ezDL: An Interactive Search and Evaluation System

Workshop on Open Source Information Retrieval (OSIR) @ SIGIR 2012

Thomas Beckers, Sebastian Dungs, Norbert Fuhr, Matthias Jordan, Sascha Kriewel, Vu Tuan Tran (Demo)

Universität Duisburg-Essen, Germany

Portland, USA, August 16th 2012
Outline

1. Introduction and Concepts
2. Architecture
3. Evaluations
4. Use Cases
5. Conclusion
What is ezDL?

ezDL is a...

- interactive search tool,
- development platform for search applications and
- system for different types of user-oriented evaluations.
Desktop frontend
Major concepts as an interactive search tool

- Digital Library Life Cycle (Paepcke)
Support for tactics and stratagems (Bates)
Proactive support
Major concepts as development platform

- Easy integration of new data sources (DLs)
- Extendable user interface
- Plugin system (OSGi)
Support for eye-tracking with dynamic Areas of Interest (AOIs)
Automated logging of user and system actions
Implementation

- Licensed under GPLv3 (other licenses also possible)
- Fully implemented with Java 1.6 (1.7 is planned) ⇝ no legacy code
- Unit tests that help maintain stability
- JavaDoc documentation available
Architecture

- **Backend**
  - Service-oriented architecture (Agents)
  - Message-driven communication (via JMS-based infrastructure)

- **Frontend**
  - Desktop (Swing)
  - Web (Google Web Toolkit)
  - Mobile devices (Android, iOS), under development
Architecture

Client 1
Client 2
Client 3
Client 4
...
Client m

MTA 1
MTA 2
MTA n
...

Directory
Query History Agent
Repository Agent
Search Agent

Communication bus (e.g. CORBA or JMS)

Backend

ACM Wrapper
IEEE Wrapper
DBLP Wrapper
Wiley Wrapper
Wrapper for data sources (e.g. DLs)

- Wrapper toolkit for HTML-based sources
  - currently used for ACM DL, SpringerLink, PubMed . . .
- Connectors for common IR systems
  - Apache Lucene/Solr
  - Terrier, Lemur, MG4J (in development)
The desktop frontend architecture

- Search Tool
- pLib Tool
- Detail Tool
- Query History Tool
- Docking Framework
- Plugin System
- Import/Export
- Logging
- Communication

Internal Messaging Dispatcher

Client → Backend

MTA 1 → MTA 2 → MTA n
Extending ezDL

- New data sources
  - Patent databases: Google Patents, ESPACENET, CLEF-IP
  - Medical databases
  - ...

- New features
  - tool (desktop frontend)
  - backend agent (if required)
Extending ezDL

- New visualisations
  - e.g. Hypergrid visualisation of search results
1. Introduction and Concepts

2. Architecture

3. Evaluations

4. Use Cases

5. Conclusion
Built-in support for user evaluations

- Logging user actions

Example

User searches for information retrieval in 2 connected DLs (DBLP, ACM)

event:
  name: "search"
  clientTimestamp: 1/4/2012 15:26:32,1234
  timestamp: 1/4/2012 15:26:32,3456
  sequenceNumber: 10
  parameters:
    query: "information retrieval"
    sources: dblp, acm
Built-in support for user evaluations

- Tracking AOIs
Built-in support for user evaluations
Built-in support for user evaluations

- Fixed layout on screen
- Loading predefined perspectives
- Splash screen for choosing evaluation settings
Evaluation infrastructure

ezDL frontend
- Questionnaire tool
  - Evaluation web app.
  - Experiment
  - IR engine
  - logging data
  - eye tracking data
  - questionnaire answers

ezDL backend
- Eye tracking tool
- Experiment database

Beckers et al. (UDE)
zDL
August 16th 2012
Evaluation infrastructure (cont’d)

Evaluation web app.

- Start
- Training Task
- Pre-Exp Questionnaire
- Pre-Task Questionnaire
- Task
- Post-Task Questionnaire
- Post-Exp Questionnaire
- End

Log Viewer

Analysis Tool (Example)

Collected Data

ezDL System

Log Data

Session Data

Questionnaire & Session Data

Database

Beckers et al. (UDE)
1 Introduction and Concepts

2 Architecture

3 Evaluations

4 Use Cases

5 Conclusion
Khresmoi project

- Multilingual + multimodal search in the medical domain
- Use cases
  - General public
  - Clinicians and general practitioners
  - Radiologists
Use Cases

Web-based frontend for general public

Search query: HEA tumor staging notation

Detection Kits, Chemical Agent, Colorimetric Tube

- N-acetyl-L-asparaginase
- Phenylalanine-hydroxyethylamino-poly(L-asparaginyl-2-methyl-1-oxo-
  riboflavin)
- Poly(hydroxyethylaspartamide-co-dimethylaminopropylaspartamide)
- Poly(1-acrylactyl-co-hydroxyethyl acrylate)
- Poly(hydroxyethylaspartamide-co-aspartic acid)

Warfarin Versus Aspirin

Clinical Trial

The purpose of this study is to determine the benefit of warfarin versus aspirin in patients with atrial fibrillation. The study will randomize patients to either warfarin or aspirin and follow them for one year. At the end of the study, the number of major bleeding events will be compared between the two groups.

Evaluation of Patients

Clinical Trial

An important mission of the National Institutes of Health Clinical Center is to evaluate new treatments for patients with serious and life-threatening illnesses. One such trial is evaluating a new drug for treating cancer. The trial will randomize patients to receive either the new drug or a placebo and follow them for one year. At the end of the study, the number of patients who respond to the new drug will be compared to those who receive the placebo.

Magnetic Resonance Imaging of the Blood Vessels of the Heart

Clinical Trial

This study is directed at evaluating patients with "myocardial ischemia" or poor blood supply. National Institutes of Health Clinical Center, 9000 Rockville Pike, Bethesda, United States

The BENEFIT Trial: Evaluation of the Use of an Antiparasitidal Drug (Benznidazole) in the Treatment of Chronic Chagas' Disease

Clinical Trial

The purpose of this study is to determine if 60 days of treatment with an antiparasitidal drug... Population Health Research Institute - McMaster University, Hamilton, Canada (ends December 2010)

The BENEFIT Trial: Evaluation of the Use of an Antiparasitidal Drug (Benznidazole) in the Treatment of Chronic Chagas' Disease

Clinical Trial

The purpose of this study is to determine if 60 days of treatment with an antiparasitidal drug... FAPEA-Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, Brazil (ends December 2010)

ESCAP: Supervised Exercise for Patients With Coronary Heart Disease in the Primary Care Setting

Clinical Trial

This study is directed at evaluating patients with "ischemic heart disease" or "myocardial infarction." National Institutes of Health Clinical Center, Rockville Pike, Bethesda, United States
Use Cases

Desktop frontend for general practitioners

Beckers et al. (UDE)

ezDL

August 16th 2012 28 / 35
Use Cases

CAIR project

- User-oriented clustering of search results
- Clustering was added to ezDL (clustering agent and adapted search tool)
- Collections:
  - Amazon Book database
  - Wikipedia
  - Web search
iPerFedPat: Federated Search for Patents
(Mike Salampasis, U. Thessaloniki, Greece)
Use Cases

iPerFedPat New Search Tools
Result Clustering, Entity Mining
Conclusion

Introduction and Concepts

Architecture

Evaluations

Use Cases

Conclusion
ezDL is a
- interactive search tool,
- development platform and
- system for user-oriented evaluations

Efficient and effective IR systems need an appropriate user interface

Framework to avoid “reinventing the wheel” when developing search UIs

Participation by other persons/groups is welcome
Thank you for your attention.