tamr

Elevating Operational Intelligence With Automated, Scalable Data Insights

Machine Generated Data Only Leads To Intelligence If It's Intelligently Processed

In today's digital world, the number of data sources, public and dark, is increasing at an exponential rate. To an intelligence agency, more data from more sources can lead to richer, higher quality intelligence. Unfortunately, human analysts can't be hired at a rate to keep up with demand for preparing these data sources for use. Machines can help to automatically parse and process data sources, but only if the generated data is of high quality. A government intelligence agency approached Tamr to automatically process millions of data sources while still ensuring the highest levels of data quality.

Specifically, the agency had an internal database on persons of interest but sought to improve their ability to integrate unstructured text sources, such as social media, news sites, and analyst reports, into that database. If this process could be automated without compromising data quality, it would give them a more accurate and comprehensive view of those people. In essence, the intelligence agency could cast a wider net to capture insight from all data available, ensuring no persons of interest 'fall through the cracks' and all available insight on known persons is captured.

Unlocking the power of the merged datasets inherently comes with significant challenges, such as:

+ Difficulty in resolving data extracted from websites -- data sources like news articles are free text and may result in a high count of extracted entities that do not easily match to database records

+ Problems with disambiguation of persons -- it's very challenging to identify separate individuals who have the same name or recognize that two people identified by different nicknames or aliases are actually the same person

## Tamr Works With Complex Datasets To Provide Critical, Comprehensive Insight

To demonstrate the capability to overcome these hurdles, in a pilot, Tamr was asked to integrate news information into a database of people. In the solution:

+ A technology partner crawled news sites and social media to conduct entity extraction (of people, places, etc.) through the use of natural language processing -- transforming free text into intelligible facts

+ Tamr then took the data coming out of the entity extractor and conducted entity resolution with the existing database. In this step, person disambiguation is extremely difficult but accomplished through Tamr's identification of relationships between people, places, organizations, events, and more. For example, Tamr could recognize that "Moamer" and "Muammar Gaddafi" are the same entity as they would both be mentioned alongside "Lockerbie" and "Benghazi". Moreover, Tamr used a unique combination of automation and expert sourcing to match people, organizations, and other entities in a rapid, scalable, and accurate manner -- ultimately significantly differentiating them from current methods

John Doe   John Doe   John Doe
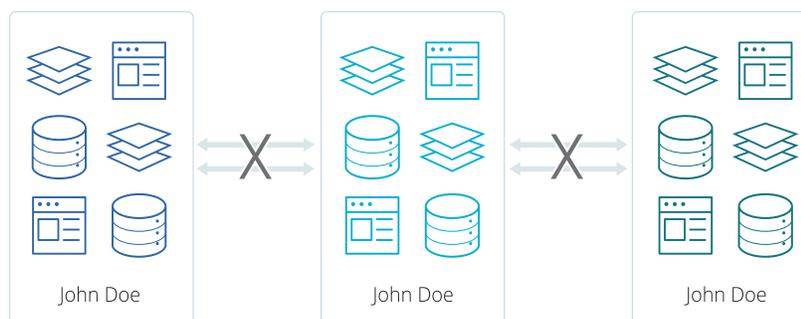
Tamr's Approach To Integration Elevates Operational Intelligence In Mere Weeks

Tamr's use of machine learning and expert sourcing in the integration of datasets for the agency enabled matching of over 20 million people from the various news sources against the existing corpus of over 1 million people. To accomplish this, Tamr created a unified list of millions of places, organizations, objects, and other entities by connecting the known references to these objects to references in the news sources. From there, Tamr processed over 100 million relationships stitching this network together. In the end, Tamr was able to use these relationships to integrate the persons mentioned on the web into the known persons data set with a 95% accuracy rate on the verification data set. This integration effort, which would have been infeasible through manual efforts, took Tamr merely a couple of weeks.
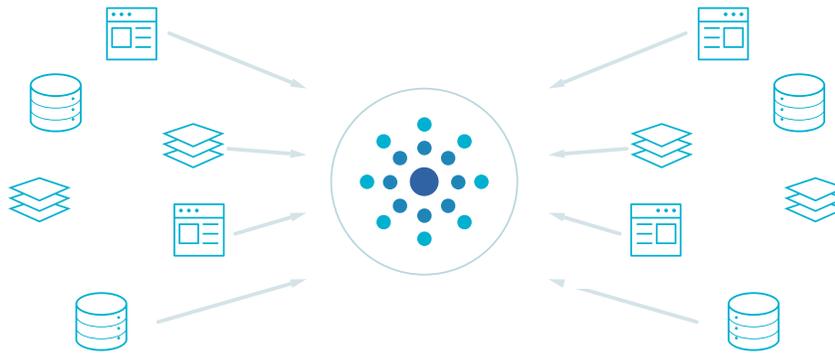
Tamr's ability to resolve entities is not limited to people. Current projects involve using Tamr to try to resolve discrete events coming out of event feeds such as GDELT, and to integrate the event data coming from multiple entity extractors. Ultimately, Tamr's entity resolution technology in conjunction with natural language processing tools can be used to integrate and understand more sources of data than ever before while maintaining the level of data quality necessary to produce actionable insights.

About Tamr

Tamr, Inc., provides a data unification platform that dramatically reduces the time and effort of connecting and enriching multiple data sources to achieve a unified view of siloed enterprise data. Using Tamr, organizations are able to complete data unification projects in days or weeks versus months or quarters.

For your own personalized Tamr demo, visit www.tamr.com.