

Ontologien

Seminar AI Tools
Wintersemester 2006/2007

Universität des Saarlandes

Christian Schaadt
Christian.Schaadt@t-online.de

Dozenten:

Dr. Dominik Heckmann

Dr. Michael Kipp

Dr. Alassane Ndiaye

Michael Feld

Abstract

Die hier vorliegende Ausarbeitung soll dem Leser ein grundlegendes Wissen über Ontologien ermöglichen. Dabei liegt der Fokus zunächst auf dem Aufbau und den Möglichkeiten von Ontologien. Die Vorteile reichen dabei von Wiederverwendung von Wissen, über Austausch von Wissen zwischen Agenten und/oder Menschen, bis hin zu erweiterten Analysen von Informationen. Das Hauptaugenmerk des Textes liegt danach auf der beispielhaften Erstellung einer Ontologie gemäß einem in sieben Schritten unterteilten Design-Prozess. Dieser Prozess stellt zunächst den Umfang des Themengebietes fest und versucht bestehende Ontologien wieder zu verwenden. Anschließend werden Begriffe gesammelt, in einer Hierarchie angeordnet und mit Eigenschaften versehen. Das Erstellen der Ontologie endet mit dem spezialisieren dieser Eigenschaften und dem Kreieren konkreter Objekte. Um dem Leser einen Anreiz und Ausblick zu ermöglichen, stellt diese Ausarbeitung einige größere und frei verfügbare Ontologien vor.

Inhaltsverzeichnis

Abstract	2
Inhaltsverzeichnis	3
1. Einleitung.....	4
Sinn von Ontologien.....	5
Was ist eine Ontologie?	5
2. Erstellen einer Ontologie.....	6
Umfangsbestimmung.....	7
Wiederverwenden von Ontologien.....	8
Begriffssammlung	8
Klassenhierarchie	9
Klasseneigenschaften.....	10
Klasseneinschränkungen.....	10
Instanzen.....	11
3. Große Ontologien	12
SUMO	12
WordNet.....	13
Open Cyc	13
GUMO	14
4. Zusammenfassung.....	15
Literaturverzeichnis.....	16

1. Einleitung

Das Problem, Wissen einer Maschine beziehungsweise einem Agenten semantisch zugänglich zu machen, hat mit dem Aufkommen der Idee des „Semantic Web“ an Wichtigkeit zugenommen. Eine auch für den Menschen leicht verständliche Möglichkeit, dieses Wissen zwischen Mensch und Maschine zu teilen, sind Ontologien. Formal definiert ist eine Ontologie „eine explizite formale Spezifikation einer gemeinsamen Konzeptualisierung“¹. Leichter verständlich wird der Begriff der Ontologie, wenn man von einem „formal definierten System von Begriffen und Relationen“ zusammen mit „Inferenz- und Integritätsregeln“ spricht². Dieses System besteht grundlegend aus einer hierarchischen Anordnung der Begriffe oder Klassen. Diese Klassen sind durch die Relationen miteinander verbunden, welche wiederum durch die Regeln eingeschränkt werden.

Der vorliegende Text beschäftigt sich nun mit der Frage, wie man auf einem einfachen Weg eine solche Ontologie erstellen kann. Hierzu wird zunächst betrachtet, wozu eine Ontologie benutzt werden kann, und dann in sieben klar strukturierten Schritten eine Bauanleitung gegeben, mit der man leicht und intuitiv zu einer eigenen Ontologie gelangt³.

Am Anfang eines solchen Design-Prozesses stehen eine Umfangsbestimmung des Themas der Ontologie und der Versuch, Teile von bestehenden Wissensnetzen wieder zu verwenden. Danach werden zum Gebiet passende Begriffe gesucht und es wird gezeigt, wie man diese Begriffe als Klassen hierarchisch anordnet und die Klassen mit Eigenschaften verbindet. Nachdem Regeln für die Eigenschaften festgelegt wurden, können noch einzelne von den Klassen abgeleitete Instanzen erstellt werden.

Zum Abschluss gibt diese Ausarbeitung einen Einblick in verschiedene größere, bereits bestehende Ontologien, die frei im Web verfügbar sind. Zu diesen Wissensrepräsentationen gehört zum einen die SUMO Ontologie, die sich besonders mit natürlicher Sprachverarbeitung beschäftigt und sich in viele Richtungen spezialisieren lässt. Eine ähnlich aufgebaute Struktur besitzt die sehr umfangreiche OpenCyc, die Agenten „gesunden Menschenverstand“ ermöglichen möchte. Desweiteren soll ein kurzer Einblick in die Wissensbasis von WordNet erfolgen, deren Hauptaugenmerk auf den lexikalischen und semantischen Beziehungen der Wörter untereinander liegt.

Sinn von Ontologien

Wie schon in der Einleitung beschrieben ist ein großes Ziel von Ontologien, ein gemeinsames Verständnis von Wissen zwischen Menschen und/oder Agenten zu ermöglichen. Existieren beispielsweise mehrere Webseiten, die alle Informationen zum Thema Medizin zur Verfügung stellen, und basieren diese Seiten alle auf der gleichen Ontologie, so können Agenten das Wissen aus diesen verschiedenen Seiten extrahieren und verwenden.

Durch Ontologien erhält man weiterhin die Möglichkeit, Wissen leicht wieder zu verwenden. Grundlegende Begriffe, wie zum Beispiel das Konzept der Zeit, werden in vielen Gebieten benötigt. Wurde dieses Konzept einmal ausführlich, das heißt mit allen Informationen über Zeitpunkte, Zeitabschnitte und Zeitmaßen, definiert, kann es von neuen Ontologien eingebunden werden und muss somit nicht neu designt, implementiert und getestet werden.

Desweiteren kann man durch Verwenden von Ontologien Wissen und Anwendungen voneinander trennen. Angenommen wir verfügen über ein Programm, das eine Maschine anhand eines Bauplans aus ihren Komponenten zusammenbaut. Erhält das Programm die Bauanleitung in Form einer Ontologie, würde es keinen Unterschied machen, welche Art von Ontologie wir dem Algorithmus zur Verfügung stellen. Eine Auto-Ontologie würde uns genau wie eine Fahrstuhl-Ontologie das entsprechende Ergebnis liefern.

Schließlich erlauben Ontologien eine intensive Analyse von Wissen. Viele Problemlösungsmethoden benutzen Ontologien als grundlegende Struktur. Auch können verschiedene Suchverfahren aufgrund des Netzwerk-Charakters der Ontologie leicht angewandt werden.

Die Voraussetzung dafür, dass diese Vorteile genutzt werden können, liegt im Design der Ontologie. Aus diesem Grund wird in den nachfolgenden Kapiteln ein möglicher Aufbau eines Designprozesses näher beleuchtet.

Was ist eine Ontologie?

In der Literatur gibt es verschiedene Definitionen von Ontologien. Im Sinne dieses Dokuments ist eine Ontologie eine formale, explizite Beschreibung von Konzepten in einem bestimmten Themengebiet. Diese Konzepte werden auch Klassen genannt. Dazu kommen Eigenschaften für jedes Konzept, die die verschiedenen Attribute der Klasse beschreiben (bezeichnet werden diese Eigenschaften als „slots“ des Konzepts), und schließlich noch Einschränkungen der „slots“ (die sogenannten „facets“). Eine Ontologie zusammen mit einer Menge von Instanzen bildet eine Wissensbasis.

Klassen sind der Grundstock jeder Ontologie. Sie beschreiben Konzepte aus dem Themengebiet. Zum Beispiel repräsentiert eine Wein-Klasse alle Weine. Spezifische Weine würden durch Instanzen dieser Klasse dargestellt werden. Der Rotwein im Glas beim Mittagessen ist somit eine Instanz des Rotwein-Konzeptes. Weiterhin können Klassen Unterklassen besitzen, die Konzepte repräsentieren, die

spezifischer sind als die Oberklasse. So kann man der Oberklasse ‚Wein‘ zum Beispiel die Unterklassen ‚Rotwein‘ und ‚Weißwein‘ zuordnen.

Die „Slots“ beschreiben Eigenschaften von Klassen und Instanzen. Betrachten wir zum Beispiel: „Dieser Wein ist trocken und kommt aus dem Nahe-Gebiet“. In diesem Fall hätten wir zwei Eigenschaften, die den Wein beschreiben. Zum einen den Zuckergehalt (trocken) und die Herkunftsregion (Nahe-Gebiet). Genauso könnte man der Wein-Klasse die Slots Geschmack, Körper und Hersteller zuweisen.

Alle Instanzen und Unterklassen des Wein-Konzeptes erben die Eigenschaften ihrer Oberklasse. Somit verfügen auch alle spezifischen Weine über einen Hersteller-Slot, der als Wert natürlich nur eine Winzerei annehmen darf. In diesem Fall haben wir also eine Begrenzung einer Eigenschaft (facet), nämlich die Einschränkung der Hersteller-Eigenschaft auf Instanzen der Winzerei-Klasse.

Mit dieser Betrachtung einer Ontologie erkennt man auch gleich die Kernpunkte, die beim Erstellen einer solchen wichtig sind. Man muss die Klassen der Ontologie definieren, diese Klassen in einer Taxonomie (das heißt Ober- und Unterklassen) anordnen, Eigenschaften definieren, erlaubte Werte für diese Slots einschränken und schließlich Instanzen zu den Klassen erstellen. Auf diese verschiedenen Schritte wird nun etwas näher eingegangen.

2. Erstellen einer Ontologie

Bevor wir mit den eigentlichen Schritten des Design-Prozesses beginnen, sollen hier noch einige fundamentale Regeln aufgezeigt werden, die wie ein roter Faden durch das gesamte Design führen.

„Es gibt nicht den einen richtigen Weg, eine Ontologie zu erstellen – es gibt immer mögliche Alternativen“. Die beste Lösung des Prozesses hängt auch immer vom geplanten Einsatz der Ontologie ab. Desweiteren wird hier nur eine Möglichkeit vorgestellt, andere Methoden kommen sehr wahrscheinlich zu einem unterschiedlichen Ergebnis, das nicht weniger gut sein muss.

„Ontologie-Design ist immer auch ein iterativer Prozess“. Nur in den seltensten Fällen wird die erste Version der Ontologie ein perfektes Ergebnis liefern. Oft müssen noch neue Klassen eingebaut werden, verschiedene Eigenschaften anders eingeschränkt oder auch das komplette Konzept überarbeitet werden.

„Klassen in der Ontologie sollten an Begriffe aus dem Betrachtungsraum angelehnt sein“. Man sollte daran denken, dass die Ontologie eine Beschreibung der realen Welt darstellen soll. Konzepte zu wählen, die wirkliche Objekte wiederspiegeln, erleichtert den Design-Prozess ungemein.

Behält man diese goldenen Regeln im Hinterkopf, so sollte das Erstellen der Ontologie anhand der folgenden Schritte keine zu großen Probleme aufwerfen. Anhand dieser Schritte wird nun der inkrementelle Aufbau einer Ontologie über Wein und Essen skizziert.

Umfangbestimmung

Es bietet sich an in einem ersten Schritt den Umfang und den Bereich der Ontologie zu bestimmen beziehungsweise festzulegen. Dazu ist es hilfreich, sich verschiedene Fragen zu beantworten.

Was / Welchen Bereich soll die Ontologie abdecken?

Wer wird die Ontologie benutzen / weiterführen?

Für was soll die Ontologie benutzt werden?

Welche Fragen soll die Ontologie beantworten können?

Die Antworten auf diese Fragen können sich während des Design-Prozesses verändern. Man kann aber jederzeit auf sie zurückgreifen, um sich einen Überblick über den Umfang der Ontologie zu verschaffen.

In unserem Beispiel soll die Ontologie den Bereich „Wein & Essen“ abdecken. Die Ontologie könnte beispielsweise in einer Restaurant-Software benutzt werden, mit der man entscheiden kann, welche Weine besonders gut zu welchem Essen passen. Diese Wahl hätte zur Folge, dass Informationen über chemische Zusammensetzungen der Weine in der Ontologie keine Rolle spielen würden. Würden wir die Ontologie dagegen für eine Weinzeitschrift verwenden, sollten wir vielleicht auch sprachliche Eigenschaften in die Ontologie einbinden.

Ein sehr nützliches Mittel, um den Umfang genauer zu bestimmen, sind die sogenannten „Competency Questions“. Hierfür bildet man eine Liste mit Fragen, deren Antwort durch die Ontologie gegeben sein soll. Beispiele für solche Fragen könnten sein:

Welche Eigenschaften sollte man bei einem Wein beachten?

Ist Dornfelder ein Rot- oder ein Weißwein?

Passt ein Riesling zu Meeresfrüchten?

Welcher Wein passt am besten zu Grillfleisch?

Was waren gute Jahrgänge für Spätburgunder?

Eine solche Fragen-Sammlung besitzt zwei Vorteile. Zum einen kann man so vor dem Erstellen der Ontologie sehr schnell den Umfang abgrenzen, zum anderen lassen sich die Fragen im Nachhinein als sinnvollen Test verwenden, ob die Ontologie auch alle gewünschten Informationen bereitstellt.

Wiederverwenden von Ontologien

Wie bereits angesprochen ermöglichen Ontologien die Wiederverwendung von bereit gestelltem Wissen. Man muss sich deshalb die Frage stellen, wieso man das Rad neu erfinden sollte, wenn dies bereits ein anderer getan hat.

Das Wiederverwenden von Ontologien ist dabei sehr einfach und bietet eine Reihe von Vorteilen: Möchte man seine Anwendung an ein anderes Programm ankoppeln, so ist es sinnvoll die gleiche Ontologie zu benutzen und die Anbindung so enorm zu vereinfachen. Weiterhin spart die Einbindung einer bereits existierenden Wissensbasis eine Menge an Zeit und Geduld. Man kann so oft auf den Design-Prozess für grundlegende Teile seiner eigenen Ontologie verzichten und erhält zudem eine Struktur, auf der sich aufbauen lässt. Schließlich darf man nicht vergessen, dass vor allem große bestehende Ontologien in der Regel getestet sind und durch die Anbindung somit keine zusätzlichen Probleme auftreten.

Es stellt sich nun natürlich die Frage, woher man solche Ontologien bekommen kann. Es gibt im Internet eine Vielzahl bereits existierender, frei verfügbarer Ontologien, beispielsweise auf der ‚Ontolingua Ontology Library‘⁴ oder auf der ‚DAML Ontology Library‘⁵. Auf diesen Seiten findet man Ontologien fast zu allen erdenklichen Themengebieten.

Sollte man nun noch nicht den kompletten Bereich der gewünschten Ontologie abgedeckt haben, so muss man allerdings selbst den Rest des Design-Prozesses in die Hand nehmen.

Begriffssammlung

Dazu bietet sich nun an, alle Klassen zu finden, die in der Ontologie vorkommen sollen. Aus diesem Grund erstellen wir in einer Art Brainstorming eine Liste mit Begriffen, die Anteil an unserem Themengebiet haben.

In dieser Liste sollten sich alle Substantive wiederfinden, über die wir eine Aussage treffen wollen oder die wir erklären möchten. Desweiteren suchen wir alle Eigenschaften, durch die die bereits gefundenen Begriffe beschrieben werden oder die die gefundenen Begriffe besitzen. Viele dieser Eigenschaften lassen sich durch Verben beschreiben. Hilfreich kann es hier sein, sich zu fragen, was man noch über die einzelnen Begriffe der Ontologie aussagen möchte. Ziel sollte es sein, eine möglichst umfangreiche Liste von Begriffen zu erhalten. Es ist allerdings nicht weiter schlimm, wenn die ein oder andere Klasse zu Beginn des Prozesses fehlt, da im Laufe des Designs immer wieder Änderungen an bereits abgehandelten Schritten durchzuführen sind. Wie schon früher angedeutet ist die gesamte Entwicklung der Ontologie ein iterativer Prozess.

Betrachten wir unser Beispiel einer Wein-Ontologie, so könnten wir folgende Begriffe in die Liste aufnehmen: Wein, Traube, Winzerei, Rotwein, Herkunft, Farbe, Geschmack, Zuckergehalt, Hersteller, Produkte, Körper, Fleisch, Fisch und viele andere.

Klassenhierarchie

Im nächsten Schritt ist es nun das Ziel aus den so gefundenen Klassen eine taxonomische Hierarchie aufzubauen, das heißt die Klassen in einem System von Ober- und Unterklassen anzuordnen.

Hierzu suchen wir zunächst in der Liste alle Begriffe, die mit einem Objekt in der realen Welt korrespondieren. Diese Begriffe sollen nun in der Hierarchie angeordnet werden. Dazu bieten sich verschiedene Möglichkeiten an.

Die Top-Down-Methode. Diese Methode startet mit der Definition des allgemeinsten Konzepts im Themengebiet. Danach wird diese Klasse sukzessive mit Unterklassen spezialisiert. Die allgemeinste Klasse, die wir in unserem Beispiel finden, ist mit Sicherheit die Wein-Klasse. Wir suchen nun alle direkten Unterklassen dieses Konzepts und kommen so zu den Klassen Weißwein und Rotwein. Wurde diese Ebene komplett erschlossen, führen wir die Methode rekursiv für die nächste Ebene durch.

Die Bottom-Up-Methode. Wie der Name schon vermuten lässt beginnen wir diesmal mit den spezifischsten Klassen, die die Blätter unserer Ontologie bilden werden. Diese Klassen werden nun gemäß gemeinsamer Eigenschaften gruppiert und für jede Gruppe eine entsprechende Oberklasse gesucht. In der Wein-Ontologie kann man zum Beispiel mit den Klassen Dornfelder und Spätburgunder beginnen. Beide Weine sind Rotweine, wir können also beide Konzepte als Unterklassen zur Rotwein-Klasse definieren. Auch diese Methode wird rekursiv Ebene für Ebene durchgeführt.

In der Praxis bieten sich auch Kombinationen aus Top-Down und Bottom-Up an. Nicht immer ist eine konsistente Klassenhierarchie einfach zu finden und ein Ansatz, bei dem sich die beiden Methoden auf einer Mittelebene treffen, könnte hilfreich sein. Oft hängt die Wahl der Methode aber auch vom persönlichen Geschmack und der subjektiven Erfahrung des Ontologie-Designers ab.

Beim Erstellen der Klassenhierarchie sind noch einige Regeln von Bedeutung, die dabei helfen, eine gute Taxonomie zu erstellen.

Synonyme. Man sollte keine neue Klasse für ein Synonym eines Begriffes erstellen. In einer Ontologie repräsentieren die Klassen ein Konzept aus dem Themengebiet und nicht das Wort, das für dieses Konzept steht.“

Transitivität. Eine nützliche Methode, um zu testen ob eine Klassenhierarchie korrekt ist, ist zu überprüfen, ob die Transitivität der Taxonomie gewährleistet ist. Hierbei muss immer gelten: „Ist B Unterklasse von A und C Unterklasse von B, so ist C Unterklasse von A“.

Zyklen. Beim Design der Klassenhierarchie sollten Zyklen vermieden werden. Ein Zyklus liegt vor, wenn eine Klasse sowohl Unter- als auch Oberklasse eines anderen Konzeptes ist. Hat man einen Zyklus in der Taxonomie gefunden, so kann dies zwei Ursachen haben. Zum einen kann ein Design-Fehler vorliegen, das heißt eine Oberklasse-Unterklasse-Verbindung ist nicht richtig. Zum anderen

könnten alle Klassen im Zyklus äquivalent sein, und man könnte diese Konzepte durch ein einziges ersetzen.

Klasseneigenschaften

Die gerade festgelegte Klassenhierarchie wird alleine nicht ausreichen, um alle ‚Competency Questions‘ aus Schritt 1 zu beantworten. Wir müssen also weitere Aussagen über die interne Struktur der Konzepte, also die Eigenschaften der Klassen, treffen.

Nachdem im vorigen Schritt die Klassen aus unserer Begriffssammlung extrahiert wurden, sind die übrigen Begriffe in der Liste gute Kandidaten für die Klasseneigenschaften. Im Falle der Wein-Ontologie sind beispielsweise Herkunft, Farbe, Geschmack, Zuckergehalt, Hersteller und Produkte.

Aufgabe ist es nun, diese Eigenschaften den passenden Klassen zuzuordnen. Wir können so ohne Weiteres die Weinklasse mit den Slots Farbe, Geschmack, Zuckergehalt und Hersteller versehen. Dabei sollte man darauf achten, dass manche Eigenschaften auch mehrere Konzepte beschreiben können. Den Slot Herkunft kann man hier sowohl der Wein-Klasse als auch der Winzerei-Klasse zuordnen, eine Eigenschaft, die den Namen eines Konzeptes beschreibt, kann fast überall verwendet werden.

Klasseneinschränkungen

Wie schon im Kapitel „Was ist eine Ontologie?“ beschrieben, ist es oft notwendig und sinnvoll, eine Eigenschaft einzuschränken. Hierbei gibt es eine Vielzahl solcher Einschränkungen oder Facets. Die am häufigsten verwendeten sind Aussagen über Kardinalität des Slots, Werte der Eigenschaft und über Domain und Range des Slots.

Kardinalität. Die Kardinalität eines Slots gibt an, wie viele verschiedene Werte eine Eigenschaft annehmen kann. Hier wird oft nur zwischen einfacher Kardinalität (erlaubt höchstens einen Wert) und mehrfacher Kardinalität (erlaubt jede Anzahl von Werten) unterschieden. In diesem Zusammenhang hätte die Hersteller-Eigenschaft des Weines eine einfache Kardinalität, denn ein Wein hat immer genau einen Hersteller. Im Gegensatz würde man der Produkte-Eigenschaft einer Winzerei eine mehrfache Kardinalität zuweisen, um zu beschreiben, dass die Winzerei mehrere Weine herstellt.

Werte der Eigenschaft. Diese Einschränkung beschreibt, welche Art von Werten eine Eigenschaft annehmen darf. Die hierfür meist verwendeten Typen reichen von Strings, die sich besonders für Eigenschaften wie Namen eignen, über Zahlen, die nützlich für Jahrgangs- oder Preis-Slots sind, bis hin zu Aufzählungen, womit alle möglichen Werte für eine Eigenschaft festgelegt werden. Schließlich kommen noch Instanzen anderer Klassen als Werte in Frage.

Hierdurch können die Beziehungen der verschiedenen Konzepte auf einfache Weise beschrieben werden.

Domain und Range. Wurden die Werte einer Eigenschaft auf Instanzen eingeschränkt, so sollte man zu dieser Eigenschaft auch Domain und Range festlegen. Die Range bezeichnet in diesem Zusammenhang alle Klassen, deren Instanzen als Wert in Betracht kommen. Die Menge aller Klasse, auf die die Eigenschaft angewendet werden kann, wird als Domain bezeichnet. Betrachten wir als Beispiel den Slot Herkunft. Als Range bietet sich hier die Klasse Region an, als Domainklassen haben wir unter anderem die Konzepte Wein und Winzerei.

Instanzen

Nachdem nun die Einschränkungen für die Eigenschaften festgelegt wurden, ist man prinzipiell am Ende des Ontologie-Design-Prozesses angelangt. Es ist aber sinnvoll, die Ontologie durch Instanzen zu einer Wissensbasis zu erweitern. Diese Entscheidung hängt dabei nicht unwesentlich von den im ersten Schritt (siehe Umfangsbestimmung) gewonnenen Informationen ab. Desweiteren kann man auch anhand der Competency Questions und an der Art der gewünschten Anwendung recht gut erkennen, ob eine Erweiterung zu einer Wissensbasis notwendig ist. Wollen wir unsere Ontology in einer Restaurant-Anwendung einsetzen, die entscheidet, welcher Wein zu welchem Essen passt, ist die Erstellung einer Wissensbasis essentiell.

Das Kreieren von Instanzen ist dabei ein relativ einfacher Prozess. Um ein konkretes Objekt zu erstellen, wählt man erstens eine Klasse aus, erstellt dann eine leere Vorlage anhand dieser Klasse und füllt schließlich die Eigenschaften dieser Vorlage mit konkreten Werten aus. Nehmen wir beispielsweise unsere Wein-Klasse. Dieses Konzept besitzt die Eigenschaften Name, Farbe, Zuckergehalt, hergestellt aus, Hersteller, Geschmack und Herkunft. Wir erstellen die Instanz und füllen diese Eigenschaften mit Werten aus: Grünhäuser Riesling (Name), weiß (Farbe), süß (Zuckergehalt), Riesling (hergestellt aus), Grünhäuser (Hersteller), fruchtig (Geschmack), Saar (Herkunft). Dieser Weißwein ist nun ein konkretes Objekt der Wein-Klasse.

Hat man alle erforderlichen und gewünschten Instanzen erstellt, ist man am (offenen) Ende des Design-Prozesses angelangt. Oftmals wird es notwendig sein, dass die Wissensbasis gepflegt werden muss: Die Ontologie wird an aktuellere Informationen angepasst und im Nachhinein immer wieder neue Instanzen hinzugefügt.

3. Große Ontologien

Nachdem nun die einzelnen Schritte eines möglichen Ontologie-Design-Prozesses vorgestellt wurden, soll hier ein Einblick in bestehende, größere Ontologien gegeben werden. Hierbei liegt der Fokus auf den frei verfügbaren Ontologien SUMO, WordNet und OpenCyc; auch wird das Projekt GUMO vorgestellt, das von Dr. Dominik Heckmann an der Universität des Saarlandes entwickelt wird, vorgestellt.

Diese kurze Übersicht über andere Ontologien soll zwei Ziele verfolgen. Zum einen erhält man durch die Betrachtung von fremden Projekten Anregungen und Inspiration, die für das eigene Design von Vorteil sein können. Man darf nicht vergessen, dass der Entstehungsprozess einer Ontologie vor allem von der Erfahrung des Erstellers profitiert. Zum anderen kann man sich so einen Überblick über Ontologien verschaffen, die man in Schritt zwei (siehe ‚Wiederverwenden von Ontologien‘, Seite 8) in seine eigene einbinden kann.

SUMO

Eine der größten frei verfügbaren Ontologien ist SUMO, die vom IEEE-Konsortium (Institute of Electrical and Electronics Engineers) entwickelt wurde. SUMO steht für Suggested Upper Merged Ontology, das heißt eine Ontologie, die aus mehreren kleineren zusammengesetzt ist. Das Grundmodul von SUMO besteht aus 1500 Konzepten und 4000 Relationen und abgeleiteten Aussagen. SUMO besteht aber, wie schon angedeutet, aus vielen kleineren Modulen, die separat hinzugefügt werden, und der Ontologie so einen Gesamtumfang von 20000 Klassen und 60000 Aussagen ermöglichen.

SUMO wurde mit dem Ziel entwickelt, Programmen eine natürliche Sprachverarbeitung zu ermöglichen. Das bedeutet, dass auch Informationen der Objekte untereinander in den Verarbeitungsprozess einbezogen werden. Ein weiteres Ziel ist, Information Retrieval im Allgemeinen zu vereinfachen. Dies umfasst zum Beispiel das Suchen von Informationen in Datenbanken oder im Internet. Um diese Ziele mit semantischem Charakter zu ermöglichen, ist SUMO komplett an die WordNet-Ontologie angepasst, die auf den semantischen und lexikalischen Beziehungen der Wörter unter sich basiert (siehe ‚WordNet‘, Seite 13).

Durch den modularen Aufbau der SUMO Ontologie ist es möglich, das Wissen ganz gezielt in eine bestimmte Richtung zu erweitern. Die Module umfassen dabei viele erdenkliche Bereiche. Angeboten werden unter anderem Informationen über Kommunikation, Länder, Menschen, Militär, Geographie und Finanzen, aber auch Themen wie Mengentheorie, Zahlen, Physik, Graphentheorie, Zeit, Agenten und Maße.

Eine aktuelle Version von SUMO ist immer auf der offiziellen Homepage ⁶ verfügbar. Man hat zudem die Möglichkeit, sich die Ontologie online in einem Webbrowser anzeigen zu lassen ⁷.

WordNet

Ein ähnliches Ziel wie SUMO verfolgt die WordNet Ontologie. Genauer gesagt, möchte man mit WordNet natürlichsprachliche Texte für Computer verständlich machen. Dabei verfügt die Entwicklung der Princeton University allerdings einen ganz anderen Aufbau als SUMO. WordNet ist als lexikalische Datenbank ausgelegt, die allein auf den semantischen und lexikalischen Beziehungen zwischen den Wörtern basiert. Dies bedeutet, dass hier auf andere Eigenschaften zwischen den Konzepten verzichtet wurde. Man würde also Informationen, wie zum Beispiel welcher Wein zu welchem Essen besonders gut passt, in dieser Ontologie vergebens suchen. Dies soll aber nichts als Nachteil von WordNet, sondern nur als Abgrenzung von anderen Ontologien und zur Verdeutlichung des außergewöhnlichen Charakters dieser Wissensbasis, angesprochen werden.

WordNet verfügt über einen beachtlichen Umfang. Insgesamt werden an die 120000 Wörter in der Datenbank aufgeführt. Die Eigenschaften dieser Klassen beziehen sich sprachliche Konzepte, wie Synonyme, Antonyme und Oberbegriffe (Hyperonomie). Besonders interessant ist, dass die meisten dieser Wörter mit ihren Bedeutungen versehen sind. Damit kann WordNet auch als gutes Lexikon mit Zusatzinformationen benutzt werden.

Auch diese Ontologie ist frei im Internet auf der entsprechenden Seite verfügbar ⁸. Desweiteren wird noch eine Visualisierung angeboten, bei der man auch die beeindruckende Größe der Ontologie bewundern kann ⁹.

Open Cyc

OpenCyc ist die Open Source Variante der lizenzpflichtigen Cyc-Ontologie, die von Cycorp (LIT) entwickelt wird. Diese Open Source Version wurde veröffentlicht, damit sich die für die Ontologie gewählte Struktur als Standard in der Industrie durchsetzt. Der Name leitet sich von ‚encyclopedia‘ (zu Deutsch Lexikon) ab.

Entwickelt wird die Hauptontologie seit 1982 mit dem Ziel, Anwendungen logisches Schlussfolgern („Gesunden Menschenverstandes“) zu ermöglichen. Als interne Richtlinie wurde auch das Ziel genannt, man möchte das gesamte Weltwissen in der Ontologie erfassen. Der Inhalt umfasst deshalb auch große Teile von Alltagswissen und so gut wie jedes nur erdenkliche Spezialgebiet. Das komplette Paket kann dadurch mit einigen beachtlichen Zahlen aufwarten: Die Ontologie umfasst mehr als 100000 Begriffe, die mit mehr als einer Million Eigenschaften miteinander verbunden sind. Desweiteren wurden über 600 Mannjahre an Programmierung und Design in dieses Projekt gesteckt. Die frei verfügbare Variante OpenCyc wird in einer kaum abgespeckten Version seit 2002 angeboten und ständig aktualisiert.

Bemerkenswert ist, dass zur Realisierung der Ontologie eine eigene Codierungssprache, die sogenannte CycL, erfunden wurde, und das Wissen somit nicht in den üblichen Formaten vorliegt. Diese Inkompatibilität ist wohl mit ein Grund, dass das eigene Format als Standard forciert wird.

Die neueste Version der Ontologie ist jederzeit auf der offiziellen Homepage frei verfügbar ¹⁰.

GUMO

Abschließend soll hier noch kurz das Projekt GUMO von Dr. Dominik Heckmann angesprochen werden. Diese Wissensbasis befindet sich zurzeit am DFKI (Deutsches Forschungszentrum für künstliche Intelligenz) an der Universität des Saarlandes in Entwicklung.

Ziel dieser Wissensbasis ist es, eine Benutzermodellierung zu ermöglichen. Das bedeutet man möchte beispielsweise den Nutzer einer Internetseite als Instanz der Ontologie darstellen. Um dies zu verwirklichen umfasst die Ontologie Wissensgebiete wie Gegenstände, physikalische Konzepte und Klassen zur Modellierung von Zeit und Raum. Es wurden aber auch Konzepte für Fähigkeiten, Aktivitäten und Ereignisse eingebunden.

Die jeweils aktuellste Version von GUMO kann frei verfügbar von der Homepage heruntergeladen werden, oder bei Bedarf auch direkt online angesehen werden ¹¹.

4. Zusammenfassung

Ob man nun mit Ontologien dem anvisierten Ziel, ein semantisches Internet zu entwickeln, in dem Agenten logisch und menschenähnlich navigieren können, einen Schritt näher gekommen ist, wird sich mit der weiteren Entwicklung des Semantic Web zeigen. Es steht aber außer Frage, dass Ontologien für sich genommen bereits eine Vielzahl an Möglichkeiten und Vorteilen bieten. Im ersten Teil der vorliegenden Ausarbeitung werden aus diesem Grund der Aufbau einer Ontologie aus Klassen, Eigenschaften, Einschränkungen und Instanzen, sowie der Nutzen einer Ontologie näher beleuchtet. Diese Anwendungsgebiete reichen von Teilen von Wissen zwischen Agenten und/oder Menschen, über die Wiederverwendung von Informationen, bis hin zu erweiterten Analysen von Wissen.

Dieser Text verfolgt in einem weiteren großen Block das Ziel, dem Leser den Entstehungsprozess einer Ontologie näher zu bringen. Zu diesem Zweck wird ein möglicher Design-Prozess in mehreren Schritten vorgestellt. Diese Entwicklung beschäftigt sich zu Beginn mit einer Umfangsbestimmung und versucht dann, bereits fertige Ontologien wieder zu verwenden. Das eigentliche Erstellen einer Ontologie beginnt in einem dritten Schritt mit dem Sammeln von Begriffen zu dem gewählten Themengebiet. Diese Begriffe werden anschließend in einer Klassenhierarchie angeordnet und untereinander mit Eigenschaften verbunden. Ein wichtiger Schritt, um gezielt Informationen aus der Ontologie ableiten zu können, sind Einschränkungen, die auf die Eigenschaften angewandt werden. Im letzten Schritt des Prozesses werden noch konkrete Objekte, die sogenannten Instanzen, erstellt und die Ontologie somit zu einer Wissensbasis erweitert.

Der letzte Teil dieser Ausarbeitung stellt dem Leser einige Ontologien vor, die frei im Internet verfügbar sind. Die SUMO Ontologie ist eine große Wissensbasis, die sehr variabel ist, da sie in sehr viele Themengebiete gezielt erweitert und spezialisiert werden kann. Einen anderen Ansatz verfolgt die WordNet Ontologie, bei der die Eigenschaften zwischen den Klassen auf lexikalischen und semantischen Eigenschaften aufbauen. Durch den breiten Umfang und die vielen mitgelieferten Bedeutungen lässt sich diese Ontologie auch gut als Lexikon benutzen. Eine weitere vorgestellte Ontologie ist OpenCyc, die vor allem durch den riesigen Umfang und eine eigene Codierungssprache hervorsteicht. Abschließend beleuchtet der Text noch eine Ontologie, die zurzeit an der Universität des Saarlandes in Entwicklung ist, GUMO.

Als persönliche Bemerkung möchte ich noch kurz meine eigene Erfahrung mit dem vorgestellten Design-Prozess anmerken. Ziel einer Gruppenarbeit war das Erstellen einer Ontologie zu einem bestimmten Thema. Dabei haben wir die Ontologie grundlegend gemäß den hier skizzierten Schritten aufgebaut. Auch gerade wenn man zum ersten Mal eine Wissensbasis erstellt, geht die Konstruktion problemlos und überraschend schnell von der Hand. Man erhält relativ schnell überzeugende Ergebnisse, die wohl eher durch Erfahrung als durch andere Design-Prozesse verbessert werden. Insgesamt lässt sich sagen, dass diese Anleitung (besonders für Anfänger) sehr gelungen und zu empfehlen ist.

Literaturverzeichnis

Quellen:

Noy, Natalya F. / McGuinness, Deborah L.

Ontology Development 101: A Guide to Creating Your First Ontology

Stanford University, Stanford, CA, 94305

Gruber, T. R.

A translation approach to portable ontologies

Stanford University, Stanford, CA, 94305

Internet-Verweise:

DAML Ontology Library

<http://www.daml.org/ontologies/>

GUMO Homepage

<http://www.ubisworld.org/>

Ontolingua Ontology Library

<http://www.ksl.stanford.edu/software/ontolingua/>

OpenCyc Homepage

<http://www.opencyc.org/>

SUMO Homepage

<http://ontology.teknowledge.com/>

SUMO Webbrowser

<http://virtual.cvut.cz/ksmsaWeb/browser/title>

Wikipedia

<http://www.Wikipedia.de>

WordNet Homepage

<http://wordnet.princeton.edu/>

WordNet Visualisierung

<http://staff.science.uva.nl/~kamps/wordnet/>