Semantic Search for Personalized Medicine

Stefan Schulz
Medical University of Graz
• **Personalised medicine:**
  - advances in genome sequencing and molecular characterisation of patients
  - medical practices, medical interventions, disease management, prevention, medical products tailored to the individual patient
  - based on their predicted response to therapy or risk of disease

• **Major driver of biomarker research**
  - Focus on "classical" biomarkers like molecular biomarkers, imaging biomarkers etc.: well-defined
  - Not to forget: biomarkers in a broader sense
Biomarkers in very broad sense

- **Features that are useful for patient profiling across a patient's medical history – together with classical biomarkers**
  - Family history
  - Past diseases
  - Therapies, complications
  - Past response to drugs
  - Clinical trajectory patterns
  - Patterns of lab values or other clinical observables
  - Correlations between different features

- **Data!**

What do they have in common?
St.p. TE eines exulc. sek.knot.SSM (C43.5) li US dors. 5/11 Level IV, 2,41mm Tumordurchm. Sentinzel LK ing. li. tumorfr.
<table>
<thead>
<tr>
<th>Ambiguities</th>
<th>BSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>bovine serum albumin</td>
<td>benzene sulphonic acid</td>
</tr>
<tr>
<td>biomolecular sequence analysis</td>
<td>behavior setting assessment</td>
</tr>
<tr>
<td>bismuth sulfite agar</td>
<td>body surface area</td>
</tr>
<tr>
<td></td>
<td>background spike activity</td>
</tr>
</tbody>
</table>
### Synonyms, spelling variants

<table>
<thead>
<tr>
<th>Colon-Ca</th>
<th>Kolon-Ca</th>
<th>Kolon-karzinom</th>
<th>Colon-carzinom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon-Karzinom</td>
<td>Kolonkrebs</td>
<td>Dickdarm-krebs</td>
<td>Dickdarm-Ca</td>
</tr>
<tr>
<td>Malignom des Kolon</td>
<td>Dickdarm-karzinom</td>
<td>Dickdarm-Ca</td>
<td>Bösartige Neubildung am Dickdarm</td>
</tr>
<tr>
<td>Bösartiger Dickdarm-tumor</td>
<td>maligne Neoplasie des Dickdarms</td>
<td>Karzinom des Dickdarms</td>
<td>malignes NPL des Colon</td>
</tr>
</tbody>
</table>
Misspellings

Atorvastatin Atorvastartin Atarvastatin Atavastatin Atorvastin Atorvaststin Atrovastation

Cal-D-Vita CalDVita Caldvita Cal-de-Vita Caldevita CalDeVita KalDvita Kal-D-Vita Kal-D-vita

Euthyrox Euhyrox Ethyrox Eutrhyrox Eutyrox Euythrox Euthyrepx Euthyox Euthyox Euthryox Euthrox Euthorox Eurthryrox

Finasterid Fenasterid Finastared Finastarid Finasakrid Finastrid Finesterid Fenastred Finestret Finsaterid Finasterid Fianasterid

Lisinopril Lisinipril Lisimopril Lisinoprol Lisinoril Lisionopril Lisionpril Lisniopril Lisonopril

Pantoprazol Pantopazol Pantobrazol Pantopranzol Panzoprazol Pantazol Pantbrazol Bantoprazol

Simvastatin Simvastat Sinvastatin Simvastastin Simvastain Simvastad Simbastatin Simavstatin Simavastatin Simastatin Symvastatin Simvastation Simvaststin Simvatatin Simvatin Simvatstain Simvstatin
From language to structure

<table>
<thead>
<tr>
<th>Code (e.g. SNOMED CT)</th>
<th>Value</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>2547300000</td>
<td>Superficial spreading malignant melanoma of skin</td>
<td>392521001</td>
</tr>
<tr>
<td>301889008</td>
<td>Excision of malignant skin tumor</td>
<td>392521001</td>
</tr>
<tr>
<td>47224004</td>
<td>Skin of posterior surface of lower leg7771000</td>
<td>Left</td>
</tr>
<tr>
<td>81827009</td>
<td>Diameter</td>
<td>2.41</td>
</tr>
<tr>
<td>258673006</td>
<td>millimeter</td>
<td></td>
</tr>
<tr>
<td>258403002</td>
<td>Lymph node level IV</td>
<td></td>
</tr>
<tr>
<td>94339008</td>
<td>Secondary malignant neoplasm of inguinal lymph nodes</td>
<td>15240007</td>
</tr>
<tr>
<td>2667000</td>
<td>Absent</td>
<td></td>
</tr>
</tbody>
</table>
Semantic Search in Cbmed
Semantic Search in Cbmed

Project 1.2: Innovative use of Information for Clinical Care and Biomarker Research

Semantic Data Repository

open MEDOCS

Context-sensitive Semantic Annotation

De-identification

Quick View

Coding

Profiling / Prediction

Recruiting

CBmed Research Use Case

Pseudonymized Access & Reidentification Service

Human Language Resources

BioBank Graz

Medizinische Universitats Graz

SAP
Semantic Search in Cbmed

Project 1.2: Innovative use of Information for Clinical Care and Biomarker Research

Text mining / Computational linguistics / Semantic Sciences / Human language technologies: tools and resources
Human language resources

- **Terminologies / Lexicons**
  - Domain terms and their relations, e.g. synonyms: "SSMM" = "Superficially spreading malignant melanoma"

- **Ontologies**
  - Axiomatic descriptions of entities of a domain, e.g. lower leg part of lower extremity

- **Grammar / Rules**
  - e.g. which word endings are indicative for a certain content, e.g. "-itis" for inflammation, "-ectomy" for surgical removal

- **Corpora**
  - Clinical texts, scientific publications, the Web
    - Annotated: e.g. words by grammatical categories, text passages by context (e.g. diagnosis, family history, lab, medication)
    - Not annotated: which words are collocated, which ones are frequent, which not used for training probabilistic models

English language well supported by resources, German language isn't
Simple example for semantic search
Conclusion

- Unstructured **textual information** in electronic health records constitutes a "**gold mine**" that may provide important information for personalized medicine and biomarker research via **semantic search**

- The mining of appropriate content requires sophisticated tools and resources

- Resources are scarce for languages other than English and need to be built or enhanced

- CBmed addresses this challenge in Project 1.2 "**Innovative use of Information for Clinical Care and Biomarker Research**"
Interested in more information?

• Contact me:
  Stefan Schulz
  Institute for Medical Informatics, Statistics and Documentation
  stefan.schulz@medunigraz.at

• Visit our poster:
  Schulz, Kreuzthaler, Huppertz, Sargsyan, Leodolter, Pedevilla, Becker, Ackbar, Fasching, Pieber:
  "Innovative use of Information for Clinical Care and Biomarker Research"
# Terminology resources

<table>
<thead>
<tr>
<th>Sup spr mel</th>
<th>Superficial spreading melanoma (morphologic abnormality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSM - Sup ficial spreading melanoma</td>
<td>Superficial spreading melanoma (morphologic abnormality)</td>
</tr>
<tr>
<td>SSMM - Superficial spreading melanoma</td>
<td>Superficial spreading melanoma of skin (disorder)</td>
</tr>
<tr>
<td>Superficial spreading melanoma of skin</td>
<td>Superficial spreading malignant melanoma of skin (disorder)</td>
</tr>
<tr>
<td>SMM - Superficial spreading melanoma of skin</td>
<td>Superficial spreading malignant melanoma of skin (disorder)</td>
</tr>
</tbody>
</table>

## Parents

- Malignant melanoma of skin (disorder)
- Primary malignant neoplasm of skin (disorder)

## Superficial spreading malignant melanoma of skin (disorder)

SCTID: 254730000

- 254730000 | Superficial spreading malignant melanoma of skin (disorder) |
  - Superficial spreading malignant melanoma of skin
  - Superficial spreading melanoma of skin
    - SSMM - Superficial spreading malignant melanoma of skin
    - SMM - Superficial spreading malignant melanoma of skin

Finding site → Skin structure
Associated morphology → Superficial spreading melanoma
Internal R&D Projects

Area 1: Data & Technologies
- 1.1 Clinical information system
- 1.2 Semantic data management
- 1.3 Knowledge discovery
- 1.4 Next generation sequencing
- 1.5 Metabolomics
- 1.6 Immunology
- 1.7 Health technology assessment
- 1.8 Digital Pathology
- 1.9 Clinical MALDI applications

Area 2: Cancer
- 2.1 CTCs for disease monitoring
- 2.2 Tracking the trace
- 2.3 Blood-based cancer diagnostics
- 2.4 Minimal residual disease and CAR T-cells
- 2.5 Eukaryotic initiation factors
- 2.7 ADX models

Area 3: Metabolism & Inflammation
- 3.1 Diabesity
- 3.2 Cardiovascular disease
- 3.3 Biosensors
- 3.4 Bone metabolism
- 3.5 Fertility
- 3.6 Liver function
- 3.7 Inflammation and fibrosis
- 3.8 NAFLD
- 3.9 Microbiome-gut-brain
- 3.10 Sepsis
- 3.11 Fungal infections
- 3.12 Electrochemical biomarker detection