



NATIONAL SECURITY AGENCY TECHNOLOGY TRANSFER PROGRAM

Office of Research & Technology Applications

US Patent #8,799,339

NSA Reference #1742

AVAILABLE FOR LICENSING

Measuring Similarity Between Data Sets



BIG DATA

This technology enables users to measure similarity between data sets without needing to know how the sets interact with each other. Other approaches require knowledge of the sets' interaction using the Jaccard index. This process is more efficient and accurate because it runs searches despite insertion of duplicate information, resulting in greater flexibility of search parameters and more efficient processing.

This capability is critical for managing and sorting through immense quantities of data and can enhance big data analytics in multiple fields, including finance, genetics, and law enforcement.

FEATURES

Compares sets without knowing points of intersections

Runs searches despite insertion of duplicate information

BENEFITS

Greater flexibility of search parameters

More efficient processing

POTENTIAL APPLICATIONS

- Knowledge management
- Genetic analysis
- Financial forensics
- Social network analysis

DEVELOPMENT STAGE

EARLY

MID-STAGE

COMPLETE

This technology is **complete**. A technical briefing is available upon request.

National Security Agency
Office of Research & Technology Applications

9800 Savage Road, Suite 6843
Ft. Meade, MD 20755-6843

(866) 680-4539
tech_transfer@nsa.gov
www.nsa.gov/techtransfer