





eine bildgesteuerte Suche nach beispielsweise ähnlichen Krankheitsfällen nur dann möglich ist, wenn die entsprechende Bildsemantik (Krankheiten, Anatomien, Charakteristiken) auch an den entsprechenden Bildregionen annotiert ist und die Verknüpfung zu den Befundungstexten und entsprechenden Textpassagen hergestellt ist. Eben das, was eine "semantische Suchmaschine" von Google unterscheidet, in der nicht mit ganzen Texten und Bildunterschriften, sondern dem dahinter verborgenen Domänenwissen gearbeitet wird.

### Automatische Detektion und Segmentierung von Organen und Landmarken

Verfahren des Maschinellen Lernens bilden die Basis der Bildanalyse von MEDICO<sup>[3][1]</sup>, die als Computerprogramme die automatische Detektion von Bildinhalten liefert, die mit RadSpeech dann durch Benutzerinteraktion verifiziert und ergänzt werden können. Maschinelles Lernen als weiteres Teilgebiet der KI bedeutet hier, dass der Computer durch Präsentation zahlreicher annotierter Beispieldaten lernt, Objekte wie einen Lungenknoten, einen Nierenstein oder ein Herz im Bild wiederzuerkennen. Der Vorteil dabei ist, dass durch das maschinelle Lernen unterschiedliche Modalitäten (CT, MRT, Ultraschall) in gleicher Art und Weise verarbeitet werden können. Es ist jeweils nur ein erneutes Training, jedoch kein Programmieraufwand, notwendig. Primäres Ziel der Bildanalyse ist es, die Bilder in ihre anatomischen Grundbestandteile (Organe und Gewebe) aufzuteilen, diese semantisch zu verknüpfen und zu speichern. Das entwickelte automatische Bildanalyseverfahren arbeitet dabei hierarchisch, indem zunächst der dargestellte Bildausschnitt eruiert wird, um die Grundbestandteile zu identifizieren, z.B. Gefäßabzweigungen oder Knochenenden, gesucht wird und schließlich Organe und Gewebe, z.B. Lymphknoten, detektiert werden. Diese können dann mittels geeigneter Desktop-basierter semantischer Befundungsprogramme und neuerdings mit RadSpeech durch eigene semantische Annotationen des Arztes komplettiert und zu einem KI-gestützten Diagnose-Tool ausgebaut werden.

### Intelligente KI-gestützte Diagnose

Die KI-Maschine soll auf Basis der Detektion und Segmentierung von Organen automatisch krankhafte Veränderungen erkennen und dem Arzt per Sprachdialog eine Warnmeldung geben: "Leber wurde mit Blutgefäßen und Gewebe gescannt; es wurden krankhafte Veränderungen festgestellt.", wobei der Arzt weiterführende Fragen stellen kann und per Sprache und Zeigegesten Befundungsannotationen setzen kann. RadSpeechs Entwicklungen der letzten 3 Jahre beschäftigten sich mit diesen medizinischen Funktionalitäten, eingebettet in intelligente multimodale Interaktion<sup>[5][8]</sup>.

### Intelligente Multimodale Interaktion in der Medizin

Wie angedeutet, sind heutzutage medizinische Bilder

in der Krankheits-Diagnose und Differenzierung, Eingriffsplanung oder zur Überwachung der Behandlung jedes einzelnen Patienten unverzichtbar geworden. Um die KI-gestützte Diagnose, die unserer Arbeitshypothese entspricht, realisieren zu können, muss die nötige Bildsemantik auch im Bild annotiert werden. Jedoch, wie lässt sich dies in die radiologische Arbeit integrieren, ohne den Arbeitsablauf des Arztes zu sehr zu stören?

RadSpeech ist ein mobiler Arbeitsplatz für Radiologen. So, wie sich viele Radiologie-Spezialisten ihren Arbeitsplatz der Zukunft vorstellen könnten, wenn es darum geht, Bildregionen zu annotieren und gleichzeitig eine intuitive semantische Suche ähnlicher Fälle möglich zu machen: mit sprachlicher Interaktion im mobilen Umfeld. Per Sprachbefehl können Radiologen oder behandelnde Ärzte elektronische Krankenakten mit den dazugehörigen Bilddateien abrufen, durchsuchen und per Touchgeste auf dem iPad beliebig anordnen. Bilder aus traditionell radiologischen und tomografischen Untersuchungen können

semantisch annotiert und über die Ontologie-basierte semantische Suche wieder gefunden werden. Die Ontologie-basierte Bildsuche und Annotation bietet dabei die Grundlage zukünftiger computer-gestützter klinischer Entscheidungsfindung und Diagnose, durch die intuitive Benutzbarkeit unterstützt. RadSpeech kann den Arbeitsablauf des Arztes in Situationen genutzt werden, in denen keine feste Arbeitsstation zur Verfügung steht, z.B. bei der Visite oder in Besprechungen.

Mit herkömmlichen Benutzeroberflächen können bei der Bildsuche, bei der Interpretation von Bildern oder bei der Annotation von Bildern nur die Bilder durchsucht werden, bei der Interpretation von Bildern allerdings bieten diese Oberflächen keine weitergehende Assistenz an. In RadSpeech wird daher versucht, die nächste Generation intelligenter, skalierbarer und intuitiver Benutzerschnittstellen für die semantische Suche in medizinischen Bildverarbeitungsbereichen zu entwickeln. Ontologie-basierte Wissensrepräsentation wird dabei nicht nur für Bildinhalte genutzt, sondern auch für die komplexen Prozesse im Sprachverstehen und im Dialogmanagement. Unter Einbeziehung von qualifiziertem Fachwissen können verschiedene Sichten auf medizinische Bilder (z.B. strukturelle, funktionale und krankheitsbedingte Aspekte) explizit repräsentiert und anwendbar gemacht werden. Ein YouTube-Video erläutert die Bedienbarkeit von RadSpeech. Bei der gezeigten Interaktion stehen zwei Aspekte im Vordergrund, um eine intelligente Suche nach ähnlichen Patienten zu ermöglichen und die Diagnose zu beschleunigen: Erstens die Durchsicht von und Navigation in relevanten Patientendaten, und zweitens die Annotation radiologischer Bilder mithilfe natürlicher Sprache und Zeigegesten.

Um das Vokabular des Radiologen gut abzudecken, greift RadSpeech auf die von Experten entwickelten Ontologien zurück. Dies sind zum einen das Foundational Mo-



**Abbildung 3:**  
THESEUS Innovation Center in Berlin  
mit Prototypen von RadSpeech und  
THESEUS MEDICO Prototypen

del of Anatomy (FMA), Radlex, und ICD-10. Weiterhin werden einige Teile von Snowmed CT für klinische Daten verwendet.

Ein vollständiger Workflow könnte dann so aussehen: Ein Radiologe behandelt einen Lymphom-Krebs-Patienten, der nach erfolgter Chemotherapie zur CT-Nachfolgeuntersuchung erscheint (der behandelnde Arzt ist daran interessiert, ob die Chemotherapie angeschlagen hat). Mit der mobilen Dialogsystem-Technologie von RadSpeech kann der Arzt die Bilder selbst dann begutachten, wenn er nicht im Radiologie-Labor ist und per Sprache. Er kann also Diagnose auch abseits seines stationären Arbeitsplatzes erstellen.

Das von uns mithilfe der Dialogtechnologie des Deutschen Forschungszentrums für Künstliche Intelligenz, DFKI, entwickelte Sprachdialogsystem ist aktuell in der Lage, den folgenden Beispieldialog zwischen dem Benutzer (U) und RadSpeech (R) abzubilden.

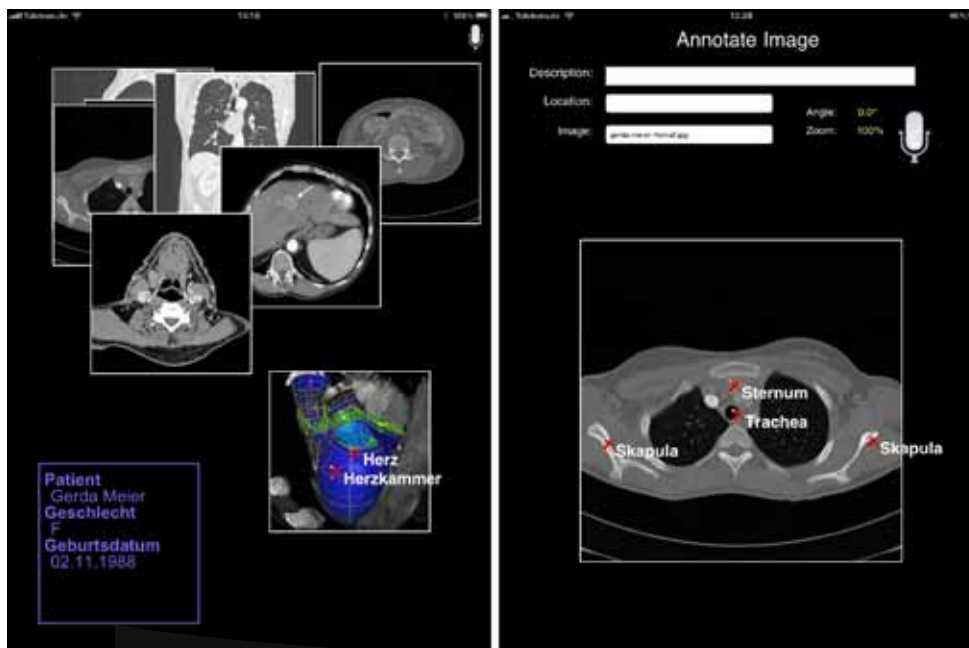
- U:** „Zeige mir die CTs der letzten Untersuchung, Patient XY.“  
**R:** Öffnet die entsprechenden CT-Bildserien.  
**U:** „Zeige mir zu diesem Patienten die inneren Organe: Lunge, Leber, dann Milz.“  
**R:** Zeigt die Patientenbilder der entsprechenden Akte, auf die sich sprachlich bezogen wird.  
**U:** „Annotiere hier (+ Zeigegeste) mit Vergrößerung der Lymphknoten.“  
**R:** „Die Region wurde mit ‘Vergrößerung der Lymphknoten’ annotiert.“

Der Befund kann zu einem späteren Zeitpunkt dann mit „Ergänze die Vergrößerung der Lymphknoten bei Patient XY mit Hodgkin-Lymphom“ zu einer Diagno-

se vervollständigt werden, die dann für die Suche ähnlicher Fälle und zum Training der Gewebe-Detektoren bereitsteht. Das wirklich Interessante neben der neuen Annotierungsfunktionalität ist der natürlichsprachliche Dialog, der durch die KI-Methoden der Computerlinguistik und die natürlichsprachliche Synthese von Antworten ermöglicht wird. In Zukunft wird der Dialog mit einem virtuellen Gesprächspartner immer mehr einem Gespräch mit einem menschlichen Gegenüber gleichen. Sprachdialogsysteme bieten also enorme Vorteile gegenüber herkömmlichen Spracherkennungssystemen, weil versucht wird, die Äußerungen des Arztes zu verstehen und entsprechend zu reagieren. Ein weiterer Vorteil von RadSpeech sind die kombinierte Sprachinteraktion mit anderen Ein-/Ausgabemedien, wie in unserem Fall die grafische Oberfläche des iPad Touchscreens.

### Diskussion

Dass neuerdings in vielen Krankenhäusern das strukturierte Berichten eingeführt wird, hat zur Folge, dass vordefinierte Formulare verwendet werden müssen. Dies empfinden beispielsweise Radiologen als Behinderung und fürchten, dass sie von den Bildern zu sehr abgelenkt werden. In dieser Situation soll dem Radiologen mit RadSpeech zukünftig ermöglicht werden, Bildannotationen im natürlichen Sprachdialog zu ergänzen bzw. zu überarbeiten. Der Fokus unserer Arbeiten lag hier, wie beschrieben, auf der Implementation eines praxistauglichen Dialogsystems für die Radiologie, das ausgiebig unter Real-Bedingungen installiert, getestet, und verbessert werden kann. Idealerweise würde dann der Radiologe Patientenbilder in einem mobilen 2D/3D-Viewer betrachten, die Bilder mittels Touchgesten manipulieren und gleichzeitig über Sprache multimodal und strukturiert befunden.



**Abbildung 4:** Radspeech multimodale Benutzeroberfläche mit Spracheingabe- und Dialogmöglichkeit.

Prof. Wolfgang Wahlster, der Chef des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI), sagte beim THESEUS Abschluss-Kongress, dass das THESEUS-Forschungsprogramm um semantifizierte Inhalte allgemein zur semantischen Wende in der Informatik beitragen wird, „Semantische Technologien sind Innovationstreiber der vierten industriellen Revolution“, so Wahlster. Der Faktor Mensch ist dabei einer von vielen Faktoren, industrielle Assistenzsysteme zur kognitiven Verstärkung und physischen Entlastung erfolgreich in Medizinanwendungen zu integrieren. Multimodale industrielle Assistenzsysteme rücken somit in den Mittelpunkt unserer Forschung in der Mensch-Maschine-Kommunikation für die Medizin. Sie bieten eine Möglichkeit, der individualisierten Medizin Vorschub zu leisten. Denn gerade im Medizinbereich ist es wichtig, bei steigender Autonomisierung durch KI-Methoden die Ärzte bestmöglich einzubeziehen. „Allerdings dürfen wir die Industrialisierung der Medizin nicht so verstehen, dass das Schicksal der Patienten der Apparatemedizin überlassen wird, sondern vielmehr im Sinne einer Produktivitäts- und Effizienzsteigerung bei qualitativ hochwertiger medizinischer Versorgung.“, sagt Prof. Herrmann Requardt (Siemens) im Zusammenhang der MEDICO-Arbeiten um die automatische Extraktion von Meta-Wissen aus heterogenen Datenquellen, semantischer, d. h. inhaltsbezogener Suche in medizinischen Datenbanken sowie adressatengerechter, datenintegrierender Präsentation und Interaktion bei einem Acatech-Symposium in Berlin.

Dass RadSpeech aus 130 THESEUS-Demonstratoren ausgesucht wurde, um Bundeswirtschaftsminister Dr. Philipp Rösler bei einem Rundgang auf der Computermesse CeBIT in Hannover präsentiert zu werden, zeigt auf, dass zukünftige Medizintechnik in der Strategie der Bundesregierung bei der deutschen Technologiepolitik im Bereich

der Informations- und Kommunikationstechnologien einen hohen Stellenwert hat. In der Pressemeldung zur RSNA 2011 hieß es: „Ausschlaggebend für den Erfolg von RadSpeech beim German HighTech Champions Award ist laut der internationalen Gutachter vor allem die Marktfähigkeit mobiler Dialogtechnologie für Medizin-relevante Anwendungen und der Reifegrad der DFKI-Technologie in diesem Segment, um auf dem internationalen Markt bestehen zu können. Die Marktfähigkeit wird insbesondere durch den rasch an Bedeutung gewinnenden US-Markt für mobile (medizinische) Anwendungen attestiert. Im Februar 2011 wurde zum Beispiel die erste iPad-Anwendung von der FDA (U.S. Food and Drug Administration) für radiologische Patientenbefundungen zugelassen. Apples neue Siri-Anwendung wird auch andere Zielgruppen für Sprachanwendungen gewinnen können und die Dialogtechnologie des DFKI knüpft mit echter Dialogführung nahtlos an den zukünftigen Markt an.“

Ist dies nicht zu viel des Lobes vor dem Hintergrund der schwierigen Fragestellungen um ärztliche Zulassungen und der generellen Problematik in Deutschland, medizinische Abläufe zu ändern und entsprechende KI- und Datenbanktechnologien in den Krankenhäusern zu installieren? Glücklicherweise sind wir darüber, einen forschungsorientierten ersten Ansatz der intelligenten sprachbasierten Mensch-Maschine-Interaktion für medizinische Anwendungen geschaffen zu haben. Der Problematik um die Verwertungsmöglichkeiten in bestehenden deutschen und internationalen Zertifizierungsprozessen von beispielsweise neuen Anzeigesystemen für die radiologische Befundung muss man sich stellen. Doch wer hätte vor 3 Jahren gedacht, dass man jetzt seit knapp einem Jahr in den USA eine iPad-App zur diagnostischen radiologischen Befundung einsetzen darf? Das neue iPad beispielsweise, das ab April 2012 auf den Markt kommt, besticht insbe-

sondere durch eine noch höhere Auflösung des Retina-Displays (2048x1536 Bildpunkte) und setzt deutlich höhere Maßstäbe bei der Anzeige von Bilddaten, was sich auf eine zukünftige EU-Zertifizierung auswirken kann. Es darf vor allem nicht vergessen werden, was Prof. Carsten Schultz in einer der letzten DZKF-Ausgaben betonte, dass sich die eingeschränkte Innovationsfähigkeit negativ auf die Implementierung medizinischer und medizintechnischer Forschungsergebnisse auswirken kann. Änderungen im Diagnose- und Behandlungsprozess ziehen in unserem speziellen Feld um „semantische“ Bildverarbeitung und Interaktion Änderungen in der Arbeitsteilung mit sich, wenn hochausgebildete IT-Experten nicht nur die internen Datenbanken pflegen, sondern mit semantifizierten Inhalten und neuen Anfragesprachen zurechtkommen und eine krankenhausübergreifende Vernetzung bereitstellen müssen.

Was ist die zukünftige Strategie, wie Klinik-übergreifende „semantifizierte“ Daten bereitgestellt werden? Prof. Manfred Dietel (Charité) sagte auf der Cebit 2012, dass vor allem der Begriff Systemmedizin gemeinsam weiterentwickelt werden muss, da eigene Datenschutzregeln in den Bundesländern die allgemeine Zusammenarbeit behindern. Wie sieht es in Europa aus? Die Sprachvielfalt in Europa behindert bislang die grenzüberschreitende medizinische Datensammlung (wie etwa Krebsregister) stark. Es drängt sich die Frage auf, ob semantische Technologien auch hier eingesetzt werden können, um solche heterogenen Daten sinnvoll zu verknüpfen, vorausgesetzt entsprechender Gesetzesänderungen zugunsten eines umfassenden Datenschutzes. Ganz wichtig zu erkennen ist das Potential, dass aus anonymisierten Datensätzen zusätzliches fallbasiertes Wissen zur Diagnose und Therapie gewonnen werden kann<sup>[10]</sup>. Beispielsweise kann man mit KI-Methoden eine abstrakte Repräsentation der Topographie der Anatomie des menschlichen Körpers herleiten. Insbesondere wird dieses Wissen dann wertvoll, wenn es mit den Ergebnissen der Segmentierungsalgorithmen verglichen wird. Werden detektierten Organen Positionen zugeordnet, die aufgrund der anatomischen Umstände nicht möglich sind (eine Blase kann nicht oberhalb des Herzens liegen), wird im ersten Schritt eine Fehlermeldung gesendet, die im zweiten Schritt der Integration unter Usability-Aspekten<sup>[6]</sup> zu einer Dialog-basierten Interaktion ausgebaut werden kann.

Ministerialrat Dr. Görderler vom BMWi, der THESEUS betreute, erörtert auf der Cebit 2012, dass dieses Jahr noch intensiver über Forschungsprojekte mit intelligenter Vernetzung, stärkerer Autonomie, und Produkte mit Eigenintelligenz debattiert wird. RadSpeech war ein Projekt, das den Transfer der DFKI-Dialogtechnologie in Industrie-relevante medizinische Anwendungsfelder in THESEUS MEDICO zum Ziel hatte, wo wir bereits einen implementierten Teil dieser Agenda prototypisch im medizinischen Umfeld erproben konnten. Ich wage zu prognostizieren, und hoffe es auch, dass diese und ähnliche

mobile Dialog-Technologien als Teil eines intelligenten Assistenzsystems als Medizinprodukt den Markt im Bereich der medizinischen Versorgung in einigen Jahren mitgestalten werden.

\* MEDICO ist ein Anwendungsszenario des Großprojekts THESEUS (2007-2012), in dem rund 30 öffentliche und industrielle Forschungspartner an Basistechnologien und Standards für die nächste Generation des Internets arbeiten ([www.theseus-programm.de](http://www.theseus-programm.de)). Das Bundeswirtschaftsministerium fördert THESEUS mit rund 100 Millionen Euro. Nochmals den gleichen Betrag müssen die beteiligten Industriepartner aufbringen.

#### LITERATUR

- [1] Marius Erdt, Matthias Kirschner, Klaus Drechsler, Stefan Wesarg, Matthias Hammon, Alexander Cavallaro. Automatic pancreas segmentation in contrast enhanced CT data using learned spatial anatomy and texture descriptors. In IEEE International Symposium on Biomedical Imaging, 2011
- [2] Hans-Peter Kriegel, Marisa Petri, Matthias Schubert, Michael Shekelyan, Michael Stockerl. Efficient Similarity Search on 3D Bounding Box Annotations. In Proceedings of the SPIE Medical Imaging Conference 2012: Advanced PACS-based Imaging Informatics and Therapeutic Applications, 2012
- [3] Sascha Seifert, Adrian Barbu, S. Kevin Zhou, David Liu, Johannes Feulner, Martin Huber, Michael Suehling, Alexander Cavallaro and Dorin Comaniciu. Hierarchical parsing and semantic navigation of full body CT data. SPIE Medical Imaging, 2009
- [4] Sascha Seifert, Sonja Zillner, Martin Huber, Michael Sintek, Daniel Sonntag, and Alexander Cavallaro. Intelligente Suche in medizinischen Bilddatenbanken. In Heuser and Wahlster (eds.), *acatech DISKUTIERT* (2011).
- [5] Daniel Sonntag. *Ontologies and Adaptivity in Dialogue for Question Answering Monograph in Studies on the Semantic Web*. ISBN: 978-3-89838-623-4 (AKA), 978-1-60750-054-4 (IOS Press), 2010, Hard Cover
- [6] Daniel Sonntag, Colette Weihrauch, Oliver Jacobs and Daniel Porta. *Usability Guidelines for Use Case Applications*. DFKI Technical Report, THESEUS CTC-WP4, Vol. 1, 4/2010
- [7] Daniel Sonntag, Marcus Liwicki and Markus Weber. Digital Pen in Mammography Patient Forms. Proceedings of International Conference of Multimodal Interaction (ICMI), 2011
- [8] Daniel Sonntag, Norbert Reithinger, Gerd Herzog and Tilman Becker. *Discourse and Dialogue Infrastructure for Industrial Dissemination*. Proceedings of the International Workshop on Spoken Dialogue Systems Technology (IWSDS), Lee et al. (Eds.): LNAI 6392, Springer 2010
- [9] Manuel Möller, Daniel Sonntag and Patrick Ernst. A Spatio-Anatomical Medical Ontology and Automatic Plausibility Checks. Book Chapter in "IC3K Selection", *Communications in Computer and Information Science (CCIS)*, Springer 2012
- [10] Sonja Zillner and Daniel Sonntag. Image Metadata Reasoning for Improved Clinical Decision Support. *Network Modeling Analysis in Health Informatics and Bioinformatics Journal*, Springer 2012



**DR. DANIEL SONNTAG**

Deutsches Forschungszentrum  
für künstliche Intelligenz (DFKI)  
Stuhlsatzenhausweg 3, D-66123 Saarbrücken,  
Tel: +49 681 857 75-5254  
E-Mail: [Sonntag@dfki.de](mailto:Sonntag@dfki.de)  
[www.dfki.de/RadSpeech/](http://www.dfki.de/RadSpeech/)