



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

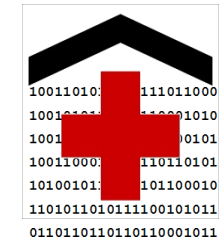
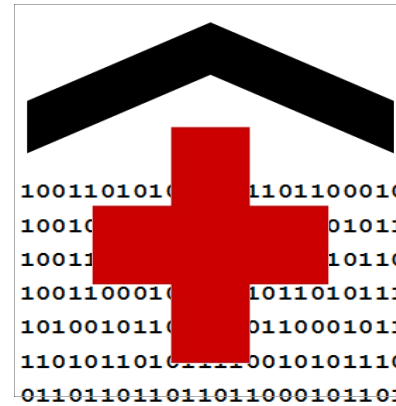
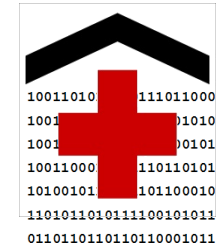
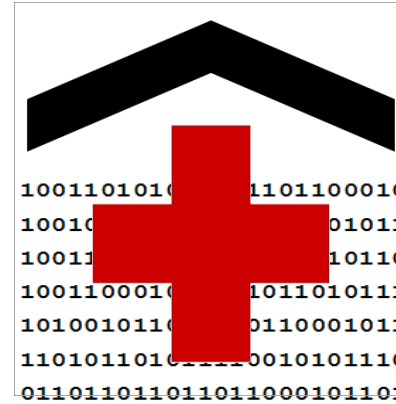
- **Features that are useful for patient profiling across a patient's medical history – together with classical biomarkers**
 - Family history
 - Past diseases
 - Therapies, complications
 - Past response to drugs
 - Clinical trajectory patterns
 - Patterns of lab values or other clinical observables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

BSA

bovine serum albumin	benzene sulphonic acid	behavior setting assessment	bell-shaped antigen
biomolecular sequence analysis	bismuth sulfite agar	body surface area	background spike activity

Synonyms, spelling variants

Colon-Ca	Kolon-Ca	Kolon- karzinom	Colon- carcinom
Colon- Karzinom	Kolonkrebs	Dickdarm- krebs	Dickdarm-Ca
Malignom des Kolon	Dickdarm- karzinom	Dickdarm-Ca	Bösartige Neubildung am Dickdarm
Bösartiger Dickdarm- tumor	maligne Neoplasie des Dickdarms	Karzinom des Dickdarms	malignes NPL des Colon

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 Eythyrox Eutrhyrox Eutyhrox Eutyrox Euythrox Euthyrex
Euthyox Euthyorx Euthryrox Euthryox Euthrox Euthorox Eurthyrox**

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

**Lisinopril Lisinipril Lisimopril Lisinoprol Lisinoril Lisionopril Lisionpril Lisniopril
Lisonopril**

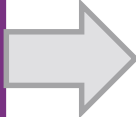
**Pantoprazol Pantopazol Pantobrazol Pantopranozol Panzoprazol Pantazol
Pantbrazol Bantoprazol**

**Simvastatin Simvastat Sinvastatin Simvastastin Simvastain Simvastad
Simbastatin Simavstatin Simavastatin Simastatin Symvastatin Simvastation
Simvaststin Simvatatin Simvatin Simvatstain Simvstatin**

From language to structure

Code (e.g. SNOMED CT)	Value	Context
254730000 Superficial spreading malignant melanoma of skin		392521001 History of
301889008 Excision of malignant skin tumor		392521001 History of
47224004 Skin of posterior surface of lower leg 7771000 Left		
81827009 Diameter 258673006 millimeter	2.41	
258403002 Lymph node level IV		
94339008 Secondary malignant neoplasm of inguinal lymph nodes		15240007 Current 2667000 Absent

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

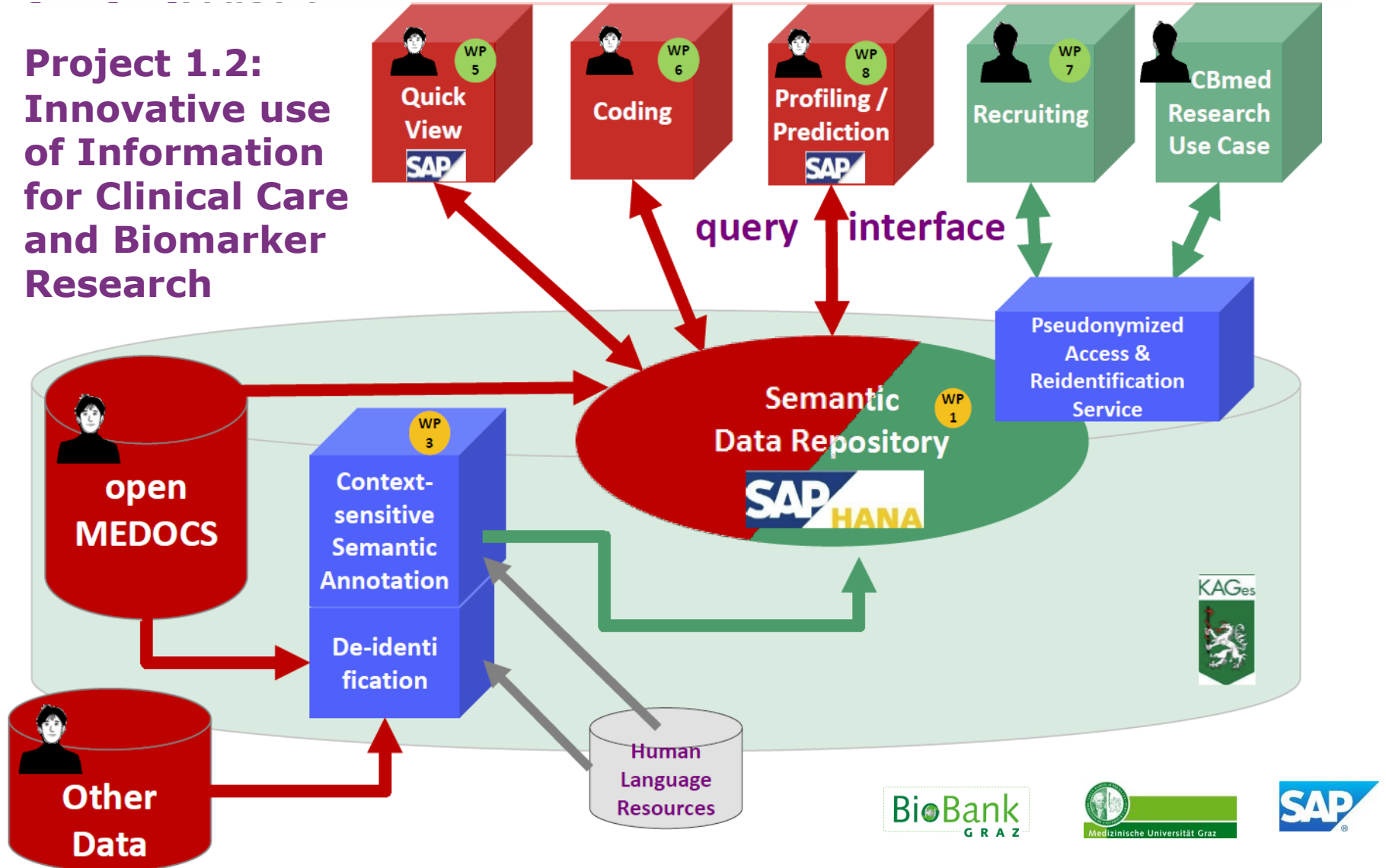


Semantic Search in Cbmed



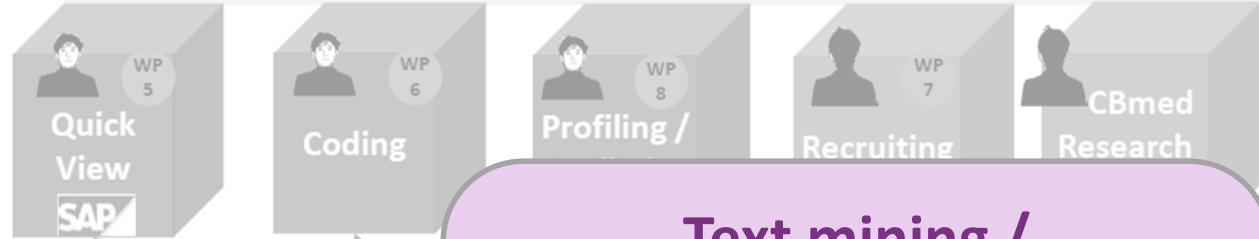
Semantic Search in Cbmed

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

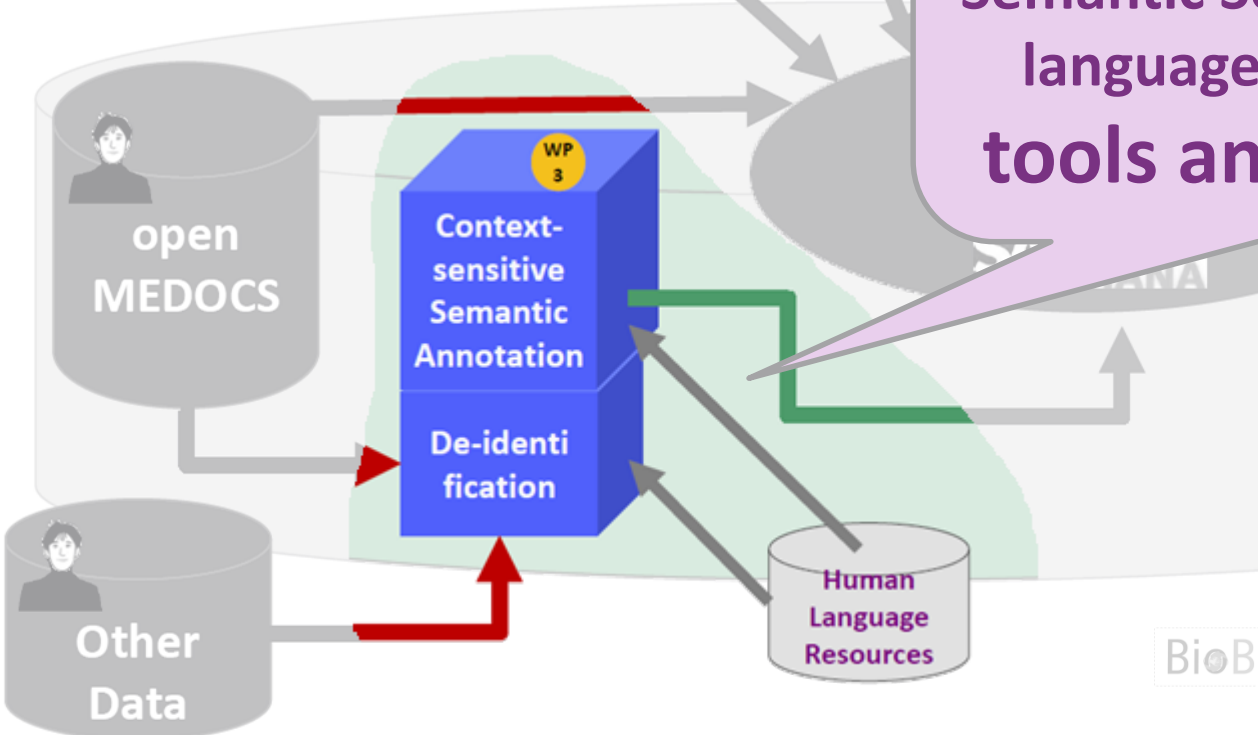


Semantic Search in Cbmed

**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**



- **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

Simple example for semantic search

In document: "Stressulcus im Corpusbereich" Query: "Magenulcus"

ANNOTATIONS

"Stressulcus im Corpus"

"Ulcus im Corpus"

"Ovarialulcus"

"Gehirnulcus"

"Uterusulcus"

"Magenulcus"

~~"Ovarialulcus"~~

~~"Gehirnulcus"~~

~~"Uterusulcus"~~

"Magenulcus"

"Corpusbereich" =_{syn}

"Corpus"

"Stressulcus" =_{syn}

"Stressulcus"

LEXICON

"Corpus luteum"

"Corpus callosum"

"Corpus uteri"

"Corpus ventriculi"

ONTOLOGY

└ Ulcus

└ Stressulcus

└ Druckulcus

(...)

part of ovary ("Ovar")

part of brain ("Gehirn")

part of uterus ("Uterus")

part of stomach ("Magen")

soweit einem duodenalen Anpart von 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

REFERENCE CORPUS

Conclusion

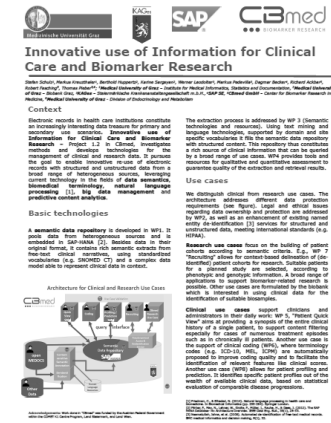
- Unstructured **textual information** in electronic health records constitutes a "**gold mine**" that may provide important information for personalized medicine and biomarker research via **semantic search**
- The mining of appropriate content requires sophisticated tools and resources
- Resources 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"**



Innovative use of Information for Clinical Care and Biomarker Research

Context

Electronic records in health care institutions contain an increasingly increasing data base for primary and secondary care activities. Increasing use of methods and devices technologies, for the management of clinical and research data, is a source of information. However, the data are often stored in unstructured and uncontrolled data from a broad range of heterogeneous sources. Integrated data handling in the form of data integration, metadata management, natural language processing, (1) big data management and predictive content analysis.

Basic technologies

A semantic data repository is developed in W3C standards. In this repository, records are stored in original format, a central rich semantic extracts from these records, clinical data are standardized according to the SNOMED CT and a complete data model able to represent clinical data in context.

Architecture for Clinical and Research Use Cases

Clinical use cases support clinicians and researchers in their work with the "real world" history of a single patient, to support content filtering especially for cases of particular importance with the support of clinical coding (SNOMED), where terminology codes (e.g. ICD-10, ICD-9, ICD-9-CM) are automatically extracted and used to facilitate the distribution of new data to the clinical system. Researchers can identify those for post-procedure analysis to identify specific patient profiles out of the records of multiple clinical data, based on statistical analysis.

Use cases

We developed clinical form research use cases. The architecture addresses different data production requirements (see Figure). Legal and ethical issues regarding data availability and protection are addressed by W3C, as well as an enhancement of existing record entry standards (1) across the structured and unstructured data, meeting international standards (e.g. HL7).

Research use cases focus on the building of patient profiles, according to semantic, clinical, and/or laboratory patient records for research. Suitable profiles for a clinical study can be selected, according to predefined specific information. Similar data applications to support biomarker-related research is possible. The data are also used to identify the records which is identified as using clinical data for the identification of suitable biomarkers.







Terminology resources

sup spr mel










Stated

10 matches found in 0.255 seconds.

 Superficial spreading melanoma	Superficial spreading melanoma (morphologic abnormality)
 SSM - Superficial spreading melanoma	Superficial spreading melanoma (morphologic abnormality)
 SSMM - Superficial spreading melanoma	Superficial spreading melanoma (morphologic abnormality)
 Superficial spreading melanoma of skin	Superficial spreading malignant melanoma of skin (disorder)
 SMM - Superficial spreading melanoma of skin	Superficial spreading malignant melanoma of skin (disorder)
 Superficial spreading malignant melanoma of skin	Superficial spreading malignant melanoma of skin (disorder)

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 melanoma of skin
- 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 Diabetesity
- 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