

# Warum Technik nicht alles ist (Referat 2b)



**Fachtagung: Suchfunktionen im Web**  
Zürich, 26. Oktober 2006

Jürg Stuker, namics

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

# Information Retrieval? – Kennen wir alle



# Information Retrieval? – Aber auch so

The screenshot shows the GoYellow.de website interface in a Microsoft Internet Explorer browser window. The address bar shows the URL: <http://www.goyellow.de/map/60329-frankfurt/gutleutstr.-96?lyrs=XsF3QMEVZkw!%2C%2BgFeVuN7f!%2CA46QI5wIbA%2CN2uG2gx5Zm!%2CKyO>. The page title is "Stadtplan gutleutstr.-96, 60329-frankfurt | GoYellow.de - Microsoft Internet Explorer".

The main content area features a map of Frankfurt, Germany, with various POI markers. A search bar at the top left contains "Str.Nr. Gutleutstr. 96" and "PLZ, Ort 60329 Frankfurt". The map shows a street grid with the Main River (Main) visible. A pop-up window titled "Restaurants und Gaststätten" is open, displaying details for "Buch-Café im Jüdischen Museum" located at Untermainkai 14, 60311 Frankfurt am Main, with phone number Tel.069 - 234921. The pop-up also includes a "Kostenlos anrufen" button and a "DB Anfahr (Tür zu Tür)" section with a "Von" input field.

On the left side, there is a sidebar with "Interessante Punkte" (Interesting Points) and a list of categories with checkboxes: Verkehr, Hotel/Restaurant, Kultur/Freizeit, Öffentliche Gebäude, Shopping, Service, Confluence Points, Gemeinden, and Wetter-Info. Below this, there are advertisements for "Signatures Veranda Restaurant im InterContinental Frankfurt" and "Le Meridien Park Hotel".

On the right side, there is a red vertical banner with promotional text: "Tagesaktuelle Superpreise", "Über 180.000 Hotels weltweit", and "Kostenlose Änderungen + Stornierungen". At the bottom of the banner is the HRS.de logo and the text "HOTEL RESERVATION SERVICE".

The footer of the page includes copyright information "© GoYellow GmbH", navigation links like "Ihr Werbeeintrag", "Login", "Alles über GoYellow", "Impressum", and "Nutzungsbedingungen", and a scale bar indicating 100 meters.

<http://www.goyellow.de/map>

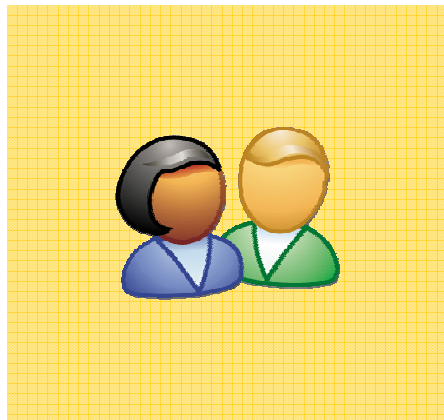
# Information Retrieval? – Oder so...

The screenshot shows a web browser window titled "ZUGer Onlineberatung - Funktionssuche - Microsoft Internet Explorer". The address bar shows the URL: [http://www.vzug.ch/v-zug/internet.nsf/vPageID/DE\\_FunktionsFinder](http://www.vzug.ch/v-zug/internet.nsf/vPageID/DE_FunktionsFinder). The page content includes:

- Navigation:** "ZUGer Onlineberatung" header with a "Home" link and a "V ZUG" logo.
- Search Progress:** "1 Geräte auswählen", "2 Konfigurations...", "3 Druckers...", "weiter".
- Left Sidebar:**
  - Geräteübersicht
  - Funktionsuche**
  - Sortiments Wegweiser
  - Preis/Leistungsvergleich
  - Gerätevergleich
- Main Content Area:**
  - Steamer** (Dropdown: "anderes Sortiment wählen")
  - Einsteigersuche** (Checkbox: "Wechseln zur Funktionssuche für Profis")
  - Dampf** (Text: "Ich möchte mit einem drucklosen Dampfverfahren arbeiten, welches sehr schonend dämpft."; Dropdown: "Wünschenswert")
  - Regenerieren** (Text: "Unregelmässige Arbeitszeiten, Sport, Kurse, Schule: Vielerorts ergeben sich für die einzelnen Familienmitglieder verschiedenste Essenszeiten. Daher ist das Programm 'Regenerieren' für mich"; Dropdown: "Sehr wichtig")
  - Profi-Backen** (Text: "Ich möchte mit meinem Steamer auch feines, knuspriges Brot backen"; Dropdown: "Weiss nicht")
  - echte Heissluft** (Text: "Backen mit echter Heissluft ermöglicht ein Arbeiten auf bis zu 3 Ebenen."; Dropdown: "Wünschenswert")
  - Button: "passendes Gerät suchen"
- Geräteliste (Right Column):**
  - Backöfen
  - Steamer
  - Herde
  - Coffee-Center
  - Mikrowellen
  - Kochfelder
  - Bedienelemente
  - Dunstabzüge
  - Gutmann-Hauben
  - Geschirrspüler
  - Kühlschränke
  - Waschautomaten MFH
  - Wäschetrockner MFH
  - Waschautomaten EFH
  - Wäschetrockner EFH
- Wissensportal (Green Box):**
  - Fragen zu Steamern**
    - Benötige ich für den Einbau eines Steamers einen separaten Wasseranschluss?
    - Was für Vorteile bringt das Electronic Steam System (ESS)?
    - Was bringt mir die Funktion "Profi-Backen" für Vorteile?
    - mehr
  - Einkaufstipps**
    - Wussten Sie, dass ZUG als erster Anbieter der Branche schon für die Zukunft gerüstet ist? Auch im Bereich der Gerätevernetzung ist ZUG führender Anbieter. Mehr Informationen dazu erhalten Sie auch unter [www.zughome.ch](http://www.zughome.ch)
    - nächster Tipp
  - Möchten Sie mehr erfahren**
    - Übersichtsseite aufrufen

<http://www.vzug.ch/onlineberater>

# Benutzeranfrage: Ein Wortvergleich



Anfrage



Dokument

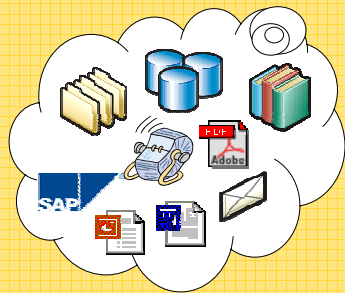
Was gibt der User ein?

Jürg Stuker's Kontaktdaten  
 in St. Gallen lauten:  
 E-Mail: [jst@namics.com](mailto:jst@namics.com)  
 Tel. +41 (0)71 / 228 67 01  
 Handy: +41 (0)79 / 252 67 01

# Ein Tag im Leben einer IR-Applikation

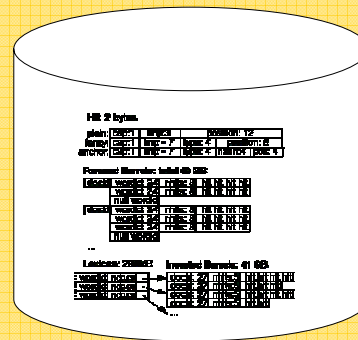
## 1. Daten

Akquisition und Speicherung aller Dokumente der Kollektion



## 2. Index

Erstellen einer effizienten Datenstruktur für die Suche



## 3. Benutzeranfrage

Finden passender Dokumente auf eine Benutzeranfrage



## 4. Resultate

Präsentation der Treffer in der richtigen Reihenfolge



# Daten: Kollektion

» Kollektion = Menge der auffindbaren Dokument

The screenshot displays the 'relevancy 6.6 Administration' web interface. The browser window title is 'relevancy 6.6 Administration - Mozilla Firefox' and the address bar shows 'http://dev.search.namics.com/admin/index.html'. The interface is divided into several sections:

- AdminTool for relevancy 6.6:** A sidebar on the left with navigation links: System Control, Information Source Browser, Sub-collections (highlighted), Add Document, System Status, Search, and Administrator Manual.
- Sub-collections:** The main content area with the heading 'Sub-collections'. Below it, a message states: 'Alter the sub-collections below and save them by pressing the "Submit" button in the applet. The changes will take effect the next time the Synchronizer is restarted.'
- Collection Tree:** A tree view on the left showing the hierarchy of collections:
  - all
    - whole collection
      - Standort
      - Format
      - Teilkollektionen
        - know.namics
        - PEP-Pool
        - Data
        - Projekte
        - News
        - Online Archiv
      - Gesamte Kollektion
        - collection entière
        - tutta la collezione

- Configuration Panel:** On the right, a configuration form for the selected sub-collection 'know.namics'. It includes:
- Selected sub-collection:** know.namics
- Defined by concept:**
- Description:** know.namics
- Parents:** Teilkollektionen
- Patterns:** ^http://my.namics.com/intern/knowhow.nsf, IN ^http://my.namics.com/intern/knowhow.nsf
- Include/Exclude:**  include,  exclude
- Search Sidebar:** A floating window titled 'Search' with a search box and a dropdown menu. The dropdown menu is open, showing options: all collections, Projekte auf Fileserver, Data auf Fileserver, know.namics (highlighted), my.namics, PEP-Pool, by last name, and by first name.

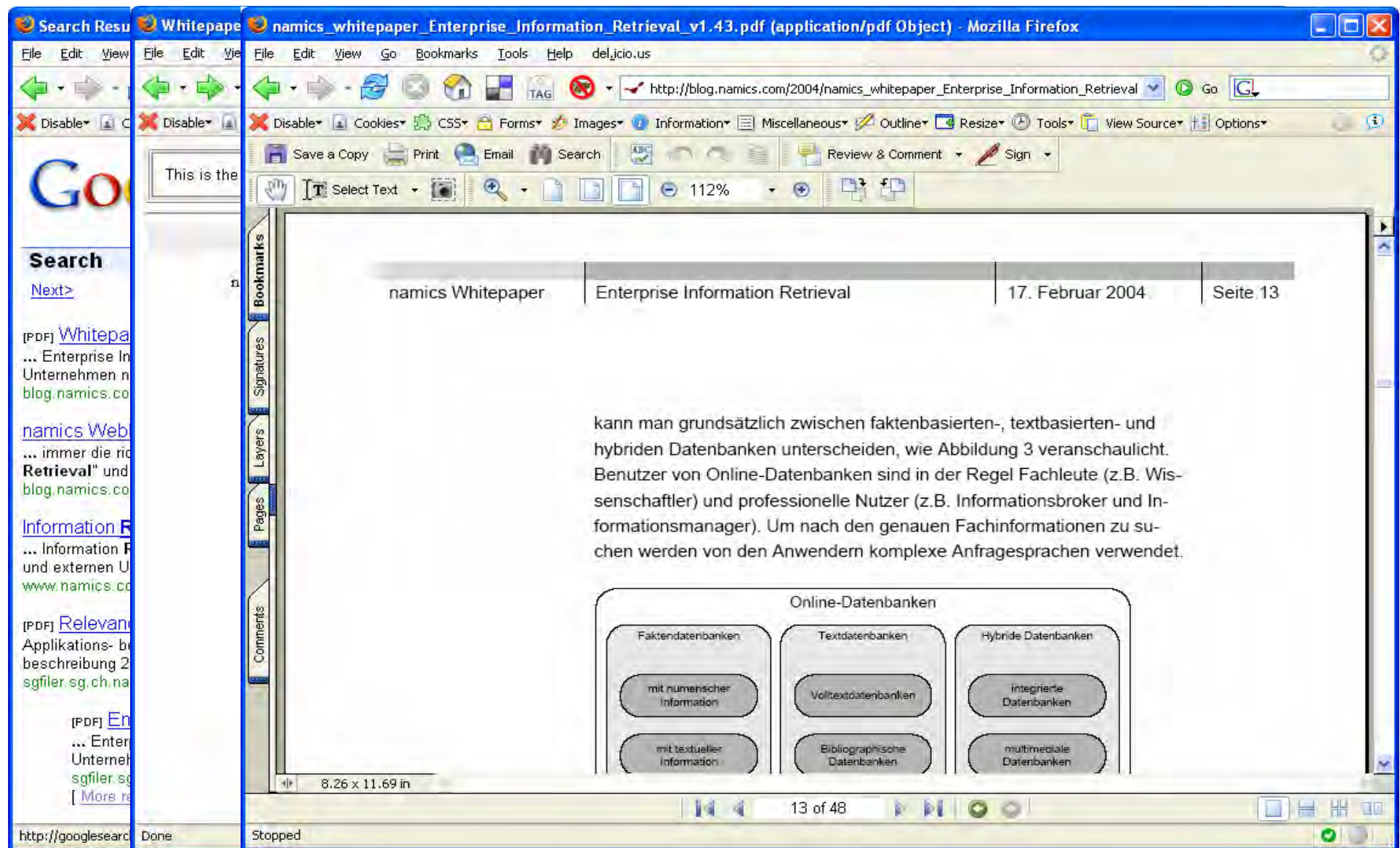
At the bottom of the applet, there are buttons for 'New sub-collection', 'Link sub-collection', 'Move sub-collection', 'Move up', 'Move down', 'Submit', and 'Reset'. A status bar at the very bottom indicates 'Applet: CollectionApplet started'.

## Daten: Formate

- » Bei einer Volltextsuche gibt der User Text ein...
- » ...die zu durchsuchenden Daten müssen im selben Format vorliegen resp. umgewandelt werden können!
- » Formate
  - html, txt, csv
  - Office: Powerpoint, Word, Excel
  - PDF: evtl. nach HTML konvertiert
  - E-Mail: Outlook, Exchange, Lotus Notes etc.
  - ERP, CRM, Datenbanken...
  - usw.
- » Probleme
  - Software-Versionen
  - Druckschutz oder Passwort
  - Kein Text z.B. Bilder



# Daten: Formate – Beispiel (indexiert)



# Daten: Formate – Beispiel (nicht indexiert)

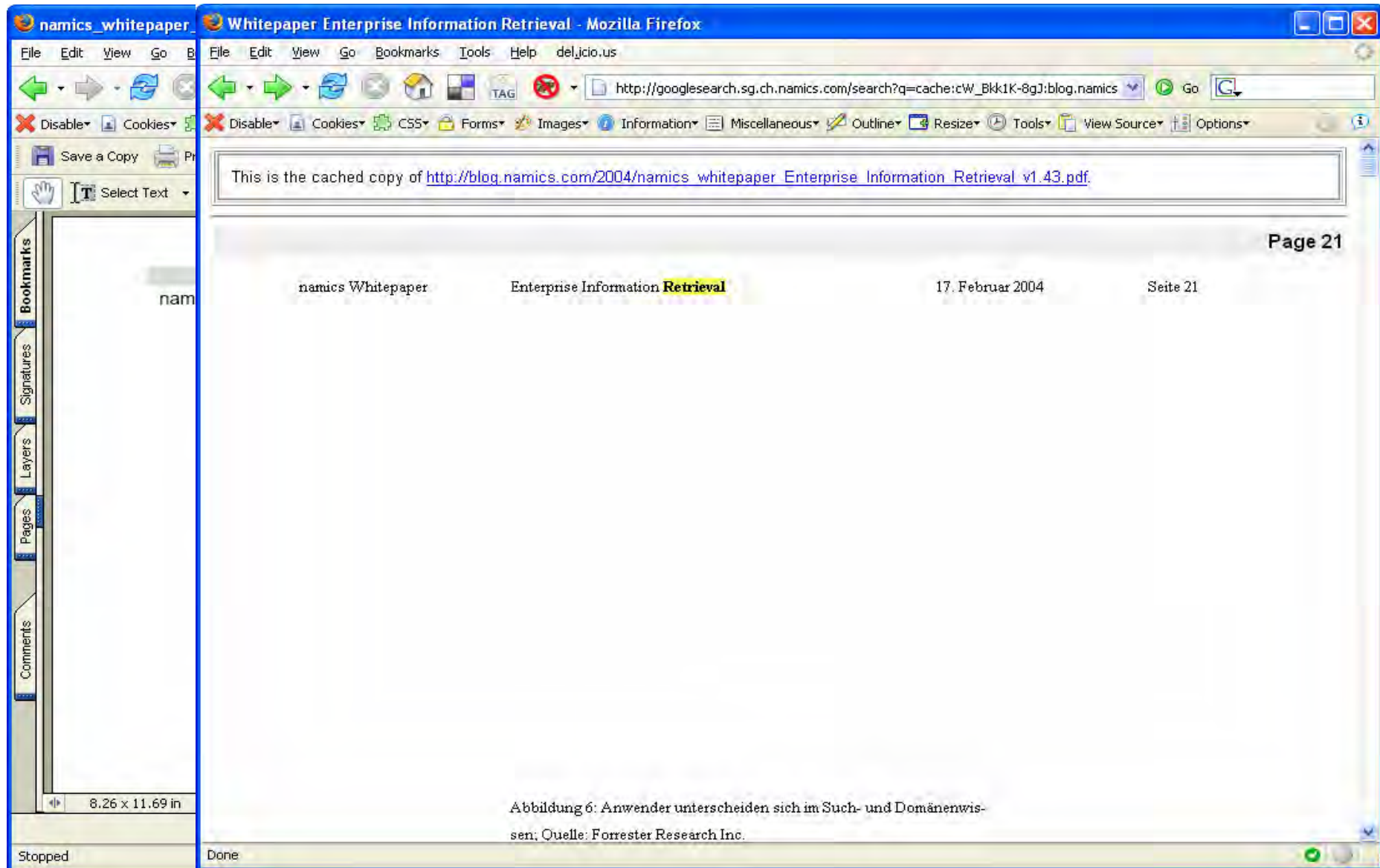
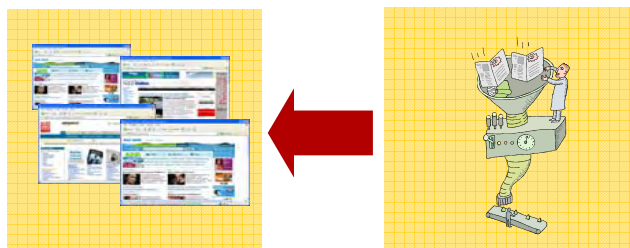


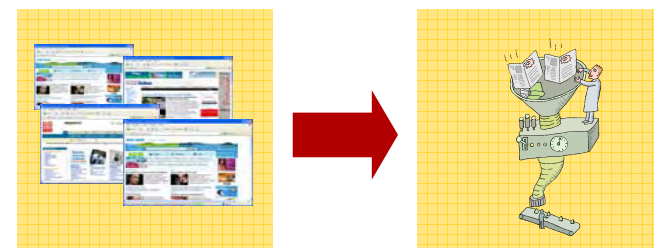
Abbildung 6: Anwender unterscheiden sich im Such- und Domänenwissen; Quelle: Forrester Research Inc.

## Index: Vollständigkeit und Aktualität

- » Wie schnell reflektiert sich eine Änderung in den Daten (z.B. Update Webpage oder Veränderung einer Datei im Dateisystem) in der IR-Applikation?
- » Wie effizient erkennt die IR-Applikation Änderungen?
  - muss die ganze Kollektion erneut durchsucht werden
  - wird die Suche aktiv benachrichtigt
- » Wie viele der potentiellen Informationselemente in der Kollektion sind durch die IR-Applikation erschlossen?



Pull-Update  
(Crawling/ Spidering)



Push-Update  
(Feeds)

# Index: Stoppworte

Erweiterte Suche Einstellungen Sprachtools Suchtipps

wer wie was wieso warum fernseh: Google Suche

Suche:  Das Web  Seiten auf Deutsch  Seiten aus der Schweiz

Die folgenden Wörter kommen sehr häufig vor und wurden daher in Ihrer Suchanfrage ignoriert: **wer wie was wieso warum.** [[Einzelheiten](#)]

Web Bilder Groups Verzeichnis

Das Web wurde nach **wer wie was wieso warum fernsehsendung** durchsucht. Ergebnisse 1 - 10 von ungefähr 26,800. Suchdauer: 0.

[Reiseportal Hessen](#)  
 ... Beiträge: **Fernsehsendung**: Wiederentdeckung regionaler Spezialitäten  
 Freigegeben am Monday, September 02 @ 13:08:55 CEST von martina. ...  
[www.reiseportal-hessen.de/article.php?sid=50](#) - 32k - [Im Cache](#) - [Ähnliche Seiten](#)

[Amazon.de: Suchergebnisse Bücher:](#)  
 Alle Ergebnisse für: Schlagwort "**Fernsehsendung**". ... Kostenlose Lieferung ab 20,00  
 EUR Bestellwert. Alle 198 Treffer für Schlagwort "**Fernsehsendung**" ...  
[www.amazon.de/.../search-handle-url/index=books-de&field-subject=Fernsehsendung/ ref=sr\\_sp\\_dp\\_sub\\_1/](#) - 74k - [Im Cache](#) - [Ähnliche Seiten](#)

[MediaLine.FAKTEN.Medialexikon.Suche nach ???Fernsehsendung". ...](#)  
 ... Marketing-Dictionary, Contacts, Medialexikon. Suche nach ???**Fernsehsendung**", **Fernsehsendung**. ... aktuelles  
 FOCUS-Magazin, Medialexikon. Suche nach ???**Fernsehsendung**",  
[medialine.focus.de/PM1D/PM1DB/PM1DBD/PM1DBDA/PM1DBDAA/ pm1dbdaa.htm?begriff=Fernsehsendung&snr=1005](#) - 92k - [Im Cache](#) - [Ähnliche Seiten](#)

[Schweizerische Multiple Sklerose Gesellschaft - Diskussionsforum ...](#)  
 Mitteilung Thema: **fernsehsendung**. Erstellt von: lisa. ... Jan. 21 05 Uhr, 19.01.03

Resultate für 'wer wie was wieso warum fernseh:'

1. [100 Blondienew](#)  
 Blondinenwitze d  
[www.tino.net/jok](#)
2. [Witze, Spass ur](#)  
 Was macht eine  
 macht man, wen  
[www.friends4fun](#)
3. [selmi.ch : : 501](#)  
 Musik Photos M  
 Witze Professor  
 Belohnung: Der  
[www.selmi.ch/in](#)

» Begriffe, welche für das Auffinden keinen (oder sehr wenig) Wert haben. Bsp. der, die, das, es etc. -> Ziel: Verkleinern des Index'

# Benutzeranfrage

Anfrage 

» Jürg



Dokument 

Jürg Stuker's Kontaktdaten  
in St. Gallen lauten:  
E-Mail: [jst@namics.com](mailto:jst@namics.com)  
Tel. +41 (0)71 / 228 67 01  
Handy: +41 (0)79 / 252 67 01

# Benutzeranfrage

Anfrage 

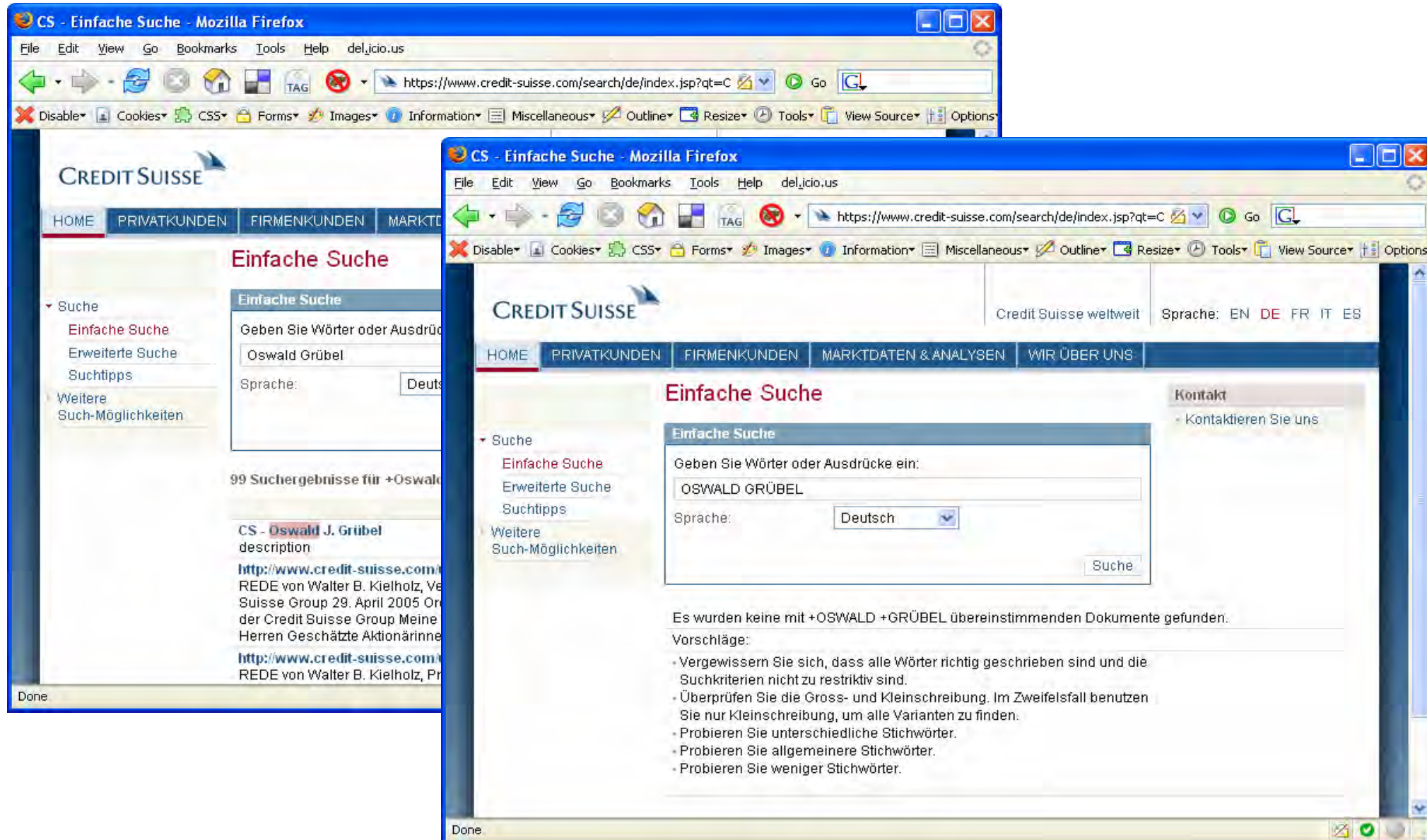
» JÜRIG



Dokument 

Jürg Stuker's Kontaktdaten  
in St. Gallen lauten:  
E-Mail: [jst@namics.com](mailto:jst@namics.com)  
Tel. +41 (0)71 / 228 67 01  
Handy: +41 (0)79 / 252 67 01

# Benutzeranfrage – Beispiel I



# Benutzeranfrage

Anfrage 

» St.Gallen

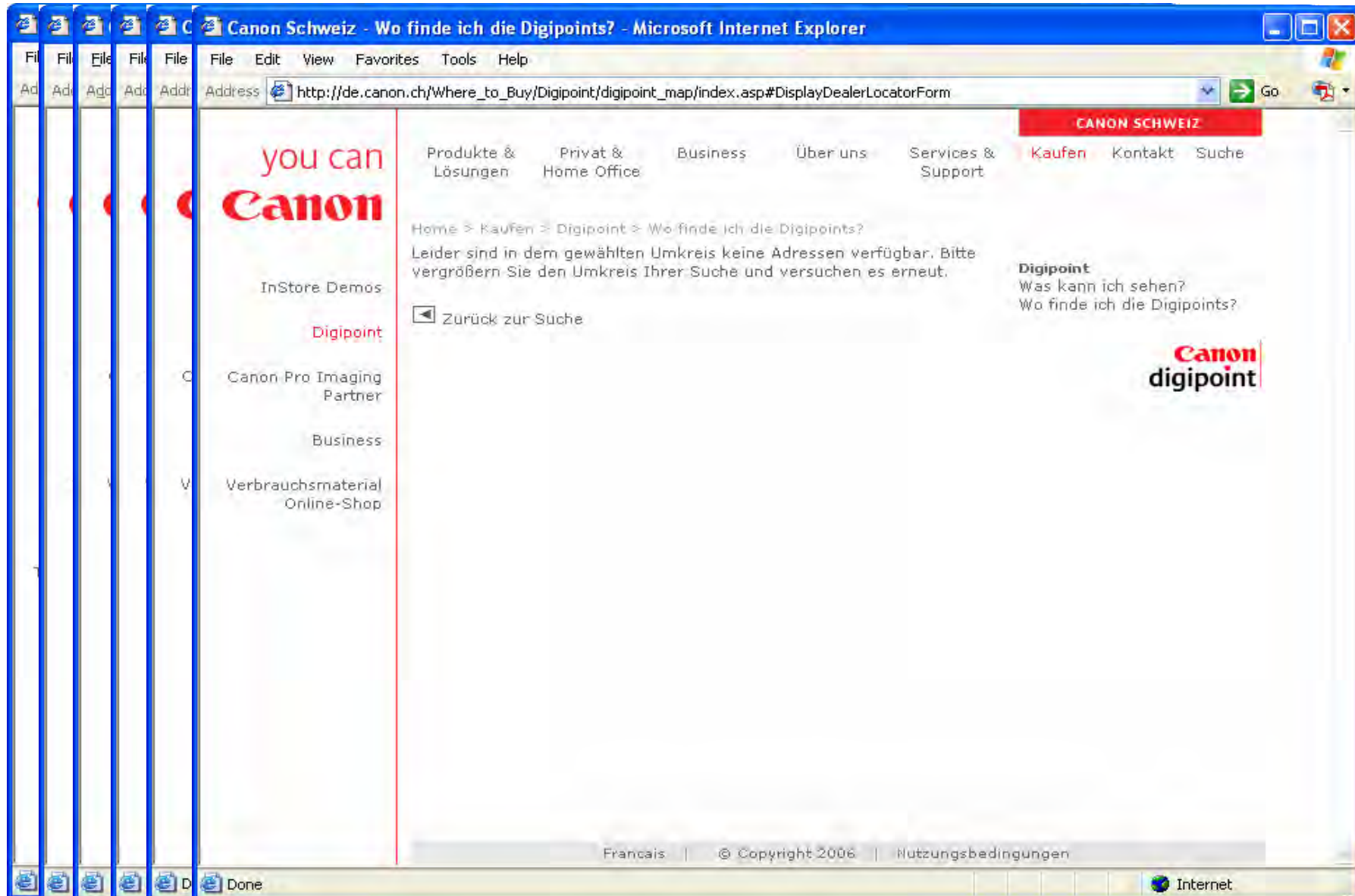


Dokument 

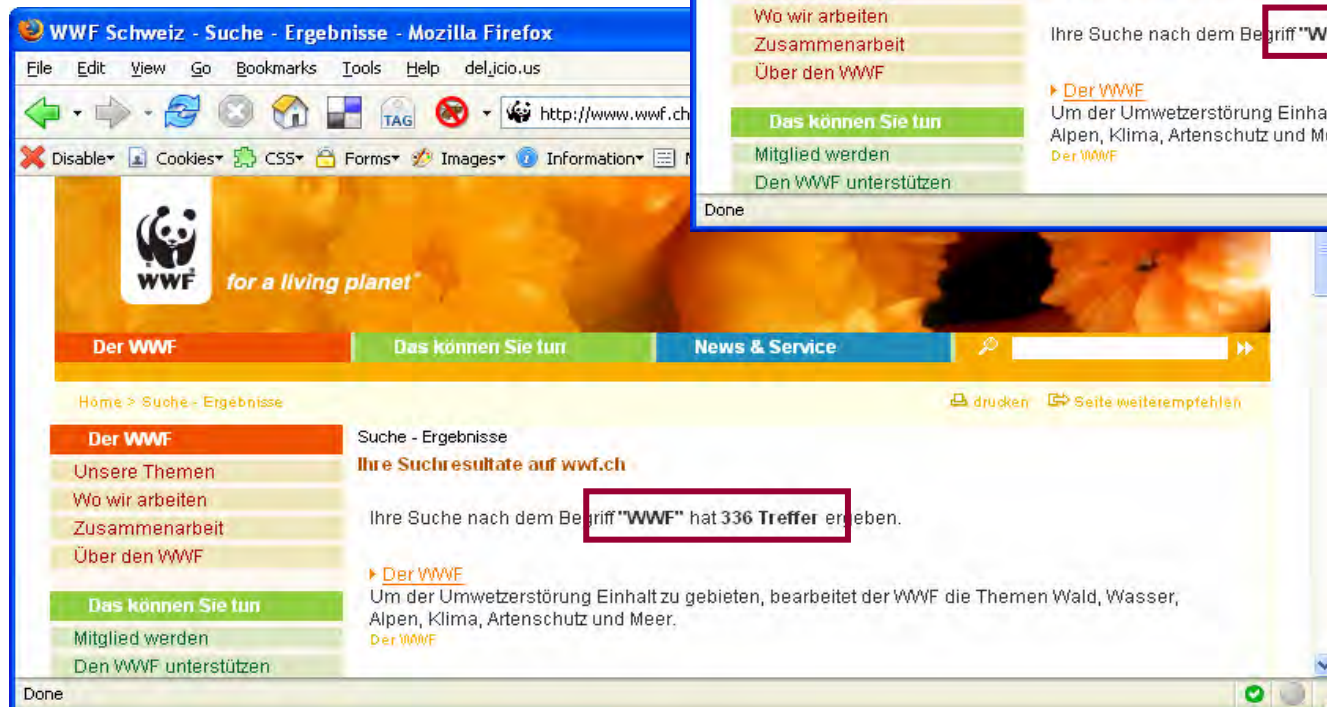
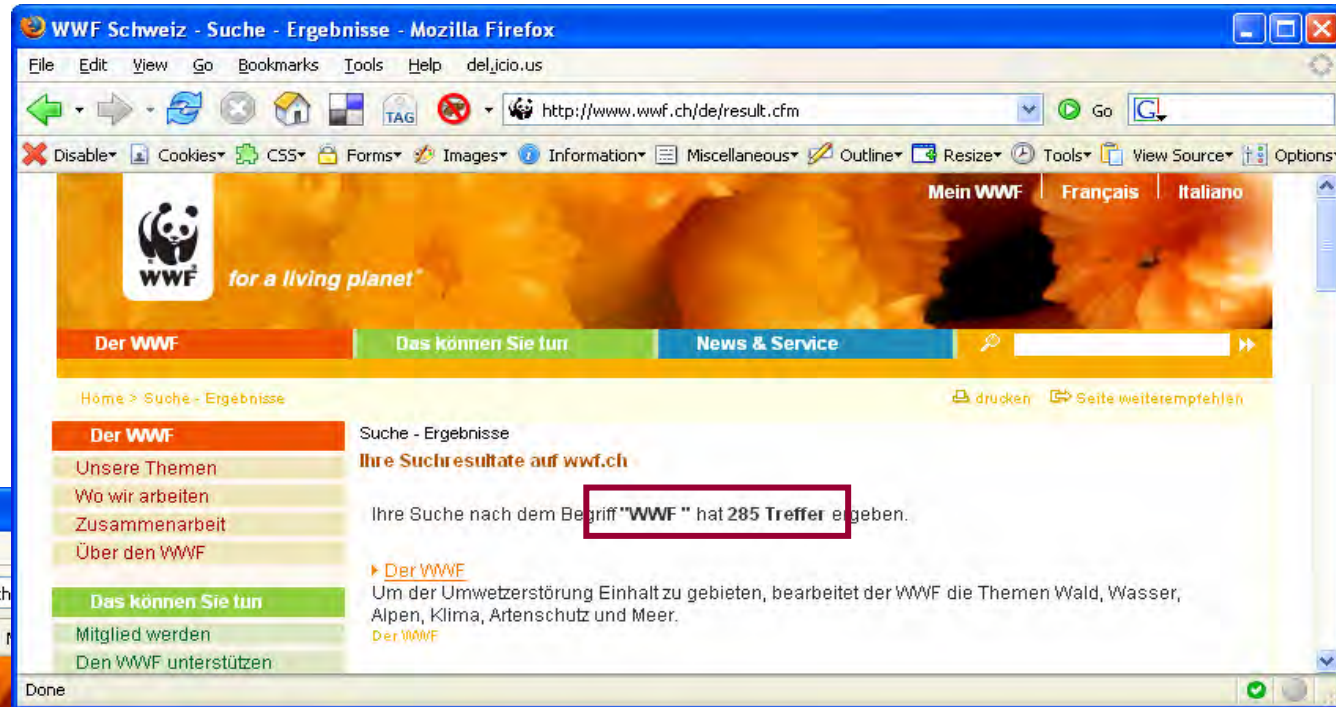
Jürg Stuker's Kontaktdaten  
in St. Gallen lauten:  
E-Mail: [jst@namics.com](mailto:jst@namics.com)  
Tel. +41 (0)71 / 228 67 01  
Handy: +41 (0)79 / 252 67 01



## Benutzeranfrage – Beispiel II



# Benutzeranfrage – Beispiel III



## Benutzeranfrage: Normalisierung

- » Normalisierung macht den Vergleich zwischen Suchanfrage und Trefferdokumenten robuster
- » Entfernen von für die Suche unnötiger Zeichen plus Rückführung auf eine einheitliche Schreibweise
- » Beispiele
  - Jürg == jürg == JÜRIG == juerg
  - Strasse == Straße
  - Rue de l'Athénée == Rue de l' Athenee
- » Kann auch schwieriger sein
  - St. Gallen == Sankt Gallen
  - Hans Peter == Hans-Peter == H.P. == H.-P. u.s.w
  - Alfred Escher-str. == Alfred Escher-Str. == Alfred-Escher-Strasse
- » Kann auch sehr schwierig sein...

# Benutzeranfrage

Anfrage 

- » 079/228 67 01
- » 228'67'01
- » ...



Dokument 

Jürg Stuker's Kontaktdaten  
in St. Gallen lauten:  
E-Mail: [jst@namics.com](mailto:jst@namics.com)  
Tel. +41 (0)71 / 228 67 01  
Handy: +41 (0)79 / 252 67 01

# Benutzeranfrage

Anfrage 

» kontakt



Dokument 

Jürg Stuker's Kontaktdaten  
in St. Gallen lauten:  
E-Mail: [jst@namics.com](mailto:jst@namics.com)  
Tel. +41 (0)71 / 228 67 01  
Handy: +41 (0)79 / 252 67 01

## Benutzeranfrage: Wortzerlegung (=Decompounding)

- » Deutsch ist eine produktive Sprache...
- » Um die Ausbeute zu verbessern, wird Wort zerlegt
  - kontaktdaten > kontakt daten
  - fussballeuropameisterschaft > fuss ball europa meisterschaft / fussball europameisterschaft
  - Achtung: washington > was hing ton
- » möglicher Ansatz
  - Wörterbuch mit ausschliesslich nicht zusammen gesetzten Worten
  - Alle Wörter innerhalb des Suchbegriffs finden
  - Variante mit den wenigsten Wortteilen auswählen
  - Bei „Gleichstand“: Kombination der häufigsten (rel. zu Kollektion oder zum Dokument)

## Benutzeranfrage: Spracherkennung

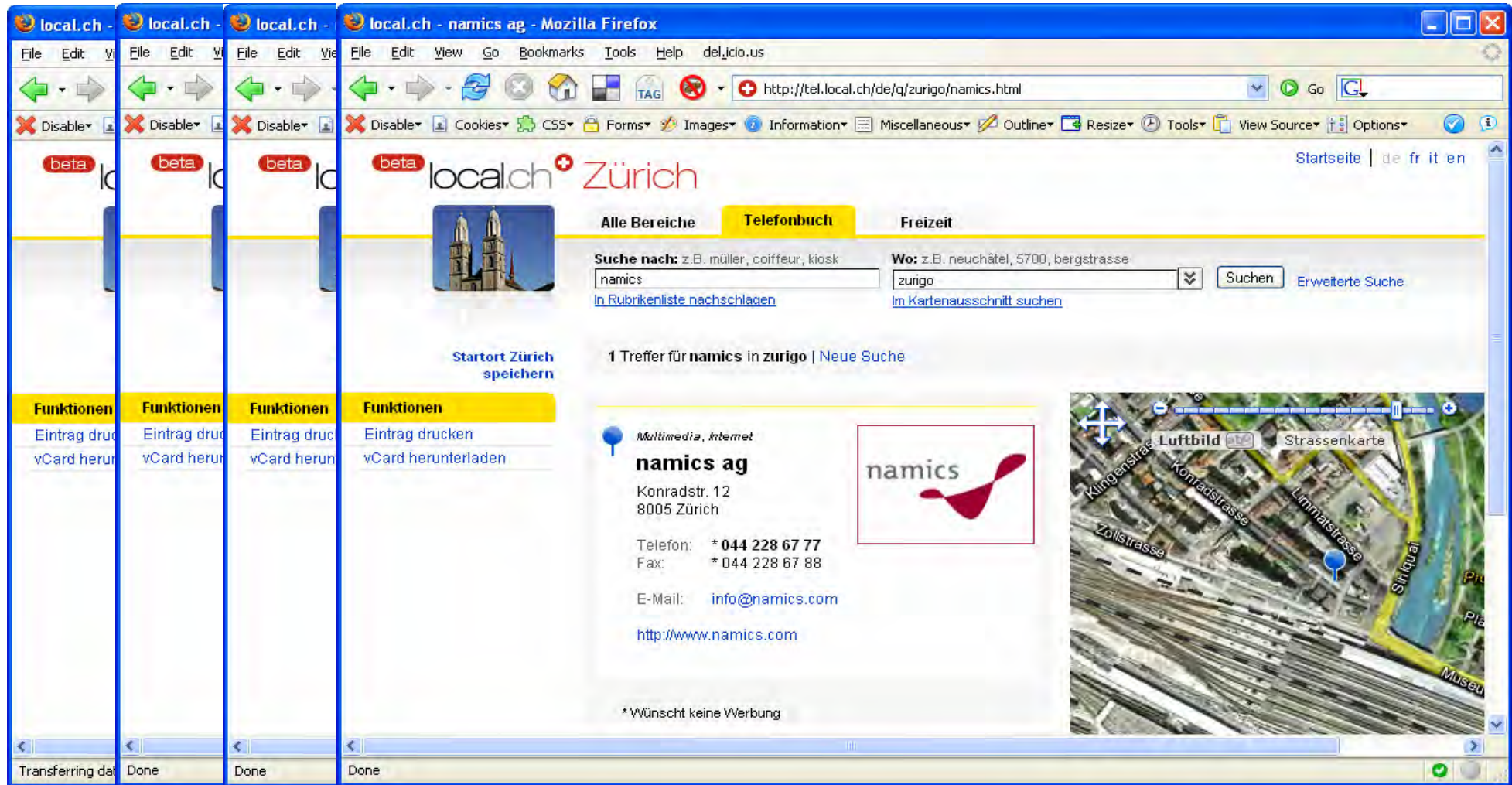
- » Die verlässliche Erkennung der Sprache ist wichtig für zahlreiche Analyseverfahren z.B. die Wortzerlegung
- » Ansätze
  - Abgabe durch Mensch (z.B.: auf der Kollektion)
  - Wörterbücher
  - Buchstabenerkennung, Codepage, META etc.
  - Häufigkeitsverteilung von Silben
  - u. a.

## Benutzeranfrage: Synonymlisten

- » Sprachvarianten
  - St. Gallen == St-Gall == San Gallo == San Gagl
- » „alternative“ Schreibweisen
  - St. Gallen == SG == 9000, 9001, 9002
- » Schreibfehler
  - Jürg Stuker == Jürg Stucker
  
- » Erstellen von Synonymlisten
  - Kontextwissen / Wörterbücher
  - Betriebsdaten z.B. Null-Treffer Queries
  - Statistisch während der Nutzung



# Benutzeranfrage: Synonymlisten – Beispiel 1



# Benutzeranfrage: Suchassistent – Beispiel 2

The screenshot shows two browser windows. The left window is the 'Portal Kanton Zürich' with a list of 10 items under 'INHALT'. The right window is 'SWISSLEX :: WESTLAW' showing a search for 'kündigung zu unzeit'. The search results are as follows:

**JurSearch - Suche in der juristischen Linkkollektion**

Suchbegriff:  [Löschen](#) | [Hilfe](#) | [Erweiterte Suche](#)

Zwischen:  und  (tt.mm.jjjj)

[Nur in ausgewählten Linkkollektionen suchen](#)

65 exakte Treffer gefunden für **kündigung zu unzeit**

- 1. Verwaltungsgerichtsentscheide BL 2000**

Übersicht Verwaltungsgericht SGS Übersicht Entscheide 2000 Basellandschaftliche Verwaltungsgerichtsentscheide 2000 3. Personalrecht 3.2 Beendigung des Angestelltenverhältnisses Sachregister Gesetzesregister Abkürzungsverzeichnis 3.2.3 **Kündigung** zur Unze... ( 07.01.2002 - Staats- und Verwaltungsrecht, Arbeits- und Sozialversicherungsrecht, Gesundheitsrecht, Kommunikations-, Transport- und Verkehrsrecht, Haftpflicht- und Privatversicherungsrecht, OR, ZGB, Landesverteidigung, Zivilprozess-, Verfahrens- und Schuldbetreibungs- und Konkursrecht, Straf- und Strafprozessrecht, Immaterialgüter- und Wettbewerbsrecht, Kartellrecht, Bank- und Finanzmarktrecht, Gesellschaftsrecht, BL, Urteile, BL - HTML - 20k )  
[http://www.baselland.ch/docs/gerichte/verwger/e.../vge2000\\_3-2-3-1.htm](http://www.baselland.ch/docs/gerichte/verwger/e.../vge2000_3-2-3-1.htm) [ähnliche Seiten](#)
- 2. 109 II 330**

Urteilkopf 109 II 330 69. Auszug aus dem Urteil der I. Zivilabteilung vom 27. September 1983 i.S. X. gegen A. AG und B. (Berufung) Regeste Arbeitsvertrag; **Kündigung** zur **Unzeit** durch den Arbeitgeber. Verlängerung der Kündigungsfrist wegen unverschuldete... ( 10.10.2006 - Staats- und Verwaltungsrecht, Arbeits- und Sozialversicherungsrecht, Gesundheitsrecht, OR, ZGB, Raumplanungs- und Baurecht, Umweltrecht, Energierecht, Zivilprozess-, Verfahrens- und Schuldbetreibungs- und Konkursrecht, Bank- und Finanzmarktrecht, Gesellschaftsrecht, Urteile, Bund - XHTML - 13k )  
<http://relevancy2.bqer.ch/cgi-bin/JumpCGI?id=BGE-109-II-330&lang=...> [ähnliche Seiten](#)
- 3. Verwaltungsgerichtsentscheide BL 2000 (Übersicht)**

Übersicht Verwaltungsgericht SGS Basellandschaftliche Verwaltungsgerichtsentscheide 2000 - Übersicht Sachregister Gesetzesregister Abkürzungsverzeichnis 2 Gemeinderecht 2.1 Einbürgerung 3 Personalrecht 3.1 Rechtsnatur des Angestelltenverhältnisses 3.2 B...

**Relevancy Assistant**

Aus der Umgebung der Suchbegriffe werden mit Algorithmen ähnliche oder verwandte Wörter vorgeschlagen, welche zur Verfeinerung der Suche anregen sollen.

Achtung: Für den Nutzer scheinbar "falsche" Vorschläge sind unvermeidlich.

Anfrage erweitern mit:  
[Art 404 OR zweckgerichtetheit](#)  
[kündigung](#)  
[ernstearbeit](#)  
[zwecksensitiv](#)  
[kündigungssystem](#)  
[Art 404 Abs 2 OR beendigungssystem](#)  
[auftragrechts](#)  
[steamboat](#)  
[BGE 78 II 1952](#)  
[unzeitigen](#)  
[kündigenden](#)  
[bananenschachteln](#)  
[saisonier](#)

Trefferliste sortiert nach:  
 Relevanz, [Datum](#)

## Benutzeranfrage VI

Anfrage 

- » St-Gall
- » Jürg Stucker
- » 9000



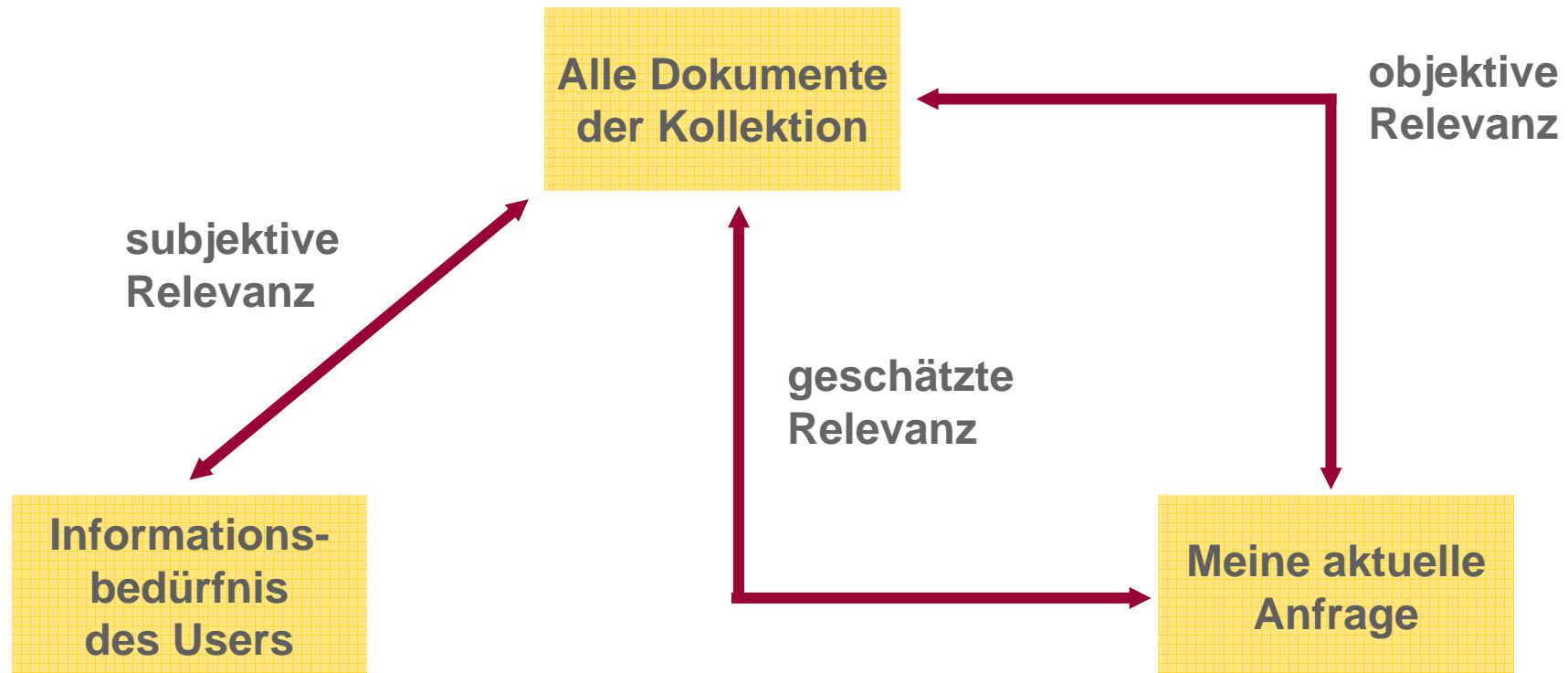
Dokument 

Jürg Stucker's Kontaktdaten  
in St. Gallen lauten:  
E-Mail: [jst@namics.com](mailto:jst@namics.com)  
Tel. +41 (0)71 / 228 67 01  
Handy: +41 (0)79 / 252 67 01

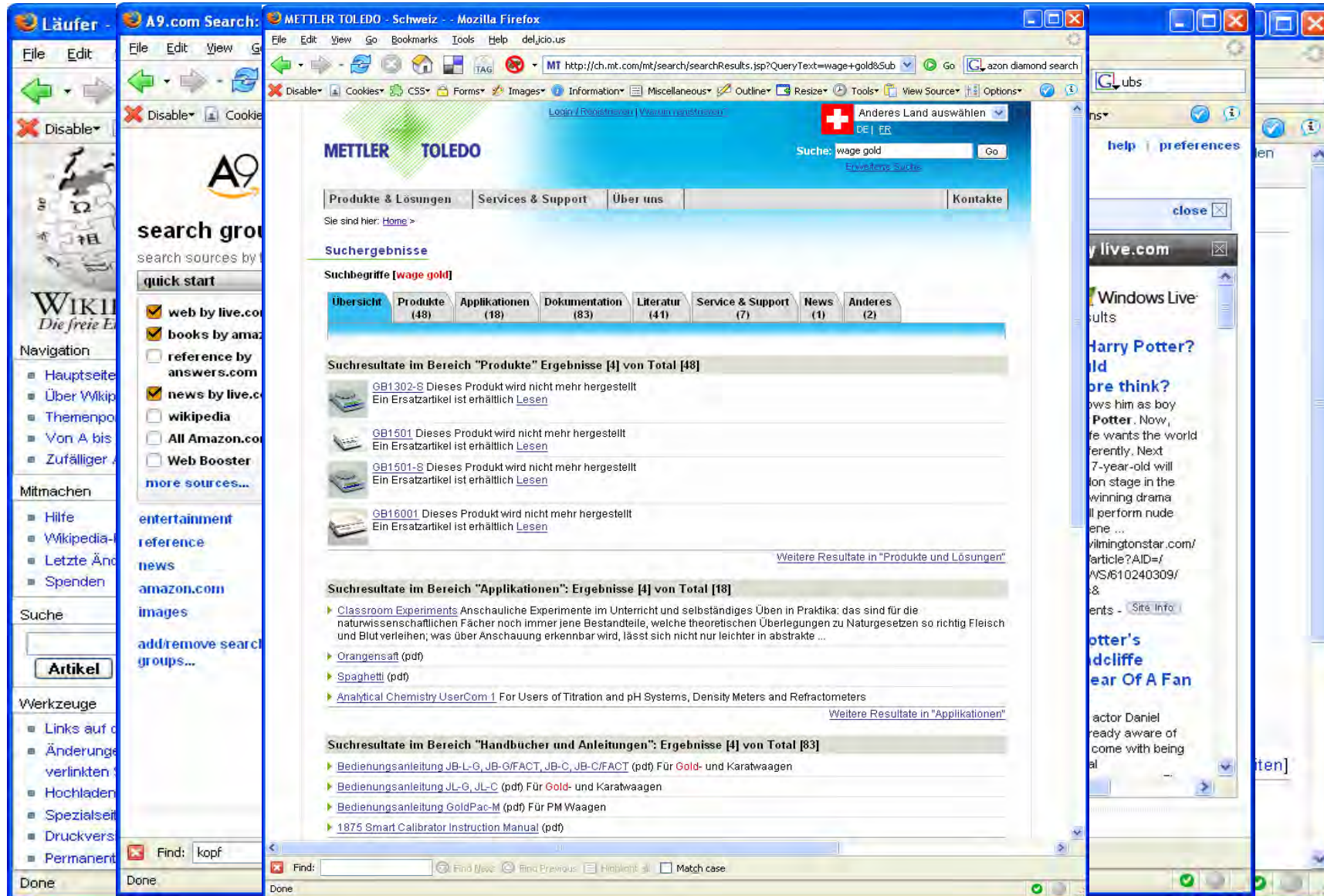
## Benutzeranfrage: Weitere Verfahren

- » So wie gezeigt, gibt es noch zahlreiche weitere Verfahren, die das Ziel verfolgen, den Wortvergleich robust zu gestalten. Diese sind abhängig von
  - den Daten
  - der Suchanforderung
  - der Sprache
  - dem User Interface
  - etc.
  
- » Nicht möglichst viele Verfahren sind wichtig (unterstützt ihre Software „Stemming“ etc.), aber die dem Einsatz angepassten...

# Resultate: Die Relevanz



# Resultate: Dem User die Entscheidung lassen



# Resultate: Eingabe steuern (Fehler vermeiden)

The image displays two screenshots of the Amazon.com Diamond Search interface. The left screenshot shows the search criteria selection page, and the right screenshot shows the search results page.

**Search Criteria (Left Screenshot):**

- Shape:** Round (selected), Princess
- Price:** \$100
- Carat:** 0.25 ct.
- Cut:** Ideal
- Color:** D, Col

**Search Results (Right Screenshot):**

You searched for:

- Shape:** Round
- Carat:** 0.25 to 6 Carat(s)
- Cut:** Ideal to Ideal
- Color:** E to D
- Clarity:** IF to FL
- Price:** \$100.00 to \$125,000.00

82 diamonds match your criteria

Sample Photo	Shape	Carat	Cut	Color	Clarity	Price	Compare (up to 20)	View Details / Buy
	Round	1.24	Ideal	D	IF	\$21,010.00	<input type="checkbox"/>	<a href="#">View / Select</a>
	Round	1.15	Ideal	D	IF	\$14,615.00	<input type="checkbox"/>	<a href="#">View / Select</a>
	Round	1.04	Ideal	D	IF	\$14,020.00	<input type="checkbox"/>	<a href="#">View / Select</a>
	Round	1.05	Ideal	E	IF	\$11,450.00	<input type="checkbox"/>	<a href="#">View / Select</a>
	Round	0.71	Ideal	D	IF	Sold	<input type="checkbox"/>	<a href="#">View / Find Similar</a>
	Round	0.71	Ideal	D	IF	\$5,030.00	<input type="checkbox"/>	<a href="#">View / Select</a>
	Round	0.63	Ideal	E	IF	\$2,615.00	<input type="checkbox"/>	<a href="#">View / Select</a>
	Round	0.55	Ideal	E	IF	\$2,385.00	<input type="checkbox"/>	<a href="#">View / Select</a>
	Round	0.42	Ideal	D	IF	Sold	<input type="checkbox"/>	<a href="#">View / Find Similar</a>
	Round	0.44	Ideal	E	IF	\$1,650.00	<input type="checkbox"/>	<a href="#">View / Select</a>
	Round	0.44	Ideal	E	IF	\$1,600.00	<input type="checkbox"/>	<a href="#">View / Select</a>

Darum ist Technik nicht alles ;-)