

Sujet de mémoire d'ingénieur CYC14

Refonte du moteur de recherche du site internet du Parlement Européen

1) Contexte de l'étude :

Le Parlement Européen mandate des consultants analyste-développeurs, dont l'auditeur fait partie, pour assurer le développement et la maintenance de ses applications internes et externes. Son site internet est considéré comme l'une de ces applications et est contraint de procéder à une évolution majeure de son moteur de recherche.

Actuellement, l'ensemble des composants permettant d'effectuer une recherche sur le site est portée par une solution payante et non ouverte de la société Google, nommée GSA (Google Search Appliance). Concrètement, il s'agit d'une machine contenant un ensemble applicatif, que l'on relie au réseau interne. Grâce à une interface d'administration, on configure la GSA pour qu'elle parcoure les pages du site internet et qu'elle indexe leur contenu. Ensuite, toutes les recherches faites par un utilisateur, sont relayées à la GSA qui renvoie les résultats de la recherche.

Bien que très imparfaite en terme de qualité attendue et répondant mal aux besoins, cette solution est jugée acceptable. Le Parlement pourrait continuer à l'utiliser mais sa licence expire en août 2018. Comme la société Google a annoncé qu'elle ne renouvellerait pas les licences car elle arrête la maintenance de son outil, le Parlement Européen est obligé de mettre en place une solution de remplacement.

L'auditeur a été chargé de mettre en œuvre un nouveau moteur de recherche, qui réponde mieux aux attentes du Parlement.

2) Sujet de l'étude :

La solution de remplacement à apporter est composée de deux parties distinctes qui contiennent chacune des problématiques autour des besoins métiers et autour de l'architecture applicative à mettre en place. L'un des composants doit parcourir le site internet pour indexer son contenu et l'autre doit exposer des services permettant à des applications identifiées d'effectuer des recherches sur le contenu des pages.

La volonté du Parlement est de bénéficier d'un module d'indexation, moins générique que la « GSA », qui soit configurable et qui puisse répondre de façon plus précise à ses besoins actuels et futurs. Il doit être mieux optimisé en terme de charge imposée au site internet, au regard de la pertinence des résultats de recherche. Il a été décidé de s'affranchir des contraintes infra-structurelles en déployant les indexes sur un cloud public. La technologie utilisée est « Elasticsearch » car elle est opensource et fait partie de la liste des technologies que l'on peut rapatrier en interne en cas de besoin.

L'application responsable de parcourir et d'indexer le site internet sera codée en Java, elle exposera une interface d'administration pour configurer les politiques d'exploration et d'indexation. Cette application sera hébergée sur le réseaux interne.

Le composant responsable d'exposer des services web de recherche aux applications tiers, sera comme les indexes, déployé sur le cloud. Il s'agira également d'une application Web développée avec la plateforme JEE. Son rôle sera de fournir une interface compréhensible pour les utilisateurs. Il sera en charge de faire le lien entre la vision métier des clients et la vision purement technique de la technologie d'indexation. Cette interface permettra également de réaliser des évolutions progressives sur les indexes, sans imposer un impact direct sur les clients. Une fois achevée, les requêtes faite à la « GSA » par les services du site internet, devront être remplacés par des appels à ces services.

3) Plan prévisionnel :

Durées exprimées en jour x homme :

- Analyse fonctionnelle et technique: **25**
- Développement: **125**.
- Déploiement et formation : **25**.

Soit un total de **175 jh**.

Comme l'auditeur travail seul et qu'il ne consacre que 75 % de son temps à ce projet, ce délai correspond à une durée réelle de 11 mois. Le projet ayant commencé en mai dernier, sa date de fin estimée est avril 2018.

4) Conditions du déroulement de l'étude :

L'étude a lieu au sein de l'unité Evolution de la Direction Technique du Parlement Européen située place de la Philharmonie à Luxembourg, sous la responsabilité de M. Olivier ISAAC (fonctionnaire responsable du projet) et de Mme Clara BOUZAS MANTE (fonctionnaire chef de service). Charles HUBER (analyste développeur) est affecté sur toutes les phases du projet, de l'analyse à la mise en production. Le tuteur CNAM sera M. Michel IANOTTO.

Auditeur CNAM	Charles HUBER	
Chef de projet	Olivier ISAAC	
Chef de service	Clara BOUZAS MANTE	
Tuteur CNAM du mémoire	Michel IANOTTO	
Responsable régional de la spécialité informatique au CNAM de Lorraine	Christian DUCLOU	