

## 2. DFG- Workshop

### 3.4. Bereitstellung/Präsentation

Beitrag von Peter Küsters

#### Bereitstellung/Präsentation

##### Index (Erfassung)

Die gespiegelten Seiten werden mit dem Programm dtSearch indexiert. DtSearch ist ein kommerzielles Programm zur automatischen Volltextindexierung. Es kann beliebig viele Indizes gleichzeitig verwenden, und jeder Index kann über ein Tera-byte Daten enthalten. Abgesehen von diesem Limit für einzelne Indizes gibt es keinerlei Beschränkung; es können beliebig viele Dateien in einem Index aufgenommen werden.

Um gute Suchergebnisse zu ermöglichen, ist es schon bei der Indexierung der Spiegelungen notwendig, eine konsistente Strategie zu verwenden, da die Optionen, die bei der Indexierung gewählt werden, sehr großen Einfluss auf die spätere Suche haben.: Einige Optionen beeinflussen die Suchgeschwindigkeit, andere die Suchtreffer. Die geschwindigkeitserhöhenden Optionen gehen auf Kosten des vom Index verwendeten Speicherplatzes. Da die Suchgeschwindigkeit hinreichend ist und keine primäre Rolle spielt, verwenden wir keinerlei geschwindigkeitsoptimierenden Optionen.

Optionen, die Einfluss auf die Suchergebnisse haben sind case-sensitive (Groß-/Kleinschreibung), und accent-sensitive (Akzente und Umlaute). Beides wird von uns nicht verwendet. Dies führt in Einzelfällen zu ungewollten Gleichsetzungen von Worten. In den meisten Fällen ist es jedoch besser für den Benutzer, wenn diese Gleichsetzung stattfindet. Die sehr oft vorkommende Großschreibung etwa in Überschriften würde bei der Option „case-sensitive“ dazu führen, dass viele Worte völlig unberechtigt unterschieden werden. Ungewollte Kollisionen durch verschiedene Bedeutung zwischen Groß- und Kleinschreibung erscheint uns als das kleinere Übel.

Die Unterscheidung „accent-sensitive“ scheint auf den ersten Blick vernünftig zu sein. Allerdings ist der Fall, dass durch einen Umlaut die Bedeutung des Wortes verändert wird, sehr selten. Gerade bei Flexionen kommt es hingegen sehr häufig vor, dass Stammformen durch Umlaute verändert werden, so dass hier eine Unterscheidung zwischen Umlaut und Ausgangsvokal kontraproduktiv ist.

Indiziert werden nicht alle Dateien. dtSearch indiziert nur Text. Hierbei kann es allerdings alle gängigen Formate verarbeiten. Bilder, Video- und Audiodateien sowie Binärdateien können allerdings nicht berücksichtigt werden. Zusätzlich zu diesen Formaten schließen wir im Indexierungsprozess auch die im Arbeitsprozess entstandenen verwaltungstechnischen Dateien aus.

##### Zugänge (Präsentation)

Die Homepage bietet zwei grundsätzlich verschiedene Möglichkeiten des Zugangs zu den gespiegelten Seiten. Zum einen gibt es die Zugänge über die Hauptseite der Spiegelung und über die tabellarische Auflistung der in einer Spiegelung vorhandenen Seiten. Zum anderen

hat man auch die Möglichkeit, die Seiten zu durchsuchen. Hier kann über einen Index gezielt nach Dokumenten gesucht werden, die bestimmte Worte enthalten. Den Index kann man auf 3 Arten erreichen, je nachdem welche Spiegelungen man durchsuchen will.

### Die Webform

Der Benutzerzugang ist sehr frei konfigurierbar. Die sogenannte Webform, welche diesen Zugang liefert, bietet ein gutes Dutzend Optionen. Wir haben uns weitgehend für die Standardoptionen entschieden. Diese bieten ein Suchfenster, verschiedene Sucharten, optionale Flexionen (Stemming), bis zu 5000 Treffer und variable Trefferanzahlen pro Seite. Von den Standardoptionen nicht übernommen wurden die phonetische Suche und die fuzzy Suche. DtSearch bietet unter den erweiterten Optionen auch die Möglichkeit, mit Hilfe eines Thesaurus eine Synonymsuche anzubieten.

### Suche in der Spiegelungsübersicht

Interessiert man sich nur für eine einzelne Spiegelung, erreicht man den entsprechenden Index über den Suche-Button dieser Spiegelung. Es erscheint eine Suchmaske, mit der man über den Index dieser Spiegelung eine Suchanfrage durchführen kann.

### Diachrone Suche

Neben der Suche über eine einzelne Spiegelung gibt es auch die Möglichkeit, über mehrere Spiegelungen eines Bundeslandes, der Bundespartei oder der Bundestagsfraktion zu suchen. Dieser Zugang ist über den Button „Diachrone Suche“ bei der jeweiligen Homepageseite zu erreichen. Hier erhält man eine Liste aller Indizes der Spiegelungen eines Bundeslandes beziehungsweise der Bundespartei oder Bundestagsfraktion. Durch gezieltes Auswählen mehrerer oder aller Indizes werden diejenigen für die Suche bestimmt, in denen man die Dokumente zu finden hofft.

### Synchrone Suche

Mit dem Button „synchronen Suche“, der über die obere Menüleiste erreichbar ist, hat der Benutzer die Möglichkeit, über sämtliche Indizes eines Jahres zu suchen. Auch hier ist es möglich, gezielt die Indizes der Spiegelungen auszuwählen, deren Inhalte potentiell die gewünschten Dokumente enthalten könnten. Dies ist zum Beispiel sinnvoll, wenn man eine Suche über mehrere Bundesländer gleichzeitig durchführen will.

### Ergebnisliste

Hat man seine Suche durchgeführt, erhält man die Trefferliste. Die Überschrift zeigt die Suchanfrage und die Anzahl der Treffer. Zu jedem einzelnen Treffer enthält man folgende Informationen:

Den Titel des Dokumentes, den Pfad des Dokumentes, die Anzahl der Treffer, das letzte Änderungsdatum, die Größe und den Anfang des Dokuments.

Zusätzlich zu diesen Kurzinformationen erhält man auch zu jedem Treffer 2 Links. Der erste Link zeigt die Seite aus dem Index mit hervorgehobenen Suchwörtern an, der zweite Link führt direkt auf den Treffer innerhalb der Spiegelung.

## Suchstrategien

Um einen Index sinnvoll zu nutzen, bedarf es einer Suchstrategie. Der Index ist kein Allheilmittel um ohne großes Überlegen brauchbare Resultate zu erhalten. Die Suche nach einem einzelnen Wort wird häufig die maximal angezeigten 5000 Treffer liefern. Eine andere Information als die, dass das Suchwort häufig verwendet wird, erhält man hierdurch nicht. Es gibt neben der einfachen Schlagwortsuche weitere Möglichkeiten, um zu besseren Suchergebnissen zu gelangen.

## Mehrere Worte

Um die Suche einzuschränken, kann man mehrere Suchworte benutzen. Um Dokumente zu finden, in denen Gerhard Schröder vorkommt, gibt man sowohl Gerhard als auch Schröder in die Suchmaske ein. Aber auch Gerhard und Schröder als Suchwort werden noch äußerst häufig auftreten. Um brauchbare Ergebnisse zu erzielen muss man die Suche weiter konkretisieren. Wenn man zum Beispiel nach Bildern sucht, so hilft es, neben Gerhard und Schröder auch nach jpg zu suchen. Dtsearch ist allerdings ein reiner Textindex, Bilder werden nicht gesondert indiziert. Also findet man bei einer derartigen Suche nur Seiten, auf denen Gerhard, Schröder und jpg als Worte verwendet werden.

Internet Archiv - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address [http://mirroring/Bundesspd/home\\_bund.htm](http://mirroring/Bundesspd/home_bund.htm)

Google Suche Rechtschreibprüfung Optionen

**SPD** Sozialdemokratische Partei Deutschlands

Internet-Archiv

Startseite Synchrone Suche Diachrone Suche Übersicht Hilfe

**dtSearch Web Search** Prev Doc Next Doc First Hit Next Hit Prev Hit Help

**Search Results**

Request: Gerhard Schröder .jpg - Bulmahn  
39 document(s) retrieved

**Score Document**

100%	<a href="#">Neujahrsempfang der Afa Dithmarschl Ministerpräsidentin Schleswig-Holstei</a>	20.11.2003	5k	Neujahrsempfang der Afa Dithmarschl Ministerpräsident
83%	<a href="#">Wahlkampfauftakt Heide Simonis - M</a>	20.11.2003	8k	Wahlkampfauftakt Heide Simonis - M Wahlkam
50%	<a href="#">N3 Heide Simonis - Ministerpräsident</a>			

Die beiden Hauptredner des Abends: Heide Simonis und Gerhard Schröder: "Unsere Stärke kommt nicht nur aus der Schwäche der anderen."

(Beispiel: Spiegelung Bundespartei SPD 20.11.2003: Gerhard Schröder .jpg -Bulmahn)

## Phrasen

Da einzelne Suchworte oft zu viele Treffer erzeugen, kann man durch Benutzung von Anführungszeichen zusammengehörige Phrasen kennzeichnen. „Willy wählen“ findet nur die

Dokumente, in denen der bekannte Slogan der Kampagne von 1972 erwähnt ist, während eine Suche nach „Willy“ und „wählen“ die maximale Anzahl Treffer liefert und wenig brauchbar ist.

Internet Archiv - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address [http://mirroring/Bundesspd/home\\_bund.htm](http://mirroring/Bundesspd/home_bund.htm)

Google Suche Rechtschreibprüfung Optionen

**SPD** Sozialdemokratische Partei Deutschlands

Internet-Archiv Startseite Synchronische Suche Diachrone Suche Übersicht Hilfe

**dtSearch Web Search** Prev Doc Next Doc First Hit Next Hit Prev Hit Help

**Search Results**  
Request: Willy wählen  
5000 document(s) retrieved

**Score Document**

100%	<a href="#">Microsoft Word - Tagungsband /IntAr/SPD_B_P_2004_12_29/www.i</a>
254	21.12.2004
723k	Stand und Perspektiven der Willy-Brandt-Berline
43%	<a href="#">Heft10_RememberingWillyBrandt.pdf /IntAr/SPD_B_P_2004_12_29/www.i</a>
111	21.12.2004
319k	HERAUSGEBER Bundeskanzler-Präsidenten
34%	<a href="#">Rede von Bundespräsident Johannes /IntAr/SPD_B_P_2004_12_29/www.i</a>
86	

Save a Copy Print Email Search Select Text Do more with Adobe PDF

76%

Bookmarks Signatures Layers Pages

**Stand und Perspektiven der Willy-Brandt-Forschung**

8,26 x 11,69 in 1 of 125

Lokales Intranet

(Beispiel: Diachrone Suche Bundespartei: Willy wählen)

(Beispiel: Diachrone Suche Bundespartei: „Willy wählen“)

### Worte ausschließen

Beim Auffinden von Dokumenten kann es hilfreich sein, bestimmte Worte auszuschließen. Dies geschieht durch voranstellen eines „-“-Zeichens. Dies kann zum Beispiel verwendet werden, ungewollte Treffer etwa aus Benutzerforen nicht anzuzeigen.

### Stemming/Flexionen

Ist die Option Stemming aktiviert, so werden nicht nur der exakt eingegebene Suchbegriff, sondern auch seine Flexionsformen gefunden. So werden zum Beispiel bei dem Suchwort „Parteitag“ auch „Parteitage“, „Parteitags“ und „Parteitages“ gefunden.

Internet Archiv - Microsoft Internet Explorer

Address: [http://mirroring/Landesverbände/mv\\_p.htm](http://mirroring/Landesverbände/mv_p.htm)

SPD-Landesverband Mecklenburg-Vorpommern

Internet-Archiv

Startseite Synchronische Suche Diachrone Suche Übersicht Hilfe

dtSearch Web Search

Prev Doc Next Doc First Hit Next Hit Prev Hit Help

**Search Results**

Request: Parteitag  
703 document(s) retrieved

**Score Document**

100%	<a href="#">SPD Rostock</a> <a href="#">/IntAr/SPD_L_MV_2005_03_09/www</a> 28 07.03.2005 26k SPD Rostock Pressemitteilungen Teil Zusammenarb
42%	<a href="#">Brosch.re-Leitantrag-rot</a> <a href="#">/IntAr/SPD_L_MV_2005_03_09/www</a> 21 07.03.2005 342k SPD AGENDA 2010 mut zur verände Juni 2003 in Berl
37%	<a href="#">SPD Kreisverband Neubrandenburg</a> <a href="#">Parteitag in Bochum</a> <a href="#">/IntAr/SPD_L_MV_2005_03_09/www</a>

§ 1 Tätigkeitsbereich und Name  
§ 2 Gliederung  
§ 3 Organe des Kreisverbandes  
§ 4 Kreisparteitag  
§ 5 Anträge  
§ 6 Aufgaben des ordentlichen **Parteitages**  
§ 7 Der außerordentliche **Parteitag**  
§ 8 Der Kreisvorstand  
§ 9 Aufgaben des Vorstandes  
§ 10 Der Erweiterte Kreisvorstand  
§ 11 Wahlkreis-konferenz  
§ 12 Schiedskommission  
§ 13 Kontrollkommission  
§ 14 Projektgruppen  
§ 15 Ortsvereine  
§ 16 Änderung der Satzung  
§ 17 Schlussbestimmungen

§ 1

Lokales Intranet

(Beispiel: MV 09.03.2005 Parteitag mit Stemming aktiv)

### Wortschöpfungen im Wahlkampf

Worte können im Laufe der Zeit ihre Bedeutung verändern, so dass ihr Gebrauch plötzlich sprunghaft ansteigt oder fällt. Über die Diachrone Suche kann dieser Wandel nachvollzogen werden. So führt eine Suche nach „Kopfpauschale“ in den Indizes bis Mitte 2003 nur zu sehr wenigen Ergebnissen. Ab Ende 2003 ist der Begriff in Tausenden Dokumenten vertreten, und wird mit der uns heute gebräuchlichen Bedeutung verwendet.

Internet Archiv - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print

Address http://mirroring/Bundesspd/home\_bund.htm

Google Suche Rechtschreibprüfung Optionen

**SPD** Sozialdemokratische Partei Deutschlands

Internet-Archiv Startseite Synchronische Suche Diachrone Suche Übersicht Hilfe

**dtSearch Web Search** Prev Doc Next Doc First Hit Next Hit Prev Hit Help

22.11.2001  
4077k  
Anträge zum Parteitag der SPD Nürnberg 19. bis 22. November 2001 SPD  
Anträge

Items 1 - 10

**Search for**  
Kopfpauschale all of the words Search

**Indexes to search**

- SPD\_B\_P\_1999\_04\_30
- SPD\_B\_P\_1999\_10\_11
- SPD\_B\_P\_2001\_11\_14
- SPD\_B\_P\_2001\_11\_22\_PT
- SPD\_B\_P\_2002\_06\_10
- SPD\_B\_P\_2002\_10\_10
- SPD\_B\_P\_2003\_04\_29
- SPD\_B\_P\_2003\_06\_06
- SPD\_B\_P\_2003\_11\_17
- SPD\_B\_P\_2004\_12\_29
- SPD\_B\_P\_2005\_09\_18\_W

**Sucharten**  
**eines der Worte oder alle Worte**  
Findet eine Liste von Worten oder Phrasen

- benutzen Sie "Anführungszeichen" zum Kennzeichnen von Phrasen
- setzen Sie + vor jedes Wort und jede Phrase die enthalten sein muß
- setzen Sie - vor jedes Wort und jede Phrase die nicht enthalten sein darf
- Beispiele:  
Banane Birne "süßer Apfel"

(Beispiel: Diachrone Suche Bundespartei „Kopfpauschale“, 10 Treffer bis Mitte 2003 )

Internet Archiv - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print

Address http://mirroring/Bundesspd/home\_bund.htm

Google Suche Rechtschreibprüfung Optionen

**SPD** Sozialdemokratische Partei Deutschlands

Internet-Archiv Startseite Synchronische Suche Diachrone Suche Übersicht Hilfe

**dtSearch Web Search** Prev Doc Next Doc First Hit Next Hit Prev Hit Help

**Search Results**

Request: Kopfpauschale  
5000 document(s) retrieved

Items 1 - 10 Next 10

**Score Document**

100%	<a href="#">SPD-Antragsbuch_BPT_Bochum03.pdf</a> <a href="#">/IntAr/SPD_B_P_2003_11_17_PT/www.spd.de/servlet/PB/show/10298019</a> 17.11.2003 2884k Anträge zum ordentlichen Bundesparteitag der SPD Ruhr Congress Bochum 17. bis 1
100%	<a href="#">SPD-Antragsbuch_BPT_Bochum03.pdf</a> <a href="#">/IntAr/SPD_B_P_2003_11_17_PT/www.spd.de/servlet/PB/salt/w65gnj1ezuf7k1ey58pj1g3xtds1q9pqa/show/10298019</a> 17.11.2003 2884k Anträge zum ordentlichen Bundesparteitag der SPD Ruhr Congress Bochum 17. bis 1
100%	<a href="#">SPD-Antragsbuch_BPT_Bochum03.pdf</a> <a href="#">/IntAr/SPD_B_P_2003_11_17_PT/www.spd.de/servlet/PB/</a>

**Sucharten**  
**eines der Worte oder alle Worte**  
Findet eine Liste von Worten oder Phrasen

- benutzen Sie "Anführungszeichen" zum Kennzeichnen von Phrasen
- setzen Sie + vor jedes Wort und jede Phrase die enthalten sein muß
- setzen Sie - vor jedes Wort und jede Phrase die nicht enthalten sein darf
- Beispiele:  
Banane Birne "süßer Apfel"

(Beispiel: Diachrone Suche Bundespartei „Kopfpauschale“, (5000 Treffer Ende 2003)

Ein Beispiel aus dem Bundestagswahlkampf 2005 zeigt, dass Wort auch auf den verschiedenen Gliederungsebenen mit sehr unterschiedlicher Intensität genutzt werden. Bei der Synchronen Suche im Jahr 2005 findet man das Wort „Merkelsteuer“ fast ausschließlich auf den Seiten der Bundesebene der SPD.

**Internet Archiv - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

Address [http://mirroring/synchrone\\_suche.htm](http://mirroring/synchrone_suche.htm)

dtSearch Web Search

**Synchrone Suche**

Internet-Archiv Startseite Synchrone Suche Hilfe

Prev Doc Next Doc First Hit Next Hit  
Prev Hit Help

**Search Results**

Request: Merkelsteuer  
0 document(s) retrieved  
No files matched the search request "Merkelsteuer"

**Search for**

Merkelsteuer

all of the words Search

**Indexes to search**

- SPD\_B\_F\_2005\_07\_13
- SPD\_B\_P\_2005\_07\_01
- SPD\_B\_P\_2005\_09\_18\_W
- SPD\_L\_BB\_2005\_11\_14
- SPD\_L\_BE\_2005\_04\_13
- SPD\_L\_BW\_2005\_12\_14
- SPD\_L\_BY\_2005\_11\_14
- SPD\_L\_HB\_2004\_11\_03
- SPD\_L\_HE\_2005\_06\_10
- SPD\_L\_HH\_2005\_05\_10

**Sucharten**

**eines der Worte oder alle Worte**

Findet eine Liste von Worten oder Phrasen

- benutzen Sie "Anführungszeichen" zum Kennzeichnen von Phrasen
- setzen Sie + vor jedes Wort und jede Phrase die enthalten sein muß
- setzen Sie - vor jedes Wort und jede Phrase die nicht enthalten sein darf
- Beispiele:  
Banane Birne "süßer Apfel"  
"süßer Apfel" -Salat +"saure Sahne"

**Exakte Phrase**

Findet eine einzelne Phrase (Anführungszeichen sind optional)

**Boolesch**

Findet eine strukturierte Gruppe von Worten oder Phrasen verbunden durch *and*, *or*, *not*, *w/*.

- Beispiele:

Lokales Intranet

(Beispiel: Synchrone Suche 2005 Fraktion und Länder „Merkelsteuer“ nicht vorhanden)

Internet Archiv - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address [http://mirroring/synchrone\\_suche.htm](http://mirroring/synchrone_suche.htm)

Suche Rechtschreibprüfung Optionen

**dtSearch Web Search**

Startseite Synchrone Suche Hilfe

Prev Doc Next Doc First Hit Next Hit Prev Hit Help

17k  
SPD Wahlkampfportal 2005 - Mehrwertsteuer SPD Wahlkampfportal 2005 • Hauptinh

66% [Manuel Müller](#)  
[/intAr/SPD\\_B\\_P\\_2005\\_09\\_18\\_W/www.spd-thueringen.de/spdnet/eichsfeld/mueller](#)

3  
12.09.2005  
2k  
Manuel Müller Nein zur "Merkelsteuer"! Manuel Müller sammelt Unterschriften

Items 1 - 10 [Next 10](#)

**Search for**

Merkelsteuer  all of the words

**Indexes to search**

- SPD\_B\_F\_2005\_07\_13
- SPD\_B\_P\_2005\_07\_01
- SPD\_B\_P\_2005\_09\_18\_W**
- SPD\_L\_BB\_2005\_11\_14
- SPD\_L\_BE\_2005\_04\_13
- SPD\_L\_BW\_2005\_12\_14
- SPD\_L\_BY\_2005\_11\_14
- SPD\_L\_HB\_2004\_11\_03

**Sucharten**

**eines der Worte oder alle Worte**

Findet eine Liste von Worten oder Phrasen

- benutzen Sie "Anführungszeichen zum Kennzeichnen von Phrasen
- setzen Sie + vor jedes Wort und jede Phrase die enthalten sein muß
- setzen Sie - vor jedes Wort und jede Phrase die nicht enthalten

Lokales Intranet

(Beispiel: Synchrone Suche 2005 Bundespartei Wahlkampfspiegelung „Merkelsteuer“ knapp 500 Treffer)