Clients : L. LECHANI, L. SOULIER, W. BAHSOUN
Fournisseurs : CURIEUX Bastien, LOUTON Julien, MOUSSET Paul

BUREAU D'ÉTUDE RECHERCHE D'INFORMATION

28/03/2014

Création d'un moteur de recherche

Sommaire

- Présentation du projet
- Démarche de développement
- Démonstration de l'outil
- Conclusion / Bilan

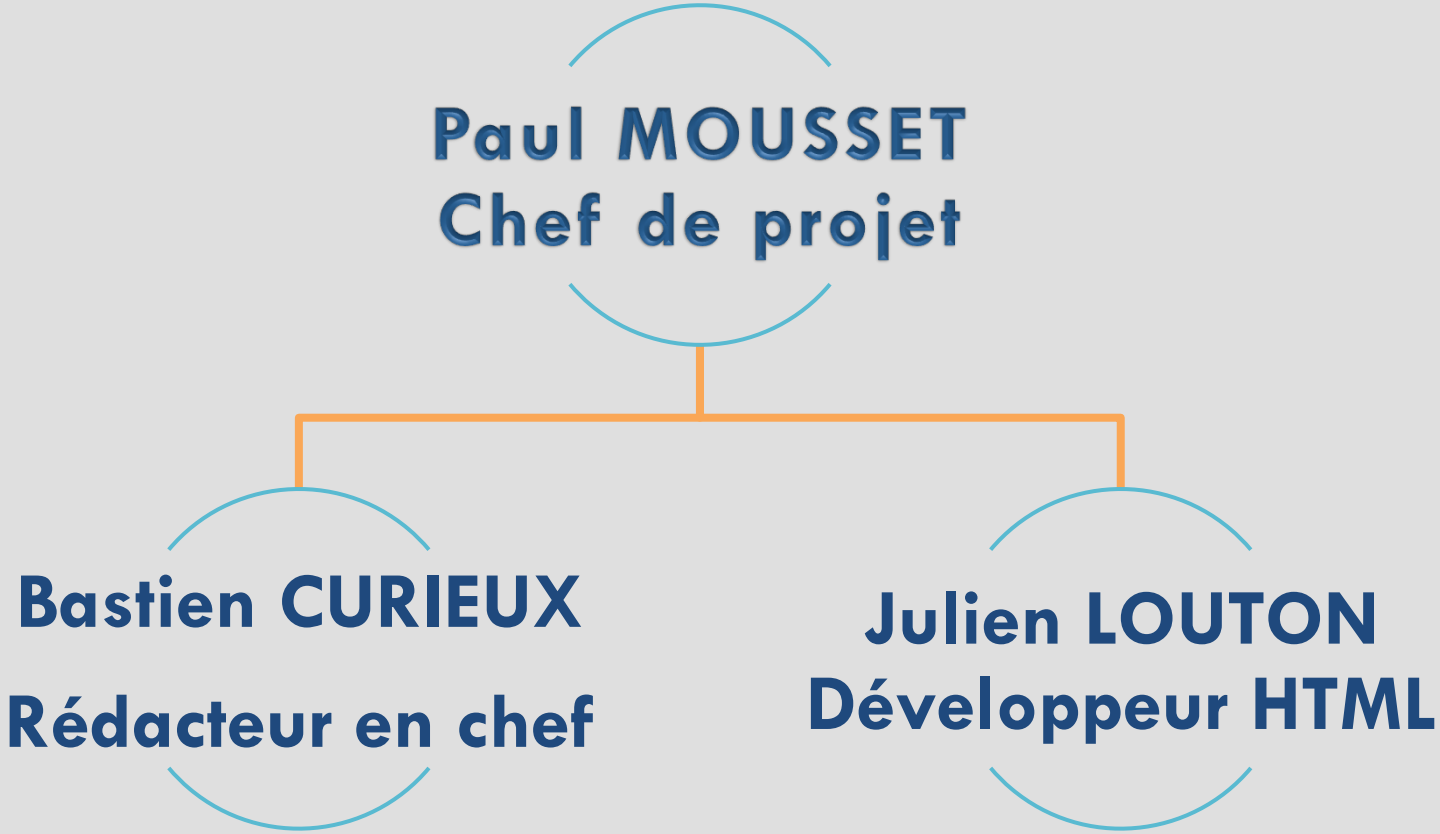
Présentation du projet

Clients : L. LECHANI, L. SOULIER, W. BAHSOUN
Fournisseurs : CURIEUX Bastien, LOUTON Julien, MOUSSET Paul

28/03/2014

Organisation

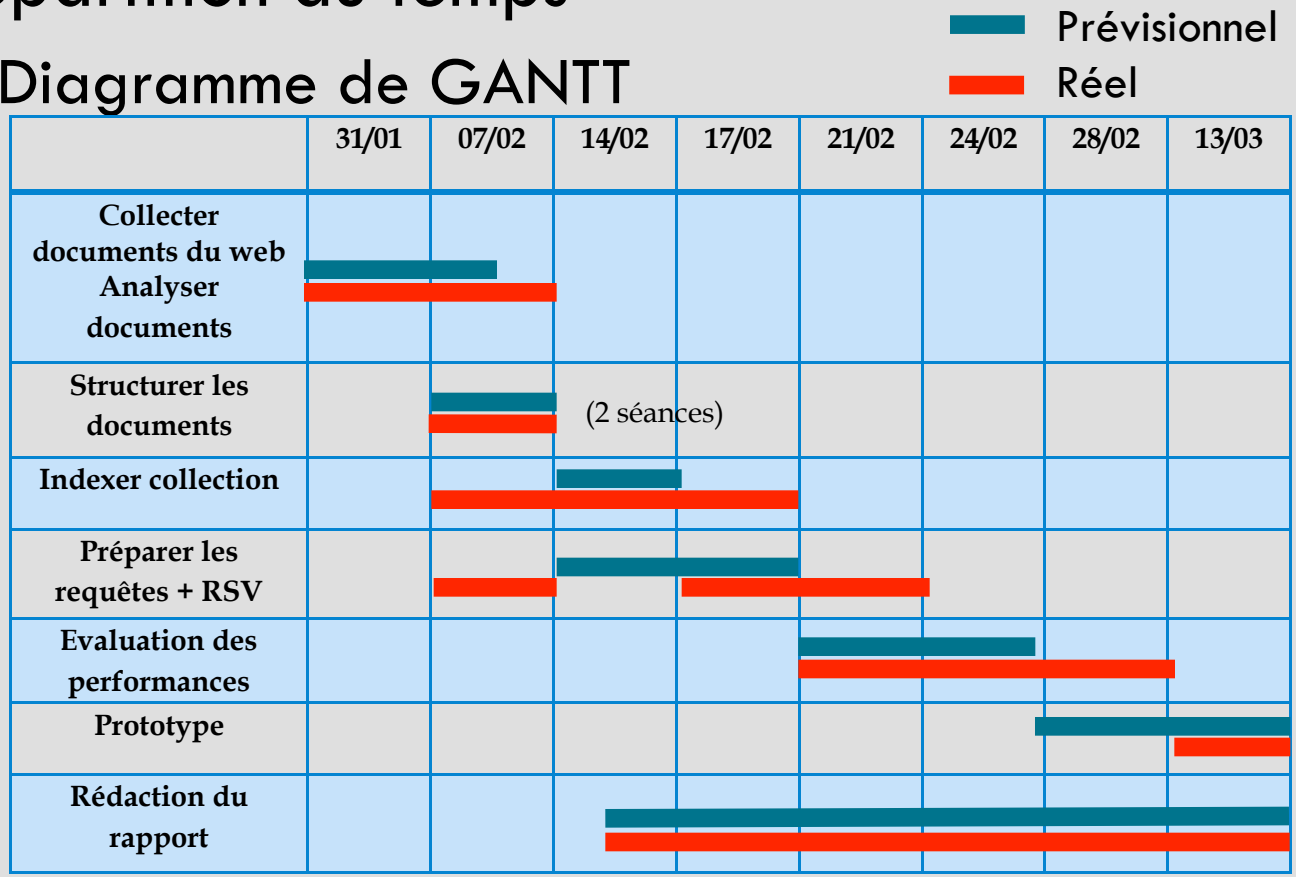
- Répartition des tâches



Clients : L. LECHANI, L. SOULIER, W. BAHSOUN
Fournisseurs : CURIEUX Bastien, LOUTON Julien, MOUSSET Paul

Organisation

- Répartition du temps
  - Diagramme de GANTT

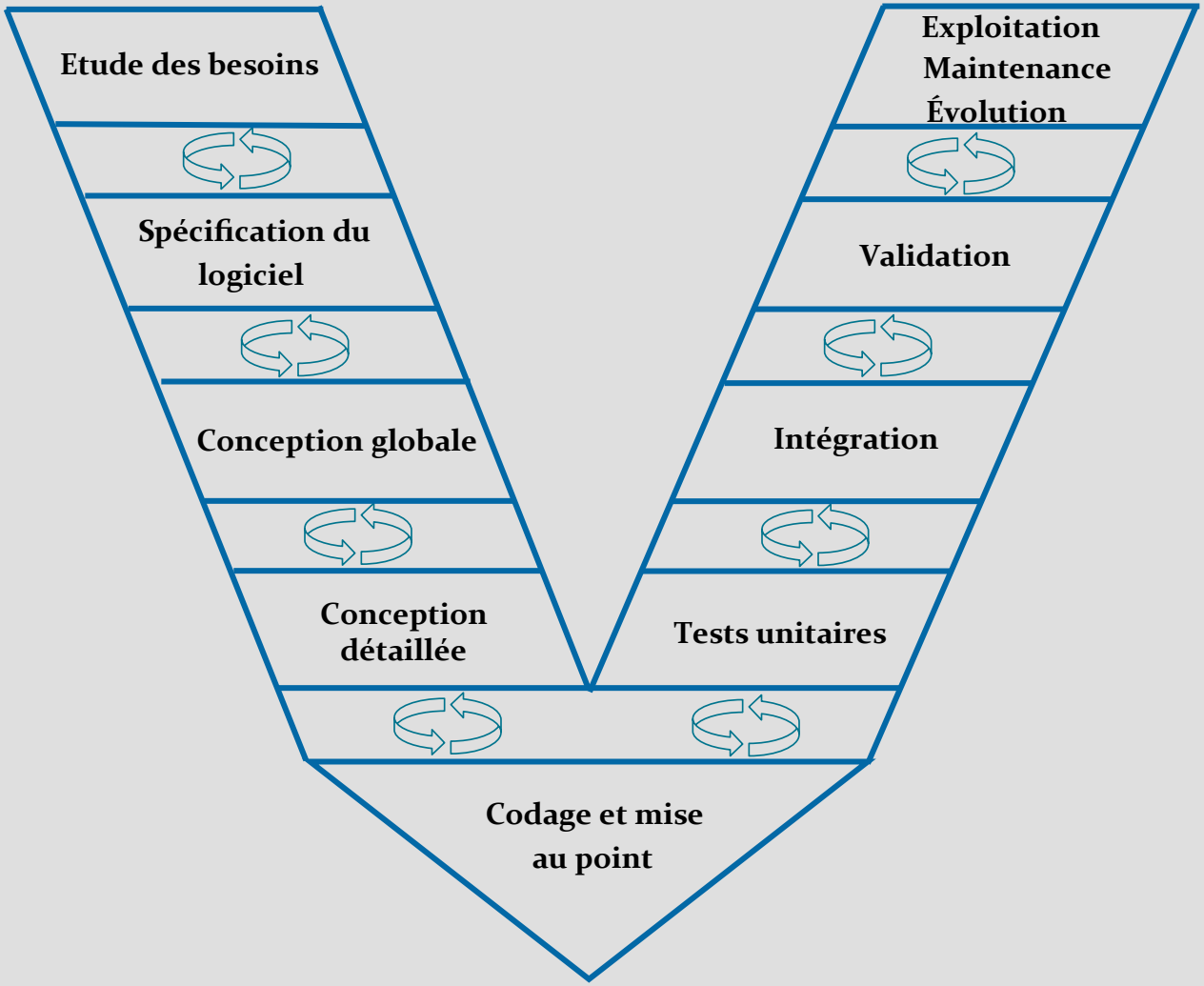


Cahier des charges

- Un module logiciel capable de sélectionner, à partir d'une collection de documents, une liste de documents pertinents en réponse à une requête utilisateur.

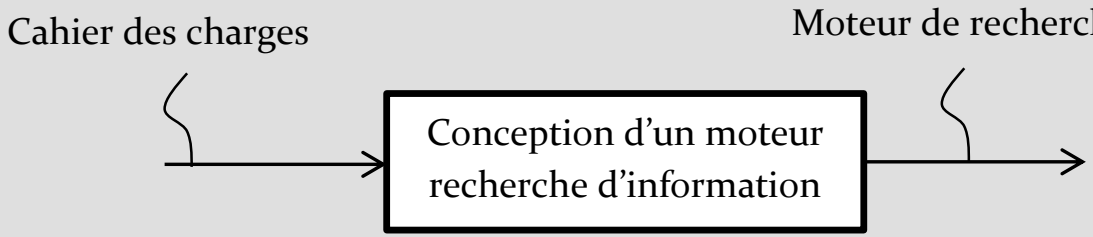
- Création d'une page d'accueil, d'un module de recherche, d'un module de statistique et d'un module administration.

Cycle de vie du logiciel

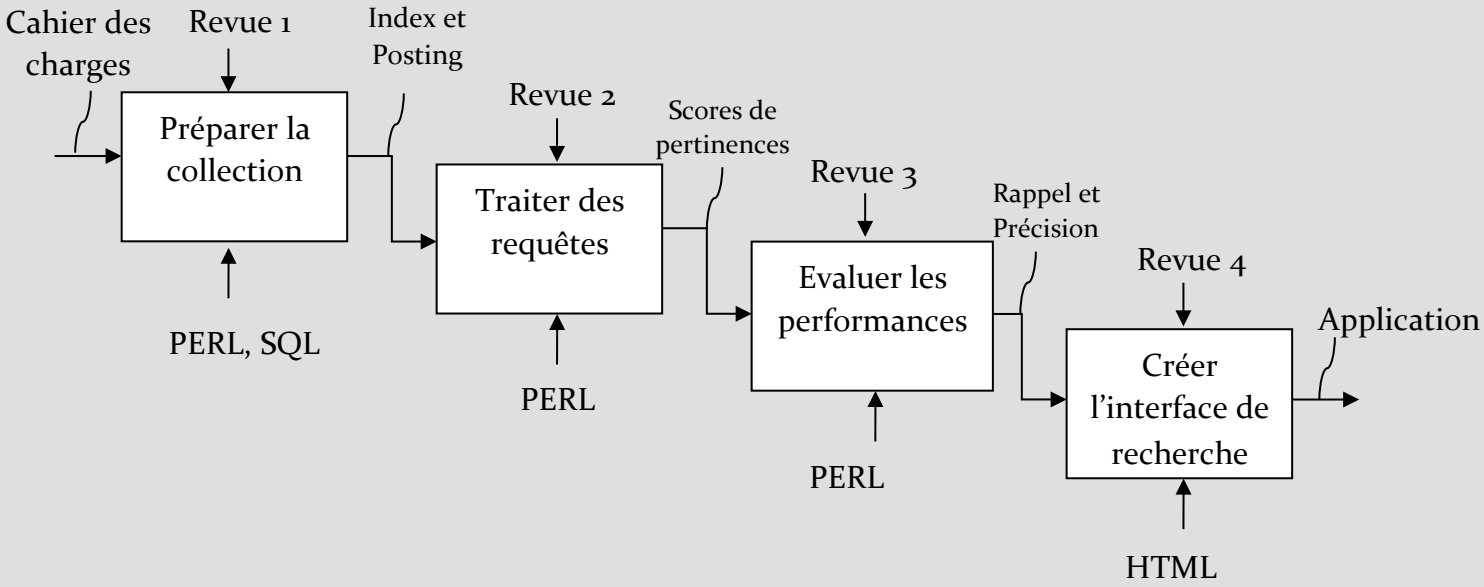


SADT

A_0


Cahier des charges
Moteur de recherche
Conception d'un moteur recherche d'information

A_1


Cahier des charges
Revue 1
Index et Posting
Revue 2
Scores de pertinences
Revue 3
Rappel et Précision
Revue 4
Application
Préparer la collection
Traiter des requêtes
Evaluer les performances
Créer l'interface de recherche
PERL, SQL
PERL
PERL
HTML

Démarche de développement

Démarche de développement

- Préparation de la collection
- Traitement des requêtes
- Evaluation des performances
- Mise en œuvre d'une interface

Préparation de la collection

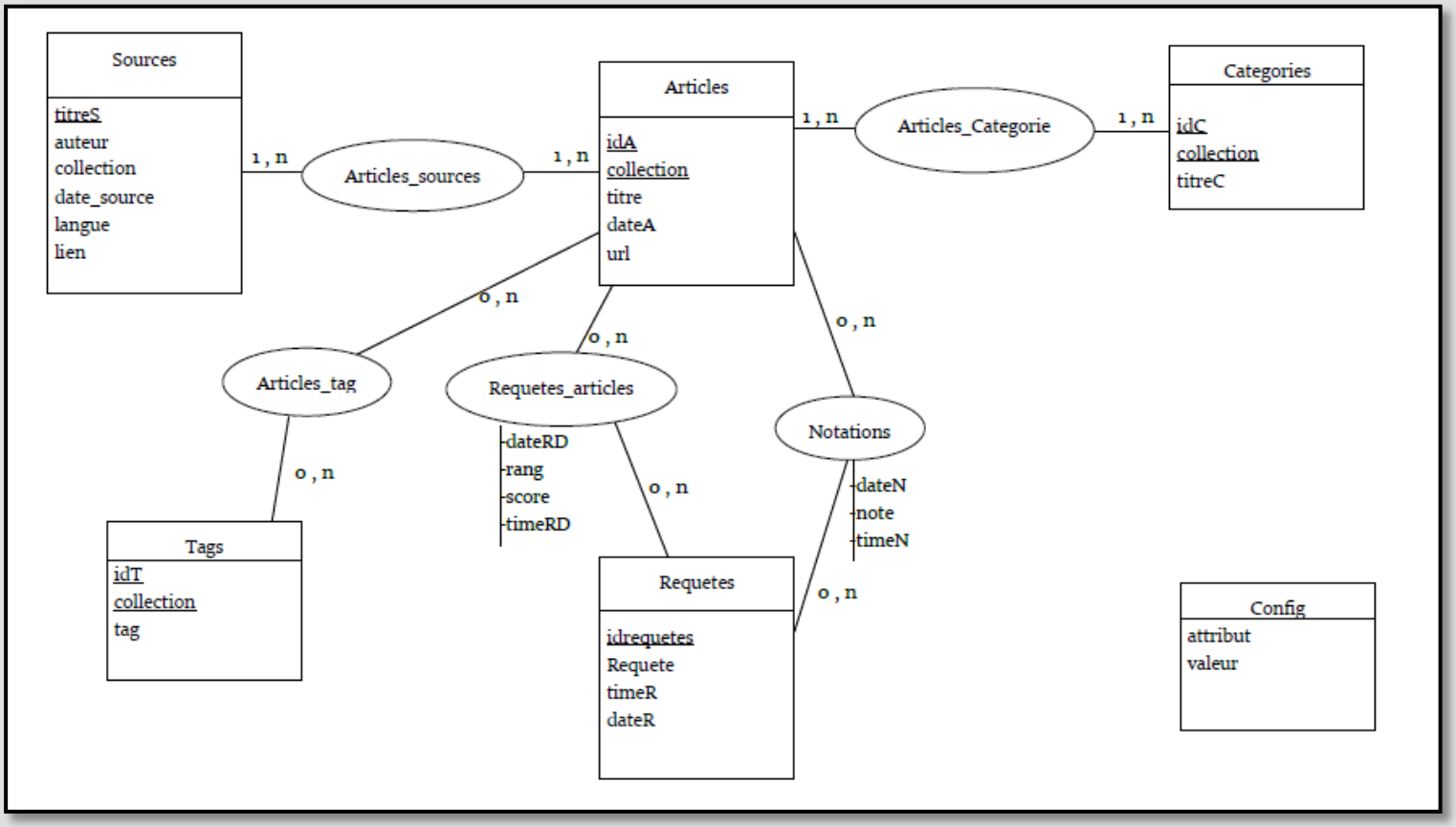
- Construction de la collection de document
  - Constituer une collection de documents HTML
  - 2 sources
    - Wikinews
    - JournalduGeek
  - Aspirateur web : HTTRACK

Préparation de la collection

- Extraction du contenu et méta-contenu grâce à PERL
  - 2 type d'info de contenu
    - Référencement (décrit le contenu)
    - Texte principal de la page
  - Algorithme d'extraction
    - Création en SQL des tables
    - Recherche des données dans les documents
    - Insertion des données dans les tables

Préparation de la collection

- Modèle Conceptuel de Données



Préparation de la collection

- Structuration des fichiers
  - Étapes de l'indexation
    - Lemmatisation
    - Index
      - idDocument terme1 : frequence, terme2 : frequence,…

```
24
02:6,10:3,134:1,16:3,18:2,191,2011:3,2012:1,2013:2,26:1,3:1,301,400000:1,51,561,6918kg:1,8:1,actuali
:1,ailleur:1,ainsi:1,allemag:1,alpha:1,america:1,ams:8,an:1,analyse:1,ans:1,antimat:2,apparei:2,artiste:1,
astrono:1,atteint:1,avait:2,avril:3,bas:1,capacit:1,capte:1,centain:1,cependa:1,chine:1,collabo:1,collect:2
,collisi:1,connue:1,constit:2,coopera:1,coree:1,cosmiqu:1,decembr:1,dernier:1,detecti:1,differe:2,direct
```

    - Posting
      - Terme idDocument : frequence,…

```
marshal     27:1,9:1,
danger      23:1,
multisp     84:1,
```

Traitement des requêtes

- Objectif :
  - Requête (besoin en information)
  - Liste des documents pertinents ordonnés

Traitement des requêtes

- Requête
- Calcul du score de pertinence
  - Méthode du produit scalaire
  - Méthode du cosinus

Evaluation des performances

- Objectif :
  - Evaluer l'efficacité du SRI
    - Sélectionner le plus de documents pertinents
    - Sélectionner le moins de document non-pertinents
  - Démarche expérimentale

Evaluation des performances

- Construction de l'échantillon test

R1	Effondrement d'un immeuble
R2	les fuites d'eau
R3	le projet énergétique
R4	manifestation SIDA
R5	accident de voiture

- Traitement des requêtes test

R1	d87, d68, d99, d88, d3, d2, d19, d86, d4, d14, d85, d13, d5, d57, d11, d35, d69, d79, d22, d17
R2	d1, d2, d71, d17, d18, d75, d70, d98, d73, d74, d15, d80, d32, d33, d21, d63, d90, d7, d26, d99
R3	d25, d76, d7, d82, d93, d1, d11, d26, d4, d6, d90, d85, d22, d5, d92, d68, d98, d74, d81, d99
R4	d95, d92, d71, d82, d70, d1, d75, d79, d93, d74, d80, d23, d98, d99, d32, d33, d21, d63, d90, d7
R5	d6, d79, d93, d35, d7, d3, d92, d74, d5, d75, d4, d77, d81, d80, d32, d33, d21, d63, d90, d71

- Evaluation manuelle des pertinences

Mise en œuvre de l'interface

- Objectif :
  - Soumettre une requête au SRI
  - Récupérer les résultats triés

Mise en œuvre de l'interface

- Besoin du client :
  - Rappel de la requête saisie
  - Affichage des documents selon leur pertinence
  - Système de notation
  - Images catégories
  - Accès aux différents modules

Démonstration de l'outil

Clients : L. LECHANI, L. SOULIER, W. BAHSOUN
Fournisseurs : CURIEUX Bastien, LOUTON Julien, MOUSSET Paul

28/03/2014

Conclusion / Bilan

Clients : L. LECHANI, L. SOULIER, W. BAHSOUN
Fournisseurs : CURIEUX Bastien, LOUTON Julien, MOUSSET Paul

28/03/2014

Conclusion / Bilan

- Réalisation d'une base de données avec une interface utilisateur dans le cadre du bureau d'étude de Concept de Recherche d'Information.

- Difficultés :
  - Concernant l'affichage des résultats.
  - Concernant la fusion.

- Améliorations envisagées :
  - Système de notation dynamique.
  - Afficher les images des catégories les plus pertinentes.

Conclusion / Bilan

- Personnel :
  - Répartition du travail
  - Collaboration avec des groupes extérieurs
  - Enseignement enrichissant
  - Résultat très satisfaisant

Conclusion / Bilan

- Personnel :
  - Répartition du travail
  - Collaboration avec des groupes extérieurs
  - Enseignement enrichissant
  - Résultat très satisfaisant

MERCI POUR VOTRE ATTENTION