

Index

A

- abstracts, patents, 113
- acquisition of information, 29
- Acquisition phase
 - VoC information, 73
 - VoC Mining, 76-77
- Addresses, email, 25
- analysis
 - Future Mining, 137
 - Innovation Mining, 119-120
 - interactions, 24-25
 - ongoing, 138
 - problem tickets, 4
 - sentiment, 107
 - taxonomies, 79-80
- analytical methodology, 9
 - Analyze, 15-16
 - Explore, 10-11
 - Understand, 11-15
- Analyze phase, 118
- Analyze process, 15
- annotations, 73, 77-78
- applications
 - customer interactions, 16, 32
 - unstructured information mining, 16-18
 - Unstructured Information Modeler, 172
 - VoC Mining, 73
- architecture, Unstructured Information Modeler, 172
- Artificial Intelligence principles, 167

- assignee(s), patents, 113
- authoritativeness, source, 135
- available information, surveys of, 28

B

- blogs, 75
 - Future Mining with, 143
- body of patent documents, 113
- branding, 149
- building feature space, 33
- business drivers, 94
 - communication, 93-94
 - for VoC Mining, 73
 - Future Mining, 134
 - Innovation Mining, 112
 - for interactional analysis, 24-25
- business information, 25
- business objectives
 - analytical methodology of, 9
 - Analyze, 15-16*
 - Explore, 10-11*
 - Understand, 11-15*
 - Future Mining, 135-136
 - Innovation Mining, 115
 - Jams, 95
 - overview of, 27
 - understanding, 74
- buttons
 - Class Tree view, 207
 - Class view, 208
 - Visualization panel, 206-207

C

- call center logs, 28
- capabilities
 - demonstration of, 28, 75
 - Future Mining, 136
- capturing business, 9
- categories
 - clusters, 37
 - cohesion, 99
 - creating, 52-57
 - customer interactions, 26
 - deleting, 59
 - distinctness, 57-58
 - Miscellaneous, 100
 - relationships, 57-61
 - renaming, 122
 - taxonomies, 29
- centroids, 35
 - visualization, 56
- characteristics
 - of Future Mining data sets, 134
 - of Innovation Mining, 112-114
 - interactional information, 25
 - structured information, 25-26
 - unstructured information, 26
- chat transcripts, 28
- claims, patents, 113
- Class Table view, 187-189
- Class Tree view, 185-187, 207
- Class view, 189-190, 208
- classes
 - deleting, 183
 - Main Class (patents), 114
 - merging, 122
 - naming, 182
 - optimizing, 185
 - saving, 185
 - secondary, 60, 123
 - similar, 181
 - sizing, 188
 - splitting, 182-185
 - subclass (patents), 114
 - subclasses, 184
- classification, 3, 16, 31, 34
 - manual, 9
 - viewing, 189-190
- classifiers, 81
- cleansing customer survey responses, 76
- clustering
 - documents, 13
 - Intuitive, 39-40
 - naming, 37
 - taxonomies, 29
 - text, 33-37
- co-occurrence tables, 42
- cohesion
 - categories, 99
 - classes, 188
 - of terms, 39
 - variation, 14
- collection, data, 76, 117
- commands, run, 175-178
- communication, 94-95
 - customers, 17
 - Jams
 - data feeds*, 95
 - future of*, 107-108
 - integration*, 98-106
 - reports*, 97
 - searching for trends*, 96
 - stages of*, 95
 - taxonomy*, 96
 - understanding business objectives*, 95
 - styles, 164-165
- computer security, 149
- content sources, identifying, 152
- converting text to feature space, 34
- corporate communication, 94-95
 - data feeds, 95
 - future of, 107-108
 - integration, 98-106
 - reports, 97
 - searching for trends, 96
 - stages of, 95
 - taxonomy, 96
 - understanding business objectives, 95
- correlations, 15
 - searching for, 30
 - between structured data and unstructured
 - categories, 41-45
- Cosine distance, 35
- cover naming, 37
- criteria, searching, 22
- customers
 - emails, 28
 - interactions, 16, 21-23
 - analyzing business drivers*, 24-25
 - characteristics of*, 25-26
 - comparing to transactions*, 23-24
 - establishing continuous processes*, 31-32
 - examples of*, 32-64
 - lessons from*, 26
 - processes*, 27-31

issue descriptions, 26
 satisfaction, 25
 surveys, 75
 voice of, 17

D

data cleansing
 Future Mining, 137
 Innovation Mining, 117
 data collection, 76
 data feeds, Jams, 95
 data formats, Unstructured Information Modeler, 172
 data mining, 2
 data sets, 134
 database queries, 10
 dates, customer interactions, 25
 DateString, 78
 deleting
 categories, 59
 classes, 183
 demographic information, 25
 demonstration of capabilities, 28, 75
 Dictionary Tool, 49, 178-181
 discussion forums, 75
 distance, 35
 distinctness, 57-58
 classes, 189
 similarities between categories, 14
 documents
 centroids, 35
 clustering, 13
 Innovation Mining, 112-114
 Analyze phase, 118
 data cleansing, 117
 demonstration of capabilities, 116
 Explore phase, 117
 finding relevant patents, 120-127
 ongoing analysis of, 119-120
 presenting results to stakeholders, 118-119
 process of, 115
 researching new products, 127
 survey of available data, 116
 Understand phase, 118
 understanding business objectives, 115
 patents, 113-114
 searching, 10
 trends, 140

domains
 analytical methodology of, 9
 Analyze, 15
 Explore, 10-11
 Understand, 11-14
 business, 9

E

Edit menu, Unstructured Information Modeler, 196, 199
 editing
 Dictionary Tool, 178-181
 feature space, 49-51
 taxonomies, 13-14, 29, 52-61, 102, 121-123
 effectiveness, 32
 evaluation of, 164-165
 VoC Mining, 82
 email
 addresses, 25
 customers, 28
 employees, voice, 17
 enterprise ecosystem, 6
 ETL (extracting, transforming, loading), 29-31
 Future Mining, 137
 Innovation Mining, 117
 Euclidean distance, 35
 evaluation
 effectiveness, 164-165
 metrics, 35
 Execute menu, Unstructured Information Modeler, 200-205
 expanding VoC Mining, 82
 expansion, 11
 Explore phase, Innovation Mining, 117
 Explore process, 10-11
 extracting data, 29

F

features
 selection, 12
 space
 creating, 33
 editing, 49-51
 text
 converting to, 34
 creating from, 45-47
 fields, queries, 10

File menu, Unstructured Information Modeler, 192-195

filed date, patents, 114

files

- Dictionary Tool, 180
- includeWords.txt, 180
- stopWords.txt, 180

final taxonomies, applying, 62, 64

financial planning, 148

fixed threshold of infrequency, selecting, 47

forecasting, 156

formatting

- categories, 52-57
- editing relationships, 57-61
- Unstructured Information Modeler, 172

forums, 94-95

- data feeds, 95
- future of, 107-108
- integration, 98-106
- reports, 97
- searching for trends, 96
- stages of, 95
- taxonomy, 96
- understanding business objectives, 95

frequency, term statistics, 56

Future Mining

- business drivers, 134
- data sets, 134
- examples of, 138-160
- ongoing analysis, 138
- stages of, 135
 - analysis*, 137
 - data cleansing*, 137
 - demonstration of capabilities*, 136
 - publishing results*, 138
 - survey of available data*, 136-137
 - taxonomies*, 137
 - understanding business objectives*, 135-136
- unstructured data, 133

G-H

geography spread, predicting the future, 142

global information space, 75-76

grant date, patents, 114

granularity, partitioning, 12

groups, 13. *See also* clustering

growth, 93-94

- of services industry, 7-8

Habitat Jam, 105-106

Help menu, Unstructured Information Modeler, 205

historical patterns, 150-157

I

IBM Global Services, 4

ID (identification), customer interactions, 25

includeWords.text file, 180

information acquisition, 29

information sources

- analytical methodology of, 9
- Analyze, 15-16
- Explore, 10-11
- Understand, 11-15

initial taxonomies

- creating, 29, 51-52
- VoC Mining
 - analysis*, 79-80
 - creating*, 78-79
 - refining*, 79

innovation, improving, 17

Innovation Mining

- business drivers, 112
- characteristics of, 112-114
- finding relevant patents, 120-127
- ongoing analysis of, 119-120
- process of, 115
 - Analyze phase*, 118
 - data cleansing*, 117
 - demonstration of capabilities*, 116
 - Explore phase*, 117
 - presenting results to stakeholders*, 118-119
 - survey of available data*, 116
 - Understand phase*, 118
 - understanding business objectives*, 115
- researching new products, 127
- value of, 114

installation

- input data format, 173-174
- Unstructured Information Modeler toolkit, 172-173

instant messaging transcripts, 28

integration, 98

- Habitat Jam, 105-106
- Values Jam, 98-104

interactions, 8

- customers, 16, 21-23
 - analyzing business drivers*, 24-25
 - characteristics of*, 25-26

comparing to transactions, 23-24
establishing continuous processes, 31-32
examples of, 32-64
lessons from, 26
processes, 27-31
 manual classification, 9
 intuitive clustering, 13, 39-40
 inventor(s), patents, 113

J-K

Jams, 94-95
 future of, 107-108
 integration, 98
 Habitat Jam, 105-106
 Values Jam, 98-104
 sentiment analysis, 107
 stages of, 95
 data feeds, 95
 reports, 97
 searching for trends, 96
 taxonomy generation, 96
 understanding business objectives, 95
 themes, 102
Java, installing, 173
k variable, 36
k-means variable, 36
key criteria, searching for, 22
keywords
 matching, 189
 searches, 183
Klokman, Kees, 164
Knowledge Engineering, 167
knowledge repositories, taxonomies as, 166-167

L-M

Landon, Alf, 72
levels of satisfaction, 85
LexisNexis, 76
lifecycles of interactions, 8
Literary Digest, 72
loading data, 29
location information, 26
loss of information, 148
Main Class, patents, 114
 manual classification, 9
mapping worry space, 145-150
market research, 72

Mars company, researching new products,
 127-131
matching keywords, 189
measurement of effectiveness, 32, 82
merging classes, 122, 181
message boards, 75
methodologies, 5
 analytical, 9
 Analyze, 15-16
 Explore, 10-11
 Understand, 11-15
Metrics, evaluation, 35
Miscellaneous category, 100
models
 applying, 155
 building, 153
monitoring
 patents, 157-160
 patterns, 81
 the future, 157
 VoC Mining, 81
multiple taxonomies, 30

N-O

naming
 classes, 182
 clusters, 37
navigating, 192
 Class Tree view, 207
 Class view, 208
 Edit menu, 196, 199
 Execute menu, 200-205
 File menu, 192-195
 Help menu, 205
 Subclass menu, 205
 Visualization panel, 206-207
newspapers, 75
objectives
 business
 overview of, 27
 VoC Mining, 74
 of capturing business, 9
**operating systems, Unstructured Information
 Modeler**, 172
operations, set, 11
opinion polling, 72
optimization
 classes, 185
 innovation, 17
other process information, 26

P

partitioning, 12

patents, 112

- documents, 113-114
- Innovation Mining, 120-127
- monitoring, 157-160

patterns

- alerts, 81
- historical, 150-157

periodicals, 75

phases of analytical methodology, 9

- Analyze, 15-16
- Explore, 10-11
- Understand, 11-15

phone numbers, 25

PlanetString, 78

polling, 72

preparation of information, 29

Preparation phase

- VoC information, 73
- VoC Mining, 76-77

problem tickets, 4

processes

- classification, 31
- ETL, 29, 31, 117
- Future Mining, 135
 - analysis*, 137
 - data cleansing*, 137
 - demonstration of capabilities*, 136
 - examples of*, 138-160
 - ongoing analysis*, 138
 - publishing results*, 138
 - survey of available data*, 136-137
 - taxonomies*, 137
 - understanding business objectives*, 135-136

Innovation Mining, 115

- Analyze phase*, 118
- data cleansing*, 117
- demonstration of capabilities*, 116
- Explore phase*, 117
- finding relevant patents*, 120-127
- ongoing analysis of*, 119-120
- presenting results to stakeholders*, 118-119
- researching new products*, 127
- survey of available data*, 116
- Understand phase*, 118
- understanding business objectives*, 115

reporting, 31

Unstructured Information Modeler toolkit, 175

- Dictionary Tool*, 178-181
- merging similar classes*, 181

- naming classes*, 182
- optimizing classes*, 185
- run command*, 175-178
- saving classes*, 185
- splitting classes*, 182-185

VoC Mining, 80-81

Proctor, Larry, 138, 150

products

- branding, 149
- information, 26
- Innovation Mining, 114
- researching new, 127-131

proposals, communication styles, 164

publish date, patents, 114

publishing results, Future Mining, 138

Q-R

queries, 10

question, term category, 99-100

recursion, 11

references, patents, 113

regular expressions, 77

relationships, 8, 57-61

renaming categories, 122

reports, 31

- Jams, 97

repositories, taxonomies as knowledge, 166-167

research

- articles, 112
- VoC, 72

Research group, 4

resolution descriptions, 26

results

- Future Mining, 138
- stakeholders, 118-119
- surveys, 106
- taxonomies, 31-32
- VoC Mining, 80

retro models, 153-155

reviews

- results, 80
- stakeholders, 31

roadmaps, monitoring the future, 157

Roosevelt, Franklin Delano, 72

RSS feeds, 75

run command, 175-178

S

satisfaction
 customer, 25
 levels, 85
saving classes, 185
scalability, 166
 Unstructured Information Modeler, 172
scatter plot visualizations, 60-61, 190-192
searching
 for correlations, 30
 documents, 10
 key criteria for, 22
 keywords, 183
 for trends, 30, 96
secondary categories, 59
secondary classes, 60, 123
security, computer, 149
selection, feature, 12
sentiment
 analysis, 107
 satisfaction, 85
services industry, growth of, 7-8
set operations, 11
similar classes, merging, 181
similar technology, identifying, 151
sizing classes, 188
snippets, 73, 78, 88
sorting categories by cohesion, 99
source authoritativeness, 135
specifications, 172
splitting classes, 182-185
Spohrer, Jim, 165
stages
 of Future Mining, 135
 analysis, 137
 data cleansing, 137
 demonstration of capabilities, 136
 publishing results, 138
 survey of available data, 136-137
 taxonomies, 137
 understanding business objectives, 135-136
 of Jams, 95
 data feeds, 95
 reports, 97
 searching for trends, 96
 taxonomies, 96
 understanding business objectives, 95
 of VoC Mining, 74
 demonstration of capabilities, 75
 surveys, 75-76
 understanding business objectives, 74

stakeholders
 results, 118-119
 reviews, 31
 VoC Mining, 80
starting Unstructured Information Modeler toolkit, 174
statistics, term frequency, 56
stemming, 49
stock phrases, 49-51
 Dictionary Tool, 181
stopwords, 49
stopWords.txt file, 180
streams, time events in, 88-90
Strong, Ray, 157
structure of transactions, 23
structured data, 41-45
structured information, 25-26
styles, communication, 164-165
Subclass menu, Unstructured Information Modeler, 205
subclasses, 114, 184
suggestion line databases, 28
summary reports, Jams, 97
surveys
 of available data
 Future Mining, 136-137
 Innovation Mining, 116
 of available information, 28
 Jams, 106
 VoC Mining, 75
 customers, 75
 global information space, 75-76
synonyms, 46, 181

T

tables
 Class Table view, 187-189
 co-occurrence, 42
taxonomies
 categories, 29
 creating, 29, 33-37, 121
 development of, 5
 editing, 13-14, 52-61, 102, 121-123
 finishing, 62, 64
 Future Mining, 137
 initial, 51-52
 Jams, 96
 as knowledge repositories, 166-167
 multiple, 30
 results, 31-32
 snippets, 88

- VoC Mining
 - analysis*, 79-80
 - creating*, 78-79
 - refining*, 79
- technology landscapes, 157
- term frequency statistics, 56
- text
 - annotations, 77-78
 - clustering, 33-36
 - feature space
 - converting to*, 34
 - creating from*, 45-47
 - editing*, 49-51
 - naming, 37
 - snippetization, 73
 - snippets, 78, 88
 - visualization, 108
- themes, 94
 - Jams, 102
- third-party vendors, 75
- time
 - customer interactions, 25
 - events in streams, 88-90
- timestamps, customer interactions, 25
- titles, patents, 113
- tools, Dictionary Tool, 49, 178-181
- transactions, 8
 - interactions, 23-24
- transcripts, 28
- transformation process, 18-19
- transforming data, 29
- trends, 15
 - documents, 140
 - Jams, 96
 - searching for, 30

U

- Understand phase, Innovation Mining, 118
- Understand process, 11-14
- Uniform Resource Locator (URL), 76
- United States Patent & Trademark Office (USPTO), 112
- unstructured categories, 41-45
- unstructured data, 133
- unstructured information, 5, 8, 26
 - applications, 16-18
 - customer interaction, 16
 - as a discipline, 167-168
 - enterprise ecosystem, 6
 - growth of services industry, 7-8

- Unstructured Information Modeler toolkit, 171-172
 - input data format, 173-174
 - installation, 172-173
 - navigating, 192
 - Class Tree view*, 207
 - Class view*, 208
 - Edit menu*, 196, 199
 - Execute menu*, 200-205
 - File menu*, 192-195
 - Help menu*, 205
 - Subclass menu*, 205
 - Visualization panel*, 206-207
 - process, 175
 - Dictionary Tool*, 178-181
 - merging similar classes*, 181
 - naming classes*, 182
 - optimizing classes*, 185
 - run command*, 175-178
 - saving classes*, 185
 - splitting classes*, 182-185
 - specifications, 172
 - starting, 174
 - views, 185
 - Class*, 189-190
 - Class Table*, 187-189
 - Class Tree*, 185-187
 - Visualization*, 190-192
- unstructured interactions, 24
- URL (Uniform Resource Locator), 76
- USPTO (United States Patent & Trademark Office), 112

V

- Values Jam, 98-104
- variables
 - k, 36
 - k-means, 36
- variation cohesion, 14
- vectors
 - documents, 35
 - space model, 13
- views
 - Class, 208
 - Class Tree, 207
 - Unstructured Information Modeler toolkit, 185
 - Class*, 189-190
 - Class Table*, 187-189
 - Class Tree*, 185-187
 - Visualization*, 190-192

- visualization, 14
 - centroid, 56
 - scatter plots, 60
 - text, 108
- Visualization panel, Unstructured Information Modeler, 206-207
- Visualization view, 190-192
- VoC (Voice of the Customer), 71
 - acquisition and preparation processes, 76-78
 - business drivers for, 73
 - characteristics of information, 73
 - effectiveness measurements, 82
 - examples of, 82-90
 - expanding, 82
 - market research, 72
 - patterns, 81
 - processes, 80-81
 - stages of, 74
 - demonstration of capabilities*, 75
 - surveys*, 75-76
 - understanding business objectives*, 74
 - stakeholders, 80
 - taxonomies
 - analysis*, 79-80
 - creating*, 78-79
 - refining*, 79
 - value of, 74
- voice
 - of customers, 17
 - of employees, 17
 - transcripts from customer calls, 28
- Voice of the Customer. *See* VoC

W-Z

- Watson Jr., Thomas J., 93
- web crawlers, 75
- websites, 75
- World Urban Forum, 105
- worry space, mapping, 145-150