

RE41 : Compte-rendu de Travaux Pratiques  
Configuration d'un moteur de recherche

Jocelyn Gourronc - Pierre Mauduit

# 1 Introduction

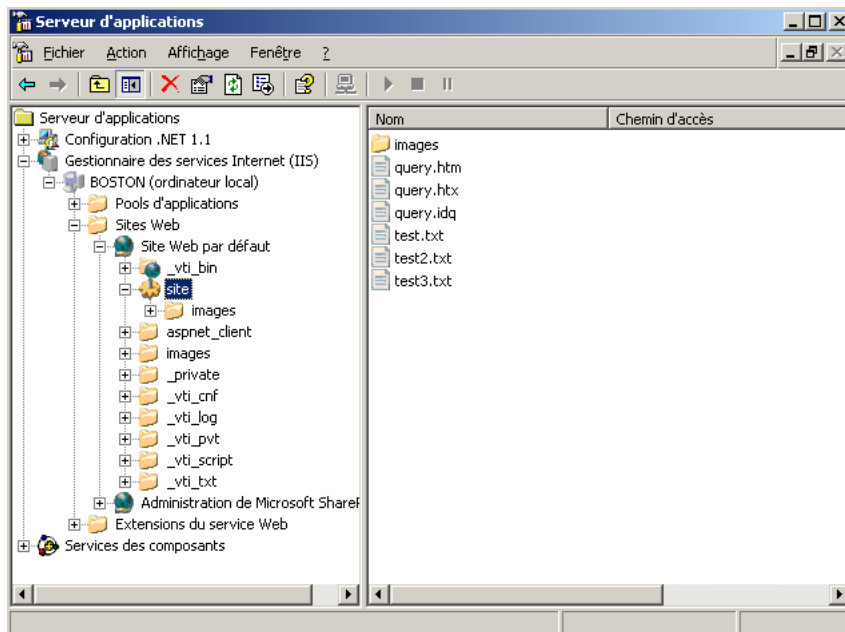
Le but de cette séance de travaux pratiques a été de déployer sous un environnement Microsoft Windows Server 2003, un serveur Web couplé d'un service d'indexation, afin de développer un moteur de recherche minimaliste.

## 2 Installation des composants

Etant donné le fait qu'un grand nombre de groupes de travaux pratiques utilisent le même matériel que nous, nous avons dû commencer par désinstaller les services IIS de Windows, par le biais de la console d'administration du serveur, ainsi que le service d'indexation, via le système d'ajout et suppression de programme. toutes ces démarches nous ont permis de partir sur une base "vierge".

Une fois les logiciels désinstallés, et le travail de nos prédécesseurs supprimés - il était en outre demandé de supprimer le répertoire racine du serveur web, situé par défaut dans C : \InetPub\, nous avons dû réinstaller IIS et le service d'indexation, sur lequel nous allons travailler tout au long du TP.

Après activation du service d'indexation dans IIS et configuration du serveur WEB IIS (cf. figure suivante), nous avons pu nous lancer dans le développement de notre site web moteur de recherche.



### 3 Développement du moteur de recherche

Le moteur de recherche a été développé en utilisant du code HTML, et des balises spéciales interprétées directement par le serveur WEB IIS, permettant la communication implicite avec le service d'indexation de Windows. Vous trouverez en annexe du présent rapport différentes captures, ainsi que le code des pages ayant permis la mise en place du service demandé dans le but de cette séance. Nous nous efforçons ici de donner l'idée générale du développement, en nous intéressant à des morceaux de code précis.

**Présentation des différents fichiers** Notre moteur de recherche était installé dans un site WEB virtuel nommé "site", dont le répertoire racine était situé physiquement dans C :\InetPub\virtuel.

Afin de faire fonctionner notre moteur, trois fichiers étaient nécessaires :

- query.htm : Fichier contenant la page d'accueil, permettant la saisie des termes de la recherche, redirigeant vers la page d'affichage des résultats.
- query.htx : Fichier présentant les résultats à l'utilisateur. Cette page jouait plus le rôle de patron, ou template en anglais, en donnant la forme générale des résultats.
- query.idq : Fichier de configuration de la requête au service d'indexation.

Nous avons en outre ajouté un logo, afin d'améliorer l'aspect esthétique des pages générées. Nous vous renvoyons aux captures en annexe pour le rendu final de notre application. Enfin, afin de tester, il a fallu ajouter du contenu à notre site, nous avons donc créé quelques fichiers au format texte contenant un texte donné, qu'il a été possible de rechercher ensuite au sein de notre moteur de recherche.

### 4 Approche critique

Nous avons utilisé ici des solutions à base de logiciels propriétaires. Le Disque d'installation nous était certes fourni, toutefois, il est nécessaire de noter que la license d'utilisation d'un tel produit coute près de 1350 euros (source : amazon.fr). Nous aurions pu utiliser des logiciels tels le serveur web Apache, un langage de script du type PHP et un SGDB tel MySQL, qui aurait considérablement amoindri les coûts, mais aurait éventuellement baissé les fonctionnalités offertes. Le service d'indexation doit en effet être capable de parcourir les documents Microsoft Office, ce que n'aurait sans doute pas été capable la solution alternative que nous proposons ici.

## 5 Conclusion

L'étude de la solution proposée au cours de cette séance ne nous permet certes pas de rivaliser avec les solutions professionnelles proposées par le géant google par exemple, toutefois nous pensons avoir rempli les objectifs demandés, étant donné que l'application était parfaitement fonctionnelle, et personnalisée.

## 6 Annexes

### 6.1 Captures d'écran





```
<html>

<!-- en tetes -->
<head>
<title>Formulaire de recherche - Jocelyn Gourronc et Pierre Mauduit</title>
</head>

<!-- corps de page -->
<body style="background: url('../images/logo.png');
background-position: top center;
background-repeat: no-repeat;
margin-top: 100px;">

<h3 style="margin-left: 20px;">Formulaire de recherche</h3>

<p style="margin-left: 40px;">

Veillez taper les termes de votre recherche dans la boite de dialogue
ci dessous : </p>

<!-- Formulaire de recherche Note : nous aurions pu faire le choix
d'envoyer le formulaire par POST au script query.idq, dans le but
de soulager la barre d'adresse, mais nous avons préféré rester
fideles aux scripts qui nous était fournis. -->

<form style="margin-left: 50px;" action="query.idq" methode="GET">
  <input size="32" type="text" name="CiRestriction" style="border: solid 1px blue;">
  <br/>
  <input type="submit" value="Rechercher" style="border: solid 1px blue;
background-color: white;">
  <input type="reset" value="Reinitialiser" style="border: solid 1px red;
background-color: #FFAAAA;">
</form>

<!-- Ces champs cachés sont nécessaires pour la recherche, lors du
passage au service d'indexation. -->
<input type="hidden"
name="ciScope" value="/"> <input type="hidden"
name="CiMaxRecordsPerPage" value="10"> <input type="hidden"
name="TemplateName" value="query"> <input type="hidden" name="CiSort"
value="rank[d]"> <input type="hidden" name="HtmlQueryform"
value="query.htm"> </body>

<!-- fin de page -->
</html>
```

```
<!-- Page d'affichage des résultats de recherche -->
<html>

<head>
<title>Résultat de la recherche</title>
</head>

<body style="background: url('../images/logo.png');
background-position: top center;
background-repeat:no-repeat;
margin-top: 150px;
margin-left: 70px;">
<p style="font-weight: bold;">

<!-- Si aucun document trouvé -->
<%if CiMatchedRecordCount eq 0%>
Aucun document correspondant a la requete "<%CiRestriction%>".
<!-- sinon -->
<%else%>
Documents <%CiFirstRecordNumber%> à <%CiLastRecordNumber%> des <%if CiMatchedRecordCount eq CiMax
RecordsInResultSet%> meilleurs
<%endif%>
<%CiMatchedRecordCount %> correspond a la requete "<%CiRestriction%>"
<%endif%>

</p>

<!-- affichage des résultats dans une liste HTML (ul / li) -->
<%begindetail%>
<ul>
  <li>
    <p style="margin: 0px;
font-weight: bold;
color: blue;">
<a href="<%filename%>"><%CiCurrentRecordNumber%> - <%if DocTitle isempty%><%filename%><%else%><%D
ocTitle%><%endif%></a>

    <p style="margin: 0px;
margin-left: 40px;
margin-right: 70px;
text-align: justify;
font-family: courier;
font-size: 6px;"> <%characterization%> </p>
  </li>
</ul>
<%enddetail%>

<!-- bouton précédent si nécessaire -->

<!--
Note : Nous avons préféré
construire nous-mêmes l'adresse de destination en utilisant des
liens, plutôt qu'un formulaire avec des boutons, qui nous
semblait plus lourd et pas nécessaires ici. Par ailleurs,
l'approche nous semblait plus pédagogiquement intéressante,
puisque ceci constitue une déviation par rapport aux scripts qui
nous étaient fournis.
-->

<%if CiContainsFirstRecord eq 0%>
<a href="./query.idq?CiRestriction=<%CiRestriction%>&CiBookMarkSkipCount=-<%EscapeRAW CiMaxRecord
sPerPage%>&CiBookMark=<%CiBookMark%>">
Page précédente
</a>
<%endif%>

<!-- bouton suivant si nécessaire -->
<br/>
```

```
<%if CiContainsLastRecord eq 0%>
<!--
<form action="query.idq" method="get">
  <input type="hidden" name="CiBookMark" value="<%CiBookMark%>" >
  <input type="hidden" name="CiBookMarkSkipCount" value="<%EscapeRAW CiMaxRecordsPerPage%>"
  >
  <input type="submit" value="Les <%CiMaxRecordsPerPage%> documents suivants" >
</form>
-->
<a href="./query.idq?CiRestriction=<%CiRestriction%>&CiBookMarkSkipCount=<%EscapeRAW CiMaxRecords
PerPage%>&CiBookMark=<%CiBookMark%>">
Page suivante
</a>

<%endif%>

<p><a style="text-decoration: none; color: blue; font-weight: bold;" href="./query.htm">Retour à
la page de recherche</a></p>

</body>
</html>
```

```
#
#
# query.idq : Paramétrage de la requête pour envoi au service
# d'indexation. Ce fichier a été légèrement modifié au cours du TP,
# afin de mettre en évidence des comportements différents, notamment
# la pagination des résultats (variables CiMaxRecordsPerPage,
# CiMaxRecordsInResult, etc ...)
#
#
[Query]
CiColumns=filename,size,rank,characterization,vpath,DocTitle,write
CiFlags=DEEP
CiRestriction =%if FreeText eq on% $contents "%CiRestriction%" %else% %CiRestriction% %endif%
CiMaxRecordsInResultSet=3
#CiMaxRecordsPerPage=%CiMaxRecordsPerPage%
CiMaxRecordsPerPage=2
CiScope=/
CiTemplate=/site/query.htx
CiSort=%CiSort%
#CiForceUseCi=true
```