Semi-automatic Generation of Active Ontologies from Web Forms

Martin Blersch, Mathias Landhäußer, and Thomas Mayer
How can we automatically add new features to intelligent assistants?
EASIER: A Framework to Connect Intelligent Assistants with Arbitrary Web Forms

- Generates 65% of the software automatically
- Correctly answers 70% of the queries with the generated software
Active Ontologies / Active Semantic Network

- Combine the modelling of domain knowledge with an execution environment
- Different node types
  - Leaf nodes
  - Non-terminal nodes
- Bottom-up processing of utterances

- Supporting new features requires extending the ontologies or even building new ones

→ Manual & labor-intensive steps

"I need a hotel room in Gothenburg from Monday to Tuesday."
"I need a **room** in **Gothenburg** for **two days**."
Related Work

Active: A Unified Platform for Building Intelligent Applications (Guzzoni)

FEATON: Builds Ontologies from Forms semi-automatically (Berlanga et al.)

OntoBuilder: Extracts Ontologies from Web Forms (Gal & Roitman et al.)

Cupid: Generic Schema Matching (Madhavan et al.)

A Clustering based Approach for Interface Matching (Wu et al.)

WISE: Automatic Integrator for Web Search Interfaces (He et al.)
Automatically Creating Active Ontologies: Overview

Form A
- **Origin**: 
- **Destination**: 
- **Departure**: 
- **Food**: Regular, Vegan, Halal

Form B
- **From**: 
- **To**: 
- **Departure**: 
- **WiFi**: ✗
Automatically Creating Active Ontologies: Grouping Related Form Elements

- Grouping similar form elements
  - Uses linguistic and structural similarities
  - Hierarchical Clustering of similar form elements

- Generation of meta form elements (global objects)
  - Merge form elements
Automatically Creating Active Ontologies: Deriving the Ontology

- One Active Ontology for each meta form (i.e., one per category)
  - Value range unclear → ask developer
  - Select AO node type
Evaluation

- Three evaluation questions
  - How good is the field matching?
  - What is the degree of automation?
  - Are the generated AOs capable to answer user queries?

- Data Source
  - 58 web forms from the UIUC Web Integration Repository
  - Three categories: airfare, automobile, and book search

- Queries
  - 61 queries for the airfare domain
  - 40 subjects (20 of them are native English speaker)
Evaluation: Field Matching

<table>
<thead>
<tr>
<th>Category</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_{0.5}$ Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airfare</td>
<td>90.6</td>
<td>21.0</td>
<td>54.4</td>
</tr>
<tr>
<td>Automobile</td>
<td>90.6</td>
<td>37.3</td>
<td>70.5</td>
</tr>
<tr>
<td>Book</td>
<td>98.4</td>
<td>46.4</td>
<td>80.4</td>
</tr>
</tbody>
</table>

- Highest precision for the book search domain
- Recall of clustering must be improved
## Evaluation: Degree of Automation

<table>
<thead>
<tr>
<th>Category</th>
<th>Total</th>
<th>Manual</th>
<th>Autom.</th>
<th>Autom. [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airfare</td>
<td>126</td>
<td>29</td>
<td>97</td>
<td>77%</td>
</tr>
<tr>
<td>Automobile</td>
<td>41</td>
<td>23</td>
<td>18</td>
<td>44%</td>
</tr>
<tr>
<td>Book</td>
<td>49</td>
<td>24</td>
<td>25</td>
<td>51%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>216</td>
<td>76</td>
<td>140</td>
<td><strong>65%</strong></td>
</tr>
</tbody>
</table>

- Automatically generated 77% of the elements needed for the airfare domain
- Needed the help of developers in only 35% of the cases
Evaluation: Query Answering

- How many query elements were correctly identified by the sensor nodes?
  - 61 queries from airfare domain
    - E.g. "Book a flight from Frankfurt to Paris."
    - Post-processing of given values
    - Expected: origin, destination, departure
    - Ask user for missing information

- Results
  - Recall: 75%
  - Queries Completely recognized
    - Only mandatory information: 77.4%
    - Mandatory and optional information: 35.5%
Evaluation: Query Answering (2)

<table>
<thead>
<tr>
<th>Field</th>
<th>Correct</th>
<th>Wrong</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin</td>
<td>191</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Destination</td>
<td>185</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Departure</td>
<td>124</td>
<td>24</td>
<td>52</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>48</td>
<td>52</td>
</tr>
</tbody>
</table>

- How many queries were correctly identified by the sensor nodes?
  - 26 complete queries (out of 61 queries)
  - 10 analyzed
  - 20 web forms (airfare domain)

- Results
  - 7 queries were recognized correctly
  - 2 were not recognized (EASIER asked the user)
  - 1 was recognized incorrectly
Conclusion and Future Work

- EASIER automates the process of building AOs
  - Automatically generates 65% of the AO's sensor nodes
  - High precision in field matching (90.6 - 98.4%)

- Queries
  - Correctly answers 70% of the queries
  - Asks for missing information

- Future Work
  - Improve field matching performance (HTML5, ARIA, …)
  - Integrate complex field mappings
  - Better domain knowledge (Wikipedia, Cyc)
References


References


BACKUP
The EASIER Active Server Architecture

EASIER Active Server

Evaluation Engine

Fact Store

Active Ontologies

Cat. 1

Cat. 2

DM

Broker

Dialog Interface

Communication Interface

Service 1

Service 2

Service n

API Call

Data Flow

Inspired by [Guz08]