



Corporate Semantic Search - 3

Suchen nach Komplexen Zusammenhängen und Personalisierung im Unternehmenskontext

Olga Streibel, Radoslaw Oldakowski

Inhalt

- Semantische Suche:
 - Suche im Web
 - Semantische Suche
 - Aspekte der Suche im Unternehmenskontext
- Suche nach komplexen Zusammenhängen:
 - Suche der Zukunft
 - Suche in nicht-semantischen Daten
 - Learning and Correlation Ansatz
- Personalisierung:
 - Einführung
 - Nutzer und Anforderungsanalyse

Suche im Web

- Stichwortbasiert
 - google.com, msn.com, web.de, yahoo.com, ...

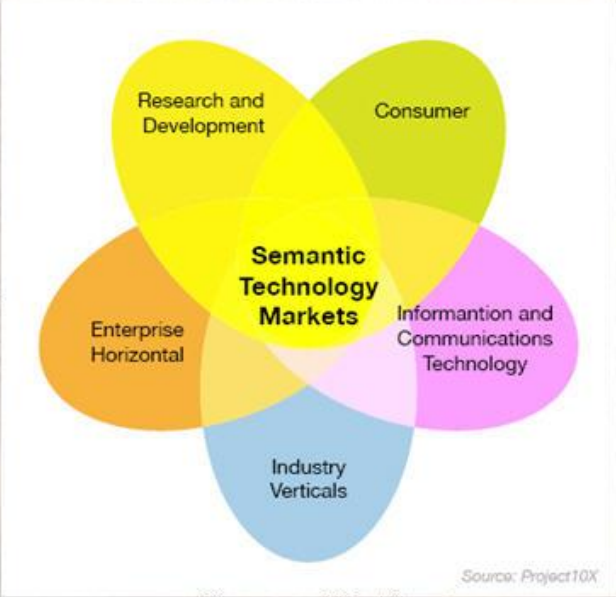
- Social search
 - answers.yahoo.com, wer-weiss-was.de
 - twitter.com (social networking und microblogging),
 - ...

- Semantische/NLP Suche
 - Cognition, Hakia, Powerset, TrueKnowledge, ...

Suche im Web: semantisch

- *cognition.com*

Where Semantics are Applied



Cognition is focused on licensing in these markets

Source: Project10X
(Source: www.Project10x.com)

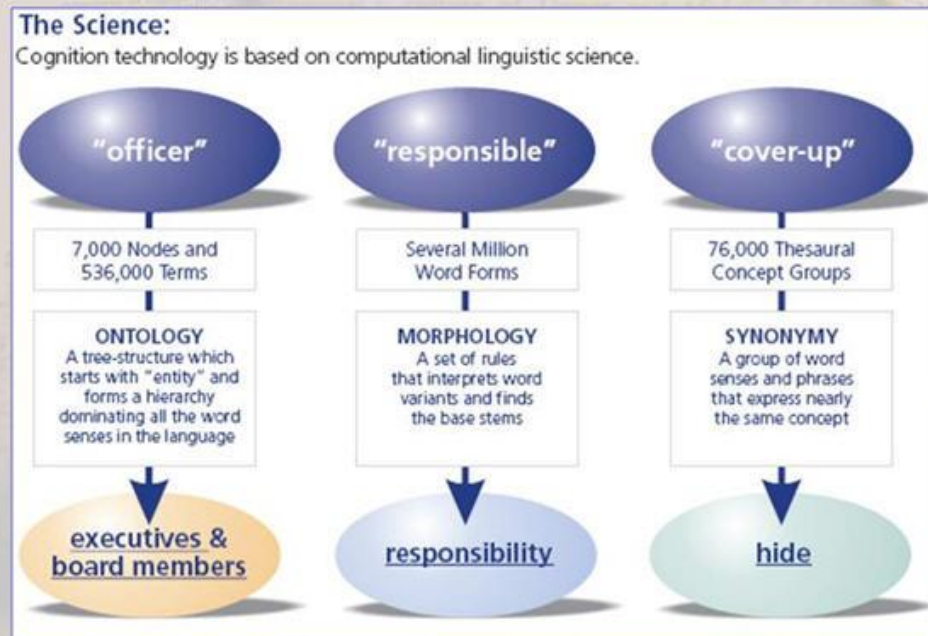
COGNITION
GIVING TECHNOLOGIES NEW MEANING™

June 2008 © 2008 Cognition Technologies, Inc. 3

Suche im Web: semantisch

- *cognition.com*

Cognition: The Science of Understanding



COGNITION
GIVING TECHNOLOGIES NEW MEANING™


June 2008


© 2008 Cognition Technologies, Inc.

5


Suche im Web: semantisch

- *hakia.com*

BETA



 See below the Semantics gallery by hakia. What else do you want to know about Semantics?

Profile on Academia

[Semantics - Wikipedia, the free encyclopedia](#) 


Semantics (Greek *semantikos*, giving signs, significant, symptomatic, from *sema*, sign) refers to the aspects of meaning that are expressed in a language, code, or other form of representation.

<http://en.wikipedia.org/wiki/Semantics>

[Semantics](#) 


A perennial problem in semantics is the delineation of its subject matter. The term meaning can be used in a variety of ways, and only some of these correspond to the usual understanding of the scope of linguistic or computational semantics.

<http://cslu.cse.ogi.edu/HLTsurvey/ch3node7.html>

[Semantic Studios | Publications | Semantics](#) 

Semantics is Peter Morville's column about information **architecture** and strategy.

<http://semanticstudios.com/publications/semantics/>








en.wikipedia.org

On This Page

- [Profile on Academia](#)
- [Image Search](#)
- [Leaders of the Field](#)
- [Founders and Innovators](#)
- [Leading Academic Programs](#)
- [Great Books and Works](#)
- [Journals and Magazines](#)
- [Textbooks](#)
- [Resources, Archives, and Organizations](#)
- [Bibliography](#)

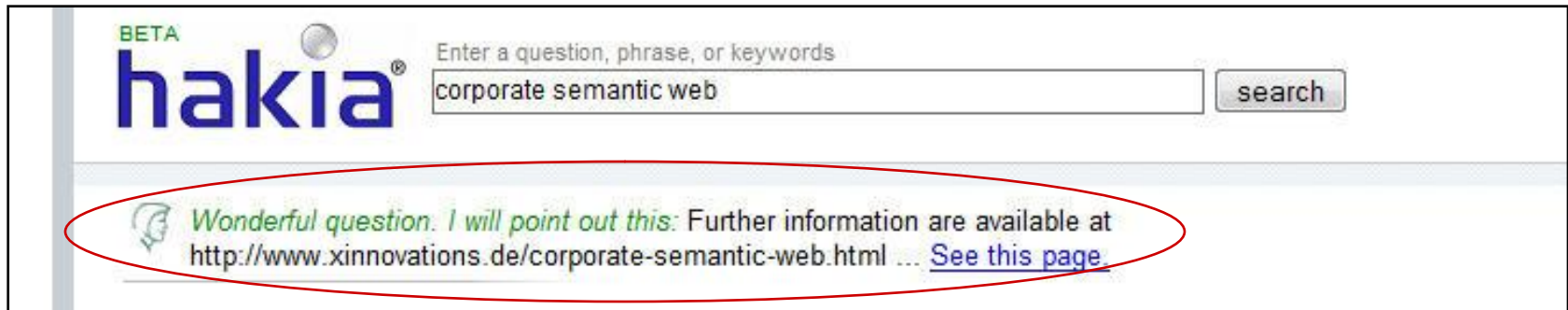
Image Search

images

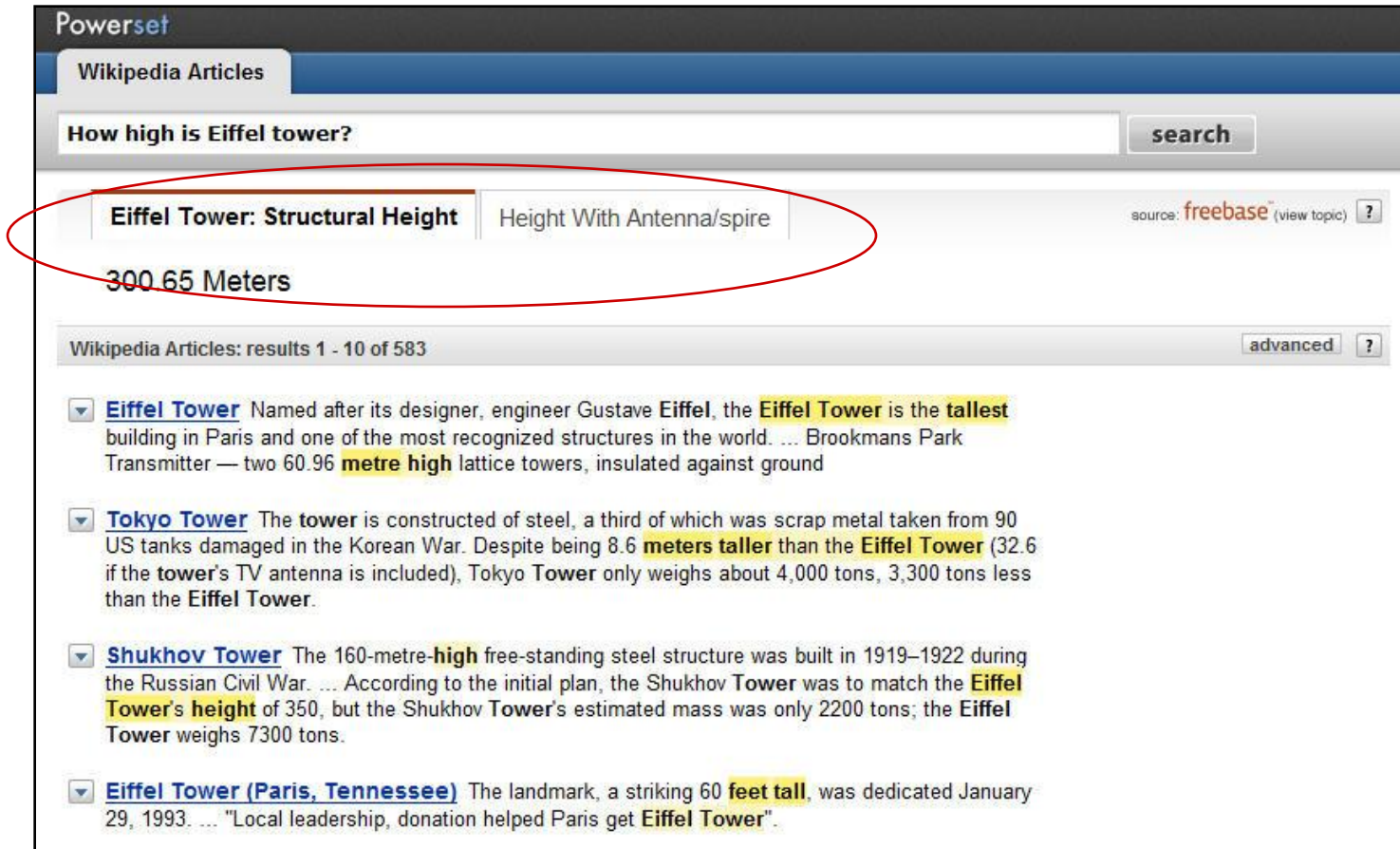
Suche im Web: semantisch

- *hakia.com*
 - zutreffende Vorschläge



Suche im Web

- *powerset.com*
 - Antwort



Powerset

Wikipedia Articles

How high is Eiffel tower?

Eiffel Tower: Structural Height | Height With Antenna/spire source: **freebase**™ (view topic) ?

300.65 Meters

Wikipedia Articles: results 1 - 10 of 583 ?

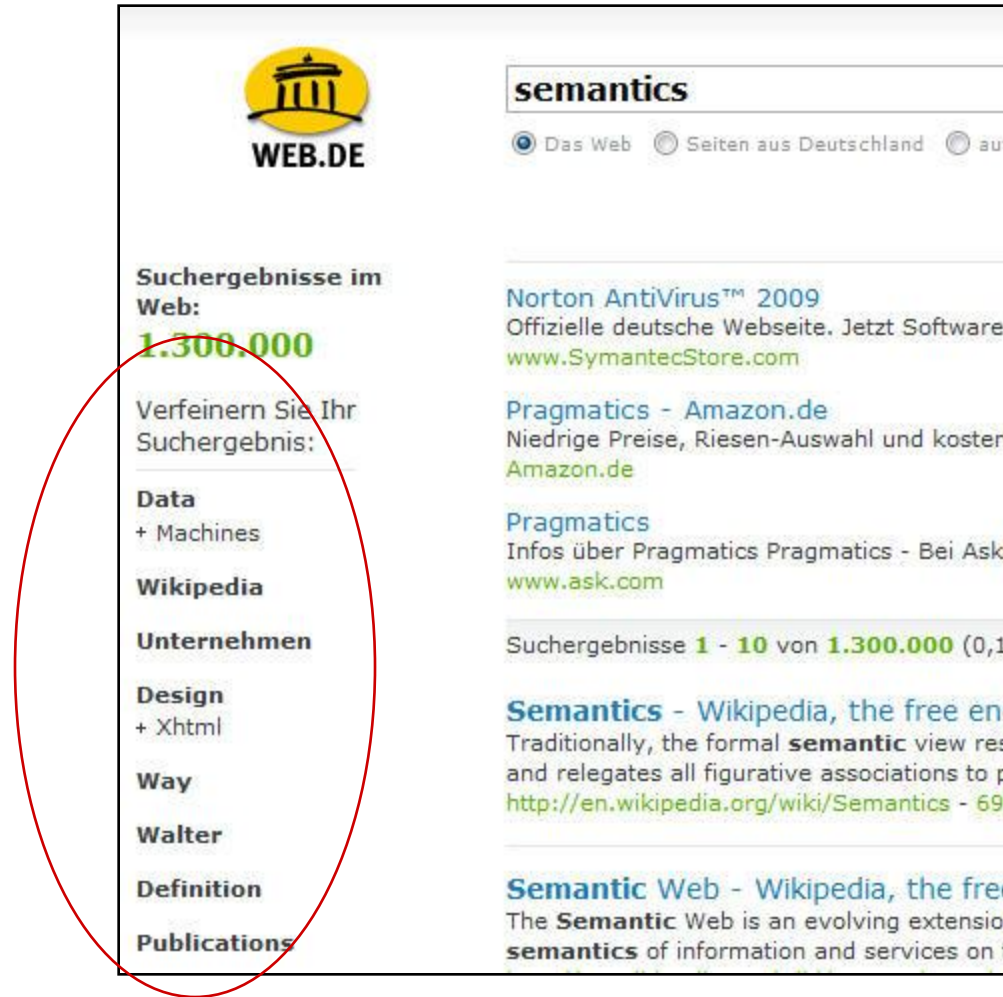
- ▼ **Eiffel Tower** Named after its designer, engineer Gustave **Eiffel**, the **Eiffel Tower** is the **tallest** building in Paris and one of the most recognized structures in the world. ... Brookmans Park Transmitter — two 60.96 **metre high** lattice towers, insulated against ground
- ▼ **Tokyo Tower** The **tower** is constructed of steel, a third of which was scrap metal taken from 90 US tanks damaged in the Korean War. Despite being 8.6 **meters taller** than the **Eiffel Tower** (32.6 if the **tower's** TV antenna is included), Tokyo **Tower** only weighs about 4,000 tons, 3,300 tons less than the **Eiffel Tower**.
- ▼ **Shukhov Tower** The 160-metre-**high** free-standing steel structure was built in 1919–1922 during the Russian Civil War. ... According to the initial plan, the Shukhov **Tower** was to match the **Eiffel Tower's height** of 350, but the Shukhov **Tower's** estimated mass was only 2200 tons; the **Eiffel Tower** weighs 7300 tons.
- ▼ **Eiffel Tower (Paris, Tennessee)** The landmark, a striking 60 **feet tall**, was dedicated January 29, 1993. ... "Local leadership, donation helped Paris get **Eiffel Tower**".

Suche im Web: Visualisierung

- Visualisierungsaspekte:
 - Cluster: *web.de, clusty.com*
 - Vorschau: *exalead.com*
 - Graphbasierte Tag-Clouds: *quintura.com*

Suche im Web: Visualisierung

- *web.de*



WEB.DE

semantics

Das Web Seiten aus Deutschland auf

Suchergebnisse im Web:
1.300.000

Verfeinern Sie Ihr Suchergebnis:

- Data**
+ Machines
- Wikipedia**
- Unternehmen**
- Design**
+ Xhtml
- Way**
- Walter**
- Definition**
- Publications**

Norton AntiVirus™ 2009
Offizielle deutsche Webseite. Jetzt Software
www.SymantecStore.com

Pragmatics - Amazon.de
Niedrige Preise, Riesen-Auswahl und kosten
Amazon.de

Pragmatics
Infos über Pragmatics Pragmatics - Bei Ask
www.ask.com

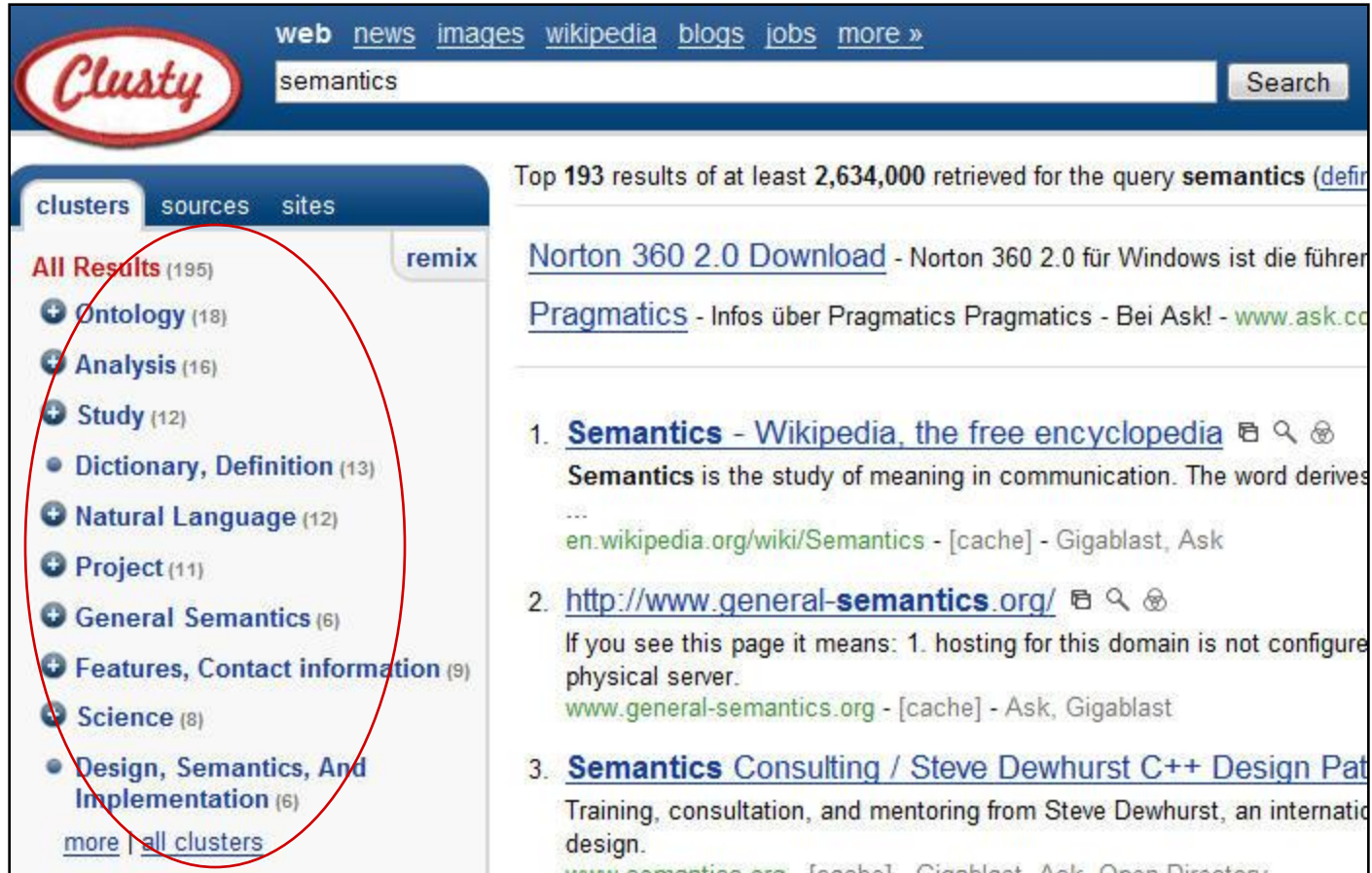
Suchergebnisse **1 - 10** von **1.300.000** (0,1

Semantics - Wikipedia, the free enc
Traditionally, the formal **semantic** view res
and relegates all figurative associations to p
<http://en.wikipedia.org/wiki/Semantics> - 69%

Semantic Web - Wikipedia, the free
The **Semantic** Web is an evolving extension
semantics of information and services on t

Suche im Web: Visualisierung

- *clusty.com*



The screenshot shows the Clusty search interface. At the top, there is a navigation bar with links for 'web', 'news', 'images', 'wikipedia', 'blogs', 'jobs', and 'more'. A search bar contains the query 'semantics' and a 'Search' button. Below the search bar, there are tabs for 'clusters', 'sources', and 'sites'. The 'clusters' tab is active, showing a list of clusters under 'All Results (195)'. A red circle highlights this list, which includes:


- Ontology (18)
- Analysis (16)
- Study (12)
- Dictionary, Definition (13)
- Natural Language (12)
- Project (11)
- General Semantics (6)
- Features, Contact information (9)
- Science (8)
- Design, Semantics, And Implementation (6)

 At the bottom of the cluster list are links for 'more' and 'all clusters'. To the right of the clusters, there is a 'remix' button. The main content area displays 'Top 193 results of at least 2,634,000 retrieved for the query semantics'. The first few results are:

- [Norton 360 2.0 Download](#) - Norton 360 2.0 für Windows ist die führende...
- [Pragmatics](#) - Infos über Pragmatics Pragmatics - Bei Ask! - www.ask.com
- [Semantics - Wikipedia, the free encyclopedia](#) - **Semantics** is the study of meaning in communication. The word derives from... en.wikipedia.org/wiki/Semantics - [cache] - Gigablast, Ask
- <http://www.general-semantics.org/> - If you see this page it means: 1. hosting for this domain is not configured on a physical server. www.general-semantics.org - [cache] - Ask, Gigablast
- [Semantics Consulting / Steve Dewhurst C++ Design Patterns](#) - Training, consultation, and mentoring from Steve Dewhurst, an international design. www.semantics.org - [cache] - Gigablast, Ask, Open Directory

Suche im Web: Visualisierung

- *quintura.com*

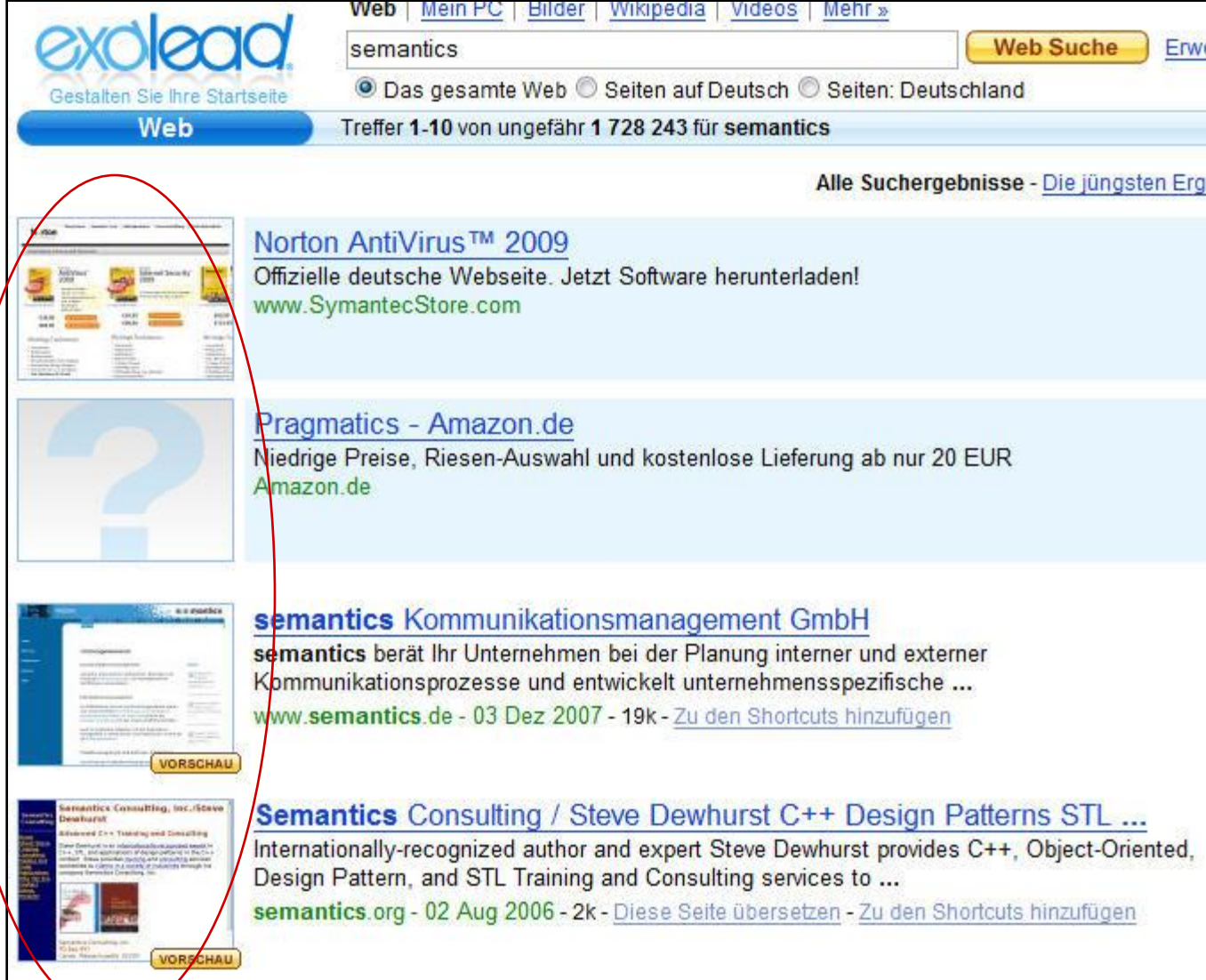


The screenshot shows the Quintura search engine interface. The search bar contains the text "corporate semantic web". Below the search bar, there are navigation buttons for "Web", "Images", "Video", and "Amazon". A red circle highlights a word cloud visualization of search results. The words in the cloud include: investment, radoslaw, ontology, solutions, enterprise, technology, web, corporate, semantic, semantic web, approach, applications, requirements, resources, milestone report, cache, manage, and underpants. To the right of the word cloud, there is a list of search results:

1. [Corporate Semantic Web - Home](#)
Further information are available at <http://www.x>
Semantic Web Milestone Report ...
<http://www.corporate-semantic-web.de/>, 14Kb
2. [A Multi-Agent Platform for a Corporate Seman](#)
... to manage a **corporate** memory in the form of
Corporate Semantic Web (2002) [6 citations]
<http://citeseerx.ist.psu.edu/...ummary?doi=10.1>
3. [Corporate Semantic Web - Radoslaw Oldako](#)
... **Corporate Semantic Web** Milestone Report
Corporate Semantic Web : Requirements and
<http://www.corporate-semantic-web.de/oldakows>
4. [Semantic Web and Multi-Agents Approach to](#)
... and organise in order to turn them into active
semantic web engine – Corby, Faron ...
<http://citeseerx.ist.psu.edu/...mary?doi=10.1.1>
5. [Use Case: WEASEL, Vodafone R&D Corporat](#)
Semantic Web Use Cases and Case Studies.
Benefits of Using **Semantic Web** Technology .
<http://www.w3.org/.../sweo/public/UseCases/Vo>
6. [Integrating external sources in a corporate sen](#)
annotated to integrate the **corporate semantic**
Philosopher Doctorate Thesis in Informatics, 7 .
http://www-2.cs.cmu.edu/...km2003/Cao_Gandd
7. [Managing enterprise applications as dynamic re](#)
... applications where we implemented **corpora**
corporate semantic web and the ...

Suche im Web: Visualisierung

- *exalead.com*



The screenshot shows the search interface of exalead.com. At the top, there are navigation links for 'Web', 'Mein PC', 'Bilder', 'Wikipedia', 'Videos', and 'Mehr'. A search bar contains the term 'semantics', and a yellow 'Web Suche' button is to its right. Below the search bar, there are radio buttons for 'Das gesamte Web' (selected), 'Seiten auf Deutsch', and 'Seiten: Deutschland'. A blue bar indicates 'Treffer 1-10 von ungefähr 1 728 243 für semantics'. On the right, there is a link for 'Alle Suchergebnisse - Die jüngsten Ergebnisse'. The search results are listed in a table-like format with three columns: a thumbnail, a title, and a description. A red oval highlights the first three results. The first result is for 'Norton AntiVirus™ 2009' with a thumbnail showing the software interface. The second result is for 'Pragmatics - Amazon.de' with a large blue question mark thumbnail. The third result is for 'semantics Kommunikationsmanagement GmbH' with a thumbnail showing a document. Below this, there are two more results, each with a 'VORSCHAU' button and a thumbnail. The first of these is for 'Semantics Consulting / Steve Dewhurst C++ Design Patterns STL ...' with a thumbnail showing a book cover.





Web | [Mein PC](#) | [Bilder](#) | [Wikipedia](#) | [Videos](#) | [Mehr »](#)

semantics Web Suche [Erwe](#)

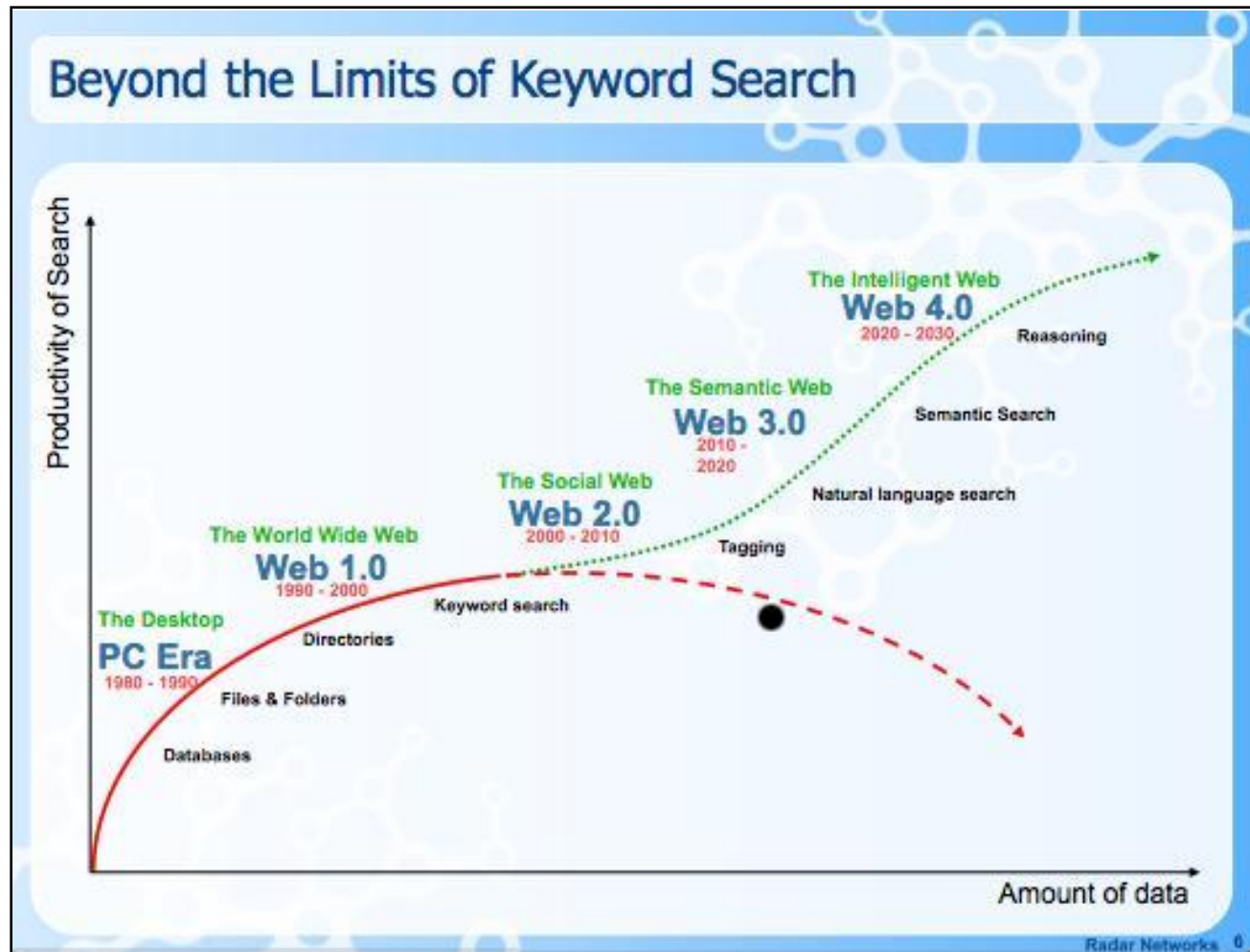
Das gesamte Web Seiten auf Deutsch Seiten: Deutschland

Web Treffer 1-10 von ungefähr 1 728 243 für semantics

Alle Suchergebnisse - [Die jüngsten Ergebnisse](#)

	Norton AntiVirus™ 2009 Offizielle deutsche Webseite. Jetzt Software herunterladen! www.SymantecStore.com
	Pragmatics - Amazon.de Niedrige Preise, Riesen-Auswahl und kostenlose Lieferung ab nur 20 EUR Amazon.de
	semantics Kommunikationsmanagement GmbH semantics berät Ihr Unternehmen bei der Planung interner und externer Kommunikationsprozesse und entwickelt unternehmensspezifische ... www.semantics.de - 03 Dez 2007 - 19k - Zu den Shortcuts hinzufügen
	Semantics Consulting / Steve Dewhurst C++ Design Patterns STL ... Internationally-recognized author and expert Steve Dewhurst provides C++, Object-Oriented, Design Pattern, and STL Training and Consulting services to ... semantics.org - 02 Aug 2006 - 2k - Diese Seite übersetzen - Zu den Shortcuts hinzufügen

Evolution der Suche im Web



Quelle: Nova Spivack, Semantic Web Talk at *The Next Web Conference*, 2008

Aspekte der Suche

Form der Suchanfrage

Qualität der
Suchergebnisse

User Involvement

Suchziel/ Suchobjekt

Suchbestände: Daten,
Texte, etc.

Visualisierungsaspekte

Semantische Suche - Klassifikation

- Klassifikationskriterien der semantischen Suchverfahren
 - Architektur
 - stand-alone vs. meta search
 - Ankopplung
 - enge vs. Lose Kopplung der Dokumente und Ontologien
 - Transparenz
 - transparent, interaktiv (recommendation), hybrid
 - User context
 - hard-coded vs. lernbasiert
 - Query Modifikation
 - manuell vs. query rewriting
 - Ontologie-Struktur (semantic reusability)
 - Ontologie-Technologie (technological reusability)

(Klassifikationskriterien nach Mangold, Ch. *A Survey and Classification of Semantic Search Approaches*, Int. J. Metadata Semant. Ontologies, 2007)

Semantische Suche – Forschungsrichtungen

- **Forschungsrichtungen** (nach Eetu Mäkelä *Survey of Semantic Search Research* 2005)
 - Modifikation traditioneller stichwortbasierter Suche mit Semantic Web Technologien
 - Query Erweiterung bzw. Einschränkung basierend auf Hintergrundwissen (Synonyme, Homonyme, Meronyme)
 - Matching der Such-Stichwörter auf Ontologiekonzepte
 - Basic concept location:
 - Instanzzuordnung zu Konzepten (Ankopplung)
 - Complex constraint queries:
 - Benutzerunterstützung in Queryformulierung
 - Problem solving:
 - Wissensbasis und Problembeschreibung

Corporate Semantic Search

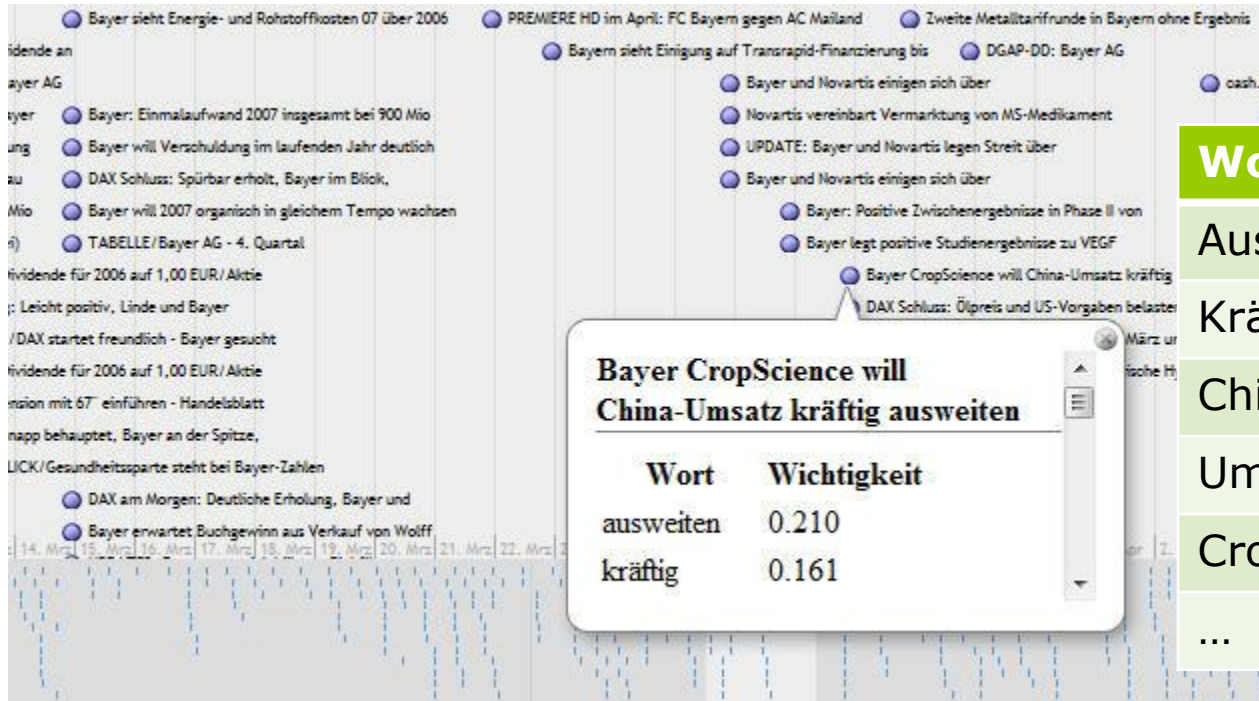
- Anwendung der semantischen Suche im Unternehmenskontext
- Aspekte von Corporate Semantic Search
 - Daten
 - Natürliche Sprache, numerische Daten, Multimedia
 - Heterogenität
 - strukturiert (z.B. Ontologien, relationale Datenbanken)
 - semi-strukturiert (z.B. Texte in XML-Format)
 - unstrukturiert (z.B. E-Mails)
 - User
 - Informationsbedarf orientiert sich an Unternehmenskontext
 - Aufgaben / Ziele
 - Suche nach Dokumenten, Menschen, Multimedia
 - Suche nach komplexen Zusammenhängen (z.B. Trends)
 - Methoden

Corporate Semantic Search

- Suche der Zukunft:
 - Alternative Suchformen: lokale, mobile, soziale Suche
 - Benutzer und Benutzersicht stehen im Vordergrund der Suche
 - Hyper Mining, Trend Mining
- Trend Mining - Eine komplexe Aufgabe:
 - Expertenwissen
 - Erkennung komplexer Zusammenhänge
 - Lernprozess
 - Korrelation hybrider Information
 - Erkennung von Trends

Suche nach komplexen Zusammenhängen in nicht-semantischen Daten

- Einfache Statistik: *(Timeline mit <http://simile.mit.edu/>)*



Wort	Wichtigkeit
Ausweiten	0.210
Kräftig	0.161
China	0.141
Umsatz	0.113
CropScience	0.111
...	...

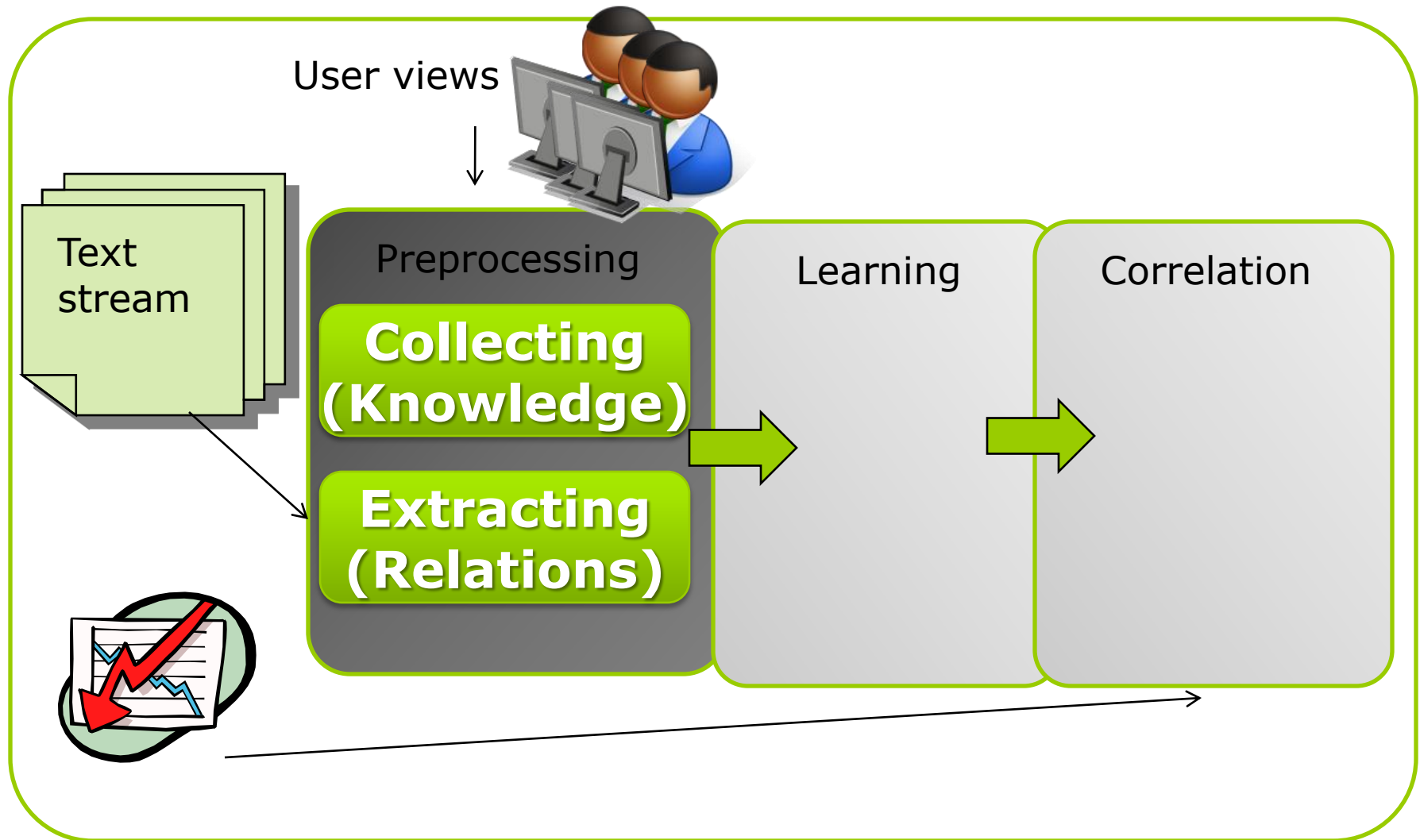
(Textkorpus dank neofonie GmbH)

(Wichtigkeit nach TFIDF berechnet)

„FRANKFURT (Dow Joes)--Bayer CropScience will in China in den kommenden Jahren kräftig wachsen. Das Unternehmen plane mittelfristig seinen Jahresumsatz in der Volksrepublik von aktuell 65 Mio EUR auf über 100 Mio EUR zu ...“

- Expertenwissen integrieren:
 - Generierung der semantischen Relationen aus Expertenfeedback (Tagging mit *Extreme Tagging Ansatz*)
(Vlad Tanasescu, Olga Streibel *Extreme Tagging: Emergent Semantics through the Tagging of Tags* ESOE bei ISWC2007)
- Suchbestände vorverarbeiten:
 - NLP-Verfahren (z.B. *Stanford Statistical Parser*)
(<http://nlp.stanford.edu/software/lex-parser.shtml>)
- Umgang mit hybrider Information:
 - Korrelation der Information aus Texten mit der Information aus numerischen Daten

Learning and Correlation Approach



Collecting:

- Informationsgewinnung und Wissensextraktion aus den Tags
- Unterstützung durch Extreme Tagging

Extracting:

- Extraktion von Triplets aus Textströmen

Benutzersicht:

- Expertensicht stets im Vordergrund



Trend Mining: Anwendungsgebiete

- Finanzmarktanalyse
- Konkurrenzbeobachtung
- Opinion Mining/Sentiment Mining
- Marktforschung

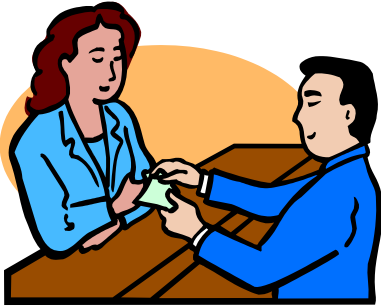
Personalisierung der Suche

- *Entwicklung von Methoden und Modellen zur Repräsentation von Informationen über Nutzer und deren Verwendung für die individuelle Anpassung der Suche*
- Adaptierbar vs. adaptiv
- Frühere Forschung (seit Mitte der 90er)
 - Adaptive Hypermedia Systeme
 - Problem: Flexibilität, Re-usability
 - Recommendations-Systeme
 - basierend auf Web-mining Technologien (Content und Nutzung)
 - Problem: Lernalgorithmen erfordern eine kritische Masse an Daten

Personalisierung der semantischen Suche

- Rolle der Semantic Web Technologien
 - Semantische Repräsentation der Informationen
 - Semantische Repräsentation der Nutzer
 - Bessere Umsetzung der Personalisierung

- Aktuelle Forschung
 - Adaptive Hypermedia
 - eLearning
 - Peer-to-peer
 - Informationsportale
 - Web Service discovery & composition
 - ...
 - Unternehmenskontext wird selten berücksichtigt



Kunden

- Anbieter personalisierter Dienste werden bevorzugt
- Wettbewerbsvorteil durch besseren Service → Kundenbindung

B2B-Partner, Investorinnen, ...

- Personalisierter Zugriff auf relevante Unternehmensdaten
→ Reduktion der Transaktionskosten, Steigerung der Produktivität



Mitarbeiterinnen

- Verbringen 15-35% ihrer Arbeitszeit mit der Suche nach Informationen (IDC Report. Enterprise Search Technology. Portals Magazine 12/2003)
- Unterschiedlicher Informationsbedarf je nach
 - Abteilung, Rolle, Erfahrung, Ziele, Kontext einer aktuellen Aufgabe

Wann macht Personalisierung Sinn?

- Große Datenmengen
 - wenn einfache Filterregeln nicht ausreichen
 - z.B. Sortierung der Dokumente nach der Projektzugehörigkeit
 - Nutzerprofil-basiertes Ranking
- Neue Informationen vorhanden
 - Vermeidung von erneuten Suchanfragen
- Unpräzise Suchanfragen
 - Nutzerprofil wird zur Eingrenzung der Suchergebnisse herangezogen
- Zeit/Aufwand-Trade-off
- Initialisierung einer Such-Session
 - Startseite mit relevanten Informationen (Bsp. Projektmanager)

- Berücksichtigung der Nutzer-Heterogenität
 - Abteilung, Rolle, Erfahrung, Ziele, Kontext der Aufgabe
- Integration von Informationsquellen
 - Unternehmensportale, Unternehmensontologien, öffentliche Websites, lokale Dateien, E-Mails, usw.
- Nachvollziehbarkeit und Kontrolle
 - Personalisierte Suchergebnisse müssen nachvollziehbar sein
 - Möglichkeit der Änderung des eigenen Suchprofils bzw. Abschaltung der Personalisierung
- Niedriges User-involvement
 - (semi-)automatische Profilgenerierung und Update
- Performance und Skalierbarkeit



Vielen Dank!

<http://www.corporate-semantic-web.de>