

DOAG Konferenz 2010  
Kongress-Zentrum Nürnberg

16. November 2010

# Lolap – Linguistische Olap

**Effiziente Erweiterung der Suchfunktion  
in Oracle in Bezug auf die deutsche Sprache**

Vosshall / Faix

# Mitarbeiter & Arbeitsgruppen

- *W. Vosshall*
- *Dipl. Kaufmann*
- *Idee und Aufgabe*
- *M. Faix*
- *Diplom Biologe – Codierung und Recherche*

# Ausgangslage

- **Isaac Asimov – I Robot**
- 23
- Abstract
- Relationale Datenbanken sind im Grenzwertfall eindeutig. Im Paradies der normalisierten Datenbank gibt es weder doppelte Datenhaltung noch unsaubere Daten. Die Abfragesprache SQL basiert auf der Annahme der Eindeutigkeit. In exakten Wissenschaften wie der Mathematik mag man das Paradies erreichen. Je weiter man in den Bereich der Lebenswissenschaften eindringt desto mehr stösst man in das Fegefeuer der unsauberen und doppelten Daten vor. Ob das im Bereich der Sicherheitsdatenblätter in der Chemie Rezepturen sind oder auf der anderen Seite die Auswertung von Arztbriefen. Ein menschliches Gehirn kann Mehrdeutigkeiten, Unsauberkeiten in der Sprache und Wortverdrehungen erkennen. Eine Maschine nicht. Das Sinnhafte Erfassen von Texten ist dementsprechend Thema der KI.

# Was haben wir gemacht?

- Wir haben zwar keine Mind Machine produziert, mit unserem Oracle Package LoLap können wir aber mindestens die Trefferquote von Anfragen sinnvoll erhöhen. Zudem können OLAP Mechanismen davon profitieren. Erfolgreich haben wir Lolap im Cleansing Prozess eingesetzt. Wir haben die Funktion in den Transform Prozess (T) des ETL Verfahrens eingebettet.
- Textorientierte Materialien können so der Auswertung mit OLAP Programmen wie Hyperion geöffnet werden. Es bieten sich so der weiter Aufschluss von Materialien in Orbis an.

# Methoden

- Unser Package erleichtert die Suche in Volltextsegmenten. Records werden nach bestimmten Kriterien mit in die Ergebnismenge übernommen. Wir haben zunächst die Soundex Funktion ersetzt. Unsere Funktion „Koelsch“ entspricht dem Aufruf von Soundex, das Ergebnis auf der Funktion entspricht aber der Definition des Koelner Algorithmus.
- `klang:= Koelsch(„String“)`
- Der Koelner Algorithmus kommt dem Klang des geschriebenen Wortes näher als die Funktion Soundex, die auf das Lautbild der amerikanischen Sprache optimiert ist. Das Ergebnis der Funktion kann dann auch zur Transformation im ETL Prozess genutzt werden.
- Unser Package ist in PL/SQL geschrieben, damit die allgemeine Lesbarkeit des Codes gewährleistet ist. Zugleich kann man sich des Java Pool in Oracle sparen.
- Eine ungetestete Java Variante liegt vor.
- Unser Package enthält zudem einen graphischen Theasaurus auf Basis des transformierten Klangwerts. In den Live Sciences sind Synonyme geläufig. Das gilt speziell auch für die Medizin – und dort auch für die Dimensionen.
- Die Begriffe Flusssäure, Fluss-Säure und HF sind Synonym zu betrachten.

# Methode – Die Funktion

- lol\_syn(„string§)
- gibt eine Liste von Synonymen aus. Mit der Funktion IN kann man dann auf Existenz prüfen.
- Wir testen die hierarchische Organisation des Thesauruses unter Verwendung von „Connect By“ Operator. Viele Dimensionen in der Biologie sind in Hierarchien geordnet.
- Das Lolap Paket enthält auch einen kleinen Stringverbinder, man kann damit alle Felder eines Record analysieren.
- Diese beiden Ebenen des Lolap Paketes funktionieren und sind im Einsatz. Lolap kann so eine Bereicherung einer relationalen Datenbankstruktur in eine Textdatenstruktur hinein sein. Im Gegensatz zu Textmining Tools wie die Tools von Google sind so auch Datenanalysen möglich.
- Wir arbeiten nun an einer fehlertoleranten Erweiterung mit Hilfe des Needle Man Wunsch Algorithmus – Wortdreher würden dann auch toleriert werden.
- Gegenüber dem bewährten Package von Oracle Context gibt es Vorteile in der Transparenz und in der Geschwindigkeit. Die Synonyme sind ladbar. spezielle Arbeitsgruppen können sich so austauschen.

# Ergebnisse

- Loloap ist ein Erweiterungs- Package dass sowohl die Suche als auch die Datenanalyse nach Scheer eröffnet. Zwar wird kein Textverständnis erzeugt aber doch die Erschließung von Texten ermöglicht – auf einfache und smarte Art und Weise. Im medizinischen Bereich böte sich Einbettung des UMS Methathesaurus
- speziell für Orbis an.

# Ergebnisse

- Die Arbeit kann als Funktion für statistische Analysen wie auch für schnelle und genaue Recherchen genutzt werden



## Beispiel:

# Temperatur-Eingabe in verschiedenen Ländern

- Von der Datenbank zur Anzeige:



## Drei-Ebenen-Struktur

**Backend**

Datenbank:  
ORACLE 10g

**Server**

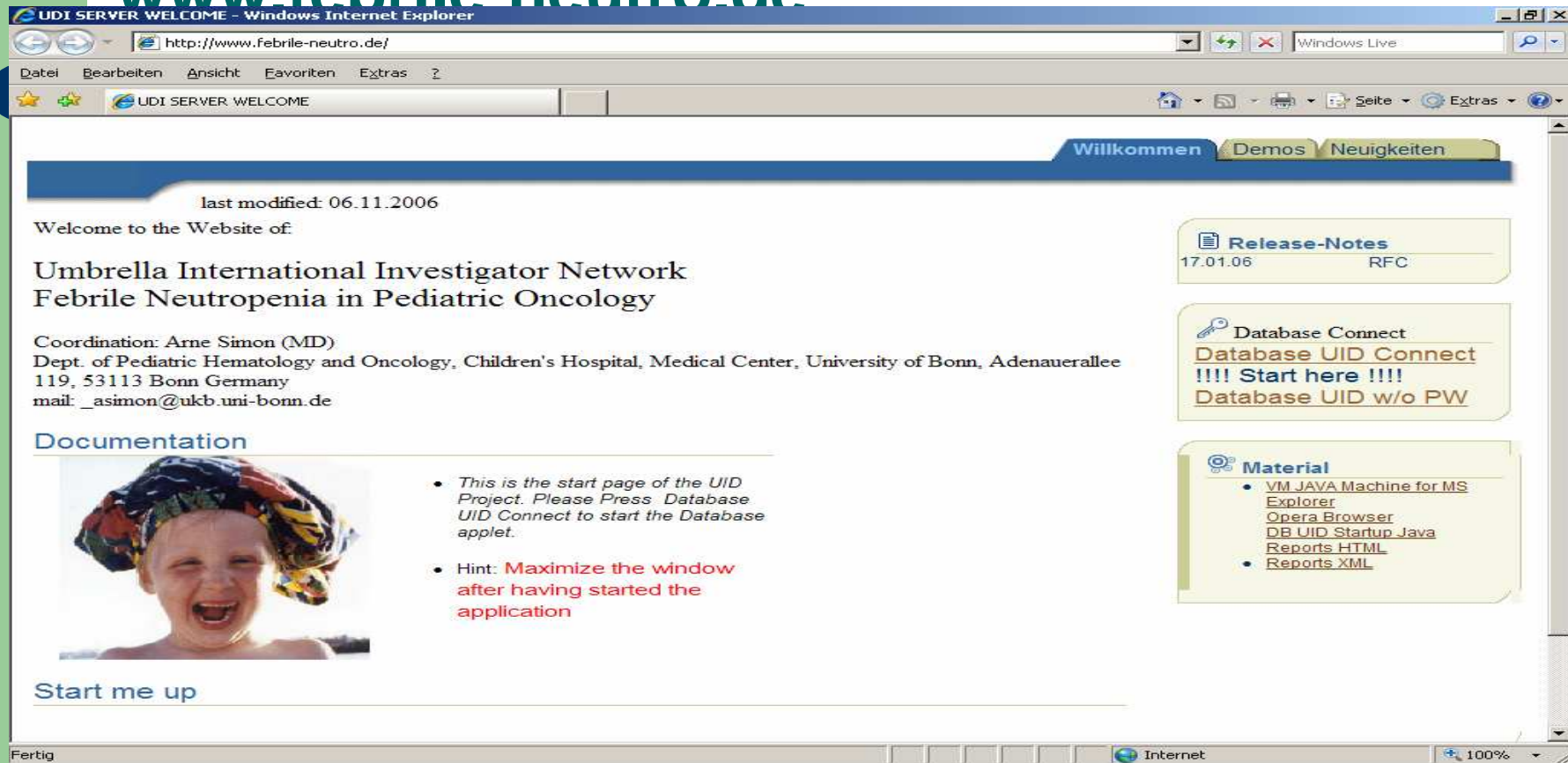
ORACLE  
Application Server  
(ohne BI)

**Frontend**

Client  
(Browser + Java)

- nur Browser-Management
- kein Feedback an die Datenbank
- Datenbank-Unabhängigkeit

# Screenshot: „Live and Alive“ www.febrile-neutro.de



# Screenshot: Datenbankerfassung

Oracle Application Server Forms Services - Windows Internet Explorer

http://kooplog.eisfair.net/forms/frmservlet?form=uid.fmx&userid=z9075m/z9075m@uid

Oracle Application Server Forms Services

UDI Project

Basic Patient Data | Basic Treatment | Reassessment | Physical Exam and RF | Lab. Values | Outcome | Special Data

Patient Id: 2907

Current\_User: Z9075M

Date\_of\_event:

Date\_of\_birth:

SurName:

Status:  Open  Compl...  Closed

SAVE\_Interim SAVE\_Complete SAVE\_Closed + Delete Logout

Admin Cultures

Admin Antibio

Admin\_Cancer

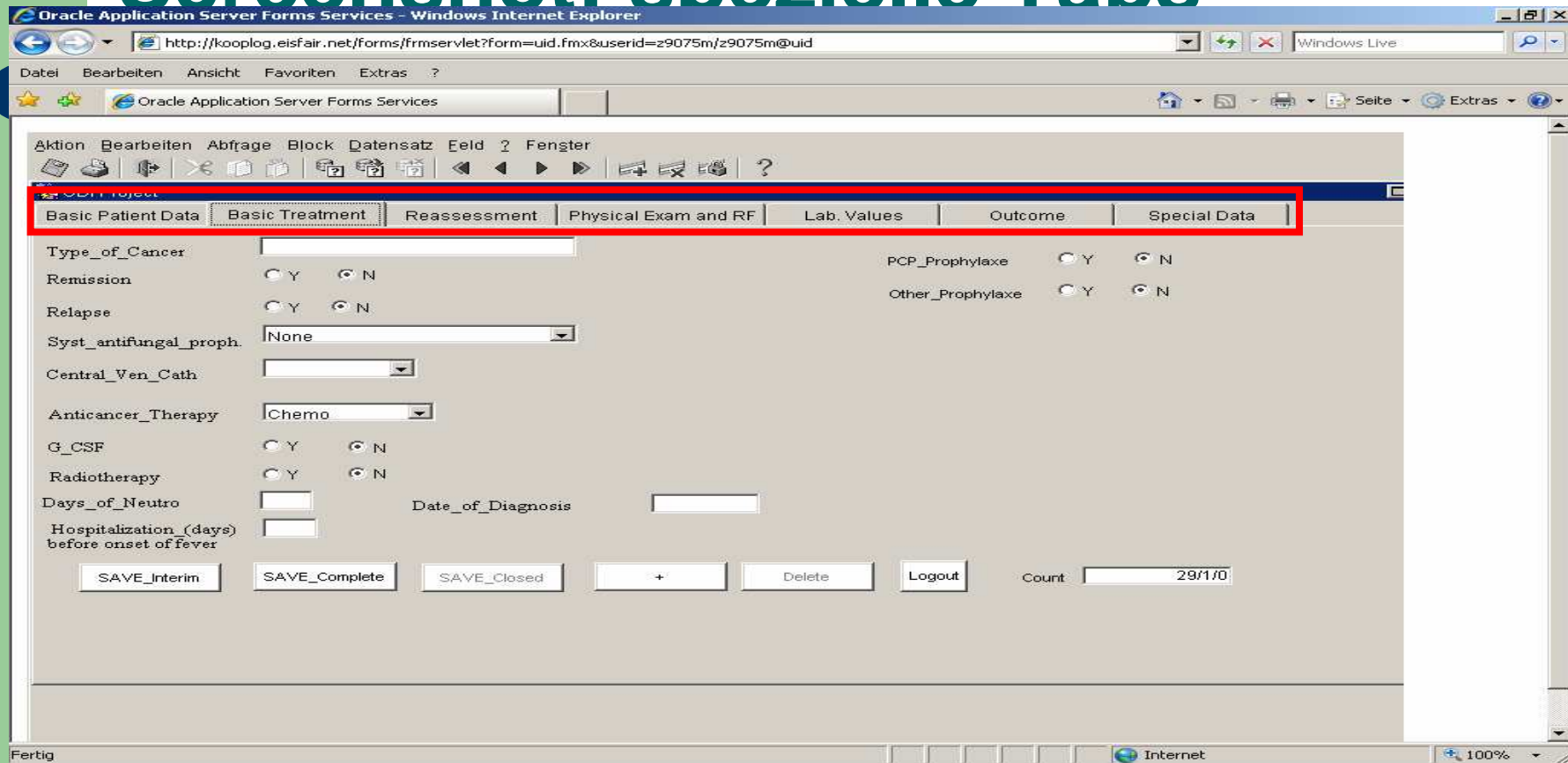
Vers.: 08/02/09 mif

Count: 29/1/0

Count: Datasets Open/Confirmed/Closed

Fertig Internet 100%

## Screenshot: spezielle Tabs



## Derzeitige Anwender

- Uni-Kinderklinik Düsseldorf
- Insel Spital Bern
- Kinderonkologie Bonn
- Bundeswehr Institute
- Defense Forces Canada
- Mil Headquarters Wiesbaden
- Nephrose-Register

## Kritikalitäten

- Firewalls
- Sicherheitsrichtlinien der Browser
- falsch konfigurierte Browser
  
- Referenzumgebung:
  - Standard-PC am Ende eines normalen Carriers (Telekom)

## Quellen

- Aktuelle Veröffentlichung:
  - Laborjournal: Ausgabe 11, Seite 98
- Diplomarbeit Wolfgang Stolz
- DOAG
- [www.febrile-neutro.de](http://www.febrile-neutro.de)
- [nephro.bitclix.de](http://nephro.bitclix.de)



*Erstellung der Präsentation:  
Markus Boll*

## Kontakt

- Adresse: IPM Koeln
- Matthias Faix
- Telefon: 0221 6778991
- E-Mail: Matthias.Faix@koeln.de
- Skype: a1594656