



Medical University of Graz

Semantic Retrieval and Navigation in Clinical Document Collections

Markus KREUZTHALER^a, Philipp DAUMKE^b and Stefan SCHULZ^a

^aInstitute for Medical Informatics, Statistics and Documentation,
Medical University of Graz, Austria

^bAverbis GmbH, Freiburg, Germany

in cooperation with

averbis
text analytics





Introduction

- ▶▶ Especially patients with **chronic diseases** and **numerous hospital stays**, **accumulate large quantities of documents** in the electronic patient record.
- ▶▶ An important document type is the doctor's discharge letter. It acts as a case based summary of clinical information and carrier of **textual** information:

Allergies Medications
Findings Temporal contexts
Symptoms Medical history
Procedures



Introduction

►► Specific peculiarities of medical language

- Ambiguous terms, acronyms, compounds, derivations, spelling variants, uncorrected spelling errors, jargon, telegram style, non-standardized numeric expressions

3. St.p. TE eines exulz. sek.knot.SSM (C43.5)
li Lab. majus. Level IV, 2,42 mm Tumordurchm.

- Context depended interpretation e.g: family history, prognosis and treatment planning



Project Aim

Automated semantic indexing of medical reports in order to increase efficient navigation (pilot study, demonstrator)

- ▶▶ Web based semantic medical report navigator
 - Support efficient navigation and search within **one patient**
 - Take into account **semantic relations** and **lexical variants**
 - **Recall oriented** parameterization of NLP-Pipeline



Related Work

Natarajan, K., Stein, D., Jain, S. & Elhadad, N. **An analysis of clinical queries in an electronic health record search utility**. International Journal of Medical Informatics 79, 515-522 (2010).

►► Introduction

- **No in-depth analysis** has been performed to understand **clinicians specific information needs** in the context of search.

►► Methods

- Columbia University, New York, NY, United States
- WebCIS: Web based clinical information system to all
 - clinical narratives
 - laboratory test results
- Integration CISearch (Lucene based simple keyword search)
 - discharge summaries, radiology reports, pathology reports

Related Work

►► Results

- Based on transaction log analysis (TLC) of 436 users (research, doctor, clinical, nurse, admin, other) within 6 months on 980 unique search queries

Table 5 – Top 5 semantic types of searches.

Semantic type	Percentage (N = 161)
Laboratory or test result	29.2%
Disease or syndrome	21.7%
Body part, organ, or organ component	8.1%
Pharmacologic substance	7.5%
Diagnostic procedure	6.2%



Data

- ▶▶ Department of Dermatology and Venerology
 - Chronic diseases: Sklerodermie, Lupus Erythematodes, Neurodermitis, Erysipel, Herpes Zoster, *Psoriasis*, **malignant Melanom**

- ▶▶ Information need for ETL-workflow:
 - Return all discharge summaries until May 2013 where the initial diagnosis of malignant melanoma was made in 2007 at the Department for Dermatology and Venerology

- ▶▶ Information need result:
 - Data set comprises 1696 pseudonymized documents of 175 patients
 - (max 82; min 1; avg ~10) documents per patient
 - Data set is exported in accordance to the *SOLR xml document specification*

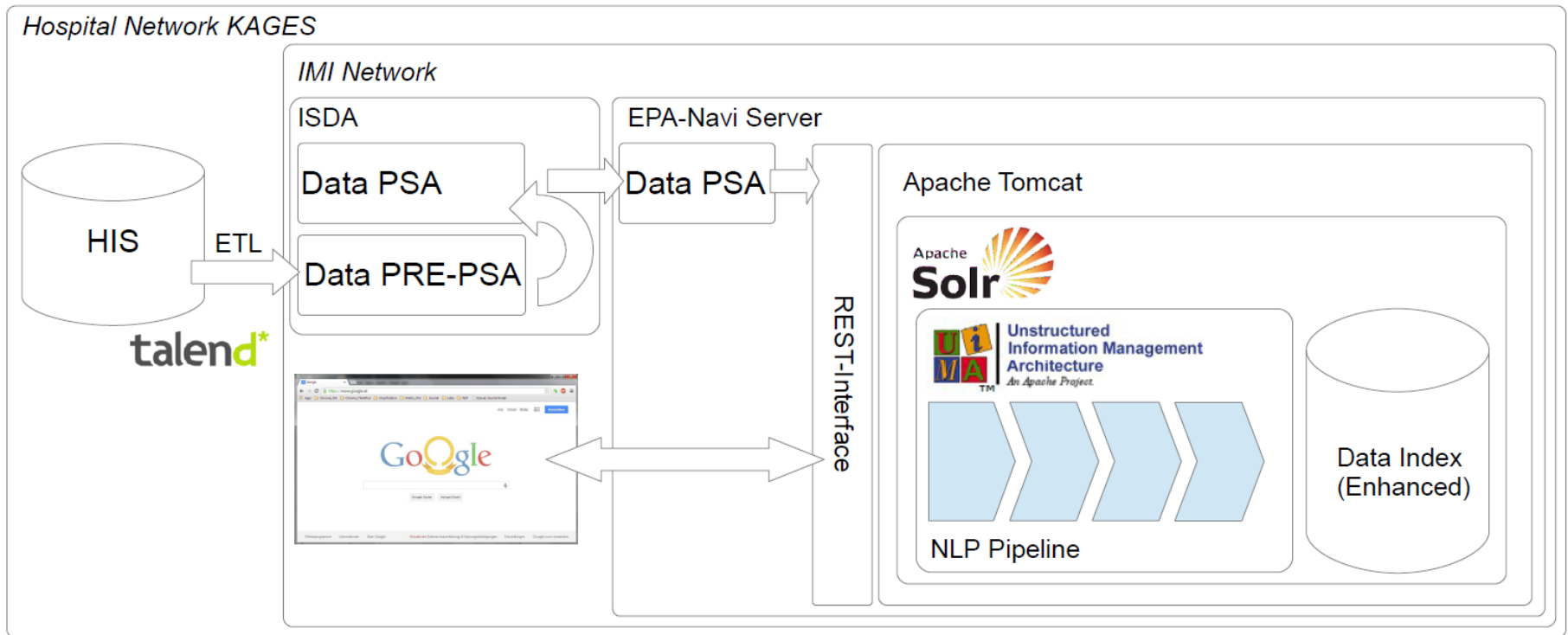
SOLR XML Document Example

```
<doc>
  <field name="AB_ART"><![CDATA[Befund - Tumorambulanz]]></field>
  <field name="AB_DATUM"><![CDATA[2000 ]]></field>
  <field name="Beurteilung"><![CDATA[Empfehlung des Tumorboards vom 2000:
Durchführung eines CT mit Kontrastmittel, evtl.
Wiederholung des PET-CTs. Wiedervorstellung in der Tumorkonferenz.]]></field>
  <field name="Diagnose"><![CDATA[Malignes Melanom Stadium IV
1997 : Primärtumor: Melanom re. Schulter (Level IV, Tumordicke 3,6
mm)
1997 : Sentinel node-Biopsie axillär re.: 1,3 cm große Metastase
eines malignen Melanoms.
1997 : Axilläre Lymphadenektomie re.: 1 LK-Metastase, 16 LK tumorfrei
1999 : Verd. auf kleine intrapulmonale Fil.
1999 : Pankreasmastasen; Nebennierenmetastasen re.; Verd. auf
Lymphknotenmetastase kaudal dem unteren Pol der re. Niere.
* V.a. Metastasen der Nieren und Nebenniere rechts
2000 : Progredienz

Bisherige Therapie:
Homöopathie und Bioresonanztherapie]]></field>
```

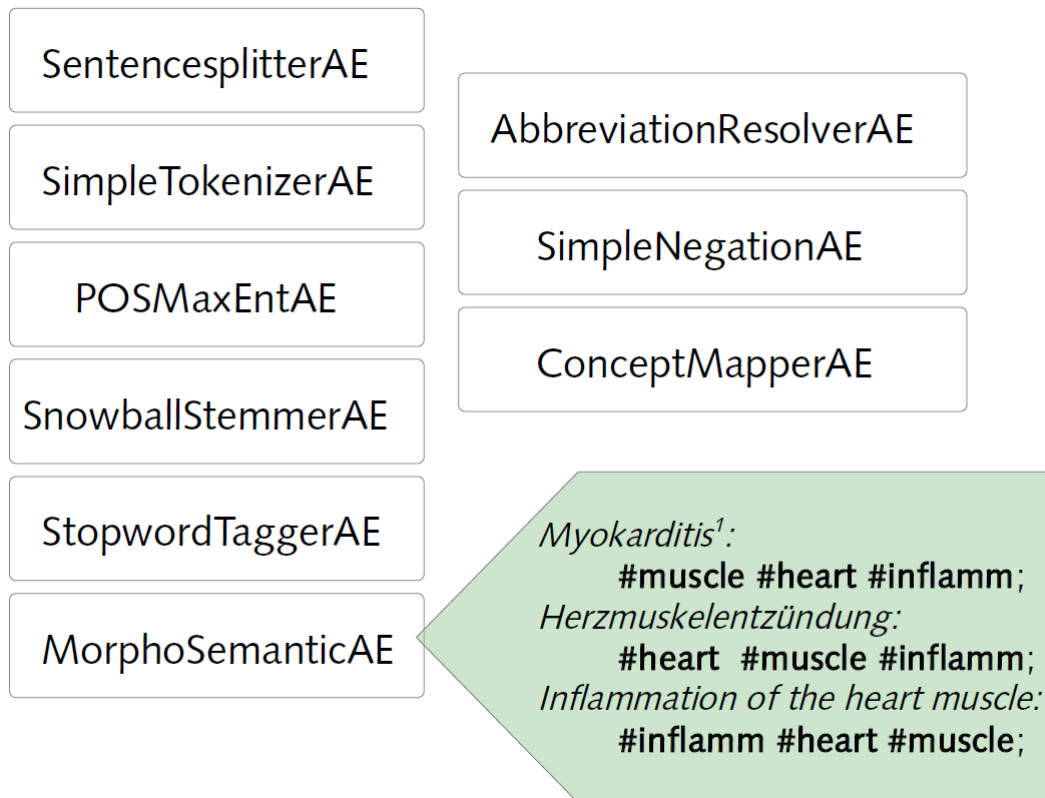

Methods

- Core frameworks in use: Apache UIMA and Apache Solr



Methods

►► NLP pipeline main components:



¹K. Markó, S. Schulz, U. Hahn: MorphoSaurus - Design and Evaluation of an Interlingua-based, Cross-language Document Retrieval Engine for the Medical Domain. *Methods of Information in Medicine* 4/2005(44): 537-545

Methods

►► Semantic type and concept mapping

Semantic Type	Percentage² (N=161)	Content Based Facets	Terminology In Use
Laboratory or test result	29.2%	Laborwerte	Averbis Lab Terminology
Disease or syndrome	21.7%	Diagnosen	ICD-10 2014
Body part, organ, or organ component	8.1%	Lokalisation	Radlex 2.0
Pharmacologic substance	7.5%	Medikationen	ABDAMED 2014
Diagnostic procedure	6.2%	Prozeduren	OPS 2014

²Natarajan, K., Stein, D., Jain, S. & Elhadad, N. An analysis of clinical queries in an electronic health record search utility. International Journal of Medical Informatics 79, 515-522 (2010).



Configuration and Annotation

Annotation Results for 40.xmi in C:\DataEPA-Navi\uima\data\output\all\xmi

Anamnese:

Patientin kommt zur Besprechung Tumorboard vom
2001: SD

Diagnose:

Malignes Melanom Stadium IV (pT3a N2a M1b AJCC 1999)
BRAF V600E negativ (33766/1997)

1997 Exzision eines Melanoms an der Schulter re (TD 3.7mm, Clark Level IV)
1997 SLNB axillär re (1+ / 1), Met.-Durchmesser 1.3 cm
1997 Axilläre Lymphadenektomie re (1+ / 17)
1999 V.a. kleine intrapulmonale Metastasen
1999 multilokuläre Metastasierung: Pancreas, Nebenniere re, LK kaudal dem unteren Pol der re. Niere)
2000 Grössenzunahme der bekannten Metastasen
2000 3 Zyklen einer Chemotherapie mit DTIC im 3-Tages-Schema;
Response: PD
2001 4 Gaben einer Immuntherapie mit Ipilimumab im Rahmen des Named Patient Program; keine NW; Response: 2001 SD (Nebenniere re regredient 1.5cm; Niere re progredient 3.7cm);
2001 SD (Status idem)
2001 SD (Status idem)

aktuell: Maintenance-Phase post Ipilimumab

Click In Text to See Annotation Detail

Annotations

Date

Date (" 1997")

- begin = 171
- end = 178
- componentId = org.apache.uima.anno
- confidence = 0.0
- value = 1997
- label = DATE
- parsedElements = null

Legend

☒ Abbreviation ☐ Conjunction ☒ Date ☐ Diagnosis ☐ DocumentAnnot...

Select All Deselect All Hide Unselected

Results - GUI

►► Core interface, main

chem|

chemotherapie

Showing 1 - 10 of 18 results

Sort by: Relevance Sort order: desc Results per page: 10

DIAGNOSEN

- Bösartiges Melan... (18)
- Chemotherapie (18)
- Chronisches Ulku... (18)
- Lymphknotenvergr... (18)
- Sekundäre bösart... (18)
- Medizinische Beh... (17)
- Situs inversus (16)
- more...

PROZEDUREN

MEDIKATIONEN

- Carboplatin (14)
- Pantoloc (11)
- Lescol (11)
- Seloken (11)
- Actrapid (7)
- Ramipril (6)
- Ramipril und Diu... (6)
- more...

LABORWERTE

LOKALISATION

CURRENT SEARCH RESTRICTIONS

PATIENTENNAME: [\[redacted\]](#) × / DIAGNOSEN: Bösartiges Melanom der Haut × / MEDIKATIONEN: Carboplatin × / chemotherapie ×

1 2 »

[\[redacted\]](#) 2008

...Zyklen einer **Chemotherapie** mit **Carboplatin** 2007 Z. n. 3 Zyklen einer **Chemotherapie** mit DTIC (C43.3 ...

[\[redacted\]](#) 2007

...erfolgte zur 2. **Chemotherapie** mit **Carboplatin**. Am 2007 wurde 500mg **Carboplatin** unter entsprechender ...

[\[redacted\]](#) 2008

...Leistenkanal beidseits. Diagnose: Laufende **Polychemotherapie** bei **malignem Melanom** der linken Schulterregion operiert ...

[\[redacted\]](#) 2008

...medial 07 St.p. 2 Zyklen einer **Chemotherapie** mit **Carboplatin** 2007 (Abbruch wegen Befundprogredienz ...

PATIENTENNAME

[\[redacted\]](#)

ARZTBRIEF DATUM

- 2007 (1)
- 2007 (1)
- 2008 (1)
- 2008 (1)
- 2008 (1)
- 2008 (1)
- 2008 (1)
- more...

Results - GUI

►► Core interface, content based facets

chem|

chemotherapie

Showing 1 - 10 of 18 results

Sort by: Relevance ▼

▼ **DIAGNOSEN**

- Bösartiges Melan... (18)
- Chemotherapie (18)
- Chronisches Ulku... (18)
- Lymphknotenvergr... (18)
- Sekundäre bösart... (18)
- Medizinische Beh... (17)
- Situs inversus (16)
- more...

► **PROZEDUREN**

▼ **MEDIKATIONEN**

- Carboplatin (14)
- Pantoloc (11)
- Lescol (11)
- Seloken (11)
- Actrapid (7)
- Ramipril (6)
- Ramipril und Diu... (6)
- more...

► **LABORWERTE**

► **LOKALISATION**

CURRENT SEARCH RESTRICTIONS

PATIENTENNAME: [Redacted] × / DIAGNOSEN: Bösartiges Melanom der Haut × / MEDIKATIONEN: Carboplatin × / chemotherapie ×

1 2 »

[Redacted] 2008

...Zyklen einer **Chemotherapie** mit **Carboplatin** 2007 Z. n. 3 Zyklen einer **Chemotherapie** mit DTIC (C43.3 ...

[Redacted] 2007

...erfolgte zur 2. **Chemotherapie** mit **Carboplatin**. Am [Redacted] 2007 wurde 500mg **Carboplatin** unter entsprechender ...

[Redacted] 2008

...Leistenkanal beidseits. Diagnose: Laufende **Polychemotherapie** bei **malignem Melanom** der linken Schulterregion operiert ...

[Redacted] 2008

...medial [Redacted] 07 St.p. 2 Zyklen einer **Chemotherapie** mit **Carboplatin** 2007 (Abbruch wegen Befundprogredienz ...



Results - GUI

►► Core interface, metadata facets

therapie

Sort by: Relevance Sort order: desc Results per page: 10

ARCH RESTRICTIONS

PATIENTENNAME: [redacted] / DIAGNOSEN: Bösartiges Melanom der Haut / MEDIKATIONEN: Carboplatin / chemotherapie

1 2 »

2008

einer Chemotherapie mit Carboplatin 2007 Z. n. 3 Zyklen einer Chemotherapie mit DTIC (C43.3 ...

2007

e zur 2. Chemotherapie mit Carboplatin. Am 2007 wurde 500mg Carboplatin unter entsprechender ...

2008

kanal beidseits. Diagnose: Laufende Polychemotherapie bei malignem Melanom der linken Schulterregion operiert ...

2008

07 St.p. 2 Zyklen einer Chemotherapie mit Carboplatin 2007 (Abbruch wegen Befundprogredienz ...

ARZTBRIEF DATUM

- 2007 (1)
- 2007 (1)
- 2008 (1)
- 2008 (1)
- 2008 (1)
- 2008 (1)
- 2008 (1)
- more...

Results - GUI

►► Core interface, details

Diagnosen:

Einleiten einer **Chemotherapie** mit Murphoran wegen Befundprogredienz
St.p TE eines ulzerierten nodulären **malignen Melanoms**(Level V, TD 1,2
cm) Schulter li. 07

St.p. Total- und Nachexzision eines **malignen Melanom** (Level IV,
Tumordicke 0.93mm) Rücken medial 07

St.p. Sentinel Node Biopsie 07

- Axilla rechts: tumorfrei
- Axilla links: metastatischer Befall

St.p. Lymphadenektomie (Metastasen) axillär links 07

Results and Discussion

- ▶ Narratives are defragmented after ETL-workflow
 - in case of non-sentence limiting line breaks '\n'.
 - in case of enumerations the line breaks should stay

E-Nr.: 00000/00 TE Schulter li: MM (oberfl. MM,
Level II; größte
vertikale TD weniger als 0,5mm, AJCC 2002: T1a). Entfernung nicht im
Gesunden. NE indiziert.

Concor 5 mg 1-0-0
Norvasc 5 mg 1-0-1
Tamsulosin 0,4 mg 0-0-1
Atarax 25 mg 0-0-1
PK Merz 1-1-0
Stalevo 150 mg 1-0-0

- readability and text format should be enhanced

Results and Discussion

►► Redundant information in-between doctoral letters of one patient

23.05.1980;23-5-1980 CT an der Kinderradiologie durchgeführt.
23.05.1980;23-5-1980 CT an der Kinderradiologie durchgeführt.
23.05.1980;23-5-1980 CT an der Kinderradiologie durchgeführt.
23.05.1980;23-5-1980 CT an der Kinderradiologie durchgeführt.
23.05.1980;23-5-1980 CT an der Kinderradiologie durchgeführt.

►► Number of different headers within doctoral letter

- 59 different observed header types in 1696 documents e.g.
 - DIAGNOSE (#1169), DIAGNOSEN (#525)
 - ausständige Befunde (#525), relevante_Befunde (#525), Dermatologischer Befund (#106), weitere Befunde (#48), Befund (#15), Laborbefunde (#13), Befunde (#3), weitere Befunde: (#1), Befund Augenklinik 23.05.1980: (#1), Auflichtmikroskopischer Befund (#1)
 - Allergien (#127), 1Allergien (#1), Allergien und Medikamente (#1)
 - (#45)

Results – Indexing Performance

►► Time – Data Size Estimation

HP Intel Core i7-3770 3.4GHz 8M HD 4000 CPU

1TB 7200 RPM 3.5 HDD

16 GB RAM

Windows 7 Professional

#Documents	Data Size (MB)	Data Size Index (MB)	Time (min)	Time (h)
1696	5.1	40.4	5.2	0.1
10000	29.9	238.2	30.7	0.5
100000	298.9	2382.1	306.6	5.1
1000000	2989.4	23820.8	3066.0	51.1



Conclusion and Outlook

►► Clinical NLP problems are addressed in scientific challenges

- SemEval 2015 Task 14: <http://alt.qcri.org/semeval2015/task14/>

"...It also comprises the task of **identifying various attributes of the disorders and normalizing their values**. We refer to this as the **template filling task**..."

"...The task aims to **combine supervised methods for text analysis with unsupervised approaches**. More specifically, the task aims to combine supervised methods for **entity/acronym/abbreviation recognition**..."

►► Precision and recall oriented information retrieval evaluation

- According to TREC 2011/2012 Medical Tracks

►► Feedback from users

- GUI
- Placement and functionality of search tools
- Retrieval performance



Conclusion and Outlook

- ▶▶ Parametrization of toolset versus use of tool set
- ▶▶ Visualized patient profile
 - Still a lot of textual information and list based views
 - Visualize and weight according to different axes e.g.:
 - Semantic categories (diagnoses, procedures, medications, other)
 - Frequency of occurrence - unexpected occurrence
 - Time interval
 - Clinical relevance
 - Enhance fast visual perception of a patient case (information visualization) with optional fast navigation
- ▶▶ ELGA
 - Enhanced use case of patient based navigation