

Plan de mémoire

Refonte du moteur de recherche du Parlement Européen

Table des matières

1.Statut et mission de l'auditeur au sein du Parlement Européen.....	3
1.1.Les institutions européennes.....	3
1.2.Le Parlement Européen.....	3
1.3.Le service EVOLUTION.....	3
1.4.La mission de l'auditeur.....	3
2.La situation initiale.....	3
2.1.Présentation et fonction de la GSA.....	3
2.2.L'architecture applicative existante.....	3
2.3.Les lacunes du système.....	3
a)Une fonction principale insatisfaisante.....	3
b)Des fonctions secondaires absentes.....	3
2.4.Une échéance à relever.....	3
3.Cahier des charges.....	3
3.1.Les besoins fonctionnelles.....	3
a)Les besoins de la direction de la communication.....	3
b)Les besoins de la direction technique.....	3
3.2.Les besoins techniques.....	3
a)Les technologies supportées et autorisées.....	3
b)Les processus qualité.....	3
c)Un projet pilote vers le Cloud.....	4
4.La gestion du projet.....	4
4.1.Budget.....	4
4.2.Planning prévisionnel.....	4
5.La conception du projet.....	4
5.1.Analyse fonctionnelle.....	4
5.2.Architecture technique prévisionnelle.....	4
5.3.État de l'art des solutions techniques applicables.....	4
a)Les web services.....	4
b)Les interfaces graphiques.....	4
c)Les systèmes d'indexation.....	4
5.4.Recherches et prototypage.....	4
a)Un POC du module d'indexation.....	4
b)La solution Azur Search.....	4
c)Les solutions Amazon.....	4

d)L'hébergement des index ElasticSearch sur le réseau du Parlement.....	4
5.5.Choix de la solution.....	4
a)Étude comparative.....	5
b)Schéma d'architecture technique définitif.....	5
5.6.Conclusion.....	5
6.La réalisation.....	5
6.1.Un développement par itération.....	5
a)De plus prioritaire au moins prioritaire.....	5
b)Des phases d'analyse approfondies et une redéfinition des besoins.....	5
c)Validation et boucle de retro-action.....	5
6.2.Développement du module d'indexation.....	5
a)La gestion des fréquences d'indexation.....	5
b)Le lien entre les pages html et le modèle de données.....	5
c)La distinction entre les liens suivis et les pages indexées.....	5
d)Politiques d'indexation et suppression des pages obsolètes.....	5
e)La gestion des pages en erreur.....	5
f)Une interface d'administration.....	6
6.3.L'implémentation des fonctionnalités de recherches.....	6
a)Résumé des concepts Elasticsearch.....	6
b)Le couplage entre la structure des index et les fonctionnalités de recherche.....	6
c)La recherche sur une phrase.....	6
d)Une recherche simultanée sur la phrase et sur les mots.....	6
e)Corrections et suggestions de « phrases ».....	6
f)Auto complétion sur l'historique des requêtes pertinentes.....	6
6.4.Développement du web service de recherche.....	6
a)Un web service REST utilisable depuis n'importe quelle source.....	6
b)La définition du modèle de donnée exposé.....	6
c)Implémentation d'une page de recherche simplifiée.....	6
6.5.Intégration du client de recherche au site internet du parlement.....	6
7.Tests et intégration continue.....	6
7.1.Test unitaires et couvertures de code.....	6
7.2.Les critères qualité de l'analyse automatique du code.....	6
7.3.Les tests d'intégration.....	6
7.4.Les tests de validation.....	6
8.La documentation.....	6
8.1.Le Document Technique d'Architecture (DTA).....	6
8.2.La documentation technique.....	6
8.3.La documentation Java et les commentaires.....	7
9.La formation utilisateurs.....	7
10.Retours d'expérience.....	7
11.Conclusion.....	7

1. Statut et mission de l'auditeur au sein du Parlement Européen

- 1.1. Les institutions européennes**
- 1.2. Le Parlement Européen**
- 1.3. Le service EVOLUTION**
- 1.4. La mission de l'auditeur**

2. La situation initiale

- 2.1. Présentation et fonction de la GSA**

(Description du service Google Search Appliance).

- 2.2. L'architecture applicative existante**
- 2.3. Les lacunes du système**
 - a) Une fonction principale insatisfaisante**
 - b) Des fonctions secondaires absentes**
- 2.4. Une échéance à relever**

3. Cahier des charges

- 3.1. Les besoins fonctionnelles**
 - a) Les besoins de la direction de la communication**
 - b) Les besoins de la direction technique**
- 3.2. Les besoins techniques**
 - a) Les technologies supportées et autorisées**
 - b) Les processus qualité**
 - *Le suivi (framework ASTRA)*
 - *Documentation*
 - *Intégration continue (Jenkins)*

- *Tests unitaires et couverture de code*
- *Revue de code automatique avec Sonar*

c) **Un projet pilote vers le Cloud**

4. La gestion du projet

4.1. **Budget**

4.2. **Planning prévisionnel**

5. La conception du projet

5.1. **Analyse fonctionnelle**

(diagramme fonctionnel, use cases)

5.2. **Architecture technique prévisionnelle**

(Schéma d'architecture provisoire)

5.3. **État de l'art des solutions techniques applicables**

a) **Les web services**

b) **Les interfaces graphiques**

c) **Les systèmes d'indexation**

(nécessite une phase de recherche approfondie)

5.4. **Recherches et prototypage**

a) **Un POC du module d'indexation**

b) **La solution Azur Search**

c) **Les solutions Amazon**

- *CloudSearch*

- *ElasticSearch service*

(Parler de la gestion des droits)

d) **L'hébergement des index ElasticSearch sur le réseau du Parlement**

5.5. **Choix de la solution**

a) Étude comparative

(malgré la complexité : on privilégie la sécurité (rapatriement possible en intranet) et la flexibilité (possibilité de couvrir un plus grand nombre de besoins) => prise de risques sur le budget prévisionnel)

b) Schéma d'architecture technique définitif

5.6. Conclusion

6. La réalisation

6.1. Un développement par itération

a) De plus prioritaire au moins prioritaire

b) Des phases d'analyse approfondies et une redéfinition des besoins

(note : Une technologie mal comprise par la maîtrise d'ouvrage et par la maîtrise d'œuvre + coûts incertains de la plateforme)

c) Validation et boucle de retro-action

(note : exemples :

fréquence parsing trop importante,

synonymes trop laborieux remplacé par la suggestion de phrase,

autocompletion fulltext remplacée par l'historique des requêtes,

Déclaration dynamique des planètes,

Refonte du mapping planete ↔ index)

6.2. Développement du module d'indexation

a) La gestion des fréquences d'indexation

b) Le lien entre les pages html et le modèle de données

c) La distinction entre les liens suivis et les pages indexées

d) Politiques d'indexation et suppression des pages obsolètes

e) La gestion des pages en erreur

(des solutions imparfaites : effort maximal)

f) Une interface d'administration

6.3. L'implémentation des fonctionnalités de recherches

a) Résumé des concepts Elasticsearch

b) Le couplage entre la structure des index et les fonctionnalités de recherche

c) La recherche sur une phrase

d) Une recherche simultanée sur la phrase et sur les mots

e) Corrections et suggestions de « phrases »

f) Auto complétion sur l'historique des requêtes pertinentes

6.4. Développement du web service de recherche

(note : Une interface qui masque la complexité)

a) Un web service REST utilisable depuis n'importe quelle source

b) La définition du modèle de donnée exposé

c) Implémentation d'une page de recherche simplifiée

6.5. Intégration du client de recherche au site internet du parlement

7. Tests et intégration continue

7.1. Test unitaires et couvertures de code

(tests d'algorithme)

7.2. Les critères qualité de l'analyse automatique du code

7.3. Les tests d'intégration

(tests techniques d'intégration)

7.4. Les tests de validation

(tests utilisateurs sur la satisfaction des besoins)

8. La documentation

8.1. Le Document Technique d'Architecture (DTA)

8.2. La documentation technique

8.3. La documentation Java et les commentaires

9. La formation utilisateurs

10. Retours d'expérience

11. Conclusion