

# Fiche de proposition de stage

## Outils de moissonnage documentaire

CETE Méditerranée, Aix-en-Provence, 09/11/09

### 1 Contexte

Le CETE (Centre d'Etudes Techniques de l'Équipement) Méditerranée est un service technique du ministère du développement durable. Le stage est proposé par le service DCEDI/TIER qui travaille dans le domaine de l'ingénierie du trafic et des systèmes d'information sur les transports. Ses activités sont décrites sur les pages <http://www.cete-mediterranee.fr/tt13/www/>

Le stage a lieu à Aix-en-Provence, pôle d'activités des Milles.

**Responsable à contacter** : Patrick GENDRE / département DCEDI

pat.gendre@developpement-durable.gouv.fr 04 4224 7687

<http://www.cete-mediterranee.fr/tt13/www>

### 2 Objectif du stage

En tant que service technique de l'état, le rôle du CETE est aussi de faire de la veille et de diffuser des connaissances aux autres services du ministère ainsi qu'aux collectivités et à la communauté technique.

En l'occurrence le domaine technique concerné ici est celui des « transports intelligents », et en particulier l'information sur les transports collectifs (site [www.predim.org](http://www.predim.org)), ainsi que les études et la simulation de trafic.

Parmi les documents que l'on souhaite capitaliser et éventuellement rediffuser sur le web figurent les avis de publication des consultations de marchés publics, les Dossiers de Consultation des Entreprises (en particulier les cahiers des charges ou Cahiers des Clauses Techniques Particulières), et les avis d'attribution.

**L'objet du stage est d'automatiser ce processus de 'moissonnage' des documents de marchés publics. (et plus généralement de veille technique).**

Les programmes seront des scripts qui pourront être exécutés en ligne de commande sous windows, il n'y aura pas besoin a priori de développer une IHM.

### 3 Méthode

Le processus à informatiser est le suivant :

-1- effectuer des requêtes thématiques par mot-clés (en commençant par le domaine 'système d'information voyageur' puis pour d'autres domaines) sur des sites (en commençant par [www.boamp.fr](http://www.boamp.fr) puis éventuellement d'autres sites), afin d'obtenir une liste de références (url) sur des appels d'offres en cours (les avis d'attribution de marché pourront être dans un 2<sup>ème</sup> temps)

-2- pour chaque élément de la liste d'appel d'offres trouvés, interroger le site de marché public adéquat en vue de récupérer les informations concernant cet appel d'offre : extraction des informations pertinentes (date, maître d'ouvrage, lieu, objet du marché...), et le cas échéant des documents de consultation (dossier complet, avis, CCTP).

-3- publier ces informations sur un outil de gestion de contenu web (de préférence spip)

Le développement sera itératif :

- commencer par spécifier, développer, tester et documenter la récupération d'information pour un seul type de requête (appels d'offres en cours), sur un seul domaine (information multimodale), un seul site (BOAMP)

- en fin de chaque itération, faire des propositions de cahier des charges pour l'itération suivante

- idéalement on devra effectuer une itération, mais la rapidité d'avancement dépendra des difficultés rencontrées et des solutions qui pourront être apportées.

Selon l'avancement des développements, de nombreuses fonctionnalités pourront compléter l'application (par exemple, en matière syndication de contenu RSS ou Atom).

En termes de langage, l'application sera de préférence écrite en python. En effet, python est un langage facile à lire (donc à maintenir), open source, bien documenté, et disposant des bibliothèques nécessaires pour développer ce genre d'applications, en particulier `httplib` (ou `httplib2`) pour dialoguer avec des sites web. La lecture et la mise en pratique des exemples décrits dans l'ouvrage « RESTful web services<sup>1</sup> » est recommandée.

Il n'y a pas de développeurs dans l'équipe accueillant le stage. En termes d'environnement de développement, la seule demande est que ce soit open source et simple afin que si ce projet aboutit, il puisse être repris et amélioré facilement ensuite par d'autres (étudiants ou prestataires). Il nous semble que les outils standard éditeur/debug de `python.org` doivent faire l'affaire. Si besoin, une forge publique pourra être mise en œuvre.

<sup>1</sup> <http://oreilly.com/catalog/9780596529260>

Il n'y a pas d'IHM à développer, les programmes fonctionneront en ligne de commande et produisent des fichiers texte analysables avec un tableur ou autre.

En termes de 'méthode de développement', une grande importance sera attachée à la documentation sans laquelle les développements effectués ne pourront être améliorés par la suite :

- document fonctionnel expliquant ce que fait / font les programmes, comment les utiliser (pour ce projet il n'y a pas besoin à proprement parler de manuel utilisateur séparé des spécifications)
- document de conception (indispensable pour la maintenance éventuelle)
- code documenté

L'application demandée comporte 3 modules , le projet est donc bien adapté à un développement en équipe, avec un étudiant pour chaque module, et éventuellement un quatrième en 'intégrateur'.

#### **4 Travail attendu, livrables**

- Prise de connaissance du contexte (marchés publics, transports, veille, diffusion de documents), et des connaissances de base sur les outils (python, spip)
- Recherche d'état de l'art sur les logiciels de veille web, en particulier open source / réutilisables
- Le cas échéant, organisation du projet si développement en équipe
- Installation de l'environnement de travail (pythons et librairies python, spip et autres outils)
- Manipulation de données et premiers tests
- Définition d'un 1<sup>er</sup> programme de travail
- Spécifications, Conception/Développement/Test et Documentation pour les 3 modules
- Livraison
- Améliorations des logiciels par itération (sur un stage de 3-4 mois, 3 itérations sont faisables/souhaitables)
- Bilan et recommandations
- Mémoire de fin de stage

#### **5 Qualités requises**

- conception d'applications web
- de préférence connaissance de python, spip
- si possible intérêt pour les déplacements et les transports
- autonomie et capacités de rédaction

#### **6 Calendrier**

Trois mois minimum