

# Graphische Entwicklung und Nutzung von Ontologien mit SemTalk in MS Office

Christian Fillies

Semtation GmbH  
Bredower Str. 145  
D-14612 Falkensee  
cf@semtation.com

Dr. Frauke Weichhardt

Beratung im Netz  
Merkurstraße 17  
D-14482 Potsdam  
fw@semtation.com

**Abstract:** Um das semantische Web aus dem Status eines Forschungsprogramms in „Real World“-Anwendungen zu überführen ist es notwendig, Szenarien zu schaffen, in denen der Nutzen dieser Technologie für den Endanwender deutlich wird. Mit SemTalk wird eine Infrastruktur für „Semantic Web enabled“ Tools innerhalb von MS-Office vorgestellt. Ein breiter Kreis von Anwendungen von der klassischen Wissensrepräsentation mit semantischen Netzen über Prozessmodellierung bis hin zur SAP IPC Produktmodellierung wird mit Technologien des Semantic Web auf eine neue Basis gestellt, indem sie auf verteilte Ontologien im semantischen Web zugreifen.

## 1 Einleitung

*“Respect for the word - to employ it with scrupulous care and an incorruptible heartfelt love of truth - is essential if there is to be any growth in a society or in the human race. To misuse the word is to show contempt for man. It undermines the bridges and poisons the wells. It causes man to regress down the long path of his evolution.”*

Joshua Allen [joshuaa@microsoft.com] 04/29/02 zitiert Dag Hammarskjöld  
auf [www-rdf-interest-request@w3.org](http://www-rdf-interest-request@w3.org)

Das Semantic Web [BHL01] befindet sich derzeit noch im Anfangsstadium. Von vielen Seiten werden große Möglichkeiten in seinem weiteren Aufbau gesehen, konkrete Anwendungen sind jedoch derzeit noch sehr selten. Die Definition von XML-Protokollen wie RDF, RDFS und DAML+OIL durch das W3C lassen eine größere Verbreitung erwarten, so dass in Zukunft mit einer zunehmenden Anzahl von Anwendungen im Semantic Web zu rechnen ist.

Diese neue Technologie wird sich zunächst innerhalb der Intranets größerer, verteilter Unternehmen ausbreiten, da hier ein großer Bedarf zur Abstimmung der Strukturen von Knowledge-Management-Systemen zwischen den verschiedenen Unternehmensteilen besteht, welcher durch die Technologien des Semantic Web befriedigt werden kann. Die Schaffung zentraler Begriffskontexte bzw. Ontologien und Prozessmodelle ist dazu eine wesentliche Voraussetzung.

Die Idee des semantischen Webs ist die gemeinsame Nutzung mehr oder weniger formalisierten Wissens via Internet. Ursprünglich stand der Gedanke, eine gemeinsame Wissensbasis für Computer zu schaffen, im Vordergrund. Die hier beschriebene Arbeit behandelt jedoch zunächst den Wissensaustausch zwischen Menschen innerhalb eines (graphischen) Wissensnetzwerks.

Schlüsselfaktor für die verteilte Wissensnutzung ist die Verwendung einer gemeinsamen Terminologie. Dieser Artikel zeigt, wie bereits jetzt existierende Semantic Web Knowledge Services, die auf RDFS basieren, in den Erzeugungsprozess semantischer Netze und Geschäftsprozesse integriert werden können. Die meisten der bereits vorhandenen online- und offline-Glossare und Editoren sind Browser, die jeweils genau einen Knoten eines Netzwerks anzeigen. Einige Werkzeuge generieren bei Bedarf hyperbolic trees oder ähnliche Darstellungen. Unsere Erfahrung mit der Kommunikation von Wissen besagt jedoch, dass komplexe Probleme besser mit Hilfe von manuell erzeugten Diagrammen beschrieben werden können, die jeweils ein spezifisches Szenario darstellen. Dazu nutzen Menschen heutzutage am liebsten entweder die flexiblen Malprogramme oder andererseits Modellierungswerkzeuge, die die Verwaltung der dargestellten Objekte auf mehr oder weniger bequeme Art erlauben. Der hier gezeigte Ansatz ermöglicht die Kombination dieser beiden Wege.

Mit SemTalk [FW01], [FW02] steht ein einfach zu bedienender Editor für Semantic-Web-Ontologien und Prozesse auf der Basis von MS Visio zur Verfügung.

SemTalk basiert auf einem offenen, graphisch konfigurierbaren Meta-Modell, so dass es leicht an verschiedene, insbesondere auch prozessorientierte, Modellwelten wie EPKs oder KSA-Modelle anzupassen ist. Die Modelle stehen den räumlich und zeitlich verteilten Benutzern mit Hilfe von MS Office XP SmartTags bei der täglichen Arbeit mit Word, Excel oder Outlook im direkten Zugriff zur Verfügung. Gleichzeitig ermöglicht SemTalk die verteilte Erstellung und Wartung solcher Modelle, indem die Koordination verteilt erstellter Modelle mit Hilfe von Namensraum-Konzepten und index-basierten Wizards ohne die sonst im Content Management üblichen zentralen Mechanismen einer auf eine Datei oder eine Datenbank zentrierten Modellierung („zentrales Dogma“) durchgeführt werden kann. Es würde wahrscheinlich kein Internet in seiner jetzigen Form geben, wenn jemals eine zentrale Instanz zur Sicherung der Konsistenz für HTML existiert hätte.

## 2 Architektur

MS Visio wird aus zwei Gründen für die graphische Präsentation verwendet: Die Anwender kennen es, und es ist vollständig programmierbar.

SemTalk benutzt - neben Visio-Dateien für die Graphik - eine XML-Datenstruktur zur Repräsentation der Modelle. Darin werden Layout-Informationen sowie objekt-orientierte Funktionalität wie Methoden und Zustände zu den Sprachelementen von RDFS hinzugefügt. RDFS selbst umfasst im wesentlichen eine XML-Notation für

Klassen und ihre Properties in der Tradition der Frame-Sprachen der AI. SemTalk nutzt im Gegensatz zu RDFS optimierte Strukturen für grundlegende Inferenz- und Vererbungsmechanismen und auch zur Graphenerzeugung. Die Property-Repräsentation ist ähnlich wie bei DAML in die Klassendefinition integriert.

Weiterhin gibt es eine Objekt-Maschine, die ein COM API bereitstellt, um das Programm auch innerhalb von MS Office-Produkten nutzen zu können. Über diese Objekt-Maschine steht ein Export / Import von RDFS und in Kürze auch DAML zur Verfügung.

Die SemTalk-Engine wird für die Definition der Semantik von bestehenden Visio-Shapes verwendet, d. h. für die Definition eines Meta-Modells für Visio Shapes. Ein Visio Shape entspricht dabei in etwa eine Klasse. Graphisch lässt sich damit festlegen, welche Shapes miteinander verbunden werden dürfen und auf welchen Typen von Visio-Seiten sie stehen dürfen. Es können Oberklassen von Shapes definiert werden. Die Beziehungen zwischen diesen Klassen werden wiederum in Visio visualisiert. Die Klassendefinitionen können dabei nach RDFS exportiert bzw. importiert werden. Einen ähnlichen Ansatz im Umfeld PowerPoint findet man bei [TB02].

SemTalk stellt die Infrastruktur zur Definition vollständiger Modellierungsmethoden innerhalb von Visio zur Verfügung. Dabei kann es sich beispielsweise um die Definition von Methoden für DAML, Produktmodellierung innerhalb von ERP-Systemen oder für BPM-Methoden wie ARIS handeln. SemTalk hat Schnittstellen zu CASE Tools wie Rational Rose und zu Prozeßmodellierungswerkzeugen wie ARIS und Bonapart. Außerdem existiert ein einfacher Report-Generator für die Erzeugung von HTML-Tabellen und XSL für die Formatierung.

Zur intelligenten Analyse der Daten wurde eine Anbindung an den Reasoning Web Service von OntoBroker der Fa. Ontoprise GmbH [St01] über das Internet entwickelt. Mit SemTalk erstellte Modelle werden dazu in F-Logic abgelegt. Ontobroker erhält dann eine Liste von Hyperlinks auf die relevanten Modelle und eine F-Logic Query. Die Lösungen der Anfrage können in die SemTalk Modelle eingefügt bzw. darin angezeigt werden. Eine ähnliche Schnittstelle gibt es für statistische Quelltextaufbereitung zum Concept Composer der Fa. TextTech GmbH [He01].

### **3 Notationen für Semantic Webs**

Visualisierungen für das semantische Web nutzen ähnliche Notationen wie sie seit den 70er Jahren für semantische Netze gebräuchlich sind. Der Unterschied liegt nur darin, dass die Modelle jetzt auf dem Internet verteilt sind und von großen Nutzerkreisen gesehen werden, während sie früher in einer geschlossenen Welt einer einzigen Lisp-Machine lagen. Es kommen dabei natürlich sofort die alten Fragestellungen wieder auf wie z.B. „What's in a link“ in der berühmten Arbeit von Woods [Wo75]. Die Frage ist dabei, wie stark formalisiert die Sprache in den Modellen sein muss, um von Inferenzmaschinen interpretiert werden zu können. Im Kontext des semantischen Web ist das äquivalent zu der in der Einleitung gestellten Frage, ob ein Web für Maschinen oder ein Web für das Knowledge Management zwischen Menschen gebaut werden soll. SemTalk bietet einerseits durch die Verwendung von Relationsklassen, die vorgegeben werden, eine maschinen-orientierte Variante, lässt aber für bestimmte Anwendungsszenarien auch eine freie ad-hoc-Definition der Relationen zu.



verwendet werden. Beispiele, die mit SemTalk realisiert wurden, sind ein graphisches Customizing des SAP Produktkonfigurators und im Agenten Umfeld die DAML Notation.

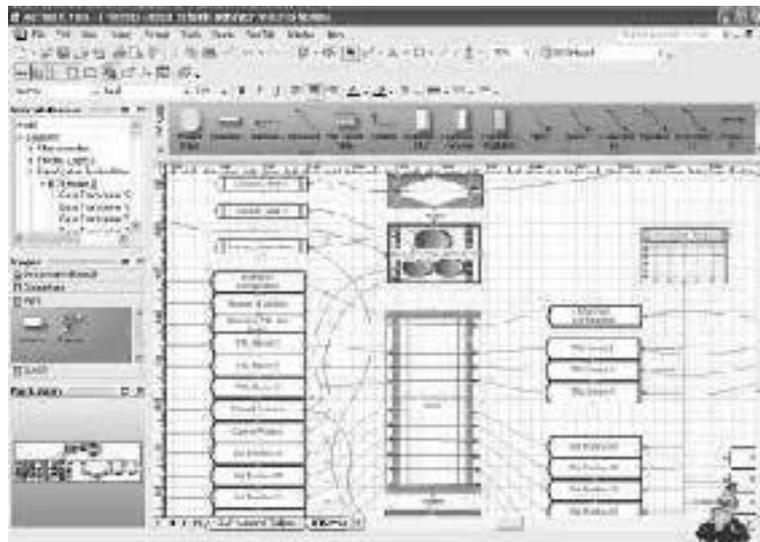


Abb. 2: Visual Product Modeller for SAP/R3 IPC

Beim Visual Product Modeller haben wir eine umgekehrte Situation wie bei den semantischen Netzen: Wohldefinierte Relationen ermöglichen es, die für den SAP Upload benötigte XML-Zwischennotation zu generieren und auf dem Web publizierte Produktmodelle intelligent zu durchsuchen.

#### 4 Tool Support für verteilte Modellierung

SemTalk enthält einen Modellierungswizard, der den Benutzer bei der Erstellung von konsistenten Modellen unterstützt. Neben der Überprüfung der Schreibweise mit Hinweisen auf Synonyme gibt es konkrete Hilfe für die objekt-orientierte Erstellung von Modellen. Da der überwiegende Anteil der Anwender nur wenig oder keine objekt-orientierten Kenntnisse besitzt, erscheint dieses Vorgehen als hilfreich.

Die primäre Aufgabe des Wizards ist es aber zu kontrollieren, ob Worte verwendet werden, die schon in anderen Modellen definiert worden sind.

Die wichtigste Philosophie des Internets und damit auch des Semantic Web ist es, Informationen nicht zu kopieren sondern zu referenzieren. Durch den Link auf eine andere Webseite verändert man deren Inhalt nicht. Das flexible Informationssystem, das auf diese Weise entsteht, hat sicherlich nicht die Konsistenz einer Datenbank, hat dafür aber die Chance dynamisch zu wachsen. SemTalk verwendet dieselbe Philosophie im Hinblick auf Modelle. Mit SemTalk wird also nicht ein einzelnes Modell sondern ein Netz von miteinander verlinkten Modellen erstellt. Verlinkt werden die Modelle sowohl untereinander als auch mit externen Modellen wie zum Beispiel branchenspezifischen

Standards. In einem solchen Fall wird eine Kopie des externen Begriffes mit einer URN des referenzierten Begriffes erzeugt. Dadurch wird es möglich, einerseits offline gemeinsame Begriffssysteme zu nutzen und andererseits die Modelle wie eine Notes-Datenbank zu replizieren.

Praktische Anwendungen dieser Technologie sind oft Referenzmodelle: Eine Versicherung erstellt ein konzernweites CRM-Modell und passt dann die länderspezifischen Vorgehensweisen daran an. Im E-Government-Bereich werden z.Z. Referenzmodelle für Verwaltungsprozesse entwickelt, die dann in den einzelnen Behörden offline ausgerollt werden können. Wichtig ist in beiden Fällen die Referenzierung einer gemeinsamen Terminologie.

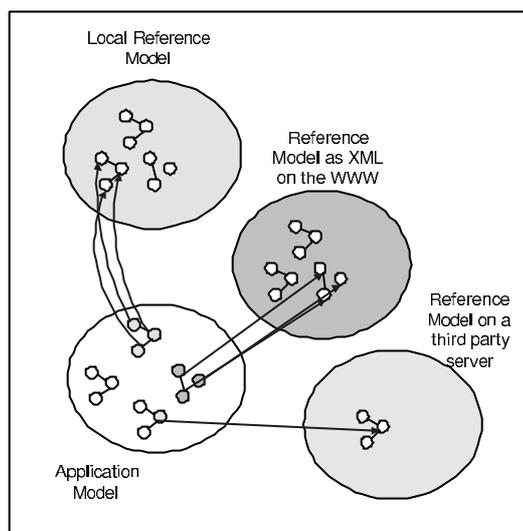


Abb 3: Hyperlinking SemTalk Models

Der Wizard nutzt zur Erkennung von Worten Index-Dateien, die von Crawlern für Modelle, Verzeichnisse oder das Semantic Web auf dem Internet erstellt wurden. Eine Index-Datei ist wiederum eine XML-Datei, die eine invertierte Liste der modellierten Objekte enthält. D. h. jedem Wort ist die Liste der Modelle zugeordnet, in denen es modelliert ist. Indexdateien werden über die SemTalk-Oberfläche nach den speziellen Anforderungen des Benutzers erzeugt. Sie stellen somit den für den Benutzer relevanten Teil des verfügbaren Semantic Webs dar. Komplexere Reasoning-Funktionalität steht durch die Anbindung von Ontobroker zur Verfügung. Mit Ontobroker können über die Sprache F-Logic komplexere Suchmuster auch über mehrere Modelle eingesetzt werden. F-Logic wird dabei zur Formulierung der Anfragen benutzt.

## 5 Praktische Nutzung des Semantic Web aus MS Office XP

Liest man diesen Text mit einem Internet Explorer 6 oder MS Word XP bemerkt man ein SmartTag auf dem Wort „SmartTag“ (violette Punkte)



Abb. 5: SmartTag action menu

SmartTags werden automatisch erkannt, während man den Text schreibt oder liest. Ein SmartTag ist ein als XML eingebetteter Marker im Text, der im Kontext des jeweiligen Lesers interpretiert werden kann und entsprechende Aktionen auslöst. SemTalk Class Recognizer SmartTags überprüfen Text, Emails, Project Pläne oder Excel sheets auf bereits modellierte Begriffe und bieten verschiedene Aktionen an. Mit der Option „HTML“ kann z.B. direkt in eine graphische Erklärung des Begriffes gesprungen werden.

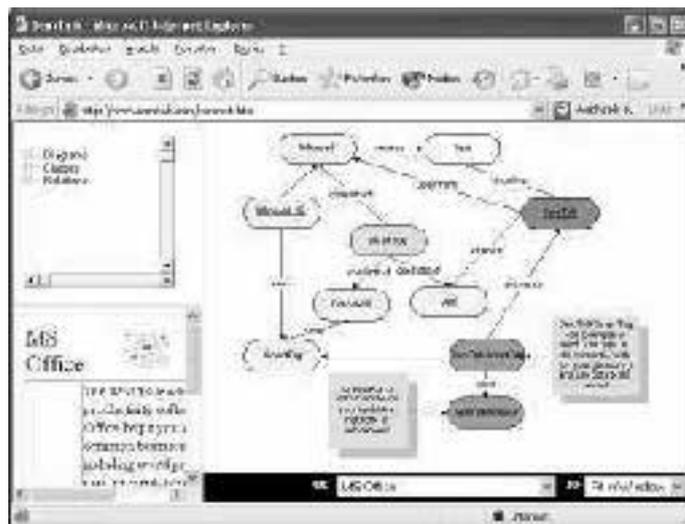


Abb. 6: HTML Visualisierung von Office Begriffen aus <http://www.semtalk.com>

Mit der Option „Make Hyperlink“ wird das SmartTag in einen statischen Hyperlink umgewandelt für diejenigen Leser, die keinen Internet Explorer, bzw. kein SemTalk Smart Tag haben. Statische Hyperlinks entsprechen aber einem „Early Binding“ also der Zuordnung des Hyperlink im Kontext des Autors. Dynamische Hyperlinks, wie z.B. SmartTags erlauben ein “late Binding” im Kontext des Lesers. Late Binding ist wesentlich spannender, da es die Transformation der Worte oder ihrer Synonyme in das

aktuelle Wissensmodell des Lesers unterstützen kann. Ein umfangreiches graphisches Glossar wurde im Laufe des Jahres 2001 bei der IT-Abteilung der Credit Suisse entwickelt. Dieses Projekt ist recht ausführlich in [FWW02] beschrieben. Falls die Anwender SemTalk installiert haben, können sie auch direkt in das entsprechende Wissensmodell in Visio navigieren und ggf. das Modell entsprechend ihren Anforderungen erweitern. SmartTag Technologie wird von uns als ein wichtiger Erfolgsfaktor für Wissensmanagement gesehen, da es proaktiv ist: Wurden bisher Dokumentationen, Glossare und Prozessbeschreibungen auf Papier bzw. statischem HTML abgelegt und vergessen, so werden sie jetzt in die täglichen Arbeitsaufläufe integriert und weisen den Nutzer mehr oder weniger von selbst auf vorhandenes Wissen hin.

## 6 Zusammenfassung

Mit SemTalk steht ein Werkzeug zur Erzeugung von Semantic Webs in XML zur Verfügung. Dabei kann auch die Entstehung großer Netze unterstützt werden, indem eine dezentrale Modellierungsweise angewendet wird. SemTalk ist derzeit bei verschiedenen Firmen im Rahmen von Wissensmanagement-Aktivitäten und zur Prozessmodellierung sowie Produktmodellierung im Einsatz. Durch die Benutzung von Office einschließlich Visio als Front-End wird so Semantic Web Technologie für einen sehr großen Anwenderkreis erschlossen.

## 7 Literatur

- [BHL01] Tim Berners-Lee, Jim Hendler, and Ora Lassila published an article about the Semantic Web in Scientific American.  
<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>
- [FW01] Fillies, C.; Weichhardt, F.; SemTalk: A RDFS Editor for Visio 2000  
Position Paper, ICCS 2001 9th International Conference on Conceptual Structures / Semantic Web Working Symposium (SWWS)
- [FWW02] Fillies, C., Wood-Albrecht, G., Weichhardt, F., A Pragmatic Application of the Semantic Web Using SemTalk. WWW2002, May 7-11, 2002, Honolulu, Hawaii, USA ACM 1-5811-449-5/02/0005
- [He01] Heyer, G.; Läuter, M.; Quasthoff, U.; Wittig, Th.; Wolff, Chr.: Learning Relations using Collocations. In: A. Maedche, S. Staab, C. Nedellec and E. Hovy, (eds.), Proc. IJCAI Workshop on Ontology Learning, Seattle/WA, 19. – 24. August 2001
- [St01] Steffen Staab and Rudi Studer, Institute AIFB—University of Karlsruhe and Ontoprise GmbH, Hans-Peter Schnurr, Ontoprise GmbH, York Sure, Institute AIFB—University of Karlsruhe: Knowledge Processes and Ontologies. In: IEEE INTELLIGENT SYSTEMS, 1094-7167/01
- [TB02] Marcelo Tallis, Robert Balzer et al. "The Briefing Associate: Easing Authors into the Semantic Web", IEEE Intelligent Systems, Jan/Feb 2002.
- [Wo75] Woods, W. (1975). What's in a link: foundations for semantic networks. In Representation and Understanding: Studies in Cognitive Science, ed. Bobrow, D. and Collins, A., Academic, New York, pp. 35-82.