

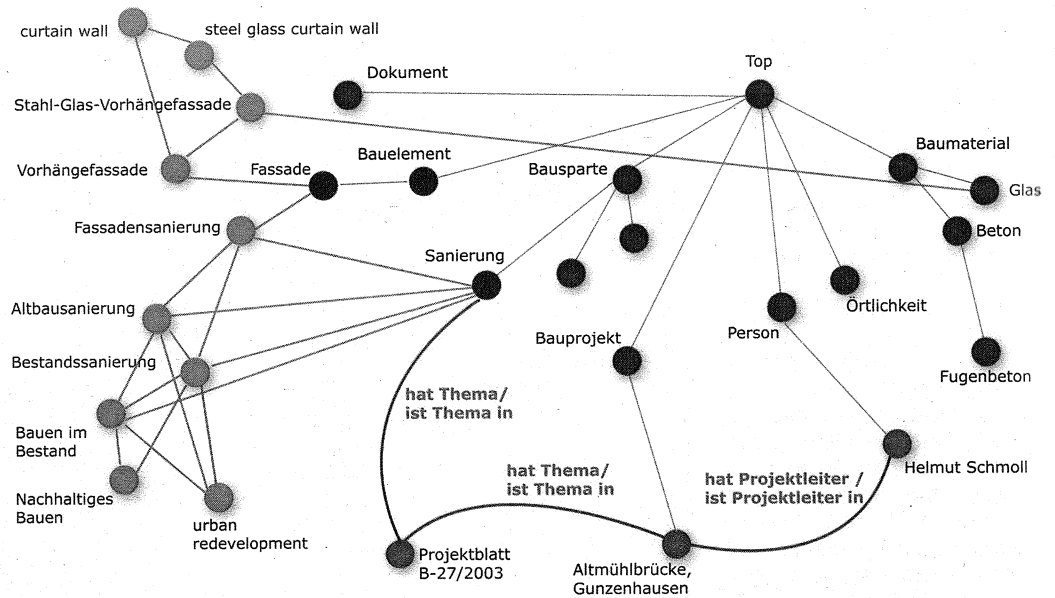
Beziehungskisten

Das Fraunhofer-Institut für Integrierte Publikations- und Informationssysteme (Fraunhofer IPSI) hat mit ConWeaver ein Werkzeug entwickelt, das Informationen automatisch aus Texten extrahiert und in eine Ontologie einfügt. Es vernetzt Informationen zu Projekten, Produkten, Kunden und Fachfragen. Davon profitiert unter anderem das Unternehmen Bilfinger & Berger für sein internationales Wissensmanagement.

Nach dem Abschluss eines Projekts verschwindet häufig ein Großteil des dort angehäuften Wissens auf Dateiservern, auf CDs im Regal oder in Papiermappen zwischen zwei Aktenordner gezwängt. Die Ablagestrukturen großer Unternehmen sind selten transparent genug, um die Träger solchen Wissens zu identifizieren. Ontologien versprechen einen Ausweg aus der Misere. Sie verknüpfen Informationen von ansonsten weit gehend unabhängig arbeitenden Wissensquellen wie einem Dokumentenmanagement- oder Enterprise-Resource-Planning-System. Wenn „Alfred Neumann“ als Kunde Gegenstand der Suche ist, weiß die Ontologie, dass beim gleichnamigen „Besteller“, „Customer“ und „Vertragnehmer“ die identische Person gemeint ist. Gleichzeitig kann sie den Kunden vom Mitarbeiter Alfred Neumann aus der Buchhaltung abgrenzen.

Aus den Daten entstehen durch den Import in eine Ontologie Entitätenetze: „Helmut Scholl“ ist Projektleiter von „Altmühlbrücke“, „Altmühlbrücke“ ist Thema in „Projektblatt B-27/2003“. Damit sich das Gerüst aus Projekten und Personen mit Kompetenzen füllt, die sich später über eine Recherche ermitteln lassen, muss die Ontologie das in Projektblättern und Dossiers gespeicherte Wissen abbilden.

Ein Indexierungsmodul analysiert die Texte nach statistischen und linguistischen Methoden. Es identifiziert zunächst das Vokabular, vor allem Komposita („Autobahnbrücke“) und Mehrwortausdrücke („Brücke über eine Autobahn“). Danach filtert es einerseits aus diesem Vokabular Entitäten heraus und ver-



Das hinter Conweaver steckende semantische Netz findet von der Altbausanierung über das Projektblatt den passenden Experten.

knüpft sie mit der Ontologie: „Dehnfuge“ ist Unterbegriff von „Fuge“, welches verwandt mit „Verfugung“ ist, welches das Projektblatt B-32/2004 thematisiert.

Um die thematische Vernetzung des Vokabulars zu analysieren, zerlegt die Software Komposita und Phrasen mit linguistischen Verfahren und übernimmt die Elemente als Synonyme, Ober-, Unter- und Differenzbegriffe in die Ontologie. Mit statistischen Verfahren ermittelt die Software Begriffspaare mit semantischer Nähe, das heißt Begriffe, die eine ähnliche Bedeutung haben wie „Bauen im Bestand“ und „Sanierung“. Dazu sucht die Software beispielsweise nach Ausdrücken, die in einem 30 Wörter umfassenden Textfenster häufig gemeinsam auftreten. Sie stellt anschließend die Richtung einer solchen Verbindung fest: Taucht das Wort „Bauen im Bestand“ fast immer zusammen mit „Sanierung“ im Text auf, „Sanierung“ aber auch in Verbindung mit vielen anderen Fachvokabeln, so ist das spezielle Thema „Bauen im Bestand“ Teil des allgemeinen Themas „Sanierung“.

Dem Anwender präsentiert sich Conweaver als Suchmaschine mit einem simplen Eingabefeld: „Kunden denken immer ausschließlich vom Interface her. Sie suchen so etwas Einfaches wie Google“, sagt Dr. Thomas Kamps, der die Entwicklung von Con-Weaver leitet. Die Bedienoberfläche präsentiert aber nicht nur

eine nach Relevanz sortierte Liste von zehn Dokumenten, sondern Einträge in verschiedenen Kategorien wie „Personen“, „Projekte“ und „Fachliteratur“. Neben der Hauptstoßrichtung, die dem Nutzer auf der linken Seite genau verrät, wonach er sucht, erhält er auf der rechten Seite über ähnliche Dinge die Möglichkeit, davon abzuweichen.

Wohin gehts?

Systeme, die kontextbezogene Rückfragen stellen oder Experten benennen, sind vielfach im Einsatz. Im ersten Fall ersetzen sie, im zweiten finden sie einen Fachmann. Einer der wichtigsten Informationsträger bleibt der Mensch: Die Informationssuche im Unternehmen endet fast immer mit dem richtigen Ansprechpartner. Ontologien besetzen eine Marktlücke, da sie den Experten ausfindig machen können. Als System zur umfassenden Beschreibung der Welt, wie es mit dem Semantic Web postuliert wird, mögen Ontologien (noch) als untauglich gelten. In eingegrenzten Aufgabebereichen sind sie klassischen Suchmethoden überlegen.

Klassifikationssysteme bringen Ordnung in große Textsammlungen. Bei der automatischen Verteilung von Briefpost in großen Unternehmen reichen wenige Klassen wie „Rechnung“, „Mahnung“, „Beschwerde“. Die Kategorisierung in 1200 Klassen

im FAZ-Archiv ergänzt das klassische Retrieval. Allein mit boolescher Suche ließe sich in den Millionen von Artikeln nicht sinnvoll recherchieren.

Hilfsmittel, die ähnliche Begriffe als semantische Wolken auf den Schirm bringen, und Programme, die Dokumente ähnlich gruppieren, zeigen Zusammenhänge auf, die eventuell auch dem versierten Archivar verborgen bleiben. Sie bringen auf Ideen und ermöglichen das Surfen auf einem semantischen Netz.

Im Unterschied zu Verfahren, die mit langen Trefferlisten von zweifelhafter Relevanz aufwarten, tragen die vorgestellten Verfahren einem erweiterten Informationsbedürfnis Rechnung. Es resultiert aus knapp kalkulierter Arbeitszeit und einer exponentiell steigenden Informationsmenge. Die Lage der Dinge verlangt nach effizienten Algorithmen.

Der Nutzer wünscht sich, was unter anderem die Suchmaschine Google so populär machte: eine einfache Bedienoberfläche. Trotzdem soll unter der Motorhaube alles Nötige stecken, um ihn mit einer perfekt auf seine Intention abgestimmten Antwort zu versorgen. Bei den vorgestellten Verfahren handelt es sich um gut funktionierende, aber nichtsdestotrotz um Nischenlösungen. Bis diese auch dem Nutzer im Web zur Verfügung stehen, wird noch einige Zeit vergehen. (akr)