

RAPPORT DE STAGE

Création d'un CMS sur mesure

Entreprise : Jean-Loup Rolland



Présenté par Rémy Drillon

17/06/2011

Tuteur enseignant : Pierre Maret

Tuteur entreprise : Jean-Loup Rolland

Table des matières

Table des matières	2
Remerciements	4
Introduction.....	5
Présentation de l'auto-entreprise Jean-Loup Rolland.....	5
Objectif	5
Qu'est ce qu'un « CMS » ?.....	5
Qu'est ce que la « SEO » ?	5
La mission	6
Présentation	6
Principales fonctionnalités	6
L'existant.....	7
Les exigences	7
Sécurité et restriction d'accès	8
Description technique	9
Présentation	9
Sécurité.....	10
Structure du CMS	13
Le panneau d'administration	14
Le fil d'actualité	14
Les rubriques	14
Les perspectives	17
Amélioration à apporter	17
La maintenance	18
Les coûts	18

Structure de la base de données.....	19
Table des illustrations	20
Bibliographie	21

Remerciements

Je souhaite remercier Jean-Loup Rolland, dirigeant de l'auto-entreprise Jean-Loup Rolland, pour son accueil et son soutien lors de la durée de mon stage au sein de son entreprise.

Je remercie également Pierre Maret qui a accepté d'être mon enseignant tuteur durant cette expérience professionnelle.

Introduction

Présentation de l'auto-entreprise Jean-Loup Rolland

L'objectif de l'AE (Auto-Entreprise) est de conseiller sur les techniques et les moyens du web, c'est-à-dire qu'elle conseille une entreprise sur un projet pour le développement d'un site ou d'une idée de développement web. La tâche de l'AE est de définir le projet en consultant ou proposant des personnes ou agences du réseau de l'AE, pouvant répondre au développement du projet en fonction des besoins et du budget.

Objectif

Ce document décrit la mission réalisée durant mon stage. Le but est d'offrir au lecteur une bonne compréhension du système développé ainsi que les difficultés qui ont dû être surmontées. La mission consiste à réaliser un CMS qui permettra à l'entrepreneur de gérer et de proposer un service aux utilisateurs.

Qu'est ce qu'un « CMS » ?

Un *système de gestion de contenu* (SGC, FR) plus connu sous le nom de CMS (*Content Management System*, EN) est un site internet doté de fonctionnalités de publication et d'une interface d'administration.

Qu'est ce que la « SEO » ?

La SEO (*Search engine optimization*, EN), pour *optimisation pour les moteurs de recherche*, vise à améliorer la compréhension du contenu d'un site internet et son référencement en regroupant un ensemble de techniques qui apporteront un maximum d'informations sur la page internet aux robots d'indexations des moteurs de recherche.

La mission

Présentation

Le CMS se veut être un système facile d'utilisation et intuitif. Il doit permettre à l'utilisateur d'avoir un maximum d'information sur les diverses rubriques mais également sur les actualités en parallèles. En arrière plan, le CMS doit être paramétrable par l'entrepreneur de façon simple et efficace.

Voici un bref exemple qui permet de comprendre le type de service que peut rendre le CMS. Admettons que l'entrepreneur met à disposition de l'utilisateur une rubrique « restaurant » et une autre « hôtel » d'une région touristique, chaque rubrique est constituée de fiches d'information sur des établissements par le biais desquelles l'utilisateur aura accès des descriptifs. Les données des fiches des deux rubriques sont croisées, cela permet à l'utilisateur de faire un lien et de trouver les établissements proches les uns des autres pour pouvoir par exemple connaître tous les moyens de restauration à proximité d'un établissement hôtelier.

Principales fonctionnalités

Cette partie présente de façon exhaustive les fonctionnalités du CMS.

Fil d'actualité

Ce contenu est généré de deux façons différentes, la première par des informations émises par l'entrepreneur, la seconde par des « Flux RSS » mise en place par l'entrepreneur.

Rubriques spécifiques

Au final, le CMS sera déployé sur un thème. Celui-ci sera composé de plusieurs rubriques qui représentent chacune une partie du thème. Les rubriques sont constituées de fiches descriptives sur des établissements.

Création de page « à la volée »

A l'aide de mot-clé, de statistiques produites par les utilisateurs et/ou avec d'autres dispositifs, certaines pages seront générées afin de regrouper plusieurs informations (rubriques, actualités) sur un ou plusieurs mots-clés par exemple.

Panneau d'administration

Il est indispensable que le CMS possède un panneau d'administration, afin de gérer les rubriques, les fiches, les avis, le fil d'actualité et d'autres paramètres.

L'existant

Divers CMS existent déjà, comme WordPress, Joomla, Drupal, SPIP, XOOPS, qui proposent pour certains la fonctionnalité de fil d'actualité à partir de Flux RSS (exemple WordPress), mais ont la fâcheuse habitude d'être assez lourd. Mais ont la fâcheuse tendance à être assez complexes à mettre en place et à utiliser pour une personne peu expérimentée et sont aussi parfois surchargés de fonctionnalités qui ne répondent pas spécialement aux besoins de l'administrateur et qui peuvent ralentir le fonctionnement.

La plupart des CMS nommés précédemment peuvent donc ne pas répondre complètement aux critères sélectionnés par l'entrepreneur. Celui qui se rapproche le plus aux exigences est Drupal. Il est simple d'utilisation, l'interface de l'utilisateur est paramétrable, il ne nécessite aucune connaissance technique particulière pour son administration. Il est également possible de rédiger ses propres modules, mais l'inconvénient est que ceux-ci dépendent du CMS, ce qui nécessite des mises à jour régulières afin de garder un système fiable et sécurisé et une adaptation de ces modules personnalisés à ces mises à jour si besoin.

Les exigences

Le CMS que l'on souhaite développer doit quant à lui être simple d'utilisation autant pour l'internaute que pour l'administrateur. Il faut que l'utilisateur lambda comprenne du premier coup son fonctionnement lors de sa première visite.

Côté administration, il faut réaliser un service pratique et rapide pour l'administrateur avec un minimum de compétences en terme de gestion.

Le site doit être autonome une fois paramétré et ne demande qu'un minimum de maintenance et de modération, ainsi qu'une insertion rapide du contenu pour ne pas perdre de temps.

Sécurité et restriction d'accès

L'accès au panneau d'administration doit être sécurisé par un identifiant et un mot de passe. Afin d'accroître la sécurité, le chemin d'accès à ce panneau ne doit être connu que par l'administrateur.

Toute donnée saisie dans des formulaires doit pouvoir éviter l'injection de code afin de se préserver contre la récupération de données. Un formulaire ne doit répondre qu'à sa fonctionnalité première.

Description technique

Présentation

MySQL

La base de données MySQL est devenue la base de données open source ¹ la plus populaire au monde grâce à sa haute performance, sa fiabilité et sa simplicité d'utilisation. Beaucoup des sociétés les plus importantes et à forte croissance telles que Google, Lafarge, EADS, Alcatel-Lucent, Nokia et YouTube, réduisent leurs coûts de manière significative en utilisant MySQL pour leurs sites Web, leurs applications critiques d'entreprise, ou en embarquant MySQL au sein de leurs solutions.

PHP : Hypertext Preprocessor

Le langage PHP est un langage de script open source dont la syntaxe est empruntée aux langages C, Java et Perl. C'est un langage utilisé pour générer des pages internet dynamiques à l'aide d'un serveur HTTP. C'est un langage, qui jusqu'à la version 4, était essentiellement impératif. Depuis la sortie de la version 5, il dispose des fonctions de langage objet.

JavaScript

JavaScript est un langage de programmation de script utilisé dans les pages internet interactives.

Flux RSS

Les Flux RSS sont des fichiers dont la syntaxe s'appuie sur le langage XML. Ces fichiers sont produits automatiquement en fonction des mises à jour d'un site internet. Ils sont souvent utilisés par des plateformes d'actualité (ex : Google news).

Microformat

Les microformats sont des conventions simples. Leur principale utilité est de décrire un type d'information spécifique, comme une personne, une organisation, un avis ou même

¹ Open source : logiciel libre de droit

des recettes. Chaque entité possède ses propres propriétés. Par exemple, une personne est associée aux propriétés suivantes : nom, adresse, fonction, société et adresse e-mail.

Ces entités complètent les renseignements déjà existants comme les métadonnées et les renforcent afin de donner des informations complémentaires aux robots qui viennent scanner les pages pour en recueillir les données. Cela permet de décrire aux programmes qui passent sur les pages dotées de ces entités la nature des informations qui s'y trouvent.

Sécurité

La garantie qu'un site est sécurisé à 100% est impossible. Tout administrateur ne peut que ralentir au maximum une tentative d'intrusion.

« URL Rewriting »

Cette technique consiste à renommer les liens des pages du site tout en masquant les variables passées en argument, ce qui permet de sécuriser d'une certaine façon le site. Cela ajoute également à l'utilisateur des informations complémentaires sur la page qu'il consulte. Pour l'utilisateur, le lien est plus compréhensible, mais pour le CMS, le lien conserve les éventuels variables et arguments.

Cette technique fait partie de la SEO, car un moteur de recherche accorde moins d'importance à une page dont l'adresse est composée de variables et de valeurs.

Par exemple si l'adresse internet d'une page parlant de la faculté Jean Monnet est *http://www.domaine.com/index.php?ville=saint-etienne&fac=jean-monnet*, l'adresse indiqué à l'utilisateur dans sa barre d'adresse de son navigateur est *http://www.domaine.com/saint-etienne/jean-monnet.htm*, le serveur http, à l'aide de l'URL rewriting, fait la traduction et peut appliquer le script permettant d'afficher le contenu. Un lien de ce type est plus lisible pour l'utilisateur, et les moteurs de recherche peuvent référencer la page avec les mots-clés « saint-etienne » et « jean-monnet ».

Les formulaires

L'utilisation première d'un formulaire est la saisie d'information afin de répondre interactivement au site.

Certaines personnes peuvent utiliser ces formulaires à d'autres fins comme obtenir des informations d'ordre privé, avoir accès à des parties du site dont elles ne sont pas censées être autorisées à entrer, comme un panneau d'administration, ou encore se faire passer pour quelqu'un d'autre. Une des techniques qu'elles utilisent est nommée *injection sql*, cela leur permet de tenter de se connecter au site sous le nom d'une autre personne. Si le formulaire ne fait pas le traitement adéquat, la personne mal avisée peut entrer sur le site comme bon lui semble. Une autre technique qui peut être utilisé dans un formulaire est d'insérer venant d'un autre site, ce qui permettra à la personne d'obtenir des informations de nature privée, ou d'injecter sur la page des scripts malveillants.

Afin d'éviter ce genre de problème, le CMS a été étudié pour qu'aucun de ces cas puissent ne puisse se produire. La première chose est que seules les personnes ayant accès au panneau d'administration connaissent son adresse. Mais admettons que cette adresse soit trouvée, le traitement des données saisies par le formulaire est transformé de manière à qu'aucune information malveillante n'y soient envoyée. Avant toute requête à la base de données, l'identifiant subit un traitement qui retire toutes les balises qui peuvent y être ajoutées, et qui transforme tous les caractères spéciaux (#, =, l'espace, etc...) ainsi que les expressions qui peuvent mal être interprétées par la SGBD (Système de Gestion de Base de Données).. Les autres formulaires dédiés aux utilisateurs ont subit le même traitement.

Les sessions

Pour conserver la session d'authentification des administrateurs, il est possible d'utiliser la fonction session de PHP. Un inconvénient de cette technique est que ses données sont sauvegardées dans un répertoire temporaire du serveur FTP. Il est possible de changer ce répertoire, mais pour une personne malveillante il est toujours plus facile d'accéder au FTP qu'à la base de données. Pour parer à ce problème, les sessions sont conservées via la base de données.

Sécurité complémentaire

En plus de ce qui a été énoncé précédemment, une sécurité complémentaire a été ajoutée. Lorsqu'un utilisateur fait une requête à une page du CMS, il faut que l'url de la page soit acceptée au préalable par l' « URL rewriting ». Si une personne malveillante arrive à obtenir les vrais liens des pages en contournant la réécriture, elle pourrait tenter d'avoir

accès à des données du CMS, voire la possibilité de l'attaquer. Pour ce faire la technique qu'elle emploierait serait d'injecter, dans le lien d'une page, des variables et un lien renvoyant sur l'un de ses propres sites.

Pour contrer ce genre d'attaque, j'ai opté pour une vérification des liens des requêtes que fait l'utilisateur. Chaque module, par exemple *fiche*, *rubrique*, *fil d'actualité*, doit être existant, c'est-à-dire que le lien permettant d'accéder à l'un de ces modules contient le nom du module.

Prenons le cas d'une rubrique. Le lien visible par l'utilisateur d'une rubrique est *http://www.domaine.com/larubrique* le lien réel que fait la requête auprès du CMS est *?mod_p=rubrique&url_r= larubrique* . Si la personne tente de mettre un module différent de *rubrique* dans ce cas présent, le CMS analyse au préalable les paramètres de la requête afin de vérifier l'existence du module. Si le module existe, le CMS regarde la composition des paramètres suivants, si l'une des deux vérifications est erronée alors l'utilisateur se retrouvera sur la page d'accueil.

Le panneau d'administration

La description du CMS se fera par le biais du panneau d'administration.

Le fil d'actualité

Pour informer l'utilisateur sur des actualités, l'administrateur a deux possibilités. Soit il les rédige lui-même, soit il utilise un Flux RSS qu'il a choisi. Le nombre d'actualités affichées est paramétrable via le panneau d'administration. Sur la page regroupant toutes les actualités, seul un aperçu de ces dernières est visible, afin d'avoir le contenu complet, il faut que l'utilisateur décide de les lire pour avoir davantage d'information.

Actualités personnalisées

L'utilisation d'actualités personnalisées permet à l'utilisateur de connaître des informations propres au site, et à ce que ces informations soient plus pertinentes que des informations tirées d'autres sources.

Flux RSS

L'utilisation des Flux RSS permet d'enrichir le contenu du site en donnant des informations complémentaires venant d'autres sources qui peuvent intéresser l'utilisateur. Il est possible d'importer plusieurs flux afin d'avoir des actualités venant de différentes sources.

Pour se faire, l'administrateur a la possibilité d'importer un flux d'un autre site via le panneau d'administration. Le contenu de ses flux est au format XML. Toutes les informations sont stockées, et sachant que les flux peuvent venir de différents sites, il faut alors les trier par date afin d'avoir les plus récentes en premier.

Les rubriques

La gestion des rubriques regroupe plusieurs choses comme la gestion de la rubrique en elle-même, la gestion des fiches associées à chaque rubrique et l'importation de fiche à partir d'une base de données existante au format XML.

Lors de la création d'une rubrique, cette dernière est cachée à l'utilisateur tant que l'administrateur ne décide pas l'inverse, qu'elle ne possède pas de fiches et que sa description n'est pas complétée.

Les fiches

Comme pour les rubriques, les fiches sont par défaut cachées à l'utilisateur. Une fiche correspond à un établissement et recense :

- Son nom
- Ses coordonnées
- Sa description
- Ses horaires
- Ses tarifs

Chacun de ces champs est prédéfini à la base, mais il pourrait être envisageable de les paramétrer via le panneau d'administration.

Les fiches sont annotées par les microformats afin de renseigner plus facilement les programmes qui passeront analyser les pages.

Importer une base de données de fiches

En plus de pouvoir ajouter des fiches individuellement, il est possible d'ajouter des fiches à partir d'une base de données existante. Sachant que les champs d'une autre base de données ne correspondront pas automatiquement ceux utilisés pour la base de données du CMS, il a fallu créer une importation générique.

Pour ce faire, il faut que la base de données à importer soit au format XML. Lors de l'importation, le panneau lit les champs et attend que l'administrateur fasse la correspondance entre les champs à importer et ceux qui sont à remplir. Une fois cette correspondance faite, le panneau remplit automatiquement la nouvelle base de données. Cette fonctionnalité laisse les fiches cachées en attendant la confirmation par un administrateur pour devenir visible pour les utilisateurs. Procéder ainsi permet de compléter les fiches même si certains champs sont incomplets.

Les avis

Les avis sont des outils pour les utilisateurs, ils permettent de se renseigner sur un établissement. Ils sont présents afin que chacun puisse donner une appréciation accompagnée d'une note sur l'établissement en question. Les fiches sont classées par établissement le mieux noté à l'établissement le moins bien noté.

Il est également possible de modérer les avis via le panneau d'administration afin d'autoriser leur publication après relecture pour ne pas avoir de commentaire injurieux ou autre condamnable par la loi.

Les perspectives

Amélioration à apporter

Croisement des données

Actuellement le croisement des données entre les diverses fiches est en développement.

Le croisement des données entre les actualités et les fiches n'est pas encore en place. Pour se faire, il faut faire un choix entre divers web service ou API afin d'avoir le résultat escompté. Le choix est assez vaste :

- *Term Extration Web Service* ²
- *Google Data Protocol* ³
- *OpenCalais Web Service* ⁴

Gestion des utilisateurs

Il pourrait être envisagé de faire une gestion d'utilisateurs, c'est-à-dire de définir des droits pour modérer uniquement une ou plusieurs rubriques, fiches ou encore uniquement les commentaires des avis.

Gestion de l'interface

Pour le moment le squelette du CMS est statique, l'administrateur a uniquement accès au fichier de style. Il serait possible de rédiger une page permettant à l'administrateur de modifier l'interface des pages que l'utilisateur verra.

La sécurité

La réalisation apportée au niveau de la sécurité est assez rudimentaire. Les cas traités concernent essentiellement les formulaires, afin d'éviter toute injection de code, ainsi que les liens url. Il faudra sûrement demander l'avis d'un professionnel pour savoir s'il faut modifier ou ajouter certaines choses.

² <http://developer.yahoo.com/search/content/V1/termExtraction.html>

³ <http://code.google.com/intl/fr/apis/gdata/index.html>

⁴ <http://www.opencalais.com/APIkey>

La maintenance

La principale fonction à surveiller en matière de maintenance est la modération des avis des divers utilisateurs afin de ne pas se retrouver avec des commentaires portant atteinte à autrui.

Lorsque l'API ou le Web Service sera mis en place, il faudra faire attention à ses mises à jour, afin de savoir s'il faut ou non modifier le code pour garder le bon fonctionnement du CMS. Il faut également se tenir au courant des nouveautés et modifications de PHP qui parfois rendent des fonctions obsolètes, ou sont remplacées par d'autres.

Les coûts

Pour la mise en place de ce CMS, le choix du type d'hébergement dépendra de sa fréquentation. Dans un premier temps, un simple pack d'hébergement classique suffit. Généralement, chaque pack comprend un nom de domaine, un espace disque de quelques giga octets, d'au moins une base de données d'un giga octet, de comptes mails et pour certain un trafic mensuel illimité. Le coût de ce genre de pack est inférieur à cinq euros par mois.

Si le trafic s'intensifie et que le CMS nécessite une optimisation en changeant certaines parties qui doivent être traitées différemment, il faudra sûrement s'orienter vers un serveur dédié. Le coût dans ce cas précis augmente significativement, de l'ordre d'une vingtaine d'euros pour le plus basique jusqu'à la centaine d'euros mensuel.

Structure de la base de données

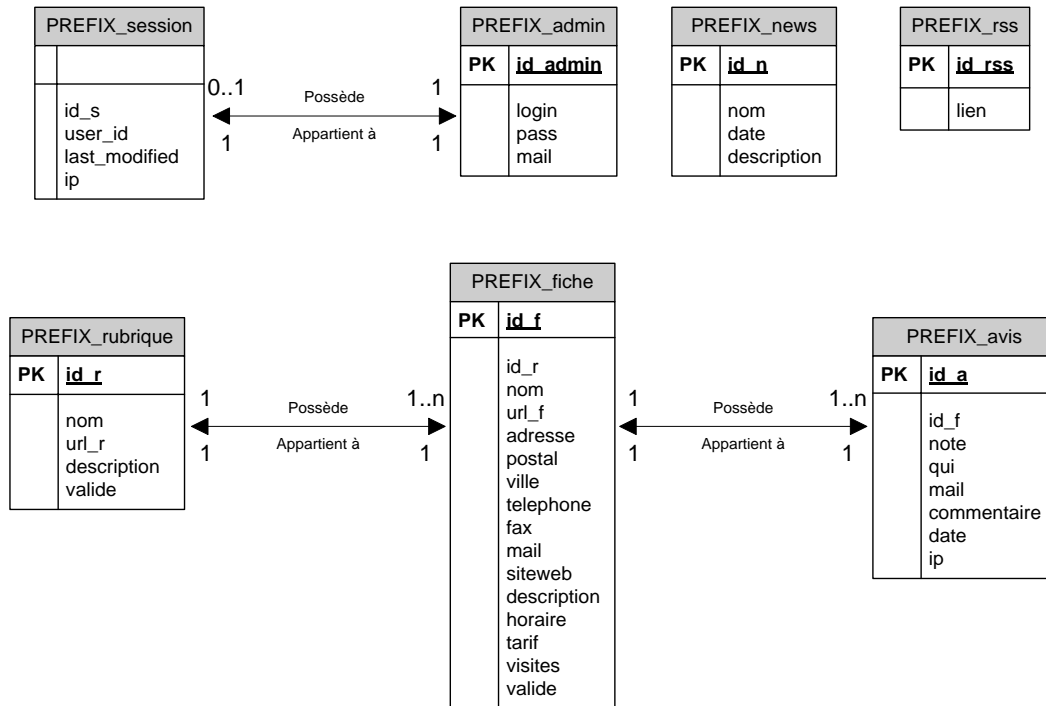


Figure 3 Structure de la base de données

Table des illustrations

Figure 1 Structure du site	13
Figure 2 Structure du panneau d'administration	13
Figure 3 Structure de la base de données.....	19

Bibliographie

Drupal – *Communauté Drupal France et francophonie* – Source : <http://drupalfr.org/>

WordPress – *WordPress Blog Tool and Publishing Platform* – Source : <http://wordpress.org/>

Joomla – Source : <http://www.joomla.org/>

SPIP – Source : <http://www.spip.net/>

XOOPS – *XOOPS Web Application System* – Source : <http://www.xoops.org/>

Microformat – Source : <http://microformats.org>