

IBM Software

# Business Analytics **Forum** 2010

See The Future Of Decision Making

**November 3-5, 2010**

Gold Coast – QLD – Australia



## Discover the Valuable Insights Hidden in Text Data

Eric Martin  
IBM





# Welcome

## Why this is Important to Know...

Text analytics helps uncover insights from people's words and use that insight to make better predictions about future outcomes.

SPSS makes it easier for customers to leverage this rich data source as part of their predictive analytics efforts.

## Speaking to you today...

**Eric Martin**

WW Predictive Analytics  
Solutions Leader  
IBM





# Information Explosion

**In volume.**  
**15 petabytes**

Amount of new information being generated every day, 8x more than the information in all U.S. libraries.

**200 billion**

More than 200 billion emails are sent every day.

**988 exabytes**

Amount of digital information that will exist in 2010—equivalent to a stack of books from the sun to Pluto and back.

**In variety.**  
**80%**

New data growth that is unstructured content, generated largely by email, with increasing contribution by documents, images, video and audio.

**30 billion**

By 2010, up to 30 billion RFID tags will be produced globally, embedded into products, passports, buildings—even animals.

**\$5.7 million**

For every 1,000 knowledge workers it employs, a company loses \$5.7 million annually in time wasted reformatting information between applications.





# Why Are We Here?

An organization's ability to learn, and translate that learning into action rapidly, is the ultimate competitive advantage.

-Jack Welch



Because people communicate with words, not numbers, it has become critical for information based businesses to be able to mine text for its meaning and to sort, analyze, and understand it in the same way that data has been tamed. In fact, the two basic types of information complement each other, with data supplying the "what" and text supplying the "why."



TextAnalytics: Software's Missing Piece?





# The Value in Unstructured Data

- “80-90% of information on the net and corporate networks is unstructured” - Goldman Sachs



|   |        |
|---|--------|
| 5 | +2.088 |
| 0 | +5.000 |
| 1 | +1.500 |
| 0 | +1.125 |
| 0 | +1.062 |

Structured Data  
(10-40%)



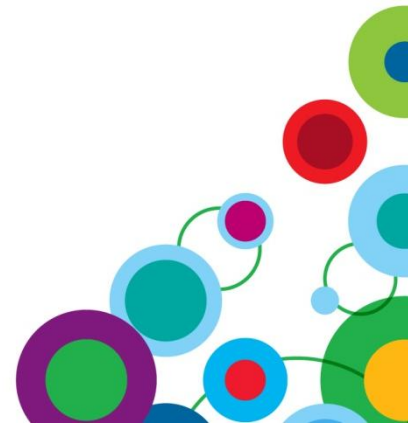
Unstructured Data  
(60-90%)





# The Importance of Text

“While data mining analyzes the contents of relational databases, taking the data in each cell of the database as building blocks for finding the patterns, trends, and relationships within the database, it ignores the bulk of information that most organizations must understand, if they are to plan and make decisions, respond to customer demands, or mitigate risks.”





# SPSS Text Analytics Business Applications

**CRM**  
Market Research  
HR  
Education

**Warranty Analysis**  
Brand/Reputation Management  
Market Intelligence  
Homeland Security  
Asset Management

**CRM**  
Market Intelligence  
Asset Management



**CRM**  
HR

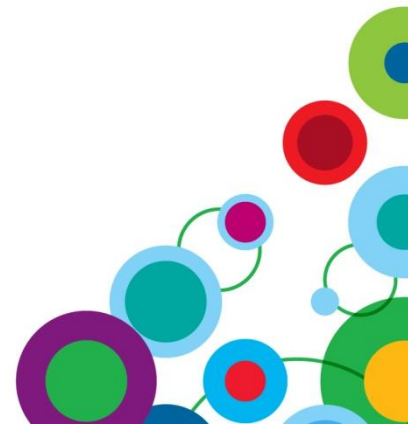
**Fraud pattern detection:**  
Insurance  
Banking  
Customs  
IRS





# Types of Analysis

- Typical analysis (reporting)
  - Measure. Compare. Report. Study.
  - Point-in-time analysis and progress dashboards
- Predictive analytics/data mining
  - Algorithms automatically “learn” significant patterns
  - Models make predictions for current/new cases
  - Insight delivered to drive better business decisions
- Adding text as a new dimension
  - Extracting usable knowledge from unstructured text data, through identification of core concepts and sentiment
  - Unlocking insights and connections between concepts and sentiment within text data
  - Increasing model accuracy (“lift”)

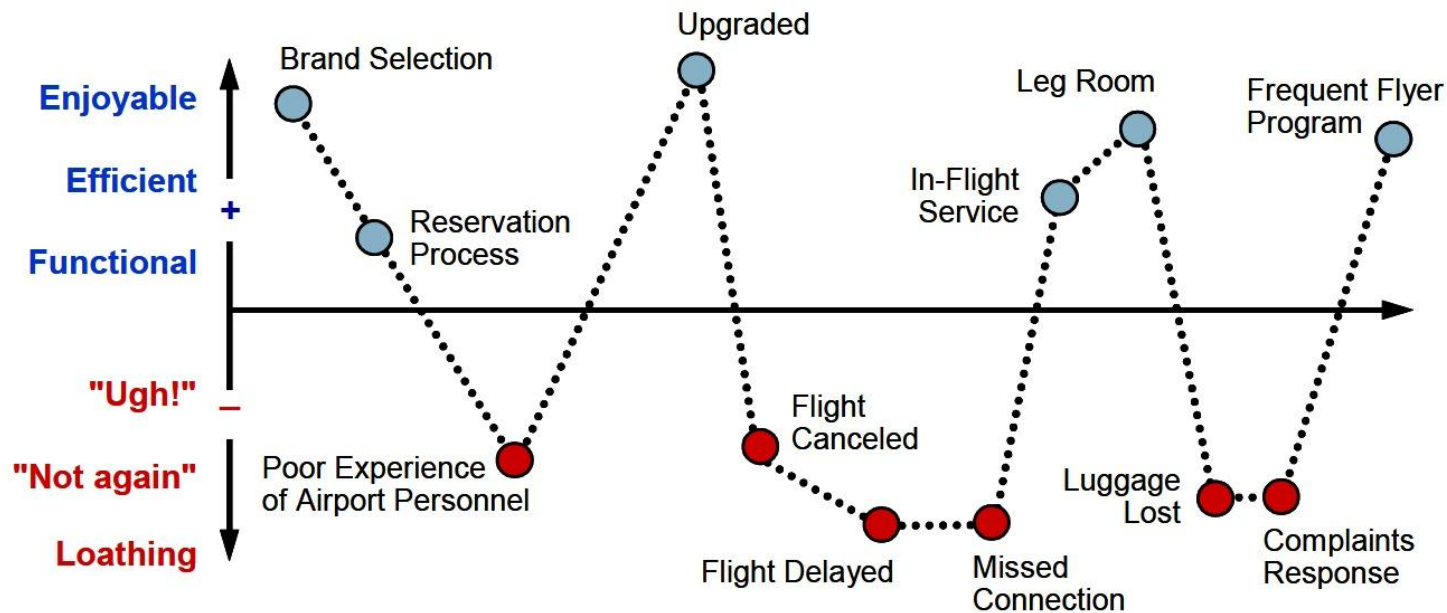






# Moment of Truth

- At every customer interaction touchpoint, there is at least one **"moment of truth"**—an interaction that can disproportionately positively or negatively affect the customer experience



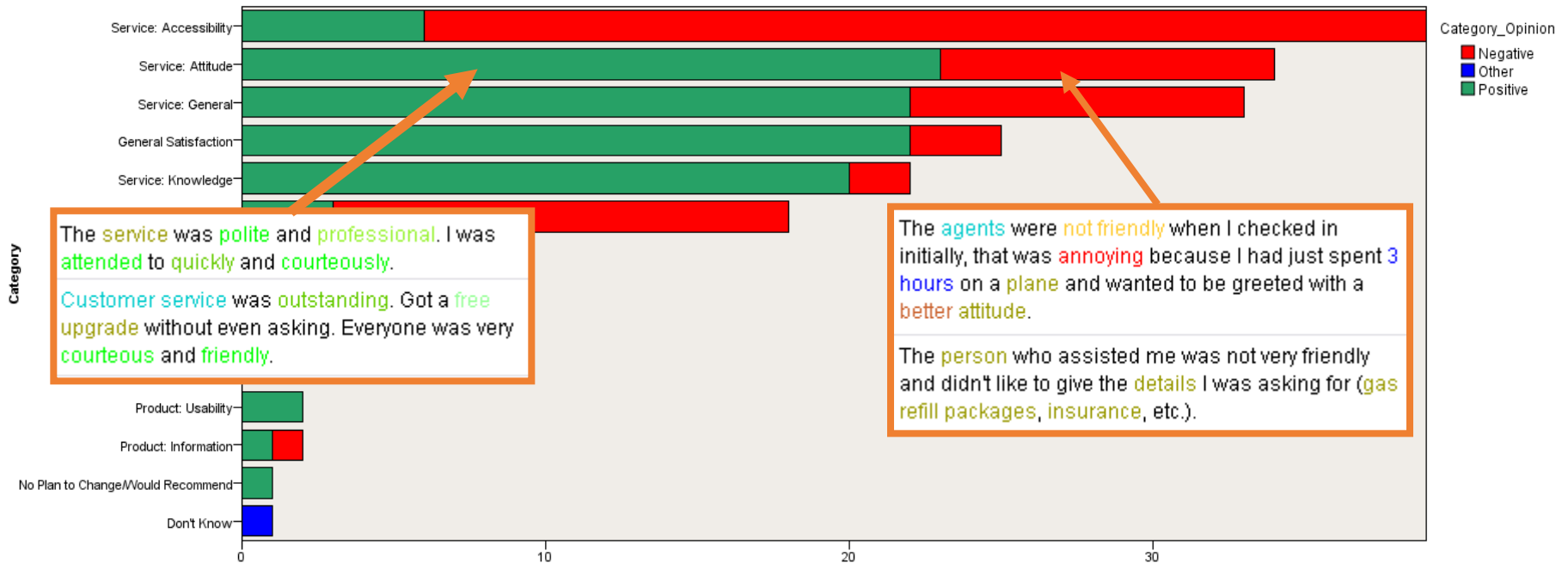
Source: Gartner (November 2007)





# Sentiment Analysis

Comments regarding customer experience:



**Sentiment Analysis** enables organizations to categorize a person's own words based on semantic relationships between **business issues** and **customer opinions**





# More Data Creates Better Predictions

## Interaction data

- Offers
- Results

**Web data**  
Up to 20%  
better  
predictions

**Text data**  
Up to 40%  
better  
predictions

## Attitudinal data

- Opinions
- Attitudes
- Up to 30%  
better  
predictions



## Descriptive data

- Attributes
- Characteristics
- Self-declared info
- (Geo)demographics

## Behavioral data

- Orders
- Transactions
- Payment history
- Usage history





# Two Approaches

**Core Capabilities**

- ✓ Structure the unstructured
- ✓ Explore linkages and associations
- ✓ Sentiment analysis
- ✓ Integrate into analytics

## IBM SPSS Text Analytics for Surveys



- Open-ended comments in surveys
- Categorize captured text for analysis
- Link text to structured survey data
- Ad-hoc analysis and hypothesis testing

## IBM SPSS Modeler Premium



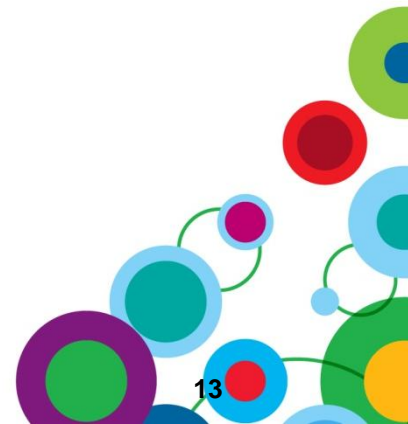
- Operational data, files, Web 2.0
- Create text models
- Combine with structured operational data
- Apply predictive modeling techniques
- Increase modeling accuracy





# Introducing IBM SPSS Modeler 14.1

- **Two *Editions***
- **IBM SPSS Modeler Professional**
  - Data access and transformation
  - Automatic data preparation
  - Modelling algorithms
  - Automatic model generation
  - Output and visualization
- **IBM SPSS Modeler Premium**
  - All of the above PLUS
  - Interactive text mining workbench

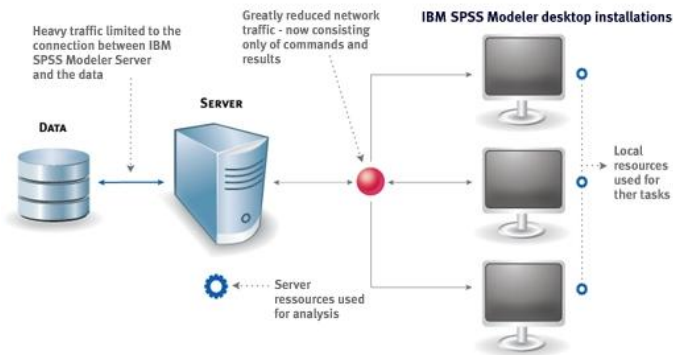
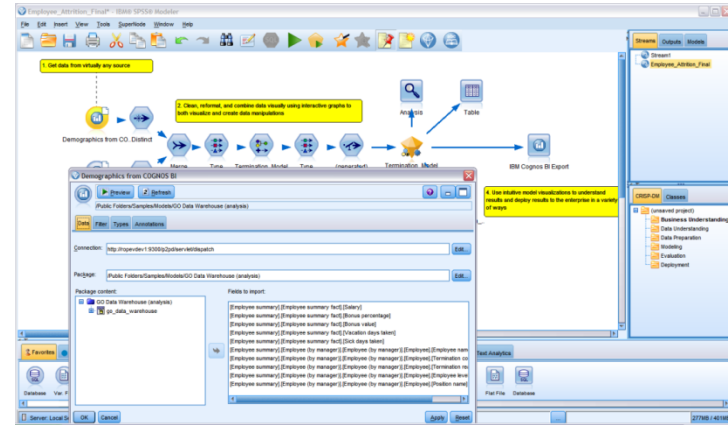




# IBM SPSS Modeler Deployment Options

## Client (Desktop)

- Access local files
- Connect to operational databases
- ★ Connect to Cognos BI
- Processing performed on local installation



## Client/Server

- Data operations/processing on server
- In-database data mining
- SQL pushback
- Modeler Batch
- ★ SuSE Linux Enterprise Server 10 (zLinux)
- ★ Inclusion in Smart Analytics System for Power (AIX) (*Professional only*)

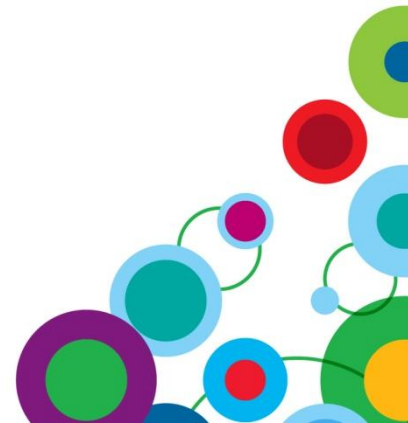
★ New!





# What's New in Text

- New industry-specific text analysis packages (TAPs)
- Ability to build and re-use hierarchical category structures
- User-friendly text-rule editor with simulation
- (for English) New semantic network for better classification results
- Ability to import and export pre-built categories for unstructured text easier
- Faster extraction built to handle large unstructured datasets
- All supported languages within the same product: Dutch, English, French, German, Italian, Japanese, Portuguese, Spanish.





# DEMO

## How Text Analytics Leverages Predictive Analytics...







# Summary

- Modeler brings repeatability to ongoing decision making with predictive modeling.
- Ability to leverage structured and **unstructured** data provides the most accurate predictive insight
- IBM SPSS Modeler
  - Provides ability to use predictive intelligence in decisions across all levels of the organization – strategic, operational and tactical
  - Provides insight into the factors that drive outcomes and the relationships therein to better mitigate risk and drive improvements
  - Intuitive visual interface, automation and simplified graphical views
  - Offers high performance, rich feature/functionality and integration
  - An open, platform agnostic architecture that “snaps into” any existing IT environment





# Summary

## IBM SPSS Modeler puts “predictive” into Predictive Analytics

**Modeler Professional** is a data mining workbench for the analysis of **structured numerical data** to model outcomes and make predictions that inform business decisions with predictive intelligence

**Modeler Premium** allows organizations to tap into the predictive intelligence held in all forms of data. Modeler Premium goes beyond the analysis of **structured numerical data alone and includes information from unstructured data such as web activity, blog content, customer feedback, e-mails, articles, and more to create the most accurate predictive models possible.**





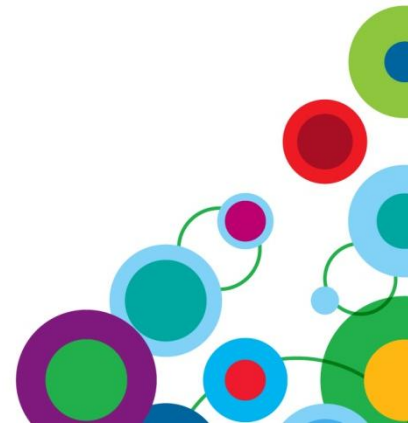
# Q&A





## Additional Resources

- [IBM SPSS Modeler](http://www.ibm.com) on [www.ibm.com](http://www.ibm.com)
  - <http://www.ibm.com/software/analytics/spss/products/modeler/>
- NEW & Updated Collateral
  - IBM SPSS Modeler Professional Spec Sheet
  - IBM SPSS Modeler Premium Spec Sheet
  - IBM SPSS Modeler and IBM Cognos 8 Business Intelligence Brochure
  - IBM SPSS Modeler and IBM InfoSphere Warehouse Brochure
- Questions about this presentation:
  - Eric Martin, [eric.martin@fr.ibm.com](mailto:eric.martin@fr.ibm.com)



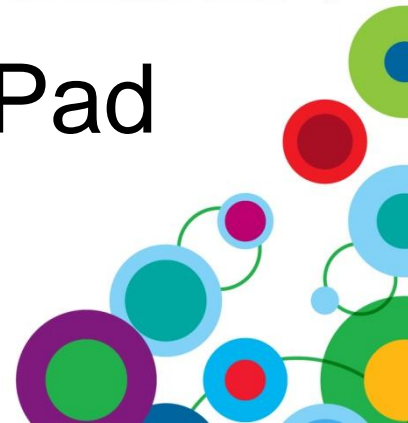
# **WIN** *an Apple® iPad*

Please remember to complete your session evaluation online at the Communication Station or point your Smart Phone browser to:

[www.spss.com/goldcoast](http://www.spss.com/goldcoast)



For a chance to win an Apple® iPad





# Disclaimer

© Copyright IBM Corporation 2010. All rights reserved.

**U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.**

**THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM’S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.**

IBM, the IBM logo, ibm.com, IBM Cognos 8 , IBM InfoSphere, IBM SPSS Modeler, IBM SPSS Statistics, and IBM SPSS Collaboration and Deployment Services are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)

Other company, product, or service names may be trademarks or service marks of others.





# APPENDIX

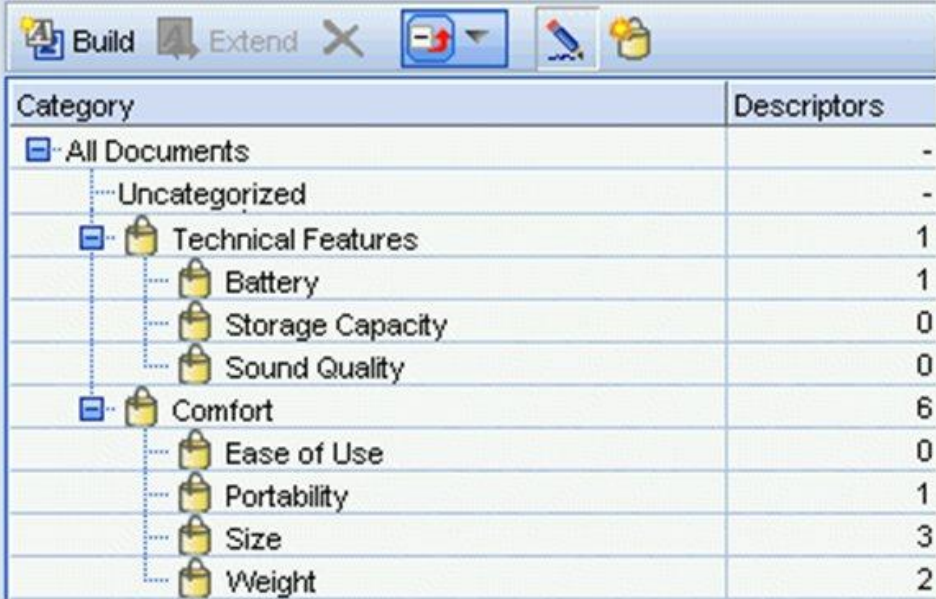
New IBM SPSS Modeler Premium Feature





# Hierarchical Category

- More accurate and more detailed classification through hierarchical categories
- Hierarchical categories can be build as part of normal extraction and classification process
- Users can modify and refine the hierarchies to better suit the level of detail they want to capture
- Can use the top level or the detail or select detail in text models



| Category                        | Descriptors |
|---------------------------------|-------------|
| [-] All Documents               | -           |
| [-] Uncategorized               | -           |
| [-] [Folder] Technical Features | 1           |
| [-] [Folder] Battery            | 1           |
| [-] [Folder] Storage Capacity   | 0           |
| [-] [Folder] Sound Quality      | 0           |
| [-] [Folder] Comfort            | 6           |
| [-] [Folder] Ease of Use        | 0           |
| [-] [Folder] Portability        | 1           |
| [-] [Folder] Size               | 3           |
| [-] [Folder] Weight             | 2           |

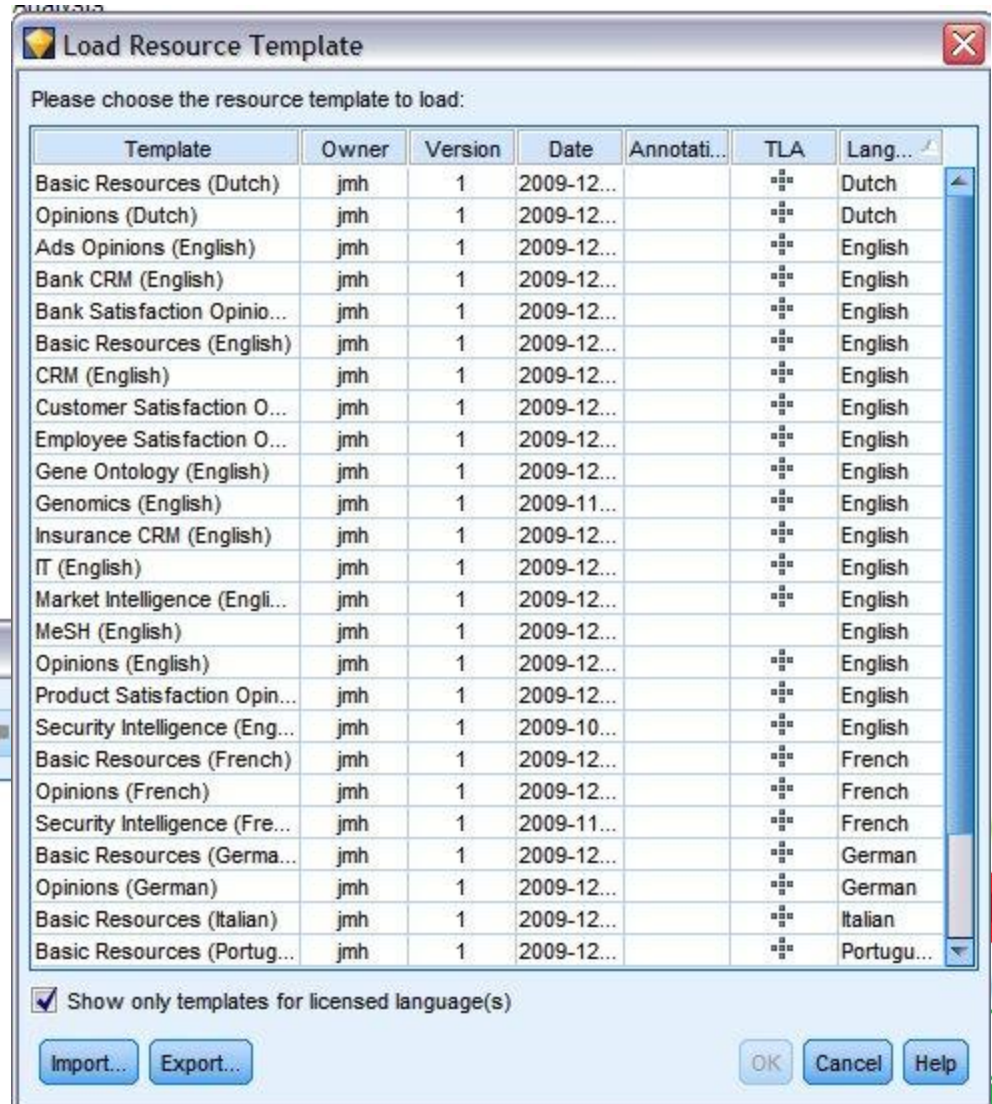
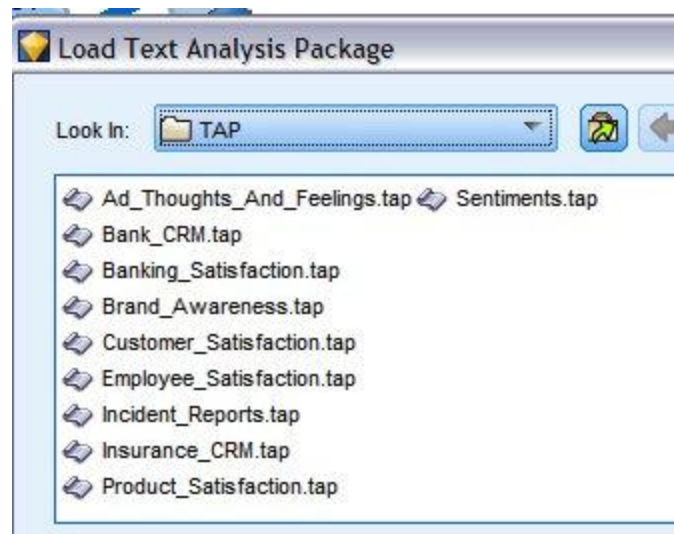




# Industry-Specific Text Templates and Text Analysis Packages



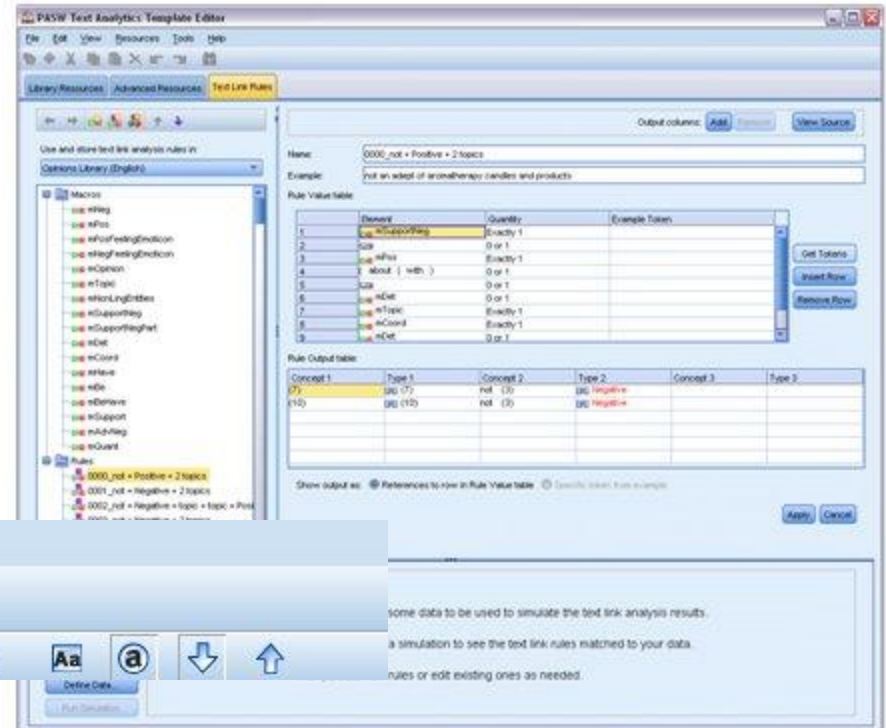
- Easier and more accurate text extraction and classification
- New semantic network
  - English-only
  - Supports hundreds of general themes and topics such as health, entertainment and politics
  - Makes classification results more accurate





# User-Friendly Text Link Rule Editor

- Create, modify and test text rules for better sentiment of facts and events analysis
- Simulate rules on a sub-set of the text to make sure it is generating the right results
- New toolbar option makes it easier to find text strings in categories, extraction results or in the data itself



# New Look and Feel Reinforces Consistency Across Portfolio

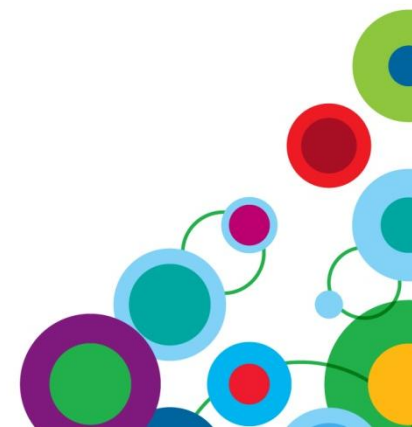


The screenshot displays the 'Interactive Workbench' software interface. The main window is titled 'Interactive Workbench - Of\_What\_do\_you\_like\_most...'. It features several panels:

- Top Left:** A table with columns 'Global', 'In', 'Type 1', and 'Type 2'. It lists various patterns and their associated types.
- Bottom Left:** A table titled 'Selected 31 patterns' with columns 'Global', 'Date', 'In', 'Concept 1', and 'Concept 2'. It shows a list of patterns and their associated concepts.
- Right Panel:** A network graph showing relationships between nodes. Nodes are represented by smiley faces and are connected by lines. A legend on the right indicates 'Global Count' with a scale from 1.0 to 3.0.
- Bottom Right:** A list of categories and their associated terms. The categories include 'memory device', 'songs', 'car', 'design', 'headphones', 'home', 'music', 'workspace', and 'screen'.

Below the main interface, there is a navigation bar with icons for 'File List', 'Web Feed', 'Text Link Analysis', 'Translate', 'Text Mining', and 'File Viewer'.

## New Text Mining Interface

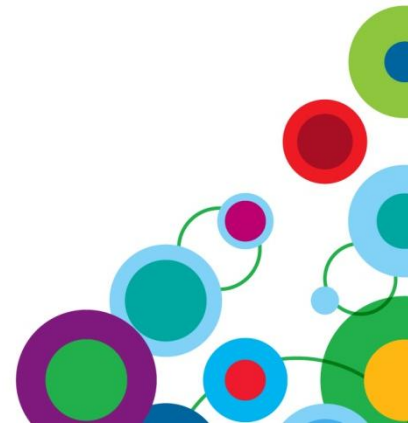
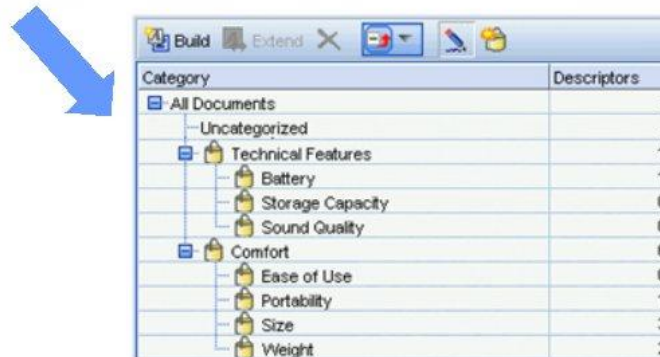
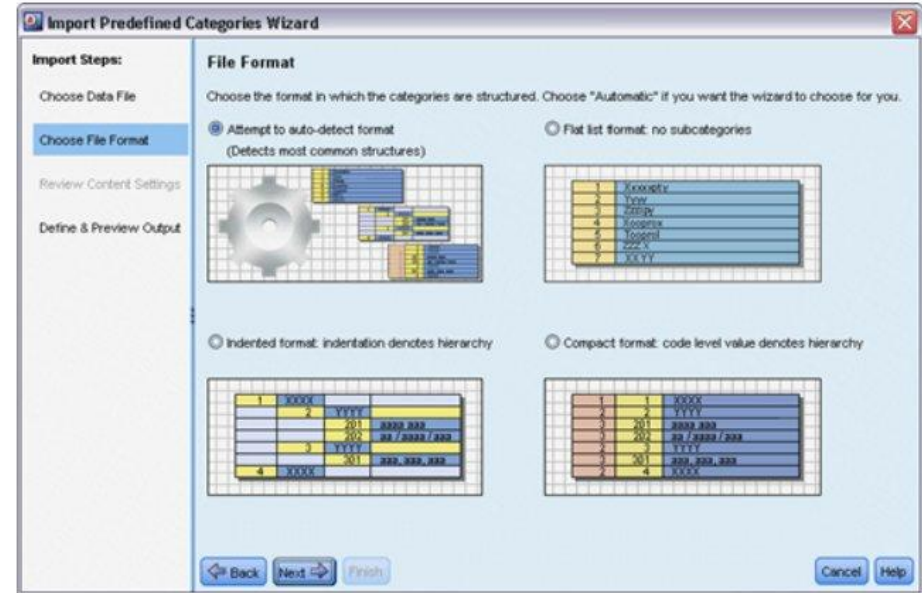




# Wizard-Based Import and Export

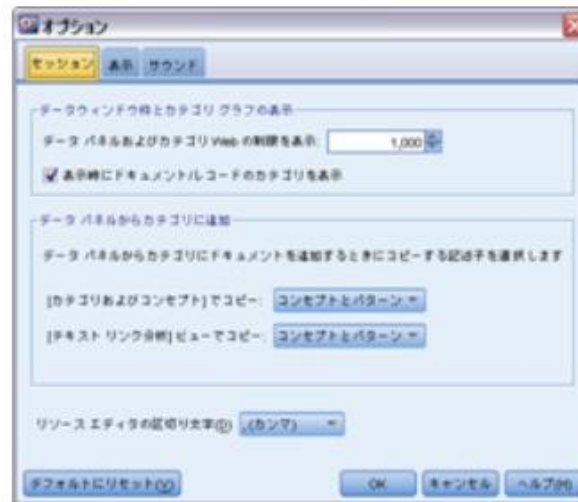
- Wizard-based import for pre-existing category structures, including hierarchies
- Import annotations and keywords
- Export to Excel from interactive session, including hierarchies

|    | A | B                    | C   |
|----|---|----------------------|---|
| 1  |   | 1 Technical Features |   |
| 2  |   |                      | 10 Battery any positive comment about battery life                            |
| 3  |   |                      | long-lasting  |
| 4  |   |                      | 11 Storage Capacity any positive comment about the amount                     |
| 5  |   |                      | 12 Sound Quality any positive comment about sound and quality                 |
| 6  |   |                      |   |
| 7  |   | 2 Comfort            |   |
| 8  |   |                      | 20 Ease of Use any positive comment indicating that the device is easy to use |
| 9  |   |                      | 21 Portability any positive comment indicating that the device is portable    |
| 10 |   |                      | handy   |
| 11 |   |                      | 22 Size any positive comment related to size, such as small                   |



# Japanese Extractor

- Enable customers working with Japanese text to leverage full capabilities of Modeler Premium
- Japanese text templates and text analysis packages available



*\* Requires purchase of Japanese Language Extractor*

