

*Knowledge management and data mining leverage relevant information from personnel and information systems, preserve corporate knowledge and add information from unorganized data sources to the explicit store of enterprise knowledge.*

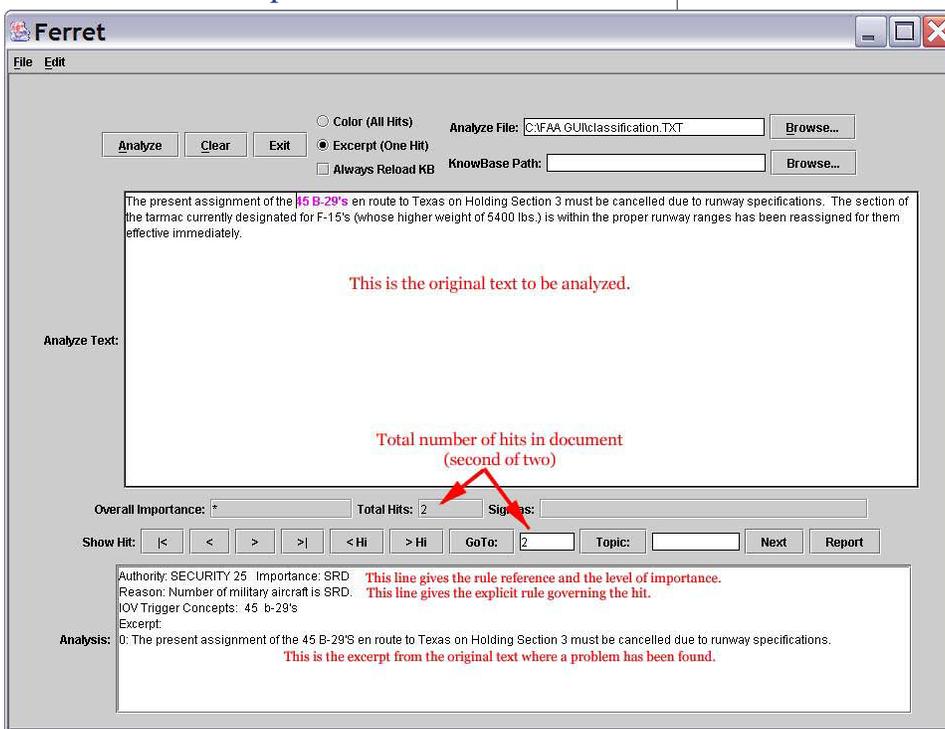
No modern business or government agency can thrive without knowledge. Amid mountains of data, relevant information must be identified, structured and directed toward work processes to facilitate or enable the achievement of goals and end-products. Knowledge management structures, maintains and provides access to knowledge within information systems. Data mining extracts relevant tacit knowledge from information system assets through data correlation.

The Y-12 National Security Complex has extensive capabilities in many facets of knowledge management, including highly secure data warehousing, expert system information-of-value determination, advanced document and content indexing and access systems and knowledge engineering. Data mining can be performed in highly automated fashion on single documents or archives of related documents, using expert systems and conceptual clustering.

## CAPABILITIES

- Model-based document life-cycle management
- Automated workflow and configuration management
- Document imaging and on-line archival
- Metadata- and content-based access control
- Automated content classification and characterization

## Unclassified Example



The screenshot shows the Ferret software interface. At the top, there are menu options 'File' and 'Edit'. Below that, there are buttons for 'Analyze', 'Clear', and 'Exit'. To the right, there are input fields for 'Analyze File:' (containing 'C:\FAA GUI\classification.TXT') and 'KnowBase Path:' with 'Browse...' buttons. There are also radio buttons for 'Color (All Hits)', 'Excerpt (One Hit)', and 'Always Reload KB'. The main window displays a text excerpt: 'The present assignment of the 45 B-29's en route to Texas on Holding Section 3 must be cancelled due to runway specifications. The section of the tarmac currently designated for F-15's (whose higher weight of 5400 lbs.) is within the proper runway ranges has been reassigned for them effective immediately.' Below the excerpt, there is a red annotation: 'This is the original text to be analyzed.' At the bottom of the window, there are fields for 'Overall Importance: 1', 'Total Hits: 2', and 'Signas:'. Below these are navigation buttons: 'Show Hit: < < > > < Hi > Hi GoTo: 2 Topic: Next Report'. At the bottom, there is an 'Analysis:' section with the following text: 'Authority: SECURITY 25 Importance: SRD Reason: Number of military aircraft is SRD. IOV Trigger Concepts: 45 b-29's Excerpt: 0: The present assignment of the 45 B-29's en route to Texas on Holding Section 3 must be cancelled due to runway specifications. This line gives the rule reference and the level of importance. This line gives the explicit rule governing the hit. This is the excerpt from the original text where a problem has been found.'

*Y-12's Q software (a.k.a. Ferrett) finds classified information by scanning a text file for sensitive terms and proposes an overall classification. The Q software can also be applied to concept networks, the next-generation of knowledge management systems.*

- Knowledge engineering, knowledge preservation and specialized knowledge base development
- Enhanced content management and integration using semantic web technologies, XML and web services
- Taxonomic conceptual clustering and natural language processing technologies
- Expert system development

---

### APPLICATIONS

---

Y-12 provides document and content management support to the customers within the U.S. Department of Energy Nuclear Weapons Complex and the U.S. Department of Defense. The application of document management and data mining technologies provides essential services in business and engineering process support, is cost-effective and enables processes and analysis that would otherwise be unfeasible. Document management and data mining support have been provided in the following areas:

- Product engineering document and drawing repositories for highly classified environments
- Record imaging and archival products
- Sensitivity level and governing regulation determination systems
- Sensitive excerpt identification and redaction systems
- Subject and category determination systems
- Need-to-know policy enforcement subsystems
- Web crawler and storage system robot to identify and summarize pages or files with “important” content
- System to monitor a document archive for important new files with automated notification
- Data mining engine to perform conceptual clustering using analyst-weighted, numerical taxonomy-based similarity coefficients
- Expert inference engine to perform automated real-time determination of information of value, possibilities and significant elaboration on unformatted text
- Knowledge base editing tools to support knowledge engineering by subject matter experts

*For more information, contact*

Gina Davis  
davisgk@y12.doe.gov  
865-963-5646 BlackBerry  
865-576-0181 Office

Connie Malarkey  
malarkeych@y12.doe.gov  
865-574-0655

Y-12 National Security Complex  
P.O. Box 2009  
Oak Ridge, TN 37831

*Visit us on the web at [www.y12.doe.gov](http://www.y12.doe.gov)*