



## **SAS Text Miner: Capitalizing on the Value Hidden in Textual Data**

---

**THE  
POWER  
TO KNOW.®**

# Text Analytics

Information  
Organization and  
Access

Predictive  
Modeling, Discover  
Trends and  
Patterns

Content  
Categorization

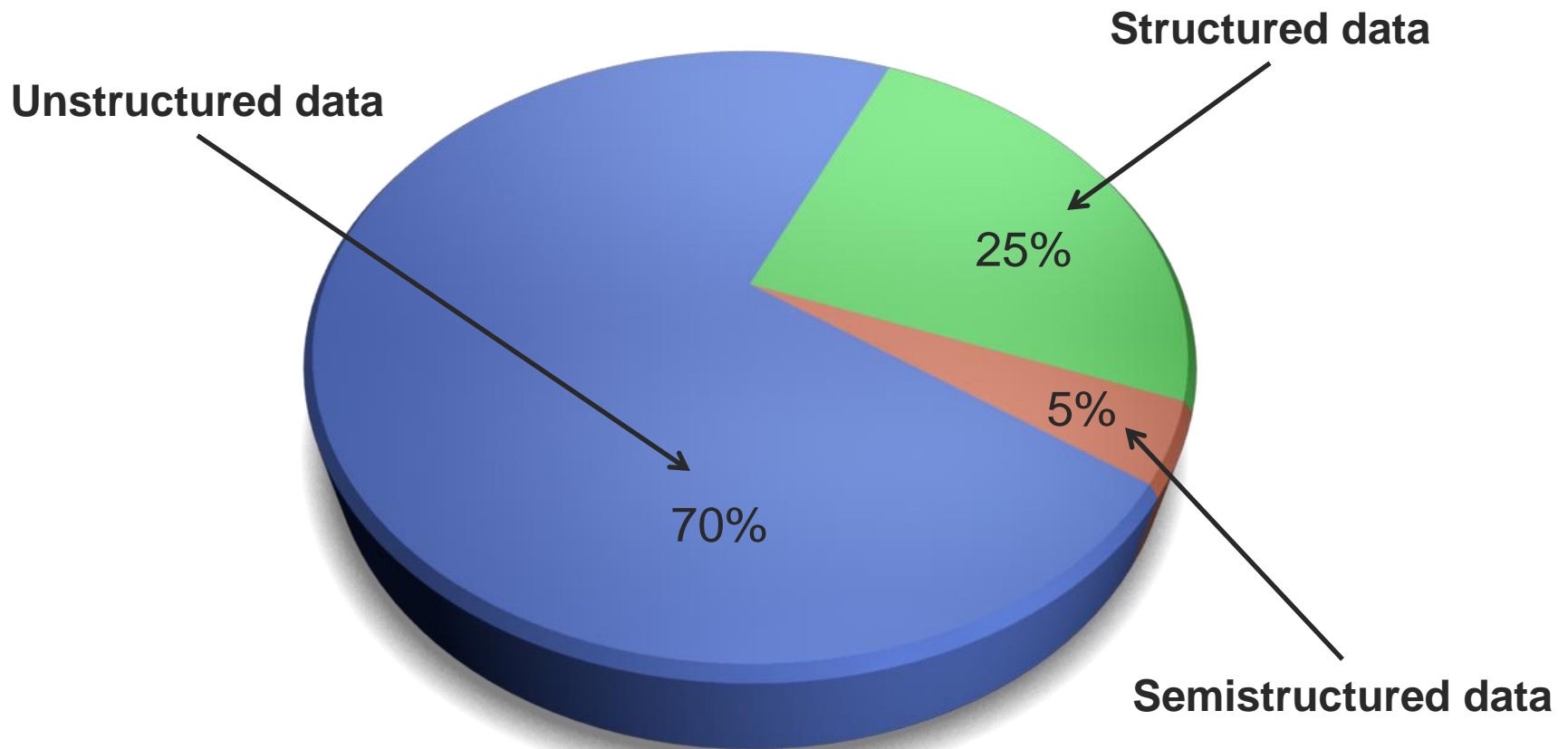
Ontology  
Management

Text Mining

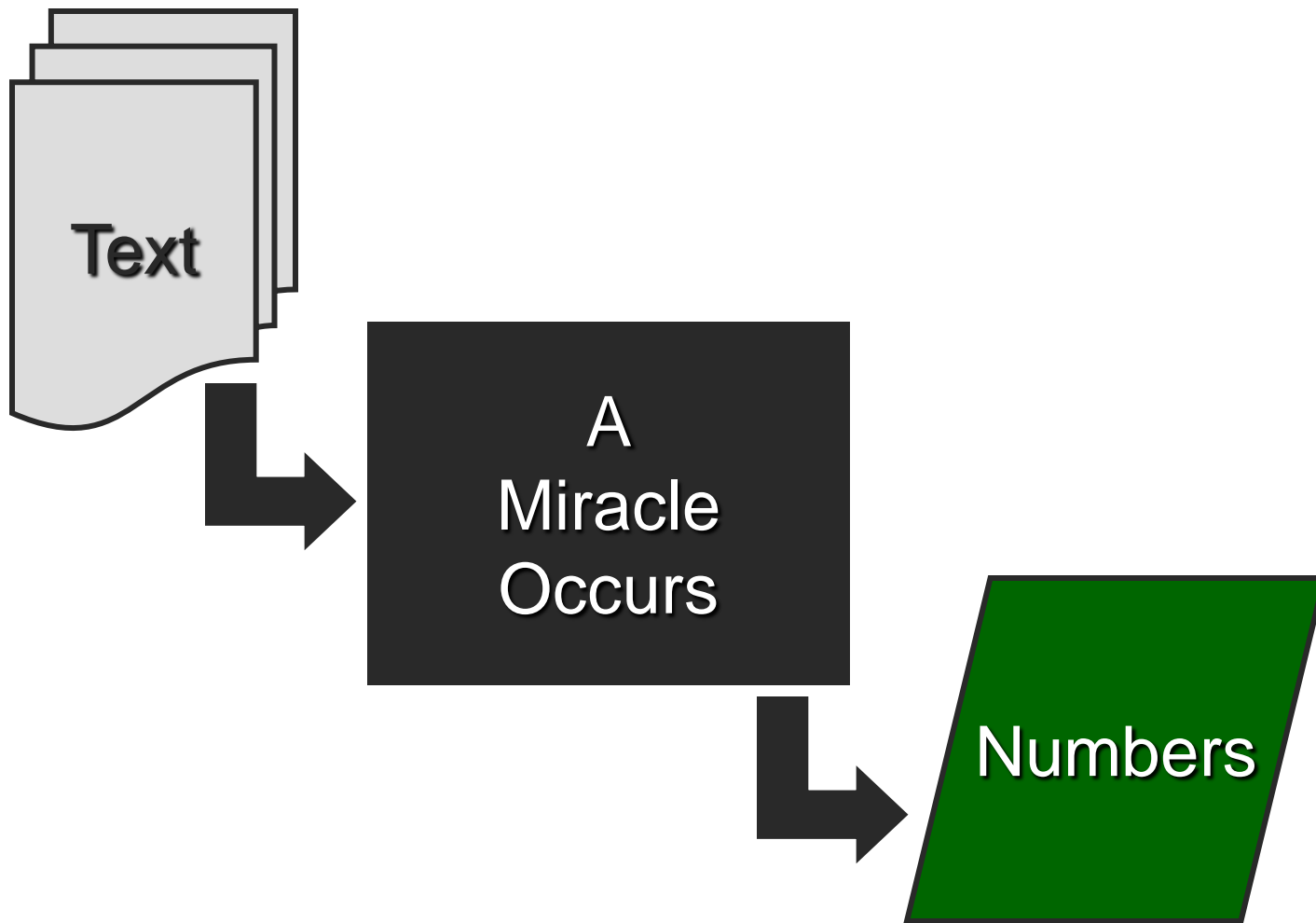
Sentiment  
Analysis

# Unstructured and Semi-structured Data

*The Dark Matter for IT*



# Text Mining



# Text Mining Definition

The process of discovering and extracting meaningful patterns and relationships from text collections



# Business Drivers for Text Analytics

**Text Analytics has numerous applications in any industry**

## Retail

Identify the most profitable customers and the underlying reasons for their loyalty.  
Brand management

## Finance

Retention of current customer base using call center transcriptions or transcribed audio.

## Insurance

Identify fraudulent claims.  
Track competitive intelligence.  
Brand management

## Manufacturing

Reduce time to detect root cause of product issues.  
Identify trends in market segments.

## Telecommunications

Help prevent churn and suggest up-sell/cross-sell opportunities for individual customers.

## Life Sciences

Identify adverse events.  
Recommend appropriate research materials.

# “Search” versus “Discover”

**Search**  
(goal-oriented)

**Discover**  
(opportunistic)

Structured  
Data

Data  
Retrieval

Data  
Mining

Unstructured  
Data (Text)

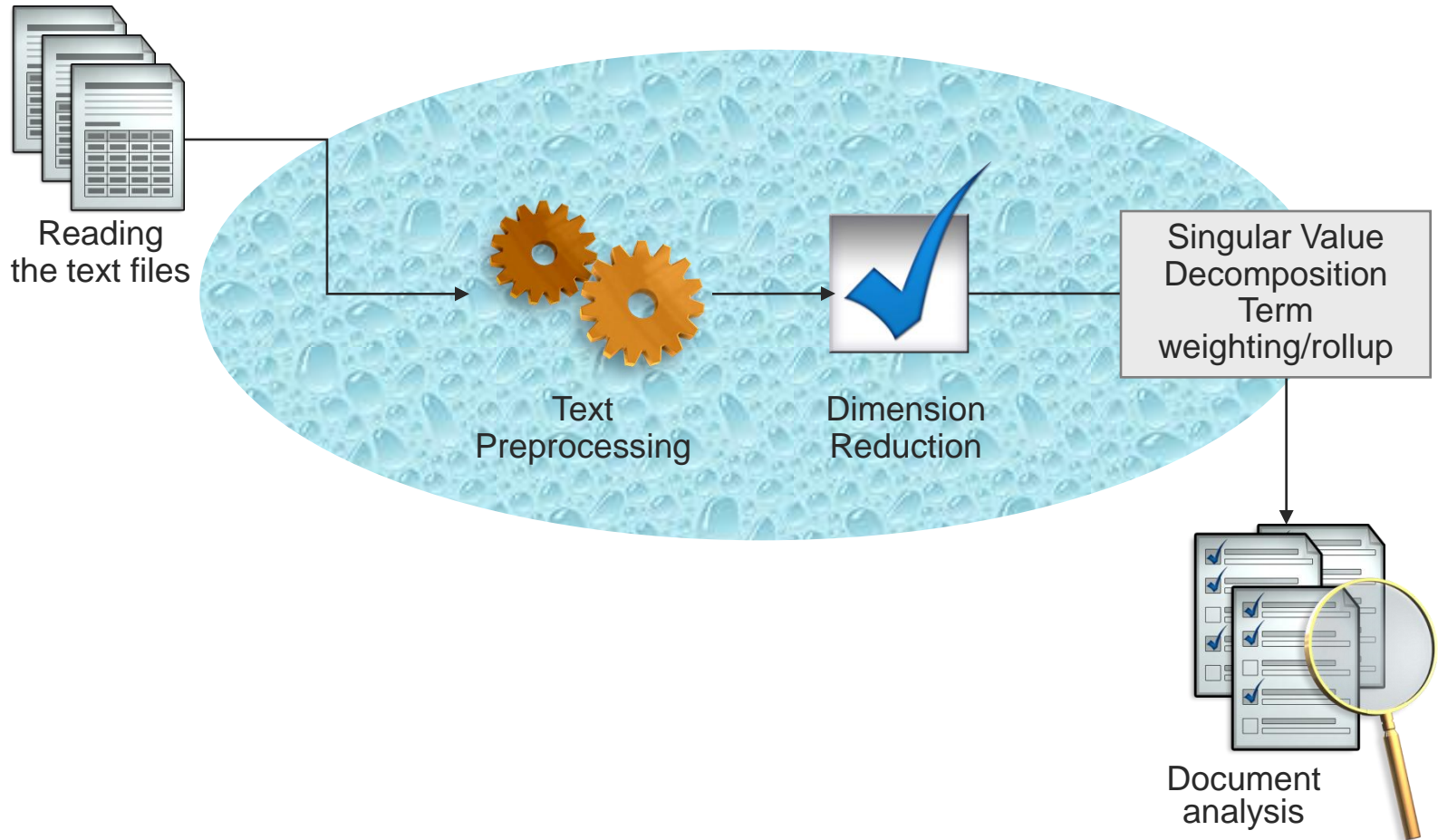
Information  
Retrieval

Text  
Mining

Source: Je Wei Liang,  
University of Texas, 1997

# Text Mining Process

S  
E  
M  
M  
A



# Text Miner 4.2 (15Dec2009)

Enterprise Miner - CSAT

File Edit View Actions Options Window Help

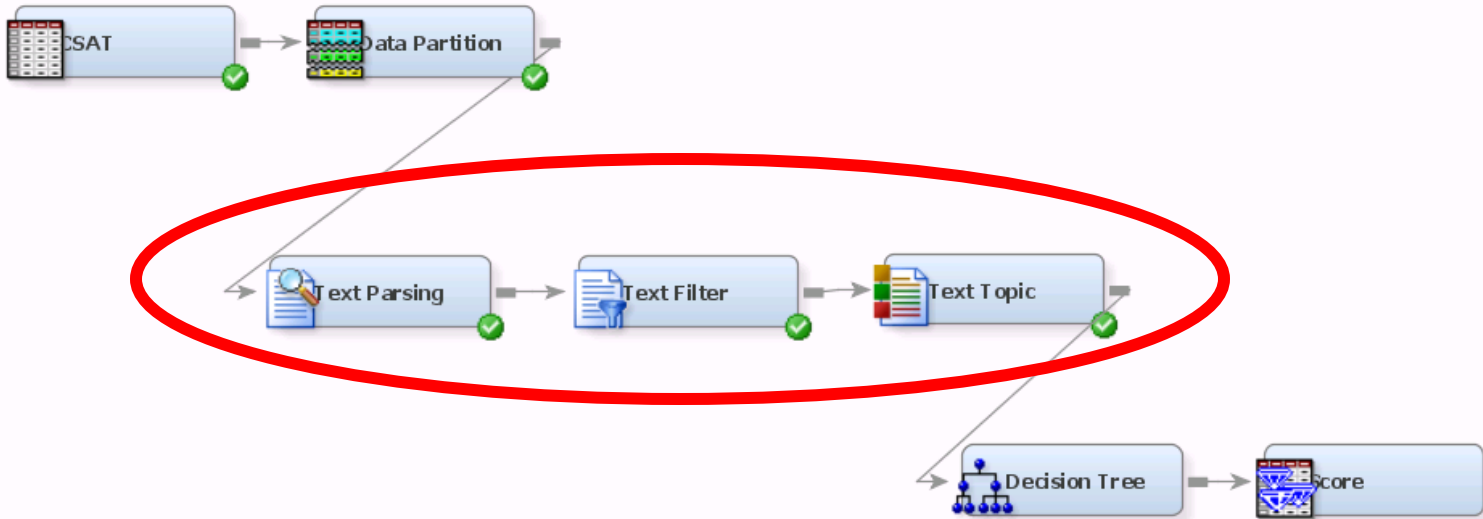
Sample Explore Modify Model Assess Utility Credit Scoring Text Mining

Customer Satisfaction

CSAT

- Data Sources
- Diagrams
  - Customer Satisfaction
  - Node names
  - Target Twitter Data

Property	Value
ID	EMWS
Name	Customer Satis
Status	Open
Notes	...
History	...



```

    graph LR
      CSAT[CSAT] --> DP[Data Partition]
      DP --> TP[Text Parsing]
      TP --> TF[Text Filter]
      TF --> TT[Text Topic]
      TT --> DT[Decision Tree]
      DT --> S[Score]
  
```

# Two Passages into the Data Mine



- Describe patterns in data



- Use known results to train/estimate values for different data

# SAS Text Miner Capabilities



- Exploratory Analysis & Visualization
  - View term statistics, identify similar documents
  - Graphically display term relationships



- Automatic Classification
  - Taxonomies – e.g., Identify Medline abstracts by hierarchies
  - Discrete groupings – e.g., Identify clusters of call center comments



- Predictive Modeling
  - Text alone – e.g., predict patients conditions based on physicians comments



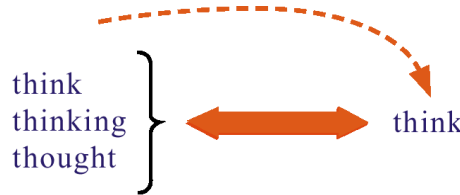
- Text and structured data – e.g., predict future purchasing based product comments combined with demographics

# Support for Multiple Languages

- English
- French
- German
- Dutch
- Italian
- Portuguese
- Spanish
- Swedish

# Text Miner Parsing

- Stemming
- Part of Speech



- Determines if the word is a common noun, verb, adjective, proper noun, adverb, etc.
- Disambiguate parts of speech when a word is used in a different context,
  - *I wish that my **bank** did not have a service charge for using other vendor ATM's.*
  - *You can **bank** on either Germany or England winning the world cup next year.*

- Entity Extraction

**Places**  
White House



**People's Names**  
James A. Goodnight



**Dates**



# Text Miner Parsing

- **Specify Start/Stop/Synonym Lists**
  - Filtering out low information words such as
    - articles (e.g. the, a, this)
    - prepositions (e.g. of, from, by)
    - conjunctions (e.g. and, but, or)
  - Consider document subject matter
  
- **Vertical dictionaries**
  - Automatically generate synonyms appropriate to your data

# Quality ≠ Quantity

- UPPERCASE
- Miss-spelings
- A.C.R.O.N.Y.M.S
- Shrt-hnd
- Pr😊f@nity
- \*Punctuation\*

# Benefits of SVD Approach

- Assumes that there is something conceptual underlying words (obscured by word choice)
- Each concept is expressed as a weighted combination of all the kept terms in the collection

# Text Clustering

- Expectation Maximization Clustering
  - Generates groups of similar documents from output of SVD
  - Fast clustering of many documents
  - E-M clustering is only available through Text Miner
- Hierarchical Clustering
- Optionally use the Self Organizing Map or Clustering node downstream in the Process Flow

# Demo...



# Success Stories

- American Honda - In addition to saving millions in warranty costs, American Honda improves safety and customer satisfaction with SAS.
- Sub-Zero Freezer Company - Freezer manufacturer uses SAS Warranty Analysis to reduce problem detection times by 50 percent.
- Hewlett-Packard - With SAS Text Miner, Hewlett-Packard reaches the next generation of CRM through analyzing customer comments.