Towards Semantic Health Assistants

Deborah L. McGuinness
Tetherless World Senior Constellation Chair
Professor of Computer Science and Cognitive Science
Director RPI Web Science Research Center
RPI Institute for Data Exploration and Application Health Informatics Lead

with support from
Evan W. Patton, James McCusker, Katie Chastain, Bassem Makni, & Heng Ji
Semantics and Healthcare by example

- Investigate how web technologies support patients and their teams navigate through complicated health issues, by supporting discovery, access, and integration of appropriate content... ultimately improving general health literacy and allowing functioning at a level higher than a person’s training.

- Investigate how semantic web technologies can be used to help form and explore drug repurposing hypotheses (along with supporting content) that may be worthy of further investigation.
Vision

Drug Repurposing

- Use semantic technologies to encode and process biological knowledge to generate hypotheses about new uses for existing drugs.
- Combine different sources with varying knowledge quality
- Track knowledge provenance
- Create actionable advisors that can explain their suggestions

Health Advisor

- Use semantic technologies to encode and process medical knowledge at different levels of abstraction for the medical professional/patient spectrum.
ReDrugS
(Repurposing of Drugs using Semantics)

Ontological Resources
Protein/Protein Interaction Ontology, Semanticscience Integrated Ontology, Gene Ontology
vocabulary, relationships

ReDrugS Quad Store
condenses

ReDrugS API
SADI-based API provides interaction network search and expansion based on consensus probabilities
queries graph

Analytical Tools
Cytoscape, R, Python, etc.

Experimental Method Assessment
Con•dence scores of experimental methods.

iRefIndex
Protein/Protein Interaction Database
curated into
curated to nanopubs
evidence to probability

ReDrugS
Drug/Protein Interactions

Researchers
pose questions
view data
explore network

Joint Work: McCusker, Solanki, Chang, Yan, Dumontier, Dordick, McGuinness

McGuinness 2/27/14
What is a Nanopublication?

NanoPub_501799_Assertion

- a direct interaction
  - has-target: SLC4A8
  - has-participant: CA2

NanoPub_501799_Attribution

- hadPrimarySource: Regulation of the human NBC3 Na+/HCO3- cotransporter by carbonic anhydrase II and PKA
- wasQuotedFrom: BioGRID
Topiramate Disease Associations: \( p \geq 0.9 \)

Repurposing Drugs with Semantics

Prev nanopub
Topiramate Disease Associations: \( p \geq 0.5 \)

Repurposing Drugs with Semantics
Inferring Probability from Evidence

Nanopublication provenance encodes the methods that we infer probabilities from.

Class: NanopubDerivedFrom_MI_0096
EquivalentClass:
  wasGeneratedBy some ‘pull down’
SubClassOf: Confidence2

Class: Confidence2
EquivalentClass: ‘has attribute’ min 1
  (‘probability value’ and
  (‘has value’ value 0.95))

McGuinness 2/27/14
Basic OWL entailments compute the probability from confidence scores.
Computing an Interaction Probability

- Find all nanopublications that assert a given interaction.
- Infer the probability of each nanopublication based on its evidence (experimental method).
- Combine multiple opinions (confidence scores) on evidence using Bayes Theorem.
- Combine multiple nanopublications using composite z-scores.
Different databases can provide the same assertions. This might be experimental replication! We model this with composite z-scores:

\[
F \left( \sum_{i=1}^{n} F^{-1} (p_i) \right)
\]

F(x): Cumulative Distribution Function (converts z-scores to probabilities)
99.98% coverage of the ~936,000 nanopubs with evidence data from iRefIndex.

Top 10 methods (86% coverage):

<table>
<thead>
<tr>
<th>Method</th>
<th>Count</th>
<th>P conf</th>
<th>M conf</th>
</tr>
</thead>
<tbody>
<tr>
<td>two hybrid [mi:0018]</td>
<td>199130</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>genetic interference [mi:0254]</td>
<td>196717</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>affinity chromatography technology [mi:0004]</td>
<td>117659</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>tandem affinity purification [mi:0676]</td>
<td>70545</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>two hybrid pooling approach [mi:0398]</td>
<td>60715</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>anti tag coimmunoprecipitation [mi:0007]</td>
<td>42249</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>pull down [mi:0096]</td>
<td>37676</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>two hybrid array [mi:0397]</td>
<td>29806</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>x-ray crystallography [mi:0114]</td>
<td>29182</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>anti bait coimmunoprecipitation [mi:0006]</td>
<td>22533</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
• 99.98% coverage of the ~936,000 nanopubs with evidence data.
Repurposing Drug Repurposing

• Now considering the health advisor – taking in unstructured data (in addition to structured data to help navigate through complicated health situations
• We can reuse the ReDrugS framework for reasoning about probabilistic evidence
• We need to model and compute trust and uncertainty in natural language processing algorithms on clinical notes and medical literature
• How can this be used for building consumer-oriented applications to improve patient-doctor interaction?
Mobile Health Advisor

User Device

PHR/EHR

Physician

CDC

Icahn School of Medicine at Mount Sinai

Joint work: Patton, Chastain, Makni, Ji, McGuinness

McGuinness 2/27/14

15
Mobile Health Monitoring

Applications

Reasoning Services

Mobile Semantic Health Integration Framework

Hardware Abstraction Layer / Device APIs

Accelerometer

GPS

Scale

Blood Pressure

Sleep

Pedometers

Heart Rate

McGuinness 2/27/14
Doctor’s Perspective

- Extract information about test results, with drill-down interface
- Highlight key information extracted from text
- Determine order of events using natural language processing and semantic integration

Test Results
- R. breast mammogram, ultrasound
- R. breast ultrasound-guided core biopsy
- Breast MRI
- L. breast mammogram and ultrasound
- L. breast MRI-guided biopsy
- Metastatic survey
- Date not available

Treatments
- Paclitaxel and trastuzumab
- 12 weekly doses from 07/16/13 until 10/01/13
- Surgical resection
- Not yet scheduled, follows "Paclitaxel and trastuzumab"
- Adriamycin®/Cytoxan®
- Four cycles, follows "Surgical resection"
- Herceptin®
- Follows "Adriamycin/Cytoxan"

People

McGuinness 2/27/14
Patient’s Perspective

Link to external resources for descriptive information
Identify possible side-effects and coping strategies
Focus on information the patient is concerned about
Explanation of reasoning and why information may be unavailable

McGuinness 2/27/14
(A few) Needs

- Methods for capturing, digitizing, and transforming multimodal information sources into structured knowledge
  - Paper documents, audio recordings, web resources, data from fitness devices, etc.
- Rich, curated resources of structured information about medications, treatments, clinical trials, genetics
  - Data from web forums and other less curated sources.
- Methods for extracting knowledge from curated resources such as PubMed and peer-reviewed publications
Medinet

- Represents medical data and knowledge as a network and integrates multiple sources of varying quality
- Constructed from both high-quality, structured sources (e.g. SNOMED-CT) and natural language processing of clinical notes and medical literature
- Question-answering using technologies based on IBM’s Watson-like technology
Integrating Medical Resources across Systems

Clinical Notes
Textual (PDF, JPEG), Audio (WAV, MP3, AIF)

Ontological Resources
SNOMED-CT, UMLS, HL7, Semantic Sensors, BioPortal

Online Resources
WebMD, PubMed, DrugBank, DBPedia, Cancer Forums/Wikis

Open Source
OCR/SR Software
Tesseract / OCRopus, Sphinx / Julius

Fact Extraction / Cross-Source Truth Finding
structured terms, relationships

Health Apps / Medical Devices

Medinet
Structured representation of global knowledge and personal health data
Provenance / trust annotations about data sources and processing

User

Smart Phone Device

Electronic Medical Records

Watson-like Question Answering

converted via

natural language

unstructured or semistructured text

share data

capture data

pose questions view data record data

submit queries answer queries

queries graph

view data record data

share data
Mobile Health Monitoring

Patient Blood Data
drag and drop characteristics to add

Glucose levels

2013-09-03
Glucose 108 milligramPerDecliter
Chloride 105 millimolPerLiter

r. Smith

Steroids

1. Aug 1. Sep

Normal range

Highcharts.com
Capture and modeling of trustworthiness of the source

Linkage to shared knowledge about the patient across multiple documents

She elected to proceed with neoadjuvant chemotherapy with 12 weekly doses of paclitaxel in combination with trastuzumab.
Linkage to external resources (e.g. drugs.com) that may be useful for inferring other relationships.

Dereferencing generic statements to past events ('received her first dose…' => start of treatment)

She elected to proceed with neoadjuvant chemotherapy with 12 weekly doses of paclitaxel in combination with trastuzumab.

2013-07-16

She received her first dose of paclitaxel and trastuzumab on Tuesday, July 16th.

A third procedure on patient, a CT scan, revealed no spread of the cancer.

DC:Date 2013-08-27
Breast Cancer Wiki

- Collect anecdotal information from breast cancer patients/survivors in natural language
- Bootstrap Medinet knowledge acquisition
- Provide a portal to ask questions of general Medinet knowledge

Recommendations

The table below lists recommendations posted to this website by anonymous users. Please consult your oncologist or general physician before implementing any of these recommendations.

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Severity</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing drains</td>
<td></td>
<td>stage 4</td>
<td>surgery (mastectomy, lumpectomy, etc.)</td>
</tr>
<tr>
<td>Tea tree oil for nail problems</td>
<td>55</td>
<td>stage 3</td>
<td>chemotherapy</td>
</tr>
</tbody>
</table>
Discussion

• Building a large interaction graph from diverse resources
• Extensible infrastructure for additional data
• Filtering, navigation/expansion
• Probabilities based on experimental methods and resources (provenance)
• Can be used for many other applications.
Conclusion

• Semantic technologies can and are changing the future of healthcare and computational biology.
• The technological barriers are small…. (but of course social/political/legal barriers exist…)
• We are looking for partners for providing and consuming data and collaborators in general.
Extra
Health Advisor Motivation

• How can web technologies support patients and their teams navigate through complicated health issues, by supporting discovery, access, and integration of appropriate content?
• How can web technologies be used to improve patient health literacy to increase information transfer of patient-doctor interactions and allow people to function at a level higher than their training?
• With ever-increasing medical knowledge, how does a medical professionals assimilate data in a timely manner?
Drug repurposing is often done by accident, or in a limited way.

New advances enable evaluation of any drug or mechanism of action against any disease or adverse event.

We are investigating how semantic web technologies can be used to find interesting connections and form hypotheses (along with supporting content) that may be worthy of further investigation.