

Textdatenbanken

Sommersemester 2020

10. Vorlesung

- Text-Genres und Korpusstatistik, Teil 1 -

Uwe Quasthoff

Universität Leipzig
Institut für Informatik

quasthoff@informatik.uni-leipzig.de

Fragestellung

Verschiedene Textgenres

- Zeitungstext
- Romane
- Gedichte
- Fachtext eines Sachgebiets
- ...

In

- eine Datenbank oder
- in verschiedene Datenbanken?

Kriterien I

Verschiedene Datenbanken sind sinnvoll, wenn

- Unterschiede untersucht werden sollen
- fremde Genres (ähnlich Rauschen) das Ergebnis verschlechtern
- genügend Daten pro Genre vorliegen
- ...

Kriterien II

Verschiedene Datenbanken sind NICHT sinnvoll, wenn

- zu wenig Daten pro Genre vorliegen
- Rauschen zunächst nicht stört und evtl. später nach Genre sortiert werden kann
- ...

Vergleich verschiedener Textgenres: Zeitung / Web / Film

Vergleichskriterien:

- Stoppwörter
- Sachgebiete
- Fachlichkeit
- Redundanz

Vergleich der top-30-Wörter Zeitung / Web

In der folgenden Tabelle werden die Top-30 Wörter aus dem Webkorpus mit dem allgemeinen deutschen Korpus verglichen. In der letzten Spalte ist die Differenz der Positionen angegeben.

```
select w.w_id-100 as rank, w.word as news, ww.word as movie,
cast(www.w_id-ww.w_id as signed) as diff from
deu_news_2007.words w, words ww, deu_news_2007.words www
where w.w_id>100 and w.w_id<=130 and ww.w_id=w.w_id and
www.word= binary ww.word;
```

rank	news	web	diff
1	der	der	0
2	die	und	1
3	und	die	-1
4	in	in	0
5	den	den	0
6	von	zu	3
7	mit	von	-1
8	das	mit	-1
9	zu	ist	6
10	auf	das	-2
11	im	für	1
12	für	auf	-2
13	Die	sich	1
14	sich	Die	-1
15	ist	des	2
16	dem	im	-5
17	des	eine	3
18	ein	ein	0
19	nicht	nicht	0
20	eine	dem	-4
21	es	werden	9
22	auch	auch	0
23	an	Sie	43
24	Der	es	-3
25	am	an	-2
26	als	sind	15
27	hat	wird	9
28	bei	als	-2
29	aus	aus	0
30	werden	oder	26

Vergleichergebnis der top-30-Wörter Zeitung / Web

Keine sehr großen Unterschiede, da Stoppwörter funktional sind und deshalb immer verwendet werden müssen. Es gibt kleine Unterschiede:

Häufiger im Web (große positive Zahlen in der letzten Spalte) sind:

Sie, sind, oder.

Erklärung für *Sie*: Der Leser wird im Web öfter direkt angesprochen, in der Zeitung eher nicht.

Vergleich der top-30-Wörter Zeitung / Film

Die Film-Texte stammen aus frei zugänglichen Untertiteln zu Filmen und Serien.

```
select w.w_id-100 as rank, w.word as news, ww.word as movie,
cast(www.w_id-ww.w_id as signed) as diff from
deu_news_2007.words w, words ww, deu_news_2007.words www
where w.w_id>100 and w.w_id<=130 and ww.w_id=w.w_id and
www.word= binary ww.word;
```

rank	news	movie	diff
1	der	ist	14
2	die	Ich	105
3	und	ich	69
4	in	nicht	15
5	den	Sie	61
6	von	du	834
7	mit	die	-5
8	das	das	0
10	auf	es	11
11	im	und	-8
12	für	der	-11
13	Die	zu	-4
14	sich	sie	26
15	ist	in	-11
16	dem	ein	2
17	des	Das	18
18	ein	wir	60
19	nicht	mir	251
20	eine	Du	1556
21	es	mit	-14
22	auch	den	-17
23	an	Was	272
24	Der	mich	215
25	am	was	94
26	als	auf	-16
27	hat	er	12
28	bei	eine	-8
29	aus	dass	3
30	werden	Wir	80

Vergleichergebnis der top-30-Wörter Zeitung / Film

Keine sehr großen Unterschiede, da Stoppwörter funktional sind und deshalb immer verwendet werden müssen. Aber sonst seltenere Stoppwörter rutschen nach oben:

Häufiger im Film (große positive Zahlen in der letzten Spalte) sind:

Du, du, Was, mir, mich Ich, Wir, ich.

Erklärung: Viel mehr wörtliche Rede im Film.

Vergleich der top- Substantive

Rang	News	Web	Film
1	Euro	Zeit	Mann
2	Prozent	Jahren	Zeit
3	Jahr	Jahr	Leben
4	Jahren	Menschen	Frau
5	Millionen	Jahre	Gott
6	Deutschland	Seite	Leute
7	Jahre	Informationen	Vater
8	Unternehmen	Deutschland	Geld
9	Menschen	Kinder	Ordnung
10	Ende	Leben	Mutter
11	Zeit	Arbeit	Nacht
12	Berlin	Stadt	Welt
13	Polizei	Ende	Mädchen
14	Dollar	Internet	Menschen
15	Angaben	Welt	Leid
16	Freitag	Unternehmen	Hause
17	Kinder	Seiten	Freund
18	Stadt	Teil	Angst
19	Platz	Bereich	Haus
20	Woche	Verfügung	Moment
21	Sonntag	Fragen	Herr
22	Welt	Frau	Arbeit
23	Mann	Thema	Kinder
24	Milliarden	Möglichkeit	Abend
25	Deutschen	Frage	Problem
26	Mittwoch	Entwicklung	Sohn
27	Donnerstag	Fall	Sache
28	Montag	Ziel	Morgen
29	Dienstag	Kunden	Jack
30	Leben	Name	Bruder

Vergleich der Sachgebiete

Im Web finden wir

- Weniger Zahlen- und Geldangaben
- Weniger Politik
- Weniger Prominente
- Mehr Internet und Navigation (Seite(n), Bereich, Teil)
- Mehr Freizeit (Kinder, Leben, Stadt)
- Mehr Pornographie

Sprachgebrauch im Web

- Häufigere persönlicher Bericht (*ich, wir*)
- Häufigere persönliche Anrede des Lesers

Wortgruppen

Die Wortschatzdatenbanken enthalten in den Wortlisten auch Wortgruppen (Eigennamen, Redewendungen u.a.).

Die Verteilung der Wortgruppen soll in den nachfolgenden Folien untersucht werden.

Wortgruppen in de100k, nicht in deweb100k

```
mysql> select w1.* from words w1 left outer join
  SG_deweb_100K_tinyCC.words w2 on(w1.word=w2.word) where
  w1.word like "% %" and isnull(w2.word) order by w1.freq
  desc limit 60;
```

w_id	word	freq			
2983	George W.	51	6991	Jacques Chirac	14
3098	George W. Bush	49	11838	US-Präsident Bill Clinton	13
5525	SPIEGEL ONLINE	31	16658	Jürgen Röber	13
5658	Oskar Lafontaine	28	10699	Wolfgang Gerhardt	13
6462	Tour de France	23	14972	Michael Glos	13
6612	US-Präsident George W. Bush	23	11479	Otto Rehgagel	13
8152	Ariel Scharon	21	13067	Arbeitslosengeld II	12
7754	Regierende Bürgermeister	20	10308	Peter Strieder	12
7562	Boris Jelzin	20	13295	Dieter Hoeneß	12
9649	mit beschränkter Haftung	18	14378	US Open	12
7414	großen Koalition	18	11694	Kanzler Gerhard Schröder	12
9205	Wolfgang Thierse	18	8933	Ulla Schmidt	12
7244	Jan Ullrich	18	10764	Love Parade	12
8247	Gregor Gysi	18	9655	Außenminister Joschka Fischer	12
7133	Theo Waigel	18	10138	Jörg Schönbohm	11
14944	Donald Rumsfeld	17	13446	Bundesagentur für Arbeit	11
11797	Welt am Sonntag	17	17037	Präsident Bush	11
9501	Peter Struck	16	10646	AC Mailand	11
9747	kommenden Montag	15	13789	Gerhard Mayer-Vorfelder	11
6884	Manfred Stolpe	15	12083	Rechtsgeschäfte mit	11
11793	Gesellschaft mit beschränkter Haftung	15	13290	Juventus Turin	11
11111	Jürgen Rüttgers	14	15213	endgültige Entscheidung	11
10816	Bundesfinanzminister Hans Eichel	14	15284	Präsident George W. Bush	11
9034	Sat 1	14	23722	7. Dezember	11
8808	Uli Hoeneß	14	14467	Axel Springer	11
11483	Ralf Schumacher	14	15751	Präsident Wladimir Putin	11
			13595	Deutschen Fußball-Bundes	10
			22348	Osama Bin Laden	10
			13841	Erik Zabel	10
			15549	Auswärtigen Amtes	10
			8071	Berti Vogts	10
			...		

U. Quasthoff

Textdatenbanken

14

Vergleich der 300k-Korpora

```
mysql> select m.*,m2.value from SG_deweb_300K_tinyCC.meta m, de300k.meta m2 where m.attribute=m2.attribute;
```

run	attribute	value	value
1	number of distinct word forms	455085	383455
1	percentage of multi word units	7.19	13.53

```
mysql> select w1.*,w4.freq,w3.freq from de100k.words w1 left outer join SG_deweb_100K_tinyCC.words w2 on(w1.word=w2.word) left outer join SG_deweb_300K_tinyCC.words w3 on(w1.word=w3.word), de300k.words w4 where w4.word=w1.word and w1.word like "% %" and isnull(w2.word) limit 30;
```

w_id	word	freq	freq	freq
2983	George W.	51	164	3
3098	George W. Bush	49	160	3
5525	SPIEGEL ONLINE	31	74	1
5658	Oskar Lafontaine	28	68	2
6462	Tour de France	23	61	3
6612	US-Präsident George W. Bush	23	67	1
8152	Ariel Scharon	21	50	NULL
7754	Regierende Bürgermeister	20	46	1
7562	Boris Jelzin	20	70	1
9649	mit beschränkter Haftung	18	44	3
7414	großen Koalition	18	55	1
9205	Wolfgang Thierse	18	49	1
7244	Jan Ullrich	18	54	2
8247	Gregor Gysi	18	50	1
7133	Theo Waigel	18	52	1
14944	Donald Rumsfeld	17	31	NULL
11797	Welt am Sonntag	17	30	2
9501	Peter Struck	16	41	3
9747	kommenden Montag	15	41	4
6884	Manfred Stolpe	15	55	1
11793	Gesellschaft mit beschränkter Haftung	15	38	3
11111	Jürgen Rüttgers	14	32	4
10816	Bundesfinanzminister Hans Eichel	14	45	1

Beobachtungen:

1. Es gibt offensichtlich mehr Wortformen in deweb als in de. Schreibfehler und/oder breitere Themen?
2. Die Anzahlen pro Wort wachsen von de100k (Spalte 3) zu de300k (Spalte 4) linear.
3. Die in deweb100k fehlenden Wörter treten in deweb300k (Spalte5) langsam auf.

Seltene Wörter

Viele Wörter sind extrem selten. Was für Wörter sind das?

Sind seltene Wörter hauptsächlich fehlerhaft? (Nein!)

Auf der folgenden Folie:

Wörter aus 1 Million Sätzen Webtext, die nicht in einer Million Sätzen Zeitungstext vorkommen, geordnet nach Häufigkeit.

LINKS: Häufigste solche Wörter.

RECHTS: Einmal vorkommende Wörter

Beobachtung:

LINKS: Kodierungsfehler und Wörter aus der Schweiz und Österreich

RECHTS: Computerwörter u.a. Fachsprache, Verarbeitungsfehler, auch orthographische Fehler.

Wörter aus deweb1M, nicht in de1M

```
mysql> select w1.* from SG_deweb_1M_tinyCC.words w1 left
outer join de1M.words w2 on(w1.word=w2.word) where
isnull(w2.word) limit 30;
```

w_id	word	freq
3757	f¼r	400
4601	Seitenanfang	319
4673	Stammtalker	312
5939	JavaScript	238
6158	°C	229
7717	¼ber	173
8668	Jänner	149
10033	½	124
10062	WS	124
10093	nach Ihren Wünschen	123
10107	mail	123
10169	Artikelliste	122
10399	ausser	119
10759	Vorheriger	114
10769	Grüsse	114
10773	gross	114
10881	Quaschnoff D.h	107
11383	Irrtümer vorbehalten	106

```
mysql> select w1.* from SG_deweb_1M_tinyCC.words w1 left
outer join de1M.words w2 on(w1.word=w2.word) where
isnull(w2.word) and w1.freq=1 limit 30;
```

w_id	word	freq
344496	hartschwänzige	1
344497	angestupst	1
344498	Anwahlgerät	1
344499	sogannanter	1
344500	Isophon	1
344502	Embargo-Bruch	1
344503	Minnesotans	1
344504	RIMINI	1
344505	dichtgewebtes	1
344506	Chrysanthemenblüte	1
344507	Fan-Turnieren	1
344508	aufgezeichnete Daten	1
344509	Hausstaubbelastungen	1
344510	CD-Repro	1
344511	Fa.Lexware	1
344512	Inselsystem)Wie	1
344513	herunterspeichern	1
344514	Websitepromotion	1

Weitere Vergleichsmöglichkeiten: verschiedene Zeitungen

Vergleich von

- Frankfurter Allgemeine
- taz
- Bild
- Neues Deutschland
- Junge Freiheit

Oder:

- Berliner Tageblatt
- Kölner Stadtanzeiger
- Hamburger Abendblatts

Weitere Vergleichsmöglichkeiten: verschiedene Autoren

Autoren mit viel Text

- Karl Marx
- Karl May
- J.W. Goethe
- Joanne Rowling
- ...

Verwandte Fragestellung: Autorenerkennung

- Wieviel Text ist dafür nötig?
- Was sind sinnvolle Parameter?

Wikipedias

- In vielen Sprachen
- Manche recht umfangreich
- Einheitliches Format
- Zum Download als fertige HTML-Seiten erhältlich

Wikipedia Eigenschaften

- Wenig Redundanz
- Fachlich breit gefächert
- Es existieren Richtlinien für Autoren
- Fast keine 1./2. Person

Projekt Gutenberg und Verwandte

Gutenberg in Zahlen

- mehr als 3.000 Bücher, (über 1.000.000 Buchseiten)
- ca. 20.000 Gedichte
- 1.800 Märchen
- 1.200 Fabeln
- 3.500 Sagen
- ca. 3,3 Millionen Seitenabrufe (Pageimpressions) / Monat

Stand: Juni 2007

Korpus Gedichte I

	Zeitungstext	Gedichte
Die 20 häufigsten Substantive	<i>Prozent, Mark, Jahr, Jahren, Uhr, Millionen, Jahre, Menschen, Deutschland, Ende, Zeit, Stadt, Berlin, Unternehmen, SPD, Frau, Mann, Leben, Milliarden, Welt, Regierung</i>	<i>Herz, Nacht, Welt, Gott, Liebe, Leben, Himmel, Hand, Zeit, Augen, Kind, Tag, Sonne, Seele, Mann, Herr, Erde, Menschen, Licht, Haus, Tod</i>
Durchschnittliche Wortlänge	12,10	7,52
Kookkurrenzen zu „quillt“	<i>Magma, Träne, Rauch, hervor, Aus, heraus, Boxen, Bauch, Briefkasten, Ritzen, dringt, Blut, Dampf, ...</i>	<i>aus, Blut, Aus, fragend, Herzen, erstaunt, suchst, Saft, Rasen, Träne, Wunden, ein, hervor</i>
Kookkurrenzen zu „Wipfel“	<i>Bäume, Baum, Fichten, Zweige, Urapanes, Tannen, Baumes, Lesesäle, Baucis, Philemon, ...</i>	<i>die, neigen, Hains, rauschen, flüstern, Eiche, durch, glühn, alle, höchsten, Durch</i>

Korpus Gedichte II

Wir können auch die auffälligsten Substantive aus Gedichten anschauen, die in Zeitungstext vergleichsweise selten sind: *Himmel, Herzen, Mutter, Schatten, Freund, Mädchen, Blumen, Lebens, Sterne, Mensch, Morgen, Wolken, Schmerz, Fenster, Freude, Wasser, Gottes, Stunde, Tränen, Jugend, Lippen, Schein, Gesicht, Garten, Frühling, Lieder, Gedanken, Tugend, Freiheit, Schönheit, Götter, Wellen, Wahrheit, Himmels, Schlaf, Nichts, Schnee, Ritter, Flügel, Dunkel, Blicke, Gesang, Gleich, Schiff, Händen, Seelen, Felsen, Flammen, ...*

Auch bei Adjektiven, Adverbien und Verben gibt es solch für Gedichte typischen Wörter: *gleich, sprach, schöne, nieder, tausend, schönen, lieben, stillen, singen, stille, grünen, bricht, schaut, sterben, frisch, blauen, steigt, schauen, schwarzen, schlägt, einsam, schien, erwacht, wilden, willst, traurig, dunkle, freundlich, langsam, treibt, weißen, schöner, schlug, heimlich, mancher, blickt, fliegen, hellen, manches, umsonst, schweben, stürzt, weiten, heiligen, ...*

Weitere Vergleichsmöglichkeit: die Dimension Zeit

Wann sind Zeitscheiben sinnvoll?

- Beobachtung von zeitlichen Veränderungen
- Neuheiten
- Trends

Zeitintervalle (Jahr / Monat / Tag)?

Erfahrungen mit:

- Wörtern des Tages (Tag)
- Wörtern der Woche (Woche)
- Neologismen (jährlich)

Wörter des Tages

Hier gibt es täglich die auffälligen Wörter des Tages

<https://wod.corpora.uni-leipzig.de/>

Damit ein Wort heute zum Wort des Tages wird, wird seine Häufigkeit heute mit der Häufigkeit in einem Vergleichszeitraum (hier: das Vorjahr) verglichen. Die folgenden Kriterien müssen erfüllt sein:

- Eine Mindesthäufigkeit heute, damit man das Wort heute wahrnimmt.
- Eine Mindesthäufigkeit im Vergleichszeitraum, damit das Wort allgemein bekannt ist.
- Die relative Häufigkeit heute muss viel höher sein als im Vergleichszeitraum.

Die so gefundenen Wörter werden danach in Gruppen eingeteilt.

- Früher durch von Hand vorklassifizierte Wörter
- Seit ca. 2015 mittels Topic Modelling.

Wörter des Tages vom 16.6.2020

← 15.06.2020 →

Card 1 (Top Left): Infizierte, Kreis, Quarantäne, Recklinghausen, Zahl, genesen, getestet, Gesundheitsamt, positiv, Infizierten, Infektionen, Fälle, Todesfälle, Personen, Landratsamt, Neuinfektionen, Kreis Recklinghausen.

Card 2 (Top Middle): geöffnet, Hygieneregeln, Ausbreitung, Masken, Hygiene-Maske, Abstandsregeln, Impfstoff, NRW, Pandemie, Corona, öffnen, Lockdown, gelten, Virus, Lockerungen, Öffnung, Mindestabstand, Infektionszahlen, Infektionsgeschehen.

Card 3 (Top Right): Nationalgarde, Polizeieinsatz, Hutchinsohn, Bottoms, George, Floyds, Atlanta, Cuomo, Afroamerikaners, Brooks, Chauvin, Rassismus, brutalen, Polizeigewalt, Afroamerikaner, Floyd, Minneapolis, Hydroxychloroquin, Malaria-Medikament, Rassismus in den USA.

Card 4 (Middle Left): Einreisende, aufgehoben, Einreisebeschränkungen, Sommerurlaub, Touristen, Reisewarnung, 15. Grenzkontrollen, Juni, Anfang Juni, Reisewarnungen, Einreise, einreisen.

Card 5 (Middle Middle): Kreis-Gesundheitsamt, Gesundheitsministerium, Mai, April, Tirol, Update, Testergebnis, baden-württembergischen.

Card 6 (Middle Right): F-15C, Videokonferenz, Montag, Lakenheath, Hoensbroech.

Card 7 (Bottom Left): Aktienoptionen, Augustus, Philipp Amthor, Amthor.

Card 8 (Bottom Middle): Wolters, Tote Mädchen lügen nicht, Maddies, McCann, Maddie, Maddie McCann.

Card 9 (Bottom Right): Formentera, Menorca, 10.900, Urlauber, Mallorca, 2312, Balearen-Inseln.

Zum Themengebiet

U. Quasthoff

Textdatenbanken

Wörter der Woche

Im Jahr 2006 bekamen wir die Anfrage, aus den *Wörtern des Tages* die *Wörter der Woche* zu ermitteln.

Es stellt sich heraus, dass ein einfaches Zusammenstellen der Wörter des Tages für sieben Tage und die Auswahl der häufigsten davon nicht zielführend ist.

Es musste ein wesentlich komplexeres Kriterium entwickelt werden.

Wörter der Woche (WdW)

Akkumulierte Wörter des Tages - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://wortschatz.uni-leipzig.de/~steresniak/WdW.html

Dieser Service befindet sich derzeit noch in der Erprobung.
Sollten Sie Verbesserungsvorschläge, Kritik oder Fragen (diese Seite betreffend) haben, so schreiben Sie bitte eine eMail an teresniak@informatik.uni-leipzig.de.

Akkumulierte Wörter des Tages vom 23.6.2007 bis 29.6.2007

Wortschatz : Wörter des Tages für heute : (BETA)

Ereignis Direktwahl · Trainingsauftakt

Organisation Hilton · HVB · UNESCO · ISAF · Ebay · Qimonda · Friedrich-Loeffler-Institut

Person Bartels · Ed · Dimitri

Sportler Federer · Allofs · Hamann · Hingis · Alinghi · Langer · Magath · Klasnic · Leichtathleten · Anna-Lena Grönefeld · Gebrselassie · Bruno Spengler · Radcliffe · Heynckes · Lukas Podolski · Michael Berrer · Maradona · Barker · Coutts · Cejka · Filippo Vollandri

Politiker

Ort Antalya · Uelzen · Venezuela

Schlagwort Geflügel · iPhones · Tarifkonflikt · Squeeze-Out · EBIT-Marge

Sport

Künstler Christiansen · Libeskind · Grass · Marilyn · Lou · Chipperfield · Victoria Beckham

Wörter der letzten 7 Tage, bis einschließlich 29.6.2007

Die Auswahl der Wörter geschieht nach Schwellwerten,
die sich aus den WdW oder aus dem Gesamtaufreten der Wörter im Auswertungszeitraum
ergeben.

© 2002 - 2004 Deutscher
Wortschatz
Alle Rechte vorbehalten

Die Daten werden aus sorgfältig ausgewählten öffentlich zugänglichen Quellen automatisch erhoben. Die Beispielsätze werden automatisch

Kriterien für WdW I

Dabei gibt es zwei Filtermechanismen.

Zum Einen muss ein WdW mehrfach WdT gewesen sein, zum Anderen muss ein WdW auch ein gewisses Mindestvorkommen in den Quellen erfüllen. Beide Schwellwerte sind Tagesmittel. Also beispielsweise wird ein Wort WdW, wenn es im Betrachtungszeitraum (üblicherweise sieben Tage) an mindestens drei Tagen WdT war. Der Schwellwert (zu erfüllende Durchschnitt) ist also $3/7$.

Analog bei der Anzahl der absoluten Vorkommen in den WdT-Quellen. Hier muss ein Wort beispielsweise 150x innerhalb einer Woche gesehen worden sein, also im Mittel rund 20x pro Tag.

Da insgesamt mehr Politiker als Künstler erwähnt werden, unterscheiden sich die Schwellwerte ja nach Kategorie.

Kriterien für WdW II

```
## filter für anzahl der tage an denen ein wort wdt war,  
## gerechnet auf eine woche (3/7. bedeutet 3x wdt in einer woche)  
  1 : 2/7., ## sportler  
  2 : 2/7., ## künstler  
  3 : 5/7., ## politiker  
  4 : 3/7., ## person  
  5 : 4/7., ## sport  
  6 : 3/7., ## organisation  
  7 : 3/7., ## ereignis  
  8 : 4/7., ## ort  
  9 : 3/7.  ## schlagwort  
  ## 10 - nicht anzeigen  
## filter für die anzahl der aufteten im beobachtungszeitraum  
## 50/7. bedeutet: das wort wurde mindestens 50x gesehen über der  
## spanne von 7 tagen - sozusagen ein mittelwert über den zeitraum  
  1 : 50/7., ## sportler  
  2 : 50/7., ## künstler  
  3 : 150/7., ## politiker  
  4 : 150/7., ## person  
  5 : 150/7., ## sport  
  6 : 100/7., ## organisation  
  7 : 60/7., ## ereignis  
  8 : 100/7., ## ort  
  9 : 100/7. ## schlagwort  
  ## 10 - nicht anzeigen
```




jetzt - das Magazin für den 18./19.03. 2006

1 | **Wochenwörter** | **Leben** | 17.03.2006 19:41

Vogelgrippe bis Herzinfarkt - die Wochenwörter

Text: **dirk-vongehlen**

Welche Begriffe bestimmten diese Woche?

In Kooperation mit der **Uni Leipzig** bietet jetzt.de jede Woche einen besonderen Service: die Wochenwörter.

Am Leipziger Institut für Informatik werten Prof. Uwe Quasthoff und sein Team jeden Tag die „Wörter des Tages“ aus. Dazu werden verschiedene Tageszeitungen und Newsdienste auf die Frage hin ausgewertet: Welche Wörter werden besonders häufig genutzt? Die Aktualität eines Begriffs ergibt sich aus seiner Häufigkeit, verglichen mit seiner durchschnittlichen Häufigkeit über längere Zeit hinweg.

jetzt.de veröffentlicht wöchentlich die Schlagworte der Woche. Hier die Begriffe, die vom 10. bis 17. März am häufigsten genutzt wurden:

