



# VIDA – Visuelle Dokumenten-Analyse

- Interaktive Benutzeroberfläche zur visuellen Darstellung von Dokumenten und den Analyse-Ergebnissen
- Analyse von Dokumenten nach vorgegebenen Kriterien, z. B.
  - Anzahl der Worte im Text
  - Entitäten wie Orte, Namen, Organisationen
- Statistische Analyse von Texten
  - Schlüsselwörter
  - N-Gramme
  - Wortassoziationen
- Automatische Landessprachenerkennung für
  - deutsche
  - englische
  - französische
  - italienischeTexte
- Zusammenfassung des Textes zum schnelleren Querlesen und Erfassen
- Ausgabe der Analyse-Ergebnisse als Tabellen und Graphiken
  - Worthäufigkeiten
  - Zusammenhänge
- Markieren von Texten nach relevanten Kriterien „Text Tagging“
- Relationen zwischen Schlüsselwörtern
- Finden von ähnlichen Texten

## Finden von wichtigen Text- Entitäten



Worte in den verschiedenen Bedeutungskategorien werden automatisch gefunden.

Die einzelnen Bedeutungskategorien werden im Text unterschiedlich markiert.

## Metadaten

Generelle Angaben zum analysierten Dokument werden in Metadaten zusammengefasst.

Diese Metadaten enthalten z.B. folgende Informationen:

- Dokumentenname
- Quellverzeichnis
- Erstellungsdatum
- Anzahl der Wörter, Zeilen und Zeichen
- Landessprache

Attribute		Value
ENV.date		10. Mai 2005
ENV.OS		Windows XP 5.1
USER.name		uueb
USER.language		de
USER.region		
USER.userDir		C:\TOPS
NUM_CHARS		16683
NUM_LINES		46
NUM_SENTENCES		140
FILE.name		einstein_de.1.txt
FILE.byteLength		16689
FILE.directory		c:\TOPS\Documents
FILE.date		Do, 01 Januar 1970
FILE.encoding		Cp1252
LANGUAGE		Deutsch - Deutschland
LANGUAGE_RESOURCES		c:\TOPS\LanguageResources\de_DE
NUM_WORD_TYPES		938
NUM_WORD_TOKENS		2310
NUM_CONTENTWORD_TYPES		777
NUM_CONTENTWORD_TOKENS		1140
AV_WORD_LENGTH		5.861471861471862
NUM TEXT LINES		22

## Wortstatistik

Eine Wortstatistik gibt Auskunft über die häufigsten Wörter bzw. über die häufigsten bedeutungstragenden Wörter, d.h. Schlüsselwörter.

Word Frequencies			Content Word Frequencies		
Rank	Word	Freque...	Rank	Word	Frequency
1	die	92	1	einstein	51
2	der	84	2	einsteins	14
3	und	63	3	zeit	10
4	er	62	4	theorie	9
5	in	55	5	relativitätstheorie	9
6	einstein	51	6	physik	9
7	im	41	7	albert	8
8	den	25	8	später	7
9	sich	24	9	universität	7
10	von	24	10	jahr	6
11	wurde	24	11	familie	6
12	an	23	12	arbeit	5
13	zu	23	13	patentamt	5
14	am	19	14	zog	5
15	mit	19	15	frau	5
16	auf	18	16	jahren	5
17	des	16	17	dezember	5
18	seine	16	18	berlin	5
19	einer	16	19	starb	5
20	dem	15	20	artikel	5
21	um	15	21	zurich	5
22	für	15	22	akademie	5
23	war	14	23	berner	5
24	nach	14	24	april	5
25	einsteins	14	25	geboren	4
26	ein	14	26	lichtablenkung	4
27	eine	13	27	erhielt	4

## Schlüsselworte – Wort-N-Gramme

Statistische Ermittlung und Auflistung aller Schlüsselwörter und der Wort-N-Gramme (häufige Folgen von Wörtern).

Keywords	Word ngrams
einstein	albert einstein
einsteins	starb einsteins
zeit	jahr später
theorie	mileva maric
relativitätstheorie	preußischen akademie
physik	berner patentamt
albert	einheitlichen feldtheorie
später	allgemeinen relativitätstheorie
universität	schweizer staatsbürgerschaft
jahr	max planck
familie	perihelbewegung des merkur
arbeit	zeit später
patentamt	

# Landessprachenerkennung

```

Metadata | KWIC | Collocations | Word lists | Ngrams | Tag Info | Log
Loading file einstein_de.1.txt
Elapsed time=0.0s

Start analysis of document.
AUTOMATIC LANGUAGE DETERMINATION
count for language 0 =35
count for language 1 =288
count for language 2 =9
count for language 3 =20
Total word count=1001 words.
Required recognition threshold=50 words.
LANGUAGE=de_DE
Language resource time=0.016s
    
```

Automatische Ermittlung der Sprache eines Textes. Die Landessprachenerkennung ist gegenwärtig für die Sprachen

- deutsch
- englisch
- französisch
- italienisch

realisiert. Weitere Sprachen können auf Anfrage realisiert werden.

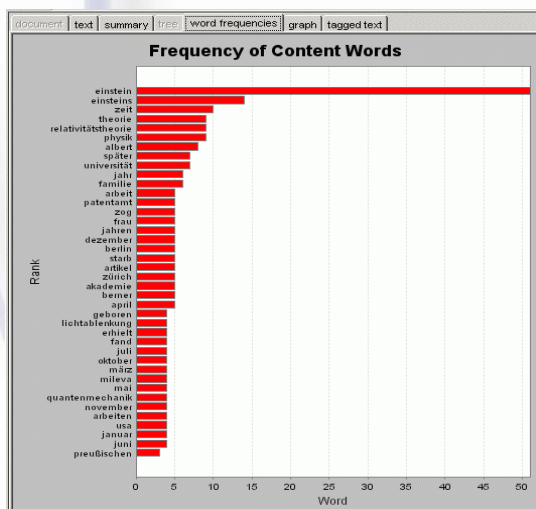
# Zusammenfassung

## Summary

- (S 1) Albert Einstein wurde am 14. März 1879 als erstes Kind der jüdischen Eheleute Hermann und Pauline Einstein, geb Koch, in Ulm geboren.
- (S 2) Im Juni 1880 siedelte die Familie nach München über, wo Hermann Einstein und sein Bruder Jakob die elektrotechnische Firma Einstein & Cie. gründeten.
- (S 3) Am 18. November 1881 wurde Albert Einsteins Schwester Maria - genannt Maja - geboren.
- (S 7) Einsteins Kindheit verlief, bis auf den für die Familie beunruhigenden Umstand, dass er erst sehr spät sprechen lernte, normal.
- (S 8) Um ihn auf die Schule vorzubereiten, erhielt er ab 1884 Privatunterricht.
- (S 9) Ein Jahr später begann er mit dem Violinunterricht.
- (S 10) Ab 1885 besuchte er die Peterserschule, eine katholische Volksschule, in München und wechselte 1888 ins dortige Luitpold-Gymnasium.

Zur Ermöglichung eines schnellen Querlesen des Textes kann eine automatische Zusammenfassung erzeugt werden.

# Visualisierung – Häufigkeit



Visuelle Darstellung der Häufigkeit der für den Text definierten Schlüsselworte.



# VIDA – Daten

---

## **Betriebssysteme**

Windows, Linux

## **Automatische Landessprachenidentifikation**

- deutsch
- englisch
- französisch
- italienisch
- weitere Sprachen auf Anfrage

## **Text-Analyse**

- englisch
- deutsch
- weitere Sprachen auf Anfrage

## **Statistische Analyse von Texten**

basierend auf Wortlisten und Regeln

## **Kein linguistisches Wissen notwendig**

## **Funktionalität**

- Wortstatistik
- Automatische Bestimmung von Schlüsselwörtern und N-Grammen
- Wortassoziationen
- Histogramm für Worthäufigkeiten
- Markierung (Tagging) von Entitäten: Eigennamen, Orten, Datum usw.
- Visualisierung der Zusammenhänge und Entitäten

## Bestellinformationen

---

<b>VIDA-Inter</b>	VIDA Software für interaktive Nutzung, inkl. graphischer Benutzeroberfläche
<b>VIDA-Prod</b>	VIDA Programm als Modul zur Integration in übergreifende Anwendungen
<b>VIDA-Conf</b>	VIDA Konfigurations-Software für Anpassung auf neue Domänen



## Unsere Ziele

### Technologie

... in der Entwicklung und in der Unternehmensführung ist auf dem neuesten Stand und repräsentiert ein hohes Niveau.



### Qualität

... in allen Bereichen unseres Unternehmens erachten wir als Voraussetzung für eine risikolose und erfolgreiche Zusammenarbeit mit unseren Kunden und Geschäftspartnern.



### Stellung im Markt

... ist geprägt von den Erfahrungen in der Signal- und Informationsverarbeitung. Wir stellen uns gerne dem Wettbewerb.



### Leistungsangebot

... ist umfassend, vollständig und bedarfsgerecht. Wir bieten als Systemhaus Standardgeräte, Systeme und Dienstleistungen an.



### Mitarbeiter

... bilden die Wurzeln des Unternehmens und erbringen die Leistung, die zur Erhaltung und zum Ausbau der technischen Basis und der vertrauensvollen Zusammenarbeit notwendig ist.



### Wachstum

... auf stabiler technischer und wirtschaftlicher Grundlage streben wir im In- und Ausland an.



### Vertrauen

... im Verhältnis zu unseren Geschäftspartnern und innerhalb des eigenen Unternehmens ist die Basis für unsere Geschäfte.



## MEDAV GmbH

GRÄFENBERGER STRASSE 32 - 34  
D-91080 UTTENREUTH

HOMBURGER PLATZ 3  
D-98693 ILMENAU

TELEFON: +49-9131-583-0  
FAX: + 49-9131-583-1  
E-MAIL: [info@medav.de](mailto:info@medav.de)  
[www.medav.de](http://www.medav.de)

w707od.0m1