

# Fachtagung Enterprise Search-Systeme Einführung Unternehmensweite Suche



Zürich  
28. November 2007

Markus Tressl, Senior Consultant

Bern, Frankfurt, Hamburg, München, St. Gallen, Zug, Zürich

# Internet- vs. Intranet-Suche.



Internet- vs. Intranet-Suche.

# Intranetsuche und Enterprise Search.



- » Langfristig konvergieren diese Begriffe
- Ein Ausgangspunkt für Unternehmensinhalte

Quelle: nach Challenges in Enterprise Search, D. Hawking, 2004

Internet- vs. Intranet-Suche.

## Was ist anders im Intranet?

- » Dokumentenanzahl
  - $\sim 10^3$ - $10^5$  gegenüber  $>10^{7+x}$
- » Dokumentenart / Homogenität
  - Unstrukturiert, <10% ist HTML
- » Quellsysteme
  - CMS, DMS, Dateisysteme, BI-Systeme, CMR, DBs,...
- » Sicherheit / Berechtigungen: Rollen und Rechte
- » Inhalte sind nicht suchmaschinenoptimiert
- » Nicht primär auf Attraktivität ausgelegt
- » Kein Spam

## Internet- vs. Intranet-Suche. Suchstrategien.

- » „Known item search“
  - 60-80% der Suchanfragen zielen auf einen schon bekannten Inhalt
  - Der Benutzer erwartet diesen Inhalt ganz oben auf der Suchergebnisliste
  
- » Gelerntes Suchen in verschiedensten Systemen
  - Verzeichnisse (Telefonbuch, Datenbanken)
  - E-Mail Systeme
  - BI-Systeme

## Internet- vs. Intranet-Suche. Suchstrategien.

- » Häufig gibt es nur eine richtige Antwort
  - Im Internet steht diese auf tausenden Seiten  
„Wie heißt die Frau von Bill Gates?“  
„Wann war der Westfälische Friede?“



The screenshot shows a Google search interface. At the top right, the user's email 'janko.zehe@gmail.com' is displayed along with links for 'Web History', 'My Account', and 'Sign out'. Below this are navigation tabs for 'Web', 'Images', 'Groups', 'News', 'Products', 'Scholar', 'Desktop', and 'more'. The search bar contains the text 'ehefrau von bill gates' and a 'Search' button. To the right of the search bar are links for 'Advanced Search' and 'Preferences'. Below the search bar, there is a 'New!' notification and a link to 'View and manage your web history'. The search results section is titled 'Web' and shows 'Results 1 - 10 of about 42,600 for ehefrau von bill gates. (0.12 seconds)'. A tip suggests searching for English results only. The first result is 'Roman's Homepage - Fun - Jokes, Fun...' with a snippet: 'page-link Revenge of a user to Bill Gates (english) ... Ehefrau 1.0. Letztes Jahr hat ein Freund von mir ein Upgrade von Freundin 1.0 nach Ehefrau 1.0 ...'. The second result is 'Gates - Wikipedia' with a snippet: 'Melinda Gates, Ehefrau von Bill Gates; Reginald Ruggles Gates, Genetiker und Botaniker. die Bezeichnung mehrerer Orte in den USA: ...'.

- Im Intranet gibt es die Seite mit der Antwort nur einmal  
„Was gibt's heute zu essen?“  
„Wie werden Visitenkarten bestellt?“  
„Wie hoch ist der aktuelle Beitragssatz?“

## Internet- vs. Intranet-Suche. Rankingstrategien.

- » Internet-Rankingalgorithmen funktionieren nicht
  - PageRank, Verlinkungen zwischen Seiten
  - Verlinkungsstruktur im Intranet ist vollkommen anders
- » Größerer Stellenwert aktueller Informationen
- » Abhängigkeiten vom Standort, Sprache, Organisationseinheit
- » Einbeziehung von geschützten Inhalten

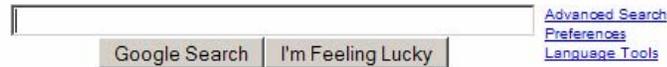
# Benutzeranforderungen.



Benutzeranforderungen.

## Die am häufigsten geäußerte Aussage:

*„Unsere Suche muss so gut sein wie Google(.com)“*



- » Dieser Wunsch drückt aus, wie der Benutzer suchen gelernt hat
  - Er gibt nur wenige Suchbegriffe ein.
  - Er benutzt die erweiterte Suche nicht.
  - Er kennt den Unterschied zur Internetsuche nicht.
  
- » Warum ist Google.com so gut?
  - Einfachheit
  - Nutzungsmessung und permanente Optimierung

# Suchmasken.

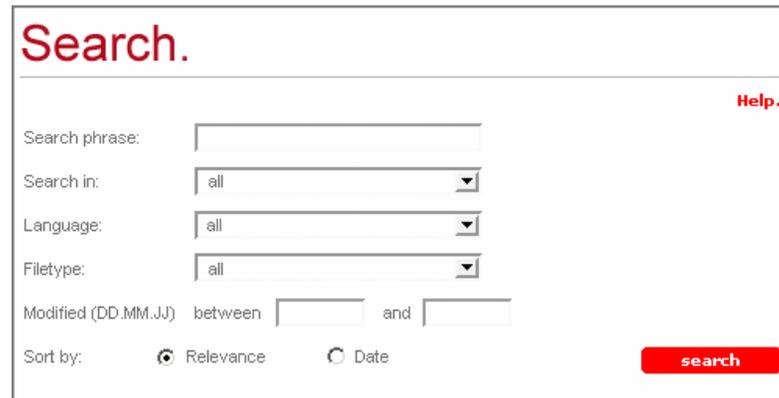


Suchmasken.

## Gestaltung und Ausprägung.

- » Suche muss von überall verfügbar sein
  - Such-Teaser im Header jeder Seite
- » Standard-Suchmaske
  - Erste Einschränkungen und Kriterien
- » Erweiterte Suchmaske
  - Detaillierte Einschränkung der Ergebnisse



# Suchmasken. „out of the box.“

**Erweiterte Suche**

Advanced Search


Suchen in   
Suchen nach

Hinweise zu Suchfunktionen finden Sie: [Tipps für die Suche.](#)

---

Lar Seite

File Perspektive

Occ Kategorie

Attribut-Auswahl

Attributname	Operator	Wert
<input type="text"/>	Enthält	<input type="text"/>
<input type="text"/>	Enthält	<input type="text"/>
<input type="text"/>	Enthält	<input type="text"/>
<input type="text"/>	Enthält	<input type="text"/>

Unterseiten aufnehmen

**Weitere Attribute**

Entsprechung  Alle  Beliebig

Powered by Google Search Appliance

Suchmasken.

## Einschränkungskriterien.

- » Must have:
  - Kollektion
  - Dateityp
  
- » Je nachdem:
  - Datum / Zeitraum
  - Autor
  - Sprache
  - Land / Standort
  - Inhaltstyp
  
- » Wichtig:
  - Den Benutzer nicht überfordern
  - Es muss transparent sein, was die Einschränkung bewirkt

# Suchergebnisseite.



# Suchergebnisseiten. „out of the box.“

## Suchergebnisse

Sortieren nach ▼ Score ▼ Name ▼ Größe

<a href="#">S</a> <a href="#">N</a> <a href="#">sg</a> <a href="#">Di</a> <a href="#">sg</a> <a href="#">zh</a> <a href="#">zh</a> <a href="#">[T</a> <a href="#">zh</a> <a href="#">ro</a> <a href="#">[T</a> <a href="#">ro</a> <a href="#">sg</a> <a href="#">sg</a> <a href="#">[T</a> <a href="#">sg</a> <a href="#">sg</a> <a href="#">[T</a> <a href="#">sg</a>	<div style="border: 1px solid #ccc; width: 100%; height: 10px; margin-bottom: 5px;"></div> <p>40% </p> <p>1944 Bytes</p>	<p><b><u>Die inxire ECM Suite</u></b></p> <p>Die inxire ECM Suite ist eine umfassende Standard-Software für das Enterprise Content Management. Sie setzt auf der Oracle Content Management-Plattform auf und nutzt aktuelle Technologien wie J2EE und XML. Alle Funktionen der inxire ECM Suite können über eine Web-Oberfläche im Browser genutzt werden. Auszeichnende Merkmale sind modulare Erweiterbarkeit, Flexibilität, Skalierbarkeit, Stabilität und Geschwindigkeit.</p>
	<div style="border: 1px solid #ccc; width: 100%; height: 10px; margin-bottom: 5px;"></div> <p>32% </p> <p>3360 Bytes</p>	<p><b><u>Glossar</u></b></p>
	<div style="border: 1px solid #ccc; width: 100%; height: 10px; margin-bottom: 5px;"></div> <p>24% </p> <p>2907 Bytes</p>	<p><b><u>Online-Hilfe</u></b></p> <p>Die inxire ECM Suite bietet umfangreiche Funktionen für ein erfolgreiches Enterprise Content Management. Die Anwendung gliedert sich in drei Funktionsbereiche: die <u>Portalansicht</u>, die <u>Bearbeitungsansicht</u> und die <u>Gruppenverwaltung</u>. Inhalte werden, wie vom Dateisystem her gewohnt, in einem hierarchischem <u>Verzeichnisbaum</u> abgelegt.</p> <p>In diesem Dokument werden die verschiedenen Bereiche der inxire ECM Suite kurz vorgestellt.</p>
	<div style="border: 1px solid #ccc; width: 100%; height: 10px; margin-bottom: 5px;"></div> <p>8% </p> <p>11877 Bytes</p>	<p><b><u>Der WYSIWYG-Editor</u></b></p> <p>Der WYSIWYG-Editor ermöglicht das interaktive Bearbeiten von Web-Artikeln. Er wird innerhalb des Artikel-Editors für einzelne Abschnitte des Textes geöffnet. Die Funktionsweise entspricht gängigen Textverarbeitungsprogrammen.</p>
	<div style="border: 1px solid #ccc; width: 100%; height: 10px; margin-bottom: 5px;"></div> <p>8% </p> <p>4635 Bytes</p>	<p><b><u>Menü Navigation - Wechseln in die Ansichten</u></b></p> <p>In diesem Dokument wird auf die Funktionen im Menü "Navigation" eingegangen.</p>

Sheet1. A, B, C, D, 1, [Referenzkarten deutsch](#), Stand: 26.02.03, 2, 3, Register, Kunde, Projekt, Bem. 4, 1, Alphabet, >> Online-Leasingberechnung auf J2EE-Basis, ...  
[sgfiler.sg.ch.namics.com/data/Administration/Backoffice%20SG/Diverses/Referenzkarten.xls](http://sgfiler.sg.ch.namics.com/data/Administration/Backoffice%20SG/Diverses/Referenzkarten.xls) - 2003-07-04 - [Text Version](#)  
[\[ More results from sgfiler.sg.ch.namics.com/data/Administration/Backoffice%20SG \]](#)

Suchergebnisseiten.

## Was gehört auf die Ergebnisliste?

- » Daumenregel: Wie bei Google.com
  - Überschrift (und zwar eine sinnvolle)
  - Snippet (mit Highlighting der Suchbegriffe)
  - URL (oder besser: klickbarer Breadcrumb-Pfad)
  - Dateityp
  - Datum
  
- » Je nachdem:
  - Dateigröße
  - Autor
  - Informationstyp
  - Kollektion, aus der der Treffer stammt

Suchergebnisseiten.

## Was gehört auf die Trefferseite?

- » Suchmaske
  - mit Wiederholung der Abfrage
- » Trefferliste
  - Blätterfunktion
  - bei annotierten und strukturierten Daten: evtl. Sortierfunktion
- » bei strukturierten und annotierten Daten: weitere Einschränkungsmöglichkeiten nach Attributen
- » Bereichseinschränkungen
- » Gruppierung nach Kategorien, Attributen

Suchergebnisseiten.

## Funktionalitäten rund um die Suche.

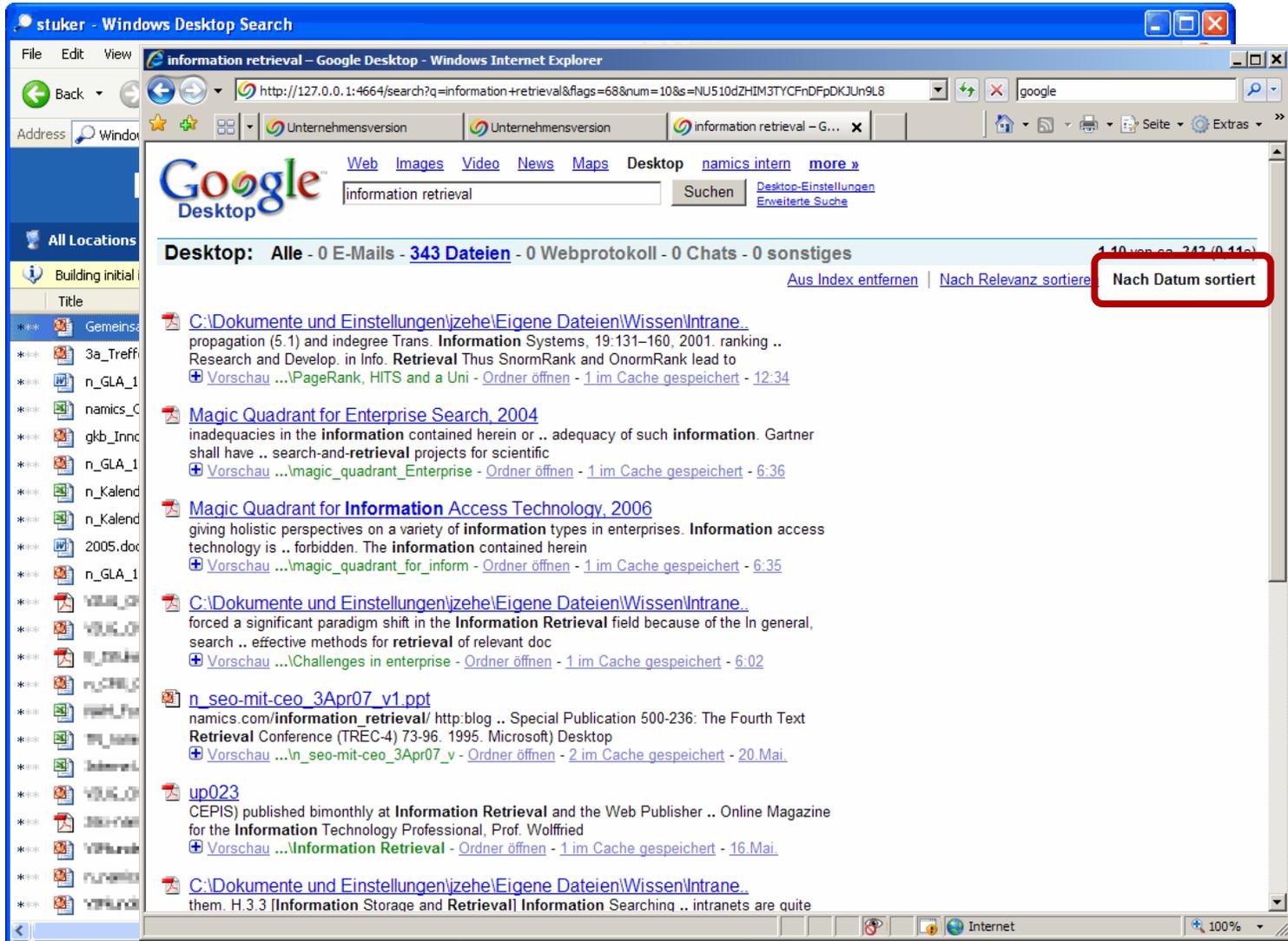
- » Bei vielen Engines im Standardset enthalten:
  - „ähnliche Seiten“
  - „Cache-Version“
  - „HTML-Ansicht“
  - „Suche speichern“
  - Automatische Zusammenfassung
  - Rechtschreibkorrektur „Meinten Sie xyz?“ → **Unverzichtbar!**
  - Synonymlisten
  
- » Eigene Features, die dem Benutzer helfen können:
  - Empfehlungen/Suchen bei der Eingabe
  - Trefferbewertungen

# Ranking.



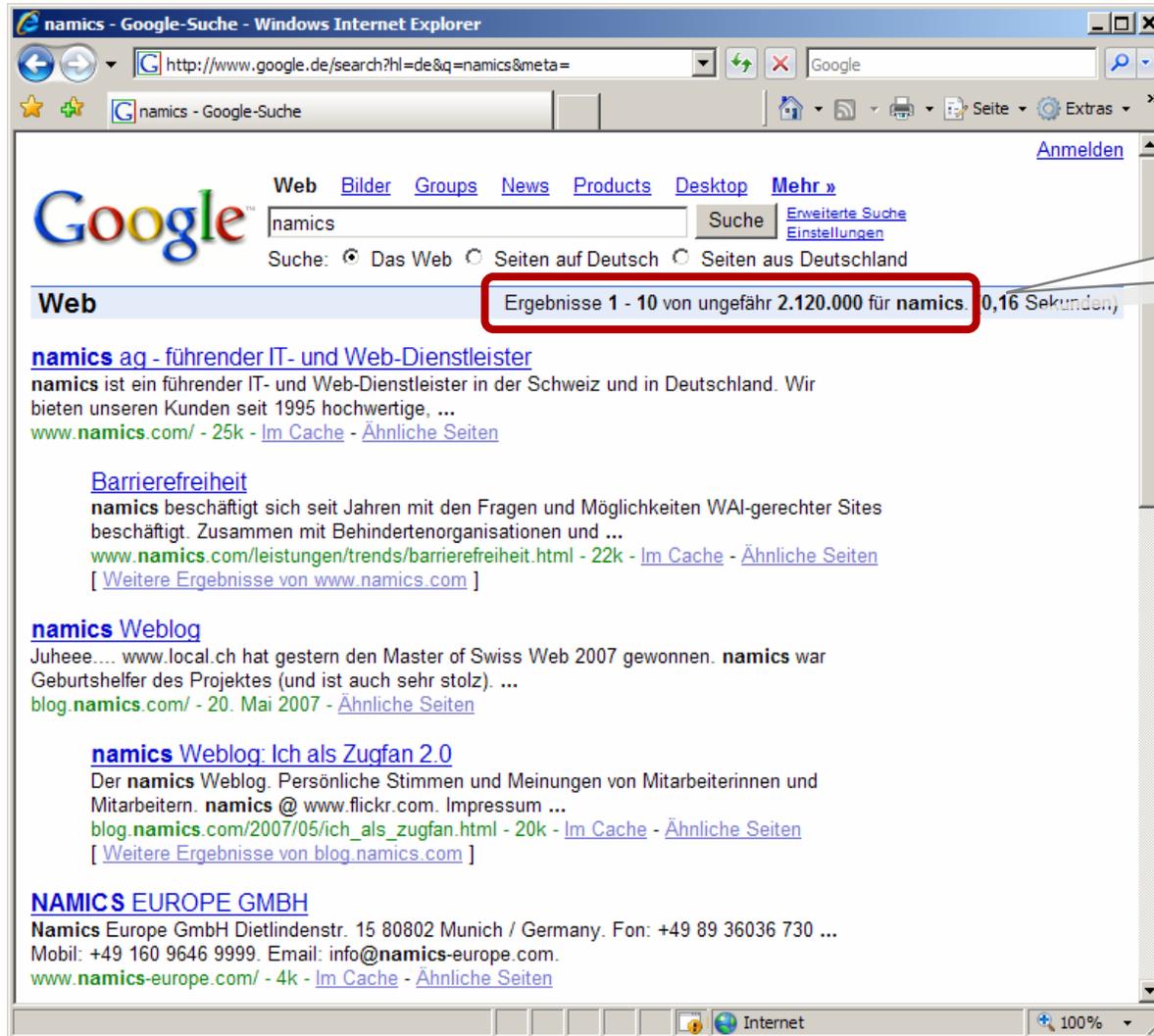
# Ranking.

## Desktop-Search: standardmäßige Sortierung nach Datum.



Ranking.

# Wie wichtig ist die Anzahl der Treffer?



namics - Google-Suche - Windows Internet Explorer

http://www.google.de/search?hl=de&q=namics&meta=

Google

namics

Suche:  Das Web  Seiten auf Deutsch  Seiten aus Deutschland

**Web** Ergebnisse 1 - 10 von ungefähr 2.120.000 für namics. 0,16 Sekunden

[namics ag - führender IT- und Web-Dienstleister](#)  
namics ist ein führender IT- und Web-Dienstleister in der Schweiz und in Deutschland. Wir bieten unseren Kunden seit 1995 hochwertige, ...  
[www.namics.com/](#) - 25k - [Im Cache](#) - [Ähnliche Seiten](#)

[Barrierefreiheit](#)  
namics beschäftigt sich seit Jahren mit den Fragen und Möglichkeiten WAI-gerechter Sites beschäftigt. Zusammen mit Behindertenorganisationen und ...  
[www.namics.com/leistungen/trends/barrierefreiheit.html](#) - 22k - [Im Cache](#) - [Ähnliche Seiten](#)  
[ [Weitere Ergebnisse von www.namics.com](#) ]

[namics Weblog](#)  
Juheee.... [www.local.ch](#) hat gestern den Master of Swiss Web 2007 gewonnen. namics war Geburtshelfer des Projektes (und ist auch sehr stolz). ...  
[blog.namics.com/](#) - 20. Mai 2007 - [Ähnliche Seiten](#)

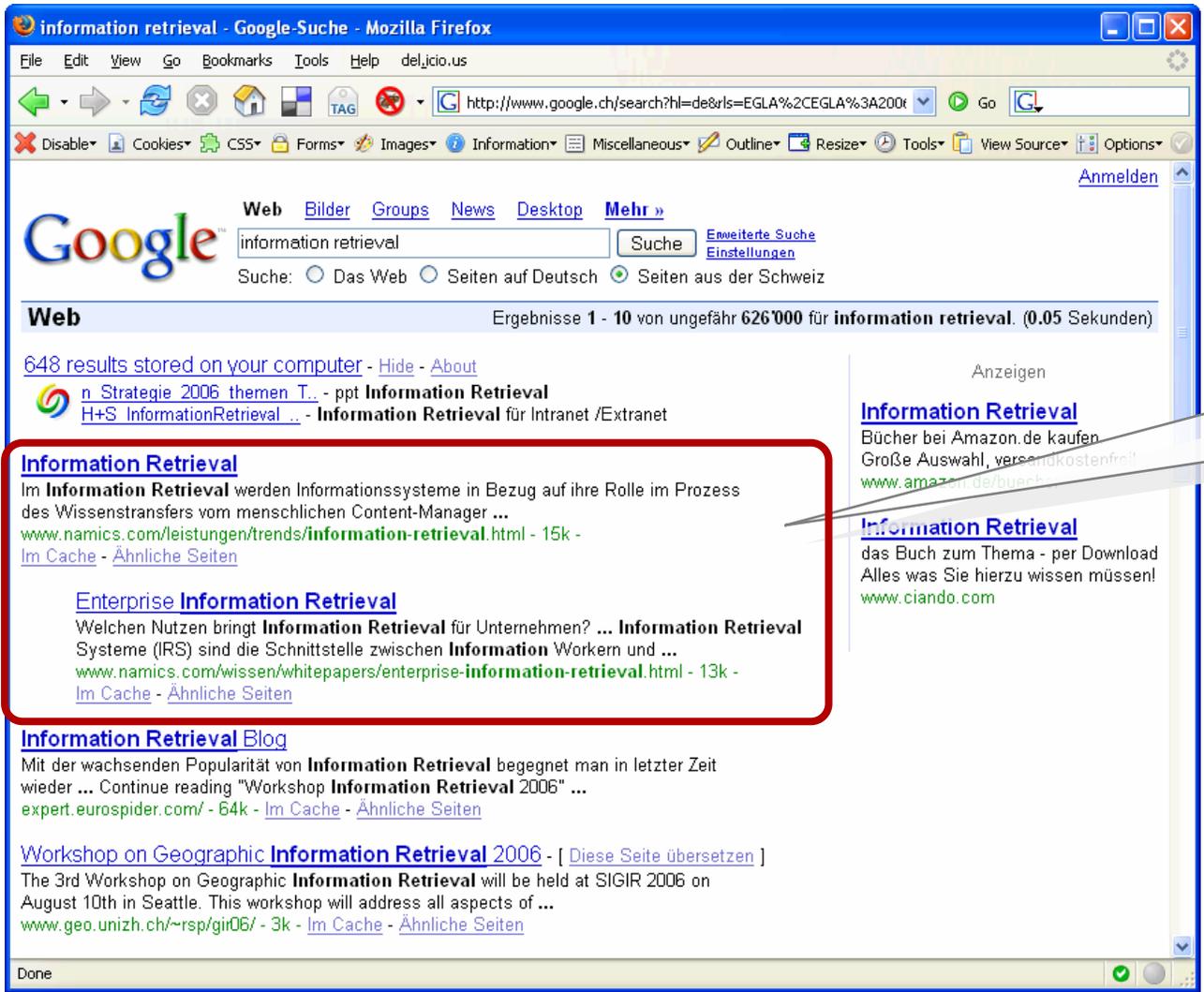
[namics Weblog: Ich als Zugfan 2.0](#)  
Der namics Weblog. Persönliche Stimmen und Meinungen von Mitarbeiterinnen und Mitarbeitern. [namics @ www.flickr.com](#). Impressum ...  
[blog.namics.com/2007/05/ich\\_als\\_zugfan.html](#) - 20k - [Im Cache](#) - [Ähnliche Seiten](#)  
[ [Weitere Ergebnisse von blog.namics.com](#) ]

[NAMICS EUROPE GMBH](#)  
Namics Europe GmbH Dietlindenstr. 15 80802 Munich / Germany. Fon: +49 89 36036 730 ...  
Mobil: +49 160 9646 9999. Email: [info@namics-europe.com](mailto:info@namics-europe.com).  
[www.namics-europe.com/](#) - 4k - [Im Cache](#) - [Ähnliche Seiten](#)

Zu viele Treffer?

Ranking.

# Was gehört auf den ersten Platz?



The screenshot shows a Mozilla Firefox browser window with the address bar containing a Google search URL. The search results for 'information retrieval' are displayed. A red box highlights the first search result, which is a document from n.namics.com. A speech bubble points to the second search result, which is an Amazon.de advertisement for books on 'Information Retrieval'.

**Information Retrieval**  
 Im **Information Retrieval** werden Informationssysteme in Bezug auf ihre Rolle im Prozess des Wissenstransfers vom menschlichen Content-Manager ...  
[www.namics.com/leistungen/trends/information-retrieval.html](http://www.namics.com/leistungen/trends/information-retrieval.html) - 15k -  
[Im Cache](#) - [Ähnliche Seiten](#)

**Enterprise Information Retrieval**  
 Welchen Nutzen bringt **Information Retrieval** für Unternehmen? ... **Information Retrieval** Systeme (IRS) sind die Schnittstelle zwischen **Information** Workern und ...  
[www.namics.com/wissen/whitepapers/enterprise-information-retrieval.html](http://www.namics.com/wissen/whitepapers/enterprise-information-retrieval.html) - 13k -  
[Im Cache](#) - [Ähnliche Seiten](#)

**Information Retrieval Blog**  
 Mit der wachsenden Popularität von **Information Retrieval** begegnet man in letzter Zeit wieder ... Continue reading "Workshop **Information Retrieval** 2006" ...  
[expert.eurospider.com/](http://expert.eurospider.com/) - 64k - [Im Cache](#) - [Ähnliche Seiten](#)

**Workshop on Geographic Information Retrieval 2006** - [ [Diese Seite übersetzen](#) ]  
 The 3rd Workshop on Geographic **Information Retrieval** will be held at SIGIR 2006 on August 10th in Seattle. This workshop will address all aspects of ...  
[www.geo.unizh.ch/~rsp/gir06/](http://www.geo.unizh.ch/~rsp/gir06/) - 3k - [Im Cache](#) - [Ähnliche Seiten](#)

**Information Retrieval**  
 Bücher bei Amazon.de kaufen  
 Große Auswahl, versandkostenfrei  
[www.amazon.de/buecher](http://www.amazon.de/buecher)

**Information Retrieval**  
 das Buch zum Thema - per Download  
 Alles was Sie hierzu wissen müssen!  
[www.ciando.com](http://www.ciando.com)

**Einverstanden?**

Ranking.

# Unterschiedliche Bedeutungen und Kontexte.



The screenshot shows a Windows Internet Explorer browser window with the address bar displaying 'http://de.wikipedia.org/wiki/L%C3%A4ufer'. The page content is the German Wikipedia article for 'Läufer'. The article title is 'Läufer' and it includes a navigation bar with 'Artikel', 'Diskussion', 'Seite bearbeiten', and 'Versionen/Autoren'. The main text explains that 'Läufer' is a term used in various contexts, including chess, sports, and biology. The left sidebar contains navigation links, a search box, and a list of tools.

**Artikel** Diskussion Seite bearbeiten Versionen/Autoren

## Läufer

Der Begriff **Läufer** bezeichnet

- einen besonderen **Teppich**, siehe **Läufer (Teppich)**
- eine Figur im Schachspiel, siehe **Läufer (Schach)**
- einen schnellen Fußboten (siehe z.B. **Läuferbrunnen**)
- einen Teil eines **Wankelmotors**
- ein Bauteil einer **Turbine**
- ein Bauteil von **elektrischen Maschinen**
- den oberen (sich drehenden) **Mühlstein**
- eine Kurzform für **Läuferschwein** (**Hausschwein** von 5–25kg)
- eine Bezeichnung für eine Hilfskraft in der **Gastronomie**, die z. B. die Bars mit Ware beliefert. ('commis')
- ein bewegliches Seil an Schiffen
- einen Teil eines **Rechenschiebers**
- ein Rad des **Seilers**
- eine schnelle Tonfolge in der **Musik**
- eine kleine Sanduhr in der **Schiffahrt**
- die Bezeichnung für einen Tropfen, der aufgrund von zu hohem Farbauftrag an der Wand herabläuft
- einen Stein im **Mauerwerk**, der längs zur Mauerrichtung verbaut ist, also mit der breiten Fläche nach außen zeigen; siehe **Backstein**; Pendant zu **Binder**
- eine Reihe gleichdicker Pflastersteine die der Länge nach entlang eines Hochbordsteines liegen **Gossenläufer**
- ein muskelgetriebenes Reiserad für ein oder zwei Personen
- ein waagerechtes Lauelement eines **Kranes** (Kurzform von Laufkatze)
- eine Firma für Büroartikel in der Nähe von **Hannover**, siehe **Läufer & Gutenberg GmbH**

**Läufer** steht im Bereich **Sport** für:

- eine Person, die eine Disziplin einer **Laufsportart** ausübt
- eine Position des Stellers beim **Volleyball**
- eine Kurzform für den Kreisläufer beim **Handball**
- eine altmodische Bezeichnung für eine Spielerposition beim **Fußball**

**Läufer** steht im Bereich **Biologie** für:

- ein **Moos** mit Lauffuß

Navigation

- Hauptseite
- Über Wikipedia
- Themenportale
- Von A bis Z
- Zufälliger Artikel

Mitmachen

- Hilfe
- Autorenportal
- Letzte Änderungen
- Spenden

Suche

Artikel Volltext

Werkzeuge

- Links auf diese Seite
- Änderungen an verlinkten Seiten
- Hochladen
- Spezialseiten
- Druckversion
- Permanenlink
- Artikel zitieren

Ranking.

## Optimierungsansätze.

- » Gewichtung von Kollektionen oder Bestandteilen verändern
- » Metadaten :-/
- » Top-Treffer
  
- » Und dabei: Messen und analysieren!
  - Ranking ist unsichtbar
  - Für Benutzer nicht quantitativ zu beschreiben

# Fragen & Antworten.



Vielen Dank für Ihre Zeit!



[markus.tressl@namics.com](mailto:markus.tressl@namics.com)