

**MELANIE
VIDA**

Visual Document Analysis

MELANIE – VIDA, Visual Documents Analysis

(Analysis and Visualisation of Texts)

VIDA – Visual Document Analysis

- Interactive user interface for visual representation of documents and analysis results
- Analysis of documents against defined criteria e. g.
 - number of words in the text
 - entities such as places, names, organizations
- Statistical analysis of texts
 - keywords
 - n-grams
 - word associations
- Automatic national language identification for texts in
 - German
 - English
 - French
 - Italian
- Text abstraction for speed reading and faster scanning of information
- Analysis results displayed in tables and graphs
 - word frequencies
 - associations
- Text tagging, where texts are highlighted according to relevant criteria
- Relations between keywords
- Finding similar texts

Finding important text entities

Words in the various meaning categories are found automatically. Different highlighting is used in the text for each of the meaning categories.

Albert Einstein was born at **Ulm**, in **Matttshberg, Germany**, on **March 14, 1879**. Six weeks later the family moved to **Musich** and he began his schooling there at the **Luitpold Gymnasium**. Later, they moved to **Italy** and Albert continued his education at **Aarau, Switzerland** and in 1896 he entered the Swiss **Federal Polytechnic School** in **Zurich** to be trained as a teacher in physics and mathematics. In 1901, the year he gained his diploma, he acquired Swiss citizenship and, as he was unable to find a teaching post, he accepted a position as technical assistant in the Swiss **Patent Office**. In 1905 he obtained his doctor's degree.

During his stay at the **Patent Office**, and in his spare time, he produced much of his remarkable work and in 1908 he was appointed Privatdozent in **Berne**. In 1909 he became Professor Extraordinary at **Zurich**, in 1911 Professor of Theoretical Physics at **Brno**, returning to **Zurich** in the following year to fill a similar post. In 1914 he was appointed Director of the Kaiser Wilhelm **Physical Institute** and Professor in the **University of Berlin**. He became a German citizen in 1914 and remained in **Berlin** until 1933 when he renounced his citizenship for political reasons and emigrated to **America** to take the position of Professor of Theoretical Physics at **Princeton**. He became a **United States** citizen in 1940 and retired from his post in 1945.

After World War II, **Einstein** was a leading figure in the World Government Movement, he was offered the Presidency of the State of **Israel**, which he declined, and he collaborated with Dr. **Chaim Weizmann** in establishing the **Hebrew University of Jerusalem**.

Metadata

General information on the analysed document is compiled as metadata. This metadata includes the following information, for example:

- Document name
- Source directory
- Creation date
- Number of words, lines and characters
- National language

Attribute	Value
ENV.date	10. Mai 2005
ENV.OS	Windows XP 5.1
USER.name	ueeb
USER.language	de
USER.region	
USER.userDir	C:\TOPS
NUM.CHARS	5357
NUM.LINES	109
NUM.SENTENCES	37
FILE.name	einstein1.txt
FILE.byteLength	5357
FILE.directory	c:\TOPS\Documents
FILE.date	Do, 01 Januar 1970
FILE.encoding	Cp1252
LANGUAGE	Englisch - Vereinigte Staaten von Amerika
LANGUAGE_RESOURCES	c:\TOPS\LanguageResources\en_US
NUM.WORD_TYPES	356
NUM.WORD_TOKENS	810
NUM.CONTENTWORD_TYPES	288
NUM.CONTENTWORD_TOKENS	402
AV.WORD_LENGTH	5.103703703703704
NUM.TEXT_LINES	96
NUM.PARAS	12

Word statistics

A set of word statistics provides information on the most frequently used words and the most frequently used words relevant to the content i.e. keywords.

Word Frequencies			Content Word Frequencies		
Rank	Word	Frequ...	Rank	Word	Frequency
1	the	59	1	theory	12
2	of	48	2	einstein	8
3	in	42	3	work	7
4	and	36	4	physics	7
5	he	34	5	relativity	6
6	his	25	6	mechanics	5
7	to	20	7	albert	4
8	a	13	8	professor	4
9	theory	12	9	problems	4
10	was	11	10	time	3
11	at	10	11	post	3
12	with	9	12	america	3
13	einstein	8	13	important	3
14	on	7	14	world	3
15	work	7	15	special	3
16	physics	7	16	berlin	3
17	relativity	6	17	nobel	3
18	as	5	18	statistical	3
19	this	5	19	year	3
20	mechanics	5	20	movement	3
21	for	4	21	continued	3
22	later	4	22	quantum	3
23	from	4	23	princeton	3
24	albert	4	24	institute	3
25	professor	4	25	swiss	3
26	an	4	26	zurich	3
27	problems	4	27	war	2

Keywords – Word n-grams

Statistical determination and listing of all keywords and word n-grams (common sequences of words).

Keywords	Word ngrams
theory	theory of relativity
einstein	special theory
work	albert einstein
physics	statistical mechanics
relativity	quantum theory
mechanics	general theory
albert	princeton new jersey
professor	continued to work

National language identification

Automatic determination of the language of a text. National language identification is currently implemented for the following languages:

- German
- English
- French
- Italian

Other languages can be implemented on request.

```

Metadata | WWW | 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=en_EN
Language resource time=0.016s
    
```

Abstraction

A summary can be generated automatically to enable speed-reading of the text.

Summary

(S 1) Albert Einstein was born at Ulm, in Württemberg, Germany, on March 14, 1879.

(S 2) Six weeks later the family moved to Munich and he began his schooling there at the Luitpold Gymnasium.

(S 3) Later, they moved to Italy and Albert continued his education at Aarau, Switzerland and in 1896 he entered the Swiss Federal Polytechnic School in Zurich to be trained as a teacher in physics and mathematics.

(S 4) In 1901, the year he gained his diploma, he acquired Swiss citizenship and, as he was unable to find a teaching post, he accepted a position as technical assistant in the Swiss Patent Office.

(S 5) In 1905 he obtained his doctor's degree.

(S 6) During his stay at the Patent Office, and in his spare time, he produced much of his remarkable work and in 1908 he was appointed Privatdozent in Bern.

(S 7) In 1909 he became Professor Extraordinary at Zurich, in 1911 Professor of Theoretical Physics at Prague, returning to Zurich in the following year to fill a similar post.

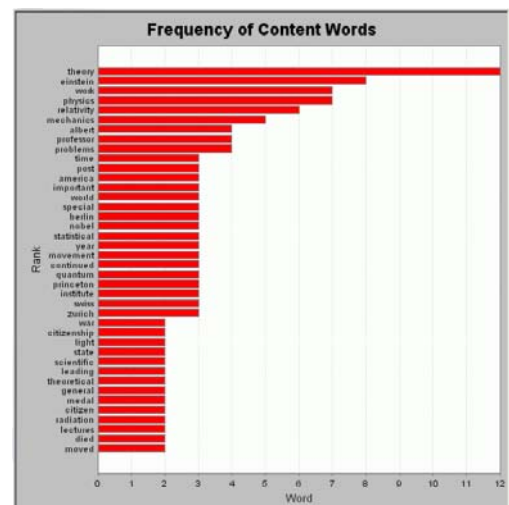
(S 8) In 1914 he was appointed Director of the Kaiser Wilhelm Physical Institute and Professor in the University of Berlin.

(S 9) He became a German citizen in 1914 and remained in Berlin until 1933 when he renounced his citizenship for political reasons and emigrated to America to take the position of Professor of Theoretical Physics at Princeton*.

(S 10) He became a United States citizen in 1940 and retired from his post in 1945.

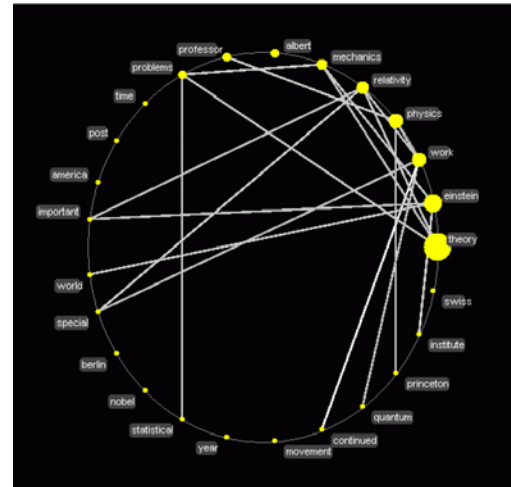
Visualisation – Frequency

Visual representation of the frequency of specified keywords in the text.



Visualisation – Associations

VIDA includes the visualisation option of an association graph for delivering faster identification of the text content.



Visualisation – Entities

VIDA provides coloured highlighting of entities and also lists entities and keywords in a table.

Rank	Word	Frequency	Rank	Word	Frequency
1	ein	42	1	theory	4
2	st	40	2	work	4
3	er	36	3	relativity	3
4	al	34	4	mechanics	3
5	er	29	5	physics	3
6	er	27	6	albert	3
7	er	21	7	princeton	2
8	er	17	8	institute	2
9	er	14	9	swiss	2
10	er	13	10	quantum	2
11	er	12	11	princeton	2
12	er	11	12	institute	2
13	er	10	13	swiss	2
14	er	9	14	quantum	2
15	er	8	15	princeton	2
16	er	7	16	institute	2
17	er	6	17	swiss	2
18	er	5	18	quantum	2
19	er	4	19	princeton	2
20	er	3	20	institute	2

Finding similar texts

VIDA provides a function that assesses the similarity of texts, for text location. This function automatically lists in a table all the texts analysed, in order of decreasing similarity. In addition, the most important matches are output in keywords.

Ähnlichkeit	Text	Übereinstimmung
1.0	einstei_n_de_1	einstei_n einstei_n zeit physik relativitätstheorie theorie albert später
0.75	einstei_n_de_3	einstei_n einstei_n zeit physik relativitätstheorie albert
0.5	einstei_n_de_2	einstei_n zeit relativitätstheorie albert
0.25	einstei_n_en_2	einstei_n albert
0.125	einstei_n_en_1	einstei_n
0.125	einstei_n_en_3	einstei_n
0.125	einstei_n_en_4	einstei_n
0.125	einstei_n_de_4	einstei_n
0	en_alqaida	
0	en_bush	

VIDA – Specification

Operating systems

Windows, Linux

Automatic national language identification

- German
- English
- French
- Italian
- Other languages on request

Text analysis

- English
- German
- Other languages on request

Statistical analysis of texts

Based on word lists and rules

No linguistic knowledge necessary Functions

- Word statistics
- Automatic determination of keywords and n-grams
- Word associations
- Histogram for word frequencies
- Marking (tagging) of entities: proper names, places, date, etc.
- Visualisation of associations and entities

Ordering information

VIDA-Inter Vida tool, to be used interactively, incl. graphical user interface

VIDA-Prod Vida program to be embedded in software environment

VIDA-Conf Vida configuration software, used for adaptation to new domain

Company Principles and Policy

Technology

... in development and company management is state-of-the-art, and represents only the best.

Quality

... in all areas of our company is regarded as the almost requirement for risk-free and successful cooperation with our customers, and business partners.

Market Position

... we are the specialists in the field of signal and data processing as well as pattern recognition, and we are glad to face competition.

Colleagues

... form the roots of our company, and give the performance required for maintaining and building the technical base, and close personal cooperation we have with our clientele.

Growth

... we strive toward a healthy, stable foundation at home and abroad.

Services

... are comprehensive and complete. As a full-system company we offer standard equipment, systems, and services.

Trust

... in the relationships to our business partners, and within our own company forms the basis of our business.



If you would like further Information about ELAMAN,
or would like to discuss a specific requirement or project, please contact us at:

Elaman GmbH
German Security Solutions
Seitzstr. 23
80538 Munich
Germany

Tel: +49-89-24 20 91 80
Fax: +49-89-24 20 91 81
info@elaman.de
www.elaman.de