

Web-Qualitätsmanagement

Grundlagen Webbasierter Analyse und Visualisierung

Seminar 2004

Inhalt:

1. Einführung

1.1. „lost in hyperspace“

1.2. Meta Index (CUI W3 catalog,
ALIWEB index)

1.3. WAVE

2. Zusammensetzung der Information

2.1. Web Robot (Web Wanderer, Web Robot)

2.2. WAISINDEX or ICE

2.3. Software Agents

2.4. Aufbau moderner Suchsystemen

Inhalt:

3. Interpretation and Klassifikation

3.1 Konzept-Klassen

3.2 Konzept-Einordnung

4. Ergebnis

5. Quellen

Einführung

1.1. „lost in hyperspace“

- ❑ Das Internet als Informationsquelle
- ❑ Das Suchen und Finden so kompliziert
- ❑ Suchmaschinen oder
- ❑ Katalog als Hilfsmittel

Einführung

1.2. *Meta Index Servers*

- ❑ CUI W3 catalog
 - Unterstützt von Universität Geneva
 - http://cui_ww.unige.ch/w3catalog
- ❑ ALIWEB index
 - Martijn Koster
 - <http://web.nexor.co.uk/doc>

Einführung

1.3. WAVE- Web Analysis and Visualisation Environment

- ❑ Erwerb von unverarbeiteten Daten
- ❑ Automatische Analyse und Klassifikation
- ❑ Visualisierung und interaktive Durchsicht

Einführung

WAVE - Prozess von Interpretation und Klassifikation

□ Analyse:

- Sammlung aller relevanten Informationen über ein Dokument im Cache
- Gliederung der Information des Dokumentes in der Komponententeile

□ Synthese:

- Identifikation der richtigen Konzept-Scalen
- Filterung der Informationen des Dokumentes durch die Konzept-Klassen
- Aufbau der Konzept-Klassen

2. Zusammensetzung der Information

2.1. Web Robot

- Web Wanderer
- Web Spider

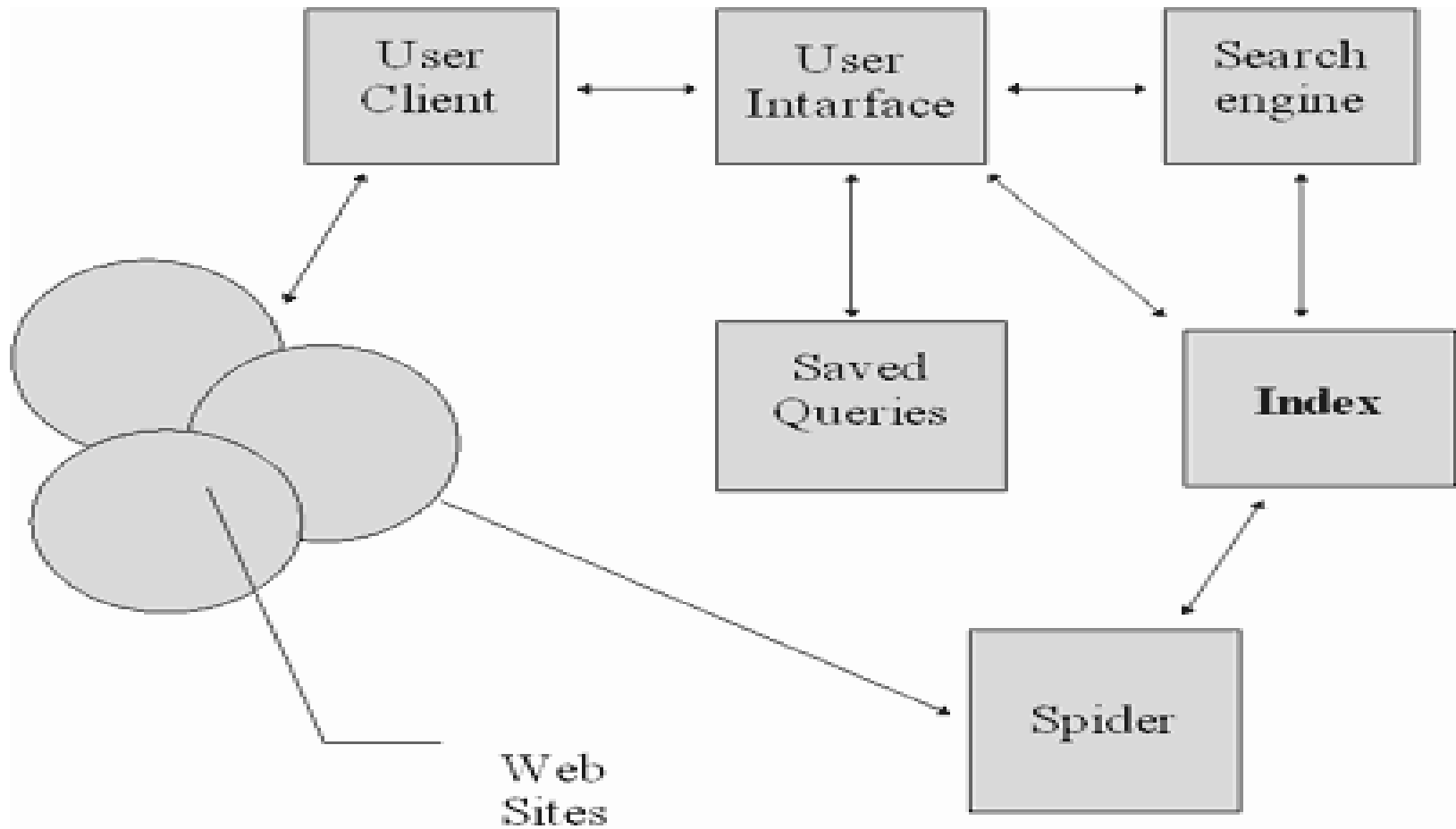
2.2. WAISINDEX or ICE

- Document indexing tools

2.3. Software Agents



2.4. Aufbau moderner Suchsystemen



3. Interpretation and Classification

3.1 Konzept-Klassen

- ❑ Formal context (G, M, I) , glm
- ❑ Extension
- ❑ Intension

3. Interpretation and Classification

3.2 Konzept - Scalin

- Vielbewertete Attribute für geographische Scaling

attributes

hemisphere

continent

country

state, province, land

city

domain

(eastern, western)

(...,europe,america:north)

(...,britain, germany,...)

(...,scotland, hessen, alaska,...)

(...,edinburgh, frankfurt, kiev,...)

- "geographical.scale.tbl"

continent:asia hemisphere:eastern

country:britain continent:europe

state:alaska country:united_states

city:frankfurt land:hessen

3. Interpretation and Classification

Location scale

city => state,

state => country,

country => continent,

continent => hemisphere

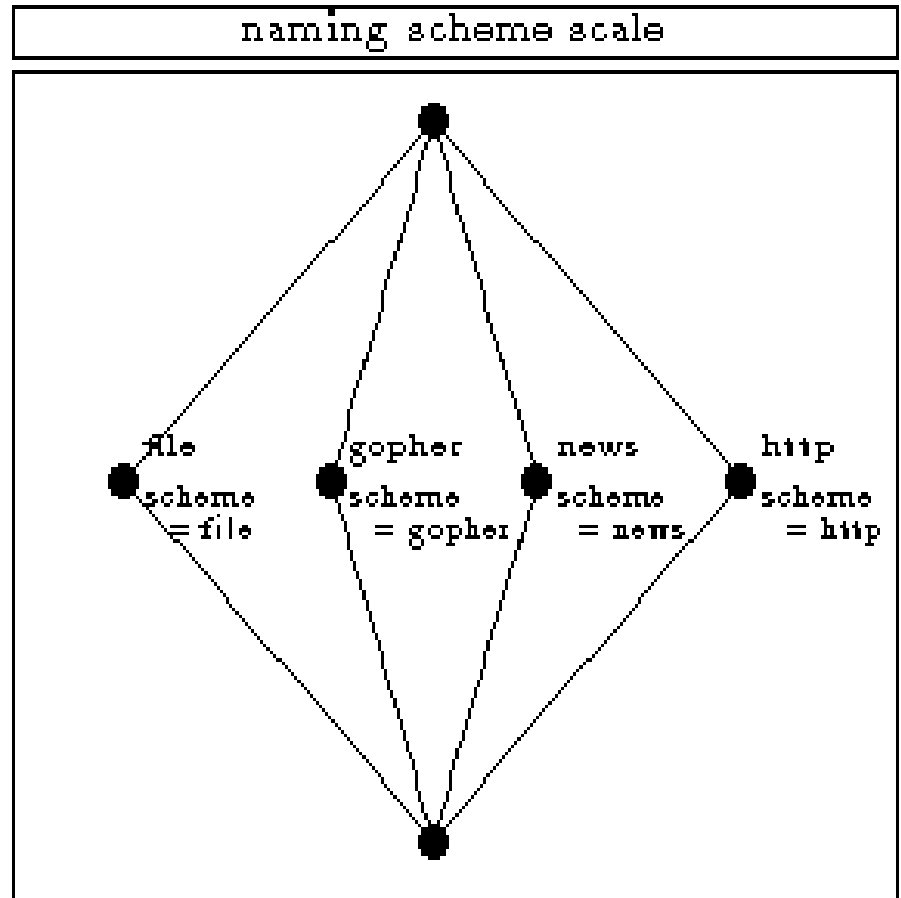
3. Interpretation and Classification

URLs - Scale

- The URL naming scheme component
- geographical scale

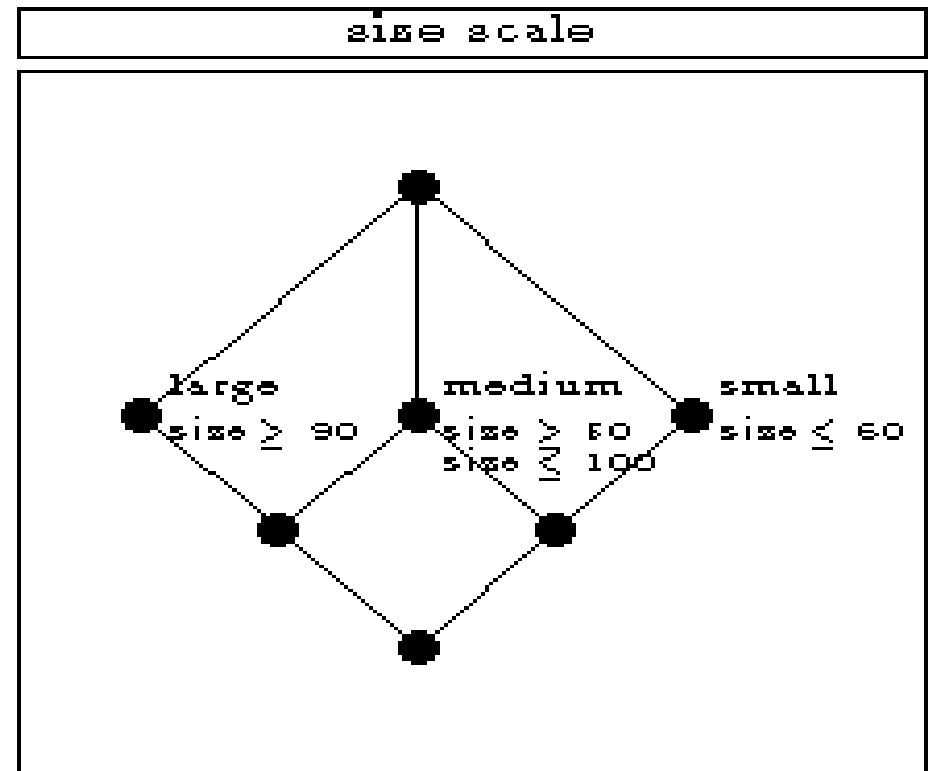
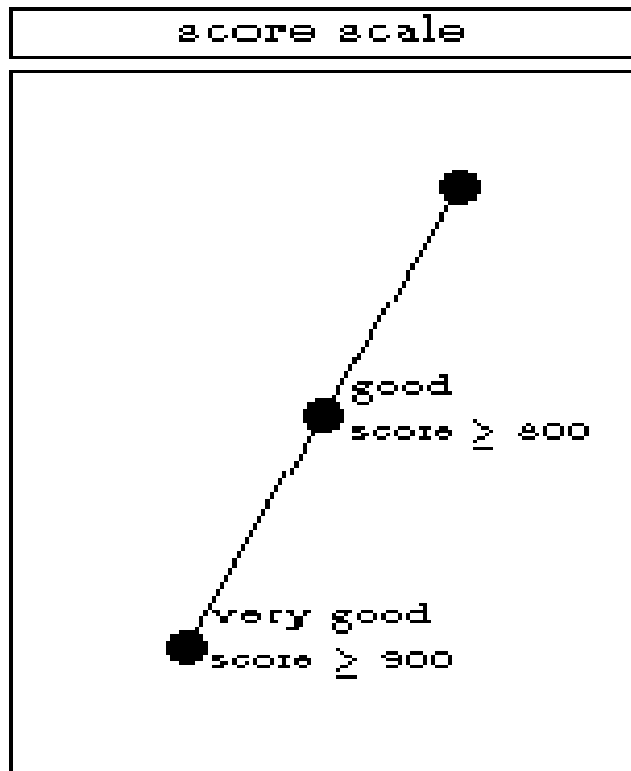
`http://www.cern.ch/ =>`
`www.cern.ch =>`
`geneva. (url => city)`

- URL path component



3. Interpretation and Classification

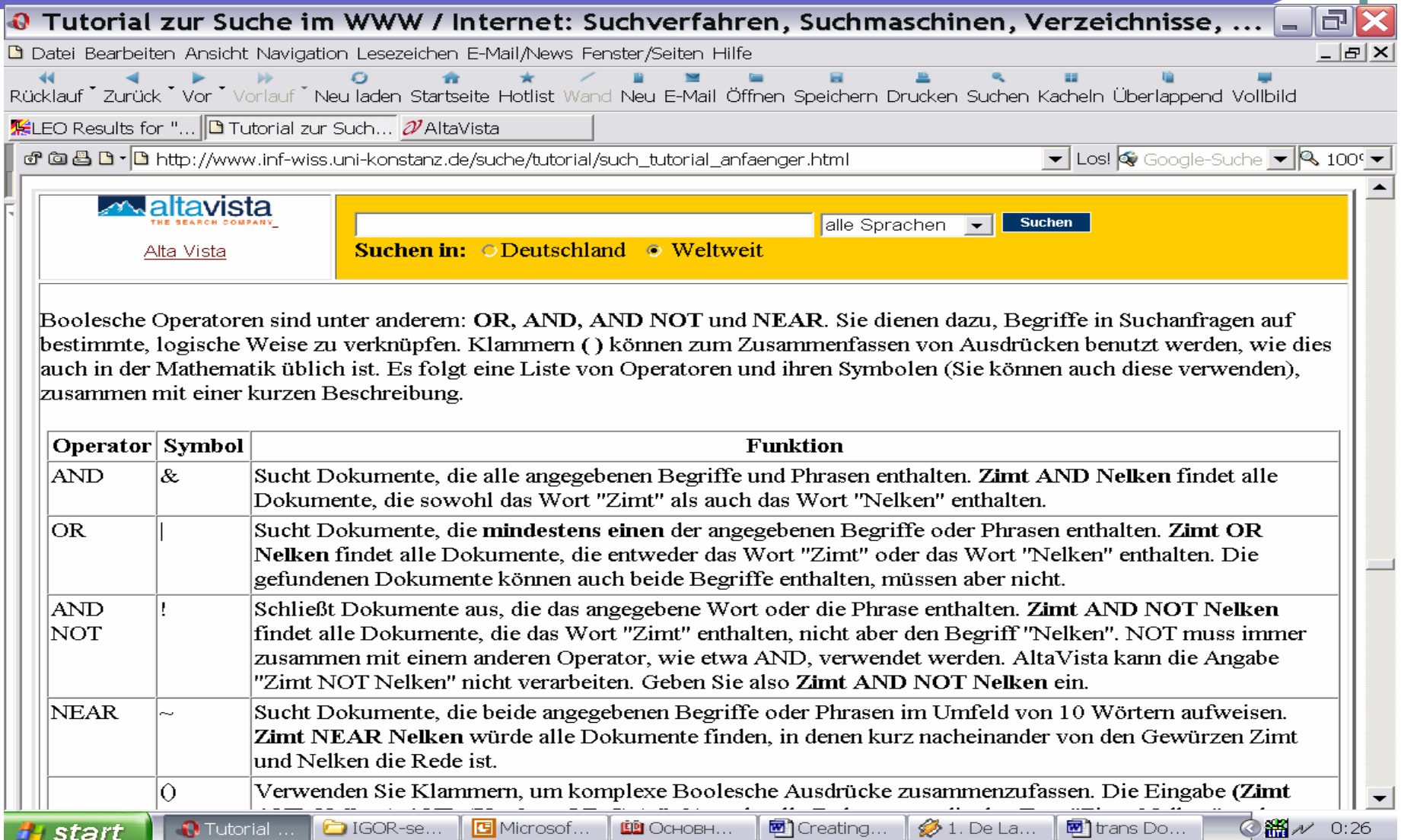
size scales, form scales



Ergebnis

- ❑ Erforschung und Entwicklung neuer Methoden zur Unterstützung der Nutzer für ein effizientes und zielgerichtetes Arbeiten im Internet.
- ❑ Begriff »Intelligente« oder »Mobile« Agenten. Diese weiterführenden Konzepte aus dem Bereich der Künstlichen Intelligenz versprechen eine weitere Professionalisierung des Umgangs mit dem Internet.

Boolesche Operatoren (AltaVista)



The screenshot shows a web browser window with the title "Tutorial zur Suche im WWW / Internet: Suchverfahren, Suchmaschinen, Verzeichnisse, ...". The address bar shows the URL "http://www.inf-wiss.uni-konstanz.de/suche/tutorial/such_tutorial_anfaenger.html". The search engine is AltaVista, and the search results page is titled "LEO Results for '...'". The search bar contains the text "Tutorial zur Such..." and "AltaVista". The search results page has a yellow header with the AltaVista logo and a search bar. Below the header, there is a section titled "Boolesche Operatoren sind unter anderem: OR, AND, AND NOT und NEAR. Sie dienen dazu, Begriffe in Suchanfragen auf bestimmte, logische Weise zu verknüpfen. Klammern () können zum Zusammenfassen von Ausdrücken benutzt werden, wie dies auch in der Mathematik üblich ist. Es folgt eine Liste von Operatoren und ihren Symbolen (Sie können auch diese verwenden), zusammen mit einer kurzen Beschreibung."

Operator	Symbol	Funktion
AND	&	Sucht Dokumente, die alle angegebenen Begriffe und Phrasen enthalten. Zimt AND Nelken findet alle Dokumente, die sowohl das Wort "Zimt" als auch das Wort "Nelken" enthalten.
OR		Sucht Dokumente, die mindestens einen der angegebenen Begriffe oder Phrasen enthalten. Zimt OR Nelken findet alle Dokumente, die entweder das Wort "Zimt" oder das Wort "Nelken" enthalten. Die gefundenen Dokumente können auch beide Begriffe enthalten, müssen aber nicht.
AND NOT	!	Schließt Dokumente aus, die das angegebene Wort oder die Phrase enthalten. Zimt AND NOT Nelken findet alle Dokumente, die das Wort "Zimt" enthalten, nicht aber den Begriff "Nelken". NOT muss immer zusammen mit einem anderen Operator, wie etwa AND, verwendet werden. AltaVista kann die Angabe "Zimt NOT Nelken" nicht verarbeiten. Geben Sie also Zimt AND NOT Nelken ein.
NEAR	~	Sucht Dokumente, die beide angegebenen Begriffe oder Phrasen im Umfeld von 10 Wörtern aufweisen. Zimt NEAR Nelken würde alle Dokumente finden, in denen kurz nacheinander von den Gewürzen Zimt und Nelken die Rede ist.
	()	Verwenden Sie Klammern, um komplexe Boolesche Ausdrücke zusammenzufassen. Die Eingabe (Zimt

Advanced Search (Google)

Tutorial zur Suche im WWW / Internet: Suchverfahren, Suchmaschinen, Verzeichnisse, ...

Datei Bearbeiten Ansicht Navigation Lesezeichen E-Mail/News Fenster/Seiten Hilfe

Rücklauf Zurück Vor Vorlauf Neu laden Startseite Hotlist Wand Neu E-Mail Öffnen Speichern Drucken Suchen Kacheln Überlappend Vollbild

LEO Results for "... Tutorial zur Such... AltaVista

http://www.inf-wiss.uni-konstanz.de/suche/tutorial/such_tutorial_anfaenger.html#1 Los! Google-Suche 100%

Google [Suchtipps](#) | [Alles über Google](#)

Erweiterte Suche

Ergebnisse finden mit **allen** Wörtern 10 Ergebnisse
mit der **genauen Wortgruppe**
mit **irgendeinem** der Wörter
ohne die Wörter

Sprache Antwortseiten, geschrieben in beliebiger Sprache

Dateiformat Ausschließlich Ausgabe von Ergebnissen des Dateiformats irgendein Format

Datum Ausgabe neuer Webseiten, aktualisiert innerhalb: keine Zeitbegrenzung

Position Antwortseiten, in denen meine Begriffe vorkommen irgendwo auf der Seite

Domains Ausschließlich Antwortseiten von der Site oder Domain
Beispiele: .org, google.com [Mehr Informationen](#)

Seitenspezifische Suche

Ähnlich Seiten suchen, die der folgenden Seite ähnlich sind
Beispiel: www.google.com/help.html

Links Seiten suchen, die einen Link auf die folgende Seite enthalten

©2002 Google

3.2 Groß-/Kleinschreibung

Quellen:

- ❑ <http://www.cbc.umn.edu/>
- ❑ <http://www.citforum.ru/>
- ❑ <http://www.devmag.net/suchmaschinen/suchfunktion.htm>
- ❑ http://www.lpm.uni-sb.de/chemie/ch_suchinternet.htm
- ❑ http://www.inf-wiss.uni-konstanz.de/suche/tutorial/such_tutorial_anfaenger.html#1
- ❑ Neuss, Christian; Höfling, Stefanie: *Lost in Hyperspace? Free Text Searches in the Web*, Proceedings of the First International World-Wide Web Conference, CERN, Switzerland, May 1994
- ❑ Koster, Martijn: *ALIWEB - Archie-Like Indexing in the WEB*, Proceedings of the First International World-Wide Web Conference, CERN, Switzerland, May 1994
- ❑ Ganter, B.; Wille, R.: *Conceptual Scaling*, Applications of Combinatorics and Graph Theory in the Biological and Social Sciences, Springer, New York, 1989
- ❑ Kahle, Brewster: *An Information System for Corporate Users: WAIS*, Thinking Machines Corporation, April 1991