

# Trend Web 3.0 in der Öffentlichen Verwaltung – Semantik, Text-Mining & Co

**CeBIT 2010**

## Es geht um:

- Charakteristika von Web 3.0
  - Z.B. Semantische Suche, Text-Mining und Tagging
- Bezug zur Öffentlichen Verwaltung, Beispiele
- Anwendung im Destatis-Kundenmanagement
- Ausblick





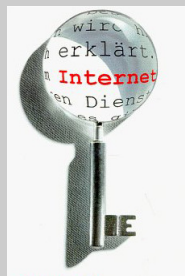
# Inhalte sind mehr als reichlich vorhanden, aber:

## Wie findet man was?

- Suchmaschinen sind „Rettung“

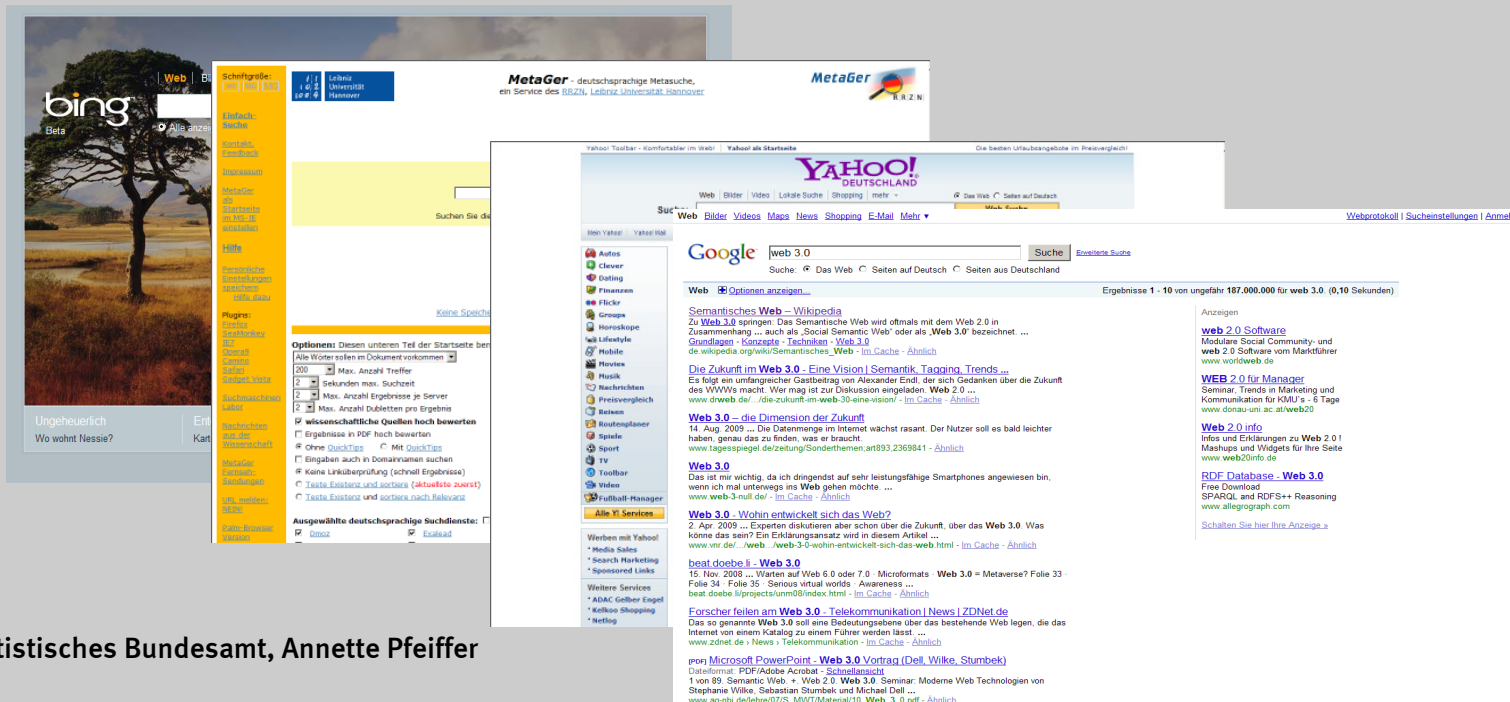
## Aber:

- Das Web wächst schneller als die Suchmaschinen mit der derzeitigen Technik indizieren können
- Heterogenität: - Viele Webseiten werden häufig aktualisiert  
- Unterschiedliche „Dokumenttypen“ (z.B. Text, Bild, Ton, Video)  
- Große Mengen unstrukturierter Daten  
- Bezug zwischen und innerhalb der Dokumente fehlt



# Heute verbreitete Suchmaschinen

- Ergebnisdarstellung als Suchlisten
- Suche nach festen Begriffen, Zeichenfolgen (einschl. Page Ranking)
- Keine Interpretation, keine Zusammenhänge
- Basis sind unstrukturierte Daten
- Wirtschaftliche Interessen (Werbung, AdWords, usw.)



# Suchbeispiel

Web Bilder Videos Maps News Shopping E-Mail Mehr ▾

Google  Suche [Erweiterte Suche](#)

Suche: ☉ Das Web ☉ Seiten auf Deutsch ☉ Seiten aus Deutschland

Web [Optionen anzeigen...](#) Ergebnisse 1 - 10 von ungefähr 882.000.000 für **auto**. (0,12 Sekunden)

<p><a href="#">mobile.de: Auto kaufen</a> www.mobile.de Deutschlands größter Fahrzeugmarkt für Neu- und Gebrauchtwagen.</p> <p><a href="#">Autos bei AutoScout24</a> www.autoscout24.de/auto Rund 1,8 Mio. Angebote einfach per E-Mail anfragen!</p> <p><a href="#">Auto.de   Günstige Autos - Gebrauchtwagen, Jahreswagen - Neuwagen ...</a> Automarkt für Gebrauchte und Neufahrzeuge, Magazin mit Autonachrichten, Fahrberichten und Autotests, Community mit Tipps und Tricks im Forum, ... <a href="#">Neuwagenkonfigurator</a> - <a href="#">Magazin</a> - <a href="#">Inserate</a> - <a href="#">Schwacke-Bewertung</a> www.auto.de/ - <a href="#">Im Cache</a> - <a href="#">Ähnlich</a></p>	<p>Anzeigen</p> <p><a href="#">Opel Autos</a> Finden Sie für jeden Bedarf und Anspruch das passende <b>Auto</b>. www.opel.de</p> <p><a href="#">Saab Automobile</a> Finden Sie Ihre Saab Modelle und entdecken Sie Fahrspaß pur. www.saab.de</p>
--	--

Suche: Auto



882.000.000 Treffer

Web Bilder Videos Maps News Shopping E-Mail Mehr ▾

Google  Suche [Erweiterte Suche](#)

Suche: ☉ Das Web ☉ Seiten auf Deutsch ☉ Seiten aus Deutschland

Web [Optionen anzeigen...](#) Ergebnisse 1 - 10 von ungefähr 13.200.000 für **pkw**. (0,12 Sekunden)

<p><a href="#">PKW kaufen bei mobile.de</a> www.mobile.de Deutschlands größter Fahrzeugmarkt- Kaufen oder für 0 € inserieren!</p> <p><a href="#">Top Neuwagen Angebote</a> www.opel.de/neuwagen Alle Opel Modelle besonders günstig kaufen, leasen oder finanzieren.</p> <p><a href="#">Pkw</a> pkw.autoscout24.de AutoScout24: Deutschlands bekanntester Automarkt!</p>	<p>Anzeigen</p> <p><a href="#">Volkswagen All-Inclusive</a> Gut ausgestattet mit Preisvorteil.* Buchen Sie das "All-Inklusiv-Paket" www.volkswagen.de/Angebot</p> <p><a href="#">SEAT Auto</a></p>
--	--

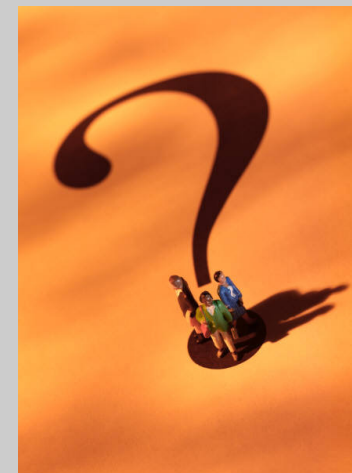
Suche: PKW



13.200.000 Treffer

# Suchergebnisse sind mehr als reichlich vorhanden, aber:

- Wie findet man verwandte Inhalte?
- Wie findet man Beziehungen zwischen Daten?
- Wie erkennt man Strukturen?
- Wie kann man Strukturen schaffen?
- Wie kann man Fragen stellen und Antworten erhalten?



# Semantic Web

- Semantik ist grundsätzlich Bedeutungslehre
- Semantic Web soll Webinhalte strukturieren, besser auffindbar machen und miteinander verknüpfen.



**Voraussetzung: Maschinenlesbarkeit**



**Damit zwischen unterschiedlichen Datensätzen maschinell Zusammenhänge herstellbar sind, müssen diese mit zusätzlichen Informationen (Annotationen) ausgestattet werden.**



**Die Annotation geschieht unter Einsatz von festgelegten Vokabularien und Ontologien, beispielsweise mittels RDF oder OWL.**

**=> Wird von speziellen Suchmaschinen verstanden**

## Neuere Suchverfahren (Beispiele)

### ■ Semantische Annotation

Daten werden mit Metainformationen ausgezeichnet

Friedrich Schiller lehrt Wilhelm Tell das Fußballspielen

<Name>Friedrich Schiller</Name> lehrt <Name>Wilhelm Tell</Name>  
 Friedrich Schiller = Name, Autor  
 <Autor>Friedrich Schiller</Autor> lehrt <Schauspiel>Wilhelm Tell</Schauspiel>  
 Wilhelm Tell = Name, Schauspiel

### ■ Text-Mining

- Erkennung durch Analyse großer Textmengen
- Sinnverwandtschaften von Begriffen und Beziehungen in bzw. zwischen Texten
  - PKW
  - Auto
  - Fahrzeug

# Was hat das mit der öffentlichen Verwaltung zu tun?

- Mittlerweile gibt es viele webbasierte Systeme in der Verwaltung:
  - Internetangebote
  - Intranet
  - E-Government: Internet als bevorzugter Kommunikations- und Vertriebskanal für bedarfsgerechte Verwaltungsdienstleistungen
  - 115-Wissensdatenbank
  
- Ziel: Informationsangebot für die breite Öffentlichkeit schaffen

# Beispiele

- Mikroformate und XML

D115-Wissensdatenbank



Halle 9,  
Stand B 60

- Topic Maps

Umweltportal Deutschland



Halle 8,  
Stand C 30

- Tagging + Text Mining

Destatis-Kundenmanagement



Halle 9,  
Stand C 48

# Tagging und Text-Mining im Statistischen Bundesamt

**Wir nutzen das webbasierte Destatis-Kundenmanagementsystem:**

- Hausweite, zentrale, transparente und einheitliche Kundendatenbank.
- Direkter Zugang vom Kontaktformular in das Kundenmanagementsystem.



Anrede\*  Herr  Frau Titel

Vorname  Nachname\*

Institution/Firma

Sie gehören folgender Kundengruppe an:\*

Straße, Hausnummer

PLZ, Ort

Land

E-Mail-Adresse\*

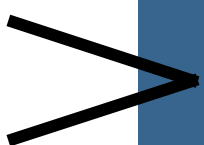
Telefon

Telefax

Ich benötige Informationen zu Deutschland und/oder seinen Regionen aus der deutschen amtlichen Statistik.  
 Ich benötige Informationen aus der amtlichen europäischen bzw. internationalen Statistik (Quellen: EUROSTAT, UN, OECD, Weltbank, IWF, u. a.).

Anfrage\*

**Der semantikgestützte Workflow identifiziert das Anliegen des Kunden und steuert die Anfrage direkt in den zuständigen Bereich innerhalb des Kundenmanagementsystems**



Statistisches Bundesamt Kundenmanagement Kundenakte

**Kundenakte** | Erweiterte Kundenakte | Verteilermanagement + Kundenprofil

**Adressinformationen**

Anrede*	Frau	Titel	
Vorname	Daniela	Nachname*	Hartmann
Abkürzung	Destatis		
Firma	Statistisches Bundesamt	Straße/Postfach	Gustav-Stresemann-Ring 11
PLZ	65180	Ort	Wiesbaden
Land*	Italien	Ländervorwahl	00960
Telefon	061175-4495	Mobil	
Fax	03018106444495	Homepage	www.destatis.de
E-Mail	daniela.hartmann@destatis.de	2. E-Mail	
Nutzerkategorie*	E-3 Statistisches Bundesamt	EU-List.kltnr.	
Rechnungssumme	0,00	Kassenzeichen	
Erstellt von	Webformulare	Export Haushalt	20.07.2009
Erstellt am	28.05.2009	Verantwortlich	- Alle -
Letzter Kontakt	05.11.2009	geschützt	<input type="checkbox"/>
		Sonderstatus	<input type="checkbox"/>
		Werbung	<input type="checkbox"/>

**Vorgänge / Vorgänge**

Vorgangs-Nr.	Eingangsdatum	Auskunftsbereich	Betreff*	Rechnungsbetrag
18057	04.11.2009 15:21		Bruttoinlandsprodukt	0,00

# Kundeninformationen nutzen

Im Destatis-Kundenmanagementsystem sind über 150.000 Anfragetexte der Kunden gespeichert

Analysen

Semantikgestützter Workflow

Haben Sie Daten zum *Mittelstand in Deutschland*? Wie hoch ist das *Wirtschaftswachstum*? Gibt es einen *Geburtenrückgang*? Wie ist die *Kaufkraft in Berlin*? Wie viele.....

.....

....

Was wird wie gefragt?

Inhalte  
Was wollen unsere Kunden?

Sprache  
Welche Sprache sprechen unsere Kunden?

Automatische und direkte Steuerung der neuen Kundenanfragen in den zuständigen Bereich innerhalb des Destatis-Kundenmanagementsystems

# ... und so funktioniert der semantikgestützte Workflow

Sehr geehrter Bundeswahlleiter,  
zum Ergebnis der Bundestagswahl habe ich folgende Frage:  
Bis zu welchem Listenplatz zogen die Bewerber aus den Landeslisten NRW  
(CDU/SPD/FDP/Grüne/Linke) in den deutschen Bundestag ein?

1. **Entfernen von Stoppwörtern**  
Ein Stoppwort ist ein inhaltsleeres Wort, also ein Term, der nichts zur Semantik eines Dokumentes beiträgt. → Stoppwortliste
2. **Gewichtungsmatrix**  
Die noch bestehenden Begriffe durchlaufen eine Gewichtungsmatrix.  
Die Gewichtungsmatrix enthält Begriffe und Begriffskombinationen, die prozentual auf die verschiedenen Terminplaner (Referate) verteilt werden. → Gewichtung

➤ Anfrage wird automatisch in den zuständigen Bereich weitergeleitet !  
Im Beispielfall an den Bundeswahlleiter.

Es genügen 124  
Begriffe, um  
Anfragen für den  
Bundeswahlleiter  
zu identifizieren.

**Wortstamm**

**Gewichtung**

Abgeordnetensitz	1
Briefwahl	1
BT Wahl	1
Bundesrat	0,1
Bundestag	0,3
Bundestagswahl	1
Bundestagswahl	1
Bundesversammlung	0,3
Bundesvorsitz	0,1
Bundeswahl	1
CDU	0,3
CSU	0,3
DieLinke	0,3
Direktkandidat	1
Direktmandat	1
Direktwahl	1
Erststimme	1
ersthwahl	1
europaparlament	0,4
Europawahl	1
FDP	0,3
Gemeindeordnung	1
Gemeinderatswahl	1
Gemeindewahl	1
Gesamtsitze	1
Gesamtwahl	1
grüne	0,3
Hare Niemeyer	1
Heimatwahlkreis	1
Infopaket Partei	1
Jungwahl	1
Kandidatenaufstellung	1
Kandidatenliste	1
Kandidierend	1
Kleinpartei	1
Kommunalwahl	1
Kreiswahlleiter	1
Länderparlament	0,1
Landesliste	1
Landeswahl	1

Landratswahl	1
Landtag	0,1
Landtagskandidat	1
Landtagswahl	1
linksextrem	0,1
linksradikal	0,1
Listenmandat	1
Listenplatz	1
Mandatsvert	1
Nachfolgepartei	1
Nachwahl	1
Nationalratswahl	1
Nichtwahl	1
NPD	0,1
Parlamentwahl	1
Partei	0,1
Parteiaus	1
Partei Bibeltreuer Christen	0,3
Partei übertritt	1
Parteienfragment	1
Parteiengesch	1
Parteiengesetz	1
Parteiengruendung	1
Parteiengruendung	1
Parteienpopularität	1
Parteienprogramm	1
Parteifinanzierung	1
Parteigrundlage	1
Parteimitglied	1
Parteiengruendung	1
parteiprogramm	1
Parteistimm	1
Parteiunterlag	1
Parteizugehörig	1
Randpartei	1
Rechtsextrem	0,1
Rechtsradikal	0,1
Regierungsmehrheit	1
Reichstagswahl	1
Republikaner	0,1

RRP	0,1
Satzung	0,1
Schill Partei	0,3
Sitzeverteilung	1
Sitzverteilung	1
Sozialdemokratische Partei	0,3
SPD	0,3
Stimmabgabe	1
Stimmbezirk	1
Stimmkreis	1
Stimmverteilung	1
Stimmzettel	1
Überhangmandat	1
vepd	0,1
Volkspartei	1
Wahlervers	1
Wahlbe	1
Wahlbogen	1
Wahlbrief	1
Wahldat	1
Wahlen	1
Wahler	1
Wahlg	1
Wahlhelf	1
Wahlkampf	1
Wahlkreis	1
Wahlhist	1
Wahllokal	1
Wahlmanipulation	1
Wahlmaschin	1
Wahlp	1
Wahlrecht	1
Wahlresultat	1
Wahlschein	1
Wahlstatistik	1
Wahlssystem	1
Wahltag	1
wahlte	1
Wahlver	1
Wahlvor	1
Wahlzettel	1
Wechselwahl	1
WRMM	0,3
Zweitstimme	1

# Semantikgestützter Workflow


## Vorteile:

- Automatisierte Verteilung von Kundenkontakten
- Herausfiltern von mehrfach gestellten Anfragen
- Kein Spam mehr
- Nur noch hausweit eine Bearbeitung



**Bisher einmalig in der öffentlichen Verwaltung.**

## Was wir erreichen wollen

- Zunächst Fragen identifizieren,  
 dann automatisiert webbasierte Antworten liefern
- Optimierte Strukturierung von Themen
- Internetsuche optimieren
- Internetangebot flexibel an sich stetig ändernde Kundenwünsche anpassen
- Begrifflichkeiten der Kundensprache anpassen

# Neugierig geworden? Sie möchten weitere Informationen?

Sie finden das Statistische Bundesamt  
auf der

**CeBIT 2010**  
**in Halle 9**  
**Stand C 48**



# VIELEN DANK FÜR IHRE AUFMERKSAMKEIT!



**Annette Pfeiffer**

**[annette.pfeiffer@destatis.de](mailto:annette.pfeiffer@destatis.de)**

**Statistisches Bundesamt, Wiesbaden**