

会計ドメインにおける RDF モデルの構築と Linked Data との連携

Constructing Accounting RDF model with Linked Data

鈴木 健太
Kenta SUZUKI

山口 高平
Takahira YAMAGUCHI

慶應義塾大学
Keio University

This paper discusses how to develop a rdf model of accounting domain. Accounting domain ontology is based on XBRL data. According to restructuring the XBRL model to RDF based model, the ontology is able to connect as linked data.

1. はじめに

現在, XBRL(eXtensible Business Reporting Language)が多くの企業で注目されている。XBRLとは, XMLを財務情報の交換に適用したものであり, XBRL で表現される財務情報は各勘定科目の金額情報だけでなく, その勘定科目の定義も含まれている。HTML と PDF によって開示されていた財務情報が, XBRL で開示されることにより, 情報の再利用性が向上したのである。

しかしながら, XBRL で厳密に財務情報の定義をされているにもかかわらず, 各勘定科目の数値の抽出のみを利用するケースが多い。また, 財務情報を詳細に分析するアプリケーションにおいても, アプリケーションごとに会計規則のルールを独自に実装することによって対応している場合がほとんどである。

XBRL の勘定科目の定義体系を活用すれば, 財務情報よりシームレスに扱うことができる。会計規則は毎年変更され, 勘定科目のスキーマも変更しなければならない。財務諸表のルールをより柔軟に管理するため, 会計ドメインオントロジーとしてこれらの関連性をモデル化し, 財務情報を活用するための基盤を構築する。また, 会計ドメインオントロジーに基づいた RDF を利用することで, Linked Data との連携可能性を高めることができる。

このような現状を踏まえ, 本稿では会計ドメインにおけるオントロジーの構築を行い, セマンティック Web のプラットフォーム上における会計データの効率的な流通方法について検討する。

2. XBRL

XBRL の基礎的な仕組みについて説明する。

2.1 XBRL の概要

XBRL は財務諸表の電子的な報告に XML を応用したものである。XBRL の導入により, 各種財務報告用の情報の作成・流通・利用が効率的に行われることが期待されている。XBRL が導入される前は, PDF 及び HTML の形式で財務報告が行われており, 構造化されたデータを取得するには高いコストが必要であった。XBRL 標準で定義されたデータは, 財務情報を利用する企業間で効率的に流通させることが可能となる。

財務情報は, 年度や業種ごとに, 項目や計算式が異なるという特徴がある。従来の紙ベースの報告では, 各データを一元的に管理しづらく, 報告書を作成するコストも高くなるという問題があった。XBRL では, 財務情報の作成・流通・文責・変換など

に適した XML の形式でデータを記述する。

日本においては, 2008 年 3 月より金融庁が XBRL を採用した新 EDINET^{*1}を稼働させた。EDINET とは有価証券報告書等の開示書類について, インターネット上においても閲覧を可能とするシステムのことである。

XBRL は, 会計, 財務関係のインターネットによる世界共通の言語という位置づけになる。そのために, 世界共通ルールとしての XBRL Specification が定められている。

しかし, 会計基準はまだ全世界共通とはなっていない。2010 年 3 月期の年度から日本においても, 一定の要件を満たす企業に対しては IFRS(International Financial Reporting Standards)に基づく連結財務諸表の作成を容認したが, 全企業が同一の基準で財務情報を開示することは未だ実現してはいない。決算書のスタイル, 表現は企業, 業種によって微妙に異なる。したがって, 基本的なルールは仕様書に基づきながら, 各国の会計基準や業界標準にあわせたテンプレートが必要になる。このテンプレートのことを, XBRL ではタクソミと呼んでいる。そしてこのタクソミに基づき, 実際の数字を記述したものをインスタンス文書と呼ぶ。

XBRL は大きく分けて2つの要素から構成される。各企業の財務情報が記述されたインスタンス文書, 項目のタグ名(語彙)を定義したタクソミ文書である。インスタンス文書には, タクソミ文書はタクソミスキーマとリンクベースによって構成されている。

2.2 タクソミスキーマ

XBRL のスキーマ記述には XML Schema が利用されている。タクソミスキーマでは, XBRL のもつデータの定義を記述している。XBRL においては名前空間及び各リンクベースファイルのインポートの機能を利用し, 財務情報定義を導入している。タクソミスキーマは3つの階層をなしており, GAAP タクソミ, 産業別タクソミ, 企業別タクソミによって構成される。各タクソミは, 5 種類のリンクベースをインポートし, 当該 XBRL インスタンスの各勘定科目や文書情報の定義に関する情報を提供する。

GAAP タクソミ及び産業別タクソミについては, 会計規則に基づいたものを金融庁が毎年 1 回発表している。企業別タクソミは各企業独自の勘定科目を定義するものであり, 開示書類ごとに個別に含まれている。

2.3 リンクベース

リンクベースには以下の 5 つがある。

・表示リンク

連絡先: 鈴木健太, 山口高平 慶應義塾大学理工学部
〒223-8522 神奈川県横浜市港北区日吉 3-14-1
E-mail: {k_suzuki,yamaguti}@ae.keio.ac.jp

*1 <http://info.edinet-fsa.go.jp/>

- ・計算リンク
- ・定義リンク
- ・名称リンク
- ・参照リンク

表示リンクでは各勘定科目の表示順を定義する。計算リンクは各勘定科目間での重みつき加算式を定義している。定義リンクは項目間の同義性や出現規則などが示されている。名称リンクは様々な言語における各勘定科目の呼称が定義されている。参照リンクは各勘定科目の会計概念定義の根拠となっている文献へのリンクが示されている。

タクソノミスキーマに定義された項目に対して、各項目間の関係や、各項目に対する追加情報などを、XLink[DeRose, S 2001]の外部リンク機能を利用して定義したものがリンクベース

である。具体的には、各勘定科目の表示順序や、計算方法、勘定科目として表示される値のラベルの定義などを行う。これらの定義は、タクソノミスキーマとは別のファイル(=リンクベースファイル)として作成される。XBRL ではこれらのリンク定義を個別のリンクベースとして、ファイルを分けて作成することができる。

XBRL を使って財務情報を作成する場合、各国の会計制度に対応したタクソノミ文書に加え、業種などで共通化されたタクソノミ文書を利用する。さらに、自社独自の情報については、自社タクソノミ文書を作成する。こうして作成されたタクソノミ文書を元にインスタンス文書を作成する。

3. 関連研究

関連研究としては、Publishing XBRL as Linked Open Data[R. Garciaら 2009]があげられ、EDGARを対象としてXBRLデータをRDFに変換している。Publishing XBRL as Linked Open DataではXBRLからRDFへの変換をXMLの構造をそのまま置き換えることによって実現しているため、XBRLの記述形式に近いRDFを出力している。しかし、このデータは財務報告データの一部のみをマッピングしたものであり、XLinkによって関連付けられているXBRLの全データを変換するには至っていない。

Representing Financial Reports on the Semantic Web [J. Baoら 2010]においては、XBRLを意味的に解釈し、XBRLのスキーマをOWLに変換を行っている。

Linked Dataとしての利用を見越した場合、使いやすいRDFモデルに基づいたデータの構築が不可欠であると考えられる。したがって、本稿ではXBRKの意味モデルに則った会計ドメインオントロジーをOWLで構築する。それに基づいた形でRDFの出力を試みる。これにより、より正確なデータモデルをXBRLから取得できる。XBRLの意味を正確に集めるためには、オントロジーの制約とルールをモデリングしなければならない。機械可読な結果を得るために、OWLを利用してオントロジーの構築を行う。また、XBRL構造に含まれている意味を失うことなく、RDFデータを出力することを試みる。XBRLに含まれているデータを一部でも失ってしまうと、データの不整合性を引き起こす可能性があるためである。

4. 提案システム

本稿では、金融庁EDINETの公開しているXBRLを対象として、RDFへの変換を行う。また、関連するタクソノミスキーマからOWLオントロジーを構築する。システムの実装にはJavaを使用し、データベースにはTDB^{*2}を利用している。

4.1 システムの全体像

本システムは以下の流れでRDFを生成する。

1. XBRL ファイルの取得
2. XBRL インスタンスから発見可能なタクソノミ集合 (DTS: Discoverable Taxonomy Set) の識別
3. 各ファイルのパーズ
4. RDF の出力
5. データベースへの格納

DTS とは、XBRL インスタンスから参照されている全てのXMLスキーマファイル及びリンクベースファイルのことを指す。XBRLをパーズするためには、XBRL インスタンスだけではなく、すべてのDTSをパーズをパーズし、各種定義情報を取得しなければならない。DTSを発見する方法は、拡張可能な事業報告言語(XBRL) 2.1[JISX7206]の第3章において、定義されている。

4.2 RDF モデル

モデルの出力に関しては、以下の環境を用いた。

- ・Java SE: version 1.6
- ・Jena 2.6.2

XBRL インスタンスから得られた情報をRDFとして出力する。会計ドメインオントロジーとの整合性を担保するため、以下の情報をアウトプットする。

- ・開示した会社
- ・対象となる書類
- ・書類が開示された日
- ・当該書類の対象期間または日付
- ・対象となる勘定科目
- ・当該勘定科目の金額
- ・金額の単位

上記の情報はすべてXBRLパーサで抽出したものを利用できる。図1に出力するRDFのモデルを提示する。各名前空間のURIは表1のとおりである。

表1 名前空間一覧

| 名前空間 | URI |
|------------------|---|
| xbrlont_instance | http://www.yamaguti.comp.ae.keio.ac.jp/xbrl_ontology/instance |
| xbrlont_property | http://www.yamaguti.comp.ae.keio.ac.jp/xbrl_ontology/property |
| xbrlont_class | http://www.yamaguti.comp.ae.keio.ac.jp/xbrl_ontology/class |
| foaf | http://xmlns.com/foaf/0.1/ |
| rdf | http://www.w3.org/1999/02/22-rdf-syntax-ns# |
| xsd | http://www.w3.org/2001/XMLSchema |
| iso4217 | http://www.xbrl.org/2003/iso4217 |

上記以外の名前空間(jpfr-t-cte等)は、EDINETで定義されている名前空間に順ずるものとする。

上記のモデルは、会社Aを大元となるリソースとして定義している。会社Aには日本語名称が入るものとする。

*2 <http://openjena.org/TDB>

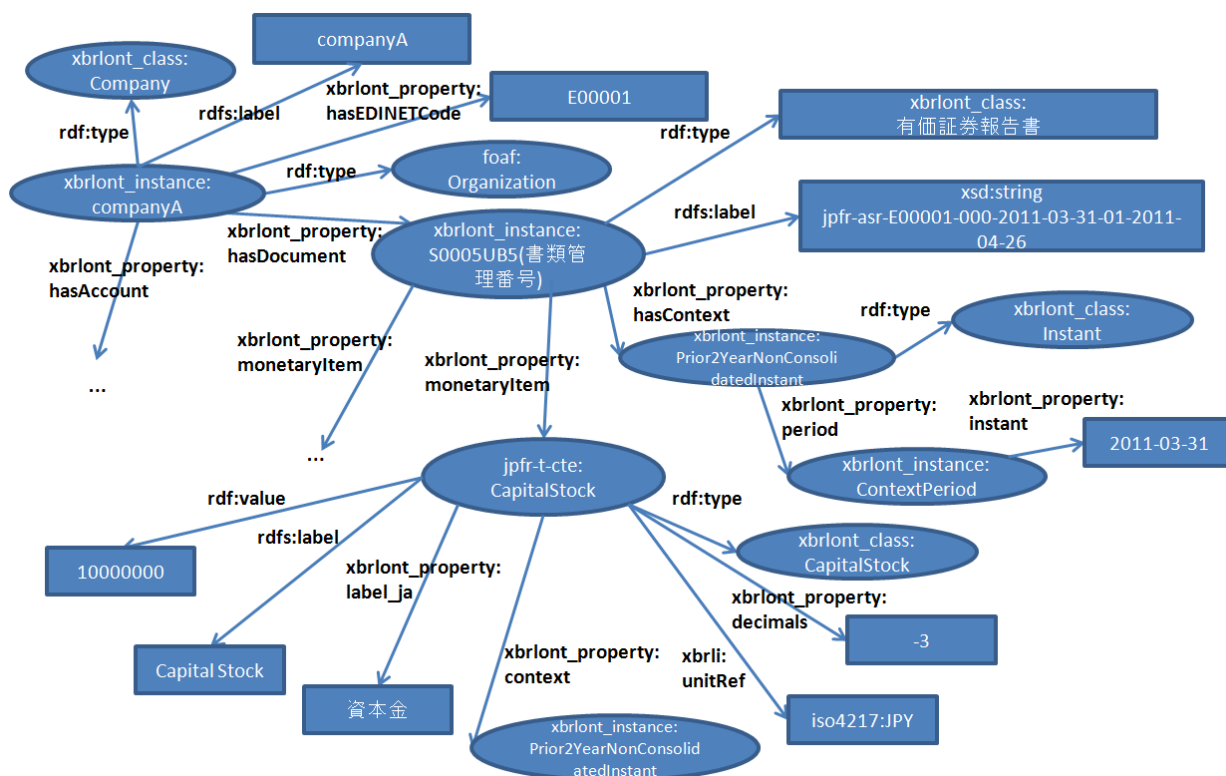


図 1: RDF モデル

hasEDINETCode 属性によって、リテラル値である E99001 と結び付けられている。これは EDINET コードを示しており、会社の存在をユニークなものとして識別する。また、会社Aは rdf:type 属性によって foaf:Organization 及び xbrlont_class:Company に結び付けられる。foaf:Organization は組織形態を持つグループを表すクラスである。この3項関係により、会社Aはある組織形態を持つグループであるということが示されている。

次に、hasAccount プロパティ以下の関係について説明する。会社Aは複数の hasAccount プロパティを持ち、目的語に各勘定科目を持つ。目的語として提示されている勘定科目は、業種ごとに異なる名前空間に属する。上記の例では、jpfr-t-cte すなわち一般商工業その他の分類に属するということを示している。各勘定科目は後述する勘定科目の owl オントロジーに基づき、各種制約を持つ。各勘定科目は以下のプロパティを持つ。

•xbrlont_property:hasAmount

値を表すプロパティである。目的語には空白ノードをもつ。この空白ノードはプロパティとして rdf:value 及び nsIns:hasUnit をもつ。rdf:value は主語リソースの「値」を示すプロパティであり、その目的語には金額がリテラル値で記述される。nsIns:hasUnit は目的語に通貨単位情報をもつ。上記の例では、iso4217:JPY すなわち日本円を示す。ユニット属性及び金額の値が対になって、「10,000,000 円」という金額を示している。

•xbrlont_property:hasInstant

対象となる日を表すプロパティである。目的語には xsd:date 型のリテラルをもつ。例では、2010年3月31日時点での現金及び預金であるということを示している。リテラルと鳴る日付は、当該勘定科目の ContextRef 要素によって参照されたコンテキスト要素における period 要素の中の Instant で指定された日付である。貸借対照表に含まれる科目はストックの要素であり、その時点での価値を持つ。したがって、hasInstant プロパティを有する。

•xbrlont_property:hasStartDate, xbrlont_property:hasEndDate
例には存在していないが、これらも勘定科目のプロパティに含まれる。hasStartDate は対象となる期間の期首日を示し、hasEndDate は期末日を示す。これらは hasInstant と同様に xsd:date 型のリテラルを持ち、目的語に日付を示す値が示される。したがって、2つのプロパティが組み合わさって期間を表すことになる。損益計算書の各勘定科目はフローの要素であり、特定期間での価値変動を表すものである。したがって、hasStartDate 及び hasEndDate プロパティを有する。

以上が勘定科目を示すリソースがもつプロパティである。hasAccount プロパティは値の設定された勘定科目に対して設定される。全ての勘定科目は上記に基づいてプロパティ及び付随する値を有する。

4.3 会計ドメインオントロジー

出力する OWL オントロジーのモデルについて説明する。会計ドメインオントロジーは個々の企業の金額情報のみを示すのではなく、勘定科目の体系を表す知識ベースである。勘定科目は各々異なった特性を持っており、それらはすべてリンクベースの中に定義されている。

ここで構築する会計ドメインオントロジーは各リンクベース文書に基づいて構成される。ドメインオントロジー中のクラスは各勘定科目の階層関係を示し、制約を与え、推論モデルを築くことを可能にする。勘定科目は各々異なった特性を持っており、それらはすべてリンクベースの中に定義されている。財務データの解釈には財務データを解釈するためのオントロジーが不可欠である。

会計ドメインオントロジーは以下の構造からなる。

- 貸借対照表に含まれる勘定科目に関するクラス階層

- ・損益計算書に含まれる勘定科目に関するクラス階層
- ・各勘定科目に関する制約
- ・`xbrlont_class:Company` クラスに対する制約

貸借対照表及び損益計算書に含まれる勘定科目のクラス階層は各表の記述形式に基づいている。全ての勘定科目に関する上位クラスは `xbrlont_class:Account` と定義した。貸借対照表に含まれる勘定科目は `xbrlont_class:BalanceSheetsAbstract` のサブクラスとし、損益計算書に含まれる勘定科目は `xbrlont_class:StatementsOfIncomeAbstract` のサブクラスとする。各勘定科目のクラス名は EDINET タクソノミで定義されているエレメント名をそのまま用いている。

次に `Company` クラスに対する制約を説明する。`Company` クラスは `hasEDINETCode` プロパティを 1 つ持ち、`hasAccount` プロパティを 1 つ以上持たなければならない。

ここではオントロジーの一例をあげた。会計ドメインオントロジーは、一般的な勘定科目及び各企業の業種の勘定科目の定義を組み合わせて構築されている。

しかし、企業別に独自に定義されたタクソノミに関しては、会計ドメインオントロジーに静的に組み入れられない。なぜなら、企業別タクソノミは各 `XBRL` インスタンスに含まれており、各書類が開示されなければルールを解釈することができないからである。もし企業別タクソノミを会計ドメインオントロジーに取り入れる必要がある場合には、各書類の分析と同時にオントロジーを拡張する必要がある。

5. 実験・評価

ここで、`Linked Data` との連携を行う。`xbrlont_class:Company` がタイプであるインスタンスを中心に行う。`Company` クラスの全インスタンスは会社に関する情報をもつオントロジーと関連付けることができる。ここでは、日本語 `Wikipedia` オントロジー[玉川ら 10]の会社情報と連動させた `SPARQL` によるクエリの一例を示す。

(例)トヨタ自動車の 2010 年度期末連結純資産及び設立年度

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wikiont_class: <http://www.yamaguti.comp.ae.keio.ac.jp/wikipedia_ontology/class/>
PREFIX wikiont_property: <http://www.yamaguti.comp.ae.keio.ac.jp/wikipedia_ontology/property/>
PREFIX wikiont_instance: <http://www.yamaguti.comp.ae.keio.ac.jp/wikipedia_ontology/instance/>
PREFIX xbrlont_class: <http://www.yamaguti.comp.ae.keio.ac.jp/xbrl_ontology/class/>
PREFIX xbrlont_property: <http://www.yamaguti.comp.ae.keio.ac.jp/xbrl_ontology/property/>
PREFIX xbrlont_instance: <http://www.yamaguti.comp.ae.keio.ac.jp/xbrl_ontology/instance/>

SELECT ?literal ?value
WHERE
{
  ?wikiont_instance wikiont_property:設立 ?literal;
    rdfs:label "トヨタ自動車".
  ?xbrlont_instance rdf:type xbrlont_class:Company ;
    rdfs:label "トヨタ自動車";
    xbrlont_property:hasDocument ?x.
  ?x xbrlont_property:monetaryItem ?y;
    xbrlont_property:hasContext ?context.
  ?context rdf:type xbrlont_class:Instant;
    xbrlont_property:period ?period.
  ?period xbrlont_property:instant "2010-03-31T00:00:00Z"^^xsd:dateTime.
  ?y rdf:value ?value;
    rdfs:label 'NetAssets';
    xbrlont_property:context xbrlont_instance:CurrentYearConsolidatedInstant.
}

```

このように複数のオントロジー間にまたがったクエリングが可能になる。

しかし、現在は全体のランキング付などを行うことができない例えば、「2010 年度期末連結純資産の額が上位 10 位以内に入る企業」というように、値の計算が必要なクエリを投げると著しくパフォーマンスが低下する。2010 年度期末決算で開示された全ての有価証券報告書について当期連結純資産の値を抽出し、これらの値のソートが必要になるため、計算量が膨大になってしまうためである。統計的な処理を行う必要のあるクエリへの対応は今後の課題である。

6. おわりに

本稿では、EDINET で公開されている `XBRL` を `RDF` に変換し、`Linked Data` として公開するための基礎的な構築を行った。`XBRL` の構造から会計分析上で必要な概念を考慮して会計ドメインオントロジーの構築を行い、その上で `RDF` を生成した。これにより、`XBRL` のデータの流動性を高めることができた。

今後の方針としては、`RESTful` な `Web` サービスとして `RDF` データを公開し、既存の `Linked Data` との連携例を増やすことが挙げられる。`XBRL` における企業のエンティティを、企業関連の `Linked Data` と結び付けることができれば、企業分析や新たなサービスの基盤の展開が容易になると考えられる。サービス開発者にとって使いやすい `API` の設計を行うため、`XBRL` のデータに対してどのようなクエリを有効にするかを適切に判断し、公開する必要がある。`XBRL` は多面的な情報をもつため、各書類の会社名、EDINET コード、該当期間、連結・非連結区分、各勘定科目ごとなど、様々な制約に基づいた問い合わせが可能である。今回生成した `RDF` は階層が深くなってしまっており、制約条件の記述がやや冗長になってしまっている。また、書類が公開されるごとに随時増大していくため、データ量が増えてもパフォーマンスを維持できるシステムの構築が不可欠である。一年分の開示書類により生成されるトリプル数は約 100 万トリプルになるため、期間や業種といった何らかの区分によってトリプルを分散させて管理する必要がある。

国際会計基準の導入が進んでいくに当たり、企業間の財務報告がある程度一元化されていることが予想される。過去に変換した `RDF` との一貫性を担保しつつ、年度横断的なクエリを投げられるようにシステムを整備していくことは今後の課題である。

参考文献

- [玉川 10]玉川 奨, 桜井 慎弥, 手島 拓也, 森田 武史, 和泉 憲明, 山口 高平: 日本語 `Wikipedia` からの大規模オントロジー学習, 人工知能学会論文誌, Vol. 25, No. 5, pp.623-636 (2010).
- [R. Garcia, R. Gil 09] Publishing `XBRL` as `Linked Open Data`, In *CEUR Workshop Proceedings*, volume 538 (2009)
- [J. Bao, G. Rong, X. Li, L. Ding 10] Representing Financial Reports on the Semantic Web - A Faithful Translation from `XBRL` to `OWL RuleML2010`, LNCS 6403, pp. 144-152 (2010)
- [DeRose, S., Maler, E., Orchard, D. 01] `XML Linking Language (XLink) Version 1.0.`, Technical report, W3C (2001)
- [JISX7206] 拡張可能な事業報告言語 (`XBRL`) 2. 1 (2005)