

Semantische Suche zur Vertriebsunterstützung

-

von Andreas Nierlich & Eddie Moench
ontoprise GmbH
info@ontoprise.de
<http://www.ontoprise.de>

Jedes Unternehmen verfügt über enorme Ressourcen an gesammeltem Know-how. Verbunden mit den Fähigkeiten der Mitarbeiter sind diese Ressourcen der wichtigste Aktivposten – nämlich die Basis für Erfolg und Wachstum. Die Herausforderung besteht darin, diese Ressourcen jederzeit für die tägliche Arbeit nutzbar zu machen. Vor allem deswegen, weil diese oftmals bereits verfügbar, für Mitarbeiter aber nicht einfach zugänglich sind. In wissensintensiven Unternehmen verbringen Mitarbeiter zu viel Zeit mit Informationsrecherche. 51 % der Mitarbeiter verwenden darauf mehr als zwei Stunden pro Tag, das sind zehn Stunden in der Woche. Dabei ist erfahrungsgemäß nur etwa jede fünfte Suche erfolgreich. Zudem werden schätzungsweise 80% aller Entscheidungen auf der Basis unstrukturierter Informationen getroffen.

Die hier dargestellte Lösung beschreibt ein semantisches Suchsystem, welches das Auffinden von Dokumenten und Datenbank-Informationen unterstützt und so Lösungen zur semantischen Integration von verteilten Datenquellen bietet. Anhand eines Praxisbeispiels werden die eingesetzten Technologien illustriert. Es zeigt den Einsatz eines ontologiebasierten Systems zur Suche und Integration von Dokumenten und Daten in einem großen deutschen Telekommunikationsunternehmen. Dabei werden verschiedene Quellen – z.B. ein DMS System und ein Intranet - integriert und in diesen mit Hilfe von Ontologien übergreifend nach Inhalten gesucht. Durch vertiefte Erläuterungen der eingesetzten Methoden und Darstellungen der verwendeten Werkzeuge wird die zentrale Aussage des Beitrages erweitert.

Einleitung

Ob Sie in der Entwicklung, im Service, im Marketing oder im Vertrieb arbeiten, ob Sie als Fachkraft ständig neue Aufgaben bewältigen oder sich als Geschäftsführer erfolgreich dem Wettbewerb stellen: Sie benötigen ständig aktuelle und relevante Informationen. Es müssen immer mehr Entscheidungen in immer kürzeren Zyklen getroffen werden. Informationen gibt es ausreichend, in Unternehmen, im Austausch zwischen Unternehmen und im Internet. Das Problem dabei ist, dass es zu viele Informationen gibt. Die Kunst besteht darin, relevante Informationen für jedermann einfach nutzbar zu machen. Um dem gerecht zu werden, müssen IT-Systeme flexibler werden, sie müssen die Anforderungen der Anwender besser verstehen und sie müssen diese schneller und besser unterstützen. Um einen einfachen und breiten Zugang zu der komplexen und stetig steigenden Menge an Wissen zu erhalten, bieten sich semantische Technologien an.

Der Einsatz dieser Technologie kann Unternehmen aller Branchen dabei unterstützen, Entwicklungs-, Service- und Vertriebsprozesse so zu gestalten, dass dadurch die Effizienz gesteigert und die Qualität erhöht werden kann. Verschiedene Projekte in der Automobilindustrie, im Maschinenbau, im öffentlichen Sektor und in der Telekommunikationsindustrie zeigen auf, welches Automatisierungs-Potenzial für die Wirtschaft in semantischen Technologien steckt: Um wettbewerbsfähig zu sein, müssen Unternehmen ständig schneller und günstiger entwickeln, die Qualität in der Entwicklung und im Service erhöhen sowie die Effizienz im Vertrieb steigern. Ontologien als Ausprägung semantischer Beschreibungssprachen sind auf dem Weg, die Geschäftssprache der Zukunft zu werden. Sie schlagen eine Brücke zwischen Menschen und Computern, zwischen Prozessen und Wissen sowie zwischen IT-Experten und Fachabteilungen und ermöglichen dadurch die gewünschten Synergie-Effekte.

Im Folgenden wird ein Best Practice der Anwendung semantischer Technologien dargestellt. Es beschreibt den Einsatz semantischer Suche zur Vertriebsunterstützung bei einem Telekommunikationsunternehmen. Abschließend wird die aktuelle und zukünftige Bandbreite der Einsatzgebiete semantischer Technologien kurz zusammengefasst.

Semantische Suche zur Vertriebsunterstützung

Jedes Unternehmen verfügt über enorme Ressourcen an gesammeltem Know-how. Verbunden mit den Fähigkeiten der Mitarbeiter sind diese Ressourcen der wichtigste Aktivposten – nämlich die Basis für Erfolg und Wachstum. Die Herausforderung besteht darin, diese Ressourcen jederzeit für die tägliche Arbeit nutzbar zu machen. Vor allem deswegen, weil diese oftmals bereits verfügbar, für Mitarbeiter aber nicht einfach zugänglich sind. In wissensintensiven Unternehmen verbringen Mitarbeiter zu viel Zeit mit Informationsrecherche. 51 % der Mitarbeiter verwenden darauf mehr als zwei Stunden pro Tag, das sind zehn Stunden in der Woche. Dabei ist erfahrungsgemäß nur etwa jede fünfte Suche erfolgreich. Zudem werden schätzungsweise 80% aller Entscheidungen auf der Basis unstrukturierter Informationen getroffen.

Seit Jahren sprechen wir von der steigenden Menge an Informationen, welche es Mitarbeitern immer schwerer macht, ihre Arbeit effizient zu erledigen und adäquate Entscheidungsunterstützung zu erhalten. Nicht nur im Internet nehmen die Informationen zu, auch innerhalb der Unternehmen steigt die Menge an verfügbaren Informationen dramatisch. Die Menge unstrukturierter Informationen verdoppelt sich alle 3-6 Monate, deren Anteil in Unternehmen beträgt mittlerweile über 70 %. Warum ist es für Anwender so schwer, die richtigen Informationen zu finden und wie kann man dieses immense Optimierungspotenzial nutzen? Erfahrungen zeigen, dass man einen Großteil der Probleme drei Fragestellungen zuordnen kann:

- Die Anwender wissen nicht, wo sie relevante Informationen suchen können, da diese auf viele verschiedene Anwendungen verteilt sind.
- Anwender sind beim Stellen adäquater Anfragen in der Regel überfordert.
- Bei der Suche erhält man zu viele Ergebnisse, in denen der Anwender die relevante Information nicht entdecken kann.

Wir wollen aufzeigen, inwiefern der Einsatz einer semantischen Suche helfen kann, diese Probleme zu minimieren. Eine semantische Suche kann durch Ontologien Verbesserungen erzielen, indem diese

- durch inhaltliche Integration einen einheitlichen Zugriff auf verteilte Quellen ermöglicht,
- durch das Einbinden von Hintergrundwissen den Anwender bei der Suche unterstützt und ihm dabei hilft, seine Suchanfragen zu verbessern sowie
- die Ergebnisse durch Knowledge Retrieval Techniken optimiert, damit die relevantesten Treffer an erster Stelle stehen.

Diese Punkte werden wir im Folgenden näher erläutern.

Ontologie-basierte Integration verteilter Quellen

Das erste Problem ist, dass der Anwender in den meisten Fällen gar nicht weiß, in welchen Quellen er die relevanten Informationen erhält, also, wo er suchen muss. Eine erste Anlaufstelle für die Suche ist der Fileserver, auf dem in einfachen Ordnerstrukturen Dateien gespeichert werden können. Da eine klassische Ordnerstruktur nur eine Dimension – die Unterordner bzw. isa-Relation im Ontologie-Jargon – und keine Mehrfachvererbung darstellen kann, ergeben sich fast zwangsläufig zahlreiche Unstimmigkeiten und Redundanzen. Als weitere Quelle steht in den meisten Firmen ein Intranet zur Verfügung, in dem meist redaktionell aufgearbeitete Inhalte angeboten werden. Dazu gesellen sich Dokumentenmanagement-Systeme, Datenbanken, ERP-, CRM-Anwendungen sowie das Intra- und Internet als scheinbar unerschöpfliche Informationsquelle.

Bei der Netzwerkeinheit eines deutschen Telekommunikationsanbieters waren wesentliche Informationen zur Angebotserstellung in einem DMS-System, einem Intranet, in verschiedenen Datenbanken sowie auf dem Fileserver verteilt.

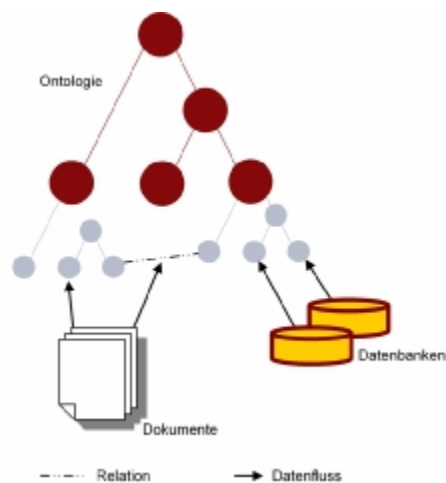


Fig. 1. Die Ontologie als Integrations-Mediator

Da eine Ontologie die Semantik einer bestimmten Anwendungsdomäne repräsentiert, ist die Terminologie der Domäne durch die Konzepte der Ontologie gegeben. Diese Terminologie beinhaltet auch Synonyme und andere lexikalische Einträge. Konzepte sind durch Beziehungen (Relationen) verbunden und durch Attribute detailliert beschrieben. Sie stehen in einer Unter-/Oberkonzeptbeziehung (isa-Relation). Komplexe Beziehungen zwischen Konzepten, deren Relationen und Attributen sind durch Regeln darstellbar. Ontologien sind somit gut zur Navigation geeignet. Der Nutzer kann über Unterkonzepte verfeinern, über Oberkonzepte generalisieren, über verlinkte Konzepte zu verwandten Konzepten gelangen und ein detailliertes Verständnis von Konzepten über die Attribute erlangen. Auf der anderen Seite stellt eine Ontologie eine integrative Sicht auf unterschiedliche heterogene Datenquellen unterschiedlichen Typs dar (vgl. Fig.1). Strukturierte Daten aus Datenquellen wie Relationale Datenbanken, Excel oder dergleichen Anwendungen bilden die konkreten Inhalte von Konzepten, deren Attribute und Relationen. Über

statistische Methoden, linguistische Verfahren und Textanalyse können selbst unstrukturierte Informationen, wie reine Texte, die konkreten Inhalte für eine Ontologie liefern. Dabei stellen die Metadaten wie z.B. Autor oder Kategorie solcher textlicher Information wiederum strukturierte Information dar.

Unterstützung des Anwenders durch die Ontologie

Ein zentrales Problem des *Information Retrieval* (IR) ist der Entwurf einer effizienten Interaktion. Das IR beschäftigt sich traditionell hauptsächlich mit dem Prozess der Repräsentation von Texten und Anfragen und des Vergleichs dieser Repräsentationen. Obwohl dieses Problem sicherlich noch immer von großer Bedeutung ist, wird der Interaktivität des Suchprozesses heute eine größere Bedeutung zugeschrieben. Darüber hinaus wurde bereits 1980 das *Anomalous State of Knowledge* (ASK) Problem beschrieben, das besagt, dass ein Benutzer mit einem Informationsproblem dieses nicht genau artikulieren kann, da ihm die genauen Kenntnisse zum Problem fehlen. Die fehlende Terminologie verursacht dann auch das Unvermögen, sein Bedürfnis einem Informationsvermittlungssystem mitteilen zu können [Belkin 1980]. Der Nutzer hat beim so genannten *Informationsbedürfnis* das Problem, die die Suche beeinflussenden Faktoren der affektiven, kognitiven und situationsbedingten Dimension in einer Suchanfrage auszudrücken [Taylor 1968]. Damit sind die meisten Anwender schlicht überfordert. Der Vorteil des File-Servers – die einfache Bedienbarkeit – ist auf der anderen Seite auch sein Problem. So ist z.B. die rein hierarchische Zuordnung zu den Ordnern oftmals nicht eindeutig. Soll man ein Dokument zeitlich (Erstellungsjahr), inhaltlich (Angebot, Projektbericht, ...) oder thematisch (Branche, Kunde, Technologie) einordnen? Mischformen in der Struktur sind nur sehr schwer und mit erheblichem Aufwand umsetzbar.

Volltextsuchen greifen in den meisten Fällen im IR zu kurz, da der Anwender in keiner Weise dabei unterstützt wird, wie er suchen kann, welche Begriffe er am besten verwendet, und er hat keine Möglichkeit der Navigation. Anwender geben im Schnitt 1,4 Begriffe pro Suchanfrage ein. Am Beispiel durchgeführt sehen wir, warum:

»Wo finde ich eine Angebotskalkulation über den Aufbau eines Funknetzes bei einem Pharma-Unternehmen?«

Wird die Suche nach den Stichwörtern „Angebot“ oder „Kalkulation“ mit hoher Wahrscheinlichkeit beispielsweise über eine Klassifikation zu Ergebnissen führen, so wird die Suche nach dem Begriff „Pharma-Unternehmen“ mit der gleichen Wahrscheinlichkeit zu keinem Ergebnis führen.

Die semantische Suche nutzt das Know-how des Unternehmens zur Erweiterung der Suchanfrage beispielsweise um die Begriffe „Bayer, Merck etc.“. Zudem wird die Suche nach dem Begriff „Funknetz“ eine Reihe von Ergebnissen liefern, jedoch keine zu interessanten Projekten, in denen von WAN, WLAN, W-LAN oder gar von „Wireless LAN“ gesprochen wird. Diese Zusammenhänge werden durch Ontologien automatisch berücksichtigt, sprich, es werden die richtigen Informationen gefunden, ohne dass der Anwender genau wissen muss, wonach er sucht. Das Ergebnis: Die Mitarbeiter erhalten eine bessere Entscheidungsunterstützung und können deutlich effizienter arbeiten.

Die Inhalte der Ontologie kommen dabei aus der Integration von CRM-Systemen, Branchenontologien und einer individuellen Domänenmodellierung. Die Zusammenhänge über Synonyme oder Homonyme werden ebenso in der Ontologie hinterlegt und können bei Bedarf automatisch über vorhandene Lexika erweitert werden.

Ontologie-basiertes Knowledge Retrieval

Eine Suche in einer Ontologie nutzt den semantischen Kontext, den die Ontologie über die Konzepte, Relationen, Attribute und Regeln definiert um treffsichere Ergebnisse in den angeschlossenen Datenquellen zu erzielen. In der Applikation stellt sich der Prozess der semantischen Suche wie folgt dar (Fig.2):

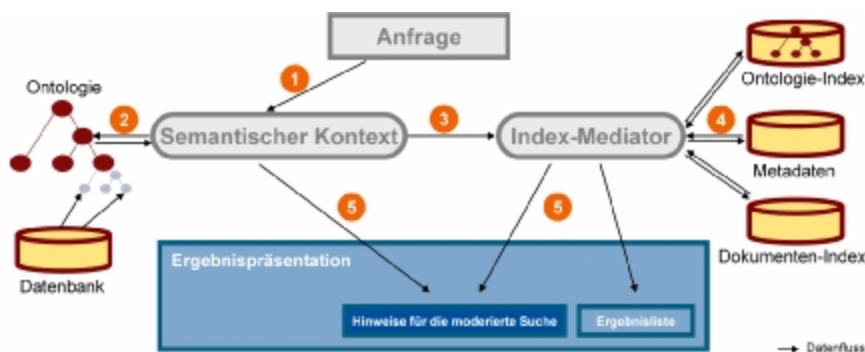


Fig. 2. Prozess der semantischen Suche zur Verwirklichung des Knowledge Retrieval

Nach der Anfrage des Benutzers (1) wird der semantische Kontext (2) ermittelt. Der Kontext ergibt sich dabei aus allen Treffern im Wissensmodell – der Ontologie – unter Berücksichtigung der externen Repräsentationen der Entitäten (z.B. Sprachen) und deren Synonyme. Bei der Kontextermittlung wird darüber hinaus eine automatische Tippfehlerkorrektur angewandt, die bei der Frage nach „Router“ die Suchanfrage zum Beispiel automatisch in „Router“ wandelt, oder dies durch Rückfrage „meinten Sie Router“ dem Benutzer überlässt (5).

Nach der Ermittlung des Kontextes sammelt der Index-Mediator unter Einbeziehung aller angeschlossenen Suchquellen, die Ergebnisse zur Kontext-erweiterten Anfrage ein. Dabei wird zwischen folgenden Kategorien unterschieden:

- **Ontologie-Index:** Die Suche in einem Ontologie-Index setzt eine semi-automatische Verschlagwortung (Annotation) der Dokumente voraus. Dieser Schritt wird von statistischen (zum Beispiel für die Ermittlung von relevanten Termen) und linguistischen Verfahren (zum Beispiel Named Entity Recognition, Informationsextraktion) unterstützt. Darüber hinaus können somit auch Terme mit den Dokumenten verknüpft werden, die nicht zwangsweise im Text vorkommen. Darüber hinaus wird für die Ontologie-Index Suche ein Ranking eingesetzt, das durch Regeln individuell an die Bedürfnisse des Kunden angepasst werden kann und deshalb die beste Qualität der Index-Suchen liefert.
- **Metadaten:** Die Suche in Metadaten ergibt durch die Integration von strukturierten Quellen wie Datenbanken mit der Ontologie, dem so genannten *Mapping*. Die Auswertung von Regeln liefert dabei implizites Wissen hinzu, Wissen also das aus bestehendem Wissen ableitbar ist, jedoch nicht explizit vorhanden war.
- **Dokumenten-Index:** Die Suche in Dokumenten verwendet den semantischen Kontext um die ursprüngliche Anfrage um Synonyme, Unter- und Oberkonzepte und definierten Relationen zu erweitern um die Qualität der Treffer zu optimieren [Moench et al. 2003].

Im nächsten Abschnitt werden diese Verfahren anhand eines Praxisbeispiels erläutert.

Praxisbeispiel SemanticMiner®

Das Telekommunikationsunternehmen nutzt einen hybriden Ansatz. Für die reine Dokumentensuche wird eine automatische Verschlagwortung (Volltextindex) in Kombination mit den Ergebnissen des Ontologie-Index – der annotierten Wissensbibliotheks-Dokumente – genutzt.

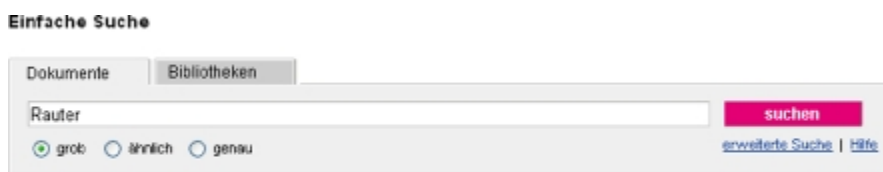


Fig. 3. Tippfehlerkorrektur im SemanticMiner des Kunden

Nach dem Abschicken der Suche wird im Domänenmodell nach einem passenden Eintrag zur Anfrage gesucht (vgl. Fig. 3). Dabei werden sowohl der Name der Entität, als auch die Synonyme der Entitäten berücksichtigt. Zusätzlich findet noch eine Tippfehlerkorrektur nach einem skalierten

Edit-Distanz-Verfahren auf Jaro [Jaro 1989], [Jaro 1995] – allgemein beschrieben in [Cohen et al. 2003] und [Chapman 2005] – Basis statt. Das Verfahren skaliert aufgrund der Tatsache, dass der Abstandswert zweier Terme unabhängig von der Länge der Terme ist. Das heißt, das Distanzmaß liefert bei einem Tippfehler von z.B. einem falschen (wie in Fig. 3) oder ausgelassenen Buchstaben einen konstanten Wert unabhängig von der Länge des Terms. Fig. 4 veranschaulicht die Erweiterung der Suche:



Fig. 4. Dokumentensuche im SemanticMiner®

Die Erweiterung der Suche basiert dabei auf – unter der Voraussetzung, dass Treffer im Domänenmodell gefunden werden – der Ver-Oderung der eingegebenen Terme. Die Eingabe beispielsweise von

Term1 Term2 -Term3

führt zu

$$(Term1 \vee \text{Synonym}_{11} \vee \text{Synonym}_{12} \vee \dots) \wedge (Term2 \vee \text{Synonym}_{12} \vee \text{Synonym}_{13}) \wedge \neg (Term3 \vee \text{Synonym}_{31} \vee \text{Synonym}_{32} \vee \dots)$$

Für den Bereich der Wissensbibliotheken wird neben der kumulierten Suche in den Dokumenten des Ontologie-Index (die Anpassungsregel für die Suche ermittelt für die Treffer eines Ordners hier die Summe der Treffer seiner Dokumente und der Dokumente seiner Unterordner usw.) – nach einer Prüfung und weiterer Annotation durch einen Redakteur – auch in den Metadaten der Bibliotheks-Ordner, die in einem CMS-System hinterlegt sind, gesucht. Beispiele solcher Metadaten sind Vertriebskanal, Region des Projekts, etc. (vgl. Fig. 5).

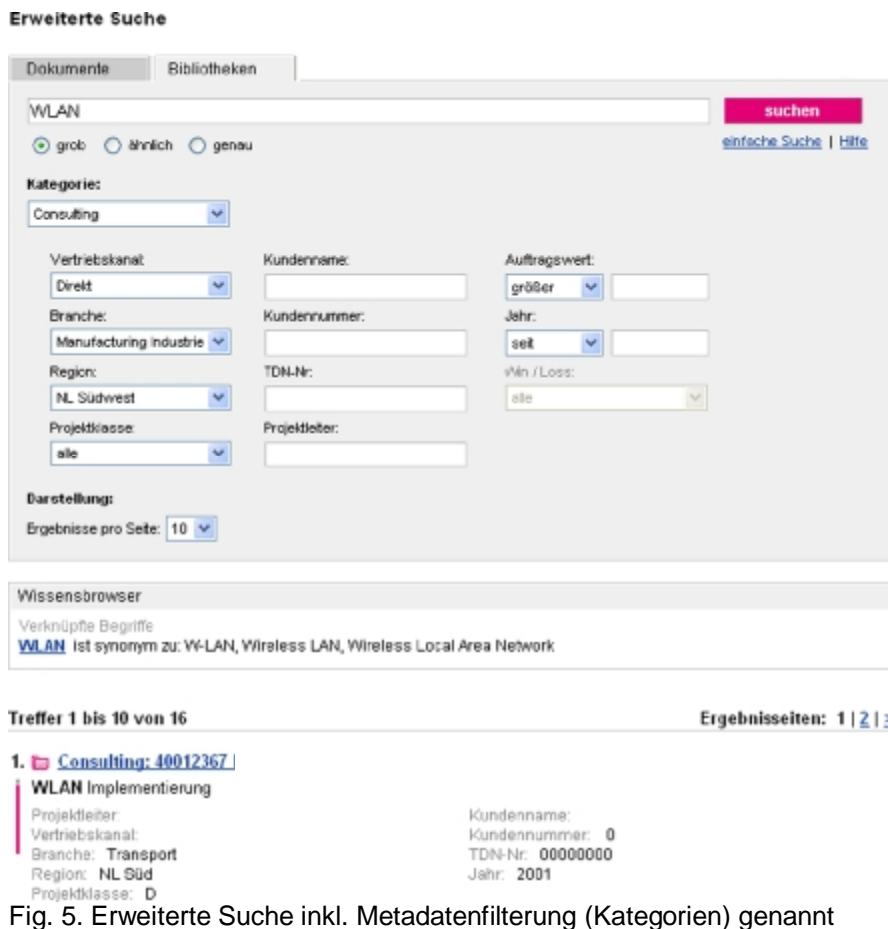


Fig. 5. Erweiterte Suche inkl. Metadatenfilterung (Kategorien) genannt

Schon durch sehr einfache Regeln, die Regional-Informationen ausnutzen, wie die Zuordnung von Städten oder Regionen zu Vertriebsregionen (in Fig. 5 beispielsweise „NL Südwest“), kann dadurch ein Mehrwert erzeugt werden.

Fazit

Semantische Technologien unterstützen den Wissensarbeiter im Unternehmen bei zentralen Arbeitsprozessen. Die Suche nach Informationen kann durch den Einsatz von Ontologien deutlich verbessert werden. Dabei ist bemerkenswert, dass der aktuelle Stand der Umsetzung von semantischen Technologien in Industrieprojekten beachtlich weit fortgeschritten ist. Die aktuellen Trends liegen insbesondere im Bereich der Verknüpfung von IT Architektur-Themen. Hier werden Service-orientierte Architekturen (SOA), Web-Services und Grid-Technologien in Verbindung mit SemanticWeb Themen gebracht. In zahlreichen Forschungsprojekten (SIMDAT, XMEDIA, NEON) aber auch bereits in Industrieprojekten werden hier interessante Ansatzpunkte bearbeitet. Ebenso arbeitet das W3C an Standards, die eine umfassende Nutzung der Technologien ermöglicht (RDF, OWL, Rules).

Die beschriebenen best practices verdeutlichen, welche Möglichkeiten heute bereits genutzt werden können - die aktuellen Trends zeigen klar die zukünftige beschleunigte Verbreitung der semantischen Technologien im industriellen Umfeld.

Referenzen

- [Belkin 1980] N.J. Belkin. *Anomalous states of knowledge as a basis for information retrieval*. The Canadian Journal of Information Science, 5:133--143, 1980.
 [Chapman 2005] Sam Chapman. SimMetrics: An open source extensible library of Similarity and Distance Metrics. <http://www.dcs.shef.ac.uk/~sam/simmetrics.html>. 2005.

- [Cohen et al. 2003] W.W. Cohen, P. Ravikumar, and S. E. Fienberg. *A comparison of string distance metrics for name-matching tasks*. In Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web, 2003.
- [Eshelman and McDermott 1986] Eshelman, L; McDermott, J.: MOLE: A Knowledge Acquisition Tool That Uses Its Head. In Boose, J.H.; Gaines, B. (Eds.): AAAI-Workshop: Knowledge-Acquisition for Knowledge Based Systems, Banff, Kanada, 1986.
- [Eshelman 1988] Eshelman, L: Mole: A Knowledge-Acquisition Tool for Cover-and-Differentiate Systems. In [Marcus 1988], 1988, p. 37-80.
- [Marcus 88] S.Marcus (Ed.): *Automating Knowledge Acquisition for Experts Systems*. Kluwer Academic Publisher, Boston, 1988.
- [Moench et al. 2003] Eddie Moench, Mike Ullrich, Hans-Peter Schnurr, Juergen Angele: *SemanticMiner - Ontology-Based Knowledge Retrieval*. In: Special Issue of selected papers of the WM2003 in the Journal of Universal Computer Science (J.UCS).
- [Jaro 1989] M. A. Jaro. *Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida*. Journal of the American Statistical Association 84:414–420, 1989.
- [Jaro 1995] M.A. Jaro. *Probabilistic linkage of large public health data files (disc: P687-689)*. Statistics in Medicine 14:491–498, 1995.
- [Taylor 1968] R.S. Taylor. *Question-negotiation and information seeking in libraries*. College and Research Libraries, 29, 178-194, 1968.