ontotext

Introduction

The Financial Industry Business Ontology (FIBO) is a standard that is being developed and published by the Enterprise Data Management Council that attempts to capture business domain knowledge using sophisticated knowledge representation techniques and linked open data technologies. It benefits from years of work by leading ontologists and technologists and the participation of a broad range of business subject matter experts. Quoting the EDMC web site,

> The Financial Industry Business Ontology (FIBO) defines the sets of things that are of interest in financial business applications and the ways that those things can relate to one another. In this way, FIBO can give meaning to any data (e.g., spreadsheets, relational databases, XML documents) that describe the business of finance.

The stated breadth of its scope and the ongoing efforts at refining its content and broadening its scope present an inherent challenge to its use. Unless one is facing a greenfield, FIBO must be seen in the context of one's specific application domains and existing businesses and their supporting information management systems. This is a nontrivial task. Even if you have detailed models of your business, data and the technology artifacts that support the business, ontology alignment is still an active area of research with few, if any, off the shelf tools to support such an effort.

There is an old shaggy dog story that features a chicken that is repeatedly sold over the course of a day, with each purchaser paying a higher price. The final purchaser consumes the chicken and encounters the original owner the next day to whom he reports that the chicken tasted good, but was nothing special. The original owner points out that the chicken was for selling, not for eating.

For FIBO to be for eating instead of selling, its conceptualizations must conform to the concepts and constraints specified in legislation, regulation and a firm's internal policies and procedures. These are all specified by humans in human languages, AKA natural language, notorious for both its expressiveness and ambiguity. Fortunately, one of the outstanding

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

features of FIBO is it's extensive English language documentation. There are more than 4,000 statements using the skos:definition predicate, more than 800 statements that identify the source of the definition and many occurrences of explanatory and editorial notes. The relationships between the words in these triples and the words in legislation, regulation, policies and procedures are the key to placing FIBO in a context.

Historical Context

FIBO's roots may be found in the Semantic Repository, which was a model specified using OWL constructs expressed through UML tooling, developed by Mike Bennett of the EDMC and first appearing in 2008. From the start, it was useful, the tabular representation being used in several firms that I'm aware of. From these humble beginnings grew the OWL representation. A fair amount of effort was expended to derive reusable 'operational ontology' RDF/OWL content from this, targeted at data domain problems such as classifying Interest Rate Swaps. The remaining upper level or 'conceptual' ontology, targeted at business domain problems, was used as the basis to inform this work and each section subsequently deprecated as the work was completed at the operational level. Since then the focus has been on the data domain as described in the EDMC's current description of the product.

The OWL representation was quickly followed by a SKOS vocabulary, which is derived directly from the ontology. These have subsequently been followed by CSV and Microsoft™ Excel format data dictionaries and a purpose built data model FIB-DM, which is a complete model transformation of the FIBO into a Conceptual Data Model. Since January of 2020 this work has progressed under an open community process, which follows rigorous and well-defined rules and principles and is organized via two groups: the FIBO Steering Group and the FIBO Community Group.

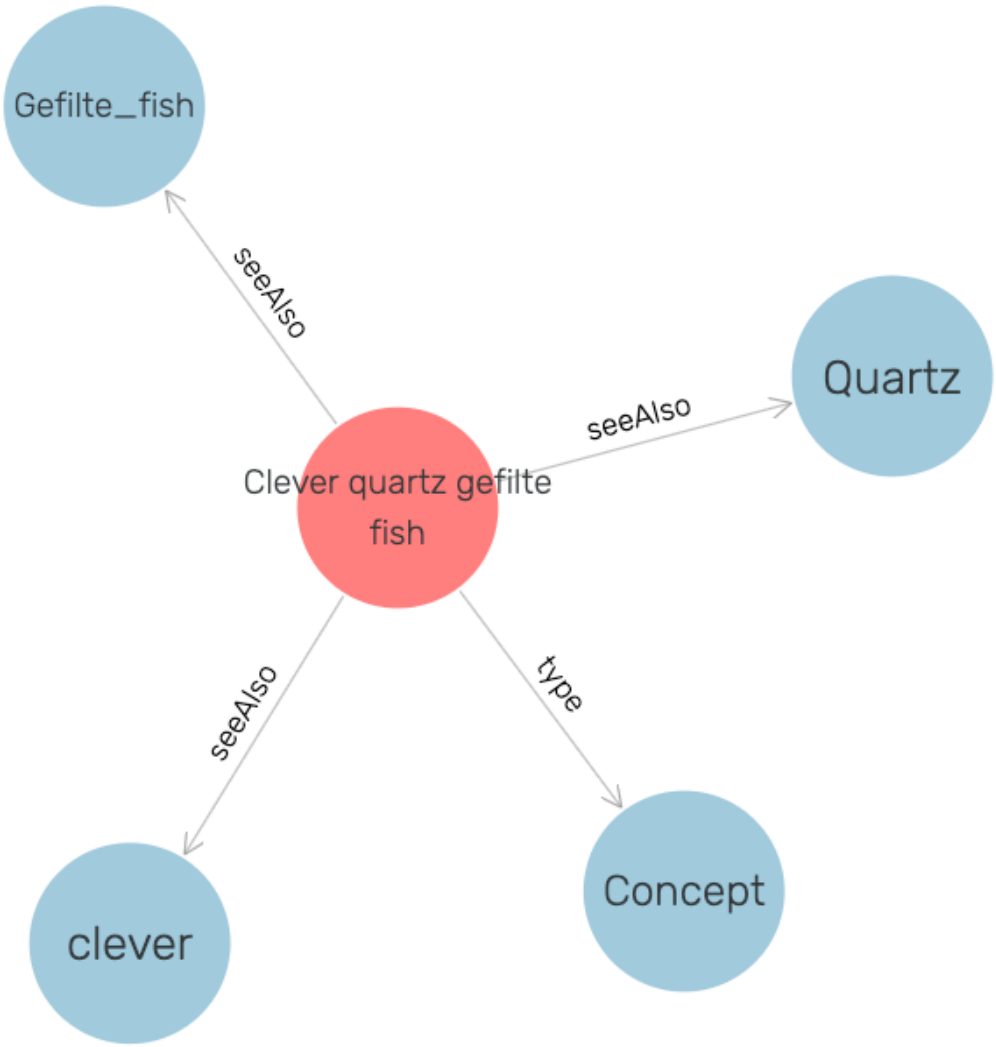
One of the important lessons from the transition from UML to OWL was that there is nothing inherent in either formalism that makes statements meaningful. One can express nonsense in a human language or in a formal language as will be demonstrated in the next section.

Precisely Specified Nonsense

The use of semantic web technologies, natural language processing tools and knowledge graph architectures does not insure the meaningfulness of content expressed with those

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

tools. By way of example, consider the concept *Clever Quartz Gefilte fish*. The gloss of this nonsensical concept is readily apparent from the words. The meaning of the words *clever* and *quartz* are readily apparent. *Gefilte fish* refers to an item of Ashkenazi cuisine. You're now all subject matter experts in *Clever Quartz Gefilte fish*. Below are the RDF/SKOS representation of the concept and a graphical representation.



Clever Quartz Gefilte fish graph

```
@prefix cam: <https://dictionary.cambridge.org/us/dictionary/english/>
```

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

```
.
@prefix fic: <https://www.industrialsemantics.com/blogpost/fic#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix wikip: <https://en.wikipedia.org/wiki/> .
fic:CleverQuartzGefilte_fish
  rdf:type skos:Concept ;
  rdfs:label "Clever quartz gefilte fish" ;
  rdfs:seeAlso cam:clever ;
  rdfs:seeAlso wikip:Gefilte_fish ;
  rdfs:seeAlso wikip:Quartz ;
skos:definition "Gefilte fish possessed of the qualities quartz and
cleverness." ;
skos:prefLabel "Clever quartz gefilte fish" .
```

Both the RDF and the graphical representation of the RDF represent the good taste of meaning without all the filling connotation or denotation. The same is all too often achieved in glossaries that contain elements such as Crack Spread Future: “An entry in the crack spread futures table”.

Discovering Context With Semantic Tooling

The cognitive space defined by the FIBO mission statement is huge. It encompasses large parts of the Consolidated Federal Register, several sections of the US Code, not to mention parts of fifty distinct state’s legislation and regulation. If the regulations of the European Central Bank, the Bank of England, the Bank of Japan, the Monetary Authority of Singapore and the Central Bank of Eswatini are included, the size of the domain exceeds what one person, or group of people can manage in their heads.

There is a variety of semantic tooling that can be brought to bear on the problem including traditional pattern based natural language processing and a number of different stochastic approaches. A good overview of traditional pattern based and ontology based techniques can be found in *Law and the Semantic Web — Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications* published by Springer. There are a number of tools that support

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

these techniques such as GATE from the University of Sheffield. But even with these tools acquiring the documents, normalizing and consistently processing parts of speech, lemmatization, etc. is a nontrivial problem.

There are commercial tools focused on narrower parts of the problem such as Quarule, which uses structured English and Rulelog to automate controls that are manually derived from the regulation and legislation that constrains a business.

At the other end of the tooling spectrum is the QuantGov effort, which is an open-source platform, a collection of unique, relational data, and a toolbox for researchers and policymakers alike. QuantGov's scope is infinitesimal compared to FIBO's, but it does provide detailed analytics over that scope without recourse to actual language understanding. It addresses the problem of quantifying large amounts of policy text for research and comprehension by using machine learning and natural language processing.

The Momma Bear Solution for Comprehending FIBO

Given the scope of FIBO and the immensity of the task of aligning it with constraints of your business, what is the middle path? GraphDB offers such a tool with its Similarity Indexes implemented as a plugin. This functionality is outside the scope of the standards compliant portion of the tool. The similarity plugin integrates the Semantic Vectors library and the underlying Random Indexing algorithm.

The algorithm uses a tokenizer to translate literals to sequences of words (terms) and to represent them into a vector space model representing their abstract meaning. A distinctive feature of the algorithm is the dimensionality reduction approach based on Random Projection, where the initial vector state is generated randomly. With the indexing of each literal, the term vectors are adjusted based on the contextual words. This approach makes the algorithm highly scalable for very large text corpora of documents, and research papers have proven that its efficiency is comparable to more sound dimensionality reduction algorithms like singular value decomposition.

As the literals are associated with subjects, this provides a mechanism for finding related subjects based upon the closeness of their associated vectors.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

There are two types of similarity indexes, text based similarity indexes and predication-based semantic indexes. Both are highly configurable. For our purposes we will only be using text based indexes. However, both are well documented with examples you can download here.

Terms in FIBO and Other Glossaries

If we are not going to analyze FIBO in the context of the world's totality of financial regulation and legislation, then what can be practicably done? As a first step, we will take a small set of words and a variety of open source glossaries including FIBO and see how the similarity works across the set of glossaries. The words in our list are, account, client, counterparty, instrument, party and security.

The set of glossaries includes the European Central Bank glossary, the Financial Control Authority Handbook glossary, the FIBO SKOS vocabulary, the International Swap Dealers Association glossary, the glossary associated with the ISO 20022 standard, the Office of Financial Responsibility glossary, and the Security and Exchange Commission glossary. Only FIBO is available in RDF. The other glossaries were transformed into SKOS/RDF glossaries from publicly available web resources. Let's look at some of the higher scoring results.

term	similarTerm	score	content	Glossary
client	Client Identifier	1	an identifier for a client	FIBO-V
counterparty	has counterparty	1	identifies a counterparty to a contract	FIBO-V
instrument	Non-investment grade debt	1	Instruments	OFR
transaction	subTransaction	1	decomposition of a BusinessTransaction into a number of sub transactions which are BusinessTransactions in their own right.	ISO

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

term	similarTerm	score	content	Glossary
client	Broker	0.813165	An individual who acts as an intermediary between a buyer and seller, usually charging a commission to execute trades. Brokers are required to seek the best execution of trades they make for clients, and if they recommend investments to clients, those investments must be suitable for the client.	SEC
party	Tender Offer	0.789894	A tender offer is typically an active and widespread solicitation by a company or third party (often called the “bidder” or “offeror”) to purchase a substantial percentage of the company’s securities. Bidders may conduct tender offers to acquire equity (common stock) in a particular company or debt issued by the company. A tender offer where the company seeks to acquire its own securities is often referred to as an issuer tender offer. A tender offer where a third party seeks to acquire another company’s securities is referred to as a third party tender offer.	SEC

As you can see, the similarity scores do not imply semantic equivalence. For example, a client and a client identifier are distinct, but highly related things. The goal here is the quest for clues as opposed to the quest for truth.

This table, and the worksheet from which it is drawn, are not directly produced from GraphDB. They were produced by using the SPARQL query that the similarity indexes provide and make available to the user. A small, to the point of being trivial, Python script was used to iterate across the list of terms using the generated query and executing the query against GraphDB’s SPARQL endpoint.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

FIBO Versus the World

WIth the builtin capability of similarity indexes and the Python script in place, the next step is to compare all of the terms defined in the FIBO vocabulary against the other glossaries. We will again look at some of the higher scoring terms.

term	similarTerm	score	content	Glossary
Personal Consumption Expenditures	PCE	1	personal consumption expenditure	ECB
Multilateral Trading Facility	MTF	1	a multilateral trading facility.	FCA
Corporate Bond	Bonds, Corporate	1	Corporate bonds are	SEC
Consumer Price Index	CPI	1	the Consumer Prices Index.	FCA
Urban Consumer Price Index	CPI	1	the Consumer Prices Index.	FCA
Over-the-counter (OTC) Transaction	Over The Counter (OTC) transaction	1	Over The Counter (OTC) transaction	ISDA
Bank For International Settlements	BIS	1	Bank for International Settlements	ECB
Cross-currency Interest Rate Swap	Cross currency interest rate swap	1	Cross currency interest rate swap	ISDA
Instrument Of Incorporation	Non-investment grade debt	1	Instruments	OFR
Instrumentality	Non-investment grade debt	1	Instruments	OFR
International Securities Identification Number	ISIN	1	International Securities Identification Number	ECB

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

term	similarTerm	score	content	Glossary
Good	Bagehot’s Dictum	0.991760607	Theory of Walter Bagehot, a 19th century writer and banker, who proposed central banks should lend freely and often against good collateral and at high interest rates to quell a financial panic.	OFR
Auditor	Auditor opinion	0.986605975	Statements auditors include in their reports on company finances. Auditors issue adverse opinions when they have concerns that the statements have not been prepared along accepted principles or that the data supporting the statements have been misrepresented. They issue clean opinions when they find no significant exceptions to accepted accounting practices and disclosure requirements. Auditors issue opinions with an	OFR

As you can see, we are once again looking at the similarity of terms and not the semantic equivalence of terms. For example, both instrumentality and instruments, and instrument of incorporation and instruments are not good semantic matches. Both of these are the result of the lemmatization of the word *instrument*. Fortunately, lemmatization can be adjusted through the GraphDB similarity index interface.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

FIBO in Your World

How can these techniques be used to put FIBO in the context of your world? Most firms involved in the financial services industry must provide, as part of their compliance with regulation, a data governance process, many of which use tools such as Solidatus or Collibra, to name just two of many available tools. These tools provide mechanisms to manage glossaries of business terms linked to the IT resources that implement them. The following steps set up the capability demonstrated above.

1. Export your glossary in one of the Excel-based formats.
2. Create a GraphDB repository to hold your glossary.
3. Import your glossary using the OntoRefine tool. Instructions may be found here.
4. Create a GraphDB repository to hold the FIBO glossary, which can be found here.
5. Create a text similarity index for the FIBO glossary.
6. Iterate through the terms in your glossary using them as search terms against the FIBO similarity indexes.

Mechanizing the Process

The final step is to mechanize this process so that you can review the proposed similarities and review them with your team to select the ones that make sense. The best way to do this is by executing the similarity queries using a script in which you iteratively substitute the terms from your glossary for the search term. Alternatively, one can use SPARQL federation to fetch the terms to be passed as parameters for similarity search. For the terms that score at or above your desired cutoff, create triples in a separate named graph or distinct repository that contain the information in the following tuple:

```
Your Repository, Your term, Similarity Score, FIBO Repository, FIBO term
```

Sort these tuples in descending order by the similarity score. This becomes the burndown list for your review team. Your review team can then annotate your glossary with links to the FIBO terms. As the FIBO glossary is linked to the FIBO ontology, you can exploit the associations with the corresponding ontology elements to enhance and enrich your data model.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

This is the second post of a series that will demonstrate how graph database engines and semantic technology can be used to deal with ontologies and data in the financial services sector.

Ontotext's GraphDB
Give it a try today!
Download Now

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com