

〒240-8501 横浜市保土ヶ谷区常盤台 79-1

役員情報データベース構築プログラムの無償公開

— HTML形式有価証券報告書から役員情報を構造化データに変換 —

国立大学法人横浜国立大学（以下、横浜国立大学）の高須悠介准教授（国際社会科学研究院）は、企業情報のなかでも注目される機会の多い役員情報（有価証券報告書の「役員状況」）に注目し、企業が開示したHTMLファイルをCSV形式の構造化データへと変換するプログラムを開発しました。

役員データに関する大規模サンプルを構築し、効率的に収集・分析する研究の一環として、このプログラムを大学等の研究・教育機関又は公共の研究機関等に属する教員、研究員等又は学生に限定して無償公開します。

【研究の背景】

情報通信技術の進展に伴い、企業のディスクロージャーは紙ベースから電子ベースへと切り替わることで、有価証券報告書に代表される企業情報は確かに容易に入手可能となっています。しかしながら、データ分析の基本である他社比較（クロスセクション分析）、異時点比較（時系列分析）および他社・異時点の同時比較（パネル分析）の際に必要な大規模なデータを効率的に収集し、分析するには相応の知識と技術が求められており、その観点からはデータが広く一般に行き渡り、容易に活用できる状況にはなっていないように思われます。本研究では企業情報のなかでも注目される機会の多い役員情報（有価証券報告書の「役員状況」）に注目し、企業が開示したHTMLファイルをCSV形式の構造化データへと変換するプログラムの開発ならびに公開を通じ、企業データのさらなる利活用の促進を試みます。

【プログラムの概要・実施例】

本プログラムでは大きく分けて2つの処理が実行されます。1つは有価証券報告書の「役員状況」のHTMLファイルを解析することで役員状況のテーブルを構造化データ（CSVファイル）へと変換することです。役員状況のテーブルはHTMLのtableタグを用いて表現されているものの、HTMLの表現には企業間で差異が存在しており、一般的な方法（例えばpythonのpandasライブラリのread_htmlメソッドなど）では情報内容を維持した形式でデータ抽出することが困難となっています。本プログラムではHTMLの解析におけるこれら問題を（確認された範囲について）解決することでtableタグによって表現された役員状況を企業間で標準化し、各企業の各決算期について1つのCSVファイルを生成します。例として図1はトヨタ自動車の2022年3月期有価証券報告書の役員状況から抽出されたCSVファイルの一部を示しています。また、この処理やデータ内容に関しては高須（2022a）で詳しく説明されています。

役職名	氏名	生年月日	略歴年月	略歴内容	任期	所有株式数(単位:株)	注記表現	氏名別表記	役員番号	役名フラグ
取締役会長(代表取締役)	内山田竹志	19460817	196904	当社入社	(注)6	450000	(注)6		1	1
			199601	当社第2開発センターチーフエンジニア就任					1	1
			199806	当社取締役就任					1	1
			200106	当社専務取締役就任					1	1
			200306	当社専務取締役就任					1	1
			200506	当社取締役副社長就任					1	1
			201206	当社取締役副会長就任					1	1
			201306	当社取締役会長就任					1	1
取締役副会長(代表取締役)	早川茂	19530915	197704	トヨタ自動車販売㈱入社	(注)6	242000	(注)6		2	1
			200506	当社広報部長就任					2	1
			200706	当社専務役員就任					2	1
			200709	トヨタモーターノースアメリカ㈱取締役社長就任					2	1
			200906	トヨタモーターノースアメリカ㈱取締役社長退任					2	1
			201204	当社専務役員就任					2	1
			201506	当社取締役・専務役員就任					2	1
			201704	当社取締役副会長就任					2	1
			201801	㈱国際経済研究所代表取締役就任					2	1

(出所) トヨタ自動車2022年3月期有価証券報告書

図1 役員の状況のテーブルをCSVファイルに変換した実施例

もう1つの処理は生成されたCSVファイルに加え、役員の状況の注記情報を抽出することで当該企業の役員（取締役および監査役）に関する各種データ（取締役人数、平均年齢、社外取締役人数など）を生成します。取締役人数や平均年齢などは上記のCSVファイルから容易に算定することができるものの、社外役員情報については事情が異なります。多くの企業では社外役員情報を役員の状況のテーブルの注記として開示しています。そのため、本プログラムでは役員の状況の注記に該当する部分を抽出し、テキストファイルを生成した上で、社外役員に関する注記の記述内容と上記CSVファイルを突き合わせて社外役員を特定しております。例として図2は日産自動車およびトヨタ自動車の2022年3月期有価証券報告書の役員の状況から抽出された社外取締役にに関するデータを示しています。また、この処理やデータ内容に関しては高須（2022b）で詳しく説明されています。

FILENAME	SECCODE	FYE	outside_directors	num_outside_directors
10043_7201_yuho_101_20220331.txt	7201	202203	木村康_19480228,豊田正和_19490628,井原慶子_19730704,永井素夫_19540304,ベルナルデルマス_19540421,アンドリュウハウス_19650123,ジェニファーロジャーズ_19630622	7
10043_7203_yuho_101_20220331.txt	7203	202203	菅原郁郎_19570306,Sir Philip Craven_19500704,工藤禎子_19640522	3

(出所) トヨタ自動車および日産自動車の2022年3月期有価証券報告書

図2 社外取締役データの生成例

本プログラムは2004年3月期から2022年3月期までの有価証券報告書の役員の状況のHTMLファイルを解析対象としており、対象となる観測値数は69,382件です。なお、本プログラムはHTMLファイルの解析プログラムであり、HTMLファイル自体は含んでおらず、プログラム利用者が別途準備をする必要があります。本プログラムで使用するHTMLファイルはアイエヌ情報センターの「eol」から取得することを想定しております（「eol」は有償データベースであり、使用にはライセンスが必要となります）。

本プログラムは日本学術振興会科学研究費若手研究「データサイエンス手法を用いた会計データベース構築とその活用」（研究代表者：高須悠介，課題番号：21K13398）の研究結果の一部です。

【プログラムの公開期間・入手・使用方法】

申込み期間：2022年12月1日から2023年4月30日まで。

入手方法

次の URL にアクセスして、下記 1 から 3 を順次実施する。

https://takasu-labo.net/application_for_board.html

1. プログラム利用申込書、プログラム利用許諾同意書、役員データベース構築プログラム使用マニュアルをダウンロードし、プログラム利用許諾同意書に同意後、ソフトウェア利用申込書を記入して次の宛先に送信する。
送信先： 横浜国立大学 産学・地域連携課 産学連携係 sangaku.sangaku@ynu.ac.jp
2. 指定された URL からプログラムをダウンロードする。
3. 役員データベース構築プログラム使用マニュアルに従い実行する。

【今後の展開】

- ・役員情報データベースプログラムの公開先を限定しない無償公開を予定
- ・EDINET から収集可能な XBRL ファイルの活用に関するプログラムを開発予定

【関連情報】

- ・高須悠介, 2022a, 「HTML データに基づく役員情報データベースの構築」『横浜経営研究』第 42 巻第 3・4 号, 77-103 頁。
- ・高須悠介, 2022b, 「「役員 の 状 況」における注記を活用した社外役員 の 特 定」『横浜経営研究』第 43 巻第 1 号, 133-150 頁。
- ・比例縮尺財務諸表描画プログラムの無償公開

(https://www.ynu.ac.jp/hus/koho/28736/34_28736_1_1_221028122534.pdf)

本件の研究に関するお問い合わせ先

横浜国立大学 大学院国際社会科学研究院 准教授 高須悠介

E-mail : takasu-yusuke-wk@ynu.ac.jp