Semantic Search for Personalized Medicine

Stefan Schulz
Medical University of Graz
Personalized Medicine

• 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. Sentinnel LK ing. li. tumorfr.
<table>
<thead>
<tr>
<th>Ambiguities</th>
<th>BSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>bovine serum albumin</td>
<td>benzene sulphonylic acid</td>
</tr>
<tr>
<td>biomolecular sequence analysis</td>
<td>behavior setting assessment</td>
</tr>
<tr>
<td>bismuth sulfite agar</td>
<td>bell-shaped antigen</td>
</tr>
<tr>
<td>body surface area</td>
<td>background spike activity</td>
</tr>
<tr>
<td>Synonyms, spelling variants</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>Colon-Ca</td>
<td>Kolon-Ca</td>
</tr>
<tr>
<td>Colon-\textit{Karzinom}</td>
<td>Kolonkrebs</td>
</tr>
<tr>
<td>\textit{Malignom des Kolon}</td>
<td>Dickdarm-\textit{karzinom}</td>
</tr>
<tr>
<td>\textit{Bösartiger Dickdarm-tumor}</td>
<td>\textit{maligne Neoplasie des Dickdarms}</td>
</tr>
</tbody>
</table>
### Misspellings

<table>
<thead>
<tr>
<th>Drug Name</th>
<th>Misspelled Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atorvastatin</td>
<td>Atorvastartin, Atarvastatin, Atavastatin, Atorvastin, Atorvaststin, Atrovastation</td>
</tr>
<tr>
<td>Cal-D-Vita</td>
<td>CalDVita, Caldvita, Cal-de-Vita, Caldevita, CalDeVita, KalDvita, Kal-D-Vita, Kal-D-vita</td>
</tr>
<tr>
<td>Euthyrox</td>
<td>Euhyrox, Eythyrox, Eutrhyrox, Eutyrox, Euythrox, Euthyreox, Euthyox, Euthyox, Euthrox, Euthorox, Eurthyrox</td>
</tr>
<tr>
<td>Finasterid</td>
<td>Fenasterid, Finastared, Finastarid, Finasakrid, Finastrid, Finesterid, Fenastred, Finestret, Finsaterid, Finasteril, Fianasterid</td>
</tr>
<tr>
<td>Lisinopril</td>
<td>Lisinipril, Lisimopril, Lisinoprol, Lisinoril, Lisionopril, Lisionpril, Lisniopril, Lisonopril</td>
</tr>
<tr>
<td>Pantoprazol</td>
<td>Pantopazol, Pantobrazol, Pantopranzol, Panzoprazol, Pantazol, Pantbrazol, Bantoprazol</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>Simvastat, Sinvastatin, Simvastasin, Simvastain, Simvastad</td>
</tr>
<tr>
<td></td>
<td>Simbastatin, Simavstatin, Simavastatin, Simvastatin, Symvastatin, Simvastation</td>
</tr>
<tr>
<td></td>
<td>Simvaststin, Simvatatin, Simvatin, Simvatstain, Simvstatin</td>
</tr>
<tr>
<td>Code (e.g. SNOMED CT)</td>
<td>Value</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>2547300000</td>
<td>Superficial spreading malignant melanoma of skin</td>
</tr>
<tr>
<td>301889008</td>
<td>Excision of malignant skin tumor</td>
</tr>
<tr>
<td>47224004</td>
<td>Skin of posterior surface of lower leg</td>
</tr>
<tr>
<td>81827009</td>
<td>Diameter</td>
</tr>
<tr>
<td>258403002</td>
<td>Lymph node level IV</td>
</tr>
<tr>
<td>94339008</td>
<td>Secondary malignant neoplasm of inguinal lymph nodes</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

St. p. TE eines exulc. sek.knot.SSM (C43.5) li US dors. 5/11 Level IV, 2,41mm Tumordurchm. Sentin nel LK ing. li. tumorfr.
Semantic Search in Cbmed

Project 1.2: Innovative use of Information for Clinical Care and Biomarker Research
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
In document:  "Stressulcus im Corpusbereich"

Query:  "Magenuulkus"

**ANNOTATIONS**

"Stressulkus im Corpus"

"Ulkus im Corpus"

\[ ? \]

**ONTOLOGY**

Ulkus

Stressulkus

Druckulkus (…)

**LEXICON**

"Corpusbereich" = syn

"Corpus"

"Stressulcus" = syn

"Stressulkus"

"Corpus luteum"

"Corpus callosum"

"Corpus uteri"

"Corpus ventriculi"

part of ovary ("Ovar")

part of brain ("Gehirn")

part of uterus ("Uterus")

part of stomach ("Magen")

**REFERENCE CORPUS**

sowohl einem duodenalen Anpart of 14 cm Länge. 2 cm aboral des Pylorus zeigt die Dünndarmwandung eine sanduhrartige Stenose. Im Magen- und Duodenallumen reichlich zähflüssiger Schleim, sanguinolent; die Schleimhaut ist insgesamt livide. Auf lamellierenden Schnitten zähfestes weißliches, teilweise nodulär konfiguriertes Gewebe, ohne das Gallengänge manifest werden
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

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

Superficial spreading malignant melanoma of skin (disorder)
SCTID: 254730000

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