Convenient Bedfellows: Natural Language Processing and Clinical Decision Support

J. Marc Overhage, MD, PhD
Chief Medical Informatics Officer
### Machine-readable data & knowledge will shift paradigms

#### Unstructured data
- **Patient-specific**
  - Diagnostic data
    - Imaging
    - In-Vitro diagnostics
    - Physical examinations
    - Genome, Proteome
    - ...
  - Patient record
    - Family data
    - Diseases

#### Disease models
- **Non-patient-specific**
  - Disease causalities
    - Infectious diseases
    - Addiction
    - Cancer
    - Diabetes
    - ...
  - Physician
    - Education
    - Experience

#### Therapy
- **Diagnosis**
  - Therapy selection
  - Treatment

- **Verification**
  - Post-treatment result evaluation
  - Aftercare

#### Action domain

---

© Siemens 2012. All rights reserved.
## Machine-readable data & knowledge will shift paradigms

### Structured data

**Patient-specific**
- Diagnostic data
  - Imaging
  - In-Vitro diagnostics
  - Physical examinations
  - Genome, Proteome
  - ...
- Patient record
  - Family data
  - Diseases

### Patient Model(s)

**Non-patient-specific**
- Comprehensive models
  - Anatomical m.
  - Physiological m.
  - Biochemical m.
  - Metabolomic m.
  - ...

### Individualized Therapy

- Therapy selection
- Treatment
- Post-treatment result evaluation
- Aftercare

### Knowledge domain

### Data domain

### Action domain

### Diagnosis & Therapy proposals

### Falsification
Overview Model

- Lab
- Personalized Models
- Bayesian Belief Networks
- Patient Data
- Inference Engine
- Knowledge Base
- Semantic Integration
- Patient Data
- Question Answering
- Evidence
- Literature
- Guidelines
- Textbooks
- Patient Data
- Learning
- EMR
- HIE
- PHR
Improving Models

**Question: Acute Myocardial Infarction 4.7**

16. Is the left ventricular systolic function (LVSF) documented as an ejection fraction (EF) less than 40% or a narrative description consistent with moderate or severe systolic dysfunction? (LVSD)

- Yes
- No

**Evidence:**

VENTRICULOGRAPHY: Ventriculography revealed an overall preserved left ventricular ejection fraction, EF 54%. There was some inferior hypokinesis. The left ventricular end diastolic pressure was normal at 4 mmHg. Central aortic pressure was 129/59 mmHg.

ANGIOGRAPHY: Adequate cine angiograms were obtained. Circulation is right dominant. The left main gives rise to the LAD and circumflex systems. The circumflex has luminal irregularities present throughout and is compromised by a small first marginal, a large second, with some luminal irregularities. The LAD has a large septal system. Around the first septal and diagonal there is a 50-70% lesion. Also prior to the takeoff of the second diagonal there appears to be a 30% lesion. The LAD continues to...
Data Class Agents

Test Result

Based on LOINC code, is test result potentially reportable?

YES

Based on result characteristics, classify test result type

Based on result characteristics, classify test result type

Numeric Result

Discrete Result

Text Result

Positivity Criteria
Result outside of known reference range

Positivity Criteria
Result one of known positives (reactive, positive, present, etc)

Positivity Criteria
Natural language processing identifies condition in positive context

Examples of positives

<table>
<thead>
<tr>
<th>Test</th>
<th>Result</th>
<th>Ref. Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead</td>
<td>17</td>
<td>0-9</td>
</tr>
<tr>
<td>Hep B Surf Ag</td>
<td>&gt;1000</td>
<td>&lt;10</td>
</tr>
<tr>
<td>CD4 Counts</td>
<td>489</td>
<td>55-1600</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-1 AB QL</td>
<td>REACTIVE</td>
</tr>
<tr>
<td>Hep C RNA PCR</td>
<td>DETECTED</td>
</tr>
<tr>
<td>Chlamydia by DNA</td>
<td>POSITIVE</td>
</tr>
</tbody>
</table>

Moderate growth of Methicillin Resistant Staph Aureus seen.

There is a left lower lobe infiltrate consistent with pneumonia. Impression: Left lower lobe pneumonia.
REMIND Knowledge Platform

Images
Patient Factors
Proteomics
Genomics
Treatment Plans

Extraction
Combine Conflicting Local Evidence
Probabilistic Inference Over Time

Extraction
Extraction
Extraction
Extraction
Extraction

Combination

Inference

Plug-in Domain Knowledge (e.g., CMS Measures)

*Not considered a Medical device.
Analyze As You Type

• The sound interpretation of a single data entry requires the reflection of how this information relates to proceeding or succeeding information.
• Highlighting the semantic relations and contexts between the data entries at different time points paves the way towards efficient and intelligent information access and analysis.
• Combining the process of entering data and data analysis the user is not slowed down in his work progress.

The user activity (typing, highlighting) defines the current focus and instantaneously triggers data analysis.

Entered findings can be added to the protocol.

All information relevant in the context of the current focus are highlighted.

Enter symptoms, findings, diagnosis.
Time series of mixed structured data like measurements and textual descriptions

- Current context (e.g. recently entered words)

- Data analytics component

- Knowledge base
- Domain model

- Interaction and context detection mechanism

- Semantic knowledge processing component

- Rendering

- Information need
  - Process: diagnosis
  - Context: fever
  - Retrieval mode: on the fly

- Patient record

- Temperature data extracted and highlighted
Providing Contextual Cues

A/P:

1) DM - due for A1c, optho consult, lytes.
2) HTN increase lisinopril, get EKG, consider ECHO
3) Hyperlipidemia - overdue for lipids, get today
4) Screening - need colonoscopy, PSA

4) Screening - need colonoscopy, PSA
Individualized Models

- Based on data from large study, >4000 patients over 10 years
  - Genetic (SNPs), clinical, lab, imaging, ...
- Analyzed 100s of potential predictive variables to identify the most informative subset
- Learned a Bayesian Network
  - predicts patients who are at high risk to become hypertensive (AUC >0.85).
- Model also suggests patient specific interventions that can reduce risk for a subset of patients,
- Undergoing multi-site validation

Overview Model

- Patient Data
- Inference Engine
- Knowledge Base
- Patient Data
- Semantic Integration
- Personalized Models
- Baysian Belief Networks
- Question Answering
- Learning
- Evidence
- Patient Data
- EMR
- HIE
- PHR
- Literature
- Guidelines
- Textbooks
- Semantic Modeling / Ontology
Knowledge Extraction and Engineering

- Search
- Explore
- Organize
- Extract
- Advise
- Aggregate

Use-cases

Text mining and information retrieval

Machine learning

Semantic modeling and reasoning

Data
- Unstructured and textual data
- Relational data
- Semantic data

Technology
Medical Knowledge Engineering Methodology (KEMM)

**Features**
- Query Pattern Derivation
- Ontology Identification
- Ontology Modularization
- Ontology Alignment and Integration
- Reasoning-based Ontology Enhancement
- Deployment and Testing
- Medical ontologies

**Challenge**
- Semi-automatic identification of knowledge requirements
- Handling of large-sized, complex and cross-linked ontologies
- Alignment of the different perceptions and understanding of a domain

**Application**
- Alignment between radiology-related OWL ontologies
- Integration of Linked Data resources in the context of Medical Image Search application
- Patient classification and extension of medical image annotations
Semantic Integration

Scope

- Access to heterogeneous knowledge sources by mediating the knowledge models and providing an integrated view of the data
- Synthesizing the understanding of a subject from different perspectives
- Incorporation of new information into a body of existing knowledge

Applications

- Horizontal, vertical, and longitudinal integration of patient data, medical information and knowledge
- Establish mappings between Ontologies of overlapping domains

Tasks

- Identify the piece of knowledge required for an application
- Discover the interrelation of knowledge pieces in the context of the application
- Fine-tuning of established mapping results by tracking user behavior
Knowledge-based, medical services using semantic annotated data

Semantic Processing
- Semantic Annotation
- Ontology Mapping
- Reasoning & Inference

Intelligent Services
- Semantic Search
- Quality Control
- Intelligent Navigation
- Image Diagnose
- Multi-modal Interaction
- Intelligent Diagnose

Heterogeneous data sources
- Medical Images
- Patient-records
- Treatment-plans
- Online Knowledge
- Expert Knowledge

Segmentation
Text Mining
Formalization
Dashboard Knowledge-driven Healthcare

- Manage cases and handle clinical conferences
- Analyze patient history with support of semantic search
- Navigate through related knowledge and guidelines
- Share experience with other clinicians in the domain
- Analyze and document symptoms and diagnosis
- Find and compare similar cases and treatments
- Explore topics related to the symptoms or disease
- Locate available experts in domain
- Manage consultation process with experts
To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease—these are our ambitions.

Sir William Osler, 1849–1919  
Father of Modern Medicine