

The Ontological Engineering Initiative (KA)²

Ausarbeitung zum Vortrag am 6.Mai 2003, Seminar Ontologien

Tobias Bosch

Einleitung

Diese Ausarbeitung fasst den Artikel "The Ontological Engineering Initiative (KA)² [KA2] zusammen.

Die Abkürzung (KA)² steht für "Knowledge Annotation Initiative of the Knowledge Acquisition Community", also eine Initiative der Gemeinde von Forschern, die sich mit dem Wissenserwerb beschäftigt, und wird von ca. 30 Forschern durchgeführt. (Im Folgenden wird die "Knowledge Acquisition Community" durch "KA-Gemeinde" abgekürzt).

Die Initiative wurde gegründet, um die folgenden drei Punkte näher zu untersuchen: Zunächst soll die Stärke und Rolle von Ontologien beim intelligenten Zugriff auf Informationen im WWW untersucht werden. Intelligenter Zugriff steht für die Autoren als Gegensatz zum Schlüsselwort-basierten Informations-Retrieval. Die Autoren bemängeln, dass Suchverfahren, die letzteres benützen, oft zu viele unerwünschte Dokumente liefern oder dass erwünschte fehlen. Außerdem erhofft sich die Initiative neue Erkenntnisse im Erstellen einer Ontologie in verteilter Art und Weise (verteilttes Ontologie-Engineering). Schließlich soll die Struktur der Gruppen und Themen der KA-Gemeinde noch besser herausgearbeitet werden, wodurch Kooperationen zwischen verschiedenen Gruppen einfacher würden, Kräfte vereint und die Wiederholung von Arbeiten verhindert werden könnten.

Dazu führt sie eine Fallstudie durch, in der zunächst eine Ontologie für die KA-Gemeinde erstellt werden soll. Anschließend sollen die WWW-Seiten der Gemeinde durch Terme aus dieser Ontologie annotiert werden und schließlich soll ein Webcrawler erstellt werden, der auf der Basis der annotierten WWW-Seiten und der Ontologie eine Anfrage intelligent durch logisches Schließen beantworten kann.

Die Umsetzung jeder dieser drei Teile der Fallstudie wird im Folgenden in einem separaten Abschnitt genauer beschrieben werden.

Erstellung der Ontologie für die KA-Gemeinde

Bedeutung von Ontologie

Die Autoren des (KA)²-Artikels verstehen unter einer Ontologie ein Modell für eine Domäne, das mit Hilfe einer logischen Sprache ausgedrückt wird. Es stellt eine bestimmte Sicht auf diese Domäne dar, die von vielen (möglichst allen) Menschen geteilt wird, die mit ihr zu tun haben. Eine Ontologie unterscheidet sich von einer Taxonomie darin, dass sie außer einer Klassenhierarchie auch noch Axiome enthält. Im Artikel wird zwar auch der Begriff der Konzeptualisierung erwähnt, er wird aber in einer ähnlichen Art und Weise wie der der Ontologie benützt, wobei der Unterschied nicht erklärt wird.

Nicola Guarino [GUA98] bezeichnet den hier verwendete Art des Begriffes Ontologie als "ontology (with the lowercase 'o')"... "in the philosophical community"¹. Er unterscheidet aber den Begriff der Konzeptualisierung von dem der Ontologie dadurch, dass die Konzeptualisierung einer Domäne die implizite Struktur und die Zusammenhänge der

1 Siehe [GUA98], Kapitel 2, 2. Absatz

Domäne ist, wogegen eine Ontologie diese Konzeptualisierung in einer logischen Sprache explizit zu formulieren versucht.

Vorgehen

Die Ontologie soll in einem verteilten Entwicklungsprozess erstellt werden. Dafür benützt die Initiative Ontolingua. Ontolingua ist eine interaktive Umgebung zum Entwickeln von Ontologien. Ontologien, die in Ontolingua geschrieben werden, benützen die Frame Ontology als Repräsentationssprache. Diese ist eine Frame-basierte Sprache, die wiederum auf KIF basiert, und besitzt als Primitiven die Ausdrücke Klasse, Subklasse, Attribut, Wert, Relation und Axiom. Es existiert auch ein Server zur verteilten Benutzung von Ontolingua (über ein Netzwerk wie das Internet), der auch einen speziellen Editor, welcher bei der Entwicklung behilflich ist, bereitstellt. Da dieser aber bei einer langsamen Netzwerkverbindung nicht geeignet ist, um große Änderungen in der Ontologie einzugeben, wird zusätzlich ODE (ontology design environment), ein Frontend für die Offline-Erstellung von Code für Ontolingua, eingesetzt.

Bei der Erstellung wurde entschieden, nur die relevantesten Konzepte der KA-Gemeinde zu modellieren, um sich nicht in Details zu verlieren und damit die Annotation von WWW-Seiten mit Ausdrücken der Ontologie möglichst einfach ist. Deshalb wurde auch keine vorgegebene Ontologie von Ontolingua benützt (wegen ihrer Detailtiefe), sondern eine eigene erstellt. Außerdem wurde die Ontologie in 7 separate Subontologien aufgeteilt, um Modularität und Wiederverwendung zu gewährleisten (z.B. Person-Ontologie, ...).

Der Entwurf der Ontologie ist auf einem Workshop (9th KEML workshop, Germany, Jan 1998) erstellt worden. Dort wurden auch von den Gruppen der KA-Gemeinde jeweils Experten eingeladen, um an der Initiative teilzunehmen. Diese bekamen dann die Aufgabe, diese begonnene Ontologie zu vervollständigen.

Annotation von WWW-Seiten durch Ausdrücke der Ontologie

Nachdem die wichtigsten Klassen und Axiome der Ontologie in Ontolingua erstellt worden sind, wird damit begonnen, WWW-Seiten zu annotieren. Durch eine solche Seite kann aber nur eine Instanz einer Klasse der Ontologie beschrieben werden, die Klassen selber oder auch Axiome müssen in Ontolingua erstellt werden.

Instanzen können also auf zwei Arten erstellt werden: Zum einen direkt in Ontolingua (über den Server, oder per ODE), oder durch annotierte WWW-Seiten. Das ist wichtig zu bemerken, da die Instanzen, die durch die WWW-Seiten gebildet werden, nicht zu der Ontolingua-Ontologie hinzugefügt werden (auch nicht im Laufe der Zeit), sondern vom Webcrawler separat betrachtet werden (siehe unten). Diese Klarstellung fehlt im Artikel und führt beim Lesen leicht zu Verständnisproblemen.

Für die Annotation wird dem Anker-Tag ("`<a>`") innerhalb von HTML-Seiten das neue Attribut "ONTO" hinzugefügt. Dieses wird von Browsern ignoriert, erlaubt aber zum einen die Angabe der Klasse, für die durch diese Seite eine Instanz erstellt werden soll, zum anderen die Zuordnung des Inhalts und der anderen Attribute des Tags zu Werten von Attributen dieser Klasse. Somit kann man die Daten der WWW-Seite, die für die Darstellung benützt werden, auch für die Definition der Attributwerte der Instanz wiederverwenden.

Diese Wiederverwendbarkeit wird in dem Artikel als entscheidendes Argument für die Einsetzung des ONTO-Attributes gegen RDF angegeben. RDF ist eine Spezifikation des W3C zur Angabe von Semantik für WWW-Dokumente, die zu der Zeit allerdings noch kein ausgereifter Standard war. In einem Folgeprojekt wird diese dann doch verwendet, wohl auch wegen ihrer Fähigkeit, allgemein XML-Dokumente zu beschreiben, was der Ansatz

der Initiative nicht leistet.

Erstellung eines Webcrawler zum Schließen auf der Ontologie und den annotierten WWW-Seiten.

Zum Schließen auf der Ontologie und den annotierten WWW-Seiten wird ein System mit Namen "Ontobroker" eingesetzt. Es besteht aus einem Webcrawler ("Ontocrawler"), einem Schließ-Mechanismus und einem Anfrage-Interface.

Innerhalb von Ontobroker wird Frame Logic (FLogic) als Repräsentationssprache eingesetzt. Hierbei handelt es sich um ein Framework mit Präd.-Logik 1.Stufe, das zusätzliche Konzepten wie Objekt, Relation, Attribut und Klasse besitzt. FLogic wird nicht für den verteilten Entwurfsprozess der Ontologie benützt, da es keinen speziellen Server und auch keine Entwicklungsumgebung dafür gibt. Außerdem ist es nicht so bekannt wie Ontolingua. Da die Ontologie aber in Ontolingua erstellt wurde, benötigt man noch Übersetzung zwischen der Sprache von Ontolingua und FLogic.

Die Aufgabe von Ontocrawler ist es, die annotierten WWW-Seiten, die bei ihm registriert wurden, zu durchsuchen und sie in FLogic für den Schließ-Mechanismus zu übersetzen.

Das Anfrage-Interface besteht aus aktiven HTML-Seiten und CGI-Skripten, die die Anfragen von Benutzern, die diese per Browser stellen, entgegen nehmen, sie an den Schließ-Mechanismus weiterleiten und schließlich das Ergebnis an den Benutzer zurückgeben. Die meisten Anfragen kann man auch benutzerfreundlich über einen Graphen oder Tabellen eingeben.

Der Schließ-Mechanismus ist kein spezifischer Ontologie-Mechanismus und besitzt zwei Informationsquellen: Die Ontologie, die aus Ontolingua in FLogic übersetzt worden ist, und die in FLogic übersetzten annotierten WWW-Seiten.

Ausblick

Es ist zu bemerken, dass heute (2003) die Initiative abgeschlossen ist. Leider funktionieren die im Artikel angegebenen Links nicht mehr, die erstellte Ontologie ist aber trotzdem noch zugreifbar unter <http://ka2portal.aifb.uni-karlsruhe.de>. Es gibt auch ein Nachfolgeprojekt von der Universität Karlsruhe, die "Karlsruhe Ontology and Semantic Web Tool Suite" (KAON). Diese ist als Open-Source-Software unter <http://kaon.semanticweb.org> erhältlich und unterstützt sogar die halbautomatische Erstellung von Ontologien durch Text-Mining-Algorithmen, die in dem Artikel noch als Ausblick genannt wurde. Außerdem benützt es RDF und ist somit auch für XML-Dokumente geeignet.

Literatur

GUA98: Nicola Guarino (1998). Formal Ontology and Information Systems. In N. Guarino (ed.), Formal Ontology in Information Systems. Proceedings of the First International Conference (FOIS'98), June 6-8, Trento, Italy (pp. 3-15). Amsterdam: IOS Press

KA2: V. Richard Benjamins, Dieter Fensel (1998). The Ontological Engineering Initiative (KA)2. In N. Guarino (ed.), Formal Ontology in Information Systems. Proceedings of the First International Conference (FOIS'98), June 6-8, Trento, Italy (pp. 287-301). Amsterdam: IOS Press