



Recurring Anomaly Detection System

Reference No. ARC-15942-1

NASA experts analyze records in International Space Station and Shuttle databases to identify recurring safety issues or concerns, and to help confirm that important problems have not been overlooked. Innovative software has been developed for discovering recurring anomalies associated with NASA missions. ReADS, or Recurring Anomaly Detection System, uses text-clustering algorithms to group loosely related reports and documents, which reduces human error and fatigue. In addition, it identifies interconnected reports, automating the discovery of possible recurring anomalies. ReADS has been integrated into a secure web-based search tool to allow users to perform their own text mining.

Benefits

- Method for context-based text mining using an XML database
- Data mining clustering viewer for selecting documents
- Data mining server that runs one or more clustering algorithms
- Easy to use system that enables faster processing time
- Enables conduct of searches on multiple databases
- Enables discovery of recurring anomalies across multiple databases, in a system or project, to determine weaknesses or high risks

Technology Details

ReADS enables users to select one or more databases to search, enter context search terms, select one or more data mining algorithms, then click on the search and/or mining text buttons. If search is selected, then the context and/or content search results are displayed in the internet browser for the databases selected. If text mining is selected, a request is sent to a text-mining server for the specified databases, context+content search terms, and clustering algorithms. Once the text mining has finished, the results are formatted in XML, uploaded to the XML database server, and an email notification is sent to the user with a URL link to the text mining results. Users can also select sets of documents from the text-mining viewer and initiate a query of those documents.

ReADS utilizes two primary servers: 1) Query-Based Search server running Netmark, and 2) Text mining clustering server running Java and Matlab. A reverse proxy is utilized for communication between the servers. Alternatively, ReADS can run on the same server or distributed load balance network of servers. Netmark is a state-of-the-art software tool for querying across multiple databases, and was previously used for similar data processing as part of the NASA Mishap and Anomaly Information System (MAIS).

Commercial Applications

- Aerospace systems
- Aircraft industry
- Information Technology Industry
- Database Management
- Telecommunications
- E-Commerce
- Health Care Information Systems
- Scientific and Engineering industry

