



## eClassifier

eClassifier is a comprehensive interactive text analysis application and runtime with capabilities for:

### • Clustering and Taxonomy Generation

- Multiple clustering algorithms including K-Means, EM, query-based clustering, and structure from associated meta-data
- Rich set of interactive manipulation capabilities including add category, delete category, subclass a category, merge categories, and drag-and-drop
- Multiple metrics and measures including cohesion, distinctness and automatic category naming

### • Text Mining and Analysis

- Various means of summarization and sorting used to quickly understand concepts in large amounts of text
- Dictionary and taxonomy Co-occurrence and dependency analysis to find hidden relationships within the information
- Trend analysis to find emerging and time dependent phenomena
- Analysis of text and associated meta-data to find relationships
- Keyword searching

### • Classification

- Multiple algorithms and meta-algorithms including rule induction with editing, statistical classifiers, support vector machines, naive-bayes, nearest centroid as well as multi-classifier voting algorithms.
- Classification rate of on the order of a million documents per hour
- Modeling and validation support thru direct inspection and automatic calculation of precision and recall.

### • Visualization

- Interactive scatter plot visualization used to visual relationships

### • Storage, Formats & Languages

- Multiple data stores including file systems, databases, lotus notes and others
- Multiple text formats including web, xml, and other text formats
- Multiple languages, although majority of experience is in English
- Scalability of taxonomy generation is up to a million documents
- Extensive dictionary tools for the discovery and editing of stop words, include words, synonyms and stock-phrases

### • Platforms, Applications & Runtime Integration

- A runtime API to integrate into other applications and work-flow. This can be used to perform real-time or batch analysis and classification.
- This technology is 100% Java, currently using Java 2 version 1.3.1

## Application Scenarios

eClassifier technology is used to derive business value from text.

Example application scenarios include:

- Patent Mining
- Customer Relationship Management Insights
- Market Intelligence
- Competitive Intelligence
- Gap Analysis
- Knowledge Discovery

## Licensing & Consulting

- Licensing terms available upon request
- Research scientists available for consulting
- Contacts:
  - Jeffrey Kreulen ([kreulen@almaden.ibm.com](mailto:kreulen@almaden.ibm.com))
  - Scott Spangler ([spangles@almaden.ibm.com](mailto:spangles@almaden.ibm.com))

