

# XBRLを活用いたデータマイニング

早稲田大学 商学学院  
金 奕群 キン エキグン  
2021年10月28日 @武蔵大学

1. XBRLの紹介とデータの入手可能性
  - 1.1 XBRLの紹介
  - 1.2 XBRLデータの入手可能性
  
2. XBRLデータのハンドリング (Rベース)
  - 2.1 XBRLデータの取得と蓄積
  - 2.2 XBRLデータから財務諸表数値の抽出
  - 2.3 XBRLデータからテキストデータの抽出

## 1.1 XBRLの紹介

XBRL の定義	eXtensible Business Reporting Language：各種事業報告用の情報（財務・経営・投資などの様々な情報）を作成・流通・利用できるように <u>標準化されたXMLベースのコンピュータ言語</u>	
	XML	Extensible Markup Language：拡張可能なマーク付け言語
	構文例	<code>&lt;ele&gt;2&lt;/ele&gt;</code>
XBRLの構文例 (連結売上高)	<code>&lt;jppfs_cor:NetSales contextRef="<b>CurrentYearDuration</b>" unitRef="JPY" decimals="-6"&gt; 310000000 &lt;/jppfs_cor:NetSales&gt;</code>	

## 1.1 XBRLの紹介

XBRL の構成	XBRL仕様 (Specification)	XBRL international Inc.により勧告され、XBRL仕様にもとづいて世界各国の規制当局がタクソノミを開発する。
	タクソノミ (Taxonomy)	財務諸表、開示書類の雛形。開示項目のタグを定義している。
		EDINETタクソノミ： 内閣府令タクソノミ 財務諸表本表タクソノミ 国際会計基準タクソノミ DEIタクソノミ
	インスタンス (Instance)	提出者は、タクソノミにもとづいてインスタンスを作成する。

## 1.2 XBRLデータの入手可能性

有価証券報告書  EDINET (過去5年分を無料で取得可能)	情報内容：	本表と一部の注記情報	ほか	
	タクソノミ：	財務諸表本表タクソノミ	内閣府令タクソノミ	
	開発元：	金融庁	金融庁	
	タグ付けの方法：	詳細タグ付け	詳細タグ付けと包括タグ付け	
決算短信  TDNetまたは東証 上場会社サービス (過去5年分を無料で取得可能)	情報内容：	サマリー情報	添付資料 (本表)	添付資料 (本表以外)
	タクソノミ：	決算短信タクソノミ	財務諸表本表タクソノミ	NA
	開発元：	東証	金融庁	-
	タグ付けの方法：	詳細タグ付けと包括タグ付け	詳細タグ付け	-

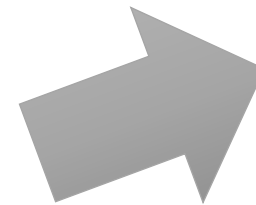
## 2.1 XBRLデータの取得と蓄積

EDINET API	利用者がEDINETの画面からではなく、プログラムを介してEDINETのデータベースから効率的にデータを取得できるAPI（アプリケーション・プログラミング・インターフェース） →リクエスト（URL）、レスポンス（内容）		
	構成	①書類一覧API：提出された書類を <b>把握する</b> ためのAPI レスポンス：日付ごとの提出書類情報（json型）	
		②書類取得API：提供された書類を <b>取得する</b> ためのAPI リクエスト：個々の書類を取得できるURL	
①書類一覧 API	リクエストの例	<a href="https://disclosure.edinet-fsa.go.jp/api/v1/documents.json?date=2019-04-01&amp;type=2">https://disclosure.edinet-fsa.go.jp/api/v1/documents.json?date=2019-04-01&amp;type=2</a>	
	レスポンスの構成	メタデータ	タイトル："提出された書類を把握するためのAPI" パラメータ：リクエストの識別子（日付、タイプ） ステータス：リクエストが成功したかどうか 等
		提出書類一覧	書類ごとのパラメータ
	提出書類情報の例	→次のページへ	

## 2.1 XBRLデータの取得と蓄積

提出書類  
情報の例

```
{  
  "seqNumber": 8,  
  "docID": "S100FJAQ",  
  "edinetCode": "E34721",  
  "secCode": "44380",  
  "JCN": "9011001070865",  
  "filerName": "株式会社W e l b y",  
  "fundCode": null,  
  "ordinanceCode": "010",  
  "formCode": "030000",  
  "docTypeCode": "120",  
  "periodStart": "2018-01-01",  
  "periodEnd": "2018-12-31",  
  "submitDateTime": "2019-04-01 09:01",  
  "docDescription": "有価証券報告書－第8期(平成30年1月1日－平成30年12月31日)",  
  ...  
  (中略)  
  ...  
},
```



seqNumber：日付ごとのファイル連番  
docID：書類管理番号  
edinetCode：提出者のEDINETコード  
secCode：提出者の証券コード  
JCN：提出者の法人番号  
filerName：提出者の名前  
fundCode：ファンドコード  
ordinanceCode：府令コード  
formCode：様式コード  
docTypeCode：書類種別コード  
periodStart：期間（自）  
periodEnd：期間（至）  
submitDateTime：提出日時  
docDescription：提出書類概要

“ordinanceCode”：“010”→企業内容等の開示に関する内閣府令  
“formCode”：“030000”→第三号様式 有価証券報告書  
“docTypeCode”：“120”→有価証券報告書

## 2.1 XBRLデータの取得と蓄積

②書類取得APIのリクエストURL	https://disclosure.edinet-fsa.go.jp/api/v1/documents/ <b>書類管理番号</b> ?type= <b>取得パラメータ</b>	
	取得パラメータ	1 提出書類本文+監査報告書のXBRLファイルを取得する
		2 PDFファイルを取得する
		3 代替書面・添付文書を取得する
		4 英文ファイルを取得する
例	https://disclosure.edinet-fsa.go.jp/api/v1/documents/ <b>S100FJAQ</b> ?type=1	
データを取得する手順（まとめ）	Step 1 書類一覧の取得	日付ごとの書類一覧リストを取得する
	Step 2 XBRLデータの取得	有価証券報告書の書類管理番号を把握してから、XBRLデータ（ZIPファイル）を取得する
	→Rスクリプト	



## 2.1 XBRLデータの取得と蓄積

XBRL ファイル の構成  (解凍後)	Audit Doc	manifest_AuditDoc.xml
		jpaud-aai-cn-001_E34721-000_2018-12-31_01_2019-04-01.xsd
		jpaud-aai-cn-001_E34721-000_2018-12-31_01_2019-04-01_pre.xml
		jpaud-aai-cn-001_E34721-000_2018-12-31_01_2019-04-01.xbrl jpaud-aai-cn-001_E34721-000_2018-12-31_01_2019-04-01_ixbrl.htm
	Public Doc	manifest_PublicDoc.xml
		jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01.xsd
		jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01_cal.xml
		jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01_def.xml EDINETタクソノミ
		jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01_lab-en.xml (拡張)
		jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01_lab.xml ↓
jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01_pre.xml 提出者別タクソノミ		
jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01.xbrl		
0000000_header_jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01_ixbrl.htm		
0101010_honbun_jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01_ixbrl.htm		
...		
中略	インスタンス	
...		
0201010_honbun_jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01_ixbrl.htm		
fuzoku		

## 2.1 XBRLデータの取得と蓄積

インスタンス の命名規則	jpcrp030000-asr-001_E34721-000_2018-12-31_01_2019-04-01.xbrl	
	jpcrp030000	jp{府令番号}{様式番号} 企業内容等の開示に関する内閣府令 第三号様式
	asr	報告書略号 有価証券報告書
	001	報告書連番（大量保有報告書等同じ様式の報告書を同日に複数提出する場合の識別番号）
	E34721	EDINETコード
	000	（シリーズファンド等）の追番
	2018-12-31	報告対象期間期末日
	01	報告書提出回数
	2019-04-01	報告書提出日
取得した後の管理 （まとめ）	Step 3 解凍	ZIPファイルを解凍する
	Step 4 XBRLデータ中身の把握	インスタンス名で提出者、提出日、決算期、提出回数を把握できるリストを作成する
	→Rスクリプト <sup>10</sup>	

## 2.2 XBRLデータから財務諸表数値の抽出

インスタンス の構成 (坂上, 2011)	①前段部分	対応するタクソノミスキーマ、名前空間を宣言
	②コンテキスト情報部分	企業情報、決算日、会計期間などを宣言
	③単位部分	通貨単位を宣言
	④データ部分	財務数値、記述的な内容を記載
	⑤脚注部分	必要に応じてアドホックな情報を記載
データ部分における 財務諸表数値情報の 記載例	連結 売上高の 記載	<pre>&lt;jppfs_cor:NetSales contextRef="<b>CurrentYearDuration</b>" unitRef="JPY" decimals="-6"&gt; 323609000000 &lt;/jppfs_cor:NetSales&gt;</pre>
	単独 売上高の 記載	<pre>&lt;jppfs_cor:NetSales id="ldFact2008058145" contextRef="<b>CurrentYearDuration_NonConsolidatedMember</b>" unitRef="JPY" decimals="-6"&gt; 210346000000 &lt;/jppfs_cor:NetSales&gt;</pre>

## 2.2 XBRLデータから財務諸表数値の抽出

インスタンス における記載	①	<jppfs_cor:NetSales contextRef="CurrentYearDuration" unitRef="JPY" decimals="-6">323609000000</jppfs_cor:NetSales>						
	②	<jppfs_cor:NetSales id="IdFact2008058145" contextRef="CurrentYearDuration_NonConsolidatedMember" unitRef="JPY" decimals="-6">210346000000</jppfs_cor:NetSales>						
整理後 のデータ		タグ 名前空間	タグ名	id属性	contextRef属性	unitRef 属性	decimals 属性	値
	①	jppfs_cor	NetSales	なし	CurrentYearDuration	JPY	-6	323609 000000
	②	jppfs_cor	NetSales	IdFact200 8058145	CurrentYearDuration_N onConsolidatedMember	JPY	-6	210346 000000

## 2.3 XBRLデータからテキストデータの抽出

	数値や内容ごとにタグを付ける	
	詳細 タグ 付け	連結売上高 の記載 <pre>&lt;jppfs_cor:NetSales contextRef="CurrentYearDuration" unitRef="JPY" decimals="-6"&gt;                 323609000000                 &lt;/jppfs_cor:NetSales&gt;</pre>
データ部分の記載例	複数の情報を1個のテキストブロックとしてまとめてタグを付ける	
	包括 タグ 付け	事業等のリスク の記載 <pre>&lt;jpcrp_cor:BusinessRisksTextBlock contextRef="FilingDateInstant"&gt;                 &amp;lt;h3&amp;gt; 2 【事業等のリスク】 &amp;lt;/h3&amp;gt;&amp;lt;div&amp;gt;&amp;lt;p style="margin-left: 18.0pt; line-height: 13.5pt"&amp;gt; 有価証券報告書に記載した事業の状況、経理の状況等に関する事項のうち、投資者の判断に重要な影響を及ぼす可&amp;lt;br/&amp;gt;能性のある事項には、                 ...                 (中略)                 ...                 &lt;/jpcrp_cor:BusinessRisksTextBlock&gt;</pre>

## 2.3 XBRLデータからテキストデータの抽出

<p>「事業等のリスク」の記載</p>	<pre>&lt;jpcrp_cor:BusinessRisksTextBlock contextRef="FilingDateInstant"&gt; &lt;h3&gt; 2 【事業等のリスク】 &lt;/h3&gt;&lt;div style="margin-left: 18.0pt; line-height: 13.5pt"&gt; 有価証券報告書に記載した事業の状況、経理の状況等に関する事項のうち、投資者の判断に重要な影響を及ぼす可&lt;br/&gt;能性のある事項には、 ... (中略) ... &lt;/jpcrp_cor:BusinessRisksTextBlock&gt;</pre>
<p>記載ルール</p>	<p>内容の中身は、HTMLで書かれている。  ただし、XBRLタグと区別するために、文字実体参照でHTMLタグを表している。</p> <p style="text-align: center;">「&amp;lt;」 → 「&lt;」  「&amp;gt;」 → 「&gt;」</p>
<p>HTMLへの変換</p>	<pre>&lt;h3&gt; 2 【事業等のリスク】 &lt;/h3&gt; &lt;div&gt; &lt;p style=""margin-left: 18.0pt; line-height: 13.5pt""&gt; &lt;U+00A0&gt;&lt;U+00A0&gt;有価証券報告書に記載した事業の状況、経理の状況等に関する事項のうち、投資者の判断に重要な影響を及ぼす可&lt;br/&gt;能性のある事項には、 以下のようなものがあります。 &lt;/p&gt; (略)</pre>

## 2.3 XBRLデータからテキストデータの抽出

HTMLの表示	<p>2【事業等のリスク】          有価証券報告書に記載した事業の状況、経理の状況等に関する事項のうち、投資者の判断に重要な影響を及ぼす可能性のある事項には、以下のようなものがあります。          なお、文中の将来に関する事項は、当連結会計年度末現在において当社グループが判断したものであります。</p> <p>(1) ○○○○○○          .....          .....</p> <p>(2) ○○○○○○          .....          .....</p> <p>(3) 継続企業の前提に関する重要事象等          当社グループは、売上高が前連結会計年度に比べ88億88百万円減少しており、営業損失88億88百万円、経常損失88億88百万円となりました。このことにより、継続企業の前提に重要な疑義を生じさせるような状況が存在しております。.....          .....          当社は、当該状況を解消すべく、.....          .....</p>	
TXTの中身 (改行、スペース削除後)	<p>2【事業等のリスク】有価証券報告書に記載した事業の状況、経理の状況等に関する事項のうち、投資者の判断に重要な影響を及ぼす可能性のある事項には、以下のようなものがあります。なお、文中の将来に関する事項は、当連結会計年度末現在において当社グループが判断したものであります。(略)</p>	
XBRLから内容の抽出 (まとめ)	Step 5 XBRLから情報内容の抽出	<p>数値内容をCSVデータへ変換</p> <hr/> <p>テキスト内容をTXTデータへ変換</p> <hr/> <p style="text-align: right;">→Rスクリプト</p>

## 参考資料：

- 金融庁. 2015. 「[次世代EDINETタクソノミの公表について](#)」.
- 金融庁. 2019. 「[EDINET API仕様書](#)」.
- 金融庁. 2021. 「[2021年版タクソノミの公表について](#)」.
- 坂上学. 2011. 『新版 会計人のためのXBRL入門』 同文館出版.

## R Package:

- [jsonlite: A Simple and Robust JSON Parser and Generator for R.](#)
- [xml2: Parse XML.](#)



ありがとうございます。