

sas[®]club

Der Business Analytics Club für SAS User



SAS[®] CONTEXTUAL ANALYSIS IN ACTION ERFAHRUNGEN AUS EINEM EIN SELBSTVERSUCH

GERHARD SVOLBA
COMPETENCE CENTER ANALYTICS
WIEN, 17. NOVEMBER 2015



- **Was macht das Produkt?**

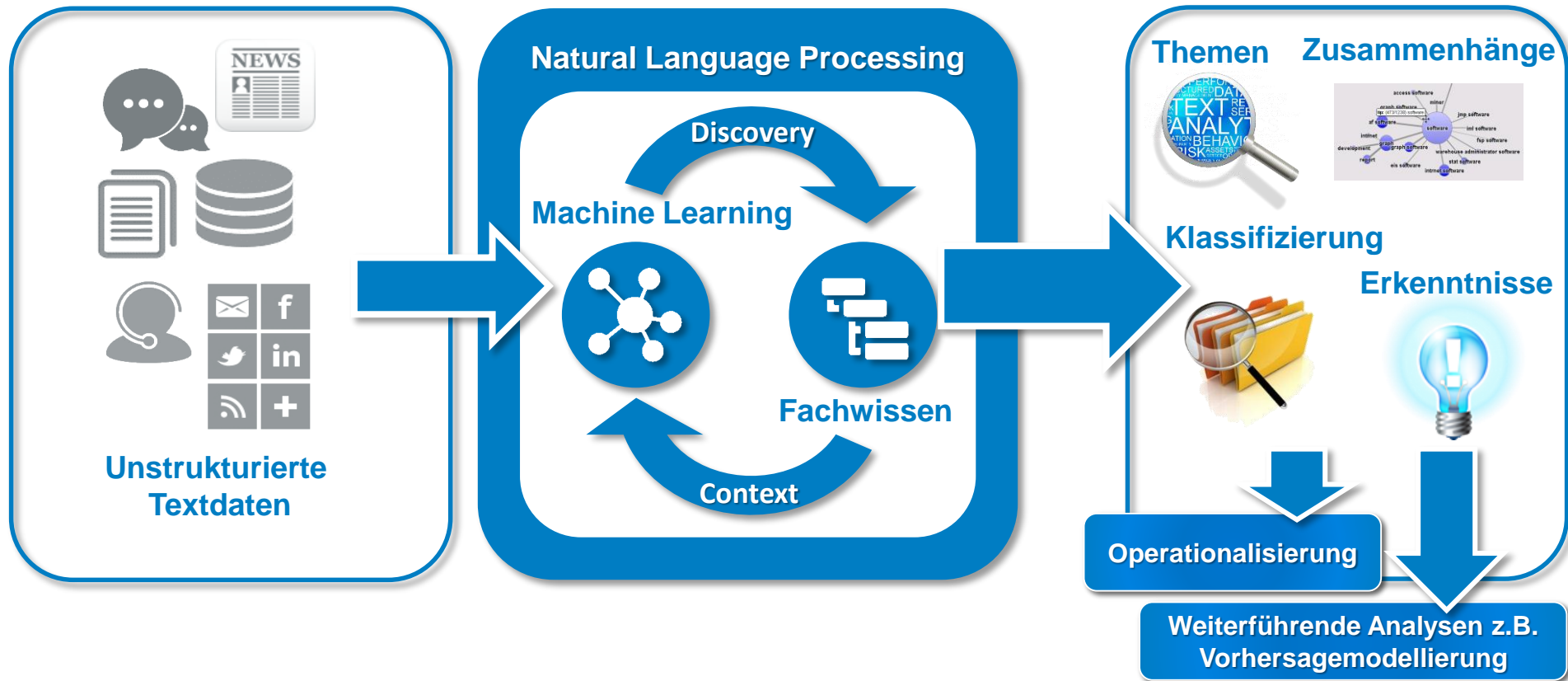
- Es erlaubt große Sammlungen von Text-Dokumenten zu analysieren, Sentiments zu identifizieren und robuste Modelle zur Kategorisierung und Extraktion von Inhalten zu erstellen.

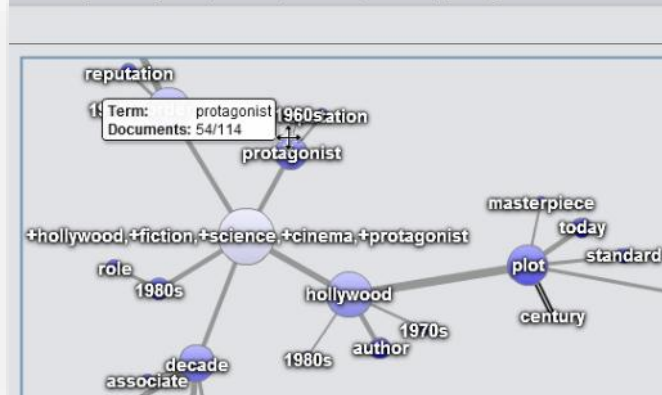
- **Wie funktioniert das?**

- Kombination von automatischer Erkennung, Machine-Learning Methoden, Linguistischer Regeln und Experten-Input zur Entwicklungen eines Kategorisierungs/Extraktions-Modells
- Automatische Identifikation von Themen in den Dokumenten, Definition von Kategorien und Überarbeitungen durch den Text-Analysten
- Interaktives Testen und visuelle Exploration über ein HTML5-Browser Interface mit Wizards und Context-sensitiver Hilfe.

- **Wie integriert sich SAS Contextual Analysis in das SAS Portfolio?**


- Integrierter Teil der SAS Plattform (SAS Metadata Server, ...)
- (Mögliche) Ergänzung zum SAS Text Miner
- Ergebnis-Darstellung mit SAS Visual Analytics, Weiterwendung in SAS Analytik Produkten





Create Rule

Name:

Format: 

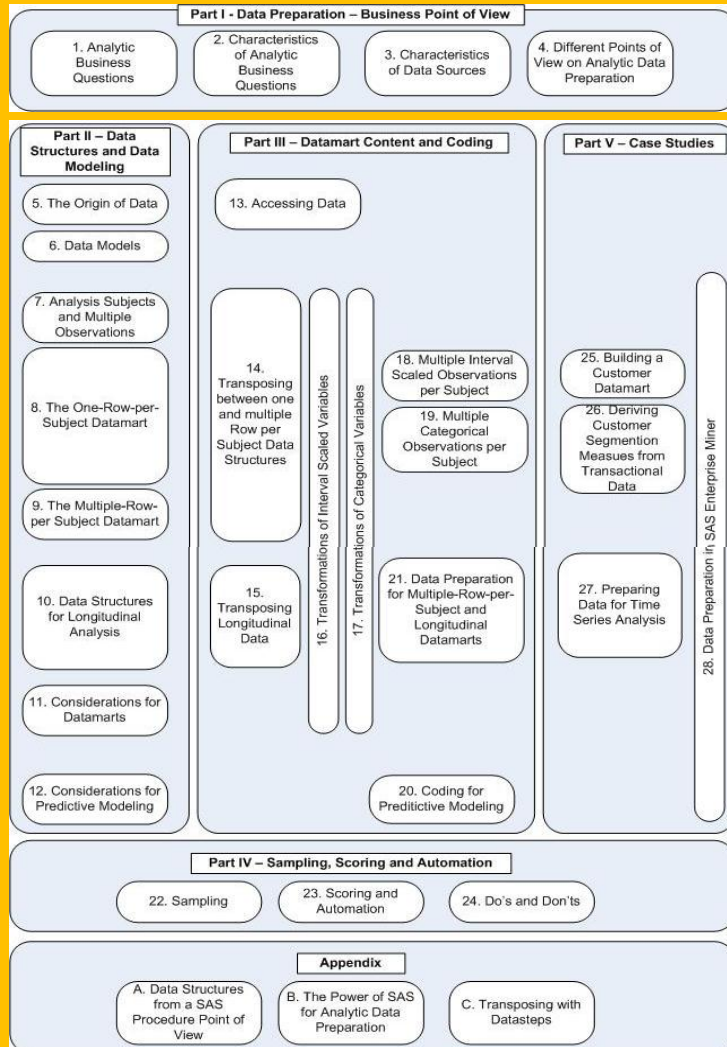
Syntax check results:

Terms and Synonymy...	Number of Docum...
▶ love	628
▶ want	593
▶ turn	587
▶ bad	587
▶ young	579
▼ great	568
great	513
greatest	65
greater	27

Topics					
<div> <div> <div></div> <div></div> <div></div> </div> <div> <div></div> <div></div> <div></div> </div> <div> <div></div> <div></div> <div></div> </div> </div> <div>?</div>					
Topics	😊 %	😡 %	😞 %	Number of Doc...	
▼ All Topics (1527)					
📄 +show,+rate,+recommend,+kid,nudity	36	59	5	155	
📄 +hollywood,+fiction,+science,+cinema,+protagonist	12	87	1	114	
📄 +comedy,+funny,+joke,+laugh,+humor	31	64	5	140	
📄 +school,+mother,+girl,+boy,+woman	21	74	5	128	
📄 +motion,+viewer,+moment,+minute,+relationship	23	74	2	168	
📄 bond,+bond,james bond,+spectre,+connery	18	76	6	33	
📄 +alien,+science,+effect,+monster,+horror	26	72	3	113	
📄 +crime,+police,+murder,+thriller,cop	9	90	1	149	
📄 +war,+soldier,+battle,+american,+military	11	85	4	119	
📄 +granger,gauge,granger movie gauge,+comedy,dr.	37	58	5	136	

DER SELBSTVERSUCH

DIE AUSGANGSBASIS: 2 BÜCHER VON SAS-PRESS



Data Quality Defined

Case Studies – Definition- Availability –
Quantity – Completeness – Correctness –
Predictive Modeling – Analytics –
Process Considerations

Profiling and Improvement

Missing Values – Time Series Data –
Across Tables – Data Quality with Analytics –
SAS Analytic Tools

Simulation Studies

Introduction – Predictive Modeling –
Time Series Forecasting

- "Unsupervised" Analyse
- Welche Themen werden gefunden?
- In welche Cluster können die Dokumente eingeteilt werden?
- Wodurch beschreiben sich diese Cluster?

★ Favorites

Desktop

Downloads

Recent Places

Libraries

Documents

Music

Pictures

Videos

Computer

Local Disk (C:)

Data (D:)

data

Gerhard

Bilder

Data

pages

DQFA und DPFA

FH-Steyr

Removed Chapters

Name ^	Date modified	Type	Size
AppA_new.docx	4/17/2012 11:00 AM	Microsoft Word Doc...	62 KB
AppB_new.docx	4/17/2012 11:03 AM	Microsoft Word Doc...	62 KB
AppC_new.docx	4/17/2012 11:04 AM	Microsoft Word Doc...	136 KB
AppD_new.docx	4/17/2012 11:07 AM	Microsoft Word Doc...	73 KB
AppE_new.docx	4/19/2012 3:04 PM	Microsoft Word Doc...	184 KB
AppendixA.new.doc	1/8/2007 11:17 AM	Microsoft Word 97 -...	284 KB
AppendixB.new.doc	1/8/2007 11:17 AM	Microsoft Word 97 -...	334 KB
AppendixC.new.doc	1/8/2007 11:17 AM	Microsoft Word 97 -...	328 KB
chap1.new.doc	1/8/2007 11:17 AM	Microsoft Word 97 -...	503 KB
chap1_new.docx	4/13/2012 11:59 AM	Microsoft Word Doc...	141 KB
chap2.new.doc	1/8/2007 11:17 AM	Microsoft Word 97 -...	294 KB
chap2_new.docx	4/13/2012 12:01 PM	Microsoft Word Doc...	104 KB
chap3.new.doc	1/8/2007 11:17 AM	Microsoft Word 97 -...	301 KB
chap3_new.docx	4/13/2012 1:11 PM	Microsoft Word Doc...	99 KB
chap4.new.doc	1/8/2007 11:17 AM	Microsoft Word 97 -...	287 KB
chap4_new.docx	4/18/2012 9:51 AM	Microsoft Word Doc...	87 KB
chap5.new.doc	1/8/2007 11:17 AM	Microsoft Word 97 -...	488 KB
chap5_new.docx	4/13/2012 1:21 PM	Microsoft Word Doc...	105 KB

SOFTWARE DEMO

AUTOMATISCHE SYNONYM-ERKENNUNG, STEMMING, STOP-LISTEN BERÜCKSICHTIGUNG

Books_Breakfast

Terms

Kept Terms

Terms and Synonyms	Number of Documents	Concept
transactional	24	
advantage	24	
overview	24	
standard	24	
analysis subject	24	NOUN_GROUP
analysis subject	19	NOUN_GROUP
analysis subjects	13	NOUN_GROUP
analysis subjects	5	PROP_MISC
analysis subject	1	PROP_MISC
monthly	24	
place	24	
leave	23	
underlying	23	
yes	23	
factor	23	
purchase	23	
otherwise	23	
simply	23	
common	23	
validation	23	
target variable	23	NOUN_GROUP
target variable	21	NOUN_GROUP
target variables	8	NOUN_GROUP
target variables	1	PROP_MISC



Dropped Terms

Documents

Terms and Synonyms	
a	
example	
and	
use	
have	
as	
where	
from	
be	
by	
following	
if	
that	
for	
of	
this	
with	
can	
to	
an	
do	
not	
more	
only	






Topics


Run












 

View

Topics



Topics	Number of Documents
▼ All Topics (59)	
 +shop,+promotion,+label,+productgroup,+pg	3
 detection,+outlier,+node,outlier detection,jmp	3
 +simulation,+training,+training data,+response,+random	4
 +record,correctness,+systematic,+bias,+database	4
 +multiple observation,+analysis subject,+entity,+account,+measurement	4
 +title,+profile,+var,+missing record,ts_profile_chain	3
 +score,historic,+historic snapshot,people,+snapshot	6
 mape,+history,+time history,mape,+disturbance	5
 f,+transpose,+weight,data,+root	6
 +access,+file,+text,+relational,+relational database	5
 +boat,+sail,wind,+race,gps	1

SOFTWARE DEMO

ANZEIGE DER DOKUMENTE PRO THEMA

File Help

Books_Breakfast

Topics Run View

Topics

Topics

Number of Documents

All Topics (59)

+shop,+promotion,+label,+productgroup,+pg3

detection,+outlier,+node,outlier detection,jmp3

+simulation,+training,+training data,+response,+random4

+record,correctness,+systematic,+bias,+database4

+multiple observation,+analysis subject,+entity,+account,+measurement4

+title,+profile,+var,+missing record,ts_profile_chain3

+score,historic,+historic snapshot,people,+snapshot6

mape,+history,+time history,mape,+disturbance5

f,+transpose,+weight,data,+root6

+access,+file,+text,+relational,+relational database5

+boat,+sail,wind,+race,gps1

Terms Documents

Topic > +simulation,+training,+training data,+response,+random

254 Data Quality for Analytics Using SAS Chapter 19: Influence of Data Correctness on Model Quality in Predictive Modeling 253 Chapter 19: Influence of Data Correctness on Model Quality in Predictive Modeling 19.1 Introduction 243 General 243 Non-visible data quality problem 244 Random and systematic bias 244 Biased values in the scoring data partition 244 19.2 Simulation Methodology and Data Preparation 245 ...

236 Data Quality for Analytics Using SAS Chapter 18: Influence of Data Completeness on Model Quality in Predictive Modeling 237 Chapter 18: Influence of Data Completeness on Model Quality in Predictive Modeling 18.1 Introduction 231 General 231 Random and systematic missing values 232 Missing values in the scoring data partition 232 18.2 Simulation Methodology and Data Preparation 233 Inserting random missing values 233...

230 Data Quality for Analytics Using SAS Chapter 17: Influence of Data Quantity and Data Availability on Model Quality in Predictive Modeling 229 Chapter 17: Influence of Data Quantity and Data Availability on Model Quality in Predictive Modeling 17.1 Introduction 219 General 219 Data quantity 220 Data availability 220 17.2 Influence of the Number of Observations 220 Detailed functional question 220 Data preparation 220 Simulation ...

216 Data Quality for Analytics Using SAS Chapter 16: Simulating the Consequences of Poor Data Quality for Predictive Modeling 217 Chapter 16: Simulating the Consequences of Poor Data Quality for Predictive Modeling 16.1 Introduction 206 Importance of predictive modeling 206 Scope and generalizability of simulations for predictive modeling 206 Overview of the functional questions of the simulations 206 16.2 Base for the Business ...

sas THE POWER TO KNOW.

Copyright © 2012, SAS Institute Inc. All rights reserved.

File Help

Books_Breakfast

Topics Run View

Topics

Topics

Number of Documents

All Topics (59)

+shop,+promotion,+label,+productgroup,+pg3

detection,+outlier,+node,outlier detection,jmp3

+simulation,+training,+training data,+response,+random4

+record,correctness,+systematic,+bias,+database4

+multiple observation,+analysis subject,+entity,+account,+measurement4

+title,+profile,+var,+missing record,ts_profile_chain3

+score,historic,+historic snapshot,people,+snapshot6

mape,+history,+time history,mape,+disturbance5

f,+transpose,+weight,data,+root6

+access,+file,+text,+relational,+relational database5

+boat,+sail,wind,+race,gps1

Terms Documents

Topic > +access,+file,+text,+relational,+relational database

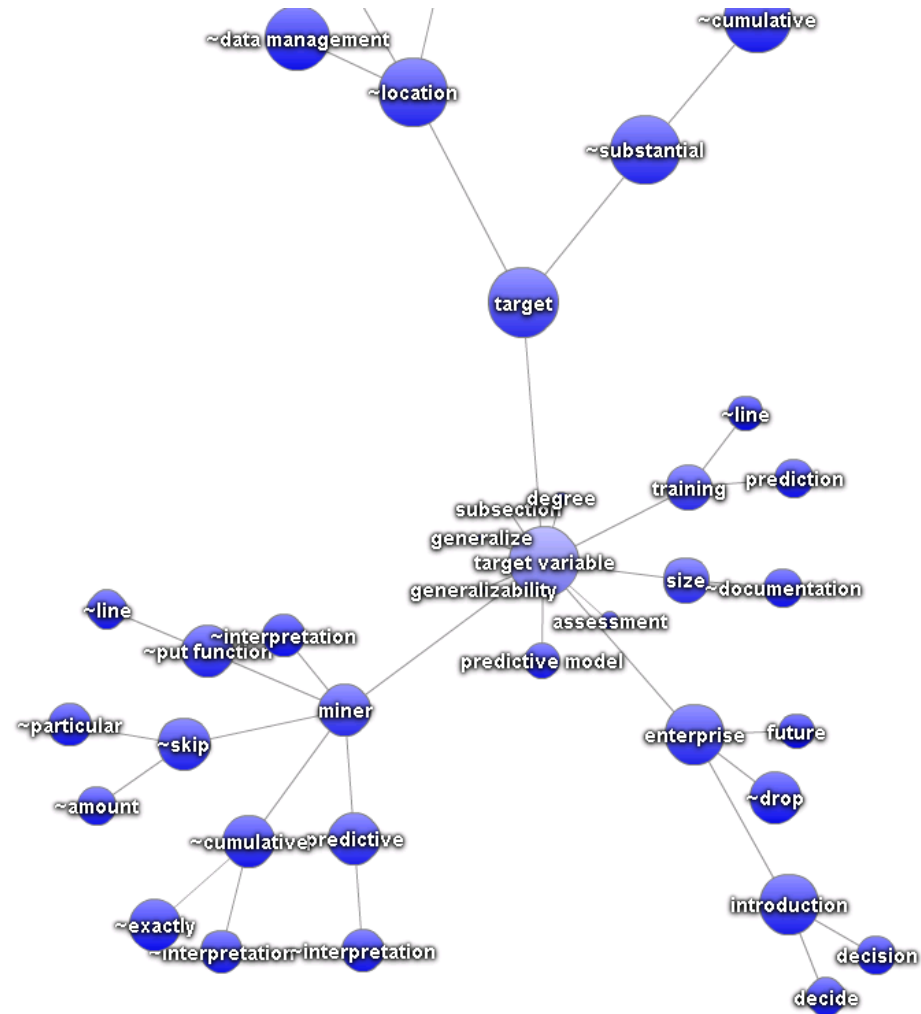
PAGE 104 [Data Preparation](#) for Analytics Using SAS Chapter 13: [Accessing Data](#) PAGE 103 Part 3 [Data Mart Coding](#) and Content Chapter 13 [Accessing Data](#) Transposing One- and Multiple-Rows-per-Subject [Data Structures](#) 115 Chapter 15 Transposing [Longitudinal Data](#) 131 Chapter 16 [Transformations of Categorical Variables](#) 161 Chapter 18 Multiple [Interval-Scaled Observations per Subject](#) 179 Chapter 19 [Multiple Categorical Variables](#) 185

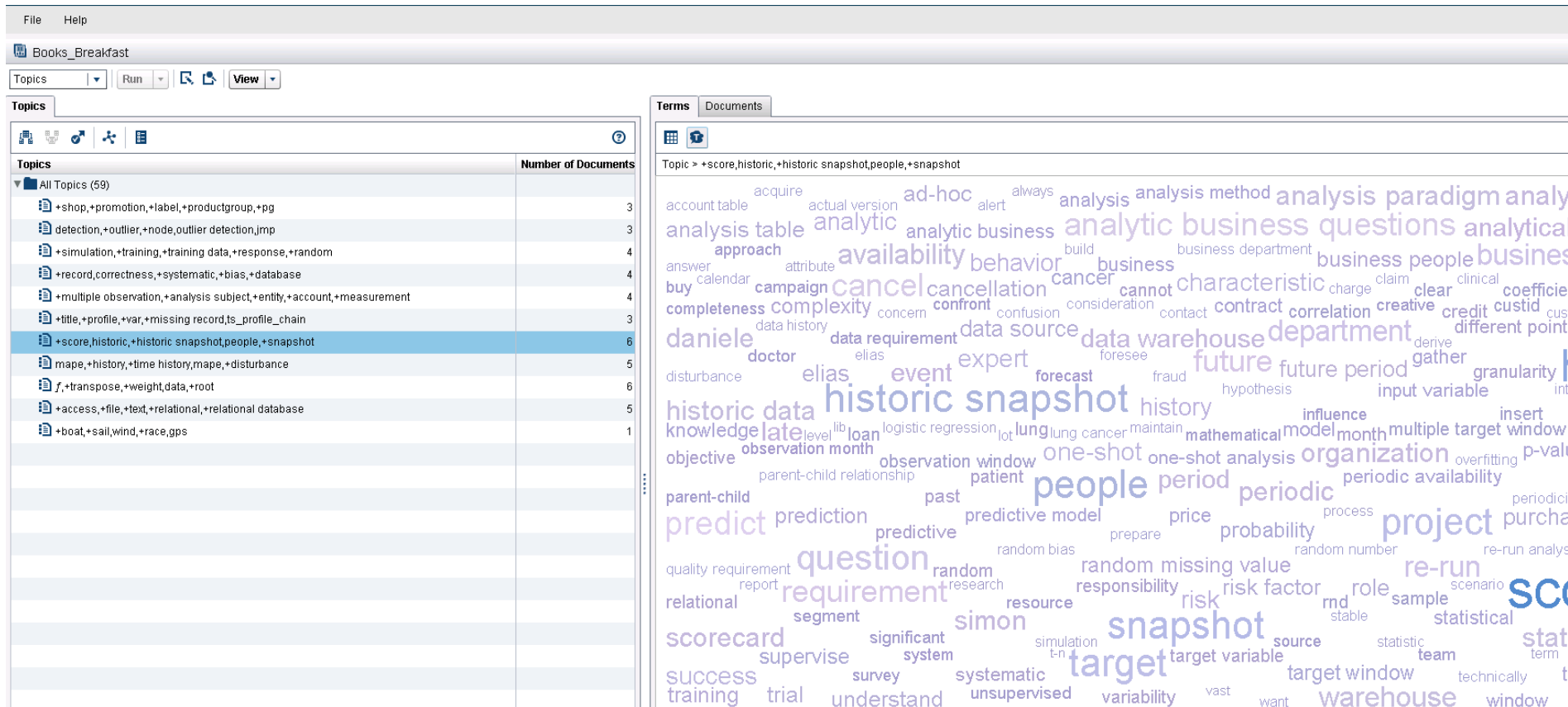
PAGE 38 [Data Preparation](#) for Analytics Using SAS Chapter 5: The [Origin of Data](#) PAGE 43 Part 2 [Data Structures](#) and [Data Modeling](#) Chapter 5 The [Models](#) 45 Chapter 7 [Analysis Subjects](#) and Multiple Observations 51 Chapter 8 The One-Row-per-Subject [Data Mart](#) 61 Chapter 9 The Multiple-Rows-per-Subject [Data Structures](#) for [Longitudinal Analysis](#) 77 Chapter 11 Considerations for [Data Marts](#) 89 Chapter 12 Considerations for Predictive [Modeling](#) 95 Introduction 97

PAGE 178 [Data Preparation](#) for Analytics Using SAS Chapter 17: [Transformations of Categorical Variables](#) PAGE 177 Chapter 17 [Transformations of Categorical Variables](#) Introduction 17.2 General Considerations for [Categorical Variables](#) 162 17.3 [Derived Variables](#) 164 17.4 Combining [Categories](#) 166 17.5 [Dummy Coding](#) 167 17.6 [Multidimensional Categorical Variables](#) 172 17.7 [Lookup Tables](#) and [External Data](#) 176 17.1 Introduction In this chapter we will deal with [transformations of categorical variables](#)

40 [Data Quality](#) for Analytics Using SAS Chapter 3: [Data Availability](#) 41 Chapter 3: [Data Availability](#) 3.1 Introduction 32 3.2 General Considerations 32 Re: [data availability](#) 32 Availability and usability 32 Effort to make [data](#) available 33 Dependence on the [operational process](#) 33 Availability and alignment in t of Historic [Data](#) 34 [Categorization](#) and examples of historic [data](#) 34 The [length](#) of the [history](#) 35 [Customer event histories](#) 35 [Operational systems](#) and [data quality](#)

PAGE 382 [Data Preparation](#) for Analytics Using SAS Appendix B: The Power of [SAS](#) for [Analytic Data Preparation](#) PAGE 381 Appendix B The Power of [SAS](#) for [Analytic Data Preparation](#) 369B.1 Motivation B.2 Overview 370 B.3 [Extracting Data](#) from [Source Systems](#) 371 B.4 Changing the [Data Mart Structure](#): Transposing 371 B.5 [Data Mart Structure](#): Transposing 371 B.6 [Data Mart Structure](#): Transposing 371 B.6 Selected Features of the [SAS Language](#) for [Data Management](#) 375 B.7 Benefits of the [SAS Macro Language](#) for [Data Management](#) 375





File Help

Books_Breakfast

Categories Run View

Categories

Categories	Document Frequency	Number of Documents
▼ All Categories		42
▶ +access,+file,+text,+relational,+relational database		5
▶ +boat,+sail,+wind,+race,gps		3
▶ +multiple observation,+analysis subject,+entity,+account,+...		4
▶ +record,+correctness,+systematic,+bias,+database		4
▶ +score,+historic,+historic snapshot,+people,+snapshot		6
▶ +shop,+promotion,+label,+productgroup,+pg		3
▶ +simulation,+training,+training data,+response,+random		4
▶ +title,+profile,+var,+missing record,ts_profile_chain		3
▶ +detection,+outlier,+node,+outlier detection,jmp		3
▶ f,+transpose,+weight,data,+root		7
▶ mape,+history,+time history,mape,+disturbance		5

Edit Rules Documents Test Rules

Category > f,+transpose,+weight,data,+root > True Positive

PAGE 390 Data Preparation for Analytics Using SAS Appendix C: Transposing with DATA Steps PAGE Transposing and Performance C.2 From a LONG Data Set to a WIDE Data Set 385 C.3 From a WIDE Dat Transactional Categories with a DATA Step 389 C.1 Transposing and Performance General We saw in C

PAGE 130 Data Preparation for Analytics Using SAS Chapter 14: Transposing One- and Multiple-Rows- Transposing One- and Multiple-Rows-per-Subject Data Structures 11514.1 Introduction 14.2 Transposing per- Subject Data Set 117 14.3 Transposing from a One-Row-per-Subject Data Set to a Multiple-Rows-p

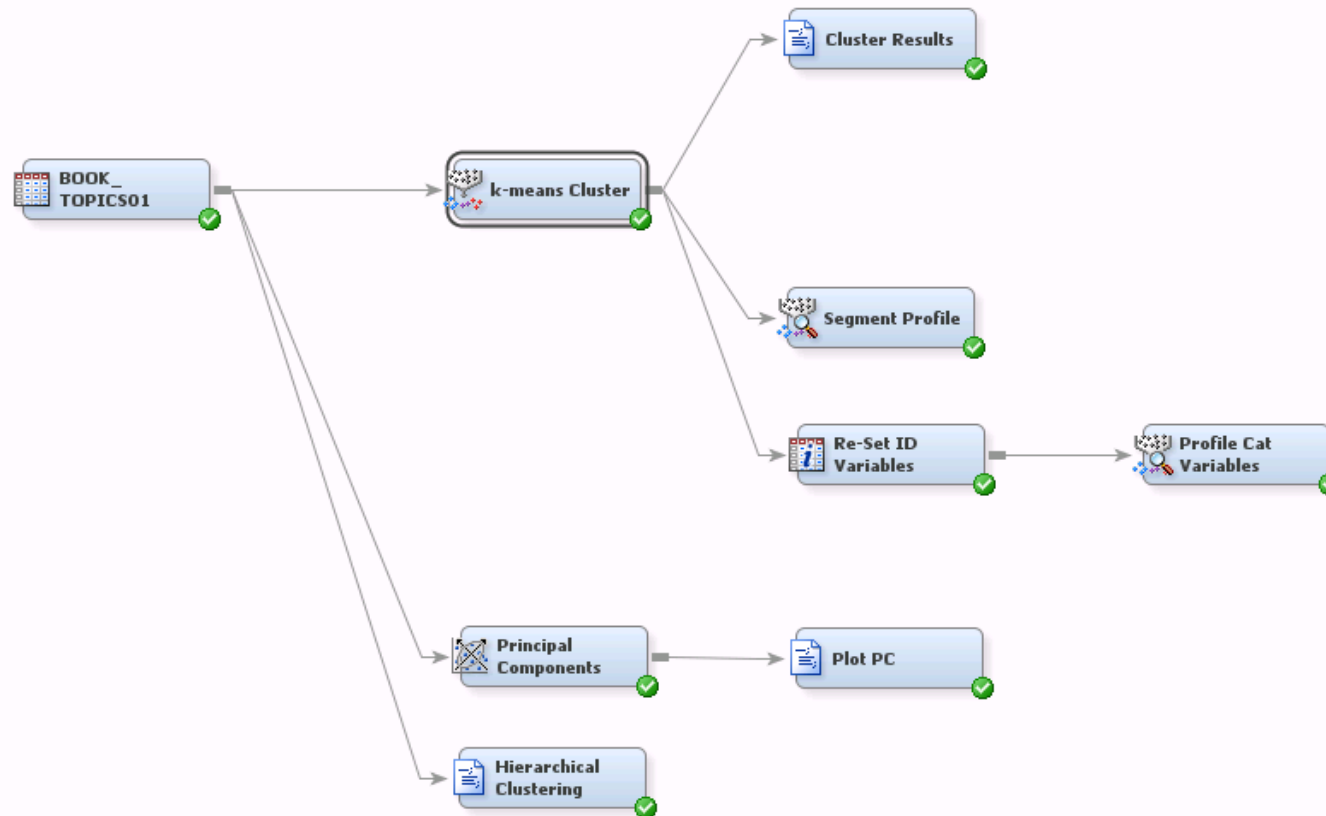
PAGE 146 Data Preparation for Analytics Using SAS Chapter 16: Transformations of Interval-Scaled Var Scaled Variables 13916.1 Introduction 16.2 Simple Derived Variables 140 16.3 Derived Relative Variables into Groups 149 16.6 Transformations of Distributions 153 16.7 Replacing Missing Values 159 16.8 Con

PAGE 200 Data Preparation for Analytics Using SAS Chapter 18: Multiple Interval-Scaled Observations per Observations per Subject 17918.1 Introduction 18.2 Static Aggregation 180 18.3 Correlation of Values 18 Standardization of Values 189 18.6 Course over Time: Derived Variables 194 18.1 Introduction If we have

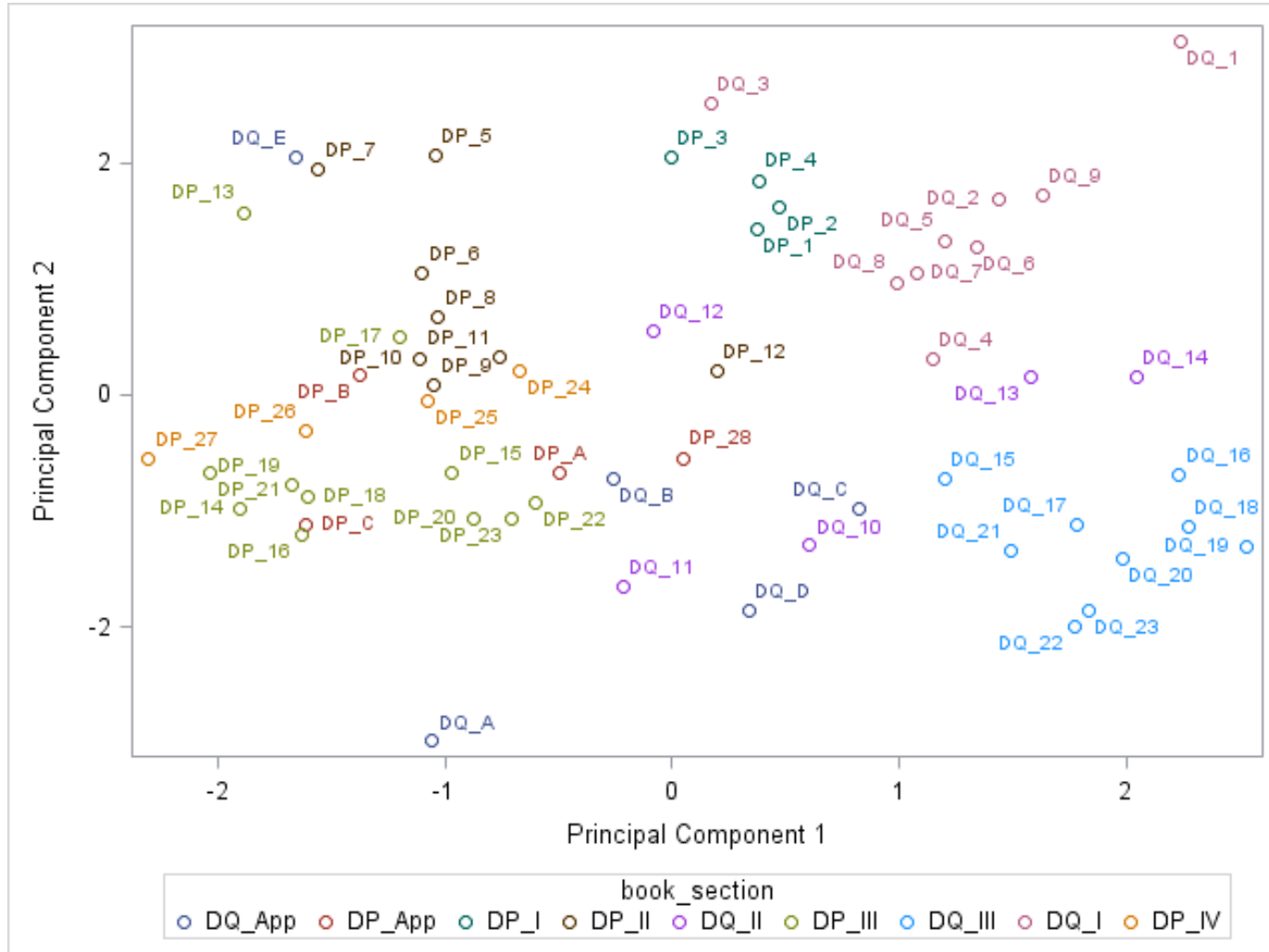
PAGE 214 Data Preparation for Analytics Using SAS Chapter 19: Multiple Categorical Observations per Observations per Subject 20119.1 Introduction 19.2 Absolute and relative frequencies of categories 202 1 19.4 Calculating Total and distinct counts of the categories 208 19.5 Using ODS to Create Different Perce

PAGE 226 Data Preparation for Analytics Using SAS Chapter 20: Coding for Predictive Modeling PAGE Introduction 20.2 Proportions or Means of the Target Variable 216 20.3 Interval Variables and Predictive M 232 20.1 Introduction 20.1.1 The Scope of This Chapter Many data mining analyses deal with predictive

	topic_raw1	topic_raw2	topic_raw3	topic_raw4	topic_raw5	topic_raw6	topic_raw7	topic_raw8	topic_raw9	topic_raw10	topic_raw11	_DOCUMENT_	TEXT	URI	N
1	0.047	0.007	0.025	0.002	0.018	0.755	-0.001	0.038	0.092	0.022	0.001	1	298 Data Quali...	file://D:\Gerha...	AppA_n
2	0.010	0.064	0.043	-0.005	0.014	0.065	0.025	0.015	0.056	0.012	0.014	2	304 Data Quali...	file://D:\Gerha...	AppB_n
3	0.011	0.050	0.276	-0.009	0.021	0.049	0.018	0.047	0.015	0.035	0.032	3	306 Data Quali...	file://D:\Gerha...	AppC_n
4	0.054	0.026	0.014	-0.053	0.018	0.097	0.023	0.363	0.048	0.015	0.010	4	318 Data Quali...	file://D:\Gerha...	AppD_n
5	0.069	0.112	0.024	0.000	0.075	0.085	0.047	0.026	0.091	-0.002	0.002	5	PAGE 368 D...	file://D:\Gerha...	Appendi
6	0.048	0.039	0.049	0.005	0.086	0.069	0.021	0.034	0.149	0.248	0.043	6	PAGE 382 D...	file://D:\Gerha...	Appendi
7	-0.002	-0.043	0.040	-0.032	0.091	0.086	0.003	0.023	0.262	0.030	0.093	7	PAGE 390 D...	file://D:\Gerha...	Appendi
8	0.031	0.016	0.028	0.061	0.736	0.018	0.053	0.019	0.028	0.056	0.032	8	320 Data Quali...	file://D:\Gerha...	AppE_n
9	0.032	0.026	0.037	0.071	0.098	0.028	0.283	0.033	0.013	0.039	0.031	9	PAGE 2 Dat...	file://D:\Gerha...	chap1.n
10	0.165	0.029	0.009	0.022	0.190	0.032	0.058	0.031	0.040	0.058	0.018	10	PAGE 88 Da...	file://D:\Gerha...	chap10.
11	0.022	0.208	0.122	0.061	0.049	0.408	0.050	0.037	0.053	0.043	0.012	11	140 Data Quali...	file://D:\Gerha...	chap10.
12	0.041	0.053	0.045	0.023	0.155	0.022	0.076	0.008	0.083	0.082	0.002	12	PAGE 94 Da...	file://D:\Gerha...	chap11.
13	0.110	0.090	-0.004	0.117	0.039	0.473	0.023	0.097	0.082	0.003	0.009	13	156 Data Quali...	file://D:\Gerha...	chap11.
14	0.038	0.023	0.121	0.025	0.055	0.015	0.159	0.021	0.033	0.024	-0.009	14	PAGE 100 D...	file://D:\Gerha...	chap12.
15	0.028	0.035	-0.030	0.173	0.134	0.064	-0.028	0.007	0.003	0.077	0.033	15	164 Data Quali...	file://D:\Gerha...	chap12.
16	0.058	0.024	0.026	0.020	0.103	0.029	0.008	0.015	0.061	0.492	0.016	16	PAGE 104 D...	file://D:\Gerha...	chap13.
17	0.007	0.377	0.092	0.110	0.075	0.064	0.082	0.072	0.069	0.034	0.048	17	178 Data Quali...	file://D:\Gerha...	chap13.
18	0.007	0.004	0.020	-0.014	0.150	0.071	0.013	0.028	0.342	0.036	0.091	18	PAGE 130 D...	file://D:\Gerha...	chap14.
19	0.033	0.483	0.124	0.061	0.049	0.115	0.074	0.072	0.035	0.072	0.150	19	198 Data Quali...	file://D:\Gerha...	chap14.
20	0.066	0.012	0.006	-0.015	0.064	0.054	0.004	0.030	0.070	0.030	0.013	20	PAGE 138 D...	file://D:\Gerha...	chap15.
21	0.011	0.038	0.223	0.089	0.020	0.037	0.014	0.120	0.025	0.021	0.045	21	Part III: Conseq...	file://D:\Gerha...	chap15.
22	0.076	0.042	0.069	0.127	0.044	0.106	0.053	0.023	0.482	0.043	0.030	22	PAGE 146 D...	file://D:\Gerha...	chap16.
23	0.024	0.091	0.430	0.053	0.048	0.039	0.104	0.125	0.001	0.027	0.032	23	216 Data Quali...	file://D:\Gerha...	chap16.
24	0.077	0.136	0.031	0.010	0.070	0.035	0.036	0.016	0.107	0.284	-0.027	24	PAGE 178 D...	file://D:\Gerha...	chap17.
25	0.027	0.041	0.458	0.037	0.025	0.022	0.049	0.068	0.026	0.009	0.020	25	230 Data Quali...	file://D:\Gerha...	chap17.
26	0.108	0.063	0.027	0.013	0.118	0.055	0.058	0.056	0.304	0.042	0.006	26	PAGE 200 D...	file://D:\Gerha...	chap18.
27	0.012	0.043	0.500	0.114	0.028	0.051	0.042	0.094	0.033	0.011	-0.004	27	236 Data Quali...	file://D:\Gerha...	chap18.
28	0.043	0.039	0.053	-0.009	0.086	0.017	0.018	-0.004	0.265	0.087	-0.017	28	PAGE 214 D...	file://D:\Gerha...	chap19.
29	0.005	0.010	0.538	0.152	0.028	0.031	0.021	0.144	0.062	0.005	0.009	29	254 Data Quali...	file://D:\Gerha...	chap19.
30	0.062	0.118	0.069	0.176	0.075	0.037	0.103	0.074	0.027	0.090	0.889	30	Part I: Data Qu...	file://D:\Gerha...	chap1_r
31	0.024	0.053	0.071	0.040	0.134	0.031	0.368	0.033	0.035	0.036	0.004	31	PAGE 20 Da...	file://D:\Gerha...	chap2.n
32	0.047	0.026	0.180	-0.007	0.019	0.086	0.069	-0.020	0.234	0.068	-0.009	32	PAGE 226 D...	file://D:\Gerha...	chap20.
33	0.060	0.064	0.188	0.037	0.036	0.052	0.069	0.445	0.017	0.028	0.044	33	262 Data Quali...	file://D:\Gerha...	chap20.
34	0.414	0.040	-0.001	0.052	0.089	0.074	0.037	0.052	0.191	0.046	0.020	34	PAGE 254 D...	file://D:\Gerha...	chap21.
35	0.023	0.048	0.081	-0.006	0.035	0.017	0.064	0.450	0.006	0.022	0.029	35	272 Data Quali...	file://D:\Gerha...	chap21.
36	0.043	0.023	0.128	-0.030	0.070	0.094	0.040	0.013	0.109	0.030	0.016	36	PAGE 256 D...	file://D:\Gerha...	chap22.
37	0.033	0.049	0.210	0.090	0.021	0.080	-0.029	0.388	0.050	0.005	-0.001	37	274 Data Quali...	file://D:\Gerha...	chap22.
38	0.061	0.070	0.129	-0.125	0.045	0.204	0.144	0.021	0.122	0.062	0.032	38	PAGE 292 D...	file://D:\Gerha...	chap23.
39	0.019	0.014	0.163	0.061	0.017	0.042	0.012	0.469	0.034	0.008	0.003	39	288 Data Quali...	file://D:\Gerha...	chap23.



1	Missing Values	10	11	A																	
2	Erzeugen des Analytic-Marts	10	11	14	15	16	18	19	20	21	22	23	24	25	26	27	28	9	A	C	B
3	Data Origin und Data Management	5	13	17	B																
4	DQ Case Studies	1																			
5	Fachliche Konzepte	1	2	3	4	12	2	3	4	5	6	7	8	9	12						
6	DQ mit Analytik und SAS	13	14																		
7	Data Quality Simulationen	15	16	17	18	19	20	21	22	23	C	D									
8	Analytic Data Mart Structures	6	7	8	E																



Sammlung von 17.205 Kunden-Feedback
Nachrichten auf einem Web-Portal



- Wie denken die Kunden über uns?
- Automatisches Routing zur relevanten Stelle?
- Monitoring über die Zeit



ANWENDUNGS- BEISPIEL 1

KUNDEN-FEEDBACK AUF EINEM WEB-PORTAL (FORTS.)

Direkter Zugriff auf diese Nachrichten (Web Crawler, Datenintegration)

Automatische Erkennung von 23 "Themen" (Topics) in diesem Nachrichten-Pool mit Hilfe von Text Mining und Natural Language Processing

Manuelle Verfeinerung/Zusammenfassung zu 21 Themen

Automatisches Ableiten der entsprechenden Kategorie pro Topic auf Basis linguistischer Regeln

Testen der Treffsicherheit dieser 21 Kategorien

Regelmäßiges, automatisches Anwenden dieser linguistischen Regeln zur Zuweisung (Routing) von Kunden-Feedbacks zu den Kategorien

Datenintegration

Analytik

Manuelle Interaktion

Analytik

Validierung, Tuning

Operatives Scoring

62.837 Vertragstexte in PDF-Form auf einem File-Server

- In welche Gruppen können die Vertragstexte gruppiert werden?
- Wie deckt sich diese Gruppierung mit anderen Vertragskategorien?
- Können bestimmte Klauseln vereinheitlicht werden?



§§



Einlesen der Daten direkt aus den PDF-Dokumenten

Automatische Erkennung von 61 Themen in den Vertragstexten.

Definition einer hierarchischen Themenstruktur mit 14 Haupt-Themen, 61 Themen und 137 Sub-Themen

Ableiten von Kategorien aus diesen Themen

Profiling der Eigenschaften der einzelnen Kategorien und Themen

Identifikation der häufigsten Inhalte der Vertragstexte und der Verteilung über die Dokumente

Datenintegration

Analytik

Manuelle Interaktion

Analytik

Validierung, Tuning

Erkenntnisgewinn

SAS TEXT ANALYTICS

CONTEXTUAL ANALYSIS

- Information Retrieval
- Automatische Themen Erkennung
- Content Kategorisierung
- Extraktion von Entitäten
- Sentiment Analyse



TEXT MINING

Erkennen von wertvollen Mustern, Themen und Einblicken aus Text-Daten

INTEGRIERTE ANALYTIK

Integration strukturierte und unstrukturierten Daten für:

- Forecasting
- Optimierung
- Predictive Modeling
- Netzwerk Analyse

- Gerne im direkten Gespräch mit uns persönlich.
- [SAS-Webinar zu SAS Contextual Analysis Demo](#)
- [Text Analytics auf der Deutschen SAS-Website](#)
- [Fact Sheet](#)
- [Beitrag](#) im Mehr-Wissen Blog vom 30. Oktober 2015

