APPLYING KNOWLEDGE BASED AI TO MODERN DATA MANAGEMENT

Mani Keeran, CFA
Gi Kim, CFA
Preeti Sharma
What we are going to discuss…

- During last two decades, majority of information assets have been digitized in the financial sector.

- However, relational databases cannot answer critical business inquiries easily because its model lacks of semantics of business processes while focusing on efficient physical mechanism of data.

- Now new ways of storing data is being matured. RDF/OWL, one of most prominent Semantic Web technologies can be one of options for enabling datastore incorporating human intelligence.

- We will present how experimental Enterprise Knowledge Graph was created out of legacy RDB tables. It will include “inferences” features in querying. We consider this “machine readable” data format as potential backend of AI mechanism in Financial Information Management.
CHALLENGES IN FINANCIAL SECTORS

Is my database intelligent enough to answer my questionnaires?
Example Questionnaire #1: Am I allowed to make this contract with them?

- Finding a counterparty for a certain trade is restricted by various regulatory requirements and internal rules.

- This process is defined as a multi-step decision making tree, which is conducted by manual workflow or predefined codes.

Is this a valid trade?
Example Questionnaire #2: What is our exposure to Lehman Brothers?

• When Lehman Brother failed, many firms were not able to aggregate exposure because their systems were not ready for this.

• An human understands “exposure” should include various relationships between two parties, but a machine needs explicit code that implements the logic to integrate various sources.
Relational Database is not Intelligent because it lacks of Semantics in its physical data models

Conceptual
- Describes Semantics of a business domain
- Captures Business Requirements in terms of Entities and Relationships

Logical
- System Model captures details of the entities in terms of attributes
- Communicates Design Details

Physical
- Technology Model/Physical Design
- Implementation of Specific Use Case
- Business metadata is separated from physical implementation

Semantics Implied in Data Model
Semantics Embedded in Application Logic
KNOWLEDGE BASED AI

Enlightening Financial Information System Using Ontology, FIBO and Triple Store
Moving to AI-enabled Data

What is AI?

“AI is the study of how to make computers do things at which, at the moment, people are better”
– Elaine Rich, 1986

What people are better in using data than databases?

✓ People understand context of data.
✓ People can associate relevant information together even when they are not obtained at the same time, same place.
✓ People can leverage both of explicit and implicit relationships of information
Knowledge Based AI

Knowledge Base

Machine Learning

Deep Learning

Reasoning

Machine Readable Business Knowledge

Knowledge from Learning
From Database to Knowledge Base

**Database**

- Structured collection of records or data. (Wikipedia)
- Organized in such a way that a program can quickly select desired pieces. (Webopedia)

**Knowledge Base**

- Machine-readable resource for the dissemination of information. A Dynamic Resource that may itself have the capacity to learn, as part of an AI expert system. (Techtarget)
- Stores knowledge in a computer-readable form, usually for the purpose of having automated reasoning applied to them. (Wikipedia)
Ontology

- Data model that captures **knowledge** of a domain as a set of **concepts & relationships** between these concepts.
- Allows relationships to become visible and not dependent on code that embeds business rules.
- Semantic Industry Standards - Resource Descriptor Framework (RDF) and Web Ontology Language (OWL) govern the construction of Ontologies
Ontology Model as Knowledge Base

- **Legal Entity**
  - **Financial Institution**
    - Type
  - **Country**
    - Is domiciled in
  - **Issuer**
    - Has an issuer
  - **Common Stock**
    - Issues
    - Buy/sell securities

- **Brokerage Firm**
  - Type
  - Transitive
  - Relationship
Triple Store as Knowledge Base Storage

Triples is the simplest instance format of RDF/OWL expression.

A Triplestore is a database for triples. Data is indexed for the graph data structure.

OWL expression

Issuer issues Common Stock

Triple Instance

Subject  Predicate  Object
Apple Inc.  Issues  Apple Common Stock (APPL)

Fund A
US
is domiciled in

has position
issues

Apple Inc.

APPL
OWL – Language of Ontology

- **Web Ontology Language** from W3C
- Computational logic-based, used to express knowledge which can be understood and processed by machines
- Ontologies described using OWL can be processed by a **reasoner** which infers knowledge based on the asserted facts in the Ontology
- Has various constructs to create Ontologies and support reasoning e.g.
  - Class
  - EquivalentClass
  - TransitiveProperty
  - FunctionalProperty
  - Rdfs:label etc.
Financial Industry Business Ontology (FIBO)

What is FIBO?
- an industry initiative to define financial industry terms, definitions and synonyms using semantic web principles such as RDF/OWL and OMG modeling standards such as UML.
- a joint effort by the Object Management Group (OMG) and the Enterprise Data Management (EDM) Council.
- includes Ontologies for Financial Instruments, Business Entities, Market Data etc.
- driven by regulatory and industry requirement for data quality and transparency.

Benefits
- provides common reference model to integrate disparate technical systems and message formats within an organization and across industry.
- aims to bring transparency in the financial system.
- aims to improve regulatory reporting by providing clear meaning of data.

Each industry is coming up with its own standard ontologies:
- CDISC : Clinical Research
- HL7 : Health Care
- ACORD : Insurance
FIBO as Knowledge Base Language for Finance

• **Any type of Legacy data can be a part**
  - Both of Structured/unstructured data can be interpreted into Ontology Language using FIBO vocabularies.

• **FIBO brings the parts together**
  - FIBO pre-defines implicit/explicit business rules between objects in standardized manner.
  - Undiscovered relationship can be added as combination of standard vocabularies using query grammar.

• **Inference: Make the sum of the parts whole**
  - Business rules are included as a part of the data model.
  - A query will leverage business rules defined in FIBO
EKG; ENTERPRISE KNOWLEDGE GRAPH
Proof of Concept
Introducing EKG – Enterprise Knowledge Graph

Enterprise Data Platform

Operational Zone
- Normalized Model
- Support Operational Reporting
- Data Distribution

Analytical Zone
- Dimensional Model
- Support Analytical Reporting
- Connected to BI Layer

EKG Zone
- Graph Data Model
- Support Semantic Query
- Extended data service with Inference feature

Integration Zone

Data Producers, Data Lake
Logical View of PoC

- Use legacy tables as sources.
- Translate the legacy data into triples and store them.
- Do queries with Business questions.
PoC Technology

Key Components

**Resource Mapper**
- Developed as an in-house Java code leveraging external libraries for triple store communication.

**Mapping Files**
- CSV format files that contain mapping rules between tables and triples.

**Triple Store**
- Implemented on AllegroGraph Free Edition (Oracle 12c was also tested)
  
**FIBO Extension**
- Relevant vocabularies picked from published FIBO version
- Additional vocabularies for extension

---

**Input:**
- 10 tables w/ +250k rows

**File feed:**
- Define mappings using Excel
- Import FIBO RDF/OWL files

**Legacy Database**

**Linux Server**
- Data Files
- Resource Mapper
- Mapping Files
- API over HTTP

**Triple Store**
- Results:
  - +1.4 million triples

* AllegroGraph 6.3 or Oracle 12c Spatial and Graph
# Rule-based Triple translation

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>FLAG</th>
<th>PREDICATE</th>
<th>OBJECT</th>
<th>CONTEXT</th>
<th>BASE_URI_S</th>
<th>BASE_URI_O</th>
</tr>
</thead>
<tbody>
<tr>
<td>bbcompid</td>
<td>Literal</td>
<td><a href="http://www.w3.org/2000/01/rdf-schema#label">http://www.w3.org/2000/01/rdf-schema#label</a></td>
<td><a href="http://franklintempleton.com/fibopoc/BloombergIdentifier#">http://franklintempleton.com/fibopoc/BloombergIdentifier#</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Column**

- **SUBJECT**: A column which has the name of Subject resource
- **FLAG**: Defines a type of a Object in a statement. URI (Bare URI), Literal (Value), Resource
- **PREDICATE**: A URI of a Predicate
- **OBJECT**: A URI, a column which have a value of Literal or a name of Object resource
- **BASE_URI_SUBJECT**: Base URI for Subject a resource
- **BASE_URI_OBJECT**: Base URI for Object a resource
- **CONTEXT**: URI of graph

**Sample value**

- `<name of a column>`
Technical Environment

1. Triple Store: **Allegrograph** – an RDF graph database

2. Visual Discovery Tool: **Gruff** – a triple store browser

3. Query Language: **SPARQL** – RDF query language

```sparql
SELECT ?contract
WHERE
{       ?contract <hasCounterparty> ?Counterparty.

?Counterparty.<isLegalEntity> ?lege }
```
USE CASE DEMO
Use Case #1: Am I allowed to make this contract with them?

Is this a valid trade?
Use Case #1: Sample Process

Account is considered a US person by CFTC, has not adhered to the Dodd Frank Protocols 1 & 2, not allowed to execute bi laterally.

Is the account not a US person per CFTC?

Account is not considered a US person by CFTC, has not adhered to the Dodd Frank Protocols 1 & 2, allowed to execute bi laterally with non-US counterparty.

Account is not compliant with EMIR protocol... EU CPs are ineligible.

Counterparty not Selectable

Counterparty not Selectable

Counterparty not Selectable