

Kapitel 3 Marktstudie

Welche Standards und Tools werden in Unternehmen eingesetzt?

Börteçin Ege ¹

¹ Technische Universität Wien
Email: info@bortecin.com
Web: <http://www.bortecin.com>

1. Semantic Web Anwendungen beginnen, in der Industrie in verschiedenen Bereichen Fuß zu fassen.
2. Hybride Semantic Web-Anwendungen sind die de facto Standard-Lösungen, da ein sehr großer Anteil der Daten immer noch in relationalen Datenbanken gespeichert ist.
3. SKOS ist meistens die erste Wahl bei der Erstellung von Thesauri.
4. W3C Standards genießen sowohl bei den Software-Herstellern als auch bei Kunden eine hohe Akzeptanz.
5. Die Bedeutung von semantischen Technologien werden in Zukunft sowohl im Bereich Big Data als auch in der Industrie 4.0 zunehmen, da sie die Integration und somit die Interoperabilität vereinfacht.

Zusammenfassung

Die Darstellung von insgesamt vier Semantic Web Projekten und die Ergebnisse einer Umfrage im Rahmen dieses Beitrags geben einen kleinen Einblick in die Welt der Hersteller von Semantic Web-Anwendungen. Die einzelne Beschreibungen der Kundenanforderungen in Projekten, die dazugehörigen technischen Umsetzungen wie die entwickelten Architekturen, verwendete Standards, und Tools, sowie konkrete Umsetzungshinweise erweitern den Horizont der Leser zusätzlich.

Aus Börteçin Ege, Bernhard Humm, Anatol Reibold (Hrsg.): "Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften". Springer-Verlag, 2015.

3.1 Einleitung

Im Folgenden werden Semantic Web Projekte dargestellt, die für den Einsatz in verschiedenen Bereichen entwickelt wurden. Die meisten davon sind bereits in der Produktion, d.h. reale Anwendungen mit echten Nutzern. Auf die Kundenanforderungen für jedes Projekt folgen dann auch die dazugehörigen technischen Lösungen bzw. die Architektur der Anwendungen und anschließend die bisherigen Erfahrungswerte der Firmen im Umgang mit verwendeten Semantic Web Standards und Tools. Gegen Ende des Beitrags erwartet den Leser die Ergebnisse einer Umfrage mit weiteren Tips und Hinweisen, dann ein Ausblick und zum Schluss eine Liste von aktuellen Semantic Web Standards und Tools.

3.2 Semantische Suche in Webarchiven (Quantinum AG)

Das erste Beispiel kommt aus der Medienbranche, welches durch die Firma Quantinum AG aus der Schweiz realisiert wurde. Im Rahmen dieses Kundenprojektes musste das Web-Archivierungstool eines Medienunternehmens abgelöst werden. Das abzulösende Tool diente dazu, eine große Anzahl Websites aus dem Bereich Politik zu archivieren und deren Inhalte für spätere Recherchen durchsuchbar zu machen.

3.2.1 Kundenanforderungen

Laut Kundenanforderungen muss das neue Web-Archivierungstool in Kombination mit einer semantischen Lösung der Quantinum AG benutzt werden, um relevante Informationen während der Suche besser aufzuspüren und damit die Recherche zu vereinfachen. In diesem Rahmen soll die Plattform sowohl strukturierte als auch unstrukturierte Daten aus beliebigen Quellen automatisch extrahieren und sie anschließend mit Daten in einem RDF-basierten Knowledge Graph verknüpfen. Für die Annotation der Informationen werden Thesauri verwendet.

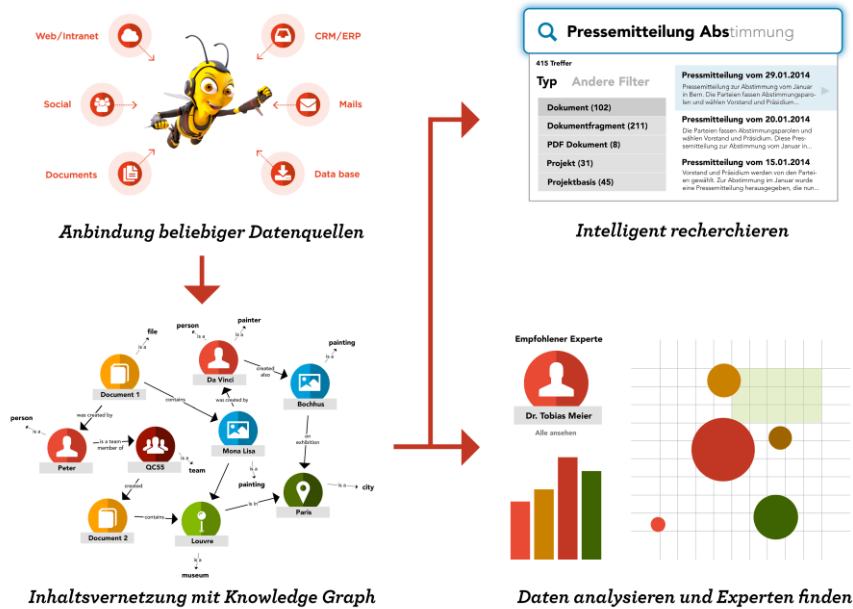


Abb. 3.1 Semantische Suche in Webarchiven im Überblick

3.2.2 Technische Umsetzung

Aufgrund ihrer Erfahrungen aus diversen Kundenprojekten setzt Quantum AG im Rahmen dieses Projektes auf die folgende Architektur und Technologien:

Die direkt einsatzfähige Quantum-Lösung nutzt die Modellierungssprachen RDF¹, RDF-S² und OWL³ und setzt das Framework OpenRDF Sesame⁴ ein. Im Gegensatz zu relationalen Datenbank-Technologien ermöglichen die verwendeten Modellierungssprachen eine flexible Wissensrepräsentation und durch die entsprechenden Inferenzmechanismen auch die Herleitung neuer Fakten. Als Triple Store wird die Java-basierte Bigdata⁵ eingesetzt, nicht zuletzt wegen ihrer Perfor-

¹ RDF (Resource Description Framework):

<http://www.w3.org/2001/sw/wiki/RDF>

² RDF Schema (RDF Vocabulary Description Language):

<http://www.w3.org/2001/sw/wiki/RDFS>

³ Web Ontology Language (OWL): <http://www.w3.org/2001/sw/wiki/OWL>

⁴ OpenRDF Sesame: <http://rdf4j.org/>

⁵ Bigdata: <http://www.w3.org/2001/sw/wiki/Bigdata>

Aus Börteçin Ege, Bernhard Humm, Anatol Reibold (Hrsg.): "Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften". Springer-Verlag, 2015.

manz (sehr hoher paralleler Schreib-/Lese-Durchsatz), hoher Scale-Out-Fähigkeit (Horizontale Skalierung) und der Kompatibilität zu W3C⁶-Standards. Die dabei verwendete Datenbank-Abfragesprache ist SPARQL⁷.

Während für die natürliche Sprachverarbeitung wegen der Verfügbarkeit aktueller NLP Algorithmen und seiner integrierten Entwicklungsumgebung das NLP-Werkzeug GATE⁸ benutzt wird, wird sowohl zur Text-Extraktion als auch zur Extraktion von bestehenden Metadaten das Toolkit Apache Tika⁹ verwendet. Alle im Rahmen des Systems verwendeten Thesauri werden mit SKOS¹⁰ implementiert. Apache Camel¹¹ dient als Enterprise Service Bus zur Datenintegration und Datentransformation. Im Rahmen des Projekts kommt der elasticsearch¹² als Suchserver zum Einsatz, der genau sowie seine Konkurrenz Apache Solr¹³, auf Apache Lucene¹⁴ basiert und in diesem Fall insbesondere wegen seiner hohen Scale-Out-Fähigkeit bevorzugt wird.

3.2.3 Erfahrungswerte

Die Quantinum AG setzt in ihren Projekten auf W3C-Standards, insbesondere auf RDF, RDF-S und OWL für die Ontologie-Modellierung, SKOS für Thesauri-Modellierung, sowie SPARQL als Abfragesprache. Außerdem empfiehlt die Quantinum AG die Verwendung von Triple Stores in Semantic Web Projekten und den Einsatz des Frameworks OpenRDF Sesame.

⁶ W3C (World Wide Web Consortium): <http://www.w3.org/>

⁷ SPARQL (SPARQL Query Language for RDF):
<http://www.w3.org/2001/sw/wiki/SPARQL>

⁸ GATE (General Architecture for Text Engineering): <https://gate.ac.uk/>

⁹ Apache Tika: <http://tika.apache.org/>

¹⁰ SKOS (Simple Knowledge Organization System):
<http://www.w3.org/2001/sw/wiki/SKOS>

¹¹ Apache Camel: <http://camel.apache.org/>

¹² elasticsearch: <http://www.elasticsearch.org/>

¹³ Apache Solr: <http://lucene.apache.org/solr/>

¹⁴ Apache Lucene: <http://lucene.apache.org/>

3.3 Semantische Analyse und Suche in Kundenspezifikationen (Ontos AG)

Das zweite Beispiel ist ein Kundenprojekt aus dem Bereich Maschinenbau, welches wieder durch eine Firma aus der Schweiz realisiert wurde, nämlich durch die Ontos AG. Der Kunde ist die Vertriebsabteilung eines Engineering-Spezialisten für Steuerungslösungen, die bis zu hundert Spezifikationen je Anfrage von internationalen Kunden verarbeiten muss.

3.3.1 Kundenanforderungen

Während der Verarbeitung der Kundenanfragen müssen die Spezifikationen zunächst im Bezug auf deren Relevanz analysiert werden, damit die entsprechenden Angebote vorbereitet werden können. In diesem Rahmen passte die Ontos AG das hauseigene Toolkit Ontos Linked Data Information Workbench (OntosLDIW¹⁵) den Projektanforderungen entsprechend an. Es wird im Rahmen des Projektes derzeit in folgenden Punkten erfolgreich eingesetzt:

- Automatische Analyse multilingualer Dokumente in verschiedenen Formaten
- Unterstützung der mehrsprachigen Wissensstrukturierung von technischen Konzepten
- Anwenderschnittstellen zur Suche von und in Dokumenten

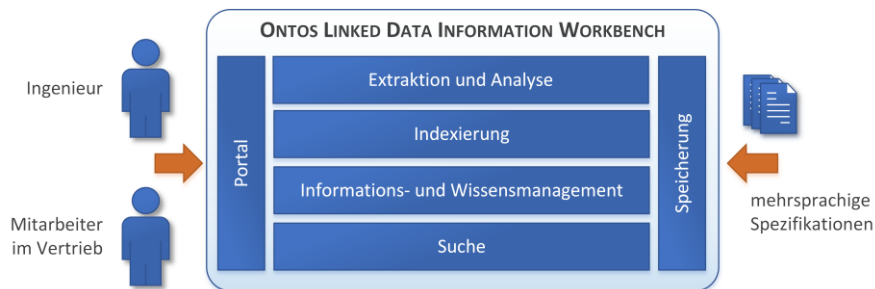


Abb. 3.2 Adaptierte OntosLDIW im Überblick

¹⁵ OntosLDIW: <http://www.ontos.com/products/ontosldiw/>

3.3.2 Technische Umsetzung

Ontos AG verwendet in seiner technischen Lösung das Toolkit Apache Tika für die Textextraktion und Homogenisierung der Dokumente ins PDF-Format, SKOS für die Modellierung des mehrsprachigen Thesaurus, RDF zur Speicherung der Metainformationen, RDF-S als Schemasprache, die hausinterne Datenbank OntoQuad¹⁶ als Triple Store und SPARQL als Abfragesprache. Um eine schnelle Suche und Filterung von Dokumenten zu gewährleisten, wird Apache Solr zusammen mit dem Semantic Vectors Package¹⁷ eingesetzt, welches verschiedene Methoden wie Latent Semantic Indexing (LSI)¹⁸ verwendet. Für das Wissensmanagement kommt der webbasierte und mehrsprachige Ontologie-Editor OntoDix zum Einsatz, der speziell für die Bearbeitung von SKOS-Vokabularen angepasst wurde. Die Suchmasken wurden unter Verwendung des Frameworks Apache Velocity¹⁹ implementiert.

Für die Spracherkennung und die Entfernung von Kopf- und Fußzeilen in Dokumenten werden die hausinterne NLP-Tools OntosMiner²⁰ (ein regelbasiertes NLP-System) sowie Eventos eingesetzt, wobei Letzteres auf Machine Learning Algorithmen beruht, die Latent Semantic Indexing (LSI) verwenden.

3.3.3 Erfahrungswerte

Ontos hält die W3C Standards bereits für ausgereift und setzt gerne auf diese, im Speziellen auf RDF sowie SPARQL und ist davon überzeugt, dass sie genau die richtigen Standards für das Web sind.

3.4 Sicherheit für Banken im Risikomanagement (VICO Research & Consulting GmbH)

Die Überprüfung bestimmter Unternehmenskunden ist mit dem Inkrafttreten von „Basel-III“ wichtiger geworden. Firmen sehen sich daher gezwungen, ihre Business Intelligence Anwendungen immer mehr mit Daten aus den sozialen Medien

¹⁶ OntoQuad: <http://www.ontos.com/products/ontoquad/>

¹⁷ Semantic Vectors Package (semanticvectors):
<https://code.google.com/p/semanticvectors/>

¹⁸ Latent Semantic Indexing (LSI):
http://de.wikipedia.org/wiki/Latent_Semantic_Analysis

¹⁹ Apache Velocity: <http://velocity.apache.org/>

²⁰ OntosMiner: <http://www.ontos.com/products/ontosminer/>

mit Geschäfts- und Marktdaten und aus dem Web anzureichern, welche jedoch größtenteils nicht nur aus unstrukturierten textbasierten Daten, sondern unter anderem auch aus Bildern und Audio- und Videomaterialien bestehen. Im Rahmen des von der Europäischen Kommission geförderten Projektes DOPA²¹ hat VICO gemeinsam mit einer großen deutschen Bank eine neuartige Lösung entwickelt, mit der Banken bestimmte Unternehmenskunden besser erkennen und somit mehr Sicherheit bei der Bewertung von Kundenrisiken erhalten können.

3.4.1 Kundenanforderungen

Das Ziel des Projektes ist mit Hilfe von semantischen Technologien eine Daten- und Auswertungsplattform für Wirtschafts- und Finanzinformationen in Europa zu schaffen, in welcher mit Extraktionsmethoden mögliche Datenquellen in sozialen Medien automatisch ausfindig gemacht werden. Somit können die aus solchen Quellen gewonnenen unstrukturierten Daten mit vorhandenen strukturierten Daten verknüpft werden, um neuen Erkenntnissen zu gelangen. Die Kundenanforderungen waren:

- Automatisierte Entdeckung verschiedener Datenquellen
- Automatisierte Verlinkung von Datensätzen und Datentöpfen, um verwandte und ähnliche Informationen aus unterschiedlichen Quellen zu verknüpfen
- Visualisierung der verknüpften Daten
- Hochskalierbare Informationsverarbeitung in Form von Data Supply Chains auf einer verteilten Umgebung, um Prozessketten effizient und intelligent zu steuern.

3.4.2 Technische Umsetzung

Wie auch in der Abbildung 3.3 dargestellt wird, werden im Rahmen der Lösung über ein Data Supply Chain Environment unterschiedliche Datenquellen beliebiger Anbieter verknüpft. Die semantische Verknüpfung liefert hierbei die Entity Annotation Engine des Partners Okkam²². Sie sorgt für eine Annotierung der vorhandenen Datensätze und somit für den semantischen Bezug zwischen unterschiedlichen Datensets. Kunden und Lösungsanbieter können darauf basierend be-

²¹ DOPA: <http://www.dopa-project.eu/index.php>

²² Okkam: <http://www.okkam.biz/>

liebige ETL (Extract Transform Load) Prozesse auf den Daten vornehmen, die hochparallelisiert ausgeführt werden.

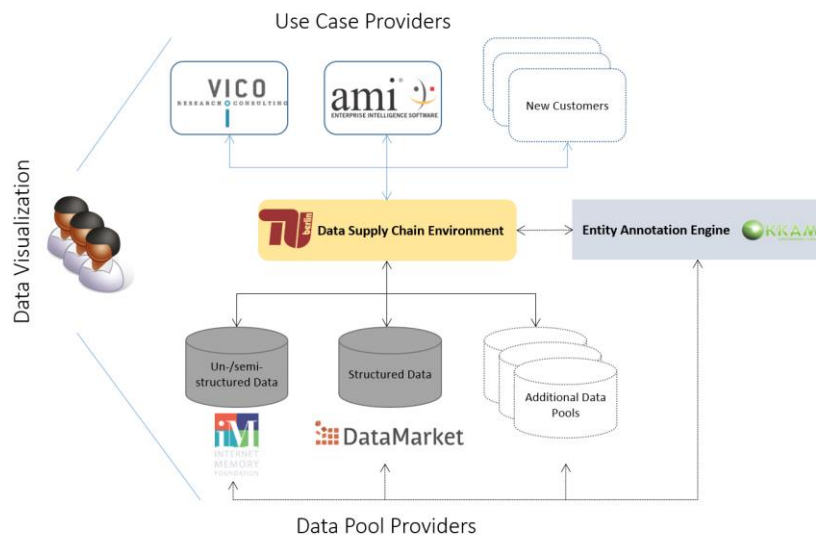


Abb. 3.3 DOPA im Überblick

Die Abbildung 3.4 beschreibt die technische Architektur von DOPA, die auf Stratosphere²³ basiert, welches vor kurzem in die Apache Incubator²⁴ Projekt unter dem Namen Apache Flink²⁵ aufgenommen wurde und unter diesem Namen weiterentwickelt wird. Die einzelnen Komponenten sind eher im Bereich Big Data als im Bereich Semantic Web angesiedelt. Wichtig ist jedoch, dass die Semantic Web Techniken zur Verknüpfung der Datenquellen unerlässlich sind und somit auf der Anwendungsebene eine der wichtigsten Rollen spielen.

²³ Stratosphere: <http://stratosphere.eu/>

²⁴ Apache Incubator: <http://incubator.apache.org/>

²⁵ Apache Flink: <http://flink.incubator.apache.org/>

Aus Börteçin Ege, Bernhard Humm, Anatol Reibold (Hrsg.): "Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften". Springer-Verlag, 2015.

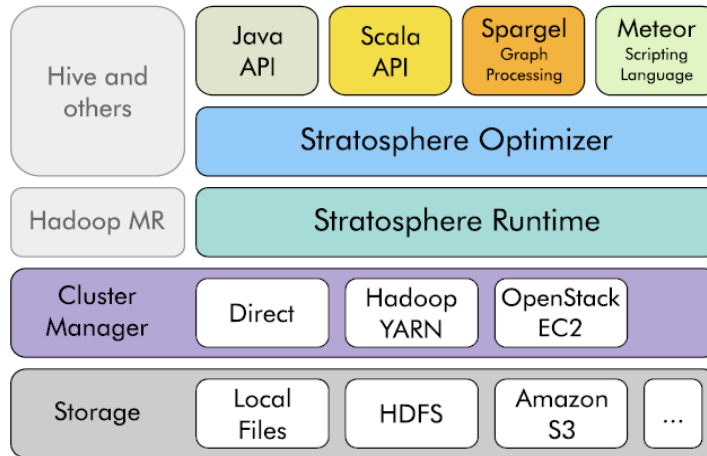


Abb. 3.4 DOPA Architektur basierend auf Stratosphere

DOPA verwendet als Teil der Matching-Lösung Ontologien, die in RDF und OWL modelliert sind. Da die DOPA Lösung hoch skalierbar ist, verzichtet das Projekt auf Triple Stores und verwendet Big Data Technologien wie bspw. Apache HBase²⁶ und Phoenix²⁷. Dennoch werden über andere Triple Stores mit SPARQL Endpoints semantisch strukturierte Daten im RDF und OWL Format dynamisch abgerufen um eine Anreicherung zu erzielen. Das Projekt ist seit April 2014 erfolgreich abgeschlossen.

3.4.3 Erfahrungswerte

Wie auch oben beschrieben setzt die VICO Research & Consulting GmbH in ihren Semantic Web Projekten auf W3C Standards wie RDF und SPARQL. Die DOPA Lösung führt durch den Partner Okkam eine eigene ID Landschaft ein, die zwar den W3C Standards entspricht, jedoch Duplikate vermeidet. Für eine hohe Skalierbarkeit werden Funktionen von Triple Stores in den Technologien Apache HBase, Phoenix und Apache Solr nachempfunden. Als Entwicklungsumgebung von Semantic Web Projekten empfehlen die Entwickler von VICO Research die Linux-Umgebung, weil gerade im Big Data Bereich serverseitige Entwicklungen in Linux über commandline Tools komfortabler gestaltet werden können. VICO

²⁶ Apache HBase: <http://hbase.apache.org/>

²⁷ Apache Phoenix: <http://phoenix.apache.org/>

Research geht davon aus, dass Ontologien in Zukunft immer mehr eine zentrale Rolle insbesondere im Big Data Bereich spielen werden, da ohne sie der Zusammenschluss von unstrukturierten und strukturierten Daten sich schwieriger gestalten würde.

3.5 Interaktive Fahrzeugdiagnose (semafora GmbH)

Ein anderes interessantes Beispiel für die technische Umsetzung von semantischen Technologien ist ein Kundenprojekt der Firma semafora GmbH im Bereich Fahrzeugdiagnose in KFZ-Werkstätten. Der Kunde ist ein international renommiertes Unternehmen für die Entwicklung von Antriebssystemen (Verbrennungsmotoren, Getrieben, elektrischen Antrieben) sowie dazugehörige Prüftechnik und Simulation. Das Unternehmen rüstet KFZ-Werkstätten und Prüfzentren mit intelligenten Testsystemen und Diagnosetechnik aus, um einen schnelleren Kundenservice und schnellere Reparaturzeiten von Fahrzeugen unterschiedlicher Hersteller durch interaktive Diagnosen zu erreichen.

3.5.1 Kundenanforderungen

Zu den wichtigsten Kundenanforderungen zählen die Nutzung der Software in der Werkstatt, die Integration eines Bestellsystems, die Anbindung an vorhandene Diagnosewerkzeuge, sowie die Übernahme der Ergebnisse für die Fehlerbehebung.

Die Firma semafora hat in diesem Rahmen das hauseigene Ratgebersystem SemanticGuide²⁸ bei dem Kunden angepasst und implementiert. SemanticGuide ist ein ontologiebasiertes Ratgebersystem, mit dem beratungs- und wissensintensive Prozesse unterstützt werden können. SemanticGuide sammelt das Wissen von Experten in einer Wissensbasis und stellt dann dieses allen Mitarbeitern des Unternehmens zur Verfügung, in diesem Falle KFZ-Servicetechnikern in Form einer geführten Fehlersuche. Nach Durchführung einer Fehlersuche kann der Techniker Feedback oder Vorschläge zur Problemfindung abgeben. Dies erlaubt die kontinuierliche Verbesserung des Fehlerfindungsprozesses: Das Ratgebersystem lernt durch seine Nutzung und nimmt so nach und nach das Expertenwissen im Team in seiner Wissensbasis auf.

Der Einsatz von SemanticGuide bringt dem Kunden die folgenden Vorteile:

²⁸ SemanticGuide: <http://www.semafora-systems.com/de/loesungen/semanticguide/>

- Schnellerer Kundenservice und somit kürzere Reparaturzeiten durch interaktive Diagnose
- Einfache Integration in bestehende Anwendungen (z.B. Teilekataloge)
- Automatische Auswahl von richtigen Ersatzteilen (automatische Berücksichtigung von Marke, Modell, Motor, Baugruppe und Bauteil)

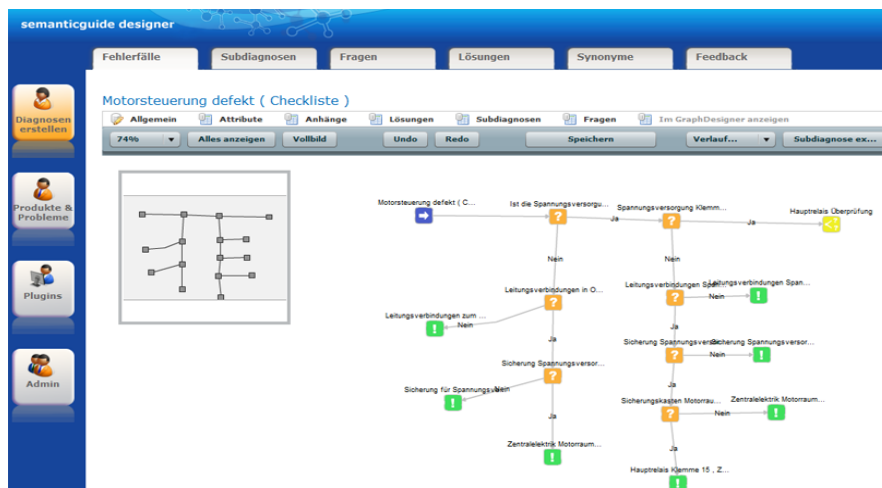


Abb. 3.5 Redaktionskomponente von SemanticGuide

Eine wichtige Besonderheit von *SemanticGuide* gegenüber anderen Ratgeber-systemen ist, dass er seine Wissensdomäne und den Diagnoseablauf vollständig in Form von Ontologien beschreibt und damit eine mächtige Grundlage für die Verarbeitung komplexer Zusammenhänge schafft, die wiederum von der semantischen Reasoning-Engine *OntoBroker*²⁹ verarbeitet werden. So können Schlüsse nicht nur aus dem Vorhandensein von Fehlerkriterien gezogen werden (die Eingrenzung von Lösungen wird durch Anwendung von Inferenzregeln auf einer Ontologie realisiert), sondern zusätzlich lassen sich durch die semantische Modellierung von Produktkategorien, Fehlertypen (Fehlercode, allgemeines Symptom), der Struktur von Baugruppen und deren Funktionszusammenhängen weitere Rückschlüsse ziehen, was die Diagnoseläufe abkürzen kann. Eine weitere Verkürzung der Abläufe lässt sich durch eine vollautomatisierte Abarbeitung der Fehlersuche auf Basis von Maschinendaten realisieren.

Mit der Redaktionskomponente von SemanticGuide können Service-Experten ihr Wissen mit wenig Aufwand in Form einer graphisch editierbaren Baum- bzw.

²⁹ OntoBroker: <http://www.w3.org/2001/sw/wiki/OntoBroker>

Netzwerkstruktur einpflegen, ohne sich über die darunterliegende semantische Struktur Gedanken machen zu müssen. Diese Vereinfachung der semantischen Modellierung für Endanwender hat sich als wichtige Voraussetzung für deren Praxistauglichkeit herausgestellt.

3.5.2 Technische Umsetzung

SemanticGuide baut auf OntoBroker auf, einer Integrationsplattform für semantische Applikationen. Über Webservice-Schnittstellen (SOA-Framework) sowie Java APIs können Anwendungen wie SemanticGuide auf semantische Daten zugreifen, die in OntoBroker verarbeitet werden. Die Integration von bestehenden Datenquellen erfolgen durch Konnektoren, die neben RDBMS (Oracle, Microsoft SQL Server, MySQL) auch Suchmaschinen, ERP-Systeme uvm. Anbinden.

Die zugrundeliegenden Reasoner kombinieren verschiedene Verfahren und erlauben die automatische Auswahl der geeignetsten Verfahren gegeben durch die Ontologie. Dies verbessert das Skalierungsverhalten gerade bei komplexen Regeln. OntoBroker unterstützt dabei alle Semantic Web-Empfehlungen des W3C: RDF(S), OWL, SPARQL sowie RIF³⁰ und verfügt über Reasoningfähigkeiten für OWL RL³¹, RDF sowie ObjectLogic³².

ObjectLogic ist eine formale Beschreibungssprache für Ontologien, die neben den in OWL und RDF(S) möglichen Sprachkonstrukten noch Regeln, Abfragen sowie sogenannte „Builtins“ unterstützt. Builtins sind prozedurale Operationen, die z.B. Arithmetik und Datenintegration realisieren und anwendungsspezifisch entwickelt werden können. ObjectLogic baut auf der Wissensrepräsentationssprache F-Logic³³ auf. SemanticGuide führt alle Logikoperationen im OntoBroker aus und verwendet dabei ObjectLogic, wobei Regeln, Abfragen und Datenintegration eine zentrale Bedeutung bei der Umsetzung haben.

3.5.3 Erfahrungswerte

Die Firma semafora setzt auf die semantische Datenintegration, bei der RDBMS-Inhalte als Instanzen einer Ontologie integriert werden können. In diesem Rahmen werden relationale Datenbanken wie Oracle, Microsoft SQL Server, MySQL und das Datenbanksystem H2 als Reasoner Backend verwendet, um Fakten (Inferenzergebnisse) zu speichern. Als semantische Abfragesprachen benutzt semafora F-

³⁰ Rule Interchange Format (RIF): <http://www.w3.org/2001/sw/wiki/RIF>

³¹ OWL RL: <http://www.w3.org/2001/sw/wiki/OWLRL>

³² ObjectLogic: <http://ontorule-project.eu/showcase/ObjectLogic>

³³ F-Logic: <http://de.wikipedia.org/wiki/F-Logic>

Logic bzw. ObjectLogic. semafora verwendet in ihren Projekten nicht nur RDF-Vokabulare, sondern auch andere semantische Vokabulare und zwar insbesondere die in F-Logic bzw. ObjectLogic erstellten (nach Angaben von semafora liefern kommerzielle Vokabular-Anbieter zudem häufig nicht in RDF). semafora entwickelt ihre Projekte unter Windows, da nicht zuletzt Windows eine größere Verbreitung in der Industrie als Linux findet. Außerdem setzt semafora in ihren Projekten auf die Inferenzmechanismen OntoBroker und KAON2³⁴, verwendet als Ontologie-Modellierungstool OntoStudio³⁵ und empfiehlt für die Erstellung von Thesauri den W3C-Standard SKOS. Laut semafora sind semantische Technologien bereits seit Jahren im Markt präsent, jedoch nur in Nischen.

3.6 Quo Vadis?

Semantic Web ist eine relativ junge Disziplin. Obwohl in diesem Bereich seit Jahren erfolgreiche Projekte realisiert sind, bleiben die meisten nicht nur der Öffentlichkeit, sondern auch den meisten Informatikern verborgen. Denn es gibt wenige Veröffentlichungen über Semantic Web Anwendungen, die in der Praxis eingesetzt werden. Oft kommt dazu, dass ein Teil dieser Projekte nur im militärischen Bereich realisiert und benutzt wird. Dies bedeutet wiederum, dass die Öffentlichkeit von diesen Projekten meistens nichts erfahren darf. So steckt die ganze Disziplin in einem Dilemma: Je weniger bekannt ist, dass es in produktiven Umgebungen laufende Semantic Web Projekte gibt, desto weniger Firmen und potentielle Kunden interessieren sich für die Semantic Web Technologien und deren Anwendungen.

Im Rahmen dieses Beitrages wird ein „Snapshot“ des Markts erfasst. Dies hilft hoffentlich dem Leser, sowohl die Architektur von Semantic Web Anwendungen zu verstehen als auch einen Überblick über die verwendeten Tools zu bekommen. Auch die im Rahmen des Beitrags durchgeführte Umfrage gibt einen Einblick in den Markt.

3.7 Umfrage-Ergebnisse

An dieser Umfrage nahmen die folgenden Firmen teil: Quantinum AG (Schweiz) semafora GmbH (Deutschland), Semantic Web Company GmbH (Österreich), Vico Research & Consulting GmbH (Deutschland) und Wolters Kluwer GmbH (Deutschland).

Aus der Umfrage geht als erstes hervor, dass in Semantic Web Projekten neben Triple Stores und der semantischen Abfragesprache SPARQL auch relationale Datenbanken und somit auch die relationale Abfragesprache SQL weiterhin ver-

³⁴ KAON2: <http://www.w3.org/2001/sw/wiki/Kaon2>

³⁵ OntoStudio: <http://www.w3.org/2001/sw/wiki/OntoStudio>

wendet werden. Die meisten Firmen empfehlen jedoch die Verwendung von nativen Triple Stores in Semantic Web Projekten und bei Bedarf auch den Einsatz von NoSQL-Datenbanken³⁶, die als Triple Stores implementiert sind. Eine der teilnehmenden Firmen gibt als Hauptargument für die Verwendung von Triple Stores die konzeptionell gegebene große Freiheit und Flexibilität im Datenbankdesign an. So werden zum einen durch die Darstellung als Graphen keine starren Datentabellen benötigt, die insbesondere bei größeren Änderungen und Erweiterungen im Lebenszyklus Nachteile haben. Zum zweiten stellt die einfache und skalierbare Erweiterung der integrierten Datenquellen eine Grundanforderung dar, welche sich mit den RDF-Tripeln gleichfalls leichter umsetzen lässt, insbesondere wenn man berücksichtigt, dass sich auch die Strukturen der Datenquellen stetig ändern und weiterentwickeln.

In der Umfrage haben die meisten Firmen auch angegeben, dass sie RDF als Basis von Semantic Web Vokabularen und Ontologien sehen und verhältnismäßig einfach und dennoch expressiv genug finden. Wie es sich aus der Umfrage auch herauskristallisiert hat, ist SKOS die beliebteste Modellierungssprache für die Erstellung von Thesauri. Als Grund gaben einige Firmen an, dass SKOS nicht nur ein etablierter Standard für Thesaurus-Entwicklung ist, sondern bei Verwendung mit anderen RDF Vokabularen und Ontologien leichter verknüpft werden kann, was wiederum die Ausdrucksstärke eines Thesaurus erweitern kann.

Aus der Umfrage geht auch hervor, dass mittlerweile auch RDFa³⁷ in Semantic Web Projekten vermehrt Verwendung findet. So haben zwei der an der Umfrage beteiligten Firmen angegeben, dass sie RDFa zur Extraktion von strukturierten Daten aus Webseiten, zur Veröffentlichung von Linked Data im Web, sowie für Semantic SEO einsetzen.

Auch die Modellierung und Verwendung von OWL-Ontologien in Semantic Web Projekten sind mehr verbreitet als angenommen. So haben vier von fünf beteiligten Firmen angegeben mit OWL-Ontologien gearbeitet zu haben. Eine Firma hat zusätzlich angegeben, dass sie OWL Fragmente dazu verwendet, um Regeln davon abzuleiten, die bei der Modellierung von Wissensgraphen u.a. für Konsistenz-Checks eingesetzt werden. Die gleiche Firma weist aber auch darauf hin, dass mit SPARQL 1.1³⁸, OWL teilweise obsolet geworden ist (z.B. durch den Einsatz von property paths).

Nicht zuletzt geht aus der Umfrage auch hervor, dass die W3C-Standards eine große Akzeptanz im Markt finden, weil die Kunden und Firmen es für wichtig halten, mit weltweit etablierten Standards zu arbeiten. In diesem Kontext wünschen sich die Firmen jedoch noch mehr Aktivitäten von W3C wie die Standardisierung von RDF-Vokabularen und die Übernahme von Praxisanforderungen in die Standards.

³⁶ NoSQL: <http://de.wikipedia.org/wiki/NoSQL>

³⁷ RDFa (RDF in Attributes): <http://www.w3.org/2001/sw/wiki/RDFa>

³⁸ SPARQL 1.1: <http://www.w3.org/TR/sparql11-overview/>

3.8 Semantic Web Standards & Tools

In den letzten Jahren sind sehr viele neue Standards und Tools entstanden sowohl für die Entwicklung von Semantic Web Anwendungen als auch rund um die Semantic Web Technologien. Hier folgt eine allgemeine Übersicht der derzeit wichtigsten Standards und Tools für die Entwicklung von Semantic Web Anwendungen.

AllegroGraph	Kommerzieller RDF Triple Store
Apache Camel	Regelbasierte Routing- und Konvertierungsengine
Apache HBase	Verteilte, skalierbare Big Data Datenbank
Apache Jena	Java-basiertes Semantic Web Framework
Apache Lucene	Programmibibliothek zur Volltextsuche
Apache Solr	Suchserver, der auf Apache Lucene basiert
Apache Stanbol	Wiederverwendbare Komponente für das Semantic Content Management
Apache Velocity	Javabasierte Template-Engine
Apache Tika	Tool zur Extraktion von Text und Metadaten
Bigdata	Kommerzieller RDF Triple Store
Cubeviz	Ein facetierter Browser für statistische Daten
DAML+OIL	Der Vorgänger der Web Ontology Language (siehe OWL)
DBpedia	Ontologie für allgemeines Weltwissen, basierend auf Wikipedia
DBpedia Spotlight	Entity Extraction Tool, das basierend auf DBpedia oder auch auf anderen Ressourcen, Entitäten aus Dokumenten extrahiert
DL-Learner	Werkzeug, um Schemata in Datensets zu identifizieren
Elasticsearch	Suchserver, der auf Apache Lucene basiert
FaCT++	OWL DL Reasoner
FIPA	Standardisierungsgremium, das sich mit den Kommunikationsgrundlagen von Software-Agenten beschäftigt
FOAF	Ontologie von Personen, deren Aktivitäten und Beziehungen zu anderen Personen und Objekten
F-Logic	Formale Sprache ähnlich zu RDF und OWL
GATE	NLP Werkzeug
GeoSPARQL	Standard der Open Geospatial Consortium zur Repräsentation und Abfrage von Geospatial Linked Data
GoodRelations	E-Commerce Vokabular in RDFa
GraphVizdb	Eine datenbankbasierte Plattform für interaktive Visua-

Aus Börteçin Ege, Bernhard Humm, Anatol Reibold (Hrsg.): "Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften". Springer-Verlag, 2015.

	lisierung von umfangreichen Graphen
GRDDL	Gleaning Resource Descriptions from Dialects of Languages: Technik/Format, um aus XHTML- und XML-Dokumenten RDF-Triples zu gewinnen
Hercules	JavaScript-basiertes Framework zur Entwicklung von Semantic Web Anwendungen
HermiT	Reasoner für OWL-Ontologien
Hypertree	Kommerzielle Javascript-Anwendung zur dynamischen Visualisierung von Ontologien
KAON2	Infrastruktur für die Verwaltung von OWL-DL, SWRL und F-Logic Ontologien
LinkedGeoData	Ein Semantic Web Projekt, um Informationen aus OpenStreetMap mit den räumlichen Dimensionen zu verknüpfen
LOD Management Suite	Ein Verwaltungstool zur Durchführung von ETL Jobs (Extract - Transform - Load)
O4Store	SQL-basierte und kommerzielle Implementierung eines Quad-Store-Datenhaltungssystems
ObjectLogic	Deduktive, objektorientierte Datenbanksprache
OntoBroker	Kommerzielle Inferenzmaschine zur Verarbeitung von Ontologien
Ontology4	Kommerzielle Semantic Web Workbench zur Erstellung und Nutzung von Ontologien
OntoMaven	Eine Erweiterung zu Apache Maven für verteilte Ontologie Entwicklung und Ontologie-basierte Software Entwicklung
OntoQuad	Kommerzieller Triple Store
OntosLDWI	Kommerzielles Linked Data Information Workbench für das Verwalten und Steuern des Linked Data Lebenszyklen
OntosMiner	Kommerzielles NLP-Tool
OntoStudio	Kommerzielles Tool zur Erstellung und Bearbeitung von Ontologien
OntoWiki	Metadaten-Tool zum Bearbeiten graphbasierter Daten
OpenLink Virtuoso	Kommerzieller RDF Triple Store
OpenRDF Sesame	Java-basiertes Semantic Web Framework
ORE	Ontology Repair and Enrichment (ORE) hilft die Inkonsistenzen in der Schemastruktur einer Ontologie zu identifizieren
Oracle Spatial 11g	Kommerzielle Graphen-basierte Version des Oracle Datenbank Management Systems
OWL	Web Ontology Language (OWL) ist eine Modellie-

Aus Börteçin Ege, Bernhard Humm, Anatol Reibold (Hrsg.): "Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften". Springer-Verlag, 2015.

	rungssprache, die auf RDF und RDF-S aufbaut
OWLIM	Kommerzieller Triple Store
Pellet	OWL 2 – Reasoner
PoolParty	Kommerzielles Metadaten-Werkzeug zum Management kontrollierter Vokabulare
Protégé	Ontologie-Editor
RacerPro	Kommerzieller OWL Reasoner und Inferenzserver
RDF	Resource Description Framework: Modellierungssprache zur Beschreibung von Metadaten
RDFa	Resource Description Framework in Attributes
RDF-S	Resource Description Framework-Schema: Beschreibungssprache für RDF
Redland	RDF Entwicklungsumgebung
RIF	Rule Interchange Format
Rsine Notification Tool	Tool zur Benachrichtigung über Veränderungen innerhalb definierter Bereiche eines Datensets
RuleML	Rule Markup Language
R2RML	Relationale Datenbanken zu RDF - Mapping Language
R2RML Parser	Tool zum Export von relationalen Inhalten als RDF-Graphen
SemanticGuide	Kommerzielles Ontologie-basiertes Ratgebersystem der Firma semafora
Sesame	Siehe OpenRDF Sesame
Silk	Matchingtool, um zwei verschiedene Datenquellen auf gleiche Inhalte zu untersuchen und diese zu verlinken
Sindice Pivot Browser	Browser, um zusammenhängende Daten über mehrere Dimensionen durch eine facetierte Suche zu analysieren
SKOS	Simple Knowledge Organisation System (SKOS): Sprache zur Modellierung von Thesauri und anderen kontrollierten Vokabularen
SPARQL	Simple Protocol and RDF Query Language (SPARQL): RDF-Abfragesprache
TopBraid Suite	Kommerzielle Semantic Web Entwicklungsumgebung
TopicFinder	Kommerzielles Tool zur automatischen Textklassifikation zur Themenerkennung und Dokumentanalyse
Triplfy	RDF-Generator, der den Inhalt einer relationalen Datenbank in die entsprechenden RDF-Triples konvertiert
Valiant	Transformationstool, mit dem die XML-basierten Daten im Batchmodus per XSLT-Script in RDF umgewandelt

Aus Börteçin Ege, Bernhard Humm, Anatol Reibold (Hrsg.): "Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften". Springer-Verlag, 2015.

	werden können
Virtuoso	Siehe OpenLink Virtuoso
W3C	World Wide Web Consortium
YAGO	Yet Another Great Ontology (YAGO): Wissensdatenbank ähnlich DBpedia

Tabelle 3.1 Semantic Web Standards und Tools

Aus Börteçin Ege, Bernhard Humm, Anatol Reibold (Hrsg.): "Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften". Springer-Verlag, 2015.

3.9 Ausblick

Eine vierte Industrie Revolution steht der Welt bevor, nämlich *Industrie 4.0*. Industrie 4.0 ist ein Zukunftsprojekt in der Hightech-Strategie der deutschen Bundesregierung, mit dem die Informatisierung der klassischen Industrien vorangerieben werden soll. Mit Industrie 4.0 sollen neue Produktionsmethoden eingeführt werden, die bisher so nicht möglich waren. Bei Industrie 4.0 handelt es sich um Intelligente Fabriken, die Produkte höchst autonom herstellen können. Um diesen Zweck zu erreichen, müssen jedoch vor allem die Produktionsmaschinen und deren Produkte untereinander vernetzt werden. Das ist durch den Einsatz semantischer Technologien erreichbar. Denn ohne semantische Interoperabilität der Maschinen ist eine intelligente Produktion nicht realisierbar [1]. Weltweit planen ein großer Teil der klassischen Industrie (z.B. die Produktionstechnik) schon heute, ihre Produktionsmethoden Schritt für Schritt auf Industrie 4.0 umzustellen.

Die Semantic Web Technologien und Anwendungen etablieren sich in der Industrie zunehmend, und ich rechne daher damit, dass die Semantic Web Anwendungen spätestens mit der Einführung der Industrie 4.0 allgegenwärtig sein werden. Die dafür notwendigen Standards sind größtenteils bereits vorhanden und sie werden insbesondere von der W3C ständig weiterentwickelt.

Nur, vom heutigen Standpunkt aus, erweist sich die Entwicklung und Wartung von notwendigen Tools für die Entwicklung von Semantic Web Projekten als Achillesferse der ganzen Disziplin, obwohl mit dem LOD2-Projekt³⁹ in letzter Zeit auch in dieser Richtung erhebliche Fortschritte erzielt werden konnte. Für die Interessenten bietet das vor kurzem erschienene Buch *Linked Open Data - Creating Knowledge Out of Interlinked Data* [2] einen guten Einblick in die Ergebnisse und im Rahmen des Projektes entwickelte Werkzeuge und Tools. Wie es jedoch auch im Beitrag *Verlage müssen sich neu erfinden* (Dirschl, C., Eck, K.) in unserem Buch betont wird, kommt immer noch ein großer Anteil von Tools nur aus dem akademischen Umfeld und ist meistens nicht für eine industrielle Nutzung gedacht. Dies gefährdet oft die Weiterentwicklung und somit die Verfügbarkeit von Tools. Aus diesen Gründen ist die Weiterentwicklung von Standards allein nicht mehr ausreichend, solange den Entwicklern und Anwendern die notwendigen Tools nicht wirklich zur Verfügung stehen. Das heißt, es gibt in diesem Bereich noch sehr viel zu tun und nachzuholen.

Danksagung

Besonderen Dank für die wertvollen Mitwirkungen bei der Erstellung der Beiträge und Umfragen an Daniel Hladky (Ontos AG), Martin Voigt (Ontos AG), Marco Kläy (Quantinum AG), Reto Trinkler (Quantinum AG), Gregor Heinrich (semafo-

³⁹ LOD2: <http://lod2.eu/Welcome.html>

ra Systems GmbH), Joachim Redmer (semafora Systems GmbH), Andreas Blumauer (Semantic Web Company GmbH), Yasan Budak (Vico Research & Consulting GmbH), Philipp Tiedt (Vico Research & Consulting GmbH), Christian Dirschl (Wolters Kluwer Deutschland GmbH), Katja Eck (Wolters Kluwer Deutschland GmbH).

Literatur

- [1] Ege, B. "4. Endüstri Devrimi" (Industrie 4.0), TÜBITAK Bilim ve Teknik Dergisi (The Scientific and Technological Research Council of Turkey), s. 26-29, Mai 2014
- [2] Auer, S., Bryl V., Tramp S., "Linked Open Data - Creating Knowledge Out of Interlinked Data", Springer Open, Juli 2014, <http://link.springer.com/book/10.1007/978-3-319-09846-3>, zuletzt abgerufen am 27.10.2014

Aus Börteçin Ege, Bernhard Humm, Anatol Reibold (Hrsg.): "Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften". Springer-Verlag, 2015.