Intelligence Surveillance of Health Care-associated Infections with SADI Semantic Web Services

Alexandre Riazanov, Artjom Klein, Arash Shaban-Nejad, Gregory Rose, Allan Forster, David L. Buckeridge, and Christopher J.O. Baker

Nosocomial condition: ICD-10 Y95

Not present or incubating prior to the patient being admitted to the hospital, but occurred within 48-72 hours after admittance to the hospital.

- ~ 1 in 10 patients have HAI
- Up to 60% of infections present after discharge
- Urinary Tract  40%
- Surgical Wound  25%
- Respiratory Tract  20%
- Bacteremia  3%
- Others  12%


Chain of infection in hospitals

Clinical Intelligence: tools for analyzing clinical data for research, surveillance and management.

Ad hoc querying is desirable in many scenarios, e.g., surveillance and cohort selection.

Should be self-service to be economical => must be semantic, so that non-technical users don't need to master SQL.

We are trying to implement semantic querying with SADI Semantic Web services that can be automatically discovered and orchestrated.
### Outline

- To semantically query one or more RDB, we write SADI services drawing data from RDB, communicating it in RDF.
- SPARQL is used to query the network of services (we currently use the SHARE query engine).
- Querying is done modulo OWL ontologies.
- This is more flexible than methods based on declarative mappings of relational data to RDF/OWL because any Java (Perl, Python, ...) code can be used in the services.
- Testing in a surveillance scenario for Hospital-Acquired Infections on a data warehouse from The Ottawa Hospital.

### Application: Hospital-Acquired Infection

- HAI are a big practical problem: thousands of deaths and $100Ms in expenses each year in Canada alone.
- Extensive surveillance helps to cope with the problem.
- HAI surveillance professionals need to make rational decisions based on patient records.
- HAI researchers need to analyse data to develop and evaluate prevention methods.
- Currently, such data analysis is mostly manual, limited in scope and costly.
Data Analysis Targets

- How many incidents of diseases caused by Serratia bacteria, were diagnosed in January 2011?
- What patients were diagnosed with SSI while they were taking corticosteroids systemically?
- How many diabetic patients were diagnosed with SSI?
- How many CAUTI diagnoses were made within 7 days from the time a test indicated the patient's serum glucose level was between 7.0 and 7.5 mmol/L?
- Conclusion: need ad hoc self-service querying of clinical data.

Experimental Setting

- Data: extract from The Ottawa Hospital warehouse, previously used in a HAI surveillance study.
- Contains potentially HAI-related data on diagnoses, surgical procedures, drug prescriptions, patient demographics, hospital infrastructure and personnel ...
- The data warehouse is a single relational DB with tables representing patients, encounters, procedures and drug prescriptions, etc
- We would like to query this data semantically, e.g., with SPARQL modulo OWL ontologies.
SADI Semantic Web Services and SHARE

- SADI Web services consume and produce RDF => no syntactic interoperability problem.
- SADI services have semantic descriptions in OWL specifying what they do; explicitly specify relations between input/output.
- If a client knows what it wants to do with some data, it can automatically find SADI services that can do this.
- Most powerful client: SPARQL query engine, e.g., SHARE. Given a SPARQL query, finds necessary services, invokes them.
- SADI services can be written on top of all kinds of data: RDB, online databases, XML repositories, RDF triplestores, etc.
- SADI services can wrap analytical resources: bioinformatics analytical software, non-semantic Web services, etc.

SADI Architecture

- SADI's Semantic Open Collaborative Registry
- SHARE Query Client
- Data Driven
- Publish ontologies to, or select ontologies from The Web.
- SADI-Compliant Web Services
- to match-make between data and semantics of each service
Semantic Query of Relational Data with SADI

• SADI services are good for data federation: multiple heterogeneous distributed autonomous resources can be queried together.

• We argue that SADI also has a lot of value as a semantic querying medium for RDB, even if it's a single RDB.

• We write a bunch of services drawing data by SQL querying from our data warehouse and query these services with SPARQL via SHARE.

Ontologies Used for Semantic Query

• Ontologies act as data schemas accessible to non-tech. users.

• Hospital-Acquired Infection Ontology
  — HAI-specific concepts: surgical site infection, catheter-associated urinary tract infection, ..
  — General health care related concepts: disease, medical test,
  — Semanticscience Integrated Ontology (SIO) upper ontology.

• Several nomenclatures turned into OWL taxonomies: ICD10 (diagnoses and procedures), DIN (Drug Identification Number), ATC and AHFS (therapeutic classifications).

• Extra Simple Time Ontology
  — Time instants, intervals and durations.
  — Standard Allen’s temporal predicates:
Our SADI Services

• Main services drawing data from the data warehouse:
  – `getDiagnosisByPatient` enumerates known diagnoses for a specified patient
  – `getProcedureByClass` enumerates instances of a specified medical procedure class from the HAI Ontology

• Online public -general purpose- services: mostly for drug and disease nomenclatures, e.g.
  – Drug product (DIN) <-> ATC Drug Class
  – ICD9 category --> more specific categories
Registered HAIKU Services

What patients were diagnosed with SSI while they were taking corticosteroids systemically?

1 SELECT DISTINCT ?Patient
2 FROM <http://unbsj.biordf.net/haiku/q1cont.rdf>
3 FROM <http://unbsj.biordf.net/haiku/cster.rdf>
4 WHERE
5 { # corticosteroids for systemic use
6   <http://unbsj.biordf.net/ontologies/atc.owl#H02>    
7   atcso:hasSubClass ?DrugClass .
8   # 'is referred to by'
10  ?DIN_Record a lsrn:DIN_Record .
11  # 'has attribute'
12  ?DIN_Record sio:SIO_000008 ?DIN_Identifier .
13  # 'has value'
15  # 'is about'
19  # Select patients with SSI:
21  # Diagnosis haio:identifies ?Incident .
22  ?Incident haio:disease_has_type haio:SSI .
23  # Temporal check ommitted
}
Temporal Reasoner: 14 SADI Services

- Allen’s predicates
  meets / ‘is met by’
  finishes / ‘is finished by’
  starts / ‘is started by’
  after / before
  during / contains
  overlaps / ‘is overlapped by’
- Service: interval duration
  getDurationOfTimeInterval
- Service: interval between two given intervals:
  getTimeIntervalBetween

Conclusions and Future Work

- SADI may be the right medium for semantic querying of RDB, at least in Clinical Intelligence: no really bad obstacles so far.
- We just started to work on real data, so a lot of service writing and experiments with queries will be forthcoming.
- HAI Ontology is being prepared for publishing: modeling clean-up, alignments.
- A usability study will be necessary and a comparison with other approaches (D2R and query rewriting).
- Temporal Reasoning Services provide real world functionality relevant for ad hoc query in the clinical intelligence domain.
Acknowledgements

NSERC CRSGN
CIHR IRSC
C-BRASS
Canadian Bioinformatics Resources
As Semantic Services
Powered by SADI
SADI
Find. Integrate. Analyze.

JOURNAL OF BIOMEDICAL SEMANTICS

www.jbiomedsem.com

Editors-in-Chief:
Dietrich Rebholz-Schuhmann (United Kingdom) and Goran Nenadic (United Kingdom)

Journal of Biomedical Semantics is an open access journal that encompasses all aspects of semantic resources used for data integration, modeling, interpretation and exploitation in biomedical research.

To submit your next manuscript to Journal of Biomedical Semantics go to www.jbiomedsem.com
JOURNAL OF BIOMEDICAL SEMANTICS

www.jbiomedsem.com

Reasons for publishing in Journal of Biomedical Semantics:

- All articles published in the journal are open access; universally accessible online without charge.
- High visibility for your work - anyone with Internet access can read your article, free of charge.
- Included in PubMed - making your work easy to find, read and cite.
- Articles are published immediately upon acceptance.
- Authors retain the copyright of their articles.
- Electronic submission and peer-review makes the whole process of publishing your article simple and efficient.
- Publishing online means unlimited space for figures, extensive datasets and video footage.
- Professional formatting meets the standards expected by researchers.