

# 有価証券報告書に含まれるデータの企業間比較における課題について

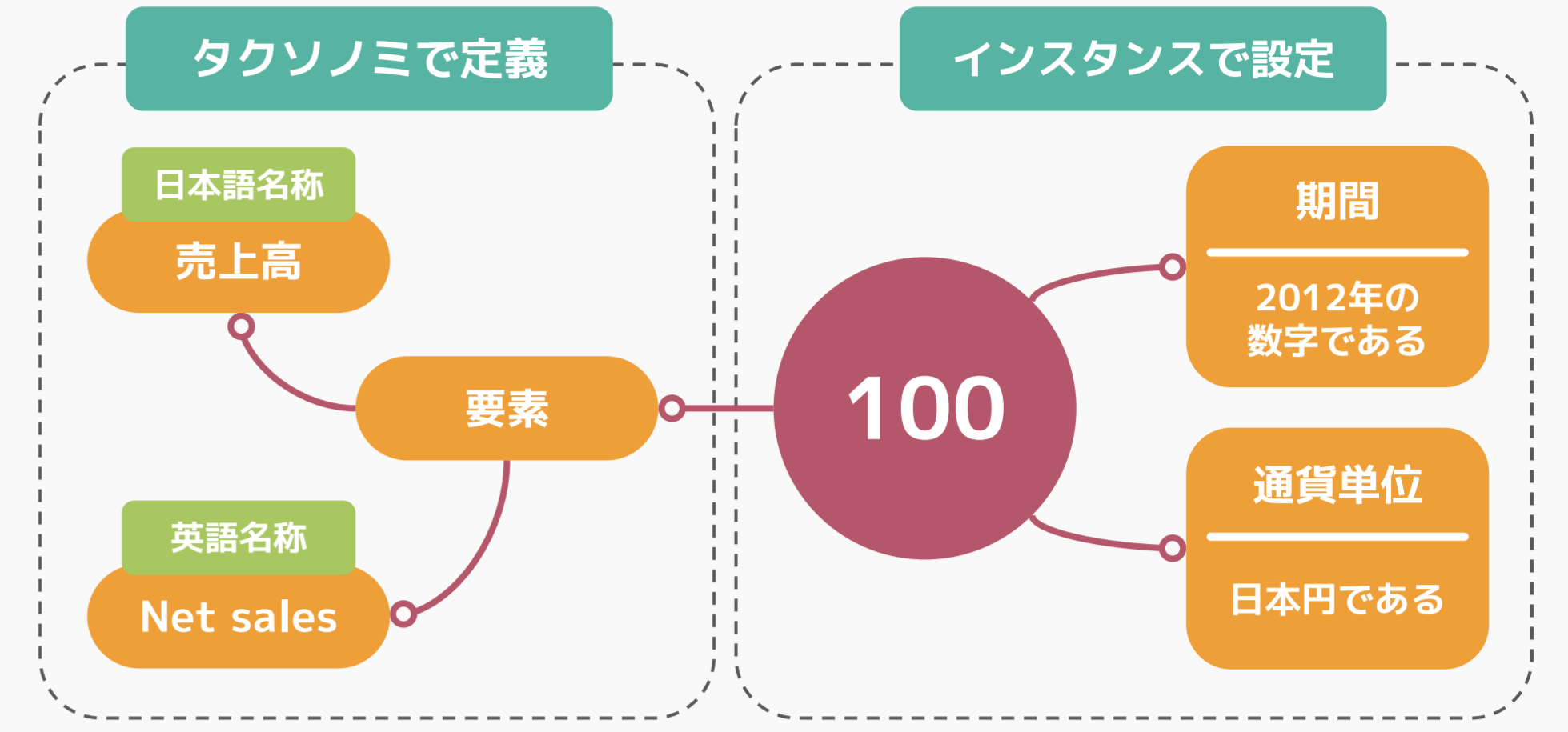
佐藤栄作 (小樽商科大学) 木村泰知 (小樽商科大学)

## 背景

有価証券報告書は、二次利用のためPDFやXBRL形式で公開されている

- XBRLに付与されているタクソノミ（要素名やその関係性の情報）とインスタンス（実際の値や年度等の情報）によって、表構造理解が可能
- しかし、タクソノミはおよそ18%の表にしか付与されていない

→ 表に含まれるデータの多くが活用困難である現状



有価証券報告書XBRLにおけるタクソノミとインスタンスのイメージ。この例では、「売上高」という日本語要素名をタクソノミが定義し、「100」という値や「2012年」という期間等をインスタンスが設定している。

## 目的

1. 有価証券報告書における「企業間の比較が可能なデータ」を定義する
2. 企業間でデータを比較する際の課題を明らかにする
3. 課題をもとに、有価証券報告書内データの企業間比較の実現に向けた解決策を提案する

## 分析・考察

### ① 「企業間の比較が可能なデータ」とは？

「要素ID+コンテキストID」が一致しているデータ

企業間の比較が可能 = 何のデータであるかが明確であり、かつ一致している  
→ XBRLの情報（タクソノミ、インスタンス）を利用して判別する

- 要素ID
- 主にデータの項目名を表す
  - 一つの文書内で一意ではない
- コンテキストID
- データの時期やメンバー要素等を表す
  - 一つの文書内で一意ではない

要素ID	項目名	コンテキストID	単位	値
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	Prior4YearDuration	円	3192900000000
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	Prior3YearDuration	円	3512909000000
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	Prior2YearDuration	円	3795992000000
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	Prior1YearDuration	円	4143505000000
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	CurrentYearDuration	円	4380209000000

大和ハウス工業

要素ID	項目名	コンテキストID	単位	値
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	Prior4YearDuration	円	1858790000000
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	Prior3YearDuration	円	2026931000000
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	Prior2YearDuration	円	2159363000000
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	Prior1YearDuration	円	2160316000000
jpcrp_cor:NetSalesSummaryOfBusinessResults	売上高、経営指標等	CurrentYearDuration	円	2415186000000

積水ハウス

比較が可能に！

要素IDとコンテキストIDを組み合わせることで、一つの文書内で一意となる

例えば…  
jpcrp\_cor:NetSalesSummaryOfBusinessResults + CurrentYearDuration  
要素ID + コンテキストID

メンバー要素によって一意になっている場合もある（詳細は②-B）

要素ID	項目名	コンテキストID	メンバー要素	値
jppfs_cor:NetAssets	純資産	CurrentYearInstant_NonConsolidatedMember	ValuationDifferenceOnAvailableForSaleSecuritiesMember	37247000000
jppfs_cor:NetAssets	純資産	CurrentYearInstant_NonConsolidatedMember	DeferredGainsOrLossesOnHedgesMember	728000000
jppfs_cor:NetAssets	純資産	CurrentYearInstant_NonConsolidatedMember	RevaluationReserveForLandMember	9119000000
jppfs_cor:NetAssets	純資産	CurrentYearInstant_NonConsolidatedMember	ValuationAndTranslationAdjustmentsMember	47095000000
jppfs_cor:NetAssets	純資産	CurrentYearInstant_NonConsolidatedMember	SubscriptionRightsToSharesMember	101000000
jppfs_cor:NetAssets	純資産	CurrentYearInstant_NonConsolidatedMember		1360805000000

### ② 企業間でデータを比較する際の問題点は？

- A. タクソノミの低付与率：約18%
- B. メンバー要素の存在
- C. 要素ID+コンテキストIDの低一致率：約4.7%
- D. 提出者別タクソノミの存在

#### A. タクソノミの低付与率：約18%

- それ以外の約82%に対しては、要素IDやコンテキストIDを取得できない
- 付与率の向上には提供者側の介入が必要であり、利用者側からのアプローチは困難である

#### B. メンバー要素の存在

- ある概念をセグメント化したもの
- タクソノミ付与データの約66%に付与
- 要素IDではなく、メンバー要素が項目名に対応している場合がある

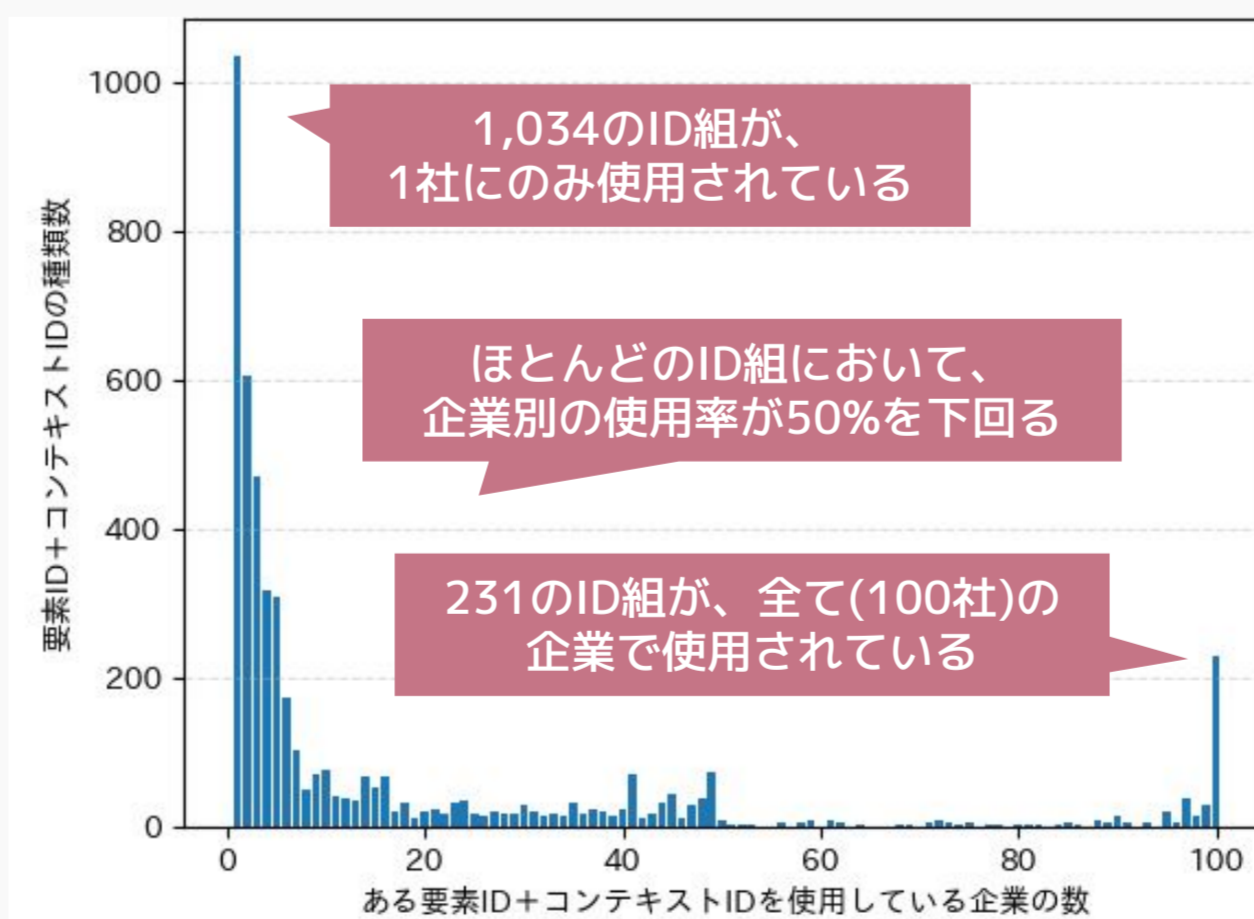
評価・換算差額等	
その他有価証券評価差額金	37,247
繰延ヘッジ損益	728
土地再評価差額金	9,119
評価・換算差額等合計	47,095
新株予約権	101
純資産合計	1,360,805

#### C. 要素ID+コンテキストIDの低一致率：約4.7%

EDINETタクソノミ 汎用的に使われているタクソノミのこと

EDINETタクソノミを対象としたところ、合計で4,907種類の要素ID+コンテキストID

全ての企業（100社）が共通して使用しているのは231組（約4.7%）に留まる

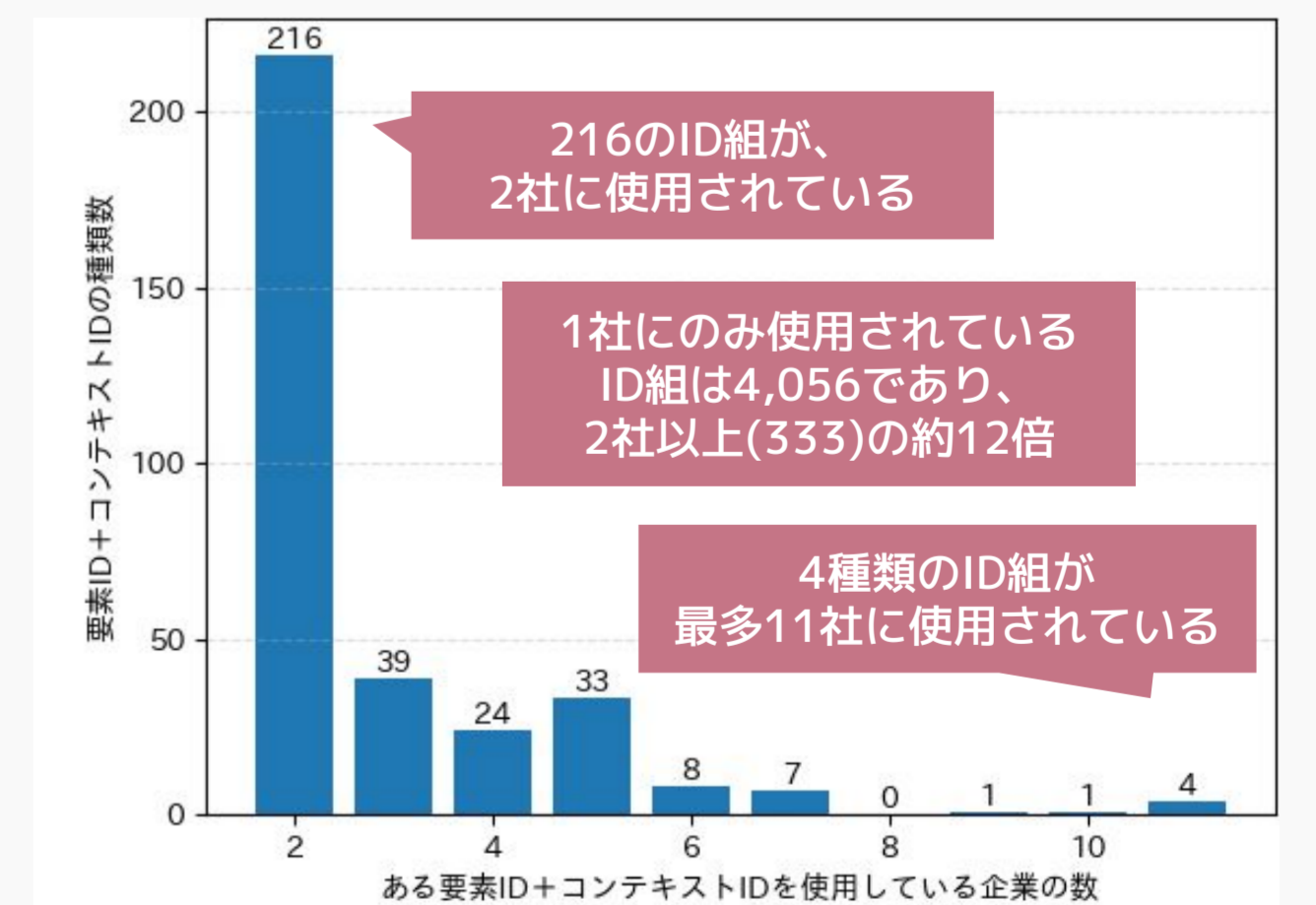


#### D. 提出者別（企業独自）タクソノミの存在

提出者別タクソノミ 企業が独自に設定するタクソノミのこと

本来は同じ要素を示すデータであっても、要素ID+コンテキストIDが一致しなければ比較が困難

類似度を用いたアプローチで解決できないか



### ③ 企業間比較の実現に向けた解決策は？

LLMを用いた質問応答のアプローチ

課題より、要素ID+コンテキストIDの一致でデータを比較することは困難

→ LLMを用いた質問応答のアプローチで解決できないか

例えば

「A社とB社では、2020年度の売上高が高いのはどちらか？」という質問に対し、「A社は〇〇、B社は■の箇所より、A社の方が高い」という回答を得たい

よって、まずは一つの企業において、所望のデータを取得したい

→ 有価証券報告書を用いたQAタスクの設定と、データセットの作成  
XBRLの情報を用いることで、大規模なデータセットの構築が期待できる

XBRL情報

質問文  
回答文

要素ID	項目名	コンテキストID	相対年度	連結・個別	期間・時点	値
jpcrp_cor:NetSales...	売上高、経営指標等	Prior4YearDuration	四期前	その他	期間	3192900000000
jpcrp_cor:NetSales...	売上高、経営指標等	Prior3YearDuration	三期前	その他	期間	3512909000000
jpcrp_cor:NetSales...	売上高、経営指標等	Prior2YearDuration	前々期	その他	期間	3795992000000
jpcrp_cor:NetSales...	売上高、経営指標等	Prior1YearDuration	前期	その他	期間	4143505000000
jpcrp_cor:NetSales...	売上高、経営指標等	CurrentYearDuration	当期	その他	期間	4380209000000

大和ハウス工業の2020年度の「売上高、経営指標等」は？

4380209000000円です。

有価証券報告書XBRLから質問文と回答文を作成するイメージ