

JP Search

Stierli Maxime

Contexte

JPIII est une application HEG qui offre un ensemble de services aux corps enseignants, aux étudiants et à l'administration.

A part des possibilités de recherche isolées, l'application ne met à disposition de l'utilisateur aucun moyen de rechercher d'une manière globale

Objectifs

- Indexer le contenu d'une base de données à des fins de recherche textuelle
- Indexer le contenu d'une base de données à des fins de classification



ORACLE
Application Express

Eléments technologiques

Oracle DB:

- Les données de l'application sont stockés dans une base de données Oracle.
- Les champs à manipuler sont de type VARCHAR2, CLOB et BLOB.

Apex 5:

- L'application preuve de concept sera développée avec Oracle Application Express (APEX) version 5.

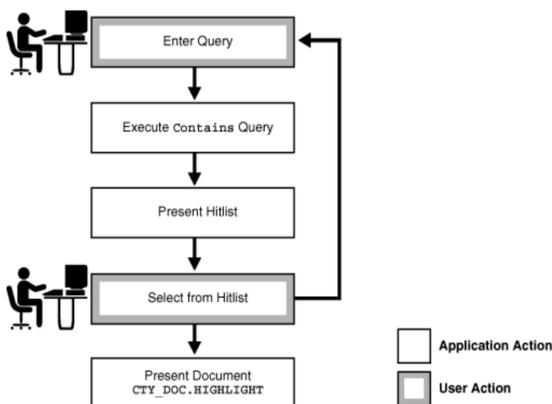
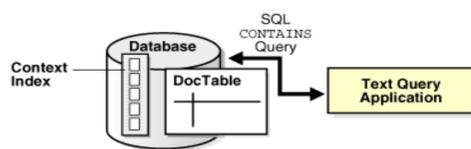
Oracle Text:

- Oracle Text est une implémentation Oracle des outils NLP (Natural Language processing)
- Il permet l'indexation et la classification par token ou par thème.

Les index Oracle Text

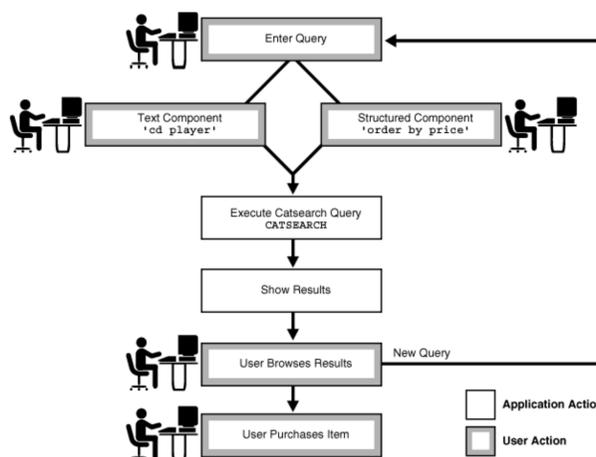
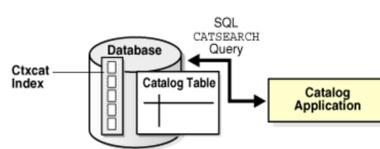
CONTEXT

- Utilisé pour les gros volumes des documents volumineux de divers formats (Word, PDF, HTML, ...).
- Mise à jour manuelle de l'index.
- Paramètres de CREATE INDEX tous supportés.
- Services de documents disponibles (mise en évidence du texte, ...).



CTXCAT

- Utilisé pour de petits documents ou fragments de texte.
- Meilleures performances en cas de requêtes mixtes.
- Mise à jour automatique de l'index .
- Supporte moins de paramètres que CONTEXT.
- Très volumineux. Prend beaucoup de temps et d'espace.
- Services de documents non supportés.



CTXRULE

- Text mining
- Classification supervisé:
- Classification non supervisé

