Enterprise Search in Medical Data

Allan Hanbury

Henning Müller

“The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°257528”
Health Information

Patient-specific information

- Narrative reports
- Structured data
- Images
- Radiology images

Knowledge-based information

- Journals
- Books
- Practice guidelines
- Taxonomies, vocabularies, ontologies, ...
- Language resources

- Personal Health Records (PHR)

- Primary – original research (in journals, books, reports, etc.)
- Secondary – summaries of research (in review articles, books, practice guidelines, etc.)
Health Records

- Health records are currently almost always only accessible by patient ID
- Could even more on optimum treatment be learned by searching and mining the huge archives of health records?
- Example: Kaiser Permanente and Vioxx
- Feufel et al. “believe that it is inefficient, and in some cases unethical, to store [population-based data] without installing mechanisms to allow access and publication in appropriate and useful ways.” (Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020, MIT Press, 2011)
Use Case: Radiology

Interaction with the system

Choose Patient study
Load Image Data into PACS Software
Setup viewing modes
Analyzing the images

Search behaviour

Pathology known?
Yes

No

Diagnosis differential diagnosis

Needed as image search not available

Switch between modalities
Scroll through images
Analyze patient history
Adjust brightness window
Switch viewing modes
Measure distances

Search books, internet, databases
Ask colleagues
Describe pathology
Radiology Example
<table>
<thead>
<tr>
<th>Rank</th>
<th>Diagnosis</th>
<th>Reliability</th>
<th>Age</th>
<th>Gender</th>
<th>Smoking</th>
<th>Quantity</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fibrose, pulmonary_fibrosis, Diagnosis reliability: 100 %, Age: 64 years old, Gender: m, Smoking: true, Quantity: 60 UPA</td>
<td></td>
<td></td>
<td>m</td>
<td>true</td>
<td>60 UPA</td>
<td>99.98%</td>
</tr>
<tr>
<td>2</td>
<td>Fibrose pulmonaire idiopathique, pulmonary_fibrosis, Diagnosis reliability: 98 %, Age: 66 years old, Gender: m, Smoking: true, Quantity: 60 UPA, Bronchoscopy transbronchial: interstitial inflammation</td>
<td></td>
<td></td>
<td>m</td>
<td>true</td>
<td>60 UPA</td>
<td>99.98%</td>
</tr>
<tr>
<td>3</td>
<td>Fibrose de la pyramide basale du lobe inférieur gauche, pulmonary_fibrosis, Diagnosis reliability: 100 %, Age: 66 years old, Gender: m, Smoking: true, Quantity: 25 UPA, remarks: 1 pack/day currently, Lung: interstitial fibrosis</td>
<td></td>
<td></td>
<td>m</td>
<td>true</td>
<td>25 UPA</td>
<td>99.98%</td>
</tr>
<tr>
<td>4</td>
<td>Fibrose interstitielle, pulmonary_fibrosis, Diagnosis reliability: 75 %, Age: 68 years old, Gender: f, Bronchoscopy transbronchial: interstitial inflammation</td>
<td></td>
<td></td>
<td>f</td>
<td>true</td>
<td></td>
<td>97.96%</td>
</tr>
<tr>
<td>5</td>
<td>Fibrose pulmonaire idiopathique, pulmonary_fibrosis, Diagnosis reliability: 100 %, Age: 62 years old, Gender: f,</td>
<td></td>
<td></td>
<td>f</td>
<td>true</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Fibrosis, pulmonary_fibrosis, Diagnosis reliability: 100 %, Age: 70 years old, Gender: m, Smoking: true, Quantity: 60 UPA, quitted since 2 years</td>
<td></td>
<td></td>
<td>m</td>
<td>true</td>
<td>60 UPA</td>
<td>94.9075%</td>
</tr>
<tr>
<td>7</td>
<td>Pneumopathie interstitielle fibrosante, pulmonary_fibrosis, Diagnosis reliability: 100 %, Age: 70 years old, Gender: m, Smoking: true, Quantity: 60 UPA, Bronchoscopy transbronchial: interstitial inflammation</td>
<td></td>
<td></td>
<td>m</td>
<td>true</td>
<td>60 UPA</td>
<td>94.9075%</td>
</tr>
<tr>
<td>8</td>
<td>Fibrose pulmonaire secondaire à une sclérose systémique, pulmonary_fibrosis, Diagnosis reliability: 100 %, Age: 59 years old, Gender: f,</td>
<td></td>
<td></td>
<td>f</td>
<td>true</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(Some) Search Tasks in Medicine

- Search for “similar” patients
  - Similar age, symptoms, treatment, etc.
  - Also potentially similar radiology image features/pathology
  - Learn how these patients were treated and what the outcome was
    - Especially important for rare diseases
  - Fits well into the common practice of consulting with colleagues
MONDAY, MARCH 7, 2011

A pleasant genetic appointment!

Bertrand takes his official ring bearer training VERY seriously!

Bertrand’s “good news” roll continues! At today’s genetic/metabolic follow-up, Bertrand was found to be “improved”, much to the puzzlement and pleasure of his geneticist.

WEIGHT CHALLENGE
05/16/11 - 36lbs. 10oz.
04/11/11 - 37lbs. 10oz.
03/14/11 - 38lbs. 10oz.
03/07/11 - 39lbs. 10oz.

STANDER CHALLENGE
Since April - 1-2 hrs. daily!
04/03/11 - 53 min.
03/29/11 - 0 min.
03/28/11 - 23 min.
03/27/11 - 105 min.
03/26/11 - 0 min.
03/25/11 - 36 min.

ABOUT BERTRAND
Born in December 2007, Bertrand is a charming, serious, young man. He lives in Salt Lake City, UT and has global developmental delays (0-6 months-old), leukodystrophy, intractable multifocal epilepsy (Doose Syndrome), peripheral neuropathy, liver fibrosis, gastroesophageal reflux.
(More) Search Tasks in Medicine

- Linking patient records to
  - Guidelines/clinical pathways
  - Literature
- Search within medical records
  - Many medical records are chronological collections of documents
  - It is difficult to find the pertinent information
Challenges

- Personal health records must be anonymised before being displayed as search results.
- The targeted end users potentially don’t want the solution!
More challenges

- Data quality and provenance
  - Health care providers are not used to entering data that will have secondary use
- The time aspect is of central importance
  - Especially time relative to events – e.g. how long has a person been on a drug
- Information should be delivered at the point of care
Khresmoi

Queries Questions

Information Answers

Semantic Data

Language Resources

Images

Books

Websites

Journals
Khresmoi

Users
- General Public
- Medical Doctors
- Radiologists

Language Resources

Queries Questions

Information Answers

Images

Books

Websites

Semantic Data

Journals
Technologies being developed

- Semantic annotation and search
- Medical image search
  - Including high-dimensional image search
- Cross-lingual search and machine translation
- Health knowledge bases

- User tests to validate technologies
Conclusions

- Hospitals have huge archives of patient data that could be put to effective secondary use
  - Stored in many formats
  - More free text than structured data
  - Multimodal information
  - Not hyper-linked
- Putting content and context together
- Privacy preservation is the biggest challenge