

**Informationsmanagement auf der Basis von Semantic-Web Techniken  
oder  
Ein Google für Entwickler**

**Information Management on the Basis of Semantic-Web Techniques, or  
A Google for Developers**

Dr. **B. Thelen**, Schenck Pegasus GmbH, Darmstadt

**N. Sevilmis**, Dr. **A. Stork**, Fraunhofer Institut für Graphische Datenverarbeitung,  
Darmstadt

**R. Castro**, Centro de Computação Gráfica, Guimarães, PT

**I. Jimenez**, **G. Marcos**, **J. Posada**, Dr. **T. Smithers**, VICOMTech, Donostia, ES

**M. Mauri**, **M. Pianciamore**, **P. Selvini**, CEFRIEL, Milano, I

**V. Zecchino**, Italdesign – Giugiaro SpA, Moncalieri (Torino), I

### **Zusammenfassung**

Eine Informationsrecherche leidet häufig an geeigneten Recherchehilfsmitteln oder der Komplexität der Recherche an sich. Die Suche einer konkreten Information anhand von Dateinamen oder dem zufälligen Vorkommen von Stichworten in Dateien ist wenig hilfreich, da die erzielbaren Treffer zu sehr dem Zufall unterliegen. Es wird die Entwicklung des Suchmaschinenprototyps WIDE vorgestellt, der die Query-Interpretation auf Semantic-Web-Techniken aufbaut und dabei Queries auf den Kontext eines Aufgabengebiets abbildet. Die Suchmaschine kann für Aufgabenfelder konfiguriert werden und ist zusätzlich in der Lage, eine Query auf verschiedene Datenquellen gleichzeitig abzubilden. Die Suchmaschine bereitet die gefundenen Ergebnisse graphisch auf und assoziiert die in der Query benutzten Begriffe mit thematisch verwandten Begriffen. Die Suchmaschine kann zur Recherche von Textdokumenten, oder Prüfstandsversuchsergebnissen in ASAM-ODS-Datenquellen genutzt werden.

### **Summary**

Information retrieval often suffers from the lack of suitable search tools or the query complexity. The search of some concrete information on the base of file names or the coincidental occurrence of key words in files is little helpful because the obtainable matches are too much subject to chance. Therefore an effective search must be based on the semantic interpretation of the query and additionally casts of the query into the context of an application domain. Here the development of the search machine prototype WIDE is presented, which builds up the query interpretation on Semantic Web techniques. The search machine can be configured for application domains and is able to map a query to different data sources in parallel. The search machine processes the retrieved results graphically and associates the concepts used in the Query with thematically related concepts. The search machine can be used to retrieve text documents or test bed results of experiments archived in ASAM-ODS data sources.

## Entwicklungsprozess und Datenrecherche

Im Verlaufe eines Entwicklungsprozesses ist das Recherchieren von Informationen eine gängige Aufgabe. Die besondere Schwierigkeit ist stets, die „richtigen“ Suchbegriffe zu kennen, unter denen die gesuchten Informationen abgelegt sind. Eine noch viel größere Schwierigkeit aber ist die, Zusatzinformationen aufzufinden, die in einem thematischen Zusammenhang mit den verwendeten Suchbegriffen stehen, die aber der benutzte Suchbegriff nicht explizit ausdrückt. Ein erfolgreiches Auffinden auch thematisch verwandter Informationen bedeutet ein besseres Ausleuchtung des Problemfeldes. Im Falle einer Motorentwicklung können die Ergebnisse eines geplanten Versuchs bereits in verschiedenen anderen Versuchen existieren, aber diese Versuche wurden unter einem anderen Aspekt durchgeführt. Gelingt es diese Versuchsergebnisse aufzufinden, lassen sich in günstigen Fällen evt. einzelne Motorversuche einsparen oder zumindest zielgerichteter gestalten. Für den suchenden Ingenieur stellt sich stets die Frage, in welchen Datenquellen soll er die benötigten Informationen suchen und vor allen Dingen, wie soll er suchen. Der letzte Punkt beinhaltet zwei prinzipielle Probleme

- die Kenntnis geeigneter Suchbegriffe
- die semantische Transformation der Suchbegriffe.

Diese Problematik sei anhand Bild 1 erläutert. Zwei Entwickler suchen Informationen über ein Fahrzeug (gleiches Konzept). Beide benutzen ihr eigenes Vokabular zur Bezeichnung des gesuchten Gegenstands, wobei die Begriffe synonymisch identisch sind, aber dennoch unterschiedliche Terminologie darstellen. Jeder assoziiert eine bestimmte Ausprägung mit dem verwendeten Begriff (eine Autoskizze, ein reales Auto). Die fachgerechte, semantische Interpretation gelingt nur, wenn zusätzlich der Kontext der Fragesteller berücksichtigt wird, d.h. in welchem Aufgabenprozess sich der jeweilige Fragesteller befindet.

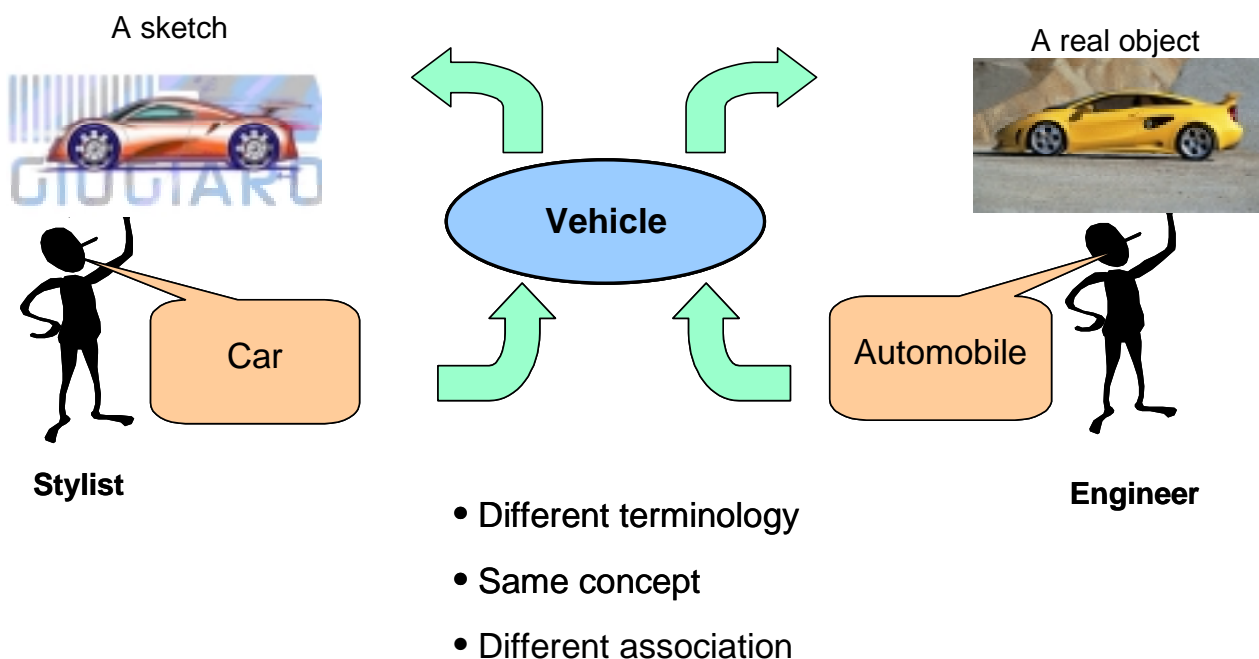


Bild 1: Sprachproblem bei der Datenrecherche  
Figure 1: The language problem in data retrieval

Suchverfahren, die auf einem reinen Textmustervergleich beruhen (text pattern matching), versagen vollständig bei der Query-Bearbeitung, wenn eine semantische Interpretation in Bezug auf einen Kontext nötig ist. Eine Alternative zu vollautomatisierten Sucherverfahren sind solche, die eine manuelle Navigation durch hierarchische Datenstrukturen unterstützen und dabei mittels Filterfunktionen Teile der Dateninhalte wegblenden, um den Suchraum einzuschränken. Hier ist das Suchproblem in erster Linie auf den Benutzer verlagert, der anhand der vorgefundenen Datenelemente selbst entscheiden muss, ob die präsentierten Informationen seiner impliziten Fragestellung entsprechen. Noch schwieriger wird die manuelle Suche, wenn Querbeziehungen zwischen Datenelementen zu berücksichtigen sind. Beispiel: die Suche nach einem bestimmten Versuch, der mit einem bestimmten Prüfling auf einem bestimmten Prüfstand ausgeführt wurde. Der gesuchte Test muss eine Reihe von Zusatzbedingungen erfüllen, die alle manuell zu verifizieren sind (Attribute & Werte, Objektrelationen). Eine derartige Suche ist manuell nicht mehr mit vertretbarem Zeitaufwand durchführbar.

Ein weiteres Suchproblem ist gegeben, wenn der Fragesteller Informationen zu thematisch verwandten Begriffen sucht, d.h., der Hintergrund der Fragestellung darin besteht, den Kontext zu einem Begriff auszuleuchten. Beispiel: Informationen über Emissionsstandards. Hier erwartet der Fragesteller als Ergebnis eine Liste mit erweiterten Begriffen, die mit dem ursprünglichen in Zusammenhang stehen. Im Falle von „Emissionsstandard“ sind das die damit assoziierten Begriffe: Gesetzestext, technische Beschreibung, Grenzwerte, Messvorschriften und Testzyklen (Bild 2).

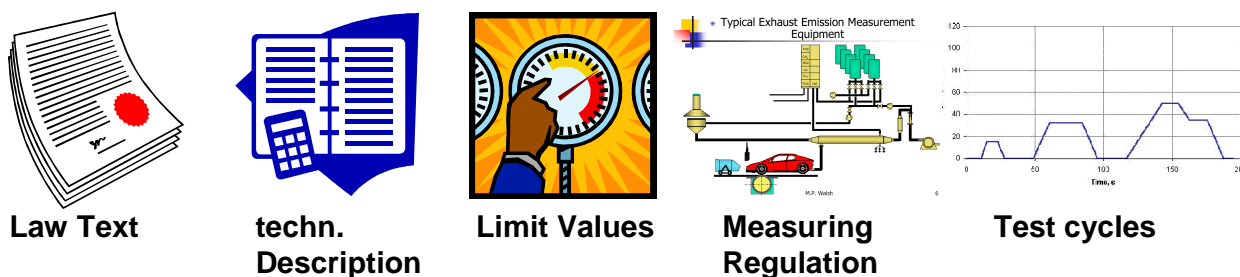


Bild 2: Assoziierte Begriffe zu Emissionsstandard  
 Figure 2: Terms related with emission standard

Mit der so erzielten Erfassung des Begriffskontexts bekommt der Fragesteller die Möglichkeit, selbst zu evaluieren, wie umfassend seine Suchabfragen beantwortet werden können, wie die Fragestellungen verstanden werden, an welche Begriffe zusätzlich zu denken ist, bzw. welche erweiterten Informationskategorien zu berücksichtigen sind. Diese Aspekte verlangen die thematische Vernetzung von Begriffen.

Die hier skizzierten Problematiken bei der Informationsrecherche verdeutlichen, welche Anforderungen eine Suchmaschine erfüllen muss, um dem Fragesteller zielgerichtet die relevanten Informationen zu liefern.

## Semantische Interpretation von Queries

Einer der Schwerpunkte bei der Entwicklung der hier vorgestellten Suchmaschine WIDE geht von der Annahme aus, dass verschiedene Benutzer aus unterschiedlichen Disziplinen oder Entwicklungsprozessen, ihr aufgabenspezifisches Vokabular pflegen, wenn sie nach gleichen oder ähnlichen Informationen suchen. Die Benutzer der Suchmaschine sollen daher nicht gezwungen sein, eine besondere Terminologie zu übernehmen. Die benutzerspezifischen Begriffe werden mittels Glossaren, Thesaurus und fachspezifischen Ontologien übersetzt (Bild 3). Agenten übernehmen die semantisch aufbereiteten Queries und beginnen die Suche in der ihnen zugeordneten Datenquelle. Nach einer fest vorgegebenen Zeit melden sie sich mit den bis dahin gefundenen Ergebnissen zurück.

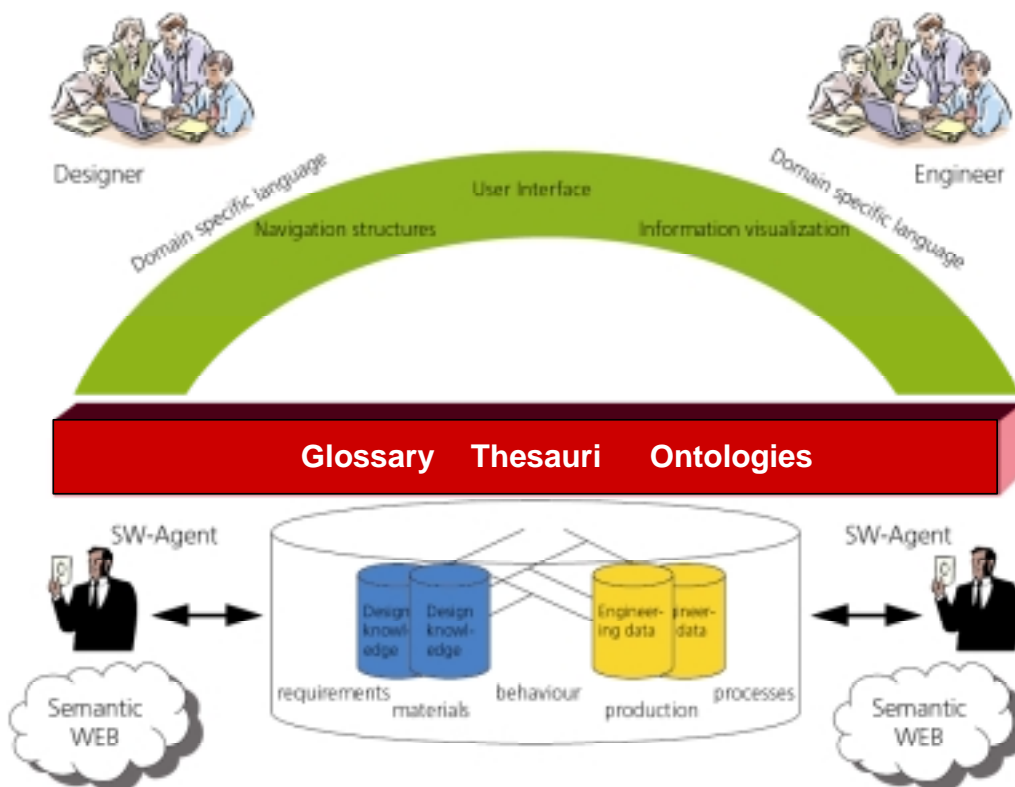


Bild 3: Neue Technologien zur Datenrecherche  
Figure 3: New technologies for data retrieval

Bei der semantischen Interpretation der Benutzerbegriffe spielen Ontologien die zentrale Rolle. Der Begriff Ontologie entstammt dem Griechischen Ontos, und bedeutet in der Philosophie „die Lehre vom Seienden, die Lehre dessen, was ist“. In der Informationstechnik wird der Begriff in einer profaneren Weise zur Strukturierung von Wissen genutzt durch

- explizite Spezifikation von Begriffen (concepts) und
- deren Beziehungen in einem Bereich (relations, domain)

Konkret bedeutet das, für ein bestimmtes Applikationsgebiet (z.B. Fahrzeug-/Motorentwicklung) eine Begriffsstrukturierung für die zu unterstützenden Entwicklungsprozesse vorzunehmen und thematisch verwandte Begriffe miteinander in Beziehung zu setzen. Bild 4 zeigt ein Beispiel. Eine Fahrzeugentwicklung ist u.a. beeinflusst von Spezifikationen zu Fahrzeugkategorien und für diese sind wiederum Spezifikationen zu Emissionsstandards maßgebend. Die in den einzelnen Fachdomänen relevanten Begriffe wurden erfasst und zueinander in Beziehung gesetzt. Die formale Beschreibung dieser Strukturen erfolgt mittels RDF(S) (Resource Description Framework) [2]. RDF ist eine Sprache zur Darstellung von Metainformation (Daten über Daten). Im wesentlichen definiert RDF ein Sprachmodell nebst Syntax, um objekt-orientiert ausgedrückte Information in XML darzustellen.

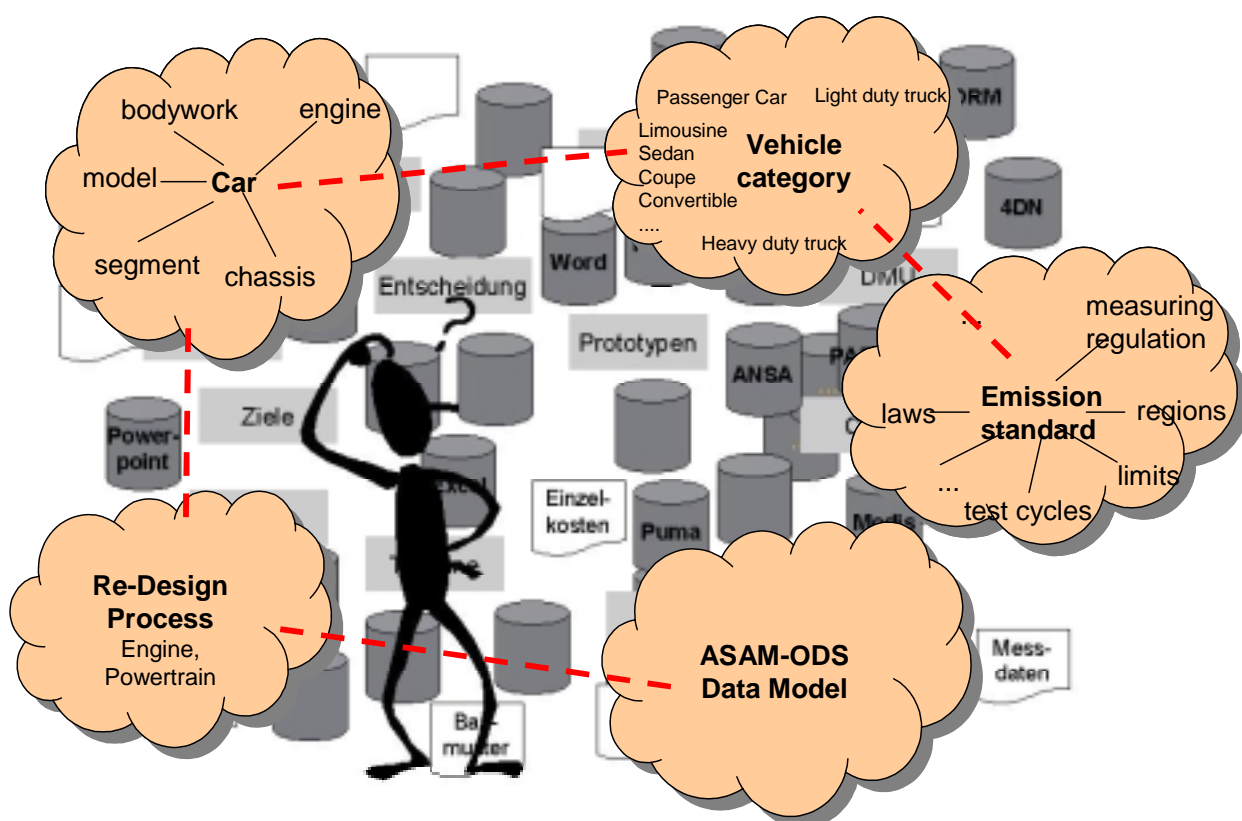


Bild 4: Strukturierung von Wissen mittels Ontologien  
 Figure 4: Structuring knowledge by ontologies

Diese Beziehungen zwischen den Begriffen werden genutzt, um von einem Begriff aus einer Query zu fachlich verwandten, weitergehenden Begriffen mittels der Ontologie traversieren zu können. Auf diese Weise ist gewährleistet, dass zu einem Begriff in einer Query, weitere in Zusammenhang stehende Begriffe gefunden werden (Bild 5). Eine zusätzliche Begriffsübersetzung mit Hilfe des Glossars und Thesaurus entbindet den Benutzer davon, das in einer Datenquelle benutzte Vokabular kennen zu müssen.

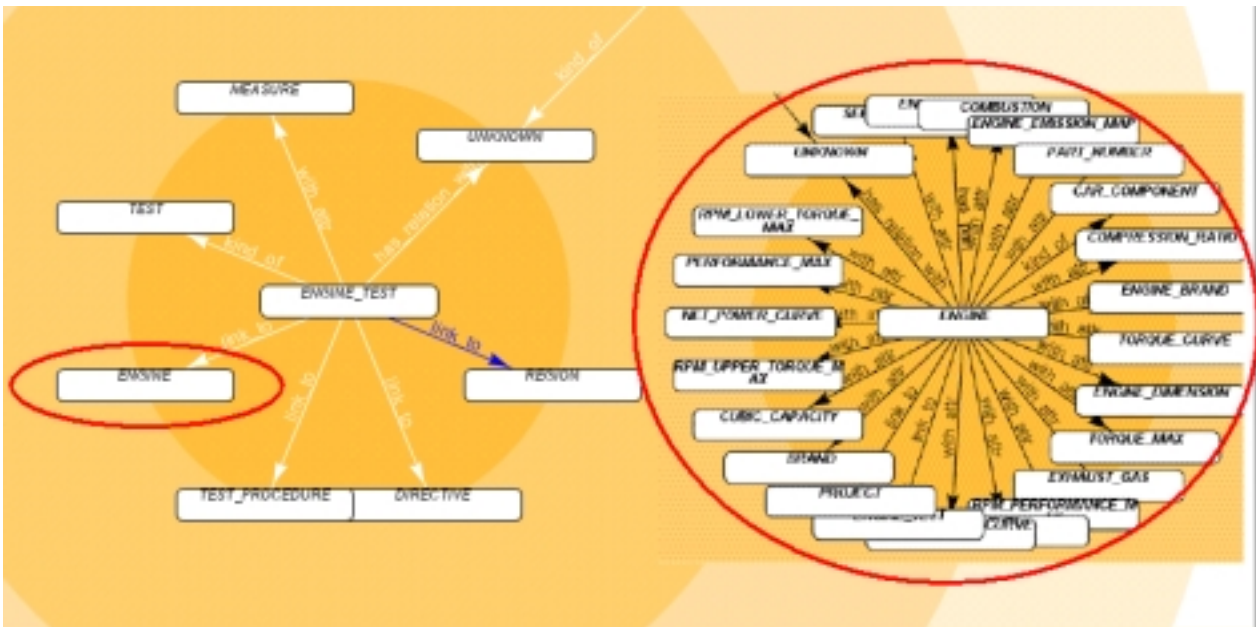


Bild 5: Strukturierung von Wissen mittels Ontologien am Beispiel Motor-Test  
 Figure 5: Structuring knowledge by ontologies demonstrated by Engine-Test

### Einbindung externer Datenquellen am Beispiel ASAM-ODS

Die Einbindung externer Datenquellen in die Suchmaschine muss drei Aspekte berücksichtigen


- die logische Datenstruktur bezüglich der Applikationen, deren Daten archiviert sind
- die implementierte, archivspezifische Datenstruktur inkl. der intern benutzten Begriffsbezeichnungen für die Datenelemente
- die Programmierschnittstelle zur Datenquelle.

Der erste Aspekt bezüglich der logischen Datenstruktur ist unabhängig zu betrachten von der besonderen Datenquelle, denn die Suchmaschine ist in der Lage, eine Benutzer-Query gleichzeitig auf verschiedenartige Datenquellen abzubilden. Aus diesem Grund sind applikations- bzw. prozessspezifische Begriffsstrukturen in der Ontologie der Suchmaschine hinterlegt. Die Handhabung der archivspezifischen Datenstrukturen ist dagegen von der systemeigenen Ontologie getrennt gehalten und befindet sich in den sogenannten „provider-spezifischen Ontologien“ und zwar für jeden Datenquellentyp. Zusätzlich existiert für jede Datenquelle ein Thesaurus und Glossar, die die Begriffsübersetzung in die in der Datenquelle verwendeten Begriffe unterstützen. Der eigentliche Zugriff auf eine Datenquelle obliegt einem datenquellenspezifischen Agenten (Provider-Agent), der die Programmierschnittstelle zur Datenquelle bedienen kann.


Die Entscheidung zum Anbinden einer ASAM-ODS Datenquelle an die Suchmaschine geschah unter dem Aspekt, dass ASAM-ODS im Forschungsbereich der Automobilindustrie ein weltweit akzeptierter Standard ist und ein für diesen

Datenquellentyp zu entwickelnder Agent allgemein einsetzbar ist. Die Spezifikation für die ASAM-ODS-Datenquelle ist in /1/ frei zugänglich. ASAM-ODS wurde für das Archivieren von Daten im Kontext des Produkttestens konzipiert. Die Grunddatenstruktur ist durch die Spezifikation vorgegeben, die Anpassung an ein Applikationsfeld geschieht kundenspezifisch. Der Kunde kann daher eigene Strukturelemente hinzufügen und diese selbst benennen (z.B. Motor, Bremse, ... etc. als Erweiterungen des in der Spezifikation allgemein vordefinierten Prüflings). Diese Spezifika sind in der ASAM-ODS Provider-Ontologie hinterlegt. Bild 6 verdeutlicht die Stufen im Übersetzungsprozess einer Benutzer-Query. Die erste Phase übersetzt die Query-Begriffe mit Hilfe der System-Ontologie, während die zweite Phase die Begriffe auf das in der ODS-Datenquelle verwendete Vokabular abbildet.

**Test with Description=Full Load Test linked\_to Engine with cylinder=4, cc= 2.0 ltr**

- System Ontology maps to Application Domain 
  - Ontologies
  - Thesaurus
  - Dictionaries

**Test is a subdomain of vehicle development**  
**Test is a synonym for a Measurement**  
**Description is an attribute of Test**

**Engine is a subdomain of vehicle development**  
**cylinder is an attribute of engine**  
**cc is an attribute of engine**  
**cc is a synonym for cubic capacity**
- Provider Ontology maps to ASAM-ODS 
  - Application Elements
  - Application Attributes
  - AE Relations

**Test is mapped to AE\_Test**  
**Engine is mapped to AE\_Engine**  
**cc is mapped to CubicCapacityRounded**

**AE\_TEST & AE\_Engine are related**

Bild 6: Semantische Übersetzung einer Query in ASAM-ODS Terminologie  
 Figure 6: Semantic translation of a user query into ASAM-ODS terminology

### Systemarchitektur

Das Benutzerinterface übernimmt die eingegebene Query, unterwirft sie einer formalen Syntaxprüfung und gibt sie danach an die Agentur weiter (Bild 7, Punkt 1). Die Agentur treibt die semantische Query-Übersetzung mit Hilfe des Meta-Levels. Der Meta-Level beherbergt die fach-/applikationsspezifischen Ontologien, den Thesaurus und Glossare. Aus einer Benutzer-Query können mehrere System-Queries entstehen (Bild 7, Punkt 2). Der Content-Level ist für die Übersetzung der System-Queries in das Vokabular einer bestimmten Datenquelle zuständig (Bild 7, Punkt 3; z.B. in das Vokabular eines ASAM-ODS-Servers). Die endgültig aufbereiteten und übersetzten Queries übergibt die Agentur an Agenten (Bild 7, Punkt 4, die das jeweilige Interface einer Datenquelle beherrschen (Bild 7, Punkt 5).

Die Agenten verfügen über ein fixes Zeitintervall, innerhalb dessen sie die bis dahin gefundenen Ergebnisse an die Agentur zurück liefern müssen. Die Agentur

strukturiert die gefundenen Ergebnisse mit Hilfe des Meta-Level und gibt sie an das Benutzer-Interface zur grafischen Aufbereitung (Bild 7, Punkt 6). Die Strukturierung der Ergebnisse geschieht anhand der Query und der Ontologie (Bild 8).

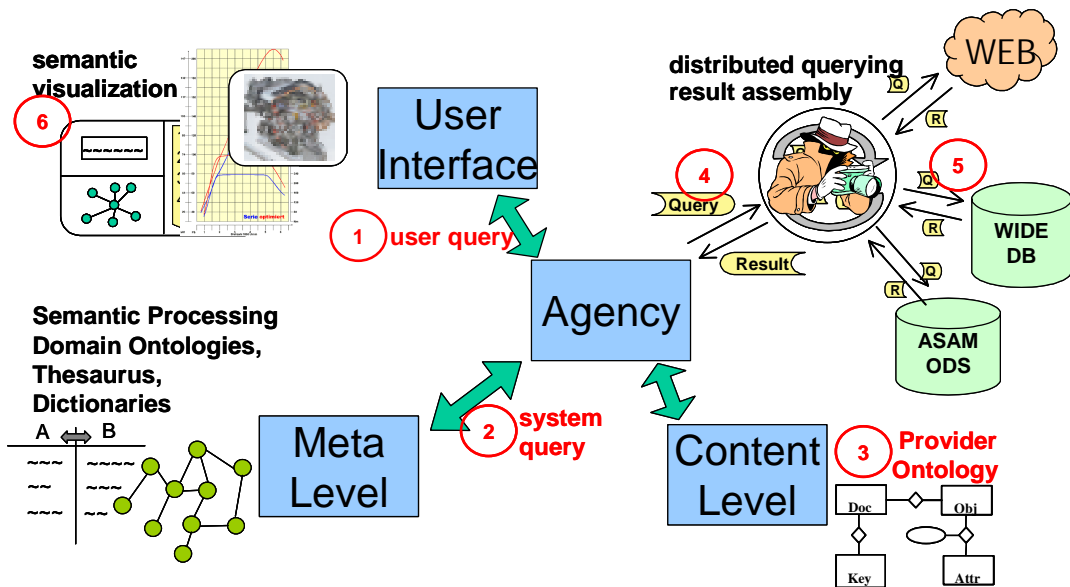


Bild 7: Bearbeitungsschritte einer Benutzer-Query  
Figure 7: Execution steps of a user query

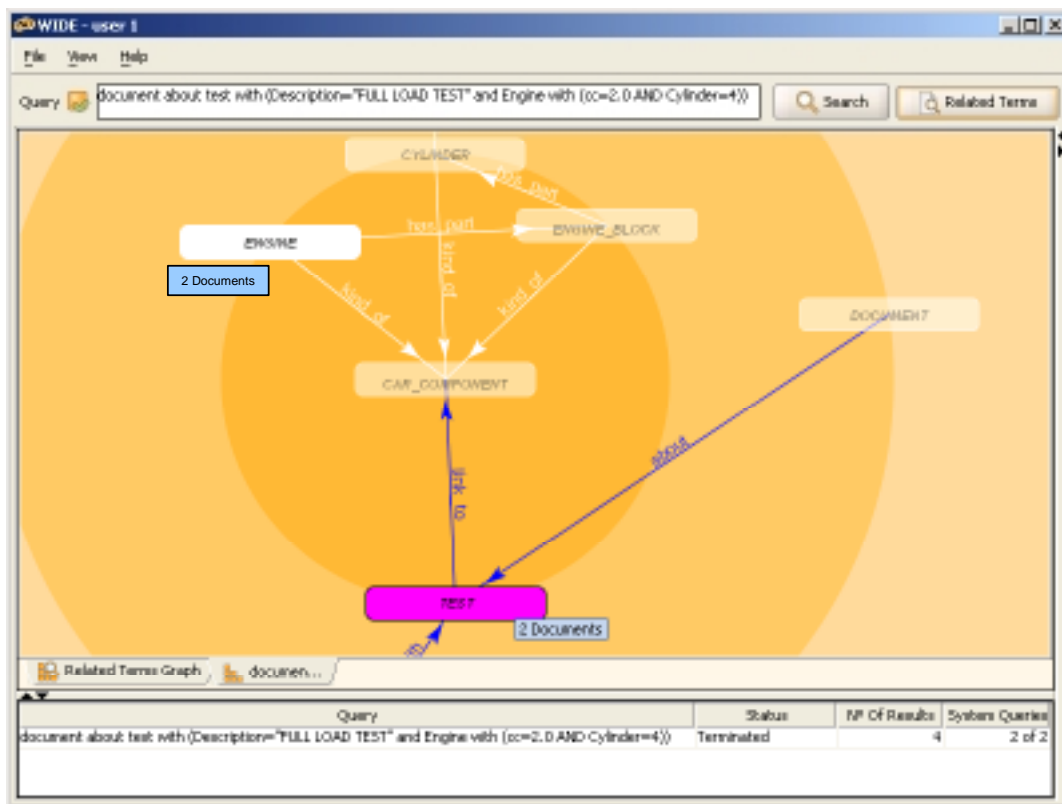


Bild 8: Ergebnisanordnung nach Begrifflichkeiten der Ontologie und der Query  
Figure 8: Graphical result arrangement according to the ontology and query relations



## Export von Suchergebnissen

Eine Recherchesitzung mit einem Suchwerkzeug wird typischerweise folgende Schritte wiederholt durchlaufen

- Formulieren einer Query
- Bewerten der Fundobjekte anhand ihrer Kurzbeschreibung bzw. Attributwerte
- Auswählen von „interessanten Fundobjekten“
- Ablegen der Fundobjekte in einer Zwischenablage

Nach Abschluss der Recherchesitzung stehen die gesammelten Fundobjekte einem Auswertewerkzeug zur weiteren Bearbeitung zur Verfügung. Der hier beschriebene Vorgang wurde in der Suchmaschine implementiert. Die Zwischenablage heißt Arbeitsbuch und dokumentiert eine Recherchesitzung mit a) Namen des Benutzers, b) Datum der Sitzung, c) die in die Suchmaschine eingegebenen Queries mit den dazu ausgewählten Fundobjekten. Die Referenzen auf die Fundobjekte können Pfadangaben zum Ablageort eines Dokuments sein (URLs) oder ASAM-Pfade. Ein ASAM-Pfad ist eine Referenz auf ein Objekt innerhalb einer ASAM-ODS Datenquelle. Diese Pfadangabe ist über den Lebenszyklus des Suchprozesses hinaus gültig.

Der Export eines Arbeitsbuches erfolgt automatisch am Ende der Recherchesitzung; der Benutzer kann auch zu jedem anderen Zeitpunkt während der Recherchesitzung den Export auslösen. Der Export des Arbeitsbuchinhalts erfolgt mittels einer XML Datei, deren Schema in einer XSD-Datei beschrieben ist. Auf diese Weise ist eine offene Schnittstelle der Suchmaschine zu Auswertewerkzeugen von Drittanbietern gewährleistet (Bild 9).

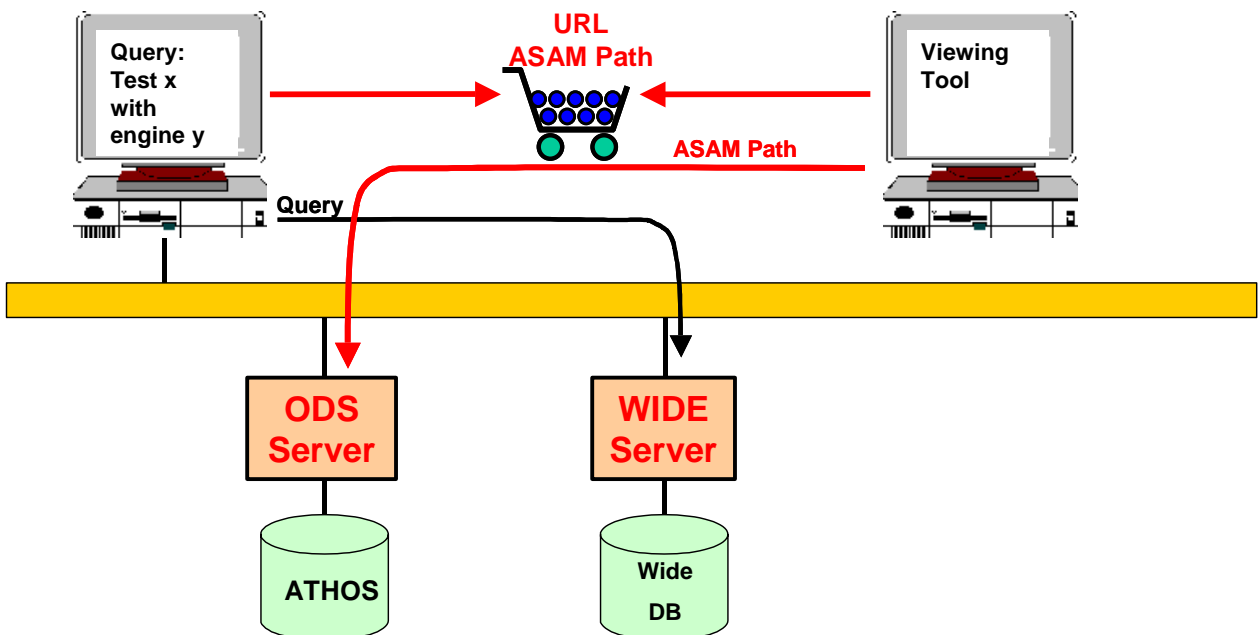


Bild 9: Export der Suchergebnisse zur Auswertung durch Drittwerkzeuge  
Figure 9: Export of search results for evaluation by 3<sup>rd</sup> party tools

Bild 10 zeigt ein Beispiel mit dem Inhalt des Arbeitsbuches. In der Recherchesitzung schickte der Benutzer zwei Queries an die Suchmaschine. Von den gefundenen Ergebnissen wählte er sechs für den Export in das Arbeitsbuch aus, zwei Ergebnisse über bestimmte Tests mit dem zugehörigen Motor und drei Dokumente über Abgasemissionstestzyklen). Alle Daten zum Test mit den zugehörigen Motordaten liegen in der ASAM-ODS Datenablage und können mit den im Arbeitsbuch enthaltenen ASAM-Pfaden für weitergehende Untersuchungen geladen werden, z.B. Anzeigen der Attribute eines Tests, bzw. Motors (Bild 10) oder Generieren von Messdatendiagrammen für im Test erfasste Prozessgrößen.

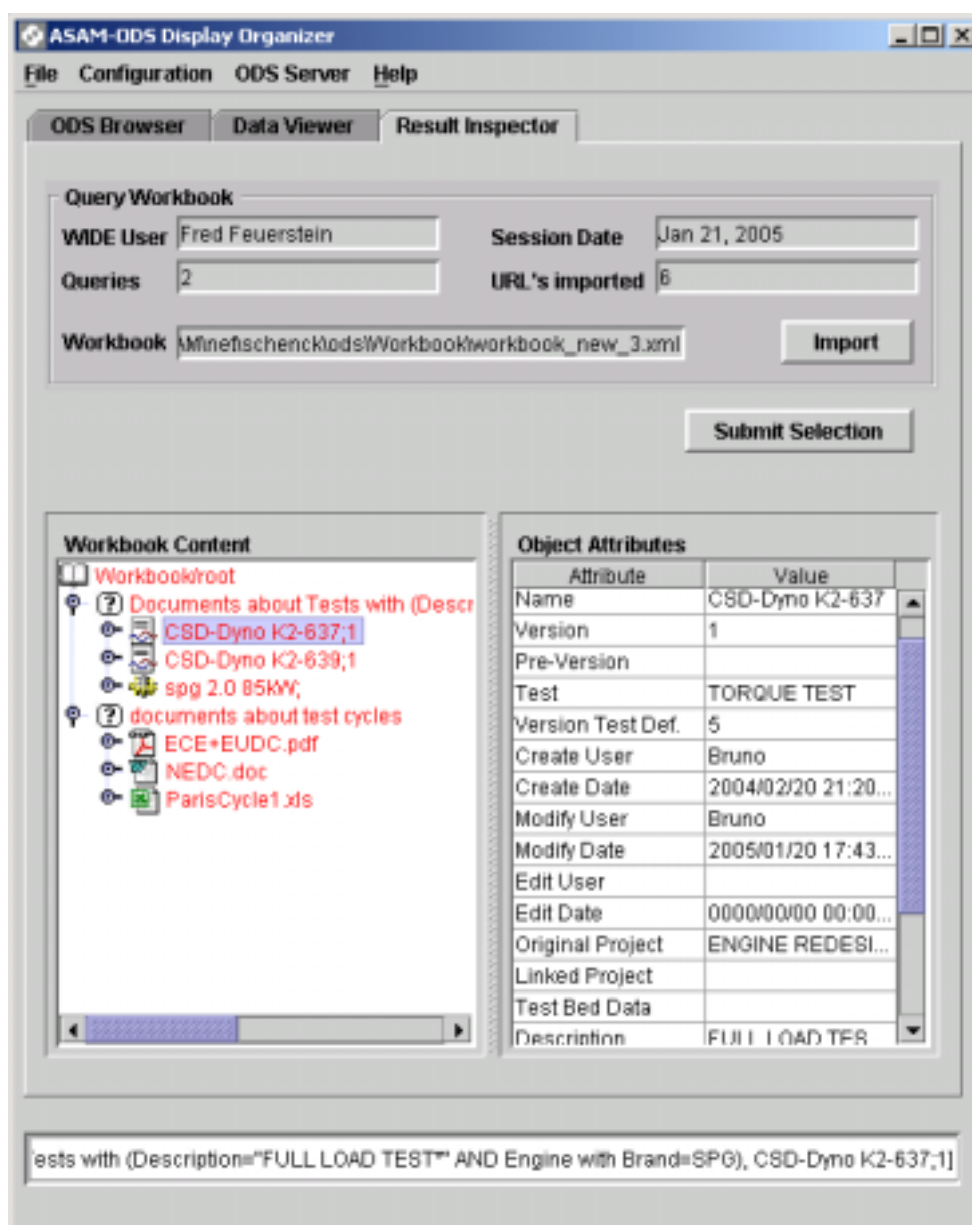


Bild 10: Beispiel für den Inhalt des eingelesenen Arbeitsbuchs  
 Figure 10: Example for an imported workbook

## Import neuer Dokumente in das Suchmaschinenarchiv

Die Suchmaschine unterstützt die Recherche von Objekten in einem ASAM-ODS Datenarchiv. Dieses Datenarchiv enthält aber direkt keine Messdatendiagramme, sondern numerische Messdaten, die ein geeignetes Auswertewerkzeug einliest, um entsprechende Diagramme zu generieren. Der Benutzer kann nun das Interesse haben, besondere Messdatendiagramme (z.B. typische Schadensbilder wie rotatorische Schwingungen als Folge von Pleuellagerschäden, ... etc.) als neue, eigenständige Dokumente zusammen mit einer Kurzbeschreibung in das Archiv der Suchmaschine aufzunehmen. Die Suchmaschine bietet hierfür eine entsprechende Schnittstelle.

Bild 11 zeigt ein Messdatendiagramm, das der Suchmaschine als neues Dokument bekannt gegeben werden soll. Die im Bild gezeigte Eingabemaske gehört zu einem Werkzeug, das die Schnittstelle zum Einfügen neuer Dokumente unterstützt. Die Einträge unter der Rubrik „Executed Test“ wurden automatisch über den zuvor ausgewählten Versuch (Bild 10) aus der ASAM-ODS Datenquelle entnommen; die Einträge unter der Rubrik „Content Document“ füllt der Benutzer aus. Das sind die Metadaten, unter denen das Dokument über die Suchmaschine später recherchiert werden kann. Ein Klick auf den Kopf „Submit to Wide Data Source“ löst die Übermittlung des Metadatenobjekts zur Suchmaschine resp. deren Datenbank aus.

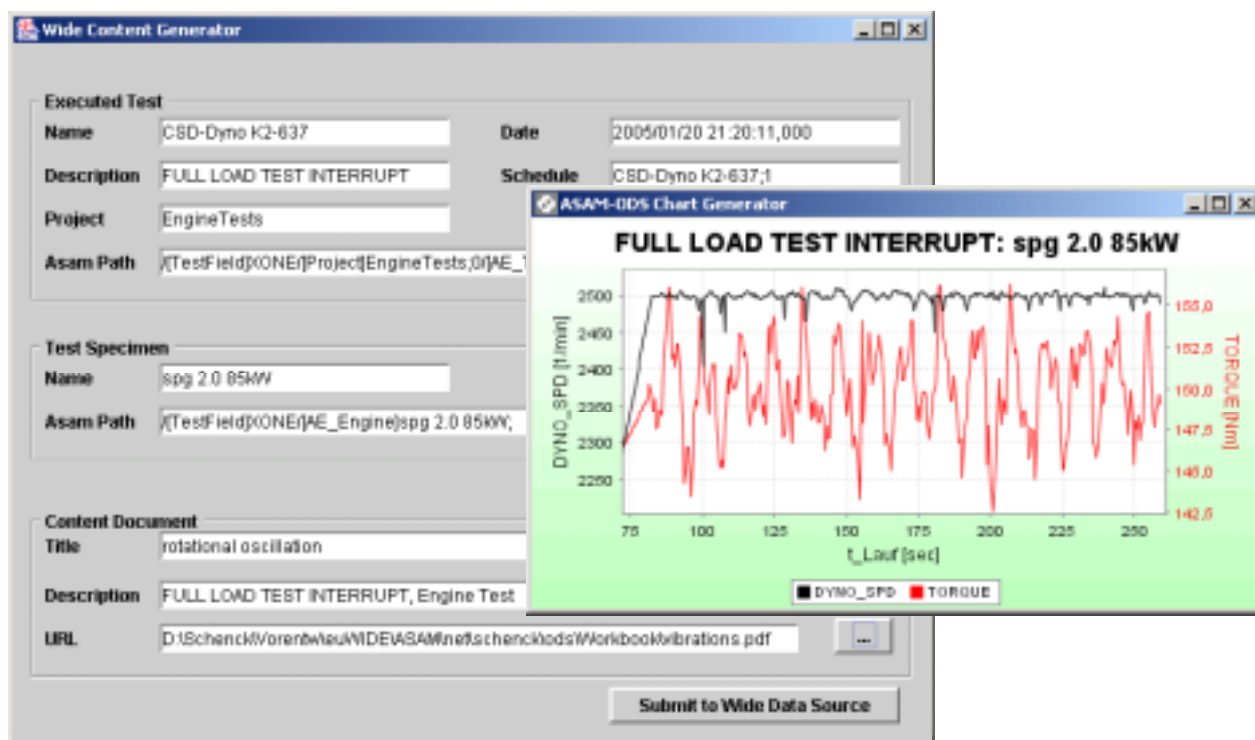


Bild 11: Importieren neuer Dokumente in das Archiv der Suchmaschine  
Figure 11: Import of new documents into the search engine's archive

## **Zusammenfassung**

Die Suchmaschine WIDE zeichnet sich durch ihre flexible Anpassbarkeit an Applikationsgebiete und Entwicklungsprozesse aus. Das Wissen der Suchmaschine ist in Ontologien gespeichert, mit deren Hilfe die Begriffe aus den gespeicherten Applikationsgebieten strukturiert und zueinander in Beziehung gesetzt sind. Eingegebene Benutzerqueries werden mittels Ontologie, Thesaurus und benutzerspezifischer Glossare semantisch interpretiert. Das hat den Effekt, dass die ursprüngliche Benutzerquery in neue Queries mit anderen Begriffen übersetzt wird. Die Suchmaschine übergibt die überarbeiteten Queries an Agenten, die die eigentliche Recherche in dem ihnen zugeordneten Datenarchiven parallel ausführen. Die gefundenen Ergebnisse werden entsprechend der Benutzerquery grafisch aufbereitet, wobei zusätzliche Konzeptbegriffe aus der Ontologie mit eingefügt werden. Diese zusätzlichen Konzeptbegriffe informieren den Benutzer über zusätzliche Dinge, die mit seiner Query assoziiert werden können.

Wegen der im Forschungs- und Entwicklungsbereich der Automobilindustrie weit verbreiteten ASAM-ODS Datenarchive, besitzt die Suchmaschine auch einen Agenten, der Ergebnisse aus Prüfstandsversuchen recherchieren kann. Die internen Spezifika des ASAM-ODS Datenarchivs muss der Benutzer nicht mehr kennen.

## **Acknowledgement**

Die Entwicklung des hier vorgestellten Prototyps der Suchmaschine wurde gefördert durch die Europäische Kommission unter dem Kennzeichen #IST-2001-34417, Semantic Web-based Information Management and Knowledge Sharing for Innovative Product Design and Engineering (the WIDE project). Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

## **Literatur**

- /1/ ASAM-ODS Spezifikation  
[http://www.asam.net/03\\_standards\\_07.php?memberlogin=](http://www.asam.net/03_standards_07.php?memberlogin=)
- /2/ Resource Description Framework (RDF), W3C Semantic Web Activity, Technology and Society Domain, <http://www.w3.org/RDF/>