

Automatische Generierung von Unternehmensontologien

Schaffung einer Unternehmenssemantik

Dr. Hans-Josef Hesse, Dr. Mohammad Abuosba,
Jürgen Lütkehellweg, Empolis GmbH¹
Markus Ott Festo AG²

***Abstract.** In dem vorliegenden Beitrag wird ein Vorgehen beschrieben, wie das in einem Unternehmen über verschiedene Systeme (ERP, CRM, SCM, PIM, PLM, ...) verteilte Wissen in Form von strukturierten Informationen in eine erste Ontologie übernommen werden kann, um mit der erstellten Ontologie die unstrukturierten Informationen (E-Mails, Protokolle, PDF-, Office-Dokumente, ...) zu erschließen bzw. semantisch zu annotieren. Durch Nutzung des vorhandenen strukturierten Wissens bei der teilweise automatisierten Ersterstellung einer Unternehmensontologie ist der zu betreibende Aufwand als eher gering zu betrachten. Hierdurch ergibt sich ein niederschwelliger Ansatz, ein Unternehmen in das semantische Zeitalter einzuführen.*

1. Ausgangssituation

Durch die zahlreichen Möglichkeiten des Web 2.0 werden täglich Unmengen von Informationen in Unternehmen aus internen und externen Quellen generiert, die so zahlreich, so unterschiedlich und so verteilt sind, dass wir nicht immer imstande sind das Richtige zu finden oder Relevantes von Irrelevantem zu unterscheiden. Dadurch ertrinken wir in Informationen und dürsten nach Wissen. Mit den Anstrengungen der letzten Jahre in Richtung semantisches Web sind durchaus neue Möglichkeiten geschaffen worden, mit deren Hilfe ein intelligentes Navigieren durch den Informations-Dschungel machbar wird. Inhalte wie Angebote, Patente, Produktbeschreibungen, Servicedokumentation, Geschäftsregeln und selbst Gerichtsurteile

¹ Empolis GmbH, Meisenstrasse 90, 33607 Bielefeld; E-Mail: hans-josef.hesse, mohammad.abuosba, juergen.luetkehellweg@empolis.com

² Festo AG, Ruitterstrasse 82, 73734 Esslingen, E-Mail: mott@de.festo.com

können so für die Lösung eines aktuellen Problems genutzt oder zur Gestaltung einer neuen Fertigungslösung eingesetzt werden. [Abu08].

Ontologien bilden in diesem Zusammenhang eine tragende Rolle. Durch Ontologien ist es möglich, domänenspezifische Fachlichkeiten abzubilden und Sinnzusammenhänge zwischen den definierten Konzepten herzustellen. Physische Produkte, die ontologisch beschrieben werden, werden hierdurch zu intelligenten Objekten. Sie haben nicht nur statische Eigenschaften sondern auch Fähigkeiten und Beziehungen zu anderen Produkten oder Komponenten. Mit diesen Ontologien ist das Navigieren, Suchen und Auffinden von Lösungswissen oder ähnlichen Konzepten wesentlich effektiver und effizienter gestaltbar, wodurch das Dickicht des Informations-Dschungel wesentlich heller und nutzbarer wird.

Die Erstellung von Ontologien selbst stellt sich jedoch heute als weitestgehend manueller Prozess dar, der teuer, langwierig und pflegeintensiv ist. Dadurch ist die Hürde für viele Unternehmen zu hoch, um diese Technologie einzusetzen und das unternehmenseigene Wissen ontologisch abzubilden.

Ziel muss es sein, das im Unternehmen bereits vorhandene strukturierte Wissen in Form von Produkttaxonomien, Normen und sonstigen Regeln zu nutzen, um automatisiert eine Ontologie aufzubauen. Dies verringert massiv die Einstiegschwelle für die Unternehmen, semantische Verfahren einzuführen und im Betrieb einzusetzen.

In dem vorliegenden Beitrag wird dieser Ansatz [Hep05] beschrieben und anhand einer Untersuchung belegt, wie Teile einer Ontologie automatisch generiert werden und im Unternehmen genutzt werden können.

2. Strukturierte Informationen in Unternehmen

In einem Unternehmen liegen diverse Informationen in strukturierter Form vor. Hierzu zählen insbesondere

- Produktinformationen, als Produkthierarchie, in PLM oder PIM-Systemen
- Kundeninformationen, in CRM oder ERP-Systemen. Strukturiert nach Branchen u.v.m.
- Unternehmensinformationen, wie Personenverzeichnisse, Berechtigungen und Fähigkeiten

Diese fachlichen Dimensionen stellen erfahrungsgemäß die „building blocks“ einer zu bildenden Unternehmensontologie dar.

Die Nutzung dieser strukturierten Informationen zur Bildung einer Ontologie soll beispielhaft im thematischen Kontext der Produkte durch den Import eines Klassifikationssystems beschrieben werden. Dieses Vorgehen kann sinngemäß dann auch auf andere strukturierte Informationen angewandt werden.

Klassifikationssysteme lassen sich heute bereits als einheitlichen „Kommunikationsweg“ über Organisationsgrenzen hinweg begreifen, da bereits in der Vergangenheit die dementsprechenden Prozess- und branchenübergreifenden Konsortien versucht haben, hierdurch die Kommunikation untereinander zu standardisieren.

Klassifikationssysteme/Informationsmodelle (eCI@ss, ETIM, proficl@ss, UNSPSC, ECCMA, ...) mit den entsprechenden Merkmalleisten, Hierarchien, Synonymen u.v.m. liefern in der Regel eine für viele Zwecke ausreichende Beschreibung von Produkten und Dienstleistungen. Dennoch stoßen auch diese Modelle an Grenzen, die in einer zu bildenden Ontologie zu berücksichtigen oder auch zu überwinden sind.

Eine nähere Betrachtung verschiedener Klassifikationssysteme (eCI@ss, Proficl@ss, ETIM, UNSPSC, ...) zeigt, dass sich diese im Wesentlichen durch folgende fünf Eigenschaften auszeichnen:

- **Produkt Klassen**
Die Klassen dienen dazu, Produkte zu kategorisieren. Dabei können als Maßstab zur Kategorisierung physikalische, chemische, elektrische oder auch andere Eigenschaften (Gebrauch, Einkauf, Branche, ...) benutzt werden. Da ein und dasselbe Produkt in verschiedenen Kontexten (Branche, Gebrauch, ...) benutzt werden kann, existiert aufgrund eines N:M Mappings hier ein hierarchisches Komplexitätsproblem, welches in der Ontologie aufgelöst werden kann.
- **Klassenhierarchie**
Die hierarchische Anordnung der Klassen folgt natürlich dem ursprünglichen Anwendungsfall des Klassifikationssystems. So wurde beispielsweise eCI@ss aus Sicht einer Einkaufs- oder Verkaufsorganisation entwickelt. Daher wurden Produkte als ähnlich angesehen, die von Herstellern einer Branche bezogen oder zusammen

(als Paket) verkauft werden könnten. Eine Ontologie kann hier weitere Sichten generieren.

- **Glossar/Wörterbuch**
Umfassendere Klassifikationssysteme besitzen für Klassen, Hierarchien und Merkmale Synonymlisten. Damit lässt sich eine umfassendere Suche wie auch Beschreibung der Produkte realisieren.
- **Merkmalsausprägung**
Die zulässigen Werte und Datentypen zur Beschreibung der Merkmale werden durch die Klassifikation definiert. Manche Klassifikationssysteme wie eCI@ss stellen zur eindeutigen semantischen Beschreibung zusätzliche Datencontainer zur Verfügung, in die freisprachlicher Text eingetragen werden kann.
- **Relation Klasse: Merkmale**
Klassifikationssysteme wie bspw. eCI@ss oder ETIM definieren eine exakte Zuordnung von Merkmalen (Attributen) zu den jeweiligen Klassen. In Abhängigkeit des Merkmales, lassen sich hier auch Relationen/Konzepte in der Ontologie ableiten.

3. Transformationsvorschrift strukturierter Informationen in Ontologien am Beispiel eines Klassifikationssystems

Die Herausforderung bei der Transformation einer Klassifikation in eine OWL-Ontologie ist die Berücksichtigung der in der Hierarchie enthaltenen semantischen Informationen. Ziel sollte es sein, möglichst vollständig die mit der Hierarchie verbundene Intention (intrinsische Semantik) zu erhalten. Da die Entwicklung der eCI@ss Klassifikation maßgeblich durch die praktischen Aspekte des eProcurements vorangetrieben wurde, sind Artikel, die sich „irgendwie“ in eine gleiche Kategorie eingruppierten ließen, untergeordnet. So wurde beispielsweise in der eCI@ss die Dockingstation eines Notebooks unter Computer eingeordnet.

Informations-, Kommunikations-, und Medientechnik

Computersystem

Notebook

Tablet PC

Notebook (nicht klassifiziert)

Personal Digital Assistant (PDA)

PDA (Stift basierend)

PDA (Tastatur basierend)

Personal Digital Assistant (PDA, nicht klassifiziert)

...

Computersystem (Teile)

Computersystem (Teile, nicht klassifiziert)

Computersystem (Zubehör)

Docking-Station (Computersystem)

Computersystem (Zubehör, nicht klassifiziert)

Computersystem (Montage)

Computersystem (Montage, nicht klassifiziert)

Eine einfache Übertragung dieser Taxonomie in eine korrespondierende Klassenstruktur der Ontologie und der gleichzeitigen Übertragung von taxonomischer Struktur mit dem Konzept `<rdfs:isSubClassOf>` wäre semantisch unsinnig; eine Dockingstation ist nämlich kein Computer sondern ein Zubehör. Dennoch ist dieses Konzept taxonomisch betrachtet hier einsortiert.

Die Herausforderung muss es sein, bei den Transformationen eines Klassifikationssystems in eine OWL-Ontologie, Konzepte wie Vererbung und Transitivität zu erhalten.

Dazu eignet sich das nachfolgend beschriebene Verfahren, welches im Rahmen des UseCase Processus³ evaluiert wird. Für jeden Knoten einer Taxonomie (Hierarchische Struktur im Klassifikationssystem) werden zwei Konzepte definiert.

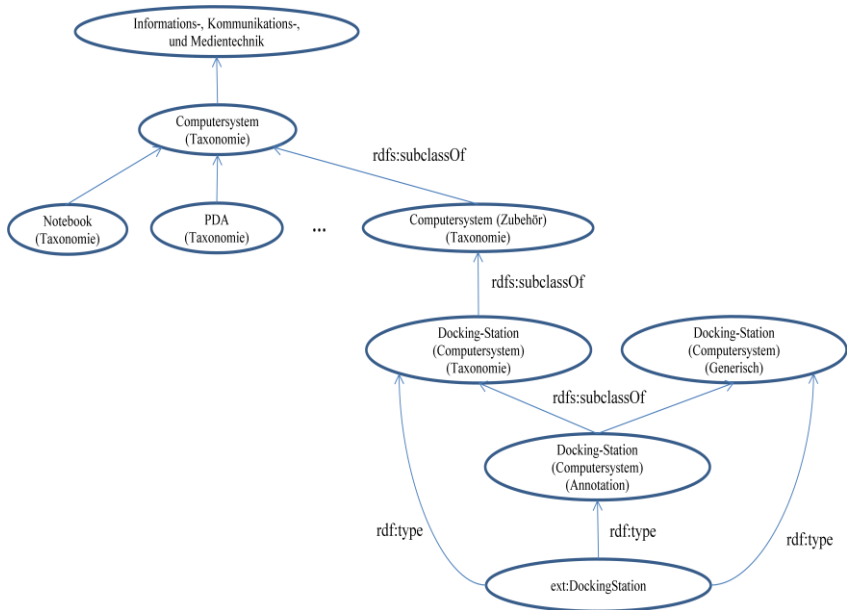
- Es wird ein Konzept definiert, welches mit der generischen Art des Knoten korrespondiert und
- das zweite Konzept reflektiert die taxonomische Struktur des Klassifikationssystems.

³ PROCESSUS ist ein Anwendungsszenario im THESEUS-Forschungsprogramm, das durch das Bundesministerium für Wirtschaft und Technologie gefördert wird. Ziel von PROCESSUS ist eine IT-gestützte Unternehmenssteuerung, die den Vergleich von Produkten, Lösungen und Geschäftspartnern sowie das Aufspüren von Problemwissen für wissensintensive Tätigkeiten ermöglicht. PROCESSUS macht die Semantik von Information in Geschäftsprozessen nutzbar – über Organisationen hinweg und entlang der Wertschöpfungsketten.

Die Elemente des taxonomischen Konzepts werden analog zur Hierarchie des Klassifikationssystems mittels der Relation `<rdfs:subClassOf>` angeordnet. Das die generische Art des Knotens reflektierende Konzept wird nicht hierarchisch einsortiert.

Anschließend werden diese beiden Konzepte durch eine weitere Klasse in Beziehung gesetzt:

- Es wird eine Annotationsklasse pro Taxonomieknoten des Klassifikationssystems gebildet, der gemäß der Relation `<rdfs:subClassOf>` vom generischen Konzept sowie vom taxonomischen Konzept erbt. Somit reicht ein einfaches `<rdfs:type>` Statement aus, um aus einem Produkt eine Instanz eines generischen wie auch taxonomischen Konzeptes zu machen.



Über die Taxonomie hinaus beschreibt ein Klassifikationssystem immer auch ein Set an Merkmalen, gruppiert nach Klasse, um festzulegen, wie ein Produkt zu beschreiben ist. Um diese Merkmalsausprägungen der Klasse in die Ontologie zu übernehmen sind diese klassenspezifisch zum jeweiligen Konzept der Ontologie zu transformieren.

Hierzu sind folgende Schritte bei der Transformation zu beachten:

- Datentypen sind auf die Datentypen der Ontologie zu mappen.
Hier : Mapping auf XSD-Datatypes
- Zusammenhang von Klasse und Merkmal als recommendedProperty abbilden,
- Zusammenhang vom Merkmal und Merkmalsausprägung als recommendedValue abbilden.

Datentypmapping:

Es ist wichtig zu wissen, dass hier eine Mappingvorschrift zu entwickeln ist, da in Systemen wie bspw. eCI@ss Hunderte verschiedener Datentypen (basierend auf ISO9735 / ISO6039) definiert sind. Hier gilt es eine praktikable Mappingvorschrift zu definieren.

Merkmale und Merkmalsausprägungen einer Klasse in die Ontologie übernehmen:

Eine zweite Mappingvorschrift definiert die Transformation der zu einer Klasse des Klassifikationssystems gehörenden Merkmale in entsprechende Annotationsproperties der zugehörigen Klasse in der Ontologie. Es wird hier empfohlen die zwei Properties „recommendedProperty“ und „recommendedValue“ anzulegen, da die Beschreibung/Defintion der Eigenschaften einer Klasse in einer Ontologie nicht starr festgelegt werden darf (Open Word Assumption).

Mit dieser beschriebenen Vorgehensweise lassen sich strukturierte Informationen in allen Ausprägungen in eine Ontologie überführen. Angefangen bei rein hierarchischen Strukturen, über Vererbungsstrukturen, hin zu den im jeweiligen Knoten definierten Attributen und deren Wertelisten.

Durch eine solche Transformation lässt sich somit relativ einfach eine erste Unternehmensontologie erstellen.

4. Intelligente Suche, durch Nutzung der Unternehmensontologie

Nach der Erstellung der Ontologie steht eine erste Wissensbasis zur weiteren Erschließung von unstrukturierten Informationen bereit. Semantische und linguistische Verfahren nutzen das Wissen, um im Unternehmen verfügbare Repositories von Dokumenten und Dateien zu parsen und deren Inhalte zu annotieren. Hierdurch werden Informationen aus unstrukturierten Dokumenten strukturiert und in ihrem semantischen Kontext auffindbar.

Die Ontologie bietet somit die Möglichkeit, strukturierte Informationen diverser Dimensionen (vgl. 2 Strukturierte Informationen in Unternehmen) zu konnektieren und hieraus einen semantischen Suchindex zu bilden. Auf Basis der Implementierung der semantischen Technologien in service-orientierte Architekturen (SOA), kann darüberhinaus eine tiefe semantische Integration in alle Unternehmensprozesse erfolgen.

Die Ontologie in Verbindung mit semantischen Technologien bietet dem Unternehmen einen qualitativen Zugriff auf Informationen auf Basis von vorhandenem Wissen. Insbesondere die Kontextermittlung auf Basis der hinterlegten Konzepte und Fakten ergibt für den Recherchierenden einen ungeahnten Mehrwert gegenüber der heutigen Volltextsuche oder anderen unternehmensspezifischen Anwendungen. Ähnlichkeiten werden auf Basis von vorhandenem Wissen ermittelt, und ergeben einen mehrdimensionalen (semantischen) Blick auf die ausgewählte Entität. Als Beispiel kann die Eingabe „mobiles Büro“ dann Informationen zu Dockingstations, PDA, Notebooks u.v.m. ermitteln.

Eine sich immer weiter entwickelnde Unternehmensontologie bietet zudem die Möglichkeit, semantisch durch die Informationen eines Unternehmens zu navigieren. Dieser intuitive Zugang zu Unternehmenswissen kann den Benutzer auf Basis des Domainwissens (Relationen und Konzepten) führen: Der Benutzer muss nicht mehr wissen, wo ein Autor seine Information abgelegt hat, der Benutzer muss nur wissen, was er sucht. Eine weitergehende Identifikation des zu findenden Wissens ergibt sich dynamisch aus dem Kontext.

5. Fazit:

Unternehmen scheuen den Schritt in die semantische Welt, da der Aufbau von Unternehmensontologien als zu kosten- und zeitintensiv erachtet wird. Die vorliegende Darstellung offenbart eine Quick-Win –Alternative, die für die semantischen Technologien notwendige Ontologie in Time-und Budget relativ automatisiert zu erstellen. Der sich hieraus für das Unternehmen ergebende Mehrwert lässt sich somit in einem kurzen ROI-Horizont sehr gut abbilden.

Literatur

[Abu08] Abuosba, M., Gaag, A., Franz, J.: „Semantic Knowledge Management – Erschließung von Lösungswissen in der Automatisierungsin-
dustrie“. Tagungsband vom 10. Kongress zum IT-gestützten Wissensma-

nagement, Frankfurt Oktober 2008. Hrsg. M. Bentele, N. Gronau, R. Hochreiter, P. Schütt, M. Weber. ISBN 978-3-7723-3990-5, S. 351-358

[Hep05] Hepp, Martin: A Methodology for Deriving OWL Ontologies from Products and Services Categorization Standards, in: Proceedings of the European Conference on Information Systems (ECIS 2005), May 26-28, 2005, Regensburg, pp. 1-12.