

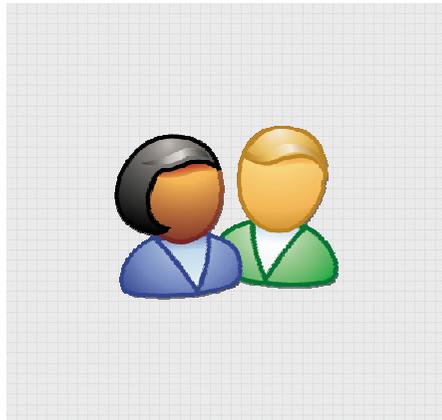
Einfach, erweitert und die anderen  
(Referat 2a)



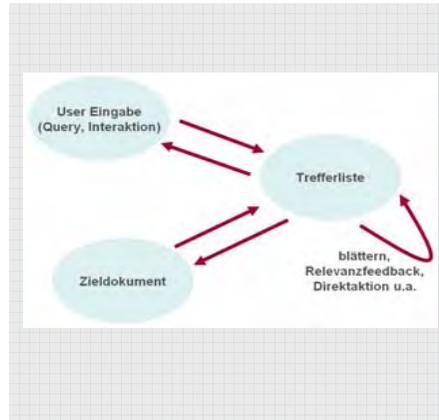
Fachtagung: Suchfunktionen im Web  
Zürich, 26. Oktober 2006

Jacqueline Badran, Zeix  
Gregor Urech, Zeix

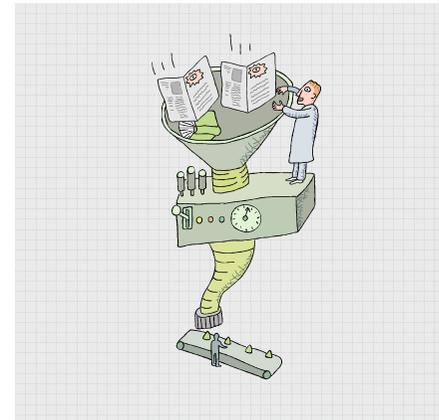
# Relevante Dimensionen: Übersicht



User



User Interface

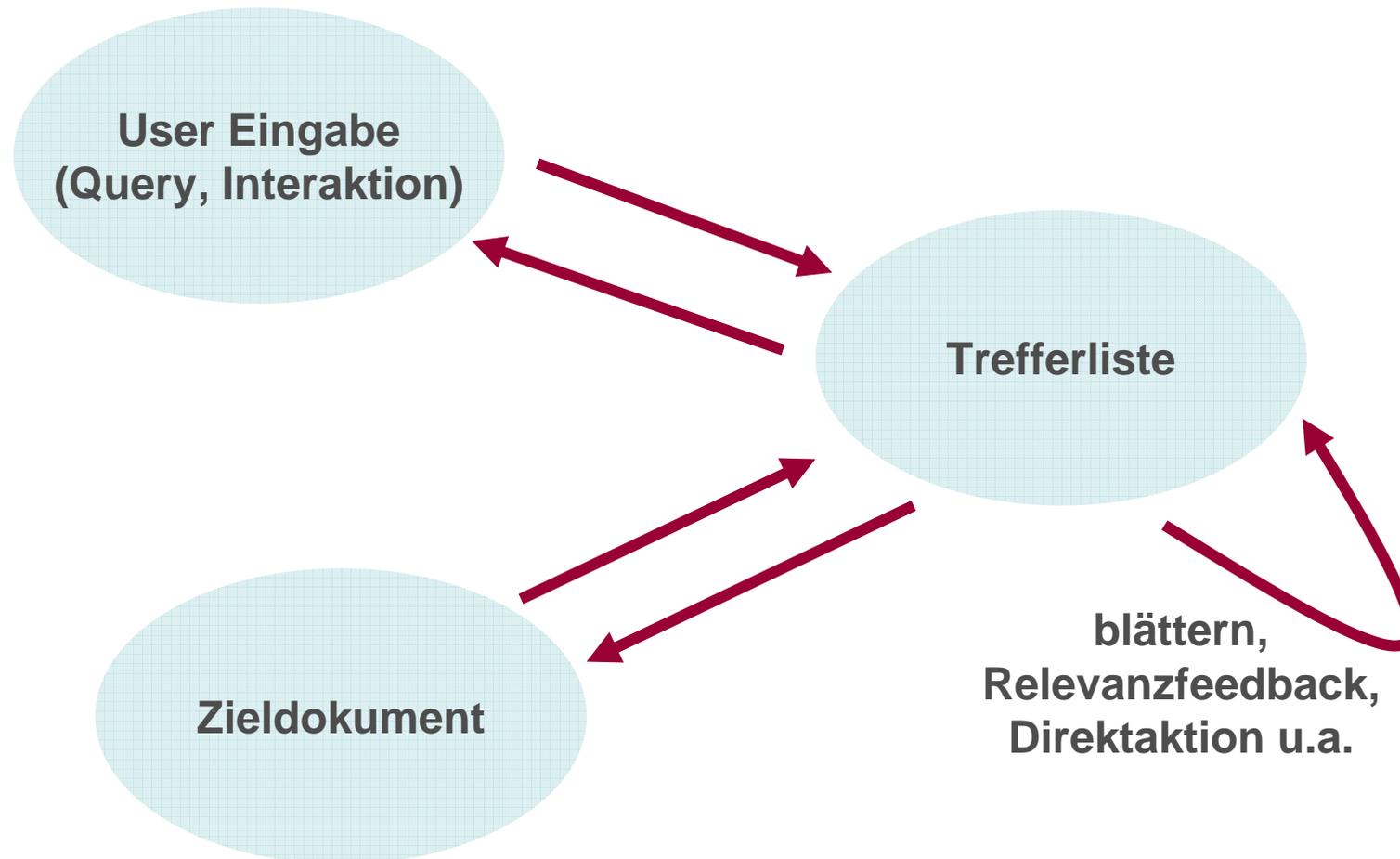


IR-  
Applikation

unstrukturiert	
strukturiert	
annotiert	

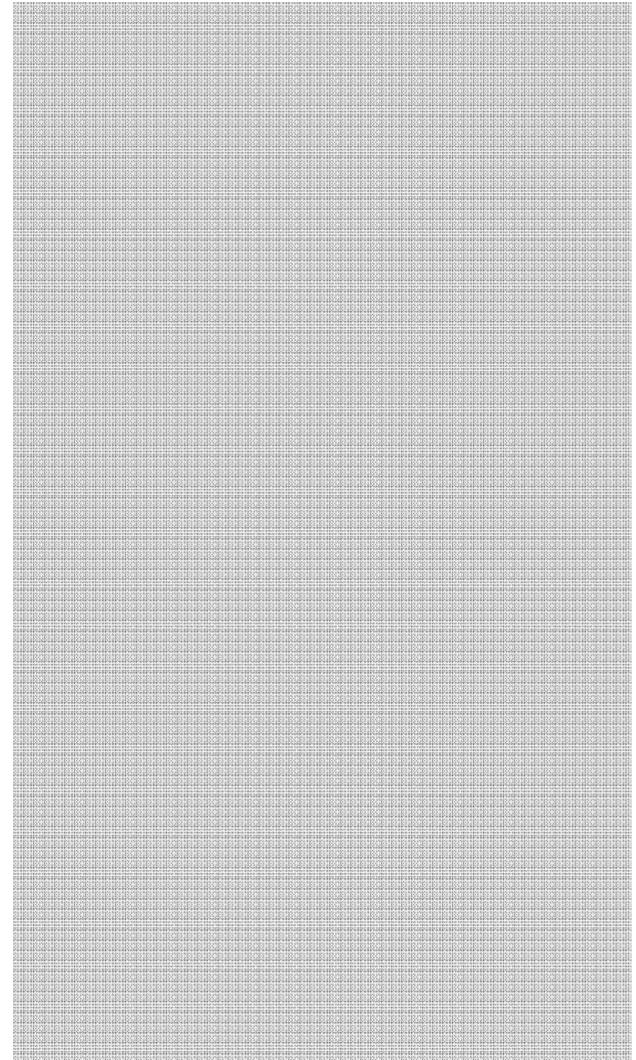
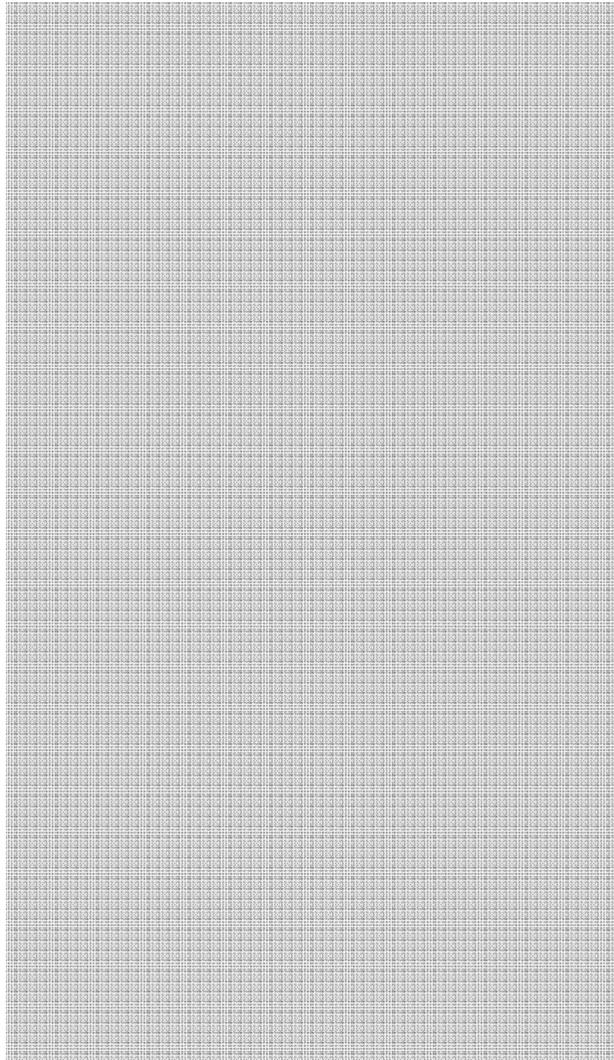
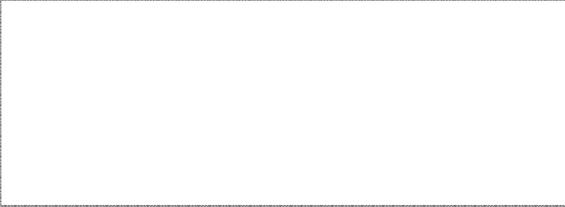
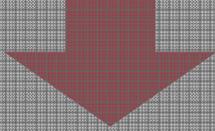
Daten

# Detail User Interface



# Unstrukturierte Daten

Unstrukturierte Daten



# Unstrukturierten Daten

» Einzelne Objekte haben keine einheitlichen Attribute

» Beispiele

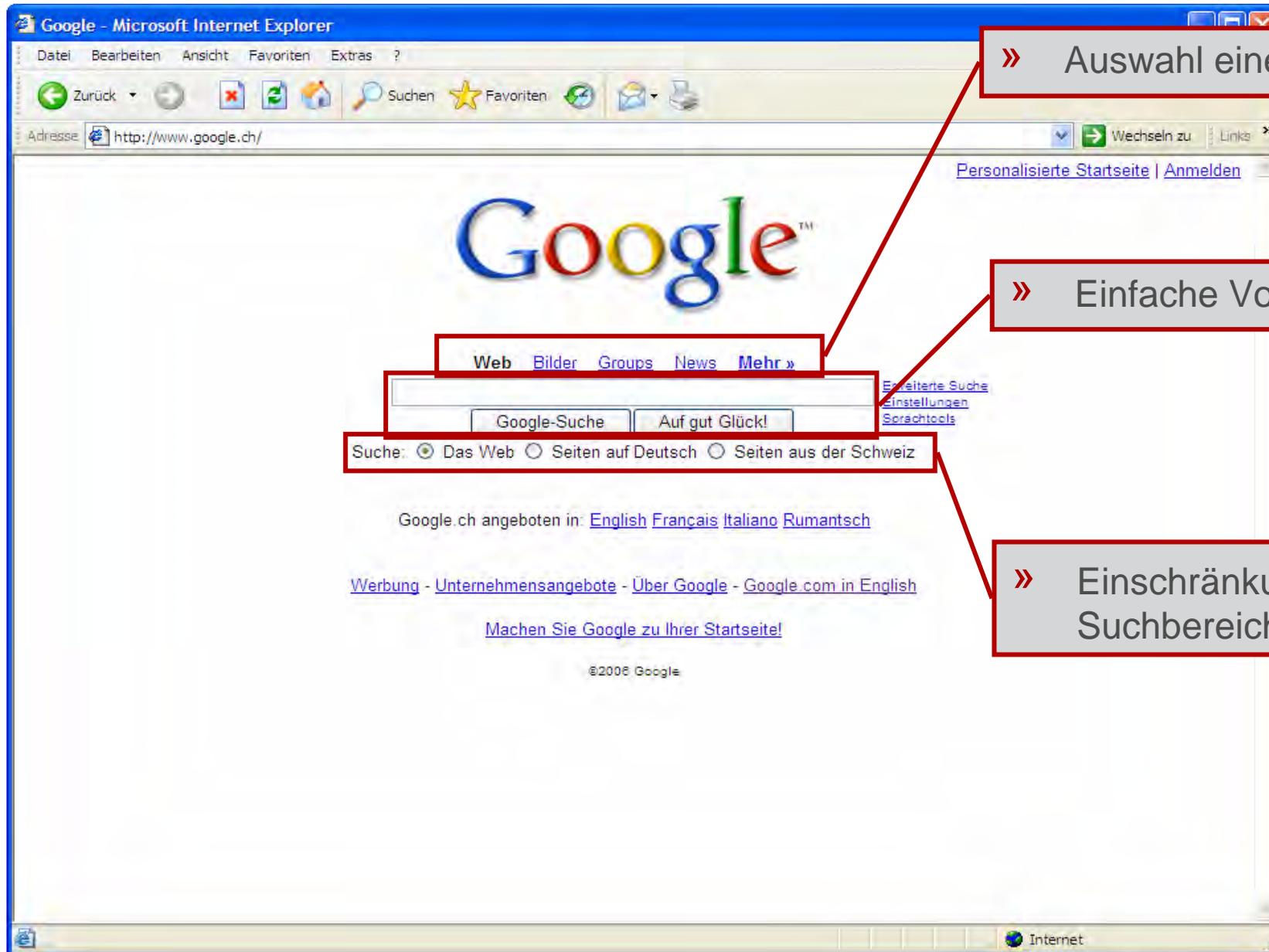
- Websites
- Webseiten



» Suche in unstrukturierten Daten nur über Volltextsuche möglich



# Suche in unstrukturierten Daten - Websuche



# Suche in unstrukturierten Daten - Sitesuche

The screenshot shows the Microsoft Internet Explorer browser window displaying the website 'Kanton Bern'. The browser's address bar shows 'http://www.be.ch/web/index'. The website header includes the logo of the Canton of Bern and the text 'Kanton Bern Canton de Berne'. Below the header, there are navigation links and a search bar. The search bar is highlighted with a red box and a callout box that says 'Einfache Suchmaske'. Below the search bar, there are two radio buttons for search scope: 'im ganzen Kanton' (selected) and 'in diesem Bereich'. This area is also highlighted with a red box and a callout box that says 'Einschränkung des Suchbereichs'. The main content area is divided into several sections: 'Der Kanton Bern', 'nach Themen', 'nach Behörden / Organisation', and 'nach Veröffentlichungen'. Below these sections, there is a large image of a landscape, a section titled 'Schlagzeilen' with a list of news items, and an 'Agenda' section with a calendar icon and a link to 'Trüelete - Das Winzerfest in Twann'. The browser's status bar at the bottom shows the current URL and the 'Internet' icon.

» Einfache Suchmaske

Suchen  
 im ganzen Kanton  
 in diesem Bereich

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

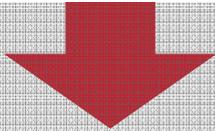
- » **Zusätzliche Einschränkungsmöglichkeiten**  
(Suchbereich, Sprache, Format des Dokuments)

## 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"

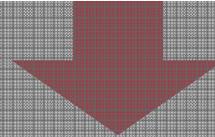
# Strukturierte Daten

Unstrukturierte Daten



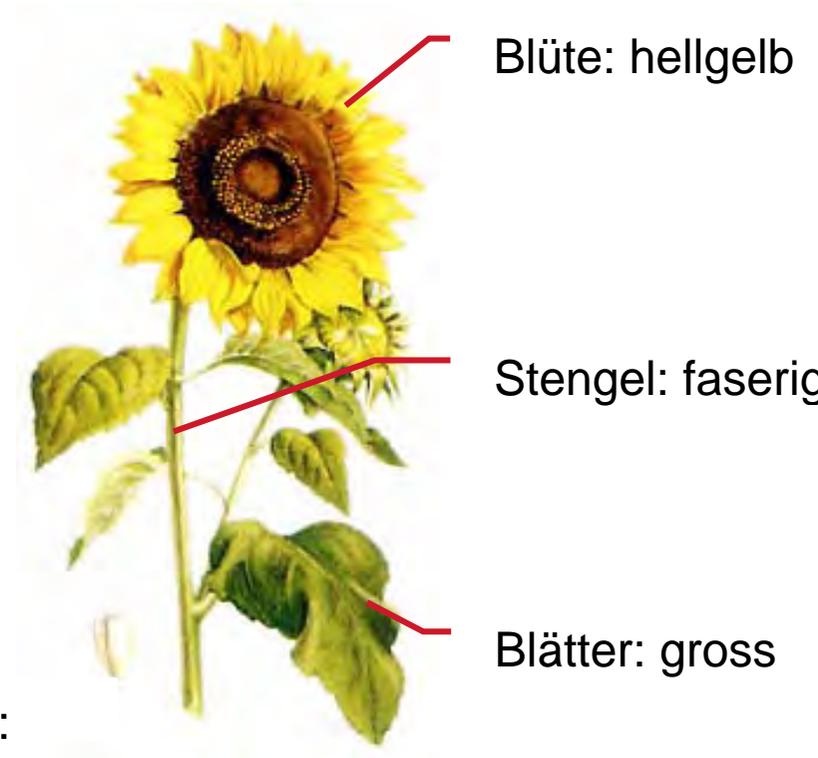
Volltextsuche

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"/>	<a href="#">View/Select</a>
	Oval	2.25	Very Good	D	IF	\$50,255.00	<input type="checkbox"/>	<a href="#">View/Select</a>
	Round	1.51	Very Good	D	IF	\$27,735.00	<input type="checkbox"/>	<a href="#">View/Select</a>
	Round	1.41	Very Good	D	IF	\$23,620.00	<input type="checkbox"/>	<a href="#">View/Select</a>
	Round	1.24	Ideal	D	IF	\$21,010.00	<input type="checkbox"/>	<a href="#">View/Select</a>
	Oval	1.58	Very Good	D	IF	\$17,260.00	<input type="checkbox"/>	<a href="#">View/Select</a>

# Suche in strukturierten Daten – Attribute in Suchmaske

The screenshot shows a Microsoft Internet Explorer window displaying the website **auto-online.ch**. The search results are for "VW GOLF Occasionen auf auto-online.ch". The search criteria are: Fahrzeug: 201, Seite: 1/11. The results table is as follows:

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.-
 no image	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.-

The left sidebar contains navigation links such as Home, Occasionen, Preise, and Über uns. The right sidebar contains advertisements for "Auto Ankauf", "ReifenDirekt.ch", and "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

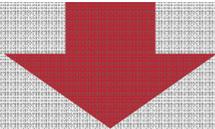


## Kombination der Attributsuche mit der Volltextsuche

- » 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

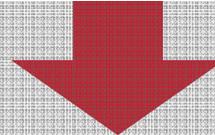
# Annotierte Daten

## Unstrukturierte Daten



## Volltextsuche

## Strukturierte Daten

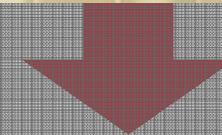


Zimmer:  bis  Mieta:  bis

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

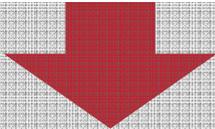
(Beispiel für semantisch inkonsistente Kategorisierung)

# 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

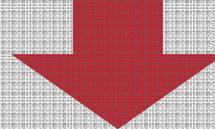
# Daten

## Unstrukturierte Daten



## Volltextsuche

## Strukturierte Daten

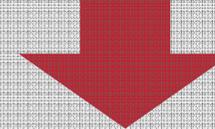


Zimmer:  bis  Mieta:  bis

Ort:  Grösse (m2):  bis

## Parametersuche

## Annotierte Daten

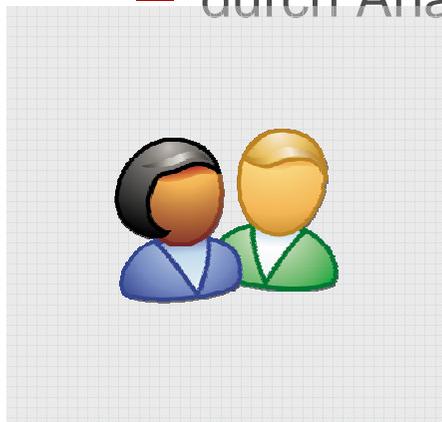


<a href="#">Staatskanzlei</a>	<a href="#">Grosser Rat</a>
<a href="#">Staatsrechnung des Kantons Bern</a>	<a href="#">Regierungsrat</a>
<a href="#">Statistik</a>	<a href="#">Finanzplan</a>
<a href="#">Stellenmarkt des Kantons Bern</a>	
<a href="#">Stellvertretungszentrale für Lehrkräfte</a>	
<a href="#">Steuererklärung online</a>	

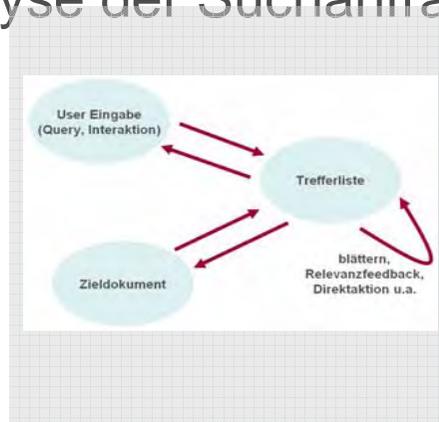
## Alphabetischer oder thematischer Index

# Merke

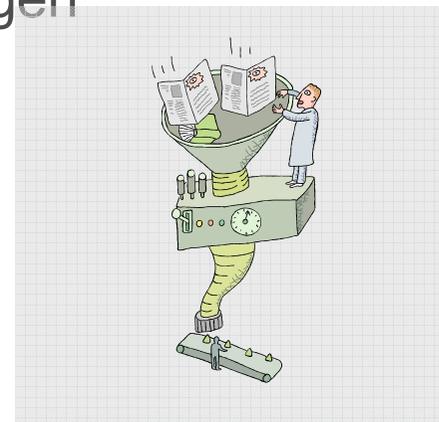
- » Suche ≠ Google
- » Strukturierten Daten verlangen nach einer strukturierten Eingabe
- » Eine gute Suche basiert auf guten Daten
- » Was die wichtigen Attribute und die richtigen Begriffe sind, erfährt man am besten
  - durch Beobachtung von Usern
  - durch Analyse der Suchanfragen



User



User Interface



IR-Applikation

unstrukturiert	
strukturiert	
annotiert	

Daten