

Die treffende Auswahl anbieten: Im Intranet (Referat 3b)

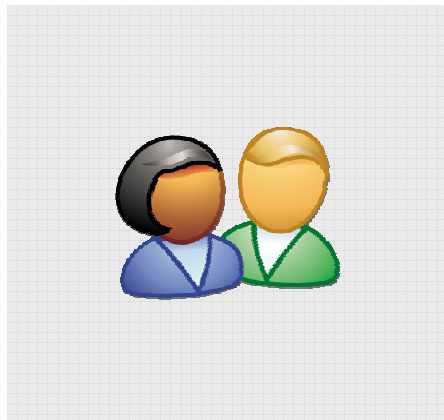


Fachtagung: Suchfunktionen im Web
Zürich, 26. Oktober 2006

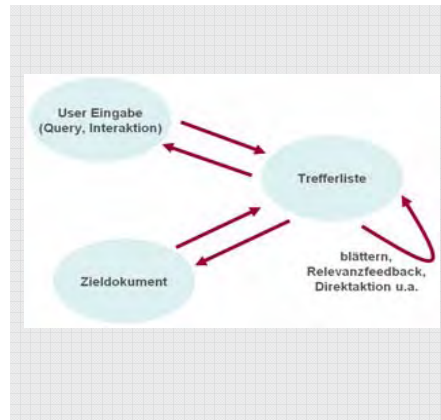
Marco Hassler, namics
Jacqueline Badran, Zeix

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

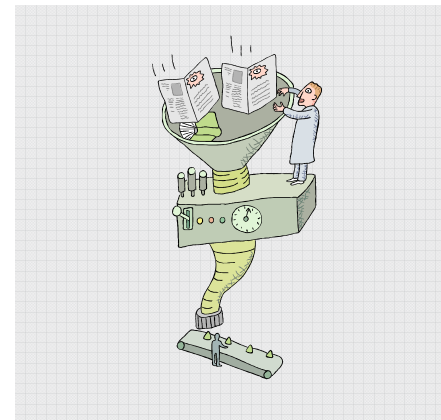
Relevante Dimensionen: Übersicht



User



User Interface

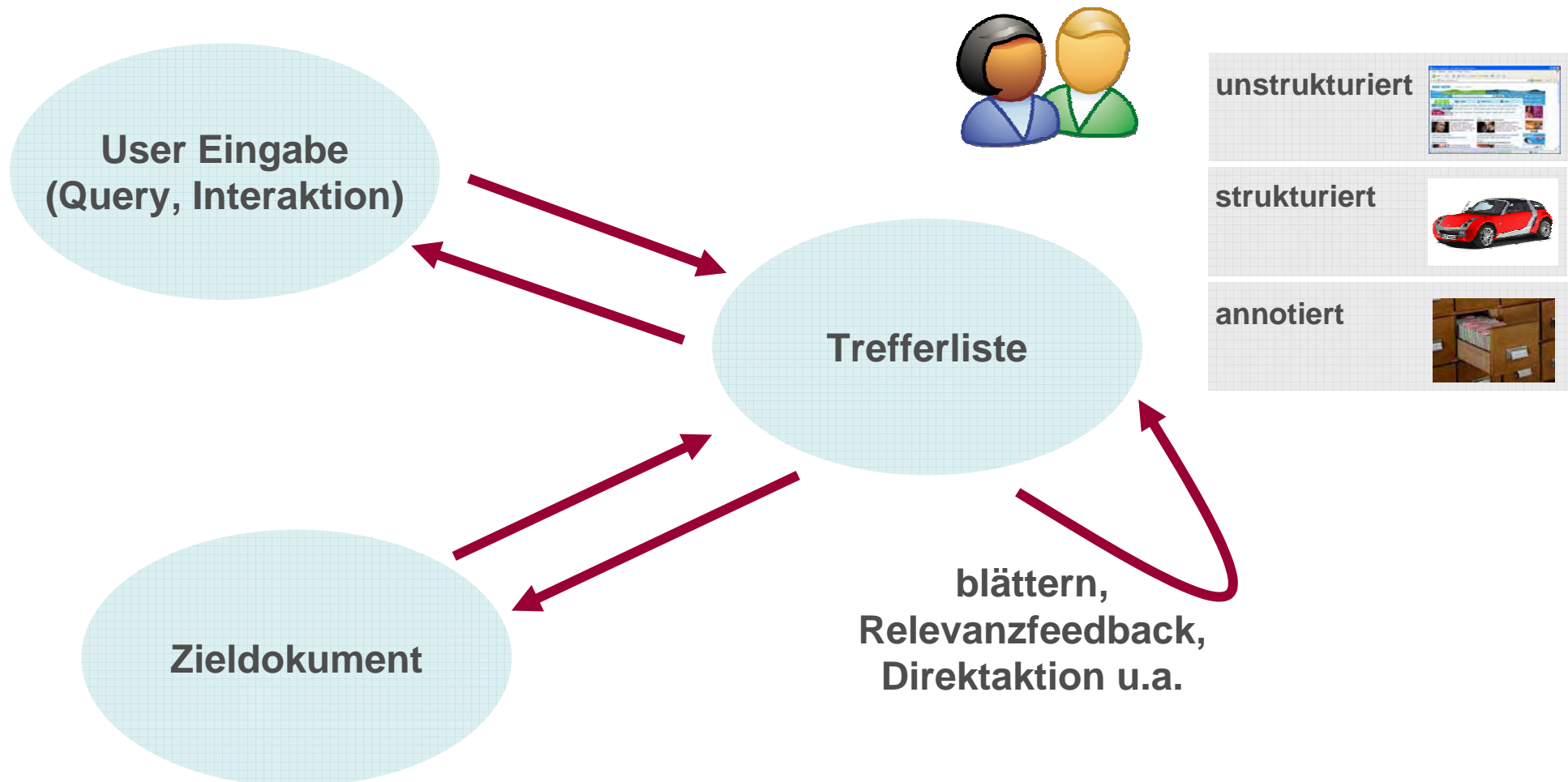


**IR-
Applikation**

unstrukturiert	
strukturiert	
annotiert	

**Daten
(Typen, Qualität)**

Schema: Detail User Interface





Hintergrund: Suchverhalten von Usern im Intranet

» Dominante Zugänge

- Navigation
- Stichwortverzeichnisse, Kataloge, alphabetische Listen
- Suche meist sekundär
- Bookmarks (oft im Browser!)
- Contentlinks
- oft organisatorischer Zugang über z.B. Abteilungen

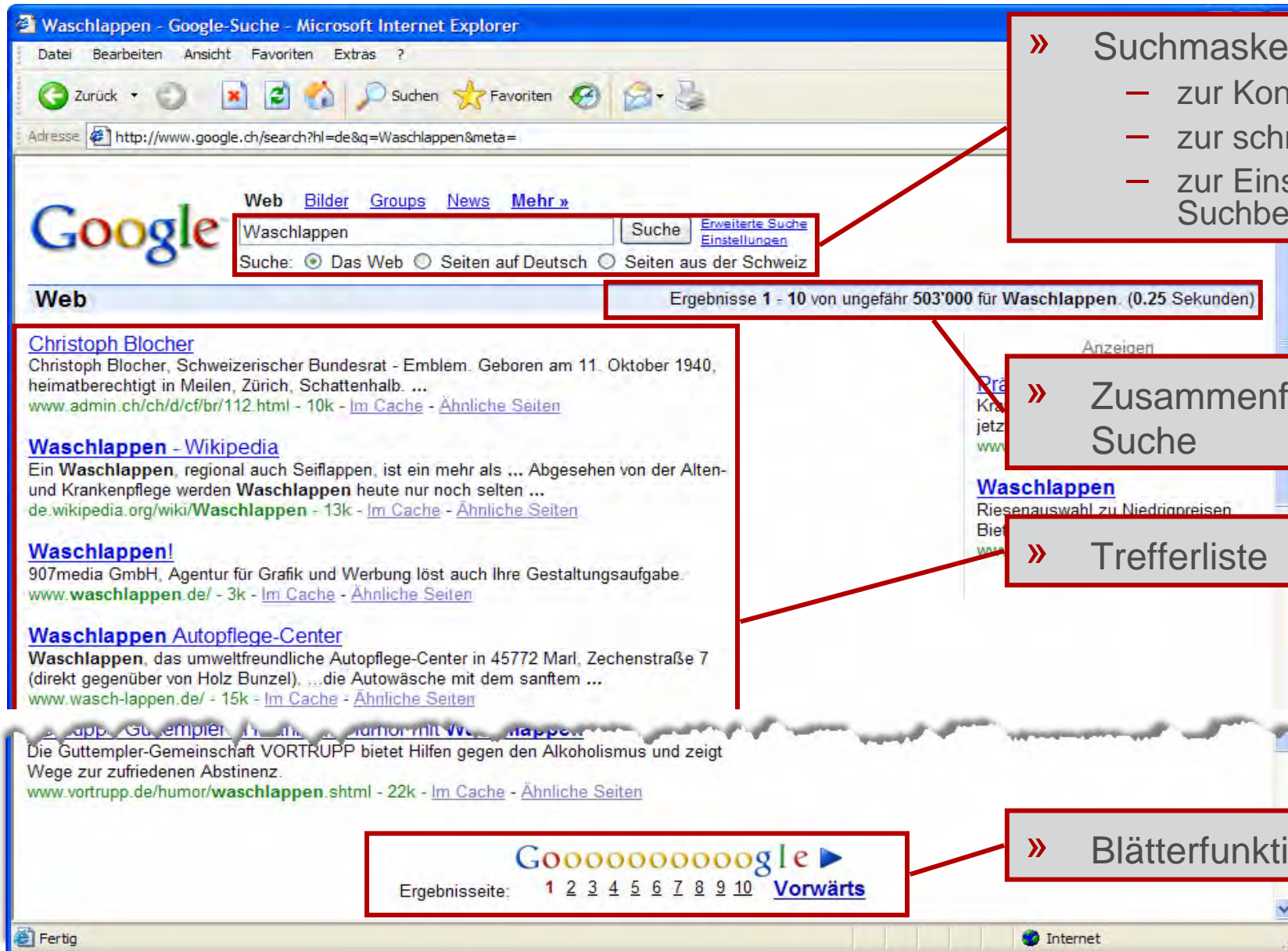
» Suchstrategien

- Wiederholung eines als erfolgreich angesehenen Wegs
- Bestehendes Wissen wird verwendet
- Wenig Experimentierfreude (sehr geringe Lernkurve)
- Geringe Frustrationstoleranz (schneller Griff zum Telefon)
- Kein Treffer: Mehr Kriterien eingeben!

» Informationsbedürfnis

- Sehr spezifisch: Mensa-Menü, spezifisches Dokument, Person etc.

Trefferseite – Websuche (unstrukturierte Daten)



» Suchmaske

- zur Kontrolle der Eingabe
- zur schnellen Iteration
- zur Einschränkung des Suchbereichs

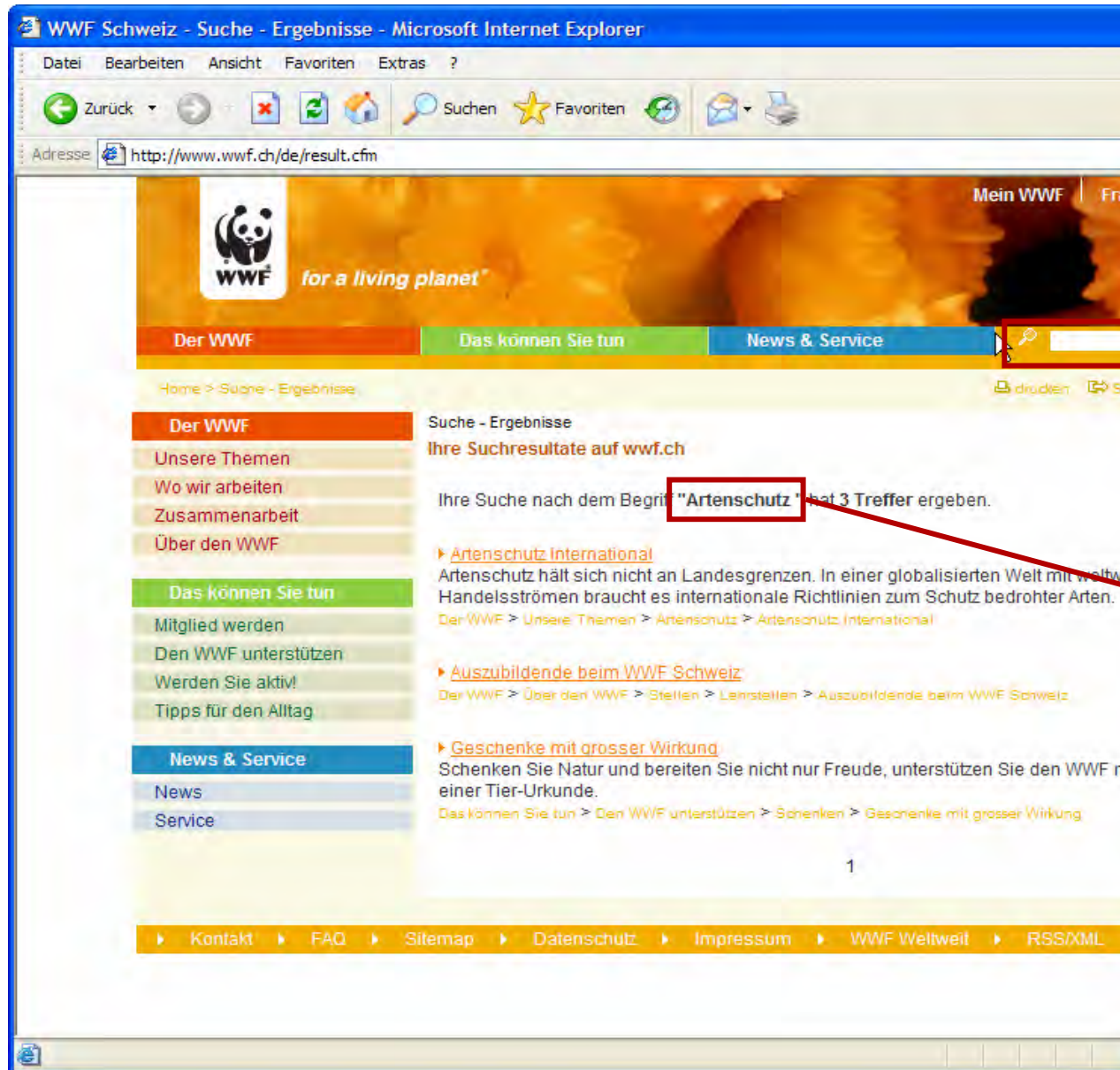
Ergebnisse 1 - 10 von ungefähr 503'000 für Waschlappen. (0.25 Sekunden)

» Zusammenfassung der Suche

» Trefferliste

» Blätterfunktion

Trefferseite – Sitesuche (annotierte Daten)



- » Gleiche Elemente wie bei Websuche:
 - Suchmaske
 - Zusammenfassung
 - Trefferliste
- » Könnte mehr Elemente haben:
 - Gruppierung
 - Sortierfunktionen

- » Hier schlecht umgesetzt:
 - Suchbegriff wird nicht wiederholt im Suchfeld
 - Klar schlechter sichtbar als im Suchfeld

Trefferseite – Shop (strukturierte Daten)

Amazon.com: James Bond: DVD - Microsoft Internet Explorer

Address: http://www.amazon.com/s/ref=nb_ss_gw/102-9462007-5749738?url=search-alias%3Ddvd&field-keywords=James+Bond

amazon.com Your Amazon.com DVD See All 35 Product Categories Your Account | Cart | Your Lists

Advanced Search | Browse Genres | Top Sellers | New & Future Releases | Television Central | Life & Learning | DVD Essen

Search DVD James Bond GO Find Gifts

DVD

"James Bond" (Related Searches: casino royale, from russia with love, superman)

Narrow or Expand Results Showing 1 - 12 of 164 Results Sort by Relevance GO

Expand Your Results Remove Keywords: James Bond

Narrow by Category

- Action & Adventure (62)
- Animation (2)
- Anime & Manga (1)
- Art House & International (24)
- Classics (52)
- Comedy (28)
- Documentary (4)
- Drama (42)
- Educational (1)
- Horror (3)
- Kids & Family (8)
- Military & War (8)
- Music Video & Concerts (2)
- Musicals & Performing Arts (1)
- Mystery & Suspense (21)

Narrow by Condition

- Used
- New
- Collectible

Browse our James Bond store in DVD

- James Bond Ultimate Collection, Vol. 1 (10 pc) (The Man with the Golden Gun / Goldfinger / The World Is Not Enough / Diamonds Are Forever / The Living Daylights)** ~ Roger Moore, Christopher Lee, Britt Ekland, and Maud Adams (DVD - 1974)

Buy new: ~~\$89.98~~ **\$58.49**

Available for Pre-order
- The James Bond Collection, Vol. 1 (Special Edition)**

Used & new from **\$56.99**
- James Bond Ultimate Collection, Vol. 2 (10 pc) (A View to a Kill / Die Another Day / The Spy Who Loved Me / License to Kill)** Christopher Walken, Tanya Roberts, and Grace Jones (DVD - 1985)

Buy new: ~~\$89.98~~ **\$58.49**

Available for Pre-order
- James Bond Ultimate Collection Vol. 3 (10pc) (Goldeneye / Live and Let Die / For Your Eyes Only / From Russia With Love / On Her Majesty's Secret Service)**

http://www.amazon.com/Ultimate-Collection-Goldeneye-Majestys-Service/dp/B0001CMSV2/sr=1-4/qid=1161718782/ref=sr_1_4/102-94

» Gleiche Elemente wie bei Websuche

- Suchmaske
- Zusammenfassung
- Trefferliste

» Zusätzlich:

- Einschränkungen nach Kategorie und nach Zustand
- Einschränkung nach Bereich (Objektarten)

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
- » Und... zeigen sie auch an, was sie nicht haben, wenn es dem Informationsbedürfnis ihrer User entspricht:
 - z.B. Im Materialshop – nicht mehr erhältlich;
 - z.B. Abteilung nicht mehr zuständig, neu Abteilung XY

Trefferseite Intranet (exemplarisch)

» auf jeder Seite

- Telefonbuch
- Sitesuche

The screenshot shows an Intranet search results page. At the top, there is a navigation bar with a menu (Datei, Bearbeiten, Ansicht, Favoriten, Extras, ?) and a toolbar with icons for navigation and search. The address bar shows 'http://intranet.firma.com'. Below the address bar, there are two search boxes: 'Telefonbuch' with fields for 'Name' and 'Abteilung', and 'Stichwortsuche' with a search field and radio buttons for 'ganze Site' and 'diesen Bereich'. A horizontal menu below these boxes contains 'Bereich A' through 'Bereich E'. The main content area has two tabs: 'Volltext' and 'Telefonbuch'. Under 'Telefonbuch', there is a search form with a text input containing 'Geldwäscherei', a dropdown for 'in allen Bereichen', a 'Suchen' button, and a 'Suchabfrage als Abo' button. Below this are radio buttons for 'alle Dokumente', 'nur Weisungen', 'nur Anleitungen', and 'nur Formulare'. The search results show 'Treffer 1-10 von 367' with pagination links. The results are grouped into sections: 'Weisungen' and 'Anleitungen'. Each section has a table with columns for 'Dokument', 'Autor', 'Datum', and 'Dateityp'. The first result in 'Weisungen' is '1 Titel des Dokuments' by Hans Meier, dated 26.10.2006, in PDF format. The second result is '2 Titel des Dokuments' by Andreas Müller, dated 18.09.2006, in PPT format. The first result in 'Anleitungen' is '3 Titel des Dokuments' by Klaus Amsteg, dated 24.10.2006, in PDF format. A red box highlights the search filters, and another red box highlights the search results table.

» Wahl der Kategorie (Objekttyp)

» Suchmaske

- Bereichseinschränkung
- Dateityp
- Kategorien
- Funktion Suchabo

Trefferseite Intranet (exemplarisch)

The screenshot shows a web browser window titled 'Intranet' with the address 'http://intranet.firma.com'. The page features a search interface with a search bar containing 'Geldwäscherei' and a dropdown menu set to 'allen Bereichen'. Below the search bar, there are radio buttons for document types: 'alle Dokumente' (selected), 'nur Weisungen', 'nur Anleitungen', and 'nur Formulare'. A pagination bar shows 'Treffer 1-10 von 367' with navigation links. The results are grouped into sections: 'Weisungen' and 'Anleitungen'. Each section contains a table with columns for 'Dokument', 'Autor', 'Datum', and 'Dateityp'. The first result under 'Weisungen' is '1 Titel des Dokuments' by Hans Meier, dated 26.10.2006, in PDF format. The second result is '2 Titel des Dokuments' by Andreas Müller, dated 18.09.2006, in PPT format. The 'Anleitungen' section shows the start of a third result by Klaus Amsteg, dated 24.10.2006, in PDF format.

» Zusammenfassung und Blätterfunktion

» Trefferliste

- Gruppierung nach Kategorien
- Sortierbar nach verschiedenen Attributen

Treffer – annotierte Daten (exemplarisch)

- » Titel
- » Autor
- » Datum
- » Dateityp

2 Titel des Dokuments	Andreas Müller	18.09.2006	PPT
Beschreibung Beschreibung Beschreibung Beschreibung Beschreibung Beschreibung Beschreibung Beschreibung Beschreibung Beschreibung Beschreibung			
Bereich C > Themen > Wirtschaftskriminalität > Titel des Dokuments			

- » Beschreibung
 - primär: redaktionell gepflegter Text (Description)
 - alternativ: Zitat aus Dokument

- » Pfad
 - wo liegt das Dokument?

Treffer – strukturierte Daten (exemplarisch)

Telefonbuch					
Vorname	Name	Telefon	E-Mail	Abteilung	Standort
Peter	Meierhans	044 255 25 55 079 555 55 55	peter.meierhans@firma.com	XY	Zürich WH1
Andreas	Zingg	061 260 25 95 079 511 00 00	andreas.zingg@firma.com	Z7	Basel RO3

- » Anzeige der wichtigsten Attribute
 - Sortierbar nach verschiedenen Attributen
 - E-Mail und evtl. Telefonnummern direkt verlinkt
- » Name verlinkt auf Detailseite

Wie sieht der optimale Suchtreffer aus?

- » Ziel: User muss möglichst schnell erkennen können, ob ein Treffer für ihn relevant ist
 - d.h. sein Informationsbedürfnis teilweise oder ganz befriedigt

- » bei **unstrukturierten** Daten
 - kurzer, sprechender Titel
 - kurze Beschreibung der Seite (aus Text)
 - Quelle
 - Fileformat
 - Autor und Datum

- » bei **annotierten** und **strukturierten** Daten
 - Anzeige der wichtigsten Attribute
 - Im Intranet immer: Titel, Autor, Datum, Dokumentart, Pfad (Breadcrumb), Kurzbeschreibung, (bei PDF, doc.etc. Seitenzahl)
 - z.B. Preis, Verfügbarkeit bei einem Intranetshop, Material

Die häufigsten Fehler bei der Trefferdarstellung

» Nichts aussagende Titel

- interne Begriffe
- Filenamen
- URLs

2. [ind7](#)
 Alle Indikatoren Öffentliche Entwicklungshilfe Bedeutung des Indikators Die Armutsbekämpfung in Entwicklungs- und Transitionsländern ist ein integrales Element der Strategie der Nachhaltigen Entwicklung und für deren Erfolg unabdingbar (Postulat 6a). De... (06.07.2006 - HTML - 60k)
<http://www.bfs.admin.ch/bfs/porta/de/ind.../ind7.indicator.70701.html>

» Schlechte Beschreibung

- z.B. Marketingtalk anstatt Information

» Darstellung der Relevanz eines Treffers

	Format	Wichtigkeit
CS - Finanzierungsablauf Sechs Schritte bis zur Auszahlung Ihrer Hypothek. Wir begleiten Sie von der ersten Kontaktnahme bis zur Auszahlung der Hypothek.	HTML	69%

» Schlechte grafische Aufbereitung

...> UBS in der Schweiz > Privatkunden > Hypotheken > Finanzierungsablauf
 → Finanzierungsablauf bei UBS Hypotheken in der Schweiz
 Mit UBS Hypotheken in sechs Schritten zum Eigenheim.
 25 Okt. 2004

Spezialfall 0 Treffer (exemplarisch)

Volltext **Telefonbuch**

Geldwätcherei in allen Bereichen

alle Dateitypen

alle Dokumente nur Weisungen nur Anleitungen nur Formulare

Keine Treffer zu **Geldwätcherei**

Meinten Sie: [Geldwäscherei](#) (37 Treffer)

» Wiederholung der Suchmaske

» Zusammenfassung der Suche

» Tipps zur Optimierung der Suche

- nur direkt verlinkte Tipps werden genutzt

Volltext **Telefonbuch**

Geldwäscherei in allen Bereichen

alle Dateitypen

alle Dokumente nur Weisungen nur Anleitungen nur Formulare

Keine Treffer zu **Geldwäscherei in Anleitungen**

Tipp: [Suchen Sie in allen Dokumenten](#) (21)

» Tipps zur Optimierung der Suche

- bei eingeschränktem Suchbereich

Rangierungsverfahren

- » Ranking = Verfahren um eine Menge von Elementen in eine Rangfolge / Reihenfolge zu bringen...
- » Häufig gehörte Aussage: *"Unsere Intranet-Suche sollte so gut sein wie Google"*

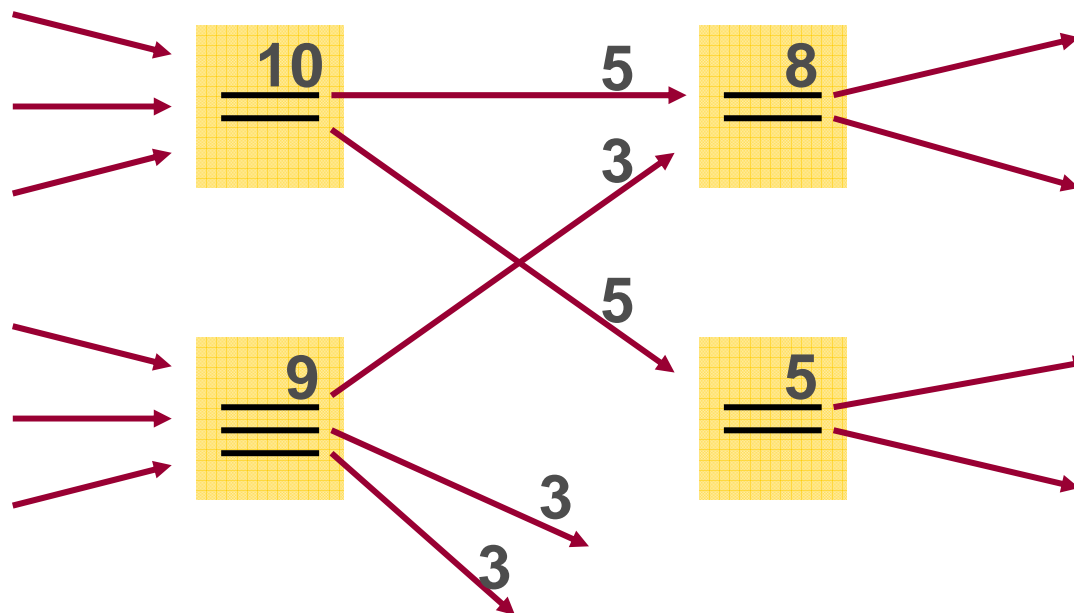


» Aber:

- Grundsätzlicher Unterschied Intranet und Google
 - Im Intranet heisst der Such-Case: „Suche genau das Dokument, das vor x Monaten dieses Thema behandelte“
 - Bei Google heisst der Such-Case: Such mir irgendein Dokument (von evtl. Tausenden), welches die Information enthält
- Warum Google so gut ist:
 - Page-Rank-Verfahren
 - Nutzungsmessung

Google.com: PageRank-Verfahren

- » Je häufiger eine Seite von einer anderen (wichtigen) Seite verlinkt ist, desto wichtiger ist diese
 - Reputation von wissenschaftlichen Papers über Zitate
 - Zufälliger Surfer



Literatur

- Page, Brin, Motwani, Terry. The PageRank Citation Ranking: Bringing Order to the Web. 1999.
- Kleinberg. Authoritative sources in a hyperlinked environment. 1998.

- » Linkkontext der abgehenden Seite wird der Zielseite zugerechnet

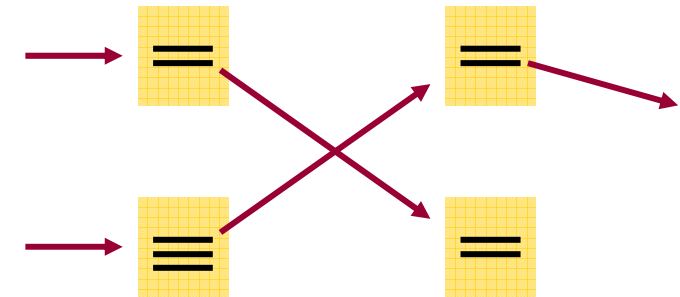
Google.com: Nutzungshäufigkeiten



Anwendbarkeit im Intranet

» PageRank

- Intranet-Seiten sind wesentlich weniger verlinkt als Internet-Seiten
→ PageRank funktioniert weniger gut

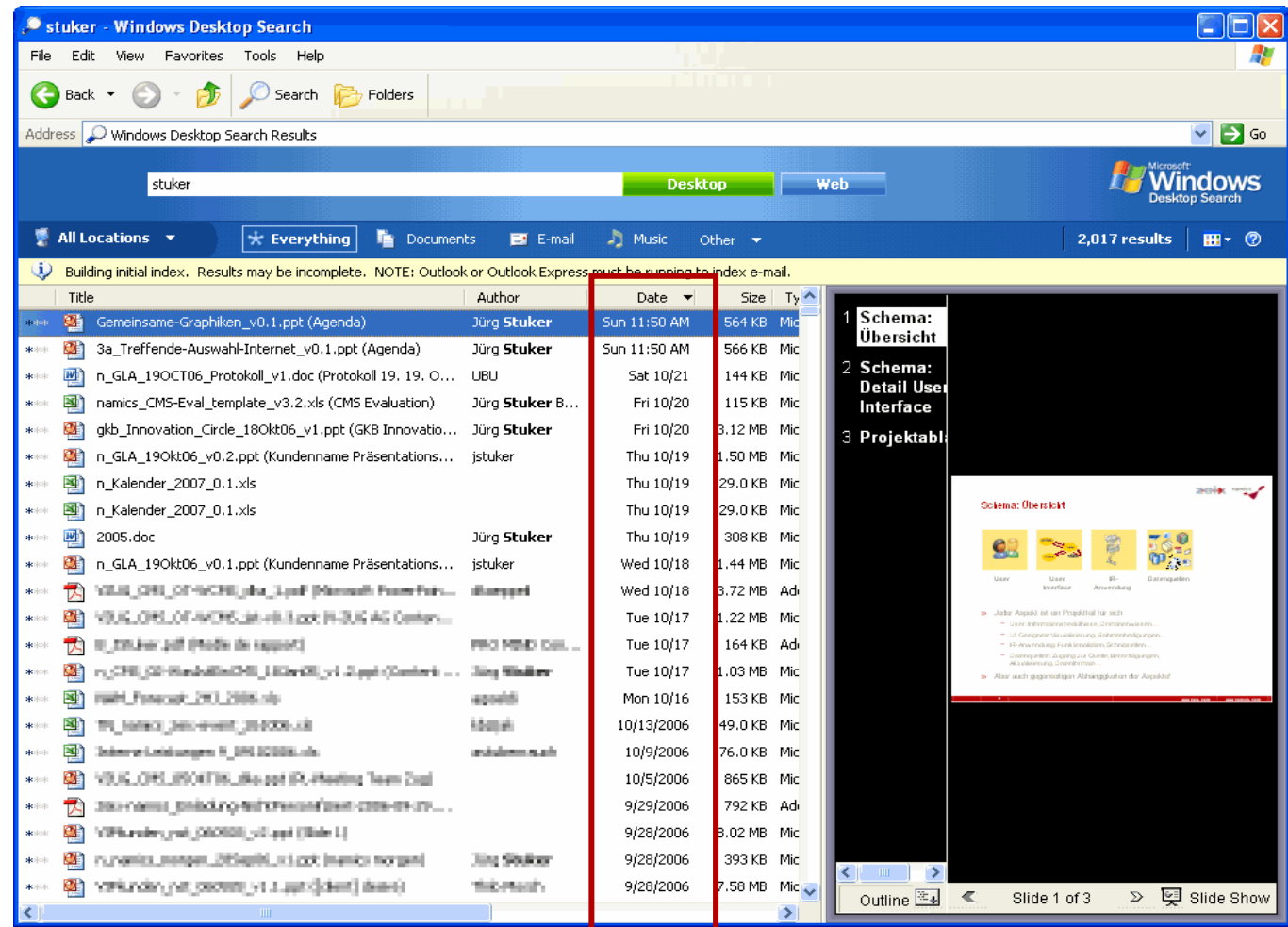


» Nutzungsdaten

- IR-Produktspezifisch, ob Nutzungsdaten automatisch für Ranking mitverwendet werden
- Von Standard IR-Applikationen meist nicht berücksichtigt
- Hohe Anforderung an IR-Produkt und Implementation

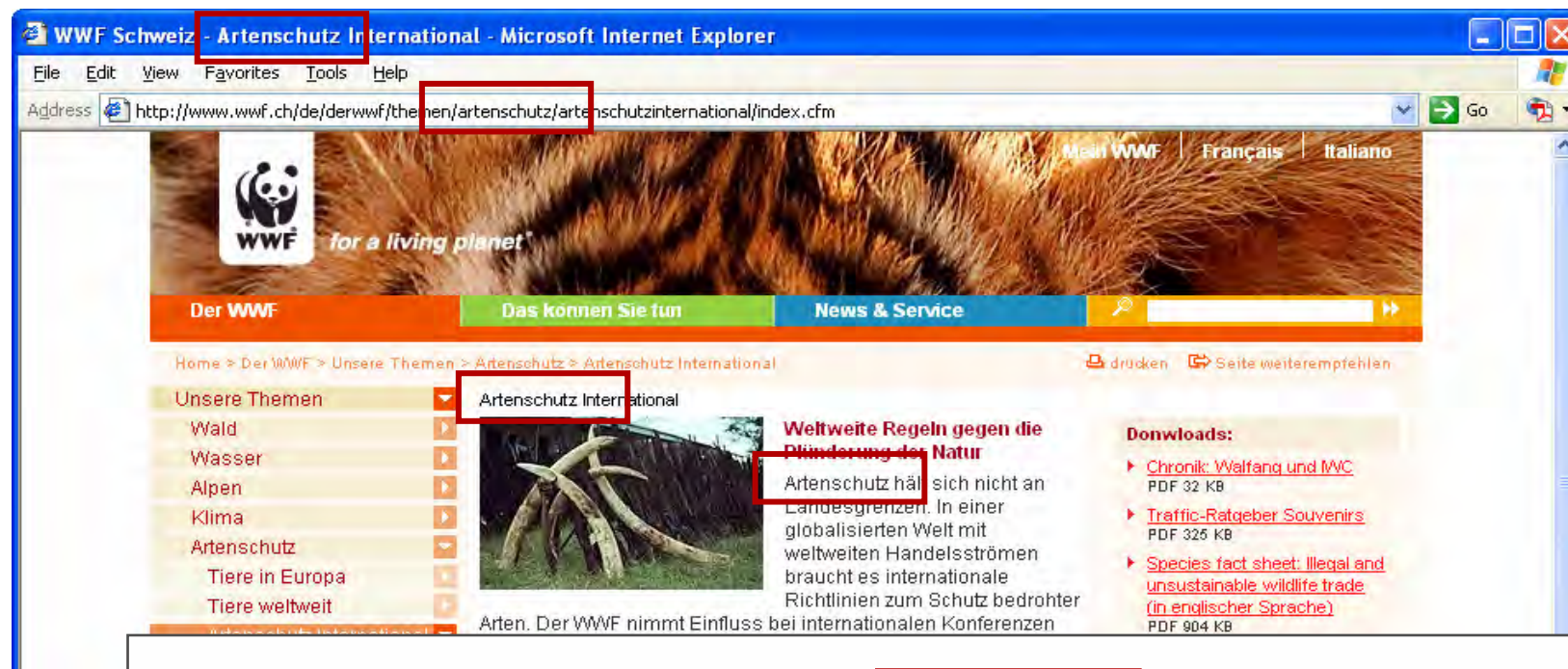
Weitere Ranking-Verfahren: Datum

- » sinnvoll, wenn Suchfunktion als Suchstrategie für aktuelle Informationen verwendet wird (z.B. E-Mails)
- » weniger sinnvoll für weiter zurück liegende Dokumente



Weitere Ranking-Verfahren: Suchwortfrequenz

- » Gewichtet nach Auffälligkeit der Platzierung (Titel, Text, URL)
- » Auch Meta-Keywords sinnvoll (Im Gegensatz zu Internet, da Spamming-Gefahr im Intranet kleiner)



```
<meta name="DESCRIPTION" content="Artenschutz hält sich nicht an Landesgrenzen.
<meta name="KEYWORDS" content="Artenschutz Artenschutzabkommen, Schutz von Tier
```

Wortfrequenzen (On Page)

$$\text{Relevanz} = \frac{\text{Häufigkeit der Suchworte auf der Trefferseite} \times \text{Auffälligkeit der Platzierung}}{\text{Häufigkeit der Suchworte in der Kollektion}}$$

Indexer Weights
Enter the importance of text relative to the body text of a document.

Title: Description: Keywords: Alt: Remote anchors:

HTML Meta Tag Names
Enter the meta tag names to use for the following purposes.

Title: Description:
 Date: Keywords:
 URL: Publisher:

Word Spam Detection Thresholds
Enter the spam detection thresholds for number times a word occurs within 100 words.

Title: Description: Keywords: Alt:
 Body:

Literatur

- Robertson, Walker, Beaulieu, Gatford, Payne. Okapi at TREC-4, in NIST Special Publication 500-236: The Fourth Text Retrieval Conference (TREC-4): 73-96. 1995.

Fazit Ranking im Intranet

- » Suche ist nur so gut wie eingepflegter Inhalt
 - Gutes Ranking lässt sich nicht rein technisch lösen
 - Iteratives Vorgehen zur Optimierung
 - Ohne sinnvollen Seitentitel/Inhalt keine sinnvollen Resultate

- » Guter Inhalt ist teilweise steuerbar
 - Annotation von unstrukturierten Daten
 - Vorgaben für Titel, Description, Keywords im CMS
 - Als Pflichtfelder (Titel) oder optionale Felder (Keywords)
 - Schulung von Autoren

The screenshot shows a CMS interface with several tabs: 'Global page properties', 'Meta information', 'Keywords', 'Units', 'Workflow', and 'VersionWarp'. The 'Meta information' tab is active, displaying the following fields:

- Title:** Artenschutz International
see 'TitleText'
- Date:** 2006-10-25
automatically updated on page modification
- Description:** Artenschutz hält sich nicht an Landesgrenzen. In einer globalisierten Welt mit weltweiten Handelsströmen braucht es internationale
- Keywords:** Artenschutz, Artenschutzabkommen, Schutz von Tierarten, Erhalt Tierarten
enter comma separated keywords
- Author:** Peter Meier
- General Expert:** (empty field)
- No Index by robots:** Yes No
Select yes if you do not want search engine robots to index this page

At the bottom right, there are 'Save' and 'Cancel' buttons.

Weitere Strategien zur Verbesserung von Trefferlisten

- Rechtschreibe-Korrekturen (Wörterbuch-Integration)
- Synonyme (Thesaurus oder manuell gepflegt)
- Top Treffer (manuell gepflegt)
- Entitätenerkennung (z.B. anhand Keyword „contact“ bei Namen)



Home Sitemap

acmeINC intranet

Acme Directory Rooms Locations Desktop

contact jim wick search

Advanced Search Search Tips / Search FAQ

include secure content

Acme Enterprise Search Results 1 - 10 Search took 0.77 seconds.

Sort by: [Date](#) / [Relevance](#)

1 result available in the Exchange phonebook

Jim P. Wick Office: (212) 555-1212 ext.2355
jpwick@acme.com Cell: (718) 555-1212

8am 9 10 11 12pm 1 2 3 4 5

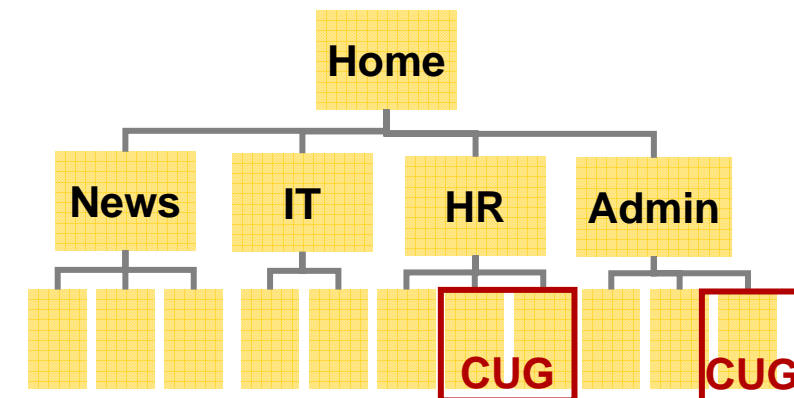
[Public Inclusion List](#)
 ... You can **contact** me at 1-800-555-1122 ext.257. ... 12/13/05, Barry **Wick**, rikwrybac@acme.com, 55577738, The books of Ian McLaren, Beside the Bonnie Briar Bush and The ...
www.corp.acme.com/publicdomain/ - 27k - 2006-02-14 - [Cached](#)

[AU Phone Messages](#)
 ... how do we **contact** you? ... Matt, Yes. 1/30/06, AccountLeads, Mr. Wang, 9747-1174 / 0284-884-200, ab ads - in the real estate industry - called and provided email **contact** ...
www.corp.acme.com/contact/au_phonemessages/ - 101k - 2006-02-15 - [Cached](#)

Spezialfall Intranet: Berücksichtigung von Berechtigungsstrukturen

» Intranets können Dokumente mit unterschiedlichen Zugriffsberechtigungen enthalten:

- News, Wissensinformationen (für jedermann zugänglich)
- GL-Sitzungsprotokolle (nur definierter Personenkreis)
- HR-Dokumente wie Lohnstatistiken (nur definierter Personenkreis)
- etc.





» Zugriff auf Dokumente selbst geschützt z.B. mittels Closed Usergroups (Login, Benutzererkennung)


» Was soll eine Suche mit solchen Bereichen machen?


- Gar nicht indizieren? → Keine Treffer aus CUG werden gefunden
- Indizieren? → Geschützte Dokumente werden gefunden → Sicherheitsproblematik
- Steigerungsform: Unterschiedliche Lese-Berechtigungen innerhalb des gleichen Dokuments


Trefferliste mit schützenswerten Inhalten

 [3_AngUnternehmen.doc](#)
AHV, IV, EO, FAK, ALV, BVG und UVG) einschliesslich der vom **Lohn** abgezogenen Arbeitnehmeranteile vollumfänglich bezahlt? Haben Sie die fälligen Beiträge, die sich aus
\\sgfileserver\projekte\Umwelt\Bank\Spit... \3_AngUnternehmen.doc - [Open folder](#) - [1 cached](#) - [Jun 29](#)

 [Lohn.xls](#)
Löhne 2006 Peter Müller CHF 4600.-- Hans Mustermann CHF 5200.-- Eva Sauber CHF 4750.--
Nicole Gross CHF 5300.--
My Documents\Archive\Projects\Wahl & Lohn\Verwalt... \Lohn.xls - [Open folder](#) - [1 cached](#) - [Mai 27](#)

 [XMAS Aktion 2004](#)
Intranet (mit entsprechender Auswertungsmöglichkeit, damit der gespendete Betrag direkt vom **Lohn** abgezogen werden kann) Eine Spendenaktion fördert die Kommunikation und trägt intern
My Documents\namics\Documents\Fürerze\11 - msp_XMAS 2004.pdf - [Open folder](#) - [1 cached](#) - [Mai 10](#)

 [Microsoft P...](#)
AG Honorieru
Gehaltssystem
My Document
[9](#)

 [Testresultat](#)
gesucht) Bere
Bonus, Entsch
und
Desktop\G5 A

Closed

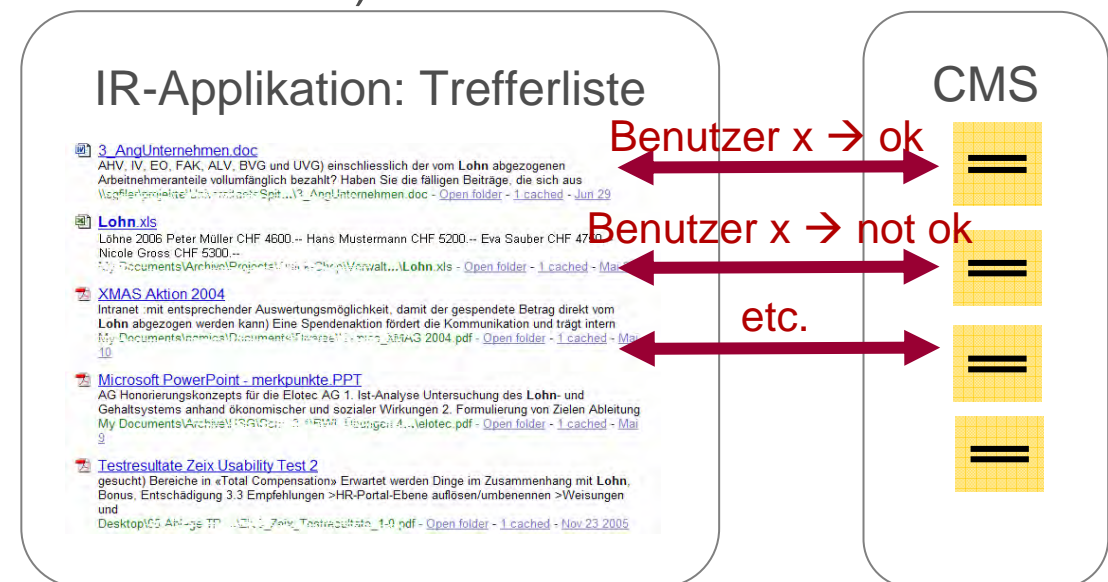
Benutzername

Passwort

[Passwort ändern](#)

Lösung für Suche nach geschützten Dokumenten

- » Retrieval Applikation muss Berechtigungen auf Dokumente kennen (bzw. abfragen können)
- » 2 Varianten:
 - Übergabe von Berechtigungsinformationen als Metadaten beim Indexieren eines Dokuments
 - Überprüfung kurz vor Auslieferung der Trefferliste, ob der Benutzer berechtigt ist (last second check)
- » Je nach Benutzerstruktur Performance-intensiv und aufwändig in der Implementation



Abschluss



- » Die Userbedürfnisse und die vorhandenen Daten-Typen und Daten-Qualität bestimmen das User-Interface
 - Richtig und gut gepflegte Inhalte sind massgebend für die Trefferqualität und das Ranking
 - Autoren User Interface ist mitentscheidend

- » Die durchsuchten Daten (Kollektion), die Suchanfrage (Query) und die Suchlogik haben einen Zusammenhang

- » Es gibt zahlreiche einfache Tricks zur Verbesserung von Trefferlisten

