

**CBmed GmbH**  
**Center for Biomarker Research in Medicine**  
**Programme: COMET – Competence Centers for Excellent Technologies**  
**Programme line: K1-Centers**  
**COMET sub-project 1.2, duration and type of project:**  
**Innovative Use of Information for Clinical Care and Biomarker Research (IICAB), 01/2015 – 12/2018, multi-firm**

## Sprachtechnologie erschließt Routinedaten für Biomarkerforschung

Um klinische Routinedaten für die Biomarkerforschung verfügbar zu machen, müssen aus den meist nur textuell vorliegenden Informationen strukturierte Inhalte extrahiert werden. Das Projekt 1.2 „*Innovative Use of Information for Clinical Care and Biomarker Research*“ verwendet Text Mining und maßgeschneiderte Terminologien, um den Klinikjargon zu erschließen, semantische Extrakte in ein Warehouse einzuspeisen und für Abfragen bereitzustellen.



### Freitextdokumentation in der elektronischen Patientenakte

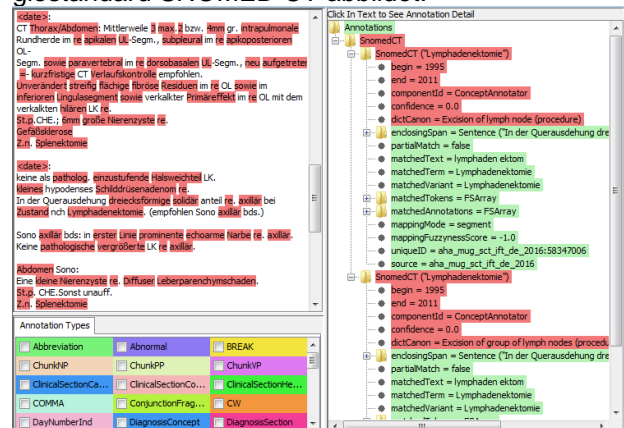
Jeder Patientenkontakt generiert eine Vielfalt von Daten, die in klinischen Informationssystemen gespeichert werden. Dies umfasst aktuelle Beschwerden, Vorerkrankungen, Krankheitsverläufe, Diagnosen, Befunde ebenso wie operative Eingriffe, Medikamente und Lifestyle-Faktoren. Ein wesentlicher Teil dieser Informationen liegen in Form von gering strukturierten Arztbriefe, Diagnoselisten und Befundberichten vor. Diese Texte weisen eine hochkompakte Fachsprache auf, die auf die Kommunikation zwischen Experten zugeschnitten ist, aber die Analyse durch Computer vor große Aufgaben stellt. Sätze wie: „*St. p. TE eines exulc. sek.knot.SSM li US dors. 5/11 Level IV 2,4 mm Tumordurchm. Sentinell LK ing. li. tumorfr.*“ sind selbst für fachfremde Mediziner schwierig zu lesen. Für die Biomarkerforschung ist es jedoch unerlässlich, relevante klinische Daten strukturiert, standardisiert und universell auswertbar vorliegen zu haben.



### Sprachtechnologien und -ressourcen

Bei CBmed arbeitet ein Team an Medizinformatikern an Softwarelösungen, welche Informationen aus klinischen Texten extrahieren, klassifizieren und in Datenbanken einspeisen.

Die 2017 entwickelte Prototypen kombinieren regelbasierte Methoden mit Verfahren des Maschinellen Lernens unter Verwendung von kommerzieller Software und Eigenentwicklungen. Eine wichtige Rolle besteht hierbei ein von der Medizinischen Universität Graz entwickeltes und gepflegtes Lexikon, welches den lokalen Medizinerjargon auf den internationalen, zukünftig auch in Österreich eingeführten Terminologiestandard SNOMED CT abbildet.



**Abb. 1: Analyse des Ausschnitts eines Arztbriefs. Hier wird das Wort „Lymphadenektomie“ erkannt und mit dem SNOMED CT-Term „Excision of groups of lymph nodes (procedure)“ annotiert**

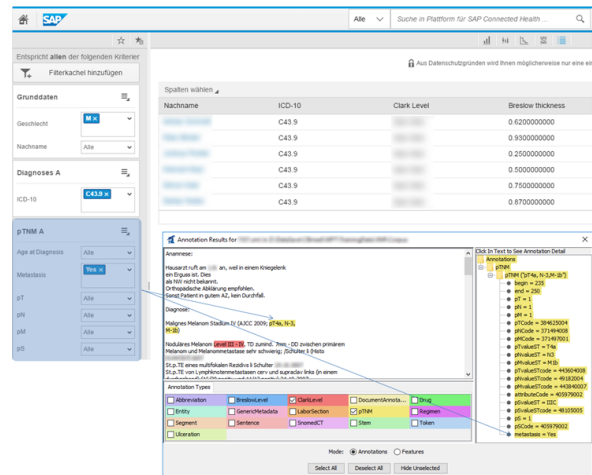


## Wirkungen und Effekte

Ziel dieser semantischen Aufarbeitung von medizinischen Routinedaten ist die Erstellung eines „strukturierten Patientenprofils“, das an internationalen Standards orientiert ist und als „Clinical Data Warehouse“ im System SAP HANA, innerhalb des Klinikinformationssystems des führenden steirischen Krankenhasträgers KAGes implementiert ist. So könnten etwa Kandidaten für klinische Studien effizient ermittelt werden oder bei Biobankrecherchen klinische Suchkriterien wie z.B. Tumorart, Schweregrad, Begleiterkrankungen, Therapien und Zeitintervalle verwendet werden.

Ein weiterer Benefit der semantischen Analyse klinischer Texte liegt in der Verbesserung der Qualität klinischer Daten. Innerhalb von weniger als zwei Jahren wurde eine komplexe und leistungsfähige Plattform zur semantischen Anreicherung klinischer Routinedaten geschaffen und an mehreren Prototypen demonstriert werden. Nutzungsszenarien, die über die Anwendung zur Biomarkerforschung hinausreichen (wie z.B. Entscheidungsunterstützung, Datenvisualisierung und Qualitätssicherung) sichern die nachhaltige Unterstützung des weiteren Aus-

baus seitens des Krankenhasträgers KAGes, der Medizinischen Universität Graz und der Softwarefirma SAP.



**Abb. 2: Anfrageoberfläche des Clinical Data Warehouse auf SAP HANA. Hier können aus Arztbriefen extrahierte und automatisch kodierte Informationen zusammen mit ihrem zeitlichen Kontext abgefragt werden.**

### Kontakt und Informationen

K1-Centre CBmed

CBmed GmbH  
Stiftingtalstrasse 5, 8010, Graz  
+43 316 385 28801  
[office@cbmed.at](mailto:office@cbmed.at)  
[www.cbmed.org](http://www.cbmed.org)

**Project coordinator**  
Univ.-Prof. Dr.med. Stefan Schulz

### Projektpartner

Organisation	Land
KAGes	Austria
SAP	Germany

**Weitere Informationen zu COMET – Competence Centers for Excellent Technologies:** [www.ffg.at/comet](http://www.ffg.at/comet)

Diese Success Story wurde von der Konsortialführung/der Zentrumsleitung zur Verfügung gestellt und zur Veröffentlichung auf der FFG-Website freigegeben. Für die Richtigkeit, Vollständigkeit und Aktualität der Inhalte übernimmt die FFG keine Haftung.