Using Knowledge Graph to improve enterprise search experience

Dmytro Dolgopolov[1] and Elena Romanova[1]

[1] FINRA (Financial Industry Regulatory Authority) Rockville, MD 20850, USA

Abstract. FINRA has many millions of documents and database records that staff need to search through to find information relevant to regulatory activities. Searching across the large set of documents and structured database records using relevance ranked text search does not present items together that the users know are related. Relevance ranking discriminates using TF/IDF, and related techniques, but does not bring together items that are not related by relevance.
The solution was to build a structured and navigable visual representation of the data returned by the underlying multiple query engines. Text mining and semantic web techniques were used extensively to build the enhanced metadata and create the linkages among the data objects needed in order to support the visual navigation paradigm. The resulting knowledge graph gives users the ability to see semantically related items.

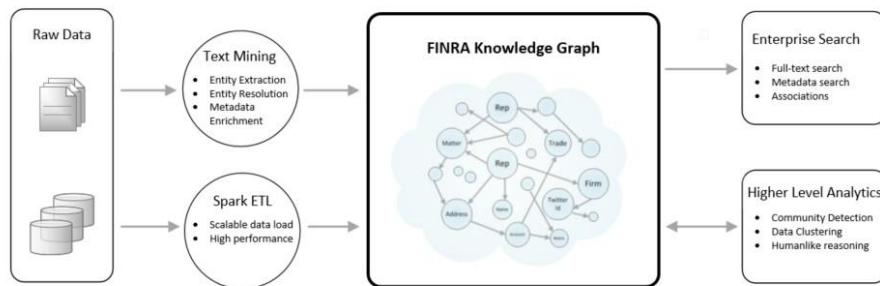**Keywords:** Knowledge Graph, Semantic Web, Text Mining, Enterprise Search, Graph Analysis, RDF store

1 Background

FINRA [1] is not-for-profit organization authorized by US Congress to protect America's investors by making sure the broker-dealer industry operates fairly and honestly. As part of its regulatory mission FINRA's staff has to review millions of structured and unstructured data elements located in numerous internal systems. This includes information found in the free style text fields as well as various documents. Investigators, examiners and analysts get easily overwhelmed with the amount of information they have to deal with. These challenges are exacerbated by the copies and 'near' duplicates. Staff spends days collecting the information required in preparation to exam or investigation. FINRA needed a solution to help users navigate through internal and external data sets collected by various systems.

2 FINRA Knowledge Graph

We combined the power of Semantic Web, Text Mining, Enterprise Search and Graph Analysis to create FINRA Knowledge Graph. Our solution uses Semantic Web to connect these technologies and make the whole to be greater than the sum of its parts. We implemented an ETL pipeline that leverages Spark's DataFrames to prepare and load

millions of records to the RDF store in less than 20 minutes. We created a scalable semantic inference engine in Spark to produce new connections across heterogeneous data. This engine derives new facts from an existing set of data using humanlike reasoning. Text mining enriches the data by extracting individuals, organizations and their features from documents and free-style comments that are then persisted to an RDF store. We are building FINRA's ontology as an extension of schema.org ontology.

As any big organization FINRA stores multiple copies of the same information. That makes it hard to retrieve important information. Our proprietary logic feeds connections between records to the Machine Learning model which creates clusters of related data elements. The quality of Enterprise Search results is now enhanced with SPARQL queries providing a better insight into millions of structured and unstructured data elements to our users. Additionally new analytics can be produced, leveraging the existing and new connections. Our community detection algorithm takes in account these new connections stored in Semantic Web to create more accurate results for cliques of 'bad actors'.

Fig. 1. Solution: FINRA Knowledge Graph

3   Conclusion

Combining Semantic Web [2], Enterprise Search, Text Mining [3], and Graph Analysis have proven to improve overall data quality and ease of data discovery and navigation. This new approach has significantly improved effectiveness of regulatory analysis.

References

1. FINRA (Financial Industry Regulatory Authority) http://www.finra.org
2. Discovering The Social Connections http://technology.finra.org/articles/discovering-social-connections.html
3. Unlocking Unstructured Data with Text Analysis http://technology.finra.org/articles/unlocking-unstructured-data-with-text-analysis.html