



IBM System i

OmniFind Text Search Server

Christian GRIERE
cgriere@fr.ibm.com

Common Romandie

03.05.2011

© IBM France 2007-2011

IBM System i



Plan de la session

- OmniFind Text Search Server for i
 - Présentation du produit
 - Nouveautés 7.1


OmniFind - Généralités

- Code produit : 5733-OMF
- Disponible depuis le : 09.05.2008
- Non facturable
- Doit être commandé
- Est installé indépendamment de DB2 for i (RSTLICPGM 5733OMF ...)
- Fonctionne avec les versions 6.1 (V1R1) et 7.1 (V1R2) de l'IBM i
- Recherche de mots de façon **linguistique** dans les colonnes de vos tables via **l'interface SQL**

OmniFind – Interface SQL

- Après installation d'OmniFind vous disposez :
 - d'un ensemble de procédures/scripts pour gérer OmniFind
 - Arrêt/démarrage des Text Server
 - Création/Maj/Altération/Suppression d'un index texte
 - Mise à jour immédiate d'un index texte
 - ...
 - deux nouvelles fonctions SQL (interface) :
 - CONTAINS
 - SCORE

Fonction contains - Comment ça marche ?


 **select * from omnifind.livres - Stn523p2(Stn...**

NO	TITRE	TEXTE
1	Le règne animal	Le chat mange les souris
2	Compagnie	Quand le CHAT n'est pas là les souris dansent
3	La cuisine facile	La souris d'agneau est un plat d'hiver
4	Les animaux	Les souris aiment le gruyère
5	Les félins	Les Chats n'aiment pas l'eau froide
6	Aventure	Le seigneur des agneaux

*select * from livres where contains(TEXTE, 'chat')=1;*

Combien de lignes ?

Fonction contains - Comment ça marche ?


 **select * from omnifind.livres where contains(TEX...**

NO	TITRE	TEXTE
1	Le règne animal	Le chat mange les souris
2	Compagnie	Quand le CHAT n'est pas là les souris dansent
5	Les félins	Les Chats n'aiment pas l'eau froide

Recherche linguistique :

chat = CHAT, chats, Chats, chatte, ... mais pas chaton

Fonction contains - Comment ça marche ?


 **select * from omnifind.livres - Stn523p2(Stn...**

NO	TITRE	TEXTE
1	Le règne animal	Le chat mange les souris
2	Compagnie	Quand le CHAT n'est pas là les souris dansent
3	La cuisine facile	La souris d'agneau est un plat d'hiver
4	Les animaux	Les souris aiment le gruyère
5	Les félins	Les Chats n'aiment pas l'eau froide
6	Aventure	Le seigneur des agneaux

*select * from livres where contains(TEXTE, 'agneau')=1;*

Combien de lignes ?

Fonction contains - Comment ça marche ?

 **select * from omnifind.livres where contains(TEX...**

NO	TITRE	TEXTE
3	La cuisine facile	La souris d'agneau est un plat d'hiver
6	Aventure	Le seigneur des agneaux

Recherche linguistique :

Ne se laisse pas impressionner par :

- l'apostrophe
- les pluriels irréguliers : agneaux

Fonction contains - Comment ça marche ?

select * from omnifind.livres - Stn523p2(Stn...

NO	TITRE	TEXTE
1	Le règne animal	Le chat mange les souris
2	Compagnie	Quand le CHAT n'est pas là les souris dansent
3	La cuisine facile	La souris d'agneau est un plat d'hiver
4	Les animaux	Les souris aiment le gruyère
5	Les félins	Les Chats n'aiment pas l'eau froide
6	Aventure	Le seigneur des agneaux

*select * from livres where contains(TEXTE, 'adorer')=1;*

Combien de lignes ?

Fonction contains - Comment ça marche ?

select * from omnifind.livres where contains(TEX...

NO	TITRE	TEXTE
4	Les animaux	Les souris aiment le gruyère
5	Les félins	Les Chats n'aiment pas l'eau froide

Recherche linguistique :

- Prend en charge :
 - les synonymes si définis via un fichier XML
 - les conjugaisons
- Les mots **STOP** (mots trop courants)


Fonction score - Comment ça marche ?

 **select * from omnifind.livres - Stn523p2(Stn...**

NO	TITRE	TEXTE
1	Le règne animal	Le chat mange les souris
2	Compagnie	Quand le CHAT n'est pas là les souris dansent
3	La cuisine facile	La souris d'agneau est un plat d'hiver
4	Les animaux	Les souris aiment le gruyère
5	Les félins	Les Chats n'aiment pas l'eau froide
6	Aventure	Le seigneur des agneaux

*select no, titre, texte,
integer(score(TEXT, 'souris OR chat^5')*1000)
as "Score" from livres
where contains(TEXT, 'souris OR chat') = 1
order by score(TEXT, 'souris OR chat^5') desc ;*

Fonction score - Comment ça marche ?

 **select no, titre, texte, integer(score(TEXT,...**

NO	TITRE	TEXTE	Score
1	Le règne animal	Le chat mange les souris	49
2	Compagnie	Quand le CHAT n'est pas là les souris dansent	49
5	Les félins	Les Chats n'aiment pas l'eau froide	4
3	La cuisine facile	La souris d'agneau est un plat d'hiver	2
4	Les animaux	Les souris aiment le gruyère	2

OmniFind - Opérateurs

AND ou *+* : *(chat + souris) = (chat souris)*

OR : *(chat OR souris)*

- : *exclusion (chat - souris)*

*** : *nom générique*

" " : *recherche exacte*

() : *(chat OR souris) AND rat*

** : *escape : recherche de (1+1):2 -> \ (1\+1\)\:2*

^ : *facteur de pondération (fonction score)*

OmniFind – Domaines de recherche

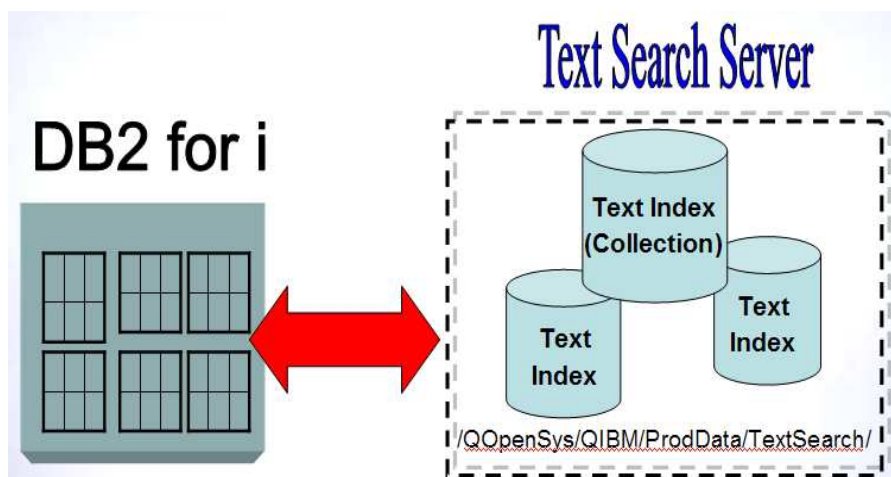
■ Types de colonne SQL :

- CHAR, VARCHAR
- BINARY, VARBINARY
- BLOB, CLOB, DBCLOB
- GRAPHIC, VARGRAPHIC
- XML (IBM i 7.1)

OmniFind – Domaines de recherche

- Familles de documents et types de documents :
 - TEXT
 - XML
 - HTML
 - INSO (INSide Out)
 - PDF
 - RFT
 - MS Excel, PowerPoint, Word
 - LOTUS 123, Freelance, WordPro
 - OpenOffice 1.1 et 2.0, Calc
 - Quattro Pro
 - JustSystems Ichitaro
 - StarOffice Calc
- 26 langues (dont les 2 chinois, le japonais et le coréen)

OmniFind – Architecture interne



OmniFind – Index texte 1/2

- Créé par un Text Server : IBM i, Linux, AIX, Windows
- Stocké dans l'IFS
- Volumétrie très différente d'un index binaire ou EVI :
 - 175.000 lignes sur une colonne VARCHAR de 32.000 caractères
 - Beaucoup + compact : 0,1 Go versus 1,7 Go (rapport : / 17)
 - Beaucoup + lent à créer : 3 heures versus 3 minutes (rapport : x 60)

OmniFind – Index texte 2/2

- Peut être créé sur :
 - une colonne d'une table DB2 for i
 - des fichiers stockés dans l'IFS (avec programmation d'une UDF, source exemple en C++ fourni)
- Maintenu de façon asynchrone :
 - Nb de modifications atteint
 - A intervalles réguliers entre plage
- Pas journalisé

OmniFind – Création d'un index texte

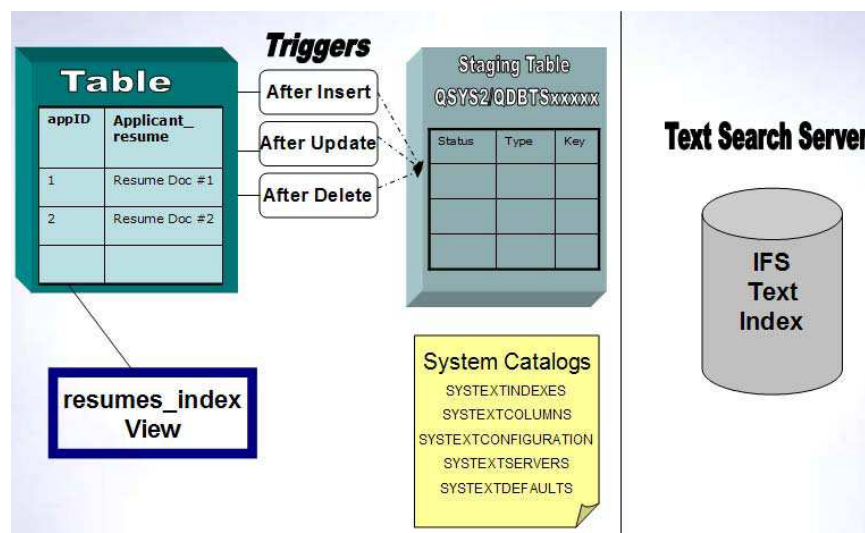
```
CALL SYSPROC.SYSTS_CREATE
('OMNIFIND', 'LIVRES_TEXTE_IDX',
'OMNIFIND.LIVRES(TEXTTE)', 'FORMAT INSO
LANGUAGE fr_FR
UPDATE FREQUENCY D(*) H(*) M(0,30)' );
```

Bibliothèque de l'index texte, nom de l'index texte

Bibliothèque.nom de la table

colonne sur laquelle l'index texte sera créé

OmniFind – Architecture interne



OmniFind – Document XML

```
bookstore>
<book genre="religion" publicationdate="1996-08-01" ISBN="0-842329-12-9"
  price="14.99">
  <title>Left Behind: A Novel of the Earth's Last Days</title>
  <author> <first-name>Tim</first-name> <last-name>LaHaye</last-name> </author>
  <slot>1234</slot>
</book>
<book genre="sports" publicationdate="2003-08-01" ISBN="1-590521-550-2"
  price="19.99">
  <title>Racing to Win</title>
  <author> <first-name>Joe</first-name> <last-name>Gibbs</last-name> </author>
  <slot>0425</slot>
</book>
</bookstore>
```

book, **title**, **author**, **slot** ... sont des éléments.
genre, **ISBN**, .. sont des attributs.

OmniFind – Recherche XML (6.1)

Exemple 1 :

```
SELECT storeid, storezip FROM books WHERE
CONTAINS(storebooks,
 '@xpath:''//title[. contains("winning")]'' ') = 1
```

Résultat : livre n° 2

Exemple 2 :

```
SELECT storeid, storezip FROM books WHERE
CONTAINS(storebooks,
 '@xpath:''/bookstore/book[@price<20.00 or
 @genre="sport"]'' ') = 1
```

Résultat : livre n° 2

OmniFind – Nouveautés 7.1 1/2

- 2 interfaces graphiques :
 - IBM Systems Director Navigator for i 7.1
 - System i Navigator 7.1
 - Fonctions supportées :
 - Liste des Text Servers
 - Arrêt/démarrage des Text Servers
 - Liste des index texte d'un Text Serveur/schéma
 - Création/MaJ/Altération/Suppression d'un index texte
- Support des namespaces niveau éléments et attributs dans les recherches XML

OmniFind – Nouveautés 7.1 2/2

- Support des comparaisons (=, <=, >=, >, < et !=) numériques avec les éléments :
 - /Book[@price > 10.00]
 - /Book[slot <= 1000] ← Nouveau
- Support des comparaisons (=, <=, >=, >, < et !=) date/datetime avec les attributs et les éléments
 - /Book[@publicationdate > xs:date("2000-01-01")]
 - /Book[purchasetime > xs:dateTime("2000-01-01T13:00:00")]
- @xmlxp remplace @xpath (toujours supporté)

OmniFind – Démonstration

- Table DB2 for i : 1304 documents (lignes)
 - 48 romans de Jules Verne
 - 33 fables de la Fontaine
 - 1218 documents d'étude niveau primaire
 - 5 livres auteurs divers
- Site de démonstration :
 - edemo.hardis.fr/OmniFind/index.jsp