

Bereitstellung unstrukturierter Daten und Verknüpfung mit strukturierten Daten in der Konzeptentwicklung von Automobilen

Andreas Konzag¹ und Frithjof Dau²

(1) BMW AG, Lehrstuhl Industrielle Informationstechnik der BTU Cottbus

(2) SAP AG

Zusammenfassung: Die effiziente Verfügbarkeit relevanter Daten und Informationen stellt für die Entwicklung von Fahrzeugkonzepten einen Erfolgsfaktor dar. Methoden und Werkzeuge des Datenmanagements können derzeit nicht zureichend die Anforderungen dieser Phase erfüllen. Die vorliegende Arbeit beschreibt Grundlagen und Methoden für ein Informationssystem, das dem Anwender Entwicklungsinformationen und Produktdaten in der Phase der Konzeptentwicklung zugänglich macht. Es werden semantische Technologien herangezogen, um in unstrukturiertem und implizitem Wissen Strukturen und Beziehungen zu erkennen, diese explizit zu modellieren und somit für den Benutzer navigierbar und zugreifbar zu machen.

1 Einleitung

Die Merkmale eines Fahrzeugs werden mit Beginn des Produktentwicklungsprozesses festgelegt. Damit am Ende der frühen Entwicklungsphasen ein in sich stimmiges sowie plausibles Gesamtkonzept für die Serienentwicklung als Eingangsgröße vorliegt, muss eine Vielzahl vernetzter Teilkonzepte entwickelt, bewertet und entschieden werden. Ein kritischer Erfolgsfaktor dafür ist die effiziente Verfügbarkeit von relevanten Produktdaten und Entwicklungsinformationen. Dies sind in den frühen Phasen weniger strukturierte Daten

wie Stücklisten als vielmehr unstrukturierte Daten, abgelegt in Informationsilos wie z. B. Office-Dokumenten in Fileshares.

Im Rahmen des Forschungsprojektes „Aletheia - Semantische Föderation umfassender Produktinformationen“ werden Grundlagen und Methoden für Informationssysteme, die dem Anwender produktbezogene Informationen aus den unterschiedlichsten Domänen, Organisationen und Communities zugänglich machen, untersucht. Hierbei wird nicht nur die explizit in den verwendeten Quellen enthaltene Information zurückgegeben, sondern es werden semantische Technologien eingesetzt, um auch Strukturen in und Beziehungen zwischen Informationen zu modellieren, um darüber hinaus implizites Wissen abzuleiten und schließlich diese Informationen für den Benutzer navigierbar und zugreifbar zu machen.

Aletheia ist eine vom BMBF geförderte ITK 2010 Leitinnovation. Am Aletheia-Forschungsprojekt sind zwölf akademische oder industrielle Partner beteiligt. Die Industriepartner kommen aus diversen Geschäftsbereichen- aus unterschiedlichen Branchen, mit einem Fokus auf unterschiedliche Zeitpunkte im Produkt-Lebenszyklus, für unterschiedliche Gruppen von Benutzern. Die Forschungsergebnisse werden in Aletheia in fünf industriellen Anwendungsszenarien umgesetzt und evaluiert. Die in diesem Artikel beschriebene Fallstudie ist eines dieser Szenarien. Es fokussiert auf die Phase der Konzeptentwicklung des Produktentwicklungsprozesses im Automobilbau. Im Folgenden werden die in Aletheia erarbeiteten Lösungen unter besonderer Berücksichtigung dieses Anwendungs-Szenarios beschrieben.

2 Entwicklung von Fahrzeugkonzepten

Der Produktentwicklungsprozess lässt sich grundlegend in die Phasen Konzept- und Serienentwicklung unterteilen. Grundlegend ist die Fahrzeugentwicklung von einer hohen Komplexität, einer starken Vernetzung der Entwicklungsthemen und vielen Beteiligten, die in einer Matrixorganisation arbeiten, gekennzeichnet. In der Konzeptphase werden den Kundenwünschen entsprechend die technischen Lösungen erarbeitet, die in der folgenden Phase zur Serienreife gebracht werden. Sie ist durch dynamische Prozesse und wenige Formalismen geprägt. Darüber hinaus stellt eine große Menge an unstrukturierten Daten für die Konzeptphase den wesentlichen Teil der Daten und Informationen dar.

Der Use Case im Forschungsprojekt Aletheia ist die Grundlage dafür, die relevanten Anforderungen an die Technologie- und Methodenentwicklung herauszuarbeiten. Hierfür wurden die zentralen Elemente des in Abbildung 1

dargestellten Geschäftsprozesses „Fahrzeugkonzeptentwicklung“ mit seinen Daten, Akteuren und Interaktionsmustern untersucht.

Dieser Prozess beschreibt auf Fahrzeugprojekt-Ebene die Ableitung funktionaler Anforderungen von kundenrelevanten Produktzielen durch den Eigenschaftsverantwortlichen und die Bereitstellung entsprechender konstruktiver Lösungen durch den Funktionsverantwortlichen. In der Verantwortung des Konzeptingenieurs liegt die Stimmigkeit und Plausibilität über alle konzeptrelevanten Entwicklungsthemen im Fahrzeugprojekt.

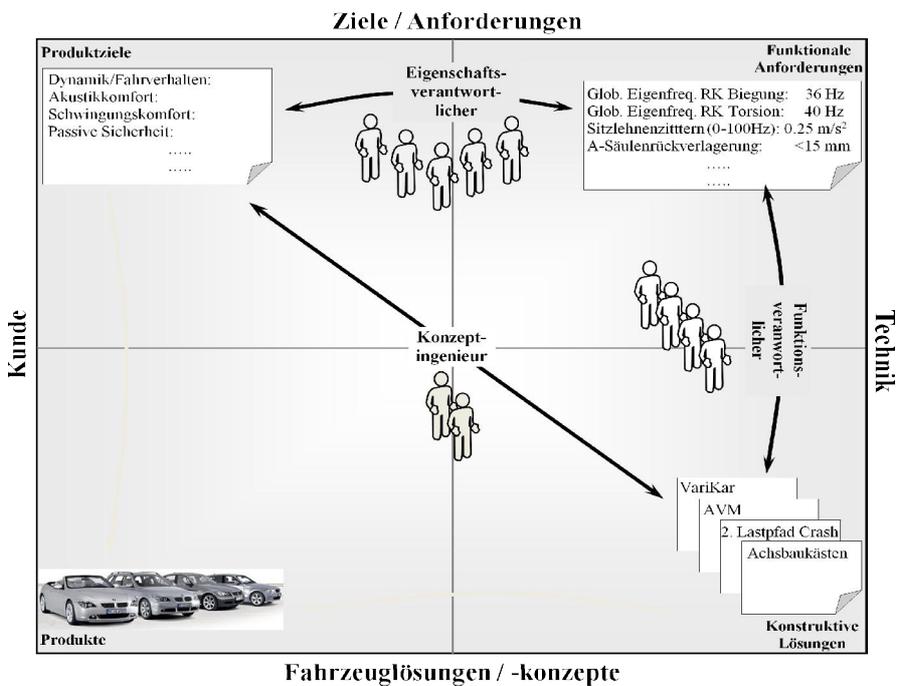


Abbildung 1: Anwendungsszenario in Aletheia – Geschäftsprozess „Fahrzeugkonzeptentwicklung“

3 Herausforderungen der Datenbereitstellung

Wie Abbildung 2 veranschaulicht, ist ein überwiegender Teil der für die Konzeptphase im Automobilbau relevanten Daten unstrukturiert. Strukturierte Daten, die das Fahrzeug beschreiben, entstehen erst im Laufe des Entwicklungsprozesses. Konventionelle Methoden und Werkzeuge des Datenmanagements, wie sie produktiv in der Serienentwicklung von Automobilen eingesetzt werden, sind aufgrund der Charakteristika der Fahrzeug-Konzeptentwicklung nur bedingt nutzbringend einsetzbar.

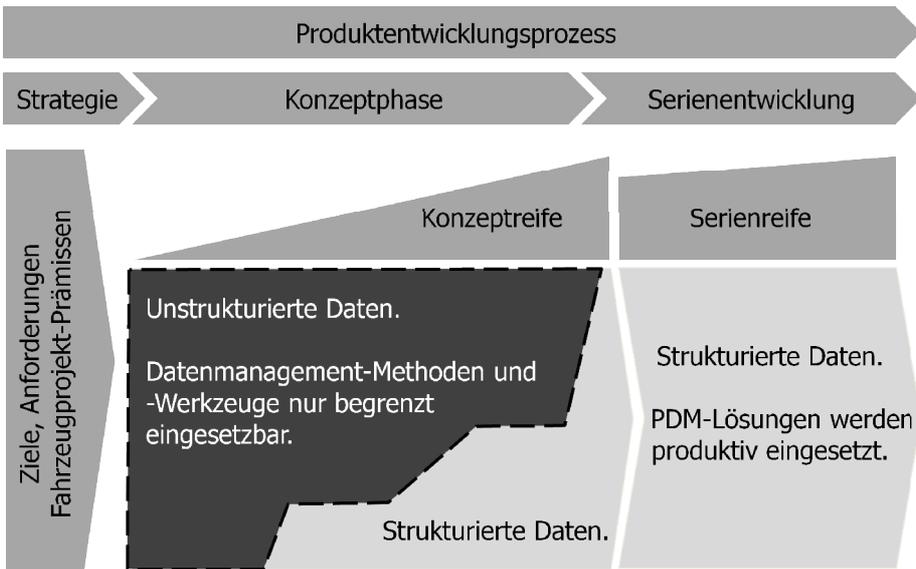


Abbildung 2: Handlungsbedarf zum Management unstrukturierter Daten in der Konzeptphase der Automobilentwicklung

Die für das Anwendungsszenario in Aletheia zu bewältigenden Herausforderungen lassen sich grob in zwei Klassen einteilen:

Föderation verschiedener Informationsquellen: Die in der frühen Produktentwicklung benötigten Daten und Informationen sind in verschiedenen Informationssilos abgelegt. Dabei sind nicht nur strukturierte Daten zu betrachten: Gerade in der Konzeptentwicklung fallen vor allem unstrukturierte Daten an, beispielsweise in Office Dokumenten. Weiterhin sind diese Daten nicht ihres Inhaltes/Kontextes entsprechend zugänglich, und der

Zusammenhang mit strukturierten Daten ist gar nicht oder nur unzureichend erfasst.

Anwenderzugriff auf die föderierten Informationen: Aus Anwendersicht ist der physikalische Speicherort von Daten und Informationen eher zweitrangig: Wichtiger sind die Informationen an sich. Zwischen den einzelnen Entwicklungsthemen, und folglich auch zwischen den entsprechenden Fakten und Dokumenten, bestehen diverse und nicht per se allen Beteiligten bekannte Abhängigkeiten. Es geht also nicht nur um das Finden einzelner Dokumente oder Fakten, vielmehr sind gerade auch die Abhängigkeiten zwischen verschiedenen Informationsfragmenten, die aus unstrukturierten sowie aus strukturierten Quellen kommen können, für den Anwender nutzbar und sichtbar zu machen. Konventionelle Systeme fokussieren sich allerdings entweder nur auf die Suche nach strukturierten Daten (hier sind beispielhaft Enterprise Search Systeme zu nennen) oder auf die Suche nach unstrukturierten Daten (zum Beispiel durch eine Desktop Suche).

Folglich müssen zur Lösung dieser Probleme als erstes auf eher technischer Ebene Mechanismen bereitgestellt werden, die einen einheitlichen Zugriff auf Daten und Dokumente aus verschiedenen physikalischen Quellen ermöglichen. Als zweites müssen aus Anwendersicht Methoden und passende Oberflächen entwickelt werden, die eine integrierte Sicht auf die föderierten Informationen und die dazwischen bestehenden Abhängigkeiten erlauben.

4 Technischer Lösungsansatz

Die grundlegende Struktur von Aletheia ist in Abbildung 3 dargestellt. Die zentrale Schicht in Aletheia ist ein semantisches Repository, in dem teilweise Fakten aus semistrukturierten Quellen sowie Informationen zu Dokumenten persistiert werden. Weiterhin erlaubt es den dynamischen (d.h. zum Zeitpunkt, in dem der Anwender Informationen abfragt) Zugriff auf strukturierte Quellen.

Wesentlich dabei ist die Modellierung der Anwendungsdomäne durch eine Ontologie [1]. Die für das Anwendungsszenario modellierte Ontologie umfasst zurzeit 72 Konzepte sowie 33 verschiedene Arten von Relationen (aka Beziehungen) zwischen den Konzepten.

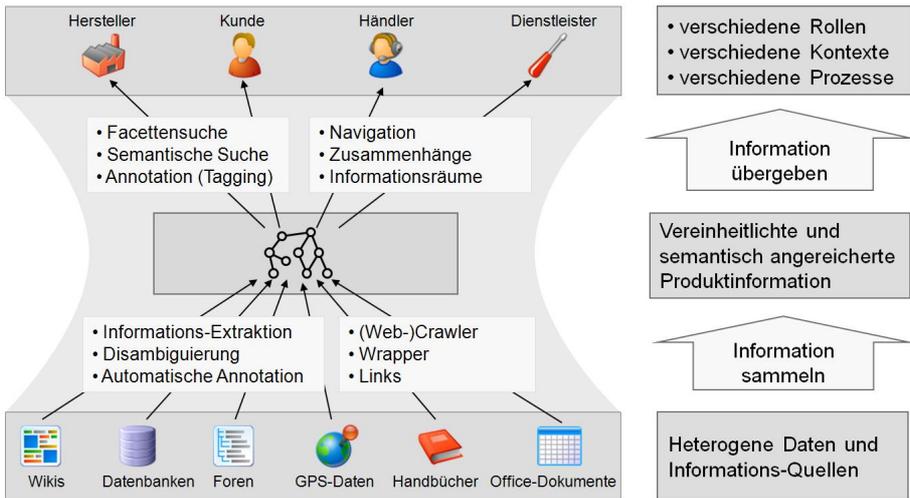


Abbildung 3: Aletheia Struktur

Das in Abbildung 3 durch ein Netz dargestellte Repository in Aletheia teilt sich auf in ein sogenanntes *semantisches Repository* und ein *Uncertain Repository*. Das semantische Repository dient dabei der Persistierung von Daten und Fakten aus strukturierten und semistrukturierten Quellen (beispielsweise Datenbanken und XML-Dateien). Das Uncertain Repository dient der Föderation von Dokumenten sowie deren (automatisch erzeugte) Annotation von Dokumenten mit Klassen der Ontologie. Diese Annotationen sind aufgrund der eingesetzten Methoden mit Unsicherheit behaftet, woraus sich der Namen des Repositories erklärt. Auf beide Repositories kann über einheitliche Web Services zugegriffen werden. Da sich insbesondere das Aletheia Frontend über diese Services auf die Informationen aus den Repositories zugreift, ist die Aufteilung in zwei Repository-Arten für den Anwender vollkommen transparent.

Semantisches Repository: Als semantisches Repository wird der OntoBroker F-Logic in der Version 5.3 eingesetzt, der in den allgemeinen OntoBroker-Server integriert ist. Das wichtigste Feature von OntoBroker F-Logic ist das *Reasoning* und das *Query-Answering* über F-Logic-Wissensbasen. Ein wichtiges Feature von OntoBroker ist die Möglichkeit, externe Datenquellen anzubinden und zur Query-Beantwortung zur Laufzeit zu berücksichtigen. Insbesondere die Fähigkeit, relationale Datenbanken anzubinden ist für Aletheia relevant, darüber hinaus können u.a. auch ERP-Systeme oder Webservices angebunden werden.

Uncertain Repository: Für dieses Repository wird Sesame eingesetzt. Sesame ist ein Open Source Framework zum Speichern, Schlussfolgern und Anfragen von RDF-Daten. Sesame verwendet RDF Schema. Die wesentliche Motivation dafür ist der Einsatz der Crawler Engine, die das aus dem Nepomuk-Projekt stammenden Aperture Framework verwendet und die einen generischen Zugriff auf verschiedenste Arten von Zugriffsprotokollen sowie unterschiedlichste Arten von Dateiformaten gestattet. Dokumentenorientierte Datenquellen, welche semistrukturierte und unstrukturierte Daten enthalten, werden über die Crawler-Engine-Komponente eingelesen, in eine einheitliche, normalisierte RDF-Repräsentation überführt und in einem Triple Store persistiert. Sie sorgt weiterhin für einen regelmäßigen Abgleich der Rohdaten mit dem persistierten Zustand.

5 Anwenderzugriff auf die föderierten Informationen

Schon Marcia Bates hat in ihrer grundlegenden Arbeit [1] anhand von Literaturrecherchen sechs verschiedene Informationszugriffarten herausgearbeitet, anhand derer Anwender sich die relevante Literatur zu einem bestimmten Thema erschließen. Beispielsweise sind für einen bereits als relevant erkannte Arbeit auch die Arbeiten, die die gegebene Arbeit zitieren oder, sich den Informationsraum in die andere Richtung erschließend, von der gegebenen Arbeit zitiert werden. Bedeutend ist hier die *Navigation* durch den Informationsraum, weniger das gezielte Finden einzelner Dokumente. Auch bei dem Gebrauch einer Suchmaschine kann in eingeschränktem Maße durch sukzessives Ändern der Suchstichworte der Informationsraum erschlossen werden. In der Tat belegen empirische Untersuchungen, dass eine reine Suche nach Dokumenten für den Anwender nicht ausreichend ist [8]. Aus diesen Gründen müssen für den Anwender folgende Punkte berücksichtigt werden:

- Aletheia muss in einer Umgebung das Finden sowohl von Dokumenten als sowie Fakten aus strukturierten Quellen ermöglichen.
- Es sind insbesondere auch die Abhängigkeiten zwischen verschiedenen Fakten und/oder Dokumenten nutzbar zu machen.
- Es muss sowohl die gezielte Suche nach Fakten und/oder Dokumenten als auch die explorative Erschließung des Informationsraums unterstützt werden.

Aus diesem Grund bietet Aletheia verschiedene Arten des Informationszugriffs. Es werden folgende Ansätze umgesetzt:

-
- Eine Volltextsuche erlaubt das Auffinden von Dokumenten anhand von Stichworten, die in den Dokumenten vorkommen.
 - Eine intelligente, semantische Suche findet Dokumente und Fakten anhand der Konzepte und deren Verknüpfungen in der Ontologie.
 - Mittels einer facettenorientierte Suche kann sich der Anwender einen Überblick über alle Fakten/Dokumente zu einem wählbaren Begriff anzeigen lassen, diesen Überblick weiter verfeinern (drill-down), sowie zu verwandten Themen navigieren.
 - Eine Darstellung von Suchergebnissen in einem semantischen Netz erlaubt das visuelle Navigieren durch den Informationsraum.

Diese Zugriffsmöglichkeiten werden im Folgenden detailliert erläutert.

Volltextsuche: Die Volltextsuche ist eine rein syntaktische Suche, die Dokumente findet, in denen die eingegebenen Suchterme vorkommen. Dieses ist die einfachste Art des Informationszugriffs in Aletheia.

Semantische Suche: Bei der semantischen Suche werden sowohl Fakten als auch Dokumente gefunden, und die eingegebenen Suchterme werden semantisch interpretiert.

Die im semantischen Repository persistierten Entitäten (Instanzen oder Konzepte) sind mit Stichworten annotiert, durch die sie gefunden werden können. Zunächst werden bei der Eingabe von Suchtermen mittels Auto-complete Entitäten angezeigt, die auf den bisher eingegebenen Suchstring zutreffen. Das ermöglicht insbesondere bei Termen, die auf mehrere Entitäten verweisen, die intendierte Entität auszuwählen. Somit werden bereits bei der Eingabe mehrdeutige Suchterme disambiguiert.

Die eingegebenen Suchterme müssen nicht zusammen auf *eine* Entität verweisen, stattdessen wird die Sucheingabe als strukturierte semantische Query interpretiert, die die Beziehungen zwischen den Instanzen der zu durchsuchenden strukturierten Datenquellen berücksichtigt. Mit anderen Worten: Die einzelnen Suchstrings können (aber müssen nicht) auf *verschiedene* Instanzen in der Ontologie verweisen, die über Beziehungen in der Ontologie miteinander verknüpft sind. Dem Benutzer wird eine graphische Darstellung der gefundenen Zusammenhänge angezeigt.

Facettenorientierte Suche: Im Gegensatz zu den bereits beschriebenen Suchmöglichkeiten bietet die facettenorientierte Suche einen eher explorativen Zugriff, der durch das Navigieren durch den Informationsraum geprägt ist.

Als Facetten werden die Klassen der Ontologie sowie deren Attribute und Attributwerte als Facetten herangezogen. Die Werte einer Facette können sowohl als Liste als auch als Cloud dargestellt werden (siehe Abbildung 4).

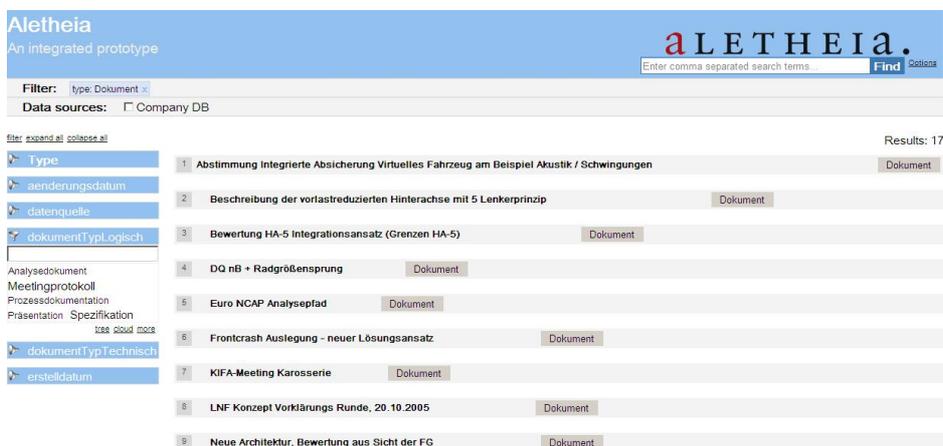


Abbildung 4: Die facettenorientierte Suche in Aletheia

Die Ergebnisliste bei einer facettenorientierten Suche sind immer Dokumente oder Fakten gleichen Types (d.h., zum selben Konzept gehörend). Konzepte sind in der Ontologie über Relationen miteinander verbunden. Im Gegensatz zu Attributen werden Relationen (zwischen Instanzen der Ontologie) für die Funktionalität „Navigate to Type“ benutzt. Mit dieser Funktionalität ist es möglich, durch den Informationsraum anhand semantischer Relationen zu navigieren (siehe Abbildung 5).

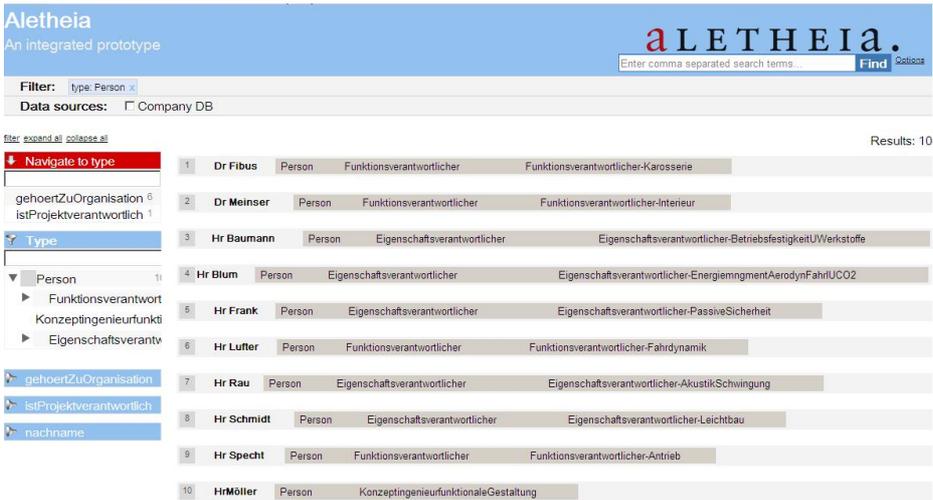


Abbildung 5: Navigation entlang semantischer Relationen in Aletheia

Semantisches Netz: Ebenso wie die facettenorientierte Suche dient auch die Darstellung als semantisches Netz dem explorativen Erschließen des Informationsraums. Noch stärker als beim facettenorientierten Ansatz werden nun nicht nur die Entitäten, sondern auch die semantischen Beziehungen zwischen diesen ausgenutzt. Im Gegensatz zur facettenorientierten Suche können nun Entitäten verschiedenen Types gleichzeitig dargestellt werden. Weiterhin werden die semantischen Beziehungen zwischen diesen Entitäten nun explizit visualisiert. Die Beziehungen dienen in diesem Ansatz nicht zum *Wechsel* der dargestellten Informationen (von einem Typ zu einem anderen), sondern zur Änderung der Menge angezeigter Entitäten. Die Darstellung als semantisches Netz ist im Rahmen von Aletheia bisher nur in einem proprietären Prototyp verfügbar und wird zur Zeit der Erstellung dieser Arbeit in das allgemeine Aletheia-System integriert. Ein Screenshot des proprietären Prototyps findet sich in [Abbildung 6](#).

Durch die oben dargestellten Arten des Informationszugriffs ergibt sich für den Anwender in der Konzeptphase folgender Nutzen: Zum einen hat er Zugriff auf die für seine Arbeit relevanten Informationen, wodurch beispielsweise die von ihm zu treffenden Entscheidungen auf einer stärkeren Informationsbasis beruhen. Zum anderen werden ihm mit Hilfe semantischer Netze Abhängigkeiten zwischen Entwicklungsthemen dargestellt, die per se nicht allen Beteiligten bekannt sind.

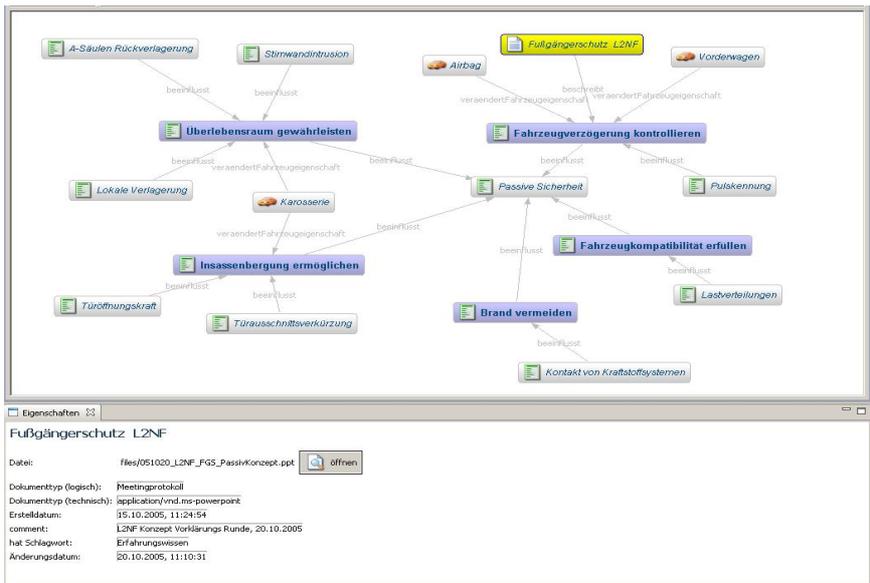


Abbildung 6: Informationsdarstellung als semantisches Netz mit Fakten aus strukturierten Quellen sowie (unstrukturierten) Dokumenten

6 Zusammenfassung und Ausblick

Die effiziente Verfügbarkeit relevanter Daten stellt für die Konzeptphase einen Erfolgsfaktor dar. Methoden und Werkzeuge des Datenmanagements können derzeit nicht zureichend die Anforderungen dieser Phase erfüllen.

Die vorliegende Arbeit beschreibt Grundlagen und Methoden für ein Informationssystem, das dem Anwender produktbezogene Informationen in der Fahrzeugkonzeptentwicklung zugänglich macht. Es werden semantische Technologien genutzt, um in unstrukturiertem und implizitem Wissen Strukturen und Beziehungen zu erkennen und zugreifbar zu machen.

Weiterführende Arbeiten sollten den rollen- und kontextspezifischen Datenbedarf verschiedener Entwickler ermitteln. Darüber hinaus ist das User-Tagging sowie semantische Annotieren von Dokumenten zu untersuchen.

Die hier beschriebenen Arbeiten wurden teilweise mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IA08001A/B gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literatur

- [1] Bates, M. J.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Review, Vol. 13, No. 5. pp. 407-424, 1989.
- [2] Becker, A.; Walther, M.; Reichert, S. und Hladik, J.: D.G4.2 Repository-Spezifikation. Technical report, ontoprise GmbH, TU Dresden, SAP AG, 2009. Öffentliches Deliverable.
- [3] Dau, F.; Hladik, J.; Becker, A.; Brookmans, S.; Korf, R.; Erdmann, M. und Niemann, M.: D.G4.1 Modellierungsmethodik und globales semantisches Modell. Öffentliches Deliverable, SAP AG, 2009.
- [4] Dau, F.; Hladik, J.; Becker, A.; Niemann, M. und Walther, M.: D.G4.3 Harmonisierungsmechanismen. Öffentliches Deliverable, SAP AG, 2009.
- [5] Fensel, D.: Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. Springer Verlag, Berlin, 2004.
- [6] Staab, S.; Studer, R.: Handbook on Ontologies (2nd edition). Springer Verlag, Berlin, 2009
- [7] Teevan, J.; Alvarado, C.; Ackerman, M. S. und Karger, D. R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. Proceedings of the SIGCHI conference on Human factors in computing systems, 415 – 422, 2004.
- [8] Yee, K.-P.; Swearingen, K; Li, K. und Hearst, M.: Faceted metadata for image search and browsing. Proceedings of the SIGCHI conference on Human factors in computing systems. Pages: 401 – 408, 2003.