require_once 'param.php';
require_once ROOT.'/user/config.php';
require_once ROOT.'/include/initialisation.php';
?>
```

Fig. 9: Source code of index.php file

The file "login.php" is used to connect to the administration section called "Back office". This file calls the methods of the class "session" of the file located in the include directory to check the connection status in an open session and returns a form of identification for people with access to the back office (administrative) Site.

```

<?php
require_once 'param.php';
require_once ROOT.'/user/config.php';
require_once ROOT.'/include/session.php';
$session = Session::getInstance();
$result = '';
if($session->isConnected() && empty($_POST))
{
    $session->deconnexion();
    header('Location: '.PATH);
    die();
}
if(!empty($_POST))
{
    $result = $session->connect($_POST['login'], $_POST['password'], false);
    if($result == true)
    {
        header('Location: '.PATH);
        die();
    }
}
include ('theme/form_login.php');
?>

```

Fig. 10: Source code login.php page

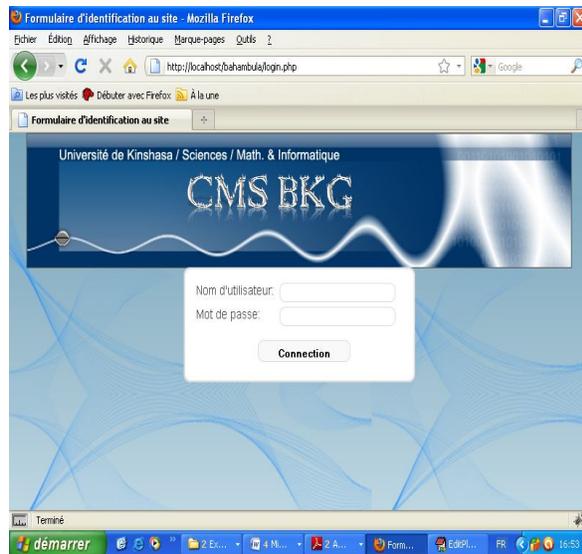


Fig 11 : Authentication form

The "inscription.php" file is the file to register in the administration section called "back office" to create an account on the site.

This file calls the methods of the class "session" of "session.php" file located in the include directory to check the connection status in an open session, and return a registration form included in the "form\_inscription file form. php" in the theme directory for people who can access the back office (administrative) of the site.

```

<?php
require_once 'param.php';
require_once ROOT.'/user/config.php';
require_once ROOT.'/include/session.php';
$session = Session::getInstance();
$result = '';
if($session->isConnected() && empty($_POST))
{
    $session->logout();
    header('Location: '.PATH);
    die();
}
if(!empty($_POST))
{
    header('Location: '.PATH);
}
include_once ('theme/form_inscription.php');
?>

```

Fig. 12: Source code of the inscription.php page

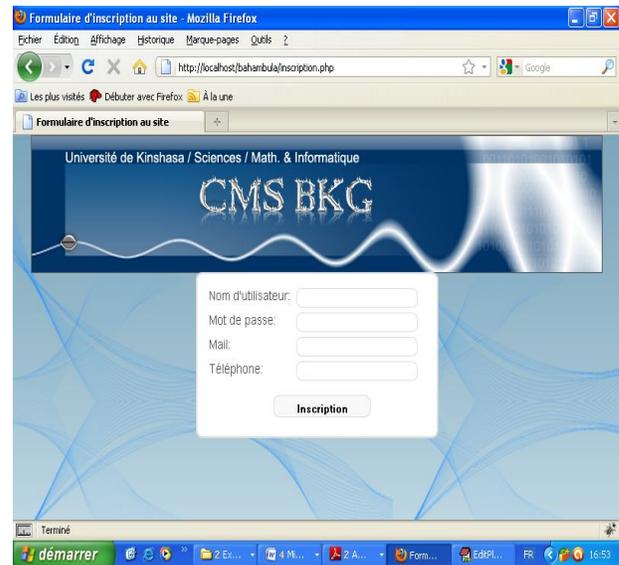


Fig. 13: Registration form

## V. CONCLUSION:

We have developed a content manager (CMS) based on an approach using metadata, in that we have helped in solving problems related to the creation, publication, archiving, validation and optimization of information retrieval for business and even for a particular. The content manager that was created is good. It includes features such as the registration of persons can access the administrative side, the creation of new content, removing content and modification of files stored in XML files and data retrieval based on key words. In the future, we plan refine this approach based on metadata, including assigning weights of importance (or applying techniques pairings) to contents to further improve the search for information.

## REFERENCES

- [1] SungRan Cho, N. Koudas and D. Srivastava, "MIX: a meta-data indexing system for XML", VLDB '05 Proceedings of the 31st international conference on Very large data bases, 2005.
- [2] D. Gorga " Métadonnées et XML – une initiation", TECFA, 2003.
- [3] P. Peccatte, " Metadata: an initiation, Dublin Core, IPTC, Exif, RDF, XMP", first version: August 2002, Last update: December 13, 2007. <http://peccatte.karefil.com/Software/Metadata.htm>
- [4] N. Tyagi, A. Solanki et M. Wadhwa. "Analysis of Server Log by Web Usage Mining for Website Improvement", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 8, July 2010.
- [5] P. Lahaye, "Les systèmes de gestion de contenu : description, classification et évaluation", MEMOIRE, Conservatoire National des Arts et Métiers, Paris, 2004.
- [6] K. Umeda, T. Ejima, and H. Nozaki, "The Development of School Websites Management System and Its Trials during School Field Work in a Distant Place", IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.11, November 2006.
- [7] "Content Management System", Programme Régional d'Actions innovatrices, Septembre 2008. <http://www.ardesi.fr/IMG/pdf/CMS.pdf>,
- [8] Khaled S. Husain " Main Requirements of Content Management System and Evolution Criteria in the University Environmental: Case Study of MARZ System of King Abdul-Aziz University" IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.5, May 2011.
- [9] R.Prasanna, M.Venkata ,P. Srinivasa ,L.Mohana and .D.Lakshmi, "An Approach for Restructuring of Web Pages", IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.2, February 2010.
- [10] Y. Nicolas, " Metadata for the French numerical theses ", 9th International Symposium on Electronic Theses and Dissertations June 7 – 10 June, 2006, Québec City / Québec, Canada.
- [11] C. Mohamad, B. Nabil, "The use of the metadata in the description and the research of the teaching resources on the WEB". C. Mohamad, B. ABADALLAH, "L'usage des métadonnées dans la description et la recherche des ressources pédagogiques sur le WEB", Colloque Veille Stratégique Scientifique & Technologique, 2004.
- [12] [Pereira, F. et al., 2008] Pereira, F.; Vetro, A.; Sikora, T. (2008); Multimedia Retrieval and Delivery: Essential Metadata Challenges and Standards. Proceedings of the IEEE, vol.96, no.4, pp.n1-744, April 2008. doi:10.1109/JPROC.2008.916384 URL:<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=44n087&isnumber=44n076>
- [13] Mama Amar, "Métadonnées et processus pour l'archivage de données médiatiques", mémoire présenté comme exigence partielle de la maîtrise en informatique, Université du Québec à Montréal, Mars 2012.