MOLTO-Patents: recent issues, solutions and perspectives

Laura Tolo̧si, Maria Mateva
Ontotext AD, Bulgaria

September 12, 2012

In the previous version of the demo, the query interface presented several NL query examples, with the purpose of assisting the unexperienced user with formulating his or her first requests to the system. However, several of the examples did not return any supporting patents containing the information requested by the query. Even trying out many queries in a row may result in answers that are not supported by patents from the collection. In order to overcome such frustrating attempts, we took two approaches. First, we changed the example queries from the demo page in order to ensure that they do return some supporting patents. Second, in the deliverable we present a summary (roadmap) of the relations present in the ontology that are also supported by patents, such that users that want to test our system can have a comprehensive set of examples to start with. In what follows we give more details on each.

1 Changes of the demo interface

The new queries that are given as examples on the interface are:

'give me all information about AMPICILLIN' – 12 documents
'give me all information about all active ingredients of BACLOFEN' –21 documents
'give me all information about all routes of administration of FAMOTIDINE' – 24 documents
'give me all information about all dosage forms of GANCICLOVIR' –24 documents
'give me the approval date of the patent for REBETOL' – 6 documents

2 Roadmap

Below we present for each type of natural-language query a general table, containing entity names that can be used in order to obtain non-empty results.

2.1 Query type: 'Give me all information about *drug/active ingredient*'

Table 1 shows drugs that are involved in triples in the ontology and are also mentioned in patents. If the *drug* is chosen among the drugs in the table, then the query 'Give me all information about *drug*' is guaranteed to return supporting patents. Table 1 is abbreviated, for a complete table run the following general SPARQL query:

```
SELECT ?drug (count(distinct ?doc) as ?count)
WHERE {
   {?s ?p ?o .
    ?s <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasName> ?n .
```

```
    ?n <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
       <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#FDA_DrugName> .
    ?n <http://www.w3.org/2000/01/rdf-schema#label> ?drug .
    }
   UNION
   {?o ?p ?s .
    ?s <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasName> ?n .
    ?n <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
       <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#FDA_DrugName> .
    ?n <http://www.w3.org/2000/01/rdf-schema#label> ?drug .
    }
   ?doc <http://proton.semanticweb.org/protonm#mentions> ?n .
}
GROUP BY ?drug
ORDER BY desc(?count)
```

A similar query for obtaining a summary of the active ingredients mentioned in the collection of patents can be obtained by the following query:

```
SELECT ?ai (count(distinct ?doc) as ?count)
WHERE {
   {?s ?p ?n .
    ?n <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
       <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#ActiveIngredient> .
    ?n <http://proton.semanticweb.org/protonsys#mainLabel> ?ai .
    }
   UNION
   {?n ?p ?s .
    ?n <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
       <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#ActiveIngredient> .
    ?n <http://proton.semanticweb.org/protonsys#mainLabel> ?ai .
    }
   ?doc <http://proton.semanticweb.org/protonm#mentions> ?n .
}
GROUP BY ?ai
ORDER BY desc(?count)
```

2.2 Query type: 'Give me all information about the active ingredients of *drug*'

Table 3 shows a small fraction of the drugs for which active ingredients are known (present in the ontology via the relation hasActiveIngredient) and these ingredients are mentioned in patent documents. The full list can be obtained with following SPARQL query, via the SPARQL interface of the demo.

```
SELECT ?drug ?l (count(distinct ?doc) as ?count)
WHERE {
    ?o <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasActiveIngredient> ?d .
    ?o <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasName> ?s .
    ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
       <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#FDA_DrugName> .
    ?s <http://www.w3.org/2000/01/rdf-schema#label> ?drug .
    ?d <http://www.w3.org/2000/01/rdf-schema#label> ?l .
    ?doc <http://proton.semanticweb.org/protonm#mentions> ?d .
    ?doc <http://proton.semanticweb.org/protonm#mentions> ?s .
}
```

```
GROUP BY ?drug ?l
ORDER BY ?drug
```

All queries including drugs listed in Table 3 are guaranteed to return supporting documents.

2.3 Query type: 'Give me all information about all routes of administration of *drug*'

A table with drugs, routes of administrations and number of documents mentioning the drug and the route of administrations can be obtained via the query:

```
SELECT ?label ?l (count(distinct ?doc) as ?count)
WHERE {
  ?s <http://www.w3.org/2000/01/rdf-schema#label> ?label .
  ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
     <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#FDA_DrugName> .
  ?o <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasName> ?s.
  ?o <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasRouteOfAdministration> ?d.
  ?d <http://www.w3.org/2000/01/rdf-schema#label> ?l.

  ?doc <http://proton.semanticweb.org/protonm#mentions> ?d.
  ?doc <http://proton.semanticweb.org/protonm#mentions> ?s
}
GROUP BY ?label ?l
ORDER BY desc(?count)
```

Table 4 shows a part of the results returned by the query above. If the user chooses one of the drugs from the table for a query of the type 'Give me all information about all routes of administration of *drug*', then there will be documents returned.

2.4 Query type: 'give me all information about all dosage forms of *drug*'

A table of all drugs, together with their dosage forms and the number of documents that contain related in formation can be obtained by running the following general SPARQL query:

```
SELECT ?label ?l (count(distinct ?doc) as ?count)
WHERE {
  ?s <http://www.w3.org/2000/01/rdf-schema#label> ?label .
  ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
      <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#FDA_DrugName> .
  ?o <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasName> ?s.
  ?o <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasDosageForm> ?d.
  ?d <http://www.w3.org/2000/01/rdf-schema#label> ?l.

  ?doc <http://proton.semanticweb.org/protonm#mentions> ?d.
  ?doc <http://proton.semanticweb.org/protonm#mentions> ?s
}
GROUP BY ?label ?l
ORDER BY desc(?count)
```

3 Changes of the queries

In the earlier versions of the demo, the query type 'give me all information about all routes of administration of *drug*' was returning confusing results. Specifically, the query was interpreted as follows: 'find in the ontology the routes of

administration of *drug*, then search for documents mentioning these routes of administration (independently from the drug)'. As a result, the user obtains for example a long list of patents mentioning various drugs that are administered orally (if *drug* is administered orally). The meaning of the initial natural-language query is different though, the user searching for information on how the *drug* can be administered.

We corrected the interpretation of the query and correspondingly, the SPARQL translation, as follows. In the new query, documents that mention both the *drug* and the route of administration must be mentioned by the document. Below is the updated query:

```
CONSTRUCT {
  ?d <http://www.w3.org/2000/01/rdf-schema#label> ?l.
  ?doc <http://proton.semanticweb.org/protonm#mentions> ?d
}
WHERE {
  ?s <http://www.w3.org/2000/01/rdf-schema#label> "FAMOTIDINE" .
  ?s <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
     <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#FDA_DrugName> .
  ?o <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasName> ?s.
  ?o <http://www.semanticweb.org/ontologies/2008/7/Ontology1218740600570.owl#hasRouteOfAdministration> ?d.
  ?d <http://www.w3.org/2000/01/rdf-schema#label> ?l.
  OPTIONAL {
      ?doc <http://proton.semanticweb.org/protonm#mentions> ?d.
      ?doc <http://proton.semanticweb.org/protonm#mentions> ?s
  }
}
```

However, the co-occurence of the two entities in the document does not guarantee that the document describes precisely the drug and its route of administration. For example, if the targetted drug is **D** and its route of administration is **oral**, and **G** is another drug, then a document containing the following sentence will be returned by the query: 'Drug **G** is administered **orally**. If taken at the same time with drug **D**, drug **G** will lead to severe side effects.' This document will be returned by the query, although there is no mention of how drug **D** is administered.

In order to avoid such false results, more complex approaches are necessary, for example methods for automated extracting of relations between annotated entities in text. Such methods exist [1] and can be considered for future developments of the project.

References

[1] McDonald et al. (2005) Simple algorithms for complex relation extraction with applications to biomedical IE. *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (ACL-05).

drug	documents
ACETIC ACID	1050
SODIUM CHLORIDE	811
TALC	749
INSULIN	711
LENTE	497
PENICILLIN	414
SODIUM BICARBONATE	402
CISPLATIN	319
MAGNESIUM SULFATE	315
CYCLOPHOSPHAMIDE	297
ADENOSINE	294
FLUOROURACIL	288
DIMETHYL SULFOXIDE	277
AMMONIUM CHLORIDE	269
DEXAMETHASONE	268
ETOPOSIDE	260
STERILE WATER	259
PACLITAXEL	244
MITOMYCIN	218
CARBOPLATIN	208
TAXOL	208
CYCLOSPORINE	207
AMPICILLIN	203
PREDNISONE	194
IBUPROFEN	192
MERCAPTOPURINE	170
MITOXANTRONE	170
ESTRADIOL	170
CYTARABINE	159
INDOMETHACIN	155
FOLIC ACID	151
PREDNISOLONE	151
IFOSFAMIDE	141
LOVASTATIN	138
THIOGUANINE	135
VITAMIN D	130
POTASSIUM CHLORIDE	130
SIMVASTATIN	128
KANAMYCIN	128
HYDROXYUREA	127
NAPROXEN	126
TESTOSTERONE	125
FLUTAMIDE	121
LIDOCAINE	119
AZATHIOPRINE	118
THIOTEPA	117
VITAMIN A	115
DACARBAZINE	115
TAXOTERE	114
GLUCAGON	110
HYDROCORTISONE	107
PROGESTERONE	106
METHYLPREDNISOLONE	104
NICOTINE	98
BENZYL BENZOATE	96
CLADRIBINE	96
PENTOSTATIN	95
PIROXICAM	94
RIBAVIRIN	94
GENTAMICIN	92
KETOPROFEN	92
FLUOXETINE	92
ACETAMINOPHEN	91
NIFEDIPINE	91
TRIAMCINOLONE	91
SODIUM THIOSULFATE	87
CAPTOPRIL	85
CHLORAMPHENICOL	84
ISOFLURANE	84
...	...

Table 1: Drugs mentioned in patents.

active ingredient	documents
CALCIUM	1458
ALCOHOL	1421
AMINO ACIDS	1138
GLYCERIN	957
SODIUM CHLORIDE	811
MANNITOL	806
TALC	749
GLYCINE	716
CITRIC ACID	663
SORBITOL	655
SULFUR	631
TYROSINE	595
PROTEASE	577
GLUTAMINE	565
PHOSPHORIC ACID	530
LACTIC ACID	465
UREA	452
ASCORBIC ACID	444
SODIUM CARBONATE	444
TARTARIC ACID	439
DEXTROSE	436
SODIUM BICARBONATE	402
BIOTIN	389
COPPER	381
CALCIUM CARBONATE	379
SODIUM SULFATE	346
SODIUM ACETATE	321
CISPLATIN	319
MAGNESIUM SULFATE	315
SODIUM CITRATE	312
CYCLOPHOSPHAMIDE	297
ADENOSINE	294
FLUOROURACIL	288
SODIUM PHOSPHATE	279
DIMETHYL SULFOXIDE	277
AMMONIUM CHLORIDE	269
DEXAMETHASONE	268
ETOPOSIDE	260
PACLITAXEL	244
MITOMYCIN	218
ASPIRIN	211
TETRACYCLINE	209
CARBOPLATIN	208
CYCLOSPORINE	207
PREDNISONE	194
IBUPROFEN	192
DOCETAXEL	191
MELPHALAN	182
SOYBEAN OIL	181
VITAMIN E	179
HYDROXYPROPYL CELLULOSE	179
ISOPROPYL ALCOHOL	177
MERCAPTOPURINE	170
ESTRADIOL	170
CETYL ALCOHOL	168
CHLORAMBUCIL	164
NITRIC OXIDE	160
CYTARABINE	159
INDOMETHACIN	155
FOLIC ACID	151
PREDNISOLONE	151
BUSULFAN	144
LIPASE	144
IFOSFAMIDE	141
ALUMINUM HYDROXIDE	141
TENIPOSIDE	140
LOVASTATIN	138
DACTINOMYCIN	138
CALCIUM CHLORIDE	136
...	...

Table 2: Active ingredients mentioned in patents.

drug	active ingredient	documents
ABILIFY	ARIPIPRAZOLE	3
ABRAXANE	PACLITAXEL	16
ACARBOSE	ACARBOSE	69
ACCOLATE	ZAFIRLUKAST	3
ACCUPRIL	QUINAPRIL HYDROCHLORIDE	3
ACCUTANE	ISOTRETINOIN	4
ACEON	PERINDOPRIL ERBUMINE	5
ACETAMINOPHEN	ACETAMINOPHEN	91
ACETAZOLAMIDE	ACETAZOLAMIDE	35
ACETOHEXAMIDE	ACETOHEXAMIDE	26
ACETYLCYSTEINE	ACETYLCYSTEINE	44
ACTH	CORTICOTROPIN	6
ACTONEL	RISEDRONATE SODIUM	1
ACTRON	KETOPROFEN	1
ACYCLOVIR	ACYCLOVIR	71
ACYCLOVIR	ACYCLOVIR SODIUM	2
ACYCLOVIR SODIUM	ACYCLOVIR SODIUM	2
ADALAT	NIFEDIPINE	3
ADENOSINE	ADENOSINE	294
ADRUCIL	FLUOROURACIL	6
ADVICOR	NIACIN	6
ADVICOR	LOVASTATIN	8
ADVIL	IBUPROFEN	3
AGENERASE	AMPRENAVIR	6
AGGRENOX	ASPIRIN	1
AGGRENOX	DIPYRIDAMOLE	1
ALA-CORT	HYDROCORTISONE	4
ALBENZA	ALBENDAZOLE	1
ALBUTEROL	ALBUTEROL	51
ALBUTEROL SULFATE	ALBUTEROL SULFATE	10
ALDACTONE	SPIRONOLACTONE	1
ALDARA	IMIQUIMOD	3
ALENDRONATE SODIUM	ALENDRONATE SODIUM	7
ALEVE	NAPROXEN SODIUM	1
ALIMTA	PEMETREXED DISODIUM	1
ALKERAN	MELPHALAN	7
ALLOPURINOL	ALLOPURINOL	29
ALLOPURINOL SODIUM	ALLOPURINOL SODIUM	2
ALOCRIL	NEDOCROMIL SODIUM	1
ALORA	ESTRADIOL	1
ALOXI	PALONOSETRON HYDROCHLORIDE	5
ALPHADERM	HYDROCORTISONE	1
ALPHAGAN	BRIMONIDINE TARTRATE	2
ALPHAGAN P	BRIMONIDINE TARTRATE	2
ALPRAZOLAM	ALPRAZOLAM	25
ALPROSTADIL	ALPROSTADIL	8
ALTACE	RAMIPRIL	7
AMARYL	GLIMEPIRIDE	2
AMCINONIDE	AMCINONIDE	11
AMIFOSTINE	AMIFOSTINE	28
AMIKACIN SULFATE	AMIKACIN SULFATE	2
AMINOCAPROIC	AMINOCAPROIC ACID	25
AMINOCAPROIC ACID	AMINOCAPROIC ACID	25
AMINOPHYLLIN	AMINOPHYLLINE	7
AMINOPHYLLINE	AMINOPHYLLINE	14
AMITIZA	LUBIPROSTONE	1
AMLEXANOX	AMLEXANOX	4
AMLODIPINE BESYLATE	AMLODIPINE BESYLATE	9
AMMONIUM CHLORIDE	AMMONIUM CHLORIDE	269
AMMONIUM LACTATE	AMMONIUM LACTATE	6
AMOXAPINE	AMOXAPINE	31
AMOXICILLIN	AMOXICILLIN	41
AMOXIL	AMOXICILLIN	1
AMPHOTEC	AMPHOTERICIN B	1
AMPHOTERICIN B	AMPHOTERICIN B	62
AMPICILLIN SODIUM	AMPICILLIN SODIUM	1
ANCEF	CEFAZOLIN SODIUM	1
ANDRODERM	TESTOSTERONE	1
ANDROGEL	TESTOSTERONE	3
ANEXSIA	ACETAMINOPHEN	1
ANEXSIA	HYDROCODONE BITARTRATE	1
ANGIOMAX	BIVALIRUDIN	1
...	...	...

Table 3: Drug names, the active ingredients of which are mentioned in documents.

drug	administration	documents
SELENIUM SULFIDE	TOPICAL	2
REBETOL	ORAL	6
PROCHLORPERAZINE	ORAL	21
PROCHLORPERAZINE	RECTAL	15
PROCHLORPERAZINE	INJECTION	21
OCTREOTIDE ACETATE	INJECTION	6
MISOPROSTOL	ORAL	17
ALBUTEROL SULFATE	ORAL	10
ALBUTEROL SULFATE	INHALATION	7
GENTAMICIN	TOPICAL	47
GENTAMICIN	INJECTION	64
NALBUPHINE	INJECTION	16
ETHOSUXIMIDE	ORAL	20
POTASSIUM CITRATE	ORAL	13
PYRILAMINE MALEATE	ORAL	5
HYZAAR	ORAL	6
NIZATIDINE	ORAL	6
VEPESID	ORAL	6
VEPESID	INJECTION	6
NEOSAR	INJECTION	5
TARCEVA	ORAL	44
MERCAPTOPURINE	ORAL	159
CARBAMAZEPINE	ORAL	68
RISPERDAL	ORAL	7
CLOTRIMAZOLE	ORAL	30
CLOTRIMAZOLE	TOPICAL	23
CLOTRIMAZOLE	VAGINAL	14
SUSTIVA	ORAL	7
SECOBARBITAL SODIUM	INJECTION	2
SECOBARBITAL SODIUM	ORAL	2
PERPHENAZINE	ORAL	20
INDOCIN	ORAL	5
INDOCIN	RECTAL	2
CHANTIX	ORAL	2
ALEVE	ORAL	2
VERAPAMIL HYDROCHLORIDE	ORAL	4
VERAPAMIL HYDROCHLORIDE	INJECTION	3
GUANFACINE HYDROCHLORIDE	ORAL	1
DEPO-PROVERA	INJECTION	1
STAPHCILLIN	INJECTION	1
TRILAFON	ORAL	1
TRILAFON	INJECTION	1
ECONAZOLE NITRATE	TOPICAL	4
TRICOR	ORAL	7
PROVERA	ORAL	3
DILAUDID	ORAL	1
PARACORT	ORAL	8
LOGEN	ORAL	13
CIPROFLOXACIN	OPHTHALMIC	17
CIPROFLOXACIN	INJECTION	40
CHOLESTYRAMINE	ORAL	50
CONCERTA	ORAL	1
WARFARIN SODIUM	ORAL	6
VITAMIN D	ORAL	100
LYRICA	ORAL	1
AMOXAPINE	ORAL	29
PANRETIN	TOPICAL	3
SYMLIN	SUBCUTANEOUS	2
FLUMAZENIL	INJECTION	12
DRONABINOL	ORAL	32
DIPROSONE	TOPICAL	1
METICORTEN	ORAL	4
...	...	...

Table 4: Drug names and routes of administration that are mentioned in documents.

drug	dosage form	documents
TALC	POWDER	570
SODIUM CHLORIDE	INJECTABLE	429
DIMETHYL SULFOXIDE	SOLUTION	276
INSULIN	INJECTABLE	265
FLUOROURACIL	SOLUTION	256
DEXAMETHASONE	SOLUTION	247
LENTE	INJECTABLE	242
STERILE WATER	LIQUID	229
SODIUM BICARBONATE	INJECTABLE	183
CYCLOSPORINE	SOLUTION	181
PREDNISONE	SOLUTION	176
FLUOROURACIL	INJECTABLE	153
IBUPROFEN	SUSPENSION	150
CISPLATIN	INJECTABLE	149
DEXAMETHASONE	INJECTABLE	147
CYCLOPHOSPHAMIDE	INJECTABLE	140
MAGNESIUM SULFATE	INJECTABLE	133
ETOPOSIDE	INJECTABLE	133
AMMONIUM CHLORIDE	INJECTABLE	127
PACLITAXEL	INJECTABLE	126
INDOMETHACIN	SUSPENSION	118
TAXOL	INJECTABLE	113
ADENOSINE	INJECTABLE	112
MITOMYCIN	INJECTABLE	109
CYCLOPHOSPHAMIDE	TABLET	107
CARBOPLATIN	INJECTABLE	102
MITOXANTRONE	INJECTABLE	101
CYCLOSPORINE	INJECTABLE	97
CYTARABINE	INJECTABLE	96
ETOPOSIDE	CAPSULE	94
NAPROXEN	SUSPENSION	92
IBUPROFEN	CAPSULE	92
DEXAMETHASONE	TABLET	88
ERYTHROMYCIN	SOLUTION	80
CYCLOSPORINE	CAPSULE	79
ESTRADIOL	TABLET	78
IBUPROFEN	TABLET	77
IFOSFAMIDE	INJECTABLE	76
INDOMETHACIN	CAPSULE	75
AMPICILLIN	CAPSULE	72
BALANCED SALT	SOLUTION	71
PREDNISONE	TABLET	70
ISOFLURANE	LIQUID	70
POTASSIUM IODIDE	SOLUTION	70
MEGESTROL ACETATE	SUSPENSION	66
THIOTEPA	INJECTABLE	65
SIMVASTATIN	TABLET	65
LOVASTATIN	TABLET	64
MERCAPTOPURINE	TABLET	62
HYDROCORTISONE	POWDER	62
TAXOTERE	INJECTABLE	61
DIAZEPAM	SOLUTION	60
DACARBAZINE	INJECTABLE	59
FLUOROURACIL	CREAM	58
TESTOSTERONE	INJECTABLE	57
THEOPHYLLINE	SOLUTION	56
FUROSEMIDE	SOLUTION	56
FLUTAMIDE	CAPSULE	55
HYDROCHLOROTHIAZIDE	SOLUTION	55
POTASSIUM CHLORIDE	INJECTABLE	55
PREDNISOLONE	TABLET	55
PENTOSTATIN	INJECTABLE	54
TRETINOIN	SOLUTION	53
FLUOXETINE	CAPSULE	53
FOLIC ACID	INJECTABLE	53
ACYCLOVIR	SUSPENSION	53
CLADRIBINE	INJECTABLE	52
HYDROXYUREA	CAPSULE	51
BENZYL BENZOATE	EMULSION	50
...	...	...

Table 5: Drug names and dosage forms that are mentioned in documents.