

Ontologiemodularisierung im Unternehmenskontext

Gökhan Coşkun

Netzbasierte Informationssysteme
Institut für Informatik
Freie Universität Berlin
Königin-Luise-Str 24-26
14195 Berlin
coskun@inf.fu-berlin.de

Abstract: Ontologie als grundlegende Technologie des Semantic Webs gewinnen zunehmend an Bedeutung. Im Unternehmenskontext ist eine wirtschaftliche Ontologieentwicklung und -wartung essentiell. Dabei spielt die Modularisierung von Ontologien sowohl für die Wiederverwendbarkeit als auch für die einfache Wartbarkeit und Erweiterbarkeit eine große Rolle. Die Erforschung der Ontologiemodularisierung im Unternehmenskontext ist sowohl für die Entwicklung der Semantic Web Technologien als auch für das Wissensmanagement im Unternehmen vielversprechend.

Einleitung

Die Bedeutungen und Inhalte der Webseiten für Maschinen zugänglich und nutzbar zu strukturieren war der grundlegende Gedanke des 2001 von Tim Berners-Lee eingeführten Semantic Web [TBL01]. Die Technologien des Semantic Web, konzipiert für das globale Internet, bringen durch ihre praktischen Ansätze mehr Transparenz und Flexibilität in die komplexen, auf heterogenen Daten basierten Informationsstrukturen des World Wide Web. Für die explizite und maschinenlesbare Darstellung von Wissen kommen im Semantic Web sogenannte Ontologien zum Einsatz. Diese sind nach T.R. Gruber eine explizite Spezifikation einer Konzeptualisierung (eines Begriffsverständnisses). Während die Semantic Web Initiative wohldefinierte Sprachen wie Web Ontology Language (OWL) für die Repräsentation von Ontologien entwickelt hat, ist die Methodologie zur Ontologieentwicklung an sich noch nicht abschließend geklärt. Es gibt mehrere Ansätze der kollaborativen Entwicklung von Ontologien, die das Internet als Plattform betrachten. Somit beruhen sie alle auf einem offenen, heterogenen, stark verteilten und dezentralen System, worin die Hauptaufgabe der Ontologien darin besteht, gemeinschaftlich einen von allen Teilnehmern anerkannten Konsens zu finden.

Die immense Größe des Internet und die unüberschaubare Anzahl von Teilnehmern erschweren die kollaborative Ontologieentwicklung und machen die Bildung eines Konsens nahezu unmöglich.

Mit der Forderung nach einer Einschränkung der Einsatzumgebung von Semantic Web Technologien zur Erforschung ihres Potentials in einem einfacher zugänglichen Umfeld entstand der Begriff Corporate Semantic Web. Dabei werden die Technologien des Semantic Web im Unternehmenskontext betrachtet. Auch aus der Sicht der Unternehmen ist dies vielversprechend, denn die Erfassung und Bereitstellung von Wissen für die bessere Zusammenarbeit und eine stärkere Wissensübertragung in unterschiedlichen Unternehmensbereichen stellen einen der wichtigsten Aspekte in der modernen, elektronischen Geschäftsabwicklung. Der bei der Erarbeitung von Ontologieentwicklungsmethodologien bislang nicht betrachtete Unternehmenskontext stellt besondere Anforderungen an die Entwicklung von Ontologien, die genauer betrachtet werden müssen.

Anforderungen an Ontologieentwicklung im Unternehmenskontext

Der Einsatz von Ontologien in einem Unternehmen muss gewissen Anforderungen genügen, damit die Integration dieser Technologie auch aus wirtschaftlicher Sicht gerechtfertigt ist. Allein der potentielle technische Mehrwert ist kein hinreichender Grund für die Einführung einer neuen Technologie. Die Investitionskosten am Anfang, aber auch die Wartungskosten während des Einsatzes, sollten geringer als der durch die neue Technologie erwirtschaftete Gewinn sein. Dies ist eine große Hürde die bewältigt werden muss. Sie wird erschwert durch die Tatsache, dass einerseits die vollständige Neuentwicklung von Ontologien sehr zeit- und kostenintensiv ist und andererseits die Wartung von großen zusammenhängenden Ontologien sehr schwer realisierbar ist. Daraus ergeben sich die folgenden Anforderungen an die Entwicklung von Ontologien im Unternehmenskontext: Zu Beginn, muss die Möglichkeit bestehen, bereits existierende Ontologien, z.B. aus dem Internet, bei der Erstellung einer ersten, einsetzbaren Version wiederzuverwenden, damit die Einführungskosten reduziert werden können. Im weiteren Verlauf müssen die Wartungskosten durch einfache Wartbarkeit und Erweiterbarkeit so gering wie möglich gehalten werden. In beiden Fällen ist die Modularisierung von Ontologien von hoher Bedeutung. Denn durch die Extraktion von relevanten Teilen aus bereits bestehenden Ontologien und die Integration dieser kann einfacher eine Startontologie erstellt werden. Außerdem können in geeigneter Art und Weise auf kleinere Module aufgeteilte Ontologien einfacher gewartet und erweitert werden.

Ontologiemodularisierung im Unternehmenskontext

Das Verständnis von dem Begriff Modularisierung ist im Zusammenhang mit Ontologien noch nicht eindeutig geklärt [ASSS07]. Existierende Techniken basieren auf unterschiedlichen, impliziten Voraussetzungen, welche explizit beschrieben werden sollten.

Mark Hall beschreibt in [MH04] Modularisierung als die Zerteilung einer Ontologie in mehrere unabhängige disjunkte Bäume, also die Partitionierung der gesamten Ontologie. Im Gegensatz dazu definiert Paul Doran in [PD06] Modularisierung als ein Mechanismus, welcher dem Ontologieentwickler die Möglichkeit bietet, aus der ursprünglichen Ontologie eine Untermenge oder ein so genanntes Ontologiemodul zu extrahieren. In [ASSS07] wird Modularisierung nicht wie in den ersten beiden Fällen als ein Prozess, bei dem die Ontologie bearbeitet wird, verstanden, sondern als ein Weg zur Strukturierung von Ontologien betrachtet. Die Konstruktion von großen Ontologien soll demnach auf der Kombination von vollständigen und in sich abgeschlossen Ontologien bestehen. Die Abbildung illustriert eine kleine Klassifikation der unterschiedlichen Modularisierungsverständnisse.

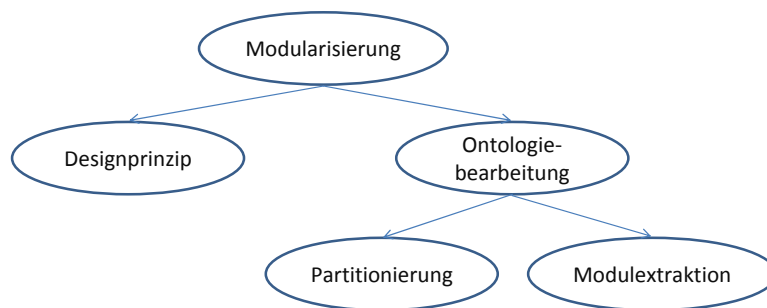


Abbildung 1 - Verständnis des Begriffs Modularisierung

In Bezug zur Ontologieentwicklung im Unternehmenskontext sind alle drei Arten der Modularisierung von Bedeutung. Modulextraktion ist vor allem bei der Wiederverwendung von bereits existierenden Ontologien wichtig, damit relevante Teile extrahiert und übernommen werden können. Bei der Integration dieser Module zu einer nutzbaren Gesamtontologie ist der modulare Ansatz als Designprinzip wichtig, vor allem für die einfachere Wartbarkeit. Abschließend ist die Partitionierung der Gesamtontologie für die Bildung von spezifischen Subontologien für die unterschiedlichen Unternehmensbereiche für effiziente Nutzbarkeit relevant. Letzteres ist vor allem möglich, da die geschlossene Umgebung des Unternehmens im Gegensatz zum World Wide Web neue Möglichkeiten bietet, wie die Überwachung der Nutzung.

Die geschlossene und kontrollierte Umgebung für den Ontologieeinsatz, wie der Unternehmenskontext, erlaubt neue Ansätze für die Modularisierung von Ontologien. Die Möglichkeit die Nutzung komplett zu überwachen in Kombination mit dem Wissen über die Umgebung an sich erlaubt z.B. die Zuordnung von Ontologieteilen an die Unternehmensbereiche wie Abteilungen und Projekte. Dabei kann durch statistische Mittel die Relevanz von Konzepten in Bezug zu bestimmten Abteilung als Maß ermittelt werden. Abgesehen von der Relevanz kann durch Überwachung wie sich die Ontologie während ihrer Nutzung entwickelt, z.B. durch Aktualisierungen, auch ermittelt werden, welche Bereiche starken Veränderungen unterworfen sind und somit möglicherweise noch keine zuverlässigen Quellen darstellen. Teile der Ontologie, die längere Zeit unverändert bleiben, könnten somit als zuverlässig markiert werden.

Relevanz und Zuverlässigkeit sind also zwei Beispiele, welche in einem Unternehmenskontext ermittelbare und messbare Größen wären, die für die Modularisierung von Ontologien genutzt werden könnten. Die Vertrauenswürdigkeit des Wissens, welches durch die Ontologie repräsentiert wird, und auch die Effizienz der Nutzung würden somit verbessert werden.

Für Entwicklung von Ontologien für den Unternehmenskontext ist also eine Methodologie notwendig, welche die erwähnten Aspekte der Modularisierung berücksichtigt und realisiert und somit auch die beschriebenen Anforderung des Unternehmens bezüglich der Einführungs- und Erhaltungskosten erfüllt. Dabei sollten die Vorteile, die der Unternehmenskontext im Gegensatz zum World Wide Web bietet, in optimaler Weise ausgeschöpft werden.

Stand der Forschung

Es gibt bislang noch keine Methodologie für die Ontologieentwicklung mit Schwerpunkt auf Modularisierung und dem Unternehmenskontext als Einsatzumgebung. Was die Modularisierung im Sinne von Zerteilen einer existierenden Ontologie angeht, wird in [SR06] eine Klassifikation eingeführt. Demnach kann die Modularisierung in drei Kategorien unterteilt werden.

Die anfragebasierten Methoden ist die erste Kategorie, welches von der Datenbankwelt inspiriert ist. Ähnlich zu einer Datenbank wird eine Anfrage an die Ontologie gestellt, woraufhin eine Antwort in Form eines Ausschnittes aus der Ontologie folgt. Einige Beispiele für diese Kategorie sind zu finden in [PS01], [VOS03]

Die zweite Kategorie heißt Netzwerkpartitionierung und folgt dem Gedanken, dass jedes System in einem gewissen Maße zerteilbar ist. Auf Ontologien angewendet, bedeutet dies, dass in einer Ontologie Konzepte existieren müssen, die eine engere Bindung zu einander haben als zu anderen. Betrachtet man die Konzepte als Netzwerkknoten und die Relationen zwischen den Konzepten als Netzwerkverbindungen, kann man möglichst einfach Algorithmen zur Segmentierung aus dem Gebiet der Netzwerke übernehmen. Im Gegensatz zu den anderen Kategorien wird bei dieser nur die Ontologie an sich und ihre Struktur als Grundlage für die Partitionierung genommen. Beispiele hierfür sind [SK04], [GPS05] und [MM03].

Extraktion durch Erweiterung ist die dritte Kategorie, welche der Netzwerkpartitionierung in sofern ähnelt, als dass sie die Ontologie auch als ein Netzwerk interpretiert. Allerdings wird hierbei der Anfang bei einem bestimmten Konzept gemacht und anschließend wird die Sicht durch alle mit dem Anfangskonzept in Beziehung stehenden Konzepte erweitert. Die ursprüngliche Ontologie bleibt erhalten und es entsteht eine neue Sicht mit Fokus auf dem Anfangskonzept. Beispiele für diese Kategorie sind [NM04] und [BW04].

Zusammenfassung und Ausblick

Die steigende wirtschaftliche Bedeutung von Erfassung und Bereitstellung des Wissens innerhalb eines Unternehmens macht die Erforschung von Ontologieeinsatz im Unternehmenskontext notwendig. Diese im Anfangsstadium befindliche Arbeit, hat das Ziel zu erforschen wie durch Modularisierungstechniken die Einführungs- und Wartungskosten von Ontologien reduziert werden können. Es soll eine Methodologie entwickelt werden, die die Modularisierung und Integration von Ontologien im Unternehmenskontext ermöglicht.

Die nächsten Schritte umfassen die Klärung eines Vorgehensmodells für die Dissertation und die Identifikation von Kriterien für die erarbeitete Lösung, so dass diese in geeigneter Art und Weise unter anderem mit Hilfe von Fallstudien evaluiert werden kann.

Danksagung: Dieser Artikel entstand als Ergebnis der Forschungsarbeiten im Vorhaben „InnoProfile: Wirtschaftsorientierte Nachwuchsforschergruppe Corporate Semantic Web“, das vom Bundesministerium für Bildung und Forschung (BMBF) und der BMBF-Innovationsinitiative für die Neuen Länder – Unternehmen Region – gefördert wird.

Literaturverzeichnis

- [ASSS07] M. d'Aquin, A. Schlicht, H. Stuckenschmidt, M. Sabou, Ontology Modularization for Knowledge Selection: Experiments and Evaluations, 18th International Conference on Database and Expert Systems Applications, 2007. DEXA '07.
- [BW04] M. Bhatt, C. Wouters, A. Flahive, J. W. Rahayu, D. Taniar, Semantic Completeness in Sub-ontology Extraction Using Distributed Methods, ICCSA 2004
- [GPS05] B. C. Grau, B. Parsia, E. Sirin, A. Kalyanpur, Automatic Partitioning of OWL Ontologies Using E-Connections, Description Logics 2005
- [MH04] M. Hall, Ontology integration and evolution, SE Data & Knowledge Engineering May 10, 2004
- [MM03] B. MacCartney, S. A. McIlraith, El Amir, T. E. Uribe, Practical Partition-Based Theorem Proving for Large Knowledge Bases, IJCAI 2003
- [NM04] N. F. Noy and M. A. Musen, Specifying Ontology Views by Traversal, International Semantic Web Conference 2004
- [PD06] P. Doran, Ontology reuse via ontology modularization, KnowledgeWeb PhD Symposium 2006
- [PS01] E. Prud'hommeaux, A. Seaborne, SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>
- [SK04] H. Stuckenschmidt, M. Klein, Structure-Based Partitioning of Large Class Hierarchies, In Proceedings of the 3rd International Semantic Web Conference 2004
- [SR06] J. Seidenberg, A. Rector, Web ontology segmentation: analysis, classification and use, Proceedings of the 15th international conference on World Wide Web, 2006
- [TBL01] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, Scientific American, 284(5): 34-43, May 2001
- [VOS03] R. Volz, D. Oberle, R. Studer, Views for light-weight Web ontologies, Proceedings of the 2003 ACM symposium on Applied computing 2003, SAC '03