

有価証券報告書の類似度による企業評価

Company evaluation based on similarity of corporate governance report

西野嘉之^{†‡}

Yoshiyuki Nishino

[†]産業能率大学

[‡]ユーレット株式会社

[†] SANN0 UNIVERSITY.

[‡] Ullet Co., Ltd.

要旨

上場企業の有価証券報告書は公開されており、企業情報を得ることができる。企業分析を行う際、その企業の課題を把握し、財務データと紐付けることで、課題に対する解決策を評価する必要がある。しかしながら、有価証券報告書内の「対処すべき課題」などのテキストデータは、重要な情報であるにも関わらず、定量的な評価をすることができない。そこで、本研究では、「対処すべき課題」の文章がどれだけ似ているかを表す類似度を人工知能 (AI) により数値化し、同一企業の財務データとテキストデータを時系列に紐づけて比較検討する方法を提案する。その結果、企業の経営方針の安定度を評価することが可能であることを示す。さらに、分析プロセスによって、人間中心の情報システムとは、人間の行動に沿って時系列に情報を整理することが重要であることを示す。

1. はじめに

上場企業は有価証券報告書を金融庁、及び東京証券取引所に提出する義務があり、それらは公開されている[1]。近年、有価証券報告書などのビジネスレポートを電子文書化することで、それらの作成の効率化や比較・分析などの二次利用を目的として、XMLの規格をベースに作られたXBRL(eXtensible Business Reporting Language)が登場した。金融庁のEDINETというサイトから、XBRLファイルをダウンロードして、「XBRLからCSVへの変換ツール」をインストールし、XBRLを取り込むことで、XBRLの中身を見ることができる。XBRLをCSVに変換すると、タグとその中身のデータをエクセルで扱うことが可能となる。これらの上場企業約3700社のデータ数年分を比較することはデータ量が膨大であり、個人投資家などが取り扱うには困難なデータ構造となっている。例えば、金融庁の「XBRLからCSVへの変換ツール」を使用して3社を比較しようとする、直近の3社の有価証券報告書のXBRLをダウンロードし、それらのファイルから比較したい勘定科目を探し、比較する必要がある。もし3社を比較するのであれば、有価証券報告書のPDFを直接目視した方が格段に早く処理することができる。これでは、少なくとも人間中心の情報システムとは言えないだろう。また、XBRLの勘定科目における売上高で全社を比較しようとする、「売上高」ではなく、保険業などは、「経常収益」となっており、業種ごとに異なる勘定科目を名寄せしなければ、二次利用することができない。XBRLをダウンロードしたのち、データの関係性の辞書を作る必要がある。また、「役員」、「大株主」、「平均年収」、「従業員数」などは、1対1のXBRLのタグになっておらず、XBRLの中をテキストマイニングして該当データを探す必要がある。さらに、「対処すべき課題」や「研究開発」といったテキストデータは各章ごとにXBRLのタグにテキスト以外のメタ情報も含まれて格納されているため、文章だけを抽出する必要がある。さらに、財務データから、企業の課題の変化を把握することは難しい。また、テキストマイニングを行う手法では、決算短信、あるいは雑誌や新聞のキーワードの露出度の分析を行い、相関関係を可視化する研究がなされているが、課題の変化を把握することはできない[2][3]。

そこで、本研究では、企業価値検索サービス「ユーレット」(Ullet)[4]のデータを用いて、すべての上場企業の財務データである勘定科目と数値だけを抽出し、さらにテキストデータからは平均年収などの数値と「対処すべき課題」を抽出するクレンジングを行う。クレンジングを行うことで、項目と値が1対1の関係性となり、データを分析することが可能となる。人間中心の情報システムは、人間の行動と、その思考に沿って考察するべきであり、「対処すべき課題」の当年度と前年比について、AIにより自然

言語処理を行った類似度を考察することで、企業内の変化を読み解く。

2. 企業価値検索サービス「ユーレット」(Uilet)

2.1. データの抽出

企業価値検索サービス「ユーレット」(Uilet)は、XBRLのデータをもとにデータを整理している。金融庁のEDINETからXBRLをダウンロードして、開いたファイルを示す(図1-1)。

XBRLはXMLタグとその中身のデータで構成されており、XBRLファイルのままでは、CSVファイルのように扱うことができない。そこで、XBRLのタグを解析し、中身を抽出する作業が必要となる。

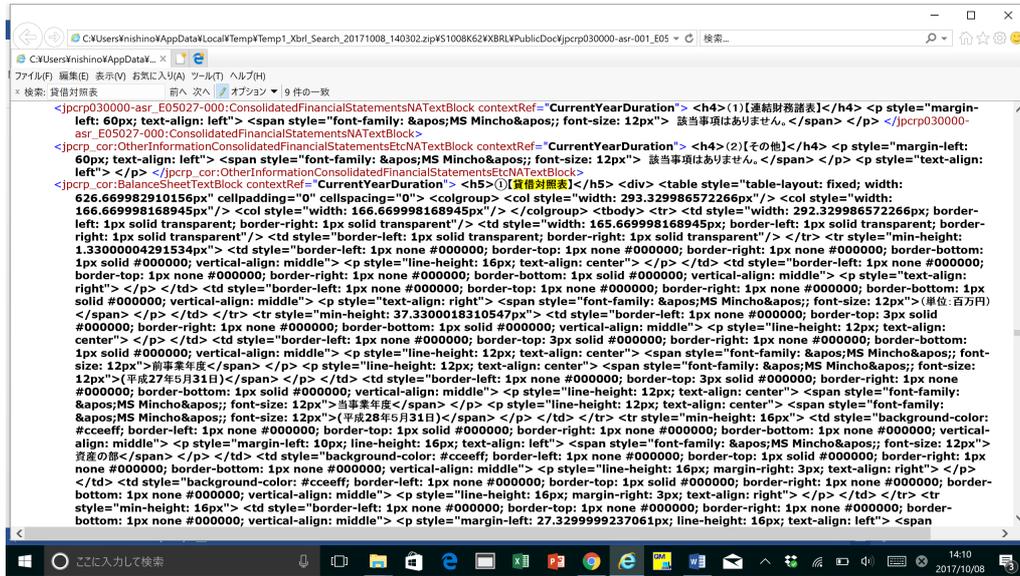


図1-1: 日本オラルクのXBRL

図1-1のXBRLファイルのデータをもとに、XBRLの項目名と、中身の値だけを抽出して、データベース(DB)に格納している。そのDBファイルから、CSVファイルで値を抽出したファイルを示す(図1-2)。

The image shows a CSV file with columns for company codes, names, industries, markets, decision dates, months, sales, and other financial metrics. The data is organized in a grid format.

A	B	C	D	E	F	G	H	I	J	K
証券コード	企業名	業種	市場	決算日	月数	売上高	当期純利益	ROA	住所	設立年月日
1409	4694 ヒー・エム・エルサービス業	東証1部	市場	2017年3月31日	12	111243	5948	5.705843981	東京都渋谷区千駄ヶ谷	1955年7月
1410	4695 マスターエン・サービス業	東証2部	市場	2017年3月31日	12	18106	576	4.986963651	東京都品川区五反田	1974年6月1
1411	4696 ユナベウエディ・サービス業	東証1部	市場	2017年3月31日	12	43908	123	0.576490439	東京都下区四谷条通	1971年4月1
1412	4699 ユナダエスコ サービス業	JASDAQ	市場	2016年7月20日	12	11590	264	2.735467827	東京都江東区本場5-8	1972年11月1
1413	4704 トレンドマイクロ情報・通信業	東証1部	市場	2016年12月31日	12	131936	24651	7.989641437	東京都渋谷区代	1989年10月2
1414	4705 リップコーポ・サービス業	JASDAQ	市場	2017年3月31日	12	3453	62	0.940391324	名古屋千種区内山3	1981年5月2
1415	4707 キック サービス業	JASDAQ	市場	2016年10月20日	12	2430	126	2.375565611	新潟市中央区新光町11	1973年2月
1416	4708 いろあいコミュニケーションサービス業	東証1部	市場	2017年3月31日	12	96188	3264	4.94028969	東京都渋谷区代	1987年6月2
1417	4709 インフォメーション情報・通信業	東証1部	市場	2017年3月31日	12	21554	654	6.19787718	東京都千代田区五番	1969年10月2
1418	4712 2ドアーズ サービス業	JASDAQ	市場	2017年3月31日	12	22355	210	0.964585917	東京都港区虎ノ門1-7	1967年7月2
1419	4714 リー教育 サービス業	東証1部	市場	2017年2月28日	12	20777	1275	11.33333333	東京都豊島区白3-1	1985年7月
1420	4716 日本オラルク 情報・通信業	東証1部	市場	2017年5月31日	12	173190	36360	18.29608868	東京都港区北青山12-5	1985年11月
1421	4718 早稲田アカデミーサービス業	東証1部	市場	2017年3月31日	12	20685	679	5.399173028	東京都豊島区池袋2-5	1974年11月
1422	4719 アルファンテ 情報・通信業	東証1部	市場	2017年3月31日	12	26806	1737	4.196970063	川崎市中原区小田田	1972年10月1
1423	4720 城南進学研究サービス業	JASDAQ	市場	2017年3月31日	12	6926	192	2.937576489	川崎市川崎区駅前本	1982年9月1
1424	4722 フェューチャー 情報・通信業	東証1部	市場	2016年12月31日	12	33653	2208	9.573776178	東京都品川区大田4-1	1989年11月2
1425	4725 CAAC Holding 情報・通信業	東証1部	市場	2016年12月31日	12	52521	2039	4.050135071	東京都中央区日本橋	1966年8月
1426	4726 フトバンクテ 情報・通信業	東証1部	市場	2017年3月31日	12	50225	1598	5.981129556	東京都新宿区西新宿6-2	1990年10月1
1427	4728 ーセ 情報・通信業	東証1部	市場	2016年8月31日	12	4874	93	1.420932009	東京都下区東区東院	1979年11月
1428	4732 ユー・エス・エスサービス業	東証1部	市場	2017年3月31日	12	67179	22909	11.60689858	愛知県東海市新町5	1980年10月2
1429	4733 オービックビジ 情報・通信業	東証1部	市場	2017年3月31日	12	23290	7669	6.50301026	東京都新宿区西新宿6	1980年12月1
1430	4734 ビーイング 情報・通信業	JASDAQ	市場	2017年3月31日	12	4936	164	2.505729565	津市桜橋1-312 周辺地	1984年9月1
1431	4735 京速 サービス業	東証2部	市場	2017年5月31日	12	13513	431	3.915690015	東京都下区馬込通	1981年4月
1432	4736 日本ラッド 情報・通信業	JASDAQ	市場	2017年3月31日	12	3222	409	16.17240016	東京都港区虎ノ門2-2	1971年6月
1433	4739 伊藤忠テクノ 情報・通信業	東証1部	市場	2017年3月31日	12	407849	21861	6.524409	東京都千代田区霞が	1979年7月1
1434	4743 アテック 情報・通信業	東証1部	市場	2017年3月31日	12	11111	906	6.07279308	東京都千代田区一善	1959年5月
1435	4745 東京個別指導サービス業	東証1部	市場	2017年2月28日	12	17909	1438	13.66270784	東京都新宿区西新宿1	1994年5月1

図1-2: ユーレットのDBから作成したCSVファイル

2.2. データの整理

2.1章で示したデータを可視化した画面を示す(図2)。図2は、ユーレット(<http://www.uilet.com>)の日本オラルクの「主な指標」である。

XBRL から項目名と値が1対1になったデータを元に Ullet では、棒グラフや円グラフを自動生成している。また、「役員」の生年月日や経歴、「大株主」、「平均年収」、「対処すべき課題」、「研究開発」といった情報も web 上で公開している[4]。

主な指標

[見方] 一般に、経営状態が良い会社はグラフが右肩上がりになります。

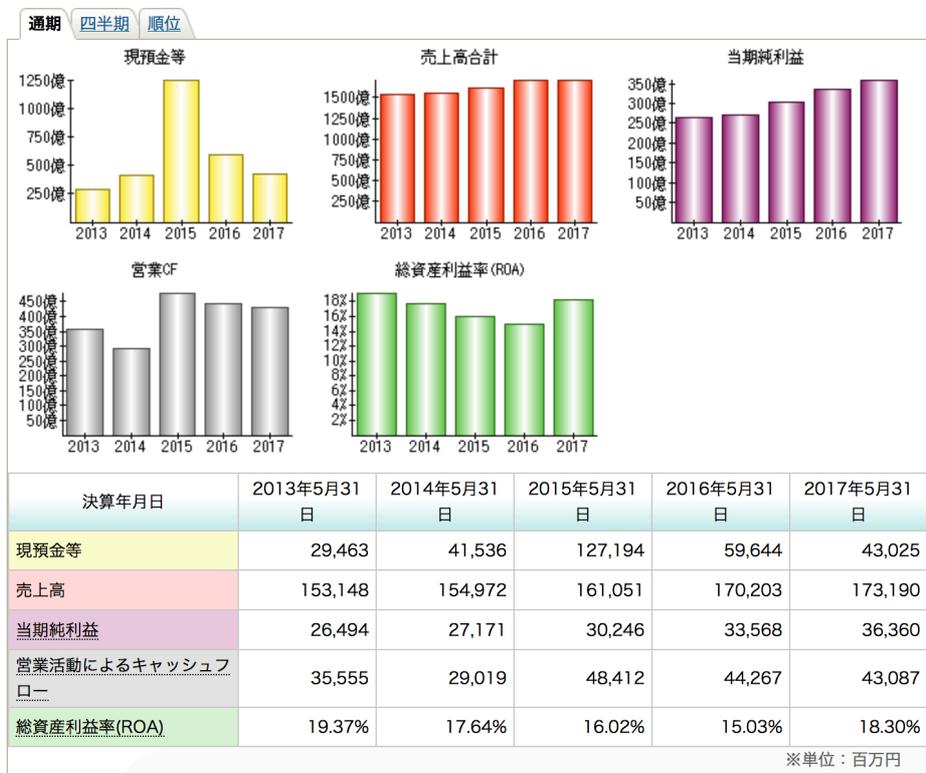


図2：日本オラクルのユーレット「主な指標」

本研究では、このようにXBRLの情報がDBで取り扱うことが可能な状態であり、尚且つ可視化されている環境が整っていることが前提となる。

3. 「対処すべき課題」の類似比較の方法

同一企業の財務データとテキストデータを時系列に紐づけて比較検討する方法を提案する。

3.1. テキストデータを比較するための条件

テキストデータを分析する場合、自由度が高いアンケートの回答のような文章は、露出回数をカウントする方法や、キーワードでクラスタリングする手法が過去に行われているが、注目すべき言葉は抽出されるが、企業の抱える課題に対する取り組みを数値化し、変化の度合いを定量的に評価することは難しい。そこで、本研究では、有価証券報告書というフォーマットが決まっており、書く内容も指導される公的文書を用いることで入力形式とその内容がある程度定型化されている条件下において、有効な方法を示す。また、他社の「対処すべき課題」のテキストデータを比較するのではなく、同じ企業を時系列で比較を行う。

3.2. 類似度の提案手法

類似度を評価する二乗和を用いて、「対処すべき課題」の類似度を算出する[5]。

各企業の当年度と前年度の「対処すべき課題」のテキストデータだけを入力値とする。それぞれの文章に対して、形態素解析を行う。分解した単語をカウントせず、配列に格納し、一致した要素 X_i の文字数の二乗を返し、集計する。前年度の「対処すべき課題」の形態素解析した要素 X_j の二乗を集計したもので正規化し、確率を算出する。こ

れにより、当年度の「対処すべき課題」に対して前年度の「対処すべき課題」がどの程度類似しているかを算出する。文字列数の値を二乗することで、一致する文字列が長い程、類似度 P は高くなる。

$$P = \frac{\sum_{j=1}^N X^2_i}{\sum_{i=1}^M X^2_j} \quad (0 \leq P \leq 1.0) \quad (1)$$

さらに、「対処すべき課題」を定量的に表し、財務データと紐づける。通常、決算報告では、当年度の決算報告を行い、来年度の「対処すべき課題」が公開される。しかし、前年度の「対処すべき課題」は当年度の有価証券報告書には記載されない。つまり、当年度の有価証券報告書には、財務データの前年比の値は存在するが、前年の「対処すべき課題」は記述されないため、評価対象外となってしまう。2017年3月31日の決算は、2016年3月31日の「対処すべき課題」と、比較検討されるべきである。これにより企業が対処すべき課題に対して取り組んだ結果、財務データはどのような結果となり、また来年度の「対処すべき課題」をどの程度前年と比べて書き換えたかを検証する必要がある。

データを読み解く、時系列を示す。

2016年記載の「対処すべき課題」は、2017年の対応策を表している。



2017年の「貸借対照表 (BS)、損益計算書 (PL)」の経済活動の結果が出る。



2017年記載の「対処すべき課題」は、2018年の対応策、すなわち未来を表している。

3.3. 財務データと「対処すべき課題」の時系列評価の提案

日本オラクルについて、提案手法を用いて説明する。ユーレットの「主な指標」を確認すると、「売上合計」と「当期純利益」は、右肩上がりである。個人投資家や転職を考えている人が見た場合、伸びている優良企業に見える。2012年から2016年の財務データと対処すべき課題を時系列に矢印で結んだ図3-1を示す。



図3-1：財務データと「対処すべき課題」の時系列

次に、日本オラクルの2012年5月期から2016年5月期の「対処すべき課題」の類似度 P を図3-2に示す。2012年→2013年は、類似度 P が100%であった。2013年→2014年は、類似度 P が29.6%となっており、大幅に「対処すべき課題」が書き変わっていると言える。その理由として、杉原新社長就任が大きいと考えられる。クラウドサービスを本格的に開始し、経営方針が大きく転換した年度である。さらに、2014年→2015年、2015年→2016年は、杉原社長の方針を貫いているのであれば、類似度 P は高いまものはずである。しかしながら、算出結果 P は2014年→2015年は66.8%、2015年→2016年は31.8%となり、2017年5月期で杉原社長は退任することに

なった。日本オラクルは、売上及び当期純利益だけを見れば、堅実に伸びている企業に見える。しかし、「対処すべき課題」と財務データを時系列で追っていくと、ビジネスモデルの変革に対応するために方針が迷走しているように見える。

さらに、貸借対照表の円グラフをユーレットから参照する。水色の現金と、紺色のその他流動資産が大きく変化していることが分かる。そこで、有価証券報告書の勘定科目で詳細を確認すると、短期貸付金が毎年変化していることが分かる。さらに、詳細な記載を有価証券報告書の「関係当事者情報」の欄を確認すると、米国オラクル社が相手先であることが分かる（図3-2）。

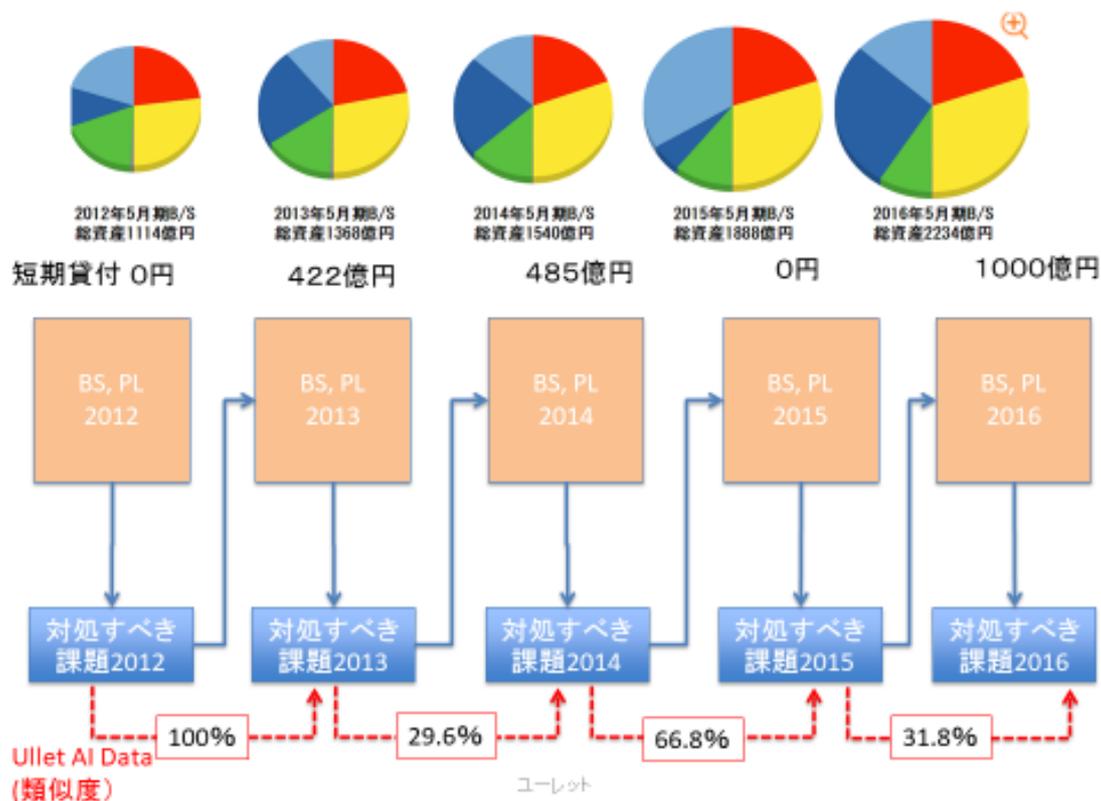


図3-2 財務データと「対処すべき課題（類似度 P）」の時系列

4. 分析結果と考察

提案手法を用いて、企業を分析した結果を示し、類似度 P と財務データの関係性について考察する。3章の日本オラクルの「対処すべき課題」の2012年→2013年の類似度 P の結果が100%であることが確認できた。当時の日本オラクルは、DB業界の王者であり、ビジネスモデルが安定していたため、「対処すべき課題」も変わることはなかったと考えられる。そこで、上場企業の売上規模など財務データの変化から企業を評価するのではなく、「対処すべき課題」の類似度 P の変化率に注目した。安定企業は「対処すべき課題」に変化がないと仮定し、調べていると複数社存在した。さらに、新興市場などの企業は成長性の高い企業ではあるが、経営としては安定期ではないため、東証一部企業の2014年から2017年の4年間で変化率が100%の企業を抽出し、「売上高」及び「当期純利益」について考察した。その結果、すべての上場企業(3645社)のうち、36社が抽出された。36社のユーレットの「主な指標」を確認すると、「4679 田谷」、「5451 淀川製鋼所」、「7601 ポプラ」以外は4期連続黒字であった。売上高でソートした上位10社を表1に示す。この結果から、「対処すべき課題」の類似度100%の企業を「当期純利益が黒字である」という条件で絞り込むことで、安定企業を算出することができる。なお、調査対象とした3645社は、ユーレットの2017年8月末時点の上場企業数である。

コード	企業名	業種	市場	売上高(百万円)	2017-2016	2016-2015	2015-2014
コード	企業名	業種	市場	売上高(百万円)	2017-2016	2016-2015	2015-2014
6503	三菱電機	電気機器	東証1部	4238666	100.0%	100.0%	100.0%
6770	アルプス電気	電気機器	東証1部	753262	100.0%	100.0%	100.0%
3231	野村不動産ホールディングス	不動産業	東証1部	569680	100.0%	100.0%	100.0%
7947	エフピコ	化学	東証1部	172858	100.0%	100.0%	100.0%
5444	大和工業	鉄鋼	東証1部	142136	100.0%	100.0%	100.0%
5481	山陽特殊製鋼	鉄鋼	東証1部	138680	100.0%	100.0%	100.0%
6146	ディスコ	機械	東証1部	134204	100.0%	100.0%	100.0%
6806	ヒロセ電機	電気機器	東証1部	115103	100.0%	100.0%	100.0%
7520	エコス	小売業	東証1部	114764	100.0%	100.0%	100.0%
8141	新光商事	卸売業	東証1部	112458	100.0%	100.0%	100.0%

表1：売上高上位10社 / 上場企業3645社 (類似度 $P=100\%$ 、4期黒字)

5. おわりに

本研究では、有価証券報告書の「対処すべき課題」に注目し、企業が課題に対して取り組む姿勢を数値化する類似度を提案した。さらに、財務データから企業評価を行うのではなく、企業の心理を表している「対処すべき課題」のテキストデータを用いて、企業を評価する方法を提案した。その結果、企業の方針に変化があれば「対処すべき課題」に変化が現れることを示した。さらに、一貫して「対処すべき課題」が全く変わらない企業が存在し、財務データと紐づけることで、安定している企業として評価できる可能性を示した。「対処すべき課題」の類似度の算出方法を応用することにより、例えば、検査・監督を行う金融庁は、有価証券報告書の届出の義務がある上場企業を人工知能(AI)によってテキストデータの変化率を自動監視することが可能である。将来の展望として、「対処すべき課題」の類似度と財務データの関係性を見つけ、個人投資家などが注目するPERやPBRといった指標との比較を行うことで、投資を行う際の銘柄選択に活用できる可能性がある。さらに、就職活動をする学生に対しては、「売上高」、「当期純利益」、「平均年収」といった一つの指標だけでなく[6]、企業の取り組んでいる課題の重要性を指導するべきだろう。情報システムにおいて、結果のデータ(財務)だけを分析するのではなく、得たい情報を補完するデータ(「対処すべき課題」の類似度)に着目することが重要であることを示した。さらに、それらを時系列で検証することが重要である。また、今回の研究で明らかとなったのは、基幹系データや情報系データと言った情報の分類ではなく、人間の経済活動に沿った時系列で情報を整理し、数値なのか、テキストなのかという分類は重要ではなく、解を得るために必要な情報を紐づけることが重要である。なぜなら、人間中心の情報システムとは、DB構造にこだわるのではなく、人間が最も自然に理解できる状態で扱う必要があるからに他ならない。それは、「できるだけ人間の行動や心境を示す情報を時系列で示すこと」だと言えるだろう。

参考文献

- [1] 大木領太, 壺井彬, 高橋正子, “環境経営ディスクロージャ: 民間調査プロジェクトへの対応”, 情報システム学会 第11回全国大会・研究発表大会, 2015.
- [2] 斎藤祐一郎, 西森丈俊, “自然言語処理を用いた企業相関関係の取得”, 情報処理学会研究報告, Vol.2010-IOT-11.
- [3] 辻 智康, “雑誌・新聞・JST データからみるよりよい企業選択へのヒント”, 情報管理, 2015, Vol.58, No. 4.
- [4] 西野嘉之, “企業を調べれば人生が変わる! -就活やビジネスを成功に導くユーレット活用術”, 産業能率大学出版部, 2016.
- [5] グェン・ベト・ハー, 帆苺讓, 石川勉, 笠原要, “単語の類似判定のための大規模概念データベース”, 情報処理学会 論文誌, Oct. 2002, Vol. 43, N. 10, pp. 3127-3136
- [6] 小野田哲弥, 西野嘉之, 熊坂賢次 (2011), “ユーレットによる就活支援”, 2011PCカンファレンス 論文集, pp.422-425.