UiO : Universitetet i Oslo

DELPH-IN 2014 – Tomar Portugal

RDF Triple Stores and a Custom SPARQL Front-End for Searching Meaning Representations

Milen Kouylekov and Stephan Oepen

What are we working with

- One of the main areas of research in our group is ‘Semantic Parsing’
- Inevitable Problem
  - Storing & Accessing Data
- Corpora
  - DeepBank (700,000 tokens)
  - WikiWoods (48 million tokens)

Semantic Annotation

- Minimal Recursion Semantics (MRS; Copestake et al., 2005)
  - underspecified logical-form meaning representation.
- In addition, we support the two reductions:
  - Elementary Dependency Structures (EDS)
    - a variable-free dependency graph
  - MRS-Derived Bi-Lexical Dependencies (DM)
    - ‘projecting’ predicate–argument information from the EDS onto individual tokens

Example - MRS

<h1,

h4:_a_q(**x6**, h7, h5), h8:_similar_a_to(e9, **x6**), h8:comp(e11, e9,_),

h8:_technique_n_1(**x6**),h2:_almost_a_1(e12, h13),

h14:_impossible_a_for(e3, h15, i16), h17 :_apply_v_to(e18, i19, **x6**, **x20**),

h21:udef_q(**x20**, h22, h23), h24:_other_a_1(e25, x20), h24 :_crop_n_1(**x20**),

h24 :_such+as_p(x26, **x20**, x27), x40 :implicit_conj(x27, x33, x38),

x31 :udef_q(x33, x32, x34), x35 :_cotton_n_1(x33), x46 :_and_c(x38, x43, x47),

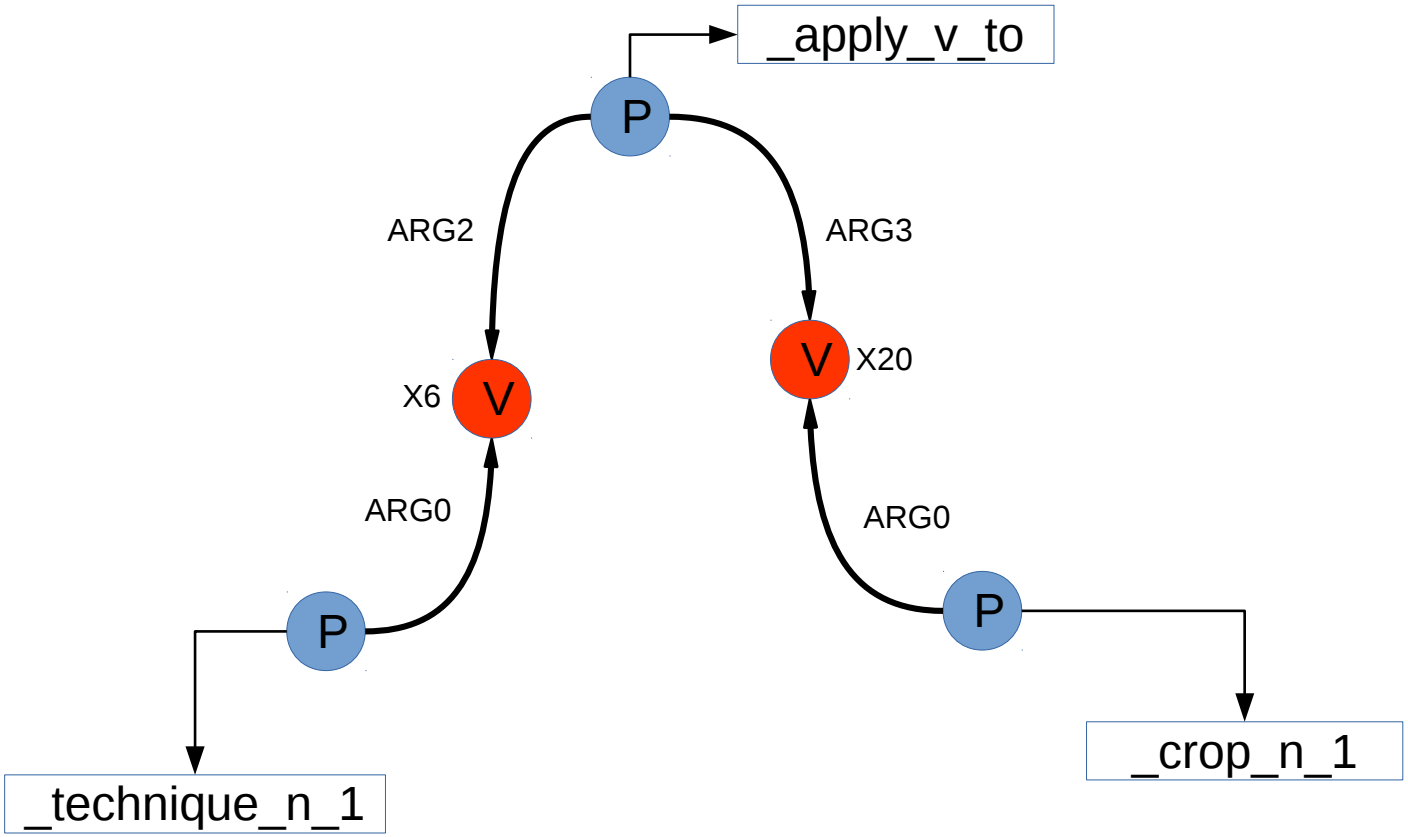
x41 :udef_q(x43, x42, x44), x45 :_soybean_u_unknown(x43),

x48 :udef_q(x47, x49, x50), x51 :_rice_n_1(x47)

{ x49 = q x51, x42 = q x45, x32 = q x35, x22 = q x24, x15 = q x17, x13 = q x14, x7 = q x8, x1 = q x2 }
>

A similar technique is almost impossible to apply to other crops, such as cotton, soybeans and rice.

Example - MRS


_apply_v_to
P
ARG2
ARG3
V
X6
V
X20
ARG0
ARG0
P
P
_technique_n_1
_crop_n_1

Example - EDS

e9:_similar_a_to<2,9>[ARG1 **x6**] **x6**:_technique_n_1<10,19>[]

e11:comp<2,9>[ARG1 e9] e12:_almost_a_1<23,29>[ARG1 e3]

e3:_impossible_a_for<30,40>[ARG1 e18] e18:_apply_v_to<44,49>[ARG2 **x6**, ARG3 **x20**]

x20:_crop_n_1<59,65>[] _2:udef_q<53,100>[BV **x20**]

e25:_other_a_1<53,58>[ARG1 x20] e26:_such+as_p<66,73>[ARG1 **x20**, ARG2 x27]

x27:implicit_conj<82,100>[L-INDEX x33, R-INDEX x38] _3:udef_q<74,100>[BV x27]

_4:udef_q<74,81>[BV x33] x33:_cotton_n_1<74,81>[]

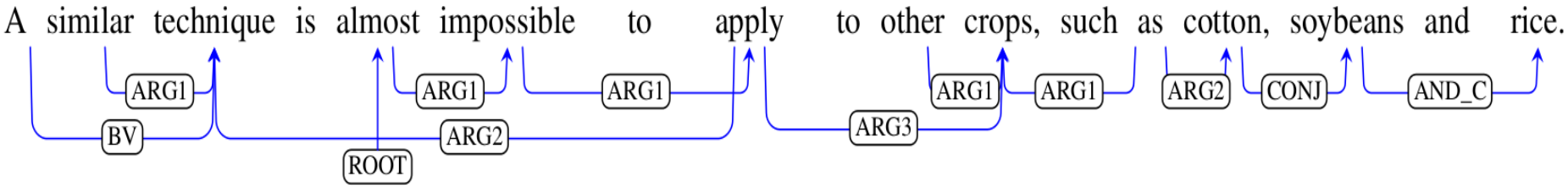
_5:udef_q<82,100>[BV x38] x38:_and_c<91,94>[L-INDEX x43, R-INDEX x47]

_6:udef_q<82,90>[BV x43] x43:_soybeans/nns_u_unknown<82,90>[]

x47:_rice_n_1<95,100>[] _7:udef_q<95,100>[BV x47]

A similar technique is almost impossible to apply to other crops, such as cotton, soybeans and rice.

Example - DM


A similar technique is almost impossible to apply to other crops, such as cotton, soybeans and rice.
ARG1
BV
ROOT
ARG2
ARG1
ARG1
ARG3
ARG1
ARG1
ARG2
CONJ
AND_C

Resource Description Framework (RDF)

- Used by the Semantic Web community
- RDF Data Model
  - statements about resources in the form of subject–predicate–object triples.
- Intrinsically represents a labeled, directed multi-graph
- Suited for knowledge representation

Search SPARQL

- Search Language for querying RDF structures
- Similar to SQL
- The query is satisfied by variable instantiation

Repositories of RDF

- Triple Stores
  - Many implementations available (free and commercial)
  - All implement SPARQL (some provide also different dialects)
  - Most are easily connected to web applications
  - Most of them provide reasoning based on description logic (OWL Ontologies)

MRS in OWL

```
@prefix : <http://wesearch.delph-in.net/rdf/mrs#> .

:Predication rdf:type owl:Class .

:Variable rdf:type owl:Class .
:I rdf:type owl:Class ;
  rdfs:subClassOf :Variable .
:E rdf:type owl:Class ;
  rdfs:subClassOf :I .

:ROLE rdf:type owl:ObjectProperty ;
  rdfs:domain :Predication ;
  rdfs:range :Variable .

:ARG0 rdf:type owl:ObjectProperty ;
  rdfs:subPropertyOf :ROLE ;
  rdfs:range :I .

:constraint rdf:type owl:ObjectProperty ;
  rdfs:range :H ;
  rdfs:domain :H .

:qeq rdf:type owl:ObjectProperty ;
  rdfs:subPropertyOf :constraint .
```

MRS in OWL

<p1> mrs:predicate “_appy_v_to”^^xsd:string
<p1> rdf:type mrs:Predication

<p1> mrs:ARG2 <**x6**>
<**x6**> rdf:type mrs:X

<p1> mrs:ARG3 <**x20**>
<**x20**> rdf:type mrs:X

<p2> mrs:predicate “_technique_n_1” ^^xsd:string
<p2> rdf:type mrs:Predication
<p2> mrs:ARG0 <**x6**>

<p3> mrs:predicate “_crop_n_1” ^^xsd:string
<p3> rdf:type mrs:Predication
<p3> mrs:ARG0 <**x20**>

EDS Search in SPARQL Example

```
PREFIX mrs:<http://www.delph-in.net/rdf/mrs#>
SELECT ?graph WHERE {
    GRAPH ?graph {
        ?p1 mrs:predicate "_apply_v_to"^^xsd:string .
        ?p2 mrs:predicate "_crop_n_1"^^xsd:string .
        ?p1 rdf:type mrs:Predication .
        ?p2 rdf:type mrs:Predication .
        ?p1 mrs:ARG3 ?var1 .
        ?var1 rdf:type mrs:X .
        ?p2 mrs:ARG0 ?var1
    }
}
```

Feasibility – Apache JENA (TDB Store)

	DeepBank			WikiWoods
	MRS	EDS	DM	EDS
Triples	20 million	2.6 million	4 million	3.4 billion
Space	3.8 GB	494 MB	744 MB	400 GB
Indexing	4 minutes	1 minute	1 minute	24 hours
Query	8 sec	3 sec	3 sec	???

Started building a test set of 10 queries with different complexity.

Querying from User Perspective

- Quering triple stores requires SPARQL
  - Requires some knowledge of how the structures are stored
- Users will prefer to query with patterns:

  _apply_v_to[ARG3 x] , x:_crop_*
- WQL: Custom query processor for family of ‘designer’ pattern languages, adapted to each representation.
- Web application for interactive search (across different graph types) and custom visualization.

DeepBank & SEMEVAL 2014 -TASK 8

wesearch.delph-in.net/sdp/search.jsp

Semeval 2014 - Task8 Help

Semantic Search Interface - Graph

Query

```
[ARG2 x, ARG* e]
e:/v*[ARG1 x]
```

Format DM PAS PCEDT **Results per Page** 20 search

Show SPARQL

Results Page 1

#20004005 Longer maturities are thought to indicate declining interest rates because they permit portfolio managers to retain relatively higher rates for a longer period .

DM PAS PCEDT

Match #1 Match #2