

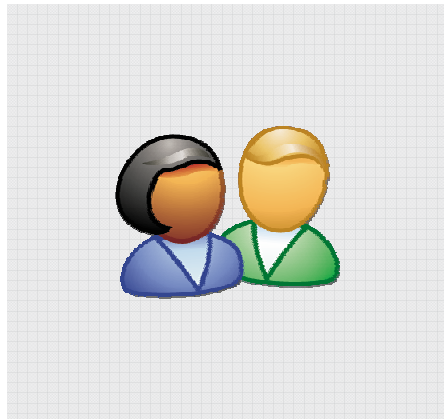
Einfach, erweitert und die anderen
(Referat 2a)



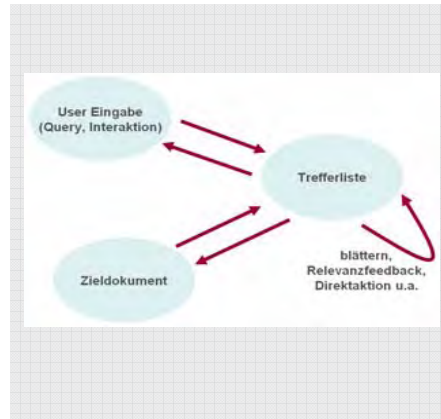
Fachtagung: Suchfunktionen im Web
Zürich, 26. Oktober 2006

Jacqueline Badran, Zeix
Gregor Urech, Zeix

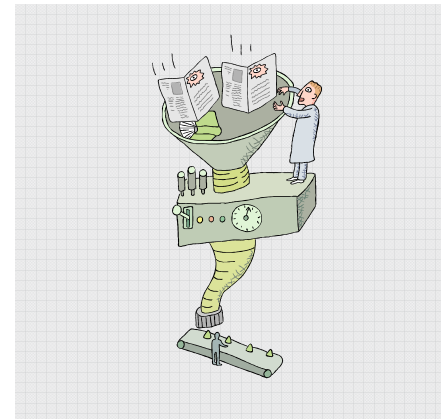
Relevante Dimensionen: Übersicht



User



User Interface



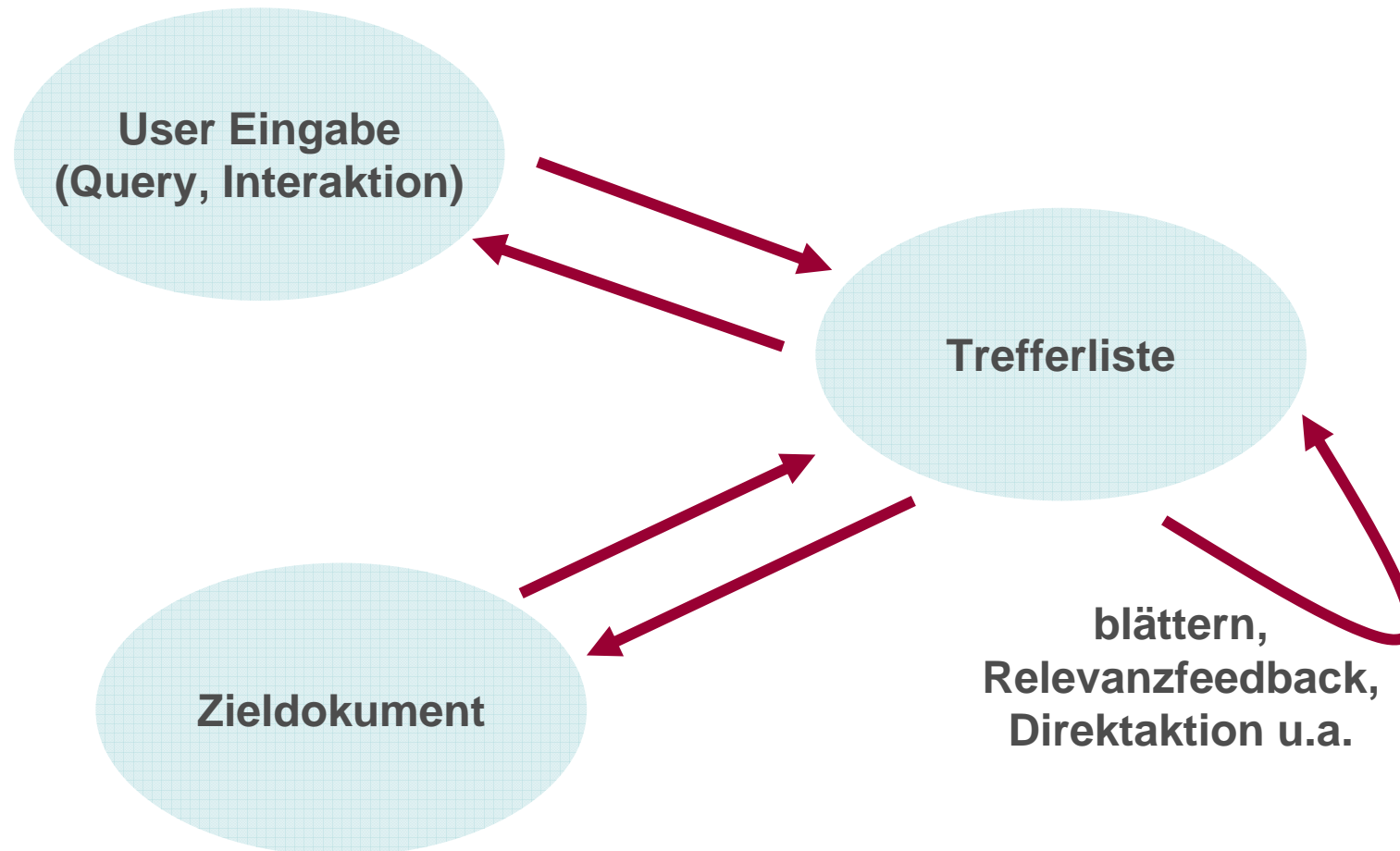
IR-
Applikation

unstrukturiert	
strukturiert	
annotiert	

Daten

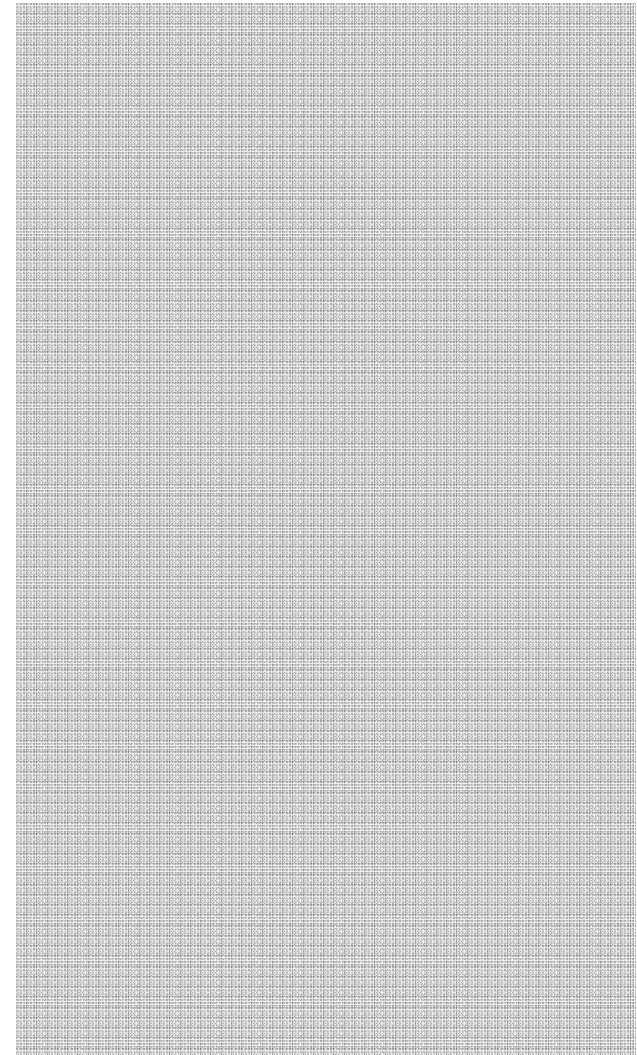
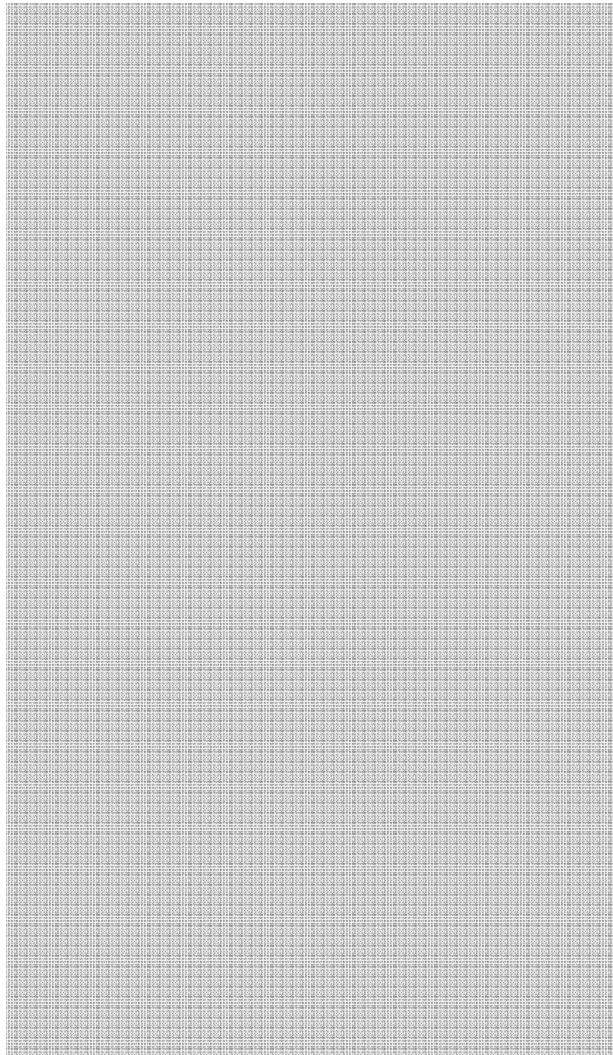


Detail User Interface



Unstrukturierte Daten

Unstrukturierte Daten



Unstrukturierten Daten

» Einzelne Objekte haben keine einheitlichen Attribute

» Beispiele

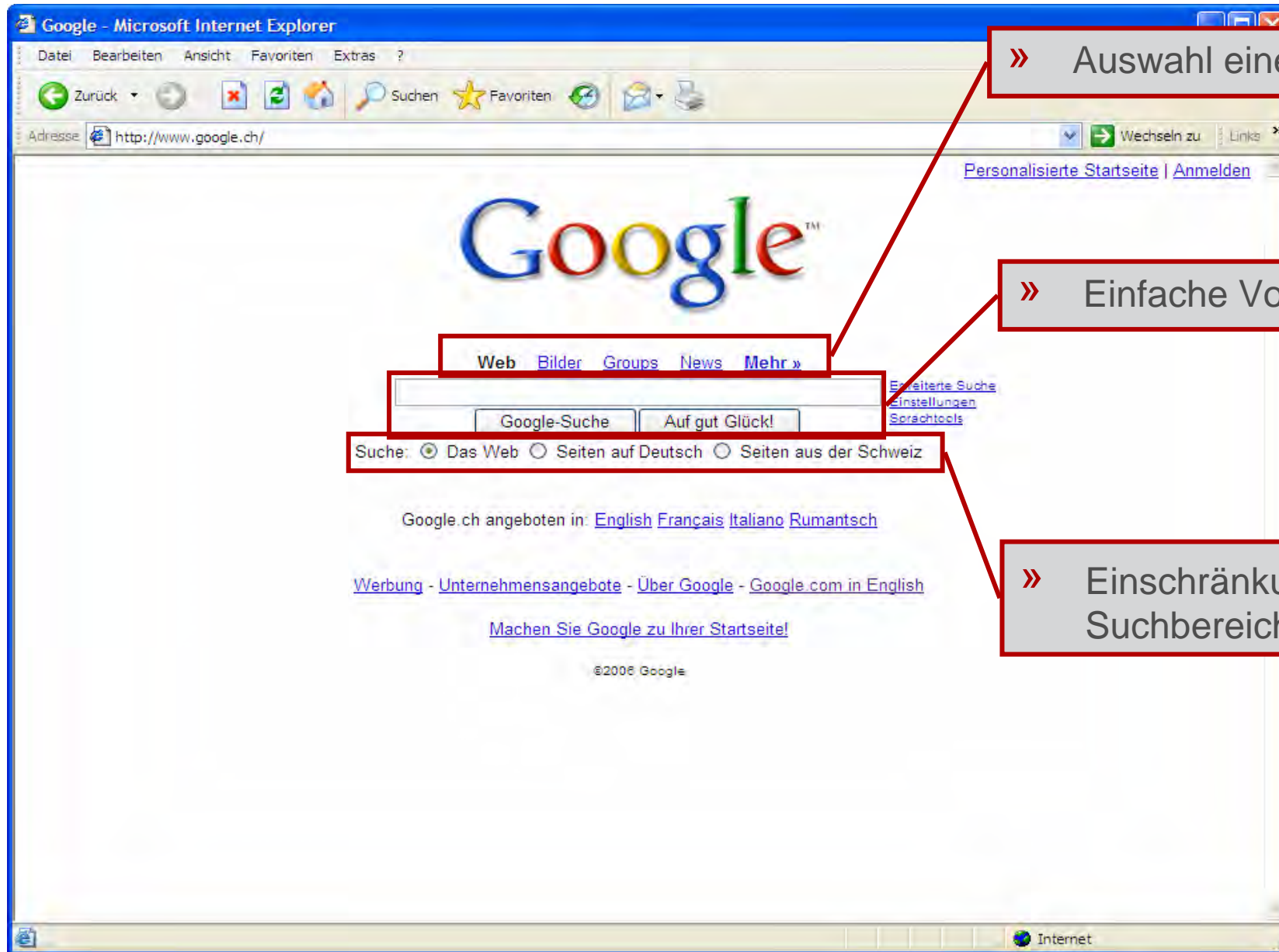
- Websites
- Webseiten



» Suche in unstrukturierten Daten nur über Volltextsuche möglich



Suche in unstrukturierten Daten - Websuche



» Auswahl einer Kategorie

» Einfache Volltextsuche

» Einschränkung des Suchbereichs

Suche in unstrukturierten Daten - Sitesuche

The screenshot shows the Microsoft Internet Explorer browser window displaying the website 'Kanton Bern'. The address bar shows 'http://www.be.ch/web/index'. The page features a search bar with a 'Suchen' button and two radio button options: 'im ganzen Kanton' (selected) and 'in diesem Bereich'. Below the search bar is a navigation menu with categories like 'Der Kanton Bern', 'nach Themen', 'nach Behörden / Organisation', and 'nach Veröffentlichungen'. The main content area includes a large landscape image, a 'Schlagzeilen' section with news items, and an 'Agenda' section with a 'Trüelete - Das Winzerfest in Twann' event. Two red callout boxes with arrows point to the search bar and the search options.

» Einfache Suchmaske

» Einschränkung des Suchbereichs



Positionierung der Sitesuche und Einschränkung

- » Suche gehört rechts oben auf jede Seite einer Website
 - seltener links unterhalb des Logos
- » Direkt ein Eingabefeld anzeigen, anstatt nur einen Link auf die Suche

- » Suchbereichs-Einschränkung
 - nur auf grossen Websites
 - mit klar unterteilbaren Bereichen
 - Default muss immer "Suche über alles" sein

Schwierigkeiten der Volltextsuche

- » Bei einem freien Eingabefeld kann der User schlecht geführt werden
- » User haben meist einen bestimmten Begriff im Kopf
 - dieser entspricht evtl. nicht der "offiziellen" Bezeichnung
- » User können die Suchanfrage schlecht variieren
 - Suchbegriff wird auch bei erfolgloser Suche selten verändert (z.B. in Synonyme oder Überbegriffe)
 - Bei 0 Treffern werden oft zusätzliche (!) Suchbegriffe eingegeben: "Ich muss genauer sagen, was ich will"
- » Bei jedem weiteren Versuch sinkt die Erfolgsquote
- » Jede Suche sollte möglichst im ersten Versuch zum Erfolg führen



First query:	51%
Second query:	32%
Third query:	18%

Quelle:
<http://www.useit.com/alertbox/20010513.html>

Anforderungen an Volltextsuche

» Hohe Toleranz

- Erkennen von Synonymen
- Automatische Rechtschreibkorrektur
- Keine Unterscheidung von Gross- und Kleinschreibung
- Gleiche Resultate für unterschiedliche Syntax (z.B. "Meier, Hans" = "Meier Hans")
- Erkennen von Operatoren (z.B. *)
 - Wildcard wird erstaunlich oft genutzt

Erweiterte Suche

- » **Erweiterte Volltextsuche** ist meistens eine Boole'sche Suche
- » in der "" sowie AND-, OR- und NOT-Verknüpfungen aufgeschlüsselt werden

mit allen Wörtern	<input type="text"/>
mit der genauen Wortgruppe	<input type="text"/>
mit irgendeinem der Wörter	<input type="text"/>
ohne die Wörter	<input type="text"/>

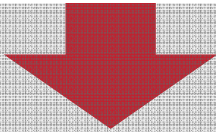
- » **Zusätzliche Einschränkungsmöglichkeiten**
(Suchbereich, Sprache, Format des Dokuments)

Nutzung der Erweiterten Suche

- » Erweiterte Suchen werden sehr selten genutzt, wenn überhaupt, erst nach erfolgloser einfacher Suche
- » Ergab die Einfache Suche 0 Treffer, wird die Erweiterte meist auch 0 Treffer ergeben
- » User denken aber oft, dass die erweiterte Suche mehr Treffer liefern müsse
- » Erweiterte (Boole'sche) Suchen werden oft nicht verstanden
- » Eine Sitesuche braucht keine "Erweiterte Suche"

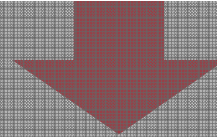
Strukturierte Daten

Unstrukturierte Daten




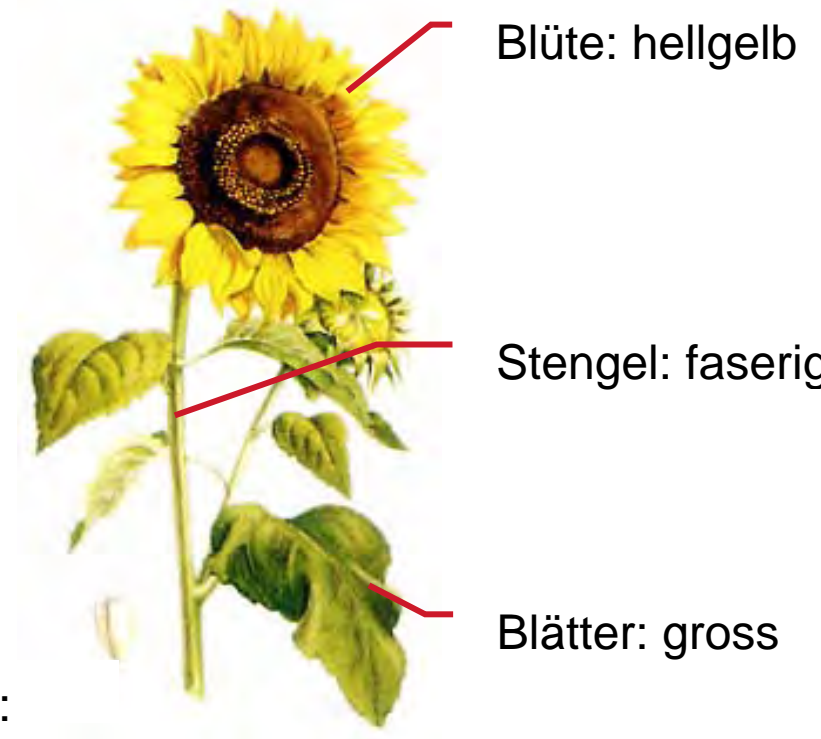
Volltextsuche

Strukturierte Daten



Strukturierte Daten

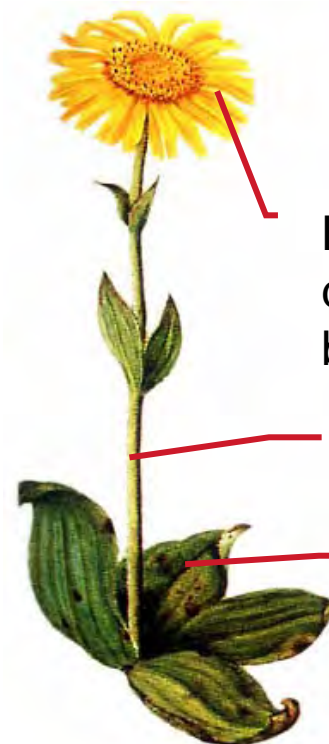
- » Alle Objekte haben einheitliche Attribute
- » Die Attribute haben eine bestimmte Anzahl Ausprägungen
- » Beispiele
 - Wohnungen
 - Autos 
 - Telefonbuch
 - etc.
- » Suche über Attributsuche



Blüte: hellgelb

Stengel: faserig

Blätter: gross



Blüte:
dunkelgelb

Stengel: gerillt

Blätter: klein

Suche in strukturierten Daten – Drill Down

The screenshot shows the homegate.ch website interface. The search criteria are as follows:

- Zimmer: 4.5 bis 5.5
- Miete: unlim. bis unlim.
- Objekt-Typ: egal
- Ort: egal
- Grösse (m2): [empty] bis [empty]
- Bezug ab: egal

The search results show 434 objects matching the criteria. The first two results are:

	Titel	PLZ Ort Adresse	Zimmer Objekttyp Etage	Bezug ab Wohnfläche	Preis
1	Sonnig - kinderfreundlich - grosszügig	8106 Adlikon b. Regensdorf Am Furtbach 17	4.5 Wohnung 1	auf Anfrage 100 m2	CHF 1'883
2	Im Grünen und doch zentral gelegen	8106 Adlikon b. Regensdorf Am Furtbach	4.5 Wohnung 1	auf Anfrage 103 m2	CHF 2'022

Suche in strukturierten Daten – Einschränkung

Amazon.com Diamond Search Results

You searched for:

- Shape:** Round, Heart, Oval, Pear, Radiant, Emerald, Marquise, Princess
- Carat:** 0.25 to 6 Carat(s)
- Cut:** Very Good to Ideal
- Color:** D to D
- Clarity:** IF to FL
- Price:** \$100.00 to \$125,000.00

62 diamonds match your criteria

Page: 1 of 3 | Next ▶

Sample Photo	Shape	Carat	Cut	Color	Clarity	Price	Compare (up to 20)	View Details/Buy
	Heart	2.52	Very Good	D	IF	\$62,905.00	<input type="checkbox"/>	View/Select
	Oval	2.25	Very Good	D	IF	\$50,255.00	<input type="checkbox"/>	View/Select
	Round	1.51	Very Good	D	IF	\$27,735.00	<input type="checkbox"/>	View/Select
	Round	1.41	Very Good	D	IF	\$23,620.00	<input type="checkbox"/>	View/Select
	Round	1.24	Ideal	D	IF	\$21,010.00	<input type="checkbox"/>	View/Select
	Oval	1.58	Very Good	D	IF	\$17,260.00	<input type="checkbox"/>	View/Select

Suche in strukturierten Daten – Attribute in Suchmaske

auto-online.ch
andere lassen wir stehen

Home
Occasionen
Auto suchen
Auto inserieren
Auto bewerten
Inserat bearbeiten
Preise
Privatperson
Garage/Händler
Werbung
Produkte
Toolbar
Videobanner
QuickInfo-private
360° Panorama
Über uns
Werbestrategie
Philosophie
Banner
Links
Rechtliches
Kontakt
Daten von eurotaxGLASS'S

VW GOLF Occasionen auf auto-online.ch

Fahrzeuge: 201
1 2 3 4 5 6 7 8 9 10 11 >>>

Suchkriterien anpassen

Modell	Jahr	Aufbau	Farbe	Kilometer	Preis
 VW Golf 1800 CL syncro	1987	Limousine	rot	170'000 km	CHF 1'500.- EUR 942.-
 VW Golf 1800 GTI	1989	Limousine	weiss	193'000 km	CHF 1'700.- EUR 1'068.-
 VW Golf 2000 GTI Edition	1996	Limousine	blau	215'000 km	CHF 1'800.- EUR 1'131.-
 VW Golf 1300 GL	1984	Limousine	braun	157'000 km	CHF 1'950.- EUR 1'225.-
 VW Golf 1800 (GL)	1987	Cabriolet	beige	140'000 km	CHF 2'500.- EUR 1'570.-
 VW Golf 1800 GTI	1988	Limousine	weiss	200'000 km	CHF 2'500.- EUR 1'570.-

Auto Ankauf
autokamela.ch
ReifenDirekt.ch

JETZT
KRANKENKASS
PRÄMIEN 20C
VERGLEICHEN U
EINE VON

Vorteile der Attributsuche

- » User wird bei der Suche in strukturierten Daten unterstützt
 - Wonach kann ich überhaupt suchen?
 - Welche Werte können die Attribute haben?
 - bzw. Wie muss ich den Suchbegriff eingeben? (Syntax)

Bestimmen Sie die Suchkriterien.

Zimmer: 4.5 bis 5.5 Miete: unlim. bis 2000

Ort: egal Grösse (m2): 150 bis

Suchen

Objekt-Typ :

- egal
- egal
- Attikawohnung
- Bauernhaus
- Chalet
- Dachwohnung
- Diverse Wohnungen
- Doppeleinfamilienhaus
- Einfamilienhaus
- Loft
- Maisonette/Duplex
- Möbl. Wohnobj.
- Reihenfamilienhaus
- Terrassenwohnung
- Villa
- Wohnung

- » Bei strukturierten Daten ist Attributsuche mit Feldern einfacher zu bedienen als Ein-Feld-Eingabemaske

4.5 Zimmer Wohnung >150m2 bis CHF 2000





Schwierigkeiten der Attributsuche und Lösungen

- » Manche User füllen immer alle Eingabefelder aus
 - Default-Wert von Attributen ist "egal" oder "alles"
 - zuerst nur die "scharfen" Attribute abfragen
 - die "unscharfen" Attribute erst als Einschränkungsmöglichkeit bei den Treffern

- » Attribute mit grosser Anzahl möglicher Werte
 - können nicht über Dropdowns, Checkboxes oder Radiobuttons abgefragt werden
 - Mögliche Alternativen
 - Clustering der Werte und sequentielle Eingabe (z.B. Produktkategorien)
 - Auswahl aus einer alphabetischen Liste (z.B. Rubrikenliste)
 - Freie Eingabe (z.B. Ortsfeld)

Anforderungen an Attributsuche

- » Keine Redundanzen in den Ausprägungen verschiedener Attribute
- » Keine Inkonsistenzen zwischen Ausprägungen eines Attributs



Kombination der Attributsuche mit der Volltextsuche

- » Oftmals ist eine Kombination von Volltextsuche und Attributsuche sinnvoll
 - weil Daten nur teilweise strukturiert sind
 - z.B. Auto mit Beschreibung
 - weil bestimmte Suchabfragen schneller über eine Volltextsuche lösbar sind (falls das gesuchte Objekt genau bekannt ist)

- » Gefahr der Kombination
 - Einige User lassen Eingabefelder nicht leer: Die Logik "wenn man nichts eingibt, sucht man nach allem" ist für viele kontraintuitiv

Annotierte Daten

Unstrukturierte Daten



Volltextsuche

Strukturierte Daten

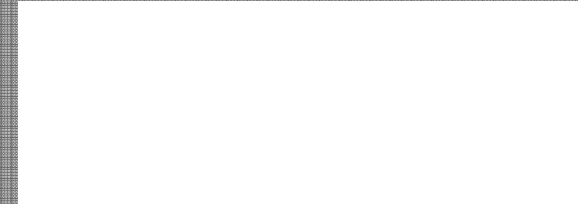
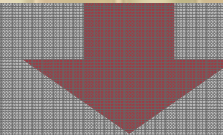


Zimmer: bis Mieta: bis

Ort: Grösse (m2): bis

Parametersuche

Annotierte Daten



Annotierte Daten

- » Mischform
- » meist unstrukturierte Daten, die manuell katalogisiert werden
- » Beispiele:
 - Bücher in Bibliothek
 - verschlagwortete Seiten einer Website
- » Suche
 - über einen Index (A – Z)
 - über Volltextsuche (die bessere Treffer liefert dank der Annotierung)



Papier

Restmüll

Verpackung

Glas

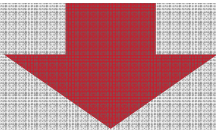
(Beispiel für semantisch inkonsistente Kategorisierung)

Anforderungen

- » Gute Synonymisierung notwendig
 - Begriff, den der User kennt, entspricht oft nicht der "offiziellen" Bezeichnung
- » Hoher manueller Initialaufwand für Verschlagwortung (Schlagworte und Stichworte)
- » Hoher manueller Initialaufwand für Erstellung geeigneter Landing Pages
- » Hoher manueller Pflegeaufwand während dem Betrieb

Daten

Unstrukturierte Daten



Volltextsuche

Strukturierte Daten



Zimmer: bis Mieta: bis

Ort: Grösse (m2): bis

Parametersuche

Annotierte Daten

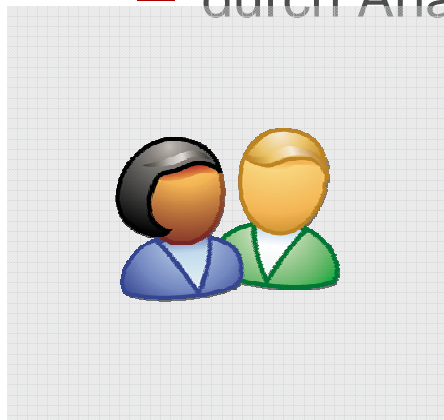


Staatskanzlei	Grosser Rat
Staatsrechnung des Kantons Bern	Regierungsrat
Statistik	Finanzplan
Stellenmarkt des Kantons Bern	
Stellvertretungszentrale für Lehrkräfte	
Steuererklärung online	

Alphabetischer oder thematischer Index

Merke

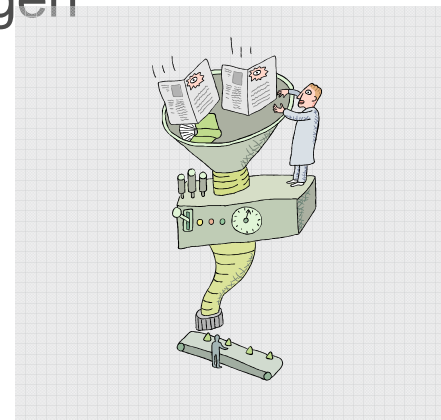
- » Suche ≠ Google
- » Strukturierten Daten verlangen nach einer strukturierten Eingabe
- » Eine gute Suche basiert auf guten Daten
- » Was die wichtigen Attribute und die richtigen Begriffe sind, erfährt man am besten
 - durch Beobachtung von Usern
 - durch Analyse der Suchanfragen



User



User Interface



IR-Applikation

unstrukturiert	
strukturiert	
annotiert	

Daten