CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Institutional Repository of the Freie Universität Berlin

Elena Pasiaru- Bontas
Netzbasierte Informationssysteme
Institut für Informatik
Freie Universität Berlin

Doktorandenworkshop
24.10-25.10.2003

„Semantic Web in der Pathologie“

1. Projektvorhaben

Das von der Deutschen Forschungsgemeinschaft DFG geförderte Forschungsprojekt „Semantic Web in der Pathologie“ erprobt neueste Technologien des Semantic Web in der Anwendung. Es wird ein prototypisches System entwickeln, das Bildbeschreibungen und Befundberichte aus der Pathologie, speziell Lungenkrankheiten, bereitstellen und verarbeiten kann. Das im September 2003 begonnene Projekt ist eine Kooperation zwischen der Freien Universität Berlin (AG Netzbasierte Informationssysteme), der Humboldt-Universität (AG Digitale Pathologie und EDV des Institutes für Pathologie der Charite) und der Universität Potsdam (AG Computerlinguistik).

Um die Idee der digitalen Pathologie zu verwirklichen ist es notwendig, spezifische Daten, wie digitale Bilder oder Befundberichte, elektronisch zu speichern und in einer Form zur Verfügung zu stellen, die eine effiziente und vielseitige Nutzung unterstützt. Der zunehmende Einsatz von digitalen Mikroskopen ermöglicht die Digitalisierung und Archivierung der entstandenen Bilder. Die entsprechenden Befundberichte werden aber manuell oder mit Hilfe von Spracherkennungsprogramme digitalisiert und abgespeichert. Es besteht heute keine praktikable Möglichkeit, das in den Berichten enthaltene hochqualitative medizinische Wissen in der klinischen Arbeit effizient wieder zu verwenden.

Das Projekt „Semantic Web in der Pathologie“ wird eine Plattform anbieten, die dieses vorhandene Expertenwissen verwendbar macht, in dem sie Pathologen erlaubt, zu einem aktuellen Fall schnell auf Vergleichsfälle mit Vergleichsbildern für die Unterstützung von Diagnostik und Differentialdiagnostik, sowie auf aktuelle klinische Informationen und statistische Daten zuzugreifen. Das Expertenwissen ist dann auch in anderen Anwendungsszenarien, wie Qualitätssicherung oder Medizinerausbildung, zugänglich.

Technologisch basiert das geplante System auf den Standards des Semantic Web wie XML, RDF, OWL und RuleML. Modellierungsseitig stützt sich das System auf anerkannte medizinische Modelle und Standards (UMLS, HL7, SCORM).

Bei der Erstellung von Befundberichten wird semantische Information durch computerlinguistische Analyse automatisch erfasst und anschließend in ein semantisches Netz eingefügt, das auf einem Pathologie-spezifischen Domänenmodell basiert. Der Datenbestand der Befundberichte wird damit semantisch erschlossen, um eine inhaltliche Recherche und Verarbeitung zu gestatten. So können beispielsweise die Bilder dann nicht nur nach „äußerlichen" Kriterien wie Formen und deren Farbe, sondern nach Bedeutung wie Art eines zu erkennenden Tumors abgefragt werden.

2. Systemarchitektur

Das System besteht aus:

- Beschreibungskomponente
- Transformationskomponente
- Fachwissenskomponente
- Anwendungskomponenten
- Integrationskomponente (siehe Abbildung 1).

Der Befundbericht und die Bildbeschreibung werden in der *Beschreibungskomponente* einer semantischen und linguistischen Analyse unterzogen. Durch Anwendung der *Fachwissenskomponente*, die die Ontologie, bereits erstellte Befunde und Bilder sowie ein Regelwerk enthält, werden Pathologen bei Fertigstellung des Befundes eine Statusmeldung über die innere Konsistenz des Befundes gegeben, bei Fragen nach Vergleichspräparaten eine Liste mit Bildern der gleichen Diagnose und Beschreibung sowie der möglichen Differentialdiagnose angezeigt. Zusätzlich können Zusammenfassungen diagnosebezogenen Daten erstellt werden (durchschnittliches Alter der PatientInnen mit dieser Erkrankung, durchschnittliche Anzahl der Wiederholungsuntersuchungen u. v. a. m.). Mittels der *Beschreibungskomponente* werden die digitalen Schnittbilder beschrieben. Die Syntax baut auf eine XML- Struktur auf. Im Rahmen des Virtuellen Mikroskops wird diese Komponente im diagnostischen Prozess zur Erstellung des Befundes verwandt. Gleichzeitig ist diese *Beschreibungskomponente* in der Lage, aus bestehenden Datenbanken, Befunde in eine XML- Struktur umzuformen. Die *Transformationskomponente* fügt den Befund mit Bildbeschreibung dem Semantischen Netz hinzu, welches in der *Fachwissenskomponente* gebildet und modelliert wird. Es werden Regeln auf das semantische Modell angewendet, dem als Domänenmodell UMLS und GeneOntology unterliegt.

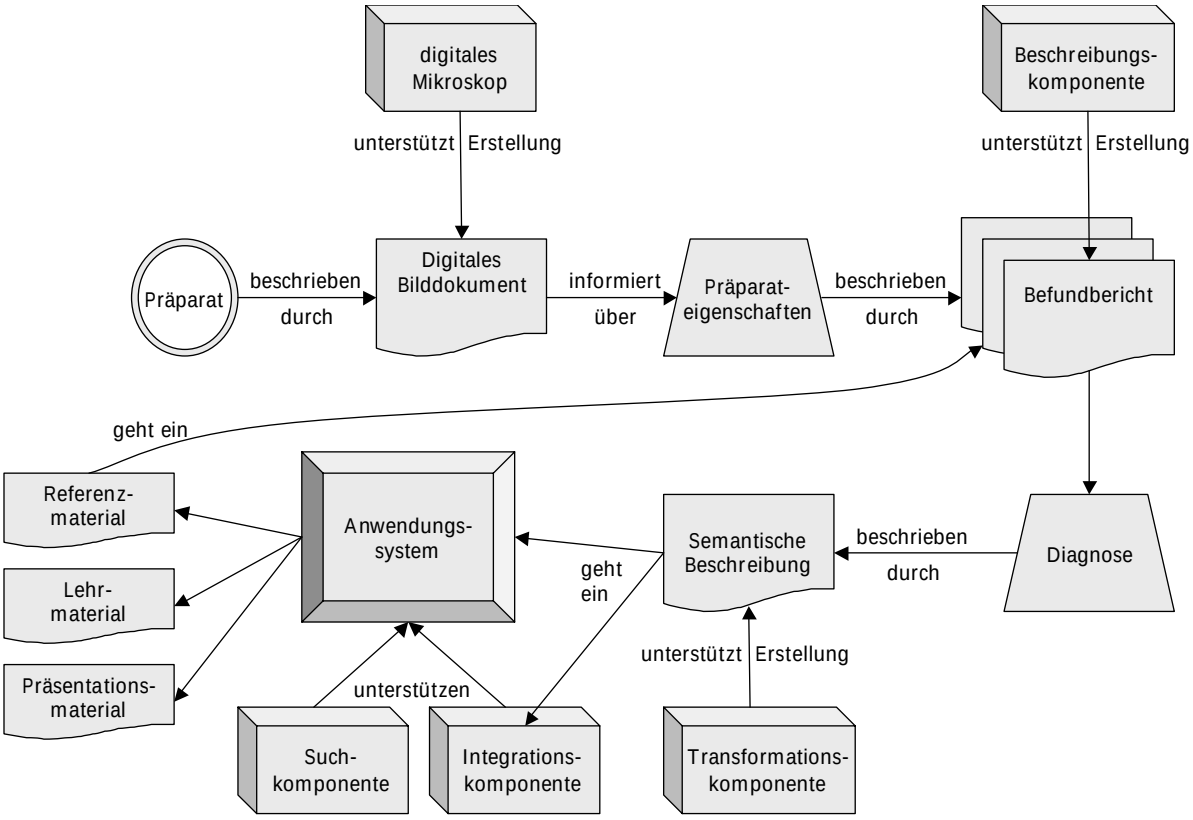

Abbildung 1: Systemarchitektur

Die *Anwendungskomponenten* stellen die Schnittstellen mit dem Benutzer dar. Dazu gehören eine Suchmaschine für Pathologie-relevante Informationen, ein Qualitätssicherungstool und eine Semantik- orientierte Pathologie Retrieval- Engine zur Unterstützung der Diagnostik, welches Informationen und Bilder zum Vergleich und zur Differentialdiagnostik zur Verfügung stellt. Ein weiteres Tool ist in der Lage, aus den hoch strukturierten Daten Analysen mit beschreibender Statistik durchzuführen mit Zusammenfassung von Informationen über Fälle, Erkrankungen und diagnostischen Schritten. Durch die modulare Bauweise dieser Softwarekomponenten können neue Funktionen ergänzt werden. Schließlich hat die *Integrationskomponente* die Aufgabe externe Wissensquellen einzubinden.

3. Forschungsfragen

Besondere Fragestellungen im Projekt sind

- die hinreichende Mächtigkeit der Semantic Web Technologien
- die geeignete Auswahl medizinischer Fachmodelle
- die Genauigkeit der inhaltlichen Klassifikation
- die Praxistauglichkeit des Systems im klinischen Alltag.

Im Folgenden wird jede Fragestellung und ihre Auswirkungen auf die Realisierung des Systems ausführlich diskutiert.

Die Semantic Web Initiative verfolgt die Idee, Web Ressourcen unter Verwendung von Techniken und Methoden der Wissensrepräsentation zu erweitern, damit sie maschinell leichter gefunden, besser genutzt und individuell zusammengestellt werden können. Um die Semantic Web zu verwirklichen ist es notwendig, dass Wissen erstens explizit und maschinenlesbar dargestellt, zweitens wieder verwendet und ausgetauscht werden kann. Ontologien, als „explizite und formale Spezifikation einer gemeinsamen Konzeptualisierung", spielen dabei eine entscheidende Rolle. Ausgehend von XML wurden Ontologie-Sprachen wie RDF(S), OIL, DAML+OIL und schließlich OWL mit unterschiedlicher Semantik und Ausdruckskraft bereits entwickelt. Es ist allerdings offen wie ein Semantic Web aussehen soll und ob Semantic Web- basierte Systeme überhaupt praxistauglich sind.

Obwohl Ontologien mittlerweile in vielen Anwendungen aus der Medizininformatik, Bioinformatik oder E- Commerce verwendet werden, werden Systeme gerade in verteilten Umgebungen mit dem Problem der semantischen Interoperabilität konfrontiert. Wie die Definition schon erläutert, einer Ontologie liegt eine gewisse Sicht, in einem bestimmten Kontext, zu Grunde. Welche Aspekte einer Domäne relevant sind, was sie konkret bedeuten und wie sie untereinender in Beziehung stehen wird entscheidend durch den Kontext geprägt. Dementsprechend ist es wichtig, dass Semantic Web Anwendungen Kontextinformationen in Wissensprozesse berücksichtigen. Die explizite Repräsentation der Kontextinformationen kann die Integration von lokalen Sichten oder die lokale Anpassung einer globalen Sicht an einem Kontext ermöglichen. Das Projekt „Semantic Web in der Pathologie" wird entscheidend mit dem Thema „Semantische Interoperabilität" konfrontiert. Die Fachwissenskomponente des Systems berücksichtigt unterschiedliche Medizinontologien, die zwar eine weite Palette an Termini und Teilgebiete der Medizin umfassen, aber teilweise auch unterschiedliche Sichten vertreten oder für eine gewisse Art von Aufgaben bestimmt waren (Verwaltungsaufgaben, Diagnose, Anatomie etc.). Die Unterschiede in der Granularität oder sogar semantische Ordnung in den Standard Ontologien aus der Medizin (UMLS, GALEN, GeneOntology, SNOMED etc.) deutet darauf hin, dass solche Wissensquellen nicht ohne weiteres integriert und verwendet werden können.

Ein weiterer offener Punkt sind die Modellierungsparadigmen der einzelnen Medizinontologien. Im Projekt soll untersucht werden in wie fern bestehende semantische Konzeptbeziehungen genutzt werden können, um eine effiziente Suche nach bildlicher und textueller Information zu ermöglichen, bzw. welche Verknüpfungen auf konzeptioneller und semantischer Ebene dafür notwendig wären. Trotz ihren beeindruckenden Umfang sind Medizinontologien oft so genannte „low-level" Ontologien im Sinne des Semantic Web. Eine unpräzise Semantik der Beziehungen und heterogene Klassifikationskriterien tragen dazu bei. Unscharfe quantitative Aussagen, wie „viele", „klein", oder beschreibende Termini, wie „oval", „irregulär" können nicht automatisch berücksichtigt werden. Außerdem sind zeitliche,

räumliche oder generell kontextuelle Informationen, die sowohl in der Diagnose, als auch in der Therapie eine wichtige Rolle spielen, nicht Teil der vorhandenen Ontologien und können dementsprechend nicht im Reasoningprozeß verwendet werden.

Die Praxistauglichkeit des prototypischen Systems spielt letztendlich die entscheidende Rolle bei der Beantwortung der projektrelevanten Forschungsfragen.