SEMANTICS ENABLED PROACTIVE AND TARGETED DISSEMINATION OF NEW MEDICAL KNOWLEDGE

Lakshmish Ramaswamy & I. Budak Arpinar
Dept. of Computer Science, University of Georgia
laks@cs.uga.edu, budak@cs.uga.edu
**Motivation**

Prilosec + Plavix

Mr. Smith

Dr. Brown
GOALS

- Cut down delays between medical discoveries and implementation via better knowledge dissemination
- Reduce information overload
- Minimal overheads on doctor establishments
- Medical information dissemination that is
  - Proactive – Push-based
  - Targeted
  - Timely
**BACKGROUND: THE PUB-SUB PARADIGM**

- Information filtering and propagation framework
- Subscriptions are continuous queries
- Used heavily in financial applications

![Diagram showing Pub/Sub Infrastructure with P1, S1, S2, and S3](image.png)
QUERY TYPES

- Topic-based
  - Queries choose topics of interest from pre-specified hierarchy

- Content-based
  - Queries on content of published items

- Type-based
  - Queries specified on object types

  - Distributed implementation to achieve scalability
**Pub-Sub: Pros and Cons**

- **Works well if**
  - Know what information will be of interest
  - Anticipate the information that will be published
  - Published information is structured
  - Queries can be crisply specified

- **Our scenario doesn’t exhibit these characteristics**
  - Scientific literature is free text
  - Cannot anticipate research
  - Too many queries or too coarse-grained queries
EHR as Basis for Info Dissemination

- Comprehensive and up-to-date (ideally !!) information about patient
  - Age, gender, physical activity, family support
  - Medical conditions, individual and family history
  - Drugs and medications, past reactions
- Can serve as basis for identifying relevant information
- Collected routinely as part of treatment process
- Fast increasing adoption
CHALLENGES

- Differences in terminologies and info representation

- Noisy EHR data
  - Incompleteness
  - Errors

- Personalization

- Scalability
  - Several thousand physicians, millions of EHRs
ARCHITECTURE

- EHR Database
- Semantic Processing
- EMR Clustering & Query Generation
- Semantic Matching & Querying
- Notification/Alert System
- Feedback-based Tuning
- Ontologies (UMLS)
- Medical Literature
  - Semantic Text Mining
  - Annotated Literature
**Semantic Annotation**

- Foundation to support powerful query and relevance mechanisms
- Enhancing text with structured domain knowledge
- The result is a set of explicit assertions indicating named-entities within them
- Term, concept & relationship identification
- We use Stanford NLP

Mr. Smith

.... Current medications include **Plavix** and **Prilosec** ...

Ontology

- Anti-blood clotting Agent
- Clopidogrel
- generic
- Plavix

Current medications include **Plavix** and **Prilosec**...
RELEVANCE DETERMINATION

- Content similarity b/w EHRs and medical articles
- Current approaches rely on syntactic similarity
  - Term vector approaches - Common words in documents and their relative importance
  - Importance measured in terms of relative frequencies (TF-IDF metric)
- Many shortcomings
  - Terminology differences
  - EHRs not likely to have significant word repetitions
  - Repetitions may not imply strong emphasis
  - Blind to relationships among words
**Semantics Distance-Based Relevance**

- Quantifies relationship strength
- Based on structure of domain ontology
  - Min hops between concepts
  - # paths between concepts
  - Weighted hop distance
- Compute semantic distance b/w concept pairs
- Aggregate semantic distances
- Can be used for relevance determination, ranking etc.
Mr. Smith

... Current medications include Plavix and Prilosec ...

Concomittent use of Clopidogrel and PPI was associated with increased risk of adverse outcomes

Ontology
SEMANTICS GRAPHS-BASED RELEVANCE

- EHRs & articles mapped to semantic graphs
- Structural correlation of graphs
  - Identify links among concepts in EMR’s and articles
  - Sub-graphs with dense cross-cutting paths signify higher degree of relevance
- More powerful but computation intensive
LOTS OF INTERESTING QUESTIONS

- Fuzzy matching
- How to associate weights with ontology links?
- How to incorporate user feedback?
  - Can it be used to strengthen/weaken relationships in ontology?
- How to scale the system?
  - Clustering EHRs
  - Distributed processing – Semantic overlays, Cloud
- Patient privacy issues
  - How much information can be exposed?
  - Where does the processing occur?