

# OpenSearch

Créer son propre moteur de recherche dans un navigateur

2016-09-23

La technologie **OpenSearch**, créée par Yahoo! est maintenant devenue un standard. Elle permet de créer son propre moteur de recherche et de l'intégrer directement dans un navigateur. Voyons comment la mettre en place.

## EXEMPLE DE MOTEUR DE RECHERCHE

Comme très souvent avec les technologies web, la définition d'un moteur de recherche opensearch passe par le XML.

Par exemple :

```
<?xml version="1.0" encoding="UTF-8"?>
<OpenSearchDescription xmlns="http://a9.com/-/spec/opensearch/1.1/">
  <ShortName>Mon moteur</ShortName>
  <Description>Mon moteur de recherche personnel</Description>
  <Image height="16" width="16" type="image/x-
icon">http://www.mon_site.web/favicon.ico</Image>
  <Tags>recherche</Tags>
  <Language>fr</Language>
  <Contact>admin@mon_site.web</Contact>
  <Url type="text/html" method="GET"
template="http://www.monsite.web/opensearch.sh?search={searchTerms}"/>
  <Url type="application/opensearchdescription+xml" rel="self"
template="http://www.mon_site.web/opensearch.xml"/>
  <InputEncoding>UTF-8</InputEncoding>
  <AdultContent>false</AdultContent>
</OpenSearchDescription>
```

Ici, nous utilisons un script **opensearch.sh**, mais nous verrons que plusieurs solutions

sont réalisables.

Voyons donc comment construire le moteur de recherche, étape par étape...

## INSTALLATION DANS LE NAVIGATEUR

Pour l'installer dans la navigateur du client, rien de plus simple ! Il suffit de l'ajouter dans la section head d'une page web de votre site :

```
<link rel="search" type="application/opensearchdescription+xml"
href="opensearch.xml" title="Mon moteur"/>
```

Notez que l'attribut **title** doit être identique à la balise **ShortName** de votre définition.

## LE MOTEUR DE RECHERCHE

A partir de là, votre moteur de recherche est intégré dans navigateur mais il ne fonctionne pas car il ne sait pas comment interpréter la requête.

Il faut donc réaliser le coeur du moteur de recherche.

### Utilisation d'un moteur de recherche externe

Le plus simple, mais le moins paramétrable, est d'utiliser un moteur de recherche externe. Par exemple celui de Google.

Pour cela, il suffit d'interroger le moteur mais en le restreignant uniquement au site local **site:www.monsite.web** :

```
<?xml version="1.0" encoding="UTF-8"?>
<OpenSearchDescription xmlns="http://a9.com/-/spec/opensearch/1.1/">
<ShortName>Recherche</ShortName>
<Description>Rechercher sur le site</Description>
<InputEncoding>UTF-8</InputEncoding>
<Url method="get" type="text/html"
```

```
template="http://www.google.com/search?q={searchTerms}+site%3Amonsieurblog.com&hl=fr"/>
</OpenSearchDescription>
```

## Création d'un moteur de recherche interne

Dans le cas où vous désirez réaliser votre propre moteur de recherche (par exemple si vous devez rechercher dans une base de données), alors il faudra construire tous les composants.

Construisons donc notre moteur.

### *La requête de base*

Une fois n'est pas coutume, la requête n'est pas envoyée à la variable *QUERY\_STRING* du navigateur mais à la variable *REQUEST\_URI*.

Transformons cette variable en *QUERY\_STRING* :

```
QUERY_STRING=$(echo ${REQUEST_URI} | awk -F'?' '{print $2}')
```

puis traitons-la de façon habituelle :

```
search=$(echo ${QUERY_STRING} | tr '&' '\n' | grep ^search= | cut -d"=" -f2 | tr '+' ' ' | ./html2utf8.sh)
```

### Remarque

`html2utf8.sh` fait référence à un shell unix de transformation du format *application/x-www-form-urlencoded* en UTF-8.

**search** contient alors le motif de recherche. Il ne reste plus qu'à rechercher ce motif dans la source en question. Une base de données, par exemple. Ce qui donnera :

```
SELECT mon_champ
FROM ma_table
WHERE
```

```
mon_autre_champ = ${search};
```

## Mise en forme

Maintenant, il faut présenter le résultat de façon à ce qu'il puisse être affiché dans un navigateur. OpenSearch permet 3 formats : RSS, Atom et HTML/XHTML.

Comme nous générons une page dynamique, nous suivrons notre habitude d'employer XML/XSLT.

Tout d'abord, la requête doit générer du XML :

```
SELECT '<resultat>' || mon_champ || '</resultat>'
FROM ma_table
WHERE
mon_autre_champ = ${search};
```

Et il ne reste plus qu'à encapsuler le tout dans du XML grâce à un shell UNIX (ou tout autre CGI) :

```
#Génération du XML
echo '<?xml version="1.0" encoding="UTF-8"?>'
echo '<?xml-stylesheet type="text/xsl" href="opensearch.xsl"?>'
echo "<application timestamp=\"$(date +%Y%m%d%H%M%S')\">"
echo '<resultats>'
request_db "select opensearch('${search}');"
echo '</resultats>'
echo '</application>'
exit 0
```

*request\_db* est une fonction d'appel à la base de données (à définir en fonction de la base de données).

Il ne reste plus qu'à réaliser la transformation **opensearch.xsl**, en récupérant notamment la partie des résultats de la façon suivante :

```
<xsl:for-each select="/application/resultats/resultat">
<tr>
<td><xsl:value-of select="."/></td>
</tr>
```

</xsl:for-each>

## Remarque

Dans cet exemple, nous avons étudié la solution avec requête SQL pour le cas d'une recherche dans une base de données. Mais il est tout à faire possible de réaliser des recherches sur les pages web (texte) grâce à des expressions régulières en langage perl, php, script unix, etc.

Par exemple : `[A-Z0-9._%+-]+@[A-Z0-9.-]+\.[A-Z]{2,4}` permettra de retrouver une chaîne de type email.

DIY

Web

Design & Infrastructure : firenode.net ♦ Site géré par EntropyCMS ♦ Entièrement développé et motorisé par des logiciels libres ♦

Licence

Contactez l'administrateur du site

Serveur fonctionnant aux énergies 100% renouvelables