PIRANHA: Intelligent Agent-based Large-scale Data Analysis

2007 R&D 100 Award Winner

Applied Software Engineering Research Group

Computational Sciences & Engineering Division

Problem Statement:

- There is a massive amount of intelligence data available that cannot be manually analyzed. Computers can provide some help in this problem, but the shear volumes of data make the most promising approaches impractical. The challenge is for a computer to sift through a large amount of data & provide a human with accurate and relevant information, not to merely allow the analyst to search over an ever increasing set of data. This requires software that is able to filter, relate, and show documents & relationships to an analyst.

Technical Approach:

- We have developed intelligence agent-based technology that allows for advanced textual analysis to be done on very large and dynamic data with unprecedented accuracy. This capability can allow for an intelligence analyst to find similar documents, eliminate or find duplicate documents, create representative samples of a set of documents, and to automatically create folders of documents related to a current interest.

Benefit:

- Using Clustering textual information provides a number of useful capabilities for an intelligence analyst. The technology that has been developed has been vetted in the scientific community, as well as a number of real world applications.

Point of Contact:

Thomas Potok
(865) 574-0834
potokte@ornl.gov