



France Labs
Open Source Enterprise Search

BIG SEARCH

Cédric ULMER
Président de France Labs
SophiaConf – 3 juillet 2012

Le Big Search dans cette session

Pas forcément Big data

- On ne parlera pas d'Hadoop (ou très peu)
- Big Search:
 - Analyser plusieurs centaines de milliers de docs,
 - temps de réponse de quelques centaines de ms,
 - Gestion des fonctionnalités modernes

WARNING ! MARKETING SLIDE !!!

A propos de France Labs

- Domaine: Les moteurs de recherche open source pour entreprise
- 2 activités:
 - Consulting/sous-traitance autour de Lucene/Solr/Constellio
 - R&D (algorithmes de ranking, connecteurs)

SEARCH le mal-aimé

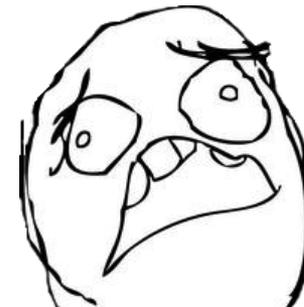
Le Cyrano des applis d'entreprise

- Utilisé dans de nombreux scénarios
- De plus en plus commun dans les usages utilisateur
- Les entreprises n'ont pas le reflexe Search
- Elles en aiment un autre (BI, KM, CMS...)

Les problématiques du Search...

... en entreprise

- ☹ Les entreprises ont de multiples défis liés au search
 - Documentation générée
 - Historique d'interactions clients
 - Données marketing
 - Catalogue produits
 - Recommandations eCommerce
 - Logistique

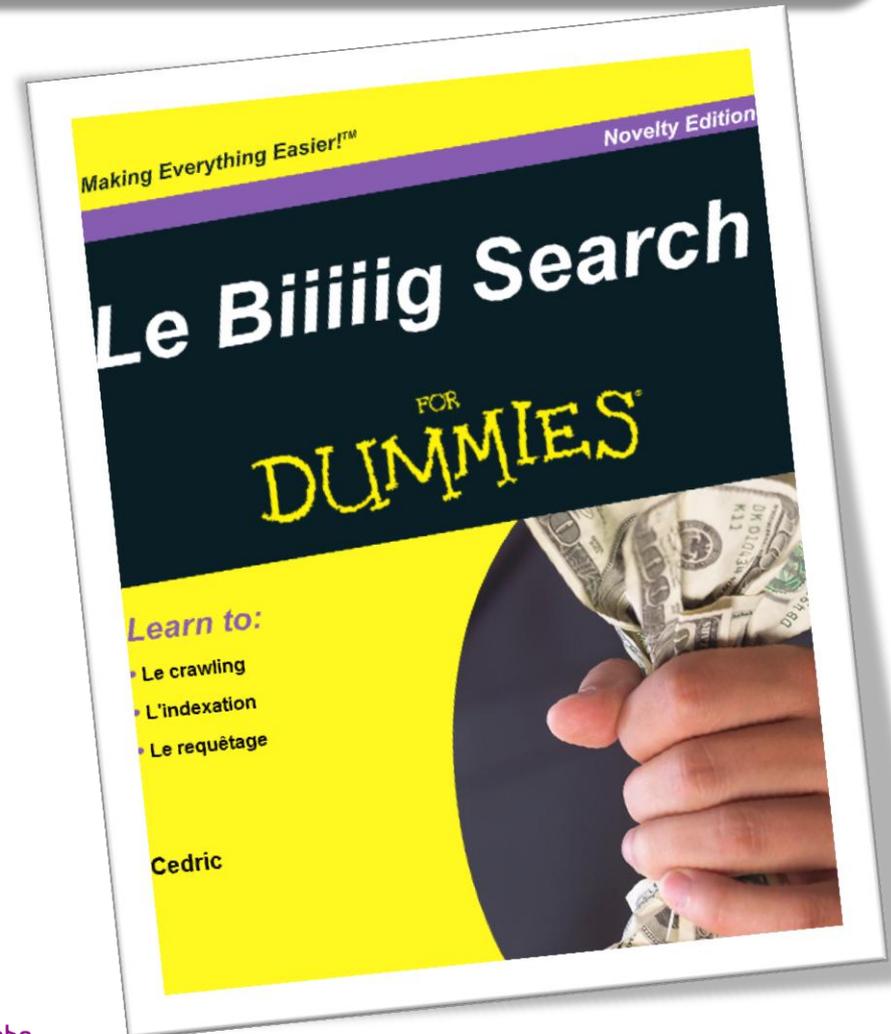


Comment ça marche ?

Le search for le nul

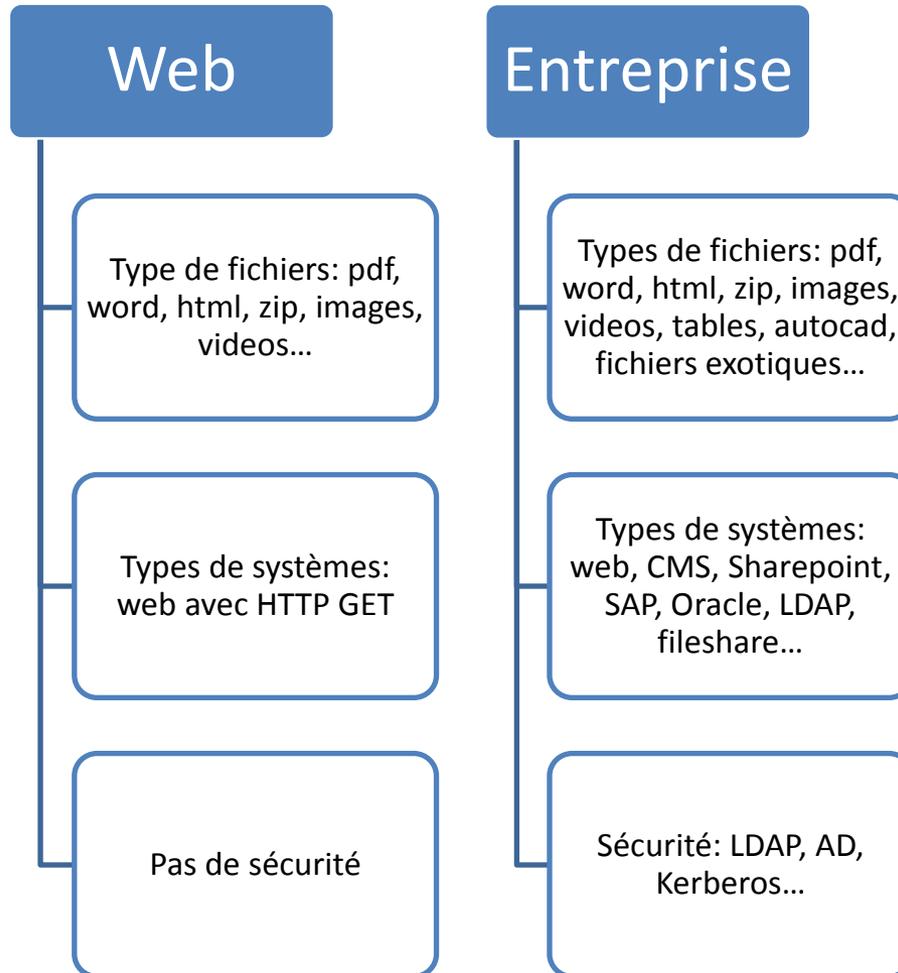
3 étapes clés:

- Le crawling
- L'indexation
- Le requêtage



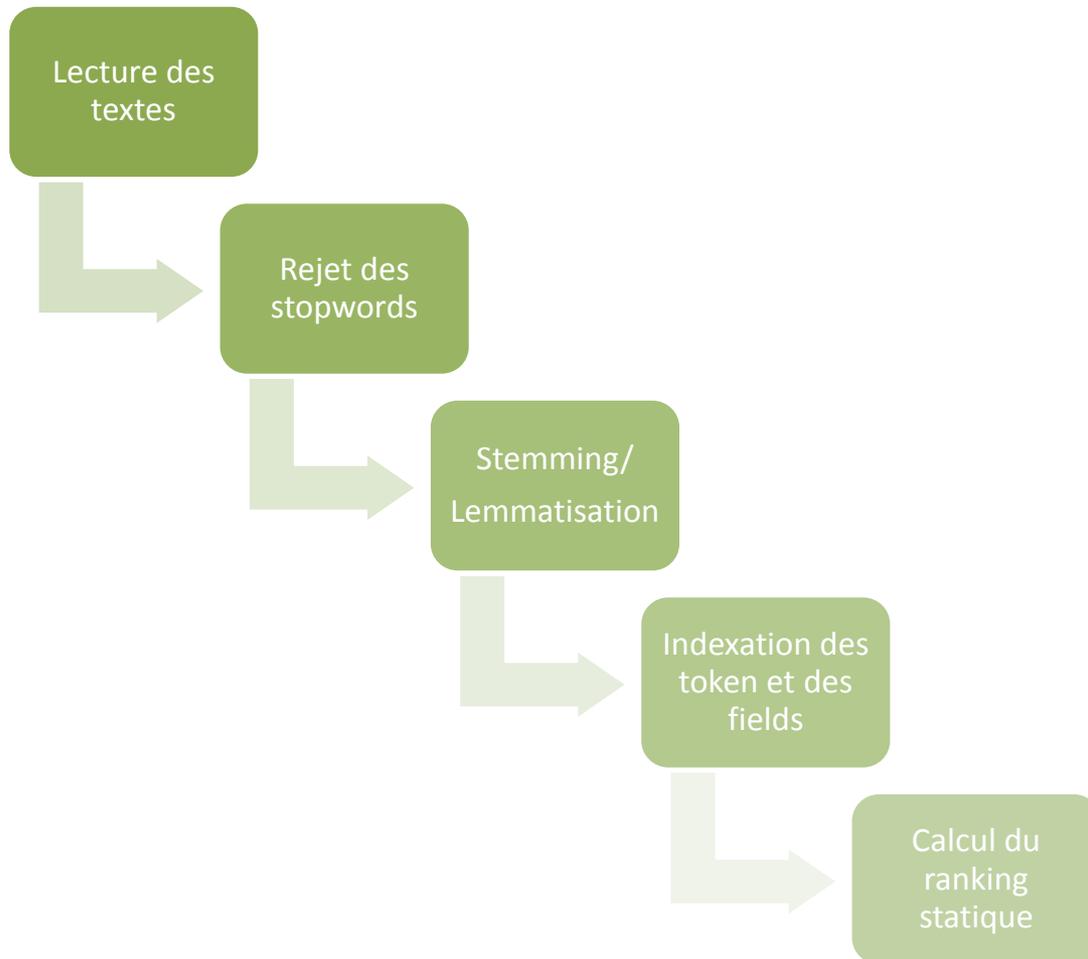
Le crawling en entreprise

Le cauchemar des connecteurs



L'indexation

La performance en background

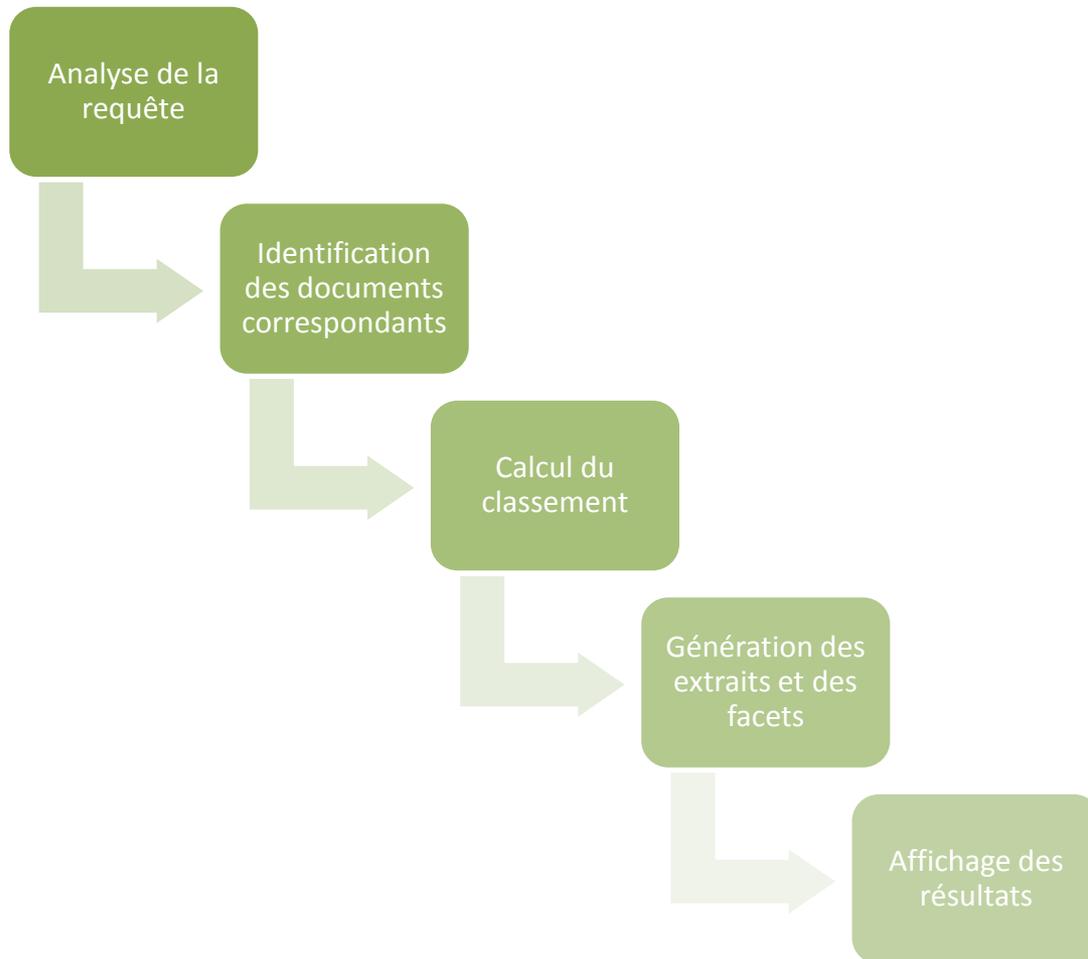


Défi big data:

- Nombre de fichiers à indexer
- Mise à jour de l'index
- Calcul du ranking

Le requêtage

La performance au runtime



Défi big data:

- Nombre de fichiers
- Calcul du ranking
- Gestion des extraits de texte

Solutions existantes

L'open source est productif

- Lucene/Solr: orienté servlet et entreprise
- Elastic Search: orienté scalabilité et distribué
- Sphinx: orienté BDD
- Constellio: orienté connecteurs et entreprise (sécurité)
- OpenSearchServer, Xapian....

Notre choix

Scotty, *support* pleine puissance!

- Problématique des entreprises: avoir un produit performant avec un support fiable.
- D'où notre choix de Lucene/Solr et Constellio:
 - Enorme communauté Lucene/Solr
 - Offre de support professionnel pour Lucene/Solr et Constellio
 - Solr 4 pensé pour le distribué



Exemples de BIG BIG Search

- LinkedIn: Calcul des recommandations en Hadoop, réinjection vers Voldemort, recherche par Lucene
- Netflix: Lucene pour le search, Hadoop pour la BI des logs. 1 TB / day
- Zoosk: Tout en Solr

Les 6 éléments à retenir

Pour votre projet de Big Search

- ① Dimensionnement
- ① Evolution de la taille
- ① Support
- ① Sécurité
- ① Disponibilité des connecteurs
- ① Interface utilisateur
- ① (France Labs ;))

Contact

Site web : www.francelabs.com

Email : cedric.ulmer@francelabs.com

Twitter : [Francelabs](https://twitter.com/francelabs)