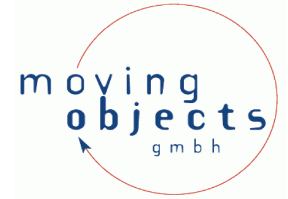


Oracle Thesauri - mehr Nutzen in der Volltextsuche

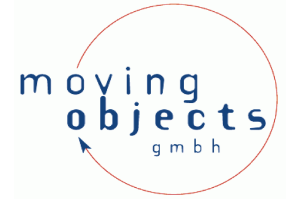
moving objects GmbH
Hamburg - Februar 2006
www.moving-objects.de

Agenda



1. moving objects gmbh
2. Thesaurus
3. Oracle Thesauri - Konzepte und Werkzeuge
 - Thesaurus & Volltextsuche
 - Beispiel Recherche
 - CTXSYS-Objekte
 - Administrations-Werkzeuge
 - Beispiele & Erfahrungen
4. Thesaurus Manager
 - Konzept & Funktionen
 - Demo
5. Nachwort

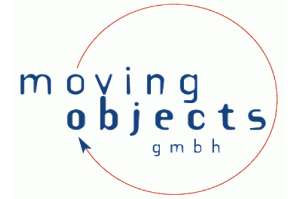
moving objects gmbh



- Eckwerte: Gründung 2001, 22 Mitarbeiter, Team seit 1992
- Inhaltliche Schwerpunkte:
 - Komplexe, datenbankgestützte Planungssysteme
 - Multimedia-Datenbanken
 - Spezialsysteme zur Informationsverteilung und -Verarbeitung
- Technische Schwerpunkte:
 - Java / J2EE -Architektur
 - Web-basierte Systeme
 - Datenbank-Entwurf und -Implementation (speziell ORACLE)
 - Client/Server-Anwendungen (4GL)
- Branchen: Fernsehen, Radio, Verlage, Umwelttechnik
- Kunden: NDR, ORF, AS, dpa, S&J, UB HH

- „Ein Thesaurus [...] ist eine geordnete Zusammenstellung von Begriffen und ihren [...] Bezeichnungen, die einem Dokumentationsgebiet zum Indexieren, Speichern und Wiederauffinden dient. Er ist durch folgende Merkmale gekennzeichnet:
 - Begriffe und Bezeichnungen werden eindeutig aufeinander bezogen („terminologische Kontrolle“), indem
 - Synonyme möglichst vollständig erfasst werden,
 - Homonyme [...] besonders gekennzeichnet werden,
 - für jeden Begriff eine Bezeichnung (Vorzugsbenennung, [...]) festgelegt wird, die den Begriff eindeutig vertritt
 - Beziehungen zwischen Begriffen [...] werden dargestellt“ (Quelle: DIN 1463)

Oracle Thesauri



- Oracle bildet Thesauri nach den Standards ISO-2788 und ANSI Z39.19 (1993) ab.
- Beziehungen:
 - Äquivalenzrelationen: Synonym (SYN), bevorzugter Begriff (PT)
 - Assoziationsrelationen: Verwandter Begriff (RT)
 - Hierarchische Relationen: Oberbegriff (BT), Unterbegriff (NT)
 - Generische Relationen: BTG bzw. NTG
 - Bestandsrelationen: BTP bzw. NTP
 - Instanzrelation: BTI bzw. NTI

Thesauri & Volltextsuche

- Indexierungsbeispiel: Franz Beckenbauer
 - Fußballspieler, Golfspieler
 - Weltmeister, Europameister, Deutscher Meister, Weltpokalsieger, ...
 - Teamchef, Vereinstrainer, Vereinspräsident, Vizepräsidenten des DFB
 - Der Kaiser, Präsident des Organisationskomitees der WM 2006
 - Kolumnist der Bild-Zeitung, Postbank-, E-Plus-, O2-Werber etc.
- Franz Beckenbauer als Thesaurus-Begriff:
 - BTI: Fußballspieler, Golfspieler
 - SYN: Der Kaiser
 - RT: Postbank, E-Plus, O2

Thesauri & Volltextsuche II

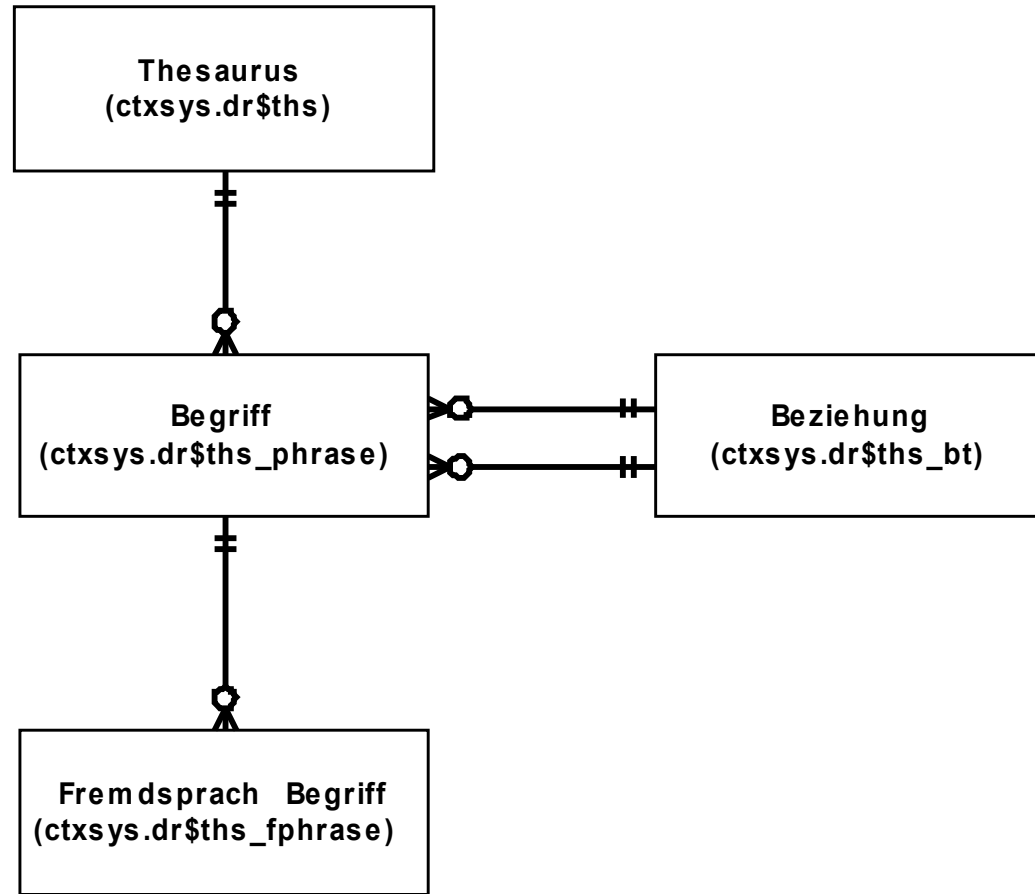
- Beispiele:
 - Suche Artikel zu „Heidi Klum“ und finde auch Artikel zu „Heidi Samuel“.
 - Die Suche nach der Stellenbeschreibung „Betriebswirt“ findet auch die des „Kaufmann“.
 - Die Recherche zum Thema „Hund“ findet auch „Dackel“

Auf die Plätze, fertig, ...

CTXSYS Objekte

- Tabellen
 - DR\$THS (Thesauri)
 - DR\$THS_PHRASE (Begriffe, Syn.-Ring, SN, Homonymhinweis)
 - DR\$THS_BT (Beziehungen)
 - DR\$THS_FPHRASE (Fremdsprachen)
- Views
 - CTX_THESAURI (Alle Thesauri)
 - CTX_THES_PHRASES (Alle Begriffe)
 - CTX_USER_THESAURI (Benutzer Thesauri)
 - CTX_USER_THES_PHRASES (Benutzer Begriffe)
 - CTX_THES_FPHRASES, CTX_USER_THES_FPHRASES (Zugriff auf Übersetzungen fehlt)

CTXSYS Thes.-Datenmodell



Administrations-Werkzeuge

- Oracle Text Manager
- Thesaurus PL/SQL-API,
d.h. also SQL-Plus bzw. SQL-Worksheet
- Ctxload
 - **Beispiel:** `ctxload -user ctxsys/ctxsys -thes -name doag -file thes_file.txt -log doag.log`
 - **Import:** `-thes`
 - **Export:** `-thesdump`
- Dokumentation
 - **Oracle® Text Reference 10g Release 2 (10.2)** Kap. 12, 14
 - **Oracle® Text Application Developer's Guide 10g Release 2 (10.2)** Kap. 9

Oracle Text Manager

- Erzeugen von Begriffen, Beziehungen
- Erfahrungen und Probleme
- Achtung: Benutzung nur über ctxsys-Account.

Auf die Plätze, fertig, ...

Thesaurus API

- **CTX_THES Package - DML:**
 - CREATE_THESAURUS, ALTER_THESAURUS, DROP_THESAURUS
 - CREATE_PHRASE, ALTER_PHRASE, DROP_PHRASE
 - CREATE_RELATION, DROP_RELATION
 - CREATE_TRANSLATION, UPDATE_TRANSLATION, DROP_TRANSLATION
- **CTX_THES Package - QL:**
 - BT, BTG, BTI, BTP / NT, NTG, NTI, NTP (Thes./Phrase/Level)
 - PT, RT, SN, SYN, TT (Thes./Phrase)
 - TR, TRSYN (Thes./Phrase/Lang.)
 - HAS_RELATION, THES_TT, OUTPUT_STYLE

Thesaurus Beispiele mit CTX_THES

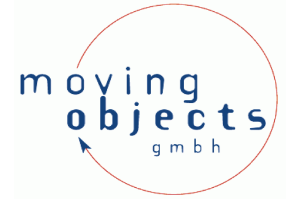
- Erzeugen von Thesauri, Begriffen, Beziehungen etc.
- Ausgabe von Begriffen und Beziehungen
- Erfahrungen und Probleme

Auf die Plätze, fertig, ...

Oracle Thesauri - Ergebnisse

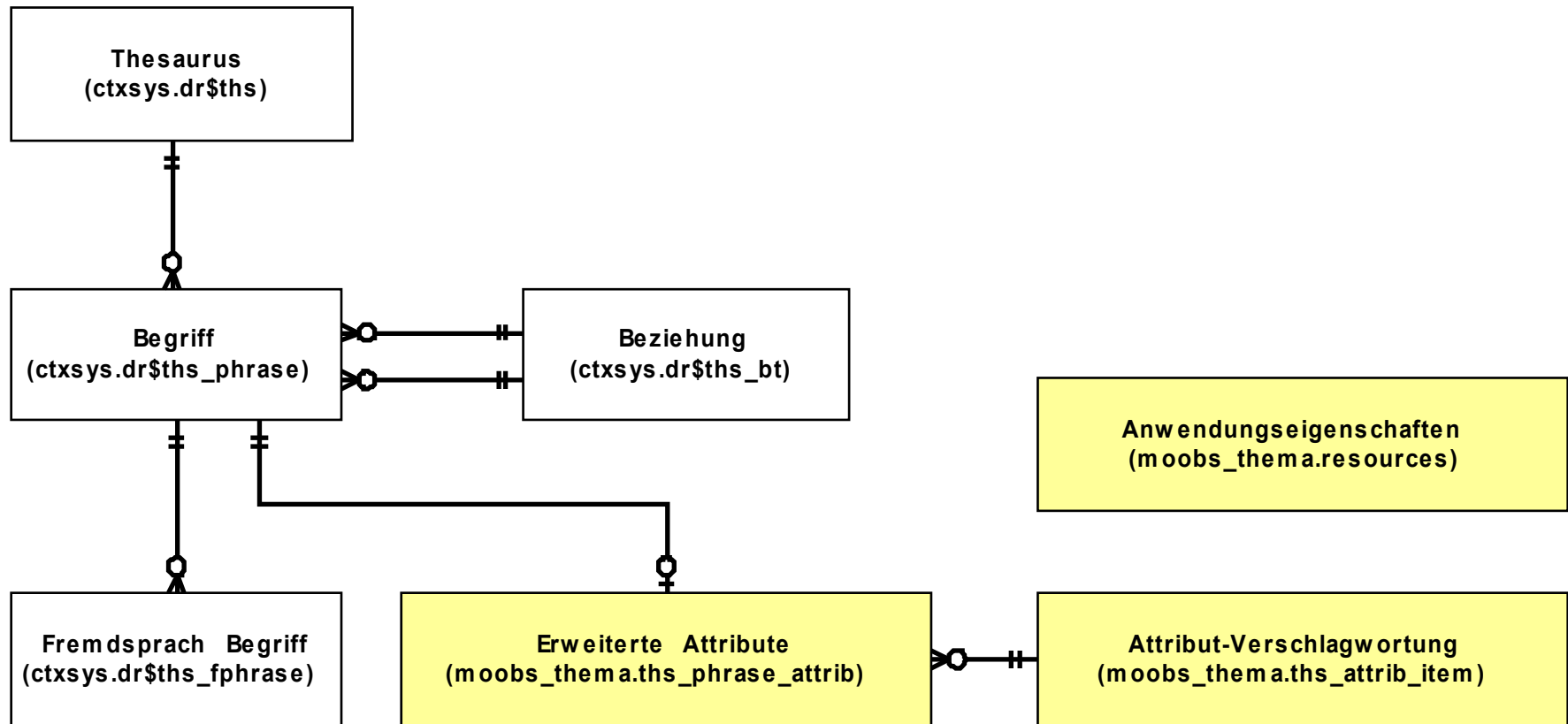
- Bedienung des Text Managers gewöhnungsbedürftig
- Keine Rechte-Rollen (alle arbeiten als „Superuser“)
- Rich-Client mit viel funktionalem Ballast
- Keine Zusatzinformationen / Freigabe

TheMa – Thesaurus Manager für Oracle®

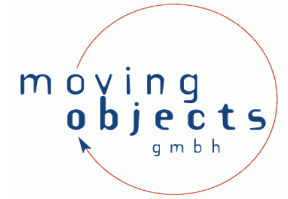


- Oracle10g-konformer webbasierter Thesaurus Manager
- Effizientes Verwalten von Thesauri
- Rechte-Rollen: Gast, Redakteur, Thes.-Manager, Admin
- Weitere Informationsfelder zu einem Begriff
z.B. Sprachschicht, Erläuterung oder Beispiel
- Schlanke J2EE-Anwendung auf Basis von OC4j bzw. JBoss 4

TheMa - Datenmodell



TheMa – Demonstration



- Erzeugen von Thesauri, Begriffen, Beziehungen etc.
- Ausgabe von Begriffen und Beziehungen
- Erfahrungen und Probleme

Auf die Plätze, fertig, ...

Nachwort



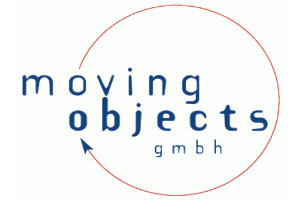
- Entwicklung von TheMa im Rahmen eines Projektes bei Axel Springer
- TheMa ist ein eigenständiges Produkt, welches käuflich erworben werden kann.
- Die Performance ist trotz großer Thesauri (27-139 tsd. Datensätze) sehr gut.
- TheMa läuft einwandfrei mit UTF-8 Oracle-Datenbanken und unterstützt UTF-8 vollständig.

Resumee

- Oracle Thesauri stellen einen wesentlichen Benefit bei der Volltextsuche dar.
Dabei gehen die Möglichkeiten der Thesaurusgestützten Suche z.B. weit über Stemming, Fuzzy oder SQE hinaus.
- Alternative Nutzung
 - Neben dem „kontrollierten Vokabular“ können Thesauri auch als „Suchkorrektur“ z.B. in e-Shops eingesetzt werden.



Fragen?



Wir danken für Ihre Aufmerksamkeit

Haben Sie noch Fragen ?

Wir stehen Ihnen gerne zur Verfügung:

info@moving-objects.de

+49 40 57 205 - 0