

Digitale Archive

von Christoph Schlieder



Anwendungen der Informatik in den Kulturwissenschaften

Während die Informationstechnologien das wissenschaftliche Arbeiten in den Natur- und Ingenieurwissenschaften tief greifend verändert haben, war ihr Einfluss auf die Kulturwissenschaften ungleich geringer. Quantitative oder stark strukturierte Datenbestände, die den Einsatz bewährter informatischer Werkzeuge (z. B. Datenbanksysteme) erlauben, spielen hier eine untergeordnete Rolle. Kulturwissenschaftliche Daten, die häufig in Form einer Sammlung von Text-, Bild- oder Tondokumenten vorliegen, stellen für Informatiker gerade deswegen eine besondere Herausforderung dar, weil sie nicht ohne weiteres maschinell auswertbar sind. Fortschritte im Bereich der Semantischen Informationsverarbeitung machen es heute möglich, auch solche schwach strukturierten Daten systematisch zu bearbeiten. Damit eröffnen sich neue Perspektiven für den Einsatz von informatischen Werkzeugen in den Kulturwissenschaften.

Die Welt der digitalen Archive wird von einem technologischen Wandel erfasst, der sich am besten am Beispiel der Informationssuche illustrieren lässt. Die gängige Lösung besteht im Einsatz von Suchmaschinen, die nach ähnlichen Prinzipien funktionieren wie die aus dem World Wide Web bekannten. Durch Testanfragen stellt man leicht fest, dass diese nützlichen Werkzeuge auch erheblichen Beschränkungen unterliegen. So errechnet beispielsweise die Suchmaschine Google zum Suchterm "Bismarck" als ersten Treffer den Webauftritt der Stadtverwaltung von Bismarck, Hauptstadt des US-Bundesstaats North Dakota. Auf Platz 2 setzt das patentierte Verfahren das 1941 gesunkene Kriegsschiff, dann folgen weitere Seiten, die auf Stadt oder Schiff verweisen, bis an 19. Stelle die Einstiegsseite der Bismarck-Stiftung erscheint. Erst auf Platz 20 findet sich Information zu Otto von Bismarck, allerdings handelt es sich um eine recht entlegene Publikation, einen Artikel aus dem Online-Magazin der deutschen Botschaft in Indien.



Prof. Dr. Christoph
Schlieder

Angewandte Informatik in
den Kultur-, Geschichts-
und Geowissenschaften

Gewiss drückt sich in der Rangfolge der Treffer keine universelle Wissensordnung aus. Sie ist vielmehr das Ergebnis der derzeitigen qualitativen und quantitativen Dominanz von Webinhalten aus den USA. Setzt sich der Zuwachs chinesischer Webseiten weiterhin fort, dann könnte die errechnete Rangfolge in ein paar Jahren eine ganz andere kulturelle Prägung aufweisen. Doch heute schon wird sichtbar, dass die herkömmliche Suchmaschinenteknologie mit ihren Verfahren zur quantitativen Bewertung der Relevanz von Dokumenten an Grenzen stößt. Eine aus Nutzersicht nahe liegende, technologisch aber erst durch jüngste informatische Fortschritte ermöglichte Verbesserung

besteht darin, die Suche durch Kategorisierung zu beschränken zum Beispiel auf Schiffe, die Bismarck heißen oder auf Stiftungen. Diese begriffliche Filterung stellt die einfachste Form der semantischen Suche dar. Komplexere Formen der semantischen Suche ermöglichen es, zu einer Anfrage nach „Museum“ auch Seiten zu finden, auf denen nur das Wort „Pinakothek“ vorkommt, also Ober- bzw. Unterbegriffe in die Suche mit einzubeziehen, wo sich das als zweckmäßig erweist. Am Ende stehen Suchwerkzeuge, die verschiedenste relationale Verknüpfungen von Suchtermen berücksichtigen können. In der Welt des Internet sind die neuen Verfahren als Semantic-Web-Technologien bekannt. Sie haben begonnen, die Labors der Entwickler zu verlassen und sich in der Praxis zu bewähren, auch außerhalb des Internet.

Selbstbeschreibende Dokumente

In Bamberg wird an Anwendungen dieser Technologien auf Problemstellungen aus den Kulturwissenschaften gearbeitet. Erstes Ergebnis ist das *Mobile Mapping System*, ein Softwaresystem zur Kartierung historischer Bauwerke mit mobilen Rechnern, das am Lehrstuhl für Informatik in den Kultur-, Geschichts- und Geowissenschaften entwickelt wurde in Zusammenarbeit mit Anwendern des Instituts für Archäologie, Bauforschung und Denkmalpflege. Digitale Karten werden mit einem elektronischen Stift direkt auf dem Bildschirm des Rechners, einem *Tablet PC*, gezeichnet. Das leichte Gerät wird über Akkus versorgt und kann Daten über ein Funknetz austauschen – eine wichtige Unterstützung für das Arbeiten auf einem Baugerüst.

Die Besonderheit der Software besteht darin, dass sie digitale Karten erzeugt, die mit Hilfe von Semantic-Web-Technologien die semantische Suche in digitalen Archiven unterstützen. Ein auf das Kartierungssystem abgestimmtes Archivsystem wird derzeit am Institut für Informationssysteme der Universität Passau entwickelt. Die Lösung für das Problem der semantischen Suche beruht auf einem einfachen Prinzip: Jedem Dokument, beim *Mobile Mapping System* sind das digitale Karten, wird eine maschinell verarbeitbare Inhaltsbeschreibung – so genannte Metadaten – mitgegeben. Semantische Suchfunktionen brauchen nur die Metadaten auszuwerten, der eigentliche Inhalt kann für sie unverständlich sein. Man bezeichnet Dokumente, die mit inhaltsbezogenen Metadaten versehen sind, auch als sich selbst beschreibende Dokumente. Neu ist an diesem Ansatz nicht die im Bibliothekswesen schon lange gebräuchliche Verwendung von Metadaten sondern deren Beschreibung mit Mitteln der formalen Logik. Dieser für die Semantic-Web-Technologien kennzeichnende Schritt ermöglicht erst komplexe Auswertungen wie das Ausrechnen von Ober- und Unterbegriffen oder von relationalen Assoziationen. Auch können Metadaten so auf Konsistenz geprüft und gegebenenfalls ergänzt werden. Dabei ist zu bedenken, dass die Bereitschaft der Nutzer, Metadaten selbst zu erstellen, z.B. durch Verschlagwortung oder gar formalsprachliche Inhaltsbeschreibung, verständlicherweise gering ist. Andererseits geben die Daten, der Prozess und der Kontext der Dokumentenerstellung häufig Hinweise auf den Inhalt. Im *Mobile Mapping System* wird dies genutzt, um die Metadaten automatisch zu erzeugen. Der Nutzer kann wie beim Arbeiten mit Papier und Bleistift vorgehen. Aus der Zeichnung erstellt das System dann ohne weiteres Zutun eine sich selbst beschreibende digitale Karte.

Das gemeinsam mit Prof. Drewello, Restaurierungswissenschaften in der Baudenkmalpflege, durchgeführte Projekt wird von der Deutschen Bundesstiftung Umwelt (DBU) finanziert, der an dieser Stelle für die Förderung herzlich gedankt sei.

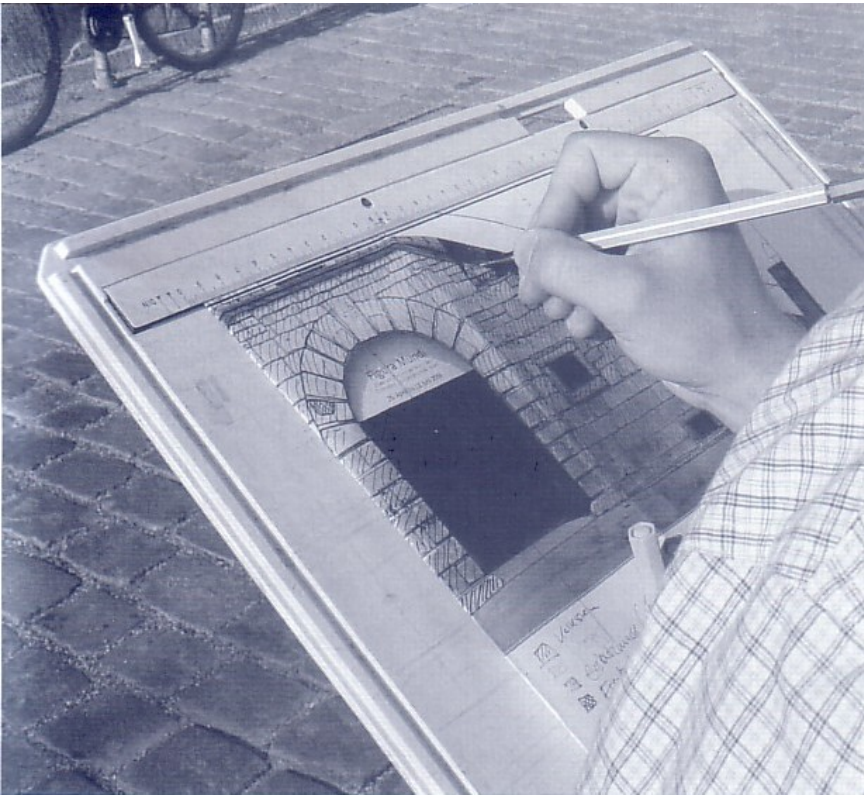


An der Passauer Dombauhütte werden regelmäßig Bestands-, Schadens- und Maßnahmenkartierungen als Planungsgrundlagen erstellt. Die Digitale Kartierung vereinfacht die Sachdatenerfassung vor Ort sowie die spätere Auswertung.

Rückkehr der Ontologie

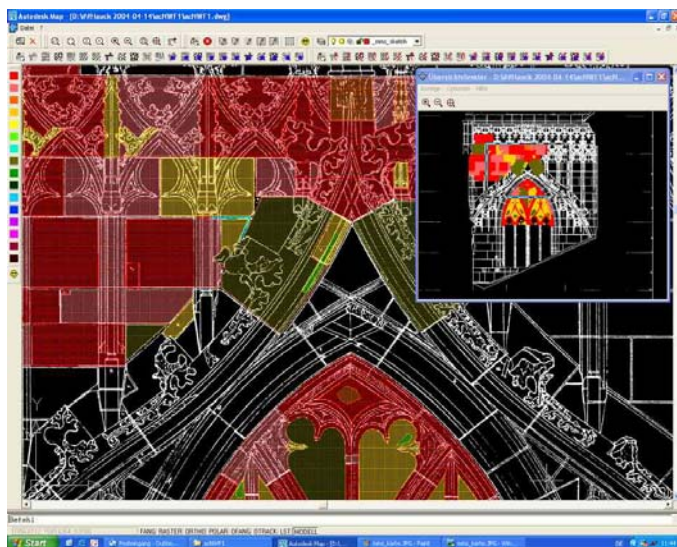
Ein Archivsystem muss nicht nur verschiedene Typen von Dokumenten verwalten, sondern vor allem unterschiedliche Nutzergruppen bedienen. Das wurde während der Erprobungsphase des Systems bei der Dombauhütte Passau und der Bamberger Restaurierungswissenschaft sehr deutlich. Die Nutzer des Archivs bilden verschiedene Informationsgemeinschaften (*information communities*), die durch Gemeinsamkeiten der Informationsverarbeitung, sei es auf der technischen, kognitiven oder sozialen Ebene charakterisiert sind. So können beispielsweise Mitarbeiter mit derselben Ausbildung oder Träger einer bestimmten Funktion in einer Organisation eine Informationsgemeinschaft bilden. Für die Gestaltung von Anwendungssystemen sind Informationsgemeinschaften von Bedeutung, weil sie häufig radikal verschiedene Sichten auf denselben Gegenstand einnehmen. Ein Mitarbeiter der Dombauhütte sieht im Passauer Dom anderes als ein Doktorand der Restaurierungswissenschaft und der wieder anderes als ein Kunsthistoriker. Sie sehen nicht nur Verschiedenes, sondern sie benennen es auch anders.

Informationsgemeinschaften entwickeln eigene Konzeptualisierungen des Gegenstandsbereichs, was sich in eigenen Fachterminologien niederschlägt. Semantische Suche muss diese vielfältigen Begriffssysteme berücksichtigen. Das geschieht durch eine informatische Modellierung der Fachterminologien, die sich der gleichen logischen Hilfsmittel bedient wie die Metadatenbeschreibung, nämlich Beschreibungslogiken (*description logics*). Dadurch kann das Archivsystem Dokumente verwalten, deren Metadaten in den Fachterminologien verschiedener Informationsgemeinschaften verfasst sind. Es ist ein weiteres Kennzeichen der Semantic-Web-Technologien, den Informationsaustausch über terminologische Grenzen hinweg mit Hilfe von Methoden der Logik zu unterstützen. Aus Anwendersicht tut sich hier eine interessante Alternative zu Bestrebungen der Vereinheitlichung und Standardisierung von Fachterminologien auf. In der chaotischen Welt des World Wide Web liegt der Gedanke an eine einheitliche Fachsprache ebenso fern wie die Annahme, dass alle Nutzer mit derselben Software arbeiten. Wichtig ist nur, dass sich verschiedene Informationssysteme austauschen können, dass sie, wie man auch sagt, interoperabel sind. Das Ziel der Interoperabilität, technisch präziser, der semantischen Interoperabilität dürfte für die transdisziplinär arbeitenden Kulturwissenschaften letztlich attraktiver sein als das der Standardisierung. Die Semantic-Web-Technologien haben ihre Wurzeln nicht nur in der mathematischen



Bei der Kartierung auf Papier können Sachdaten nur durch graphische Elemente oder Annotation von Text erfasst werden. Die Fortführung von Karten über einen längeren Zeitraum führt so schnell zu unübersichtlichen Ergebnissen.

Tradition das auf, was Herbert Schnädelbach als „phänomenologische Neubegründung von Ontologie“ entlang einer „Reihe österreichischer Bahnstationen (Bolzano, Brentano, Meinong, Husserl)“ beschrieben hat. Eine ausführlichere Würdigung der philosophischen Vorgeschichte der Semantic-Web-Technologien würde belegen, dass es sich bei Philosophie und Informatik um verschiedene aber seit längerem interoperabel arbeitende Informationsgemeinschaften handelt



Mit dem in Bamberg entwickelten Mobile Mapping System lässt sich ähnlich wie auf Papier arbeiten. Allerdings können zusätzlich Sachdaten erfasst und Metadaten generiert werden.

sondern auch in der philosophischen Logik. Eine mit logischen Mitteln beschriebenen Fachterminologie wird in der Informatik mit dem traditionsreichen Begriff „Ontologie“ bezeichnet. Es handelt sich hier nicht um einen terminologischen Zufall. Die informatische Aneignung des Begriffs hat sich vielmehr unter Mitwirkung von Philosophen vollzogen. Geprägt wurde das informatische Verständnis von Ontologie durch das Forschungsprogramm der Wissensrepräsentation, das in der Tradition der analytischen Philosophie von Willard Van Orman Quine steht, welcher die ontologischen Verpflichtungen einer wissenschaftlichen Theorie an den verwendeten formal-sprachlichen Mitteln festmacht. Blickt man hinter Quine zurück, dann tritt als Anknüpfungspunkt zur europäischen philosophischen

Kulturinformatik in Bamberg

Digitale Archive mit semantischen Suchfunktionen sind nicht der einzige Gegenstand, mit dem sich eine Angewandte Informatik der Kulturwissenschaften befasst. Die Bamberger Kulturinformatik forscht beispielsweise auch an Verfahren maschinellen Bildanalyse oder der Verarbeitung von Geodaten. Als Angewandte Informatik der Kulturwissenschaften kann sich die Kulturinformatik nicht auf ein Verständnis von Kulturwissenschaft, im Singular oder Plural, festlegen. Ihre Aufgabe besteht vielmehr darin, die Vielfalt der wissenschaftlichen Gebiete und Ansätze abzubilden und einen technologischen Beitrag zu leisten, der wissenschaftliche Arbeiten über

die Grenzen der Informationsgemeinschaften hinaus aneinander anschließbar macht. Im besten Fall schafft sie ein Medium, über das sich Wissen vermittelt und das die Entstehung von Wissen befördert. Letzteres in einem spezifisch informatischen Sinn. Die Forschung zur Semantischen Informationsverarbeitung versteht unter Information (in bewusster Abkehr vom Informationsbegriff der Shannonschen Informationstheorie) Zusammenhänge bzw. Regularitäten in Daten. Als Wissen werden die interessanten Zusammenhänge bezeichnet. Die Entscheidung darüber, welche Zusammenhänge von Interesse sind, ist keine informatische. Sie liegt bei den Fachwissenschaften. Was interessant ist, ergibt sich häufig aber nicht innerhalb einer Informationsgemeinschaft oder methodischen Position, sondern aus der Verbindung verschiedener Sichtweisen.

Eine Technologie die dieses unterstützt, in dem sie ein Medium schafft für den Austausch zwischen Informationsgemeinschaften und für die Verknüpfung von Wissen, könnte man zu Recht als Kulturtechnik bezeichnen und in eine Reihe stellen mit anderen Technologien, die die Vermittlung von Wissen revolutioniert. Technologie jedenfalls nicht, das World Wide Web noch nicht, die neue Generation der digitalen Archive befindet sich auf dem Weg dazu