DIRECTions: Design and Specification of an IR Evaluation Infrastructure

Maristella Agosti, Emanuele Di Buccio, Nicola Ferro, Ivano Masiero, Simone Peruzzo, Gianmaria Silvello

Information Management Systems Research Group
Department of Information Engineering
University of Padua, Italy

CLEF 2012 - Information Access Evaluation meets Multilinguality, Multimodality, and Visual Analytics
Rome, Italy, September 17-20, 2012
Outline

- Motivations
- The Conceptual Modeling
- The Architecture of DIRECT
- How to Access the Data: The Data Cube
- Conclusions and On-Going Work
Motivations

- IR Evaluation is challenged by variety and fragmentation
  - diverse tasks and metrics
  - heterogeneous collections
  - different systems and approaches

- We need to facilitate IR evaluation by allowing for (fair) comparisons and by preserving and providing access to the data (collections, settings, outputs) over time

- Provide visual interaction with experimental data (visual analytics)
The Main Goal

Deliver a unified infrastructure and environment for data, knowledge, tools, methodologies and the user community in order to advance the experimental evaluation of information systems.
Conceptual Modeling

We have to model an heterogeneous and highly diversified reality of interest

The first step is to identify the main functional areas of experimental evaluation
The Core

The core of the infrastructure

EVALUATION ACTIVITY

Activities aimed at evaluating applications, systems and methodologies

It includes:

Campaign
Trial
Education
The Core’ Satellites: Experimental Collection

A **traditional IR** evaluation environment

**EVALUATION ACTIVITY**

**EXPERIMENTAL COLLECTION**
The Core’ Satellites: Experiment

The **scientific data produced**

- **EXPERIMENT**
- **EVALUATION ACTIVITY**
- **EXPERIMENTAL COLLECTION**

Run Guerrilla Living
The Core’ Satellites: Experiment

RUN

GUERRILLA

LIVING

COMPONENT

SYSTEM

APPLICATION

EVALUATES

TESTS

IS PART OF

IS INTEGRATED IN

EXPERIMENT

EVALUATION ACTIVITY

EXPERIMENTAL COLLECTION

The scientific data produced by the Run Guerrilla Living Experiment evaluates the tests associated with the system composed of the application. Each component is part of the system and integrated in the application.
The Core’ Satellites: Resource Management

It handles
Resources

EXPERIMENT
EVALUATION ACTIVITY
EXPERIMENTAL COLLECTION
RESOURCE MANAGEMENT

interactions between
Users
Groups
Roles

Access Control and Concepts
The Advanced Features: Measurement

- BIBLIOGRAPHICAL
- EXPERIMENT
- VISUAL ANALYTICS
- EVALUATION ACTIVITY
- EXPERIMENTAL COLLECTION
- RESOURCE MANAGEMENT
- METADATA
- MEASUREMENT
The Advanced Features: Measurement

- Conducted From
- Measure
- Assigns
- Experiment
- Measure
- Metric
- Task
- Estimate
- Descriptive Statistic
- Pool
- Topic
- METADATA
- Statistical Test
- Statistical Analysis
- Statistical Analysis
- Elaborated From
- Calculated From
- Task
- Calculates
- Metric
- Computes
- Experiment
- Pool
- Measure
- Descriptive Statistic
- Estimate
- Task
- Calculates
- Metric
- Computes
The Advanced Features: Metadata

The **description and enrichment** of the resources

- BIBLIOGRAPHICAL
- EXPERIMENT
- VISUAL ANALYTICS
- EVALUATION ACTIVITY
- EXPERIMENTAL COLLECTION
- RESOURCE MANAGEMENT
- MEASUREMENT
- METADATA
The Advanced Features: Visual Analytics

- Bibliographical
- Experiment
- Evaluation Activity
- Experimental Collection
- Resource Management
- Measurement
- Metadata

Visual Analytics
The Architecture of DIRECT
The Architecture of DIRECT
Task, Experiments and Metrics

How users can access experimental data about task, experiments, and related metrics in order to process them?
The Data Cube

\[/task/{id.tsk}{ns.tsk}/metric\]
Slicing the Cube

\[ /task/{{id.tsk}\{ns.tsk}}/experiment/{{id.exp}}/metric \]
Slicing the Cube

(a)  

(b)  

(c)