KOIOS: Intuitive Access, Analysis, and Visualization of Structured Environmental Information

Veli Bicer: FZI Forschungszentrum Informatik, Karlsruhe, Germany
Thanh Tran: Institute AIFB, KIT, Karlsruhe, Germany
Andreas Abecker, Radoslav Nedkov: disy Informationssysteme GmbH, Germany
Large amount of environmental data

- Environmental issues stir public interests
- Growing amount of data
  - Public access through EU directive 2003/4/EC
  - PortalU (Germany) http://www.portalu.de/
  - EDP (UK) http://www.edp.nerc.ac.uk
  - Envirofacts (USA) http://www.epa.gov/enviro/index.html
- Linking data in international context
  - Local government databases of environmental part of LOD cloud
  - Linked environment data for the life sciences
Opportunity: mass dissemination and consumption of environmental data

- Increase transparency, awareness, responsibility, protection
Opportunity: mass dissemination and consumption of environmental data

- Complex results
  - CO emission values around Karlsruhe area in Germany

- Analytics
  - CO emission values around Karlsruhe area in Germany
    - Sorted by year
    - Bar chart
  - Emission values of US and Germany
    - Compare average
    - Timeline visualization
Challenges: **intuitive access and visualization** of structured environmental data and analytics

- The percentage of people who actively find environmental information is significantly lower than those who have frequent access to it!

- Complex structured queries
  - Knowledge of the underlying data / query language

- Complex structured data
  - Heterogeneity and distribution of environmental data is overwhelming

- Complex structured results
  - Understanding results and extracting relevant information / analytics are difficult tasks
Agenda

- Semantic search system, KOIOS, for **intuitive access**, **analysis**, and **visualization** of structured environmental information
- Overview and architecture
- Structured query generation from keywords
- Facet-based browsing and refinement
- Selector initialization for final result and view construction
- Implementation and deployment
- Conclusions
KOIOS – Overview

- A semantic search system
  - Exploit semantics in the data for keywords interpretation to hide complexity of query languages and data representation
  - Keyword search for searching structured data
  - Lower access barriers while enabling richness of data to be fully harnessed

- Contribution
  - Transfer research results to commercial EIS
  - Selector mechanism

- Process
  - Input: keywords
  - Facet-based refinement
  - Selector (result and view template) initialization
  - Output: query results embedded in specific views
KOIOS – Architecture

Preprocessing

User

Selector Parametrization

SQL Query Generator

Database Query Processing

Result Table

Result Compilation

Reports, maps, charts etc.

Keyword query

Facet selection

Keyword-to-element mapping

Selector

Parametrized Selector

Facets

Create augmented schema graph

Sql Query

Generate query graphs

Database Query Processing

Result

Compilation

Top-k query processing

Preprocessing

Keyword Index

Data Index

Keyword Index

Data Index

Schema Index

based on

indexes

Database

indexes

KOIOS – Architecture
Query exploration

- Retrieve data elements for keywords from index
- Constructing **query space** from keyword elements + schema
- Top-k **query exploration**
  - Subgraphs connecting keyword elements:
  - Query interpretations: keyword elements + schema information

“Karlsruhe Co Emission”
Query translation and processing

- Query interpretations mapped to a query graph
- Query graph written as queries in a particular language (SPARQL, SQL)
- Retrieve answers for automatically chosen top-k queries (alternative: manual selection)

“Karlsruhe Co Emission”

Select ?s, ?v, ?d, ?l
WHERE {
  ?s ns:description "CO Emission" .
  ?s ns:hasValue ?v .
  ?v ns:hasDimension ?d .
  ?v ns:hasLocation ?l .
  ?l ns:longname "Karlsruhe"
}
Facets generation

- Derive facets from query results (not from query!) for refinement
  - Attributes serve as **facet categories**
  - Attribute values as **facet values**
- E.g. for ?s
  - Value.year: 2005, 2006,…

(a) Query Graph
Selectors

- Selector: parameterized, predefined result and view templates
  - Data parameters: specify scope of information need, initialized to a particular values based on facet categories and values
  - Query parameter: additional data processing for analysis tasks (GROUP-BY, SORT, MIN, MAX, AVERAGE etc.)
  - Presentation parameter: visualization types (data value, data series, data table, map-based, specific diagram type, etc.)

(a) Facet-Selector Mapping
Selector initialization

- Selectors capture templates for information needs and presentation of their results
- Map facets to selectors and initialize them
  - Applicable selectors: cover facet categories
  - Initialize selectors based on facet values
  - Initialized values are captured in the WHERE clause
  - **Non-initialized parameters** are included in the SELECT clause

```
Select v.value, s.id, s.partner, l.shortname
FROM Statistics s, Value v, Location l
WHERE v.year='2005'
and s.description='CO-Emission, PKW'
and l.longname='Karlsruhe'
```
Implementation

- Open-source indexes implemented on top of Lucene
  - Keyword lookup
  - Schema retrieval
  - Top-k query graphs processing (graph pattern matching)
- Open-source semantic search modules implemented in Java
  - Query translation
  - Facets generation
  - Selector initialization
- Integration with commercial EIS (Cadenza)
  - Management of selectors
  - Retrieval of final results
  - Result presentation and visualization
Deployment

- Hippolytos project (Theseus)
  - Easy access to spatial data warehouse (disy Cadenza) built for domain of environmental administration

- Data about
  - Emission and waste
  - From the Baden-Württemberg
  - Provided by:
Facets and selectors

Thanh Tran: KOIS
ISWC 2011, Bonn, Germany
Chart-based visualization

Thanh Tran: KOIS

ISWC 2011, Bonn, Germany
Map-based visualization

Thanh Tran: KOIS
ISWC 2011, Bonn, Germany
Conclusions

- Replace predefined forms and hard-coded visualization
- Semantic search using lightweight semantics in data and schema to dynamically
  - Translate keywords to queries
  - Generate facets for results
  - Initialize result and presentation templates
- Enables intuitive
  - Access
  - Visualization
  - Analysis of environmental information!
Thanks!

Tran Duc Thanh
duc thanh.tran@kit.edu
http://sites.google.com/site/kimducthanh/