Extracting Data from the Deep Web with Global-as-View Mediators Using Rule-Enriched Semantic Annotations

Benjamin Dönz
doenz[at]ict.tuwien.ac.at
Vienna University of Technology
Institute of Computer Technology
Vienna, Austria

Harold Boley
harold.boley[at]unb.ca
University of New Brunswick
Faculty of Computer Science
Fredericton, NB, Canada
The „Deep Web“ – What is it?

- Data hidden behind search forms and interfaces
- Estimated 400-500 times more information than the indexable World Wide Web
- 77% of the content classified as structured information
- Template based – so understanding how to extract one result allows to extract them all

Examples:
- Web shops
- Classified advertising
- Miscellaneous databases
Accessing the Deep Web

- **Data access manipulation**
- **Query forwarding**
- **Direct database access**
- **Access via web form**
- **Source conversion**
- **Information extraction**

- **Extraction on demand**
- **Extraction beforehand**
Our Approach: Upper-Right Quadrant

- Processing of queries using a query forwarding approach
  - SPARQL queries as input
  - Query transformation and forwarding via mediators
  - Global-as-View mapping of local sources

- Web form interaction and information extraction
  - Extraction process based on an extensible model
  - Semantic annotations for mapping real-world Web pages to the model
  - Feature-based rules for creating annotations
Model Overview

- Query interface for submitting conjunctive queries
Model Overview

- Query interface for submitting conjunctive queries
- Result list: all valid records, but only key attribute/value pairs
Model Overview

- Query interface for submitting conjunctive queries
- Result list: all valid records, but only key attribute/value pairs
- Result detail: all attribute/value pairs, but only one record
Query Process

1. Submit Query
2. Parse & Transform
3. Unfold
4. Extract
5. Integrate Intermediate Data
6. Evaluate Query

- Extract
  - Load Query Interface (URL)
  - Fill-Out and Submit Query
  - Extract from Result List
  - Extract from Detail Pages
  - Has more Results?
Walkthrough (1)

- **Query:**
  
  "Return details of real estate offers with a rent between 800€ and 1200€ and either at least 3 rooms or 80m²"

- **SPARQL:**

  ```sparql
  WHERE {?object rdf:type realestate:RealEstateOffer;
    realestate:townname ?townname;
    realestate:offername ?offername;
    realestate:description ?description;
    realestate:rent ?realestaterent;
    realestate:rooms ?realestaterooms;
    realestate:floorSpace ?realestatefloorSpace}.
  FILTER (?realestaterent>=800 && ?realestaterent<=1200 && (?realestaterooms>=3 || ?realestatefloorSpace >= 80))
  ```
Walkthrough (2)

- **Transformation:**
  - Parse Query
  - Transform filter to Disjunctive Normal Form and split into subqueries
  - Unfold to include relevant sources via Global-as-View mappings
Walkthrough (2)

- **Transformation:**
  - Parse Query
  - Transform filter to Disjunctive Normal Form and split into subqueries
  - Unfold to include relevant sources via Global-as-View mappings

- **Result**

```sql
SELECT FROM <http://derStandard.at>: greaterOrEqual(realestate:rent,800) ∧
lessOrEqual(realestate:rent,1200) ∧ greaterOrEqual(realestate:rooms,3)
UNION
SELECT FROM <http://at.immolive24.com>: greaterOrEqual(realestate:rent,800) ∧
lessOrEqual(realestate:rent,1200) ∧ greaterOrEqual(realestate:rooms,3)
UNION
SELECT FROM <http://derStandard.at>: greaterOrEqual(realestate:rent,800) ∧
lessOrEqual(realestate:rent,1200) ∧ greaterOrEqual(realestate:floorSpace,100)
UNION
SELECT FROM <http://at.immolive24.com>: greaterOrEqual(realestate:rent,800) ∧
lessOrEqual(realestate:rent,1200) ∧ greaterOrEqual(realestate:floorSpace,100)
```
Walkthrough (3)

- Features used for identification elements
  - “Properties” of the HTML tags
  - For example id, class, value, tag path, associated text, ...

```
Query Interface
Rent >  <
Floorspace > <
Rooms >  <
Search
```

```
Features
Element_body_form_table_tr0_td1
hasID: “rent_from”
prevText: „Rent“

Element_body_form_table_tr2_td2
hasID: „flspace_to“
prevText: „<“

Element_body_form_table_tr4_td1
hasValue: „Search“
```
Walkthrough (4)

- Object-centered Datalog rules
  - Conditions: conjunction of req. features
  - Conclusions: concepts of the model
- Single efficient evaluation pass

Features

- Element_body_form_table_tr0_td1
  - hasID: "Rent_from"
  - prevText: "Rent"

- Element_body_form_table_tr2_td2
  - hasID: "flspace_to"
  - prevText: "<"

- Element_body_form_table_tr4_td1
  - hasValue: "Search"

Annotation Rules

If Element.prevText("Rent") then Element.restriction=greaterThan(realestate:rent)
If Element.hasID("flspace_to") then Element.restriction=lessThan(realestate:floorSpace)
If Element.hasValue("Search") then Element.function=submitQueryWith("click")
Walkthrough (5)

Features

- Element_body_form_table_tr0_td1
  - hasID: "rent_from"
  - prevText: "Rent"

- Element_body_form_table_tr2_td2
  - hasID: "flspace_to"
  - prevText: "<"

- Element_body_form_table_tr4_td1
  - hasValue: "Search"

Annotation Rules

If Element.prevText("Rent") then Element.restriction=greaterThan(realestate:rent)
If Element.hasID("flspace_to") then Element.restriction=lessThan(realestate:floorSpace)
If Element.hasValue("Search") then Element.function=submitQueryWith("click")

Annotations

- Element_body_form_table_tr0_td1
  - greaterThan(realestate:rent)

- Element_body_form_table_tr2_td2
  - lessThan(realestate:floorSpace)

- Element_body_form_table_tr4_td1
  - submitQueryWith("click")
Walkthrough (6)
Walkthrough (7)

- **SPARQL**

```
WHERE { ?object rdf:type realestate:RealEstateOffer;
    realestate:townname ?townname;
    realestate:offername ?offername;
    realestate:description ?description;
    realestate:rent ?realestaterent;
    realestate:rooms ?realestaterooms;
    realestate:floorSpace ?realestatefloorSpace}.
FILTER (?realestaterent>=800 && ?realestaterent<=1200 && (?realestaterooms>=3 || ?realestatefloorSpace >= 80))
```

- **Output**

<table>
<thead>
<tr>
<th>realstate* townname</th>
<th>realstate* offername</th>
<th>realstate* description</th>
<th>realstate* rent</th>
<th>realstate* rooms</th>
<th>realstate* floorSpace</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nahe Burgasse</td>
<td>Die City zum Greifen nahe</td>
<td>Die City zum Greifen nahe Das Museumsquartier ums Eck, der 1.Bez...</td>
<td>999,29</td>
<td>3</td>
<td>87</td>
</tr>
<tr>
<td>Godawagasse</td>
<td>3 Zimmer Familienwohnung mit terrasse</td>
<td>3 Zimmer Familienwohnung mit Terrasse Diese durchdachte Wohnung...</td>
<td>1109,75</td>
<td>3</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>ruhige Altbaubauung, WG-ta...</td>
<td>ruhige Altbaubauung, WG-ta... Sehr s...</td>
<td>1088</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>ERSTEZUG: Schicke Neubauung...</td>
<td>ERSTEZUG: Schicke Neubauung mit Garten zur Vermietung...</td>
<td>1150</td>
<td>3</td>
<td>84</td>
</tr>
<tr>
<td>Lugeck</td>
<td>Einzigartige Designwohnung...</td>
<td>Einzigartige Designwohnung mit Autoabstellplatz! Einzigartige Dise...</td>
<td>1184,24</td>
<td>3</td>
<td>107</td>
</tr>
<tr>
<td>Wertheimstein-Park</td>
<td>Absolute TOP-Lage - hauseigene Parkanlage! Sanierte 3-Zimmer-...</td>
<td>Absolute TOP-Lage - hauseigene Parkanlage! Sanierte 3-Zimmer-...</td>
<td>990</td>
<td>3</td>
<td>59</td>
</tr>
<tr>
<td>Muringengasse</td>
<td>Exklusiver Erstbezug im Herzen...</td>
<td>Exklusiver Erstbezug im Herzen des 12. Bezirks Exklusiver Erstbe...</td>
<td>984,16</td>
<td>3</td>
<td>77</td>
</tr>
<tr>
<td>Murinaengasse</td>
<td>Exklusives Wohnen im Herzen...</td>
<td>Exklusives Wohnen im Herzen des 12. Bezirks Exklusives Wohnen...</td>
<td>945</td>
<td>3</td>
<td>66</td>
</tr>
</tbody>
</table>
Live Presentation of Example Use Cases

**Deep Web Mediator** and examples available online: http://semann.bdoenz.com/default.aspx

#1 Plain list: Extract the average rent per town from a single site.

#2 Search and result list: Extract test results on cars of the brand Audi from a single site and return brand, model and the test conclusion.

#3 Search, list and detail page: Extract real estate offers from a single site and return details for offers with 3 or more rooms and a rent of 800€ to 1200€.

#4 Disjunctive query: Extract used car offers from a single site and return details of all offers for cars of the brand “Audi” that are priced under 12.500€ if the construction year is after 2011 or under 15.000€ if the construction year is after 2012.

#5 Union: Extract used car offers from all available sites and return details of offers for cars of the brand “Audi” that are priced under 12.500€ and have a construction year after 2011.

#6 Disjunctive union: Extract used car offers from all available sites and return details of all offers for cars of the brand “Audi” that are priced under 12.500€ if the construction year is after 2011 or under 15.000€ if the construction year is after 2012.

#7 Relations between sources: Extract average rents and real estate offers from all available sites and return those that are located in a specific town and have a lower rent/m² than the average for that town.

#8 Deep Web and local databases: Extract real estate offers from all available sites and add the type of town and population from a local dataset.

#9 Deep Web and external databases: Extract real estate offers from all available sites and add a description of the town and the population from an external SPARQL endpoint (dbPedia).
Conclusion

- Processing of queries using a query forwarding approach
  - SPARQL queries as input
  - Query transformation and forwarding via mediators
  - Global-as-View mapping of local sources

- Web form interaction and information extraction
  - Extraction process based on an extensible model
  - Semantic annotations for mapping real-world Web pages to the model
  - Feature-based rules for creating annotations
Extracting Data from the Deep Web with Global-as-View Mediators Using Rule-Enriched Semantic Annotations

Benjamin Dönz
Doenz[at]ict.tuwien.ac.at
Vienna University of Technology
Institute of Computer Technology
Vienna, Austria

Harold Boley
harold.Boley[at]unb.ca
University of New Brunswick
Faculty of Computer Science
Fredericton, NB, Canada
Use case #3

This example is situated in the domain of real estate, and asks for the name of the offer, a description, the number of rooms, the floor space and the rent of all offers with 3 or more rooms and a rent in the range of 800€ and 1200€ from a specific real estate site. To process this query, the mediator accesses the query interface of the site, sets the parameters in the fields of the web form and triggers the search function to submit the query. The returned result lists are iterated to extract the values from the list itself, but also from subpages by following the the corresponding link for each record. All extracted facts are collected in a database and presented to the user in a tabular style including a link to the page where the offer was found.

```
FROM <http://derstandard.at/anzeiger/immoweb/Immobilien-suche.aspx>
  OPTIONAL {?object <http://semannot.bdoenz.com/mediatorvocabulary#sourceURL> ?source}.
  OPTIONAL {?object <http://semannot.bdoenz.com/realestate#townname> ?realestatetownname}.
  OPTIONAL {?object <http://semannot.bdoenz.com/realestate#offername> ?realestateoffername}.
  OPTIONAL {?object <http://semannot.bdoenz.com/realestate#description> ?realestatedescription}.
  OPTIONAL {?object <http://semannot.bdoenz.com/realestate#rent> ?realestaterent}.
  OPTIONAL {?object <http://semannot.bdoenz.com/realestate#rooms> ?realestaterooms}.
  OPTIONAL {?object <http://semannot.bdoenz.com/realestate#floorSpace> ?realestatefloorSpace}.
    FILTER (?realestaterent>=800 &&
      ?realestaterent<=1200 &&
      ?realestaterooms>=3
    )
}
```
Use case #3

Deep Web Mediator: SPARQL Interface

Type

Look for
realestate#RealEstateOffer

Properties
realestate#townname
realestate#offermane
realestate#description
realestate#rent
realestate#rooms
realestate#floorSpace

from 3 to
from 900 to 1200

253 Results in 232.371s

<table>
<thead>
<tr>
<th>source</th>
<th>realestate#townname</th>
<th>realestate#offermane</th>
<th>realestate#description</th>
<th>realestate#rent</th>
<th>realestate#rooms</th>
<th>realestate#floorSpace</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Unbefristete 3 Zimmer Stils...</td>
<td>Unbefristete 3 Zimmer Stils...</td>
<td>1150</td>
<td>3</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Traumhafter ERSTBEZUG m...</td>
<td>Traumhafter ERSTBEZUG m...</td>
<td>1082,66</td>
<td>3</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>TERRASSENTRAUM nacht...</td>
<td>TERRASSENTRAUM nacht...</td>
<td>999,97</td>
<td>3</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>WG-taugliche Mietwohnung...</td>
<td>WG-taugliche Mietwohnung...</td>
<td>1094</td>
<td>3</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Jungfamilien in 1030 W...</td>
<td>Jungfamilien in 1030 W...</td>
<td>108,90</td>
<td>3</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Anschuitzgasse ERSTBEZUG NACH GENERAL...</td>
<td>ERSTBEZUG NACH GENERAL...</td>
<td>836,76</td>
<td>3</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Nussdorfer Platz Teiloebelte 3 Zimmer Bal...</td>
<td>Teiloebelte 3 Zimmer Bal...</td>
<td>950</td>
<td>3</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Die PERFEKTE 3ar-WG mit...</td>
<td>Die PERFEKTE 3ar-WG mit...</td>
<td>1100</td>
<td>3</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Ambogasse TRAUM DG-MAISONETTE...</td>
<td>TRAUM DG-MAISONETTE...</td>
<td>979</td>
<td>3</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Tölle Raumaufteilung- Kom...</td>
<td>Tölle Raumaufteilung- Kom...</td>
<td>990</td>
<td>3</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Heiligenstaedterstrasse 19. Heiligenstaedter Strass...</td>
<td>19. Heiligenstaedter Strass...</td>
<td>1000</td>
<td>3</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Leudingasse B. Bez...</td>
<td>Echte Stilaltbauwohn...</td>
<td>1130</td>
<td>3</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Schoene, helle Altbaubauwohn...</td>
<td>Schoene, helle Altbaubauwohn...</td>
<td>1006,3</td>
<td>3</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Terrassentrain in Moedling...</td>
<td>Terrassentrain in Moedling...</td>
<td>1174,41</td>
<td>3</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Stieggasse (Marishl...</td>
<td>Ruheig, kldurchflutete Al...</td>
<td>879,22</td>
<td>3</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Zwerggasse Freundliche 4-Zimmer-Woh...</td>
<td>Freundliche 4-Zimmer-Woh...</td>
<td>1011,47</td>
<td>4</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Wolfisplatz VIDEO: HELLE WOHNUNG...</td>
<td>VIDEO: HELLE WOHNUNG...</td>
<td>1054,9</td>
<td>4</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Schloess Belvedere Liebevolle sanit...</td>
<td>Liebevolle sanit...</td>
<td>1124,84</td>
<td>4</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Mosbaurere, gut aufgestellte...</td>
<td>Mosbaurere, gut aufgestellte...</td>
<td>898,34</td>
<td>3</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Hemeler Hauptstrasse... und her haetten sie auch...</td>
<td>und her haetten sie auch...</td>
<td>1102,09</td>
<td>3</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td><a href="http://de">http://de</a>...</td>
<td>Giessachbursasse...</td>
<td>...</td>
<td>1064,85</td>
<td>3</td>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>
Use case #6

This example is situated in the domain of used cars and is the combination of the previous two examples: The query requests the name, model, color, construction year, mileage and price of offers from a single site, where the brand of the car is “Audi” and that are priced under 12.500€ if the construction year is after 2010, or under 15.000€ if the construction year is after 2011. This query is split into two conjunctive subqueries and submit to all three available sites returning the union of a total of 6 queries.

```
OPTIONAL {?object <http://semannot.bdoenz.com/mediatorterminology#sourceURL> ?source}.
OPTIONAL {?object <http://semannot.bdoenz.com/cars#brand> ?carsbrand}.
?object <http://semannot.bdoenz.com/cars#brand> "Audi".
OPTIONAL {?object <http://semannot.bdoenz.com/cars#model> ?carsmodel}.
OPTIONAL {?object <http://semannot.bdoenz.com/cars#offername> ?carsoffername}.
OPTIONAL {?object <http://semannot.bdoenz.com/cars#color> ?carscolor}.
OPTIONAL {?object <http://semannot.bdoenz.com/cars#mileage> ?carmileage}.
OPTIONAL {?object <http://semannot.bdoenz.com/cars#constructionYear> ?carsconstructionYear}.
OPTIONAL {?object <http://semannot.bdoenz.com/cars#offerprice> ?carsofferprice}.
FILTER (
(?carsconstructionYear>=2010 && ?carsofferprice<=12500)
||
(?carsconstructionYear>=2011 && ?carsofferprice<=15000)
)
}
Use case #6

Deep Web Mediator: SPARQL Interface

30 Results in 77.092s

<table>
<thead>
<tr>
<th>source</th>
<th>carbrand</th>
<th>carmodel</th>
<th>carsoffename</th>
<th>carcolor</th>
<th>carmileage</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A1 1.2 TFSI Attraction</td>
<td>schwarz</td>
<td>67000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>231000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>rot</td>
<td>102954</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 2.0 TFSI &quot;1-Basis&quot; Go...</td>
<td>160300</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A3 2.0 Ambiente 2.0 TDI DPF</td>
<td>159877</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A3 2.0 TFSI 1-Basis, Go...</td>
<td>schwarz</td>
<td>155000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A3 2.0 TFSI 1-Basis, Go...</td>
<td>XENON</td>
<td>103000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A3 2.0 TFSI 1-Basis, Go...</td>
<td>rot</td>
<td>201000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A3 2.0 TFSI 1-Basis, Go...</td>
<td>schwarz</td>
<td>37324</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A3 2.0 TFSI 1-Basis, Go...</td>
<td>silber</td>
<td>123000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 2.0 TDI DPF X6...</td>
<td>222000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>schwarz</td>
<td>186000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>grau</td>
<td>334255</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>silber</td>
<td>234441</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>schwarz</td>
<td>3100</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>weiss</td>
<td>98000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>silber</td>
<td>19671</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>schwarz</td>
<td>124000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>schwarz</td>
<td>100000</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.car4you.at/Gn-">http://www.car4you.at/Gn-</a></td>
<td>Audi</td>
<td>A4 Avant 2.0 TDI DPF X6...</td>
<td>grau</td>
<td>159469</td>
<td></td>
</tr>
</tbody>
</table>

Institute of Computer Technology

TU Wien
Use case #7

This example is situated in the domain of real estate, and asks for the name of the offer, a description, the number of rooms, the floor space and the rent of all offers with 3 or more rooms in the town of "Klosterneuburg" where the rent is lower than the average rent per square meter for that town. No specific site is referenced in the query, the mediator therefore includes all sites with real estate offers and also a site containing average rents. Each of these are accessed in turn collecting the intermediate results in a database before applying the filter and returning the results. Intermediate results are only available after the average rents have been extracted and can be compared to the offers in the defined manner. Note that this type of query cannot be generated by the query wizard, but is entered directly as SPARQL

```
OPTIONAL {?object <http://semannot.bdoenz.com/mediatorvocabulary#sourceURL> ?source}.
OPTIONAL {?object <http://semannot.bdoenz.com/realestate#townname> ?realestatetownname}.
OPTIONAL {?object <http://semannot.bdoenz.com/realestate#offername> ?realestateoffername}.
OPTIONAL {?object <http://semannot.bdoenz.com/realestate#rent> ?realestaterent}.
OPTIONAL {?object <http://semannot.bdoenz.com/realestate#rooms> ?realestaterooms}.
OPTIONAL {?object <http://semannot.bdoenz.com/realestate#floorSpace> ?realestatefloorSpace}.
FILTER (REGEX(?realestatetownname,"Klosterneuburg","i") &&
?realestaterent>=800 && ?realestaterent<=1200 &&
?realestaterooms>=3 && REGEX(?cname,"Klosterneuburg","i") &&
?realestaterent/?realestatefloorSpace <(?avrent))}
```
Use case #7

Deep Web Mediator: SPARQL Interface

Semantic Access to the Deep Web

Institute of Computer Technology