

EDINET API を用いたデジタルフォレンジック学習 教材向け資料生成プログラムの開発

松浦 芽生 五味 悠一郎

本研究の目的は、デジタルフォレンジック学習教材を用いた教育によって、初学者がデジタルフォレンジック技術を容易に取得でき、情報セキュリティに対する興味が向上することを明らかにすることである。デジタルフォレンジック技術の習得には、多くのデータが含まれたディスクイメージファイルが必要とするが、このファイルを講師が短時間で用意することは難しい。先行研究ではダミーデータとして、個人情報や自動生成し、汎用的な資料を手作業で作成したが、ダミーデータ生成に多くの時間を要した。その要因は、手作業で作成した汎用的な資料にあった。そこで本稿では、自動的に汎用的な資料を生成するプログラムを EDINET API を用いて開発し、この課題を解決した。このプログラムの使用により授業の準備時間を短縮でき、講師の負担を軽減できると考える。

The purpose of this research was to clarify that the use of digital-forensics learning materials can help beginning students easily learn digital-forensics technology and increase their interest in information security. To learn digital-forensics technology, a disk-image file containing a large amount of data is necessary. However, it is difficult for instructors to prepare such a file in a short time. In a previous study, automatically generated personal information and manually created generic materials were used as dummy data. However, it took much time to create the dummy data. The reason for this was due to the manual creation of generic data. We solved this problem by developing a program that automatically generates generic materials using EDINET API. We believe that the use of this program will shorten and simplify the preparation time for classes.

1 はじめに

経済産業省による調査では日本の人口減少に伴い、IT 人材供給力は低下する可能性があることが示された。IT 人材は 2015 年時点で約 17 万人不足しており、2030 年には約 79 万人になると予測されている。情報セキュリティ人材に関しては不足数が 2014 年時点で約 8.2 万人であったが、2016 年には約 13.2 万人に増加していると推計された。情報セキュリティ人材の不足数は、情報セキュリティのニーズ拡大が予想され

ることから、今後も拡大していくと試算されている。こうした状況に対応するためには、情報セキュリティへの興味を向上させる適切な教材開発が必要であると考えられる。[1]

教材の開発にあたり、教材の使用者と範囲を想定する。教材の使用者は情報技術に対する興味をもつが、セキュリティに関する知識が不足している情報初学者を想定する。教材の範囲は「デジタル化の進展や訴訟の増大により必要性が高まっている」「IoT 化の進展によりデジタルで扱う情報が膨大化し、重要性が増していく」と考えられるデジタルフォレンジックに絞込むこととする [2]。

デジタルフォレンジックとは、種々のインシデントが発生した際に、コンピュータなどの情報処理機器上に残された証拠を確保し、将来起こり得る裁判に備えるための技術や手順のことである [3]。デジタルフォレンジック技術の習得には、多くのデータが含まれた

Development of program to generate documents for digital-forensics learning materials using EDINET API

Mei Matsuura, 日本大学理工学部応用情報工学科, Department of Computer Engineering, College of Science and Technology, Nihon University.

Yuichiro Gomi, 日本大学理工学部, College of Science and Technology, Nihon University.

ディスクイメージファイルが必要とするが、このファイルを講師が短時間で用意することは難しい。先行研究ではダミーデータに含まれる情報として、個人情報情報を自動生成し、汎用的な資料を手作業で作成したが、ダミーデータ生成に多くの時間を要した。その要因は、手作業で作成した汎用的な資料にあった。本稿では、自動的に汎用的な資料を生成するプログラムを EDINET API を用いて開発し、この課題の解決を試みた。このプログラムの使用により授業の準備時間を短縮でき、講師の負担を軽減できると考える。開発したプログラムで先行研究よりも多い量のダミーデータを短時間で生成できることを明らかにする。

2 関連研究

デジタルフォレンジックツールの使用法について学習する教材が加藤・矢野・本郷（2015）と加藤・矢野・本郷（2016）および加藤・矢野・本郷・高山（2017）によって開発されている [4] [5] [6]。この研究で開発された教材は、AOS リーガルテック社製の PC 用フォレンジックツールである FinalForensics の使用法について、ファイル検索と復元を中心に学習する動画教材である。この教材は特定のツールの使用法に限定しているほか、動画教材であるため改変が難しく、広く使用することに不向きであると考えられる。

FinalForensics はデータの保全・復元・分析・検索・分析レポート作成を一括して行うことができるパソコン用の有償フォレンジックツールである [7]。有償ツールでの学習は金銭的な負担から初学者の学習における敷居が高くなるため、無償ツールを用いた教材開発が必要であると考えられる。

3 先行研究

加藤（2022）がデジタルフォレンジック学習教材に用いる個人情報自動生成プログラムの開発を行った。ダミーデータに含まれる情報として個人情報を自動生成し、汎用的な資料を手作業で作成した。作成した個人情報のファイル数が 1 で、汎用的な資料のファイル数は 5 であった。ダミーデータの生成に要した時間は 18 分 17 秒であった。個人情報自動生成プログラムの開発により短時間で誰でも容易にデータ生成

が行えることを目標としたが、ダミーデータの生成に多くの時間を要した。加藤の研究から「ダミーデータの生成時間を短縮するためには、手作業で作成した汎用的資料を自動的に生成する必要がある」という課題が明らかになった。 [8]

4 ダミーデータの生成方法

4.1 EDINET

本研究では大量のデータを手軽に入手できるため、金融庁が公開している電子開示システム「EDINET」を使用した [9]。「EDINET」では「有価証券報告書」「半期報告書」「四半期報告書」「大量保有報告書」「臨時報告書」「その他の書類種別」が提供されている。EDINET は API が提供されており、プログラムを介して効率的に開示情報を取得できる [10]。EDINET API は「書類一覧 API」と「書類取得 API」がある。書類一覧 API では「メタデータのみ」または「提出書類一覧およびメタデータ」を取得できる。

本稿ではダミーデータとして、EDINET で公開されている資料のうち、今後の研究で必要となる「有価証券報告書」と「四半期報告書」を取得した。

4.2 ダミーデータ生成プログラム

ダミーデータ生成プログラムは Python3.10.12 で開発し、開発ツールは「Google Colabratory」を用いた。「Google Colabratory」ではランタイムのタイプが変更できるが、本稿では使用上限が GPU より大きい CPU を用いた。「Google Drive」の「マイドライブ」内に「edinet」というフォルダを作成し、edinet フォルダ内に生成した資料が保存されるようにした。

EDINET API の書類一覧 API は、リクエストパラメータとして日付のみ指定ができる。金融商品取引法第二四条によって有価証券報告書の提出期限は事業年度終了後 3ヶ月以内、四半期報告書の提出期限は四半期終了後 45 日以内と定められている [11]。事業年度を 3 月までとしている企業が多く、リクエストパラメータとして指定する日付によっては有価証券報告書が 1 つも提出されていない場合があったため、開始日と終了日を指定してその期間に提出された書類を取得するプログラムとした [12]。開発したプログ

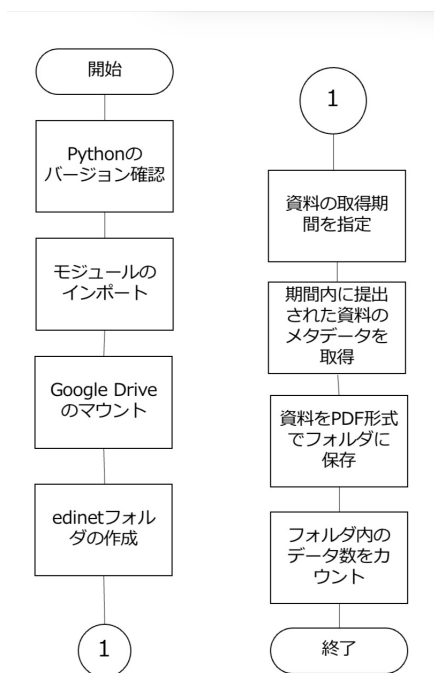


図1 開発したプログラムのフローチャート

ラムのフローチャートを図1に示す。

EDINET APIの書類取得APIはエンドポイントで書類管理番号を指定する必要がある。エンドポイントとはWebサービスやアプリケーションにおいて、外部のクライアントがAPIとの通信を行うための特定のURLで、リクエストが実行される場所を指す[13]。書類一覧APIを用いて指定した期間内に有価証券報告書が提出されていた場合、「提出者名」「提出書類概要」「docID」「日付」をリストに格納して書類管理番号を取得した。docIDは書類管理番号を示す。リストとはPythonで一覧となっているデータを扱うための型で、変数を代入したり取り出したりできる。

5 評価方法

本稿では開発したプログラムが先行研究よりも短い時間でダミーデータを生成できることを以下の方法で検証した。

- ダミーデータを先行研究の生成所要時間と同じ18分17秒生成した場合に、先行研究よりも多い量のダミーデータが生成できる。

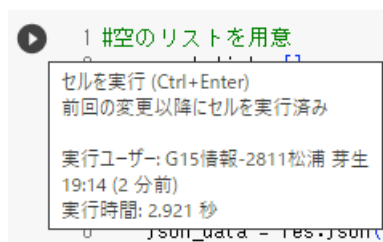


図2 Google Colaboratory セルの実行時間

	A	B	C	D	E
1		提出者名	提出書類概要	docID	日付
2	0	ヨシコン株式会社	有価証券報告書-第5 S100RB4M		2023-07-03
3	1	株式会社ダイサン	有価証券報告書-第4 S100RCCH		2023-07-06
4	2	株式会社ヤガミ	有価証券報告書-第5 S100RDQS		2023-07-13

図3 有価証券報告書提出者一覧

- ダミーデータを先行研究と同じファイル数の5生成した場合に、先行研究よりも短い時間で生成できる。

ダミーデータの生成時間が18分17秒よりも長くなるようにし、18分17秒でプログラムを停止させた。このとき edinet フォルダ内に保存されているファイル数をカウントし、ダミーデータ生成数とした。

生成所要時間はプログラムの開始時間から終了時間までを計測した。「Google Colaboratory」ではセルごとに実行時間が表示されるため、すべてのセルの実行時間を合計することで、生成所要時間とした。「Google Colaboratory」で表示されるセルごとの実行時間を図2に示す。

6 結果と考察

指定した期間内に有価証券報告書が提出されていた企業の「提出者名」「提出書類概要」「docID」「日付」が格納されたリストを、表形式で表示した結果を図3に示す。

指定した期間内に四半期報告書が提出されていた企業の「提出者名」「提出書類概要」「docID」「日付」が格納されたリストを、表形式で表示した結果を図4に示す。

リストに格納された docID を書類取得APIのエンドポイントに入力し、有価証券報告書と四半期報告書

	A	B	C	D	E
1		提出者名	提出書類種別	docID	日付
2	0	株式会社エドシーホー	四半期報告書-第74期	S100RCNH	2023-07-06
3	1	株式会社ケーヨー	四半期報告書-第86期	S100RCGA	2023-07-06
4	2	わらべや日清ホールデ	四半期報告書-第90期	S100RCE3	2023-07-06
5	3	テクノアルファ株式会社	四半期報告書-第34期	S100RCBS	2023-07-06
6	4	スターマインホールデ	四半期報告書-第26期	S100RCIS	2023-07-06
7	5	株式会社 技電産総研	四半期報告書-第63期	S100RBKT	2023-07-06
8	6	トセイ株式会社	四半期報告書-第74期	S100RCIEB	2023-07-06
9	7	日本ファイルコン株式会社	四半期報告書-第124期	S100RCPH	2023-07-06
10	8	株式会社エスコーエ	四半期報告書-第173期	S100RCSE	2023-07-06
11	9	株式会社メディアエ	四半期報告書-第209期	S100RCXI	2023-07-06
12	10	株式会社MS&Conso	四半期報告書-第12期	S100RCIR	2023-07-06

図 4 四半期報告書提出者一覧

を PDF 形式で Google Drive に保存した結果を図 5 に示す。

有価証券報告書を 5 取得するプログラムを実行した結果、得られた各セルの実行時間と合計実行時間を表 1 に示す。

プログラムを 15 回実行したところ、実行時間の平均値は 37 秒であった。

18 分 17 秒間ダミーデータ生成プログラムを実行した結果、生成できたダミーデータのファイル数を表 2 に示す。

プログラムを 10 回実行したところ、生成できたファイル数の平均値は 744 であった。

7 まとめと今後の課題

EDINET API を用いてデジタルフォレンジック学習教材に用いる汎用的な資料生成プログラムを開発した。ダミーデータ数 5 のとき、実行時間は先行研究の 18 分 17 秒から 17 分 40 秒早い 37 秒で生成することができた。プログラムを 18 分 17 秒間実行させたところ、先行研究の 5 よりも 739 多い 744 のダミーデータを生成できた。よって、本稿の目標である「開発したプログラムで先行研究よりも多い量のダミーデータを短時間で生成できること」は達成できた。

今後の課題として、誰でも容易に汎用的な資料生成が行えることを目指し、開発したプログラムの手順書を作成する。開発したプログラムと手順書を用いて被験者実験を行い、容易に資料生成が行えることを検証する。また、汎用的な資料の種類を増やすため、有価証券報告書から「財務諸表」「貸借対照表」「損益計算表」、四半期報告書から「年間売上グラフ」を自

動生成できるように資料生成プログラムを改良する。

参考文献

- [1] 商務情報政策局 情報処理振興課. "IT 人材の最新動向と将来推計に関する調査結果". 経済産業省. 2016-06-10. https://www.meti.go.jp/shingikai/economy/daiyoji_sangyo_skill/pdf/001_s02_00.pdf. (参照 2023-2-26).
- [2] 特定非営利活動法人デジタル・フォレンジック研究会. 基礎から学ぶデジタル・フォレンジック 入門から実務での対応まで. 安富潔・上原哲太郎編. 日科技連出版社. 2019. 236p. ISBN978-4-8171-9668-2.
- [3] 上原哲太郎, 桜庭信之, 白濱直哉, 野崎周作, 八槇博史, 山本清子. デジタル・フォレンジックの基礎と実践. 佐々木良一編. 東京電機大学出版局. 2017. 290p. ISBN978-4-501-55560-3.
- [4] 加藤大希, 矢野祐樹, 本郷節之. "サイバー犯罪対処能力の向上に資するデジタルフォレンジックツール活用マニュアルの作成". 工学教育研究講演会講演論文集. 2015.
- [5] 加藤大希, 矢野祐樹, 本郷節之. "デジタルフォレンジックツール使用法学習のための電子教材の開発". 工学教育研究講演会講演論文集. 2016.
- [6] 加藤大希, 矢野祐樹, 本郷節之, 高山純. "デジタルフォレンジックツール使用法学習のための電子教材の導入". 工学教育研究講演会講演論文集. 2017.
- [7] "AOS ファイナルフォレンジック". LegalSearch. <https://www.fss.jp/aos-%E3%83%95%E3%82%A1%E3%82%A4%E3%83%8A%E3%83%AB%E3%83%95%E3%82%A9%E3%83%AC%E3%83%B3%E3%82%B8%E3%83%83%E3%82%AF/>. (参照 2023-2-26).
- [8] 加藤蓮, 五味悠一郎. "デジタルフォレンジックに用いるイメージファイル自動生成システムの開発". 令和 4 年度日本大学理工学部学術講演会予稿集. <https://www.cst.nihon-u.ac.jp/research/gakujutu/66/pdf/G-27.pdf>. (参照 2023-08-07).
- [9] "EDINET について". 金融庁. 2022-11-15. <https://www.fsa.go.jp/search/20130917.html>. (参照 2023-03-16).
- [10] "EDINET API". e-Gov ポータル. <https://www.e-gov.go.jp/digital-government/api/edinet.html>. (参照 2023-03-16).
- [11] "金融商品取引法". e-Gov 法令検索. https://elaws.e-gov.go.jp/document?lawid=323AC0000000025_20230614_505AC0000000053. (参照 2023-08-10).
- [12] "(3) 決算期別法人数". 国税庁. <https://www.nta.go.jp/publication/statistics/kokuzeicho/hojin1997/11.htm>. (参照 2023-08-16).
- [13] shimizu chioka. "徹底解説: API エンドポイントとは?それをテストする方法は?". API-DOG. 2023-06-14. <https://apidog.com/jp/blog/api-endpoint-and-its-testing/>. (参照 2023-08-10)

表 1 実行時間

	セル	1	2	3	4	5	6	7	合計
実行回数		実行時間[秒]							
1		0.296	0.944	22.919	0.012	5.119	0.025	4.733	34.048
2		0.011	0.898	21.542	0.028	6.117	0.015	7.102	35.713
3		0.287	0.675	23.391	0.024	6.659	0.022	7.116	38.174
4		0.235	0.828	24.007	0.011	9.225	0.025	7.825	42.156
5		0.004	0.840	32.156	0.031	9.247	0.026	11.119	53.423
6		0.004	1.334	24.008	0.005	5.497	0.029	7.255	38.132
7		0.518	0.615	22.635	0.018	3.520	0.007	5.860	33.173
8		0.005	1.007	24.337	0.028	4.894	0.015	5.014	35.300
9		0.011	1.394	21.143	0.021	3.715	0.012	6.450	32.746
10		0.014	0.889	23.431	0.006	7.118	0.025	5.006	36.489
11		0.288	1.093	24.031	0.025	5.774	0.016	7.368	38.595
12		0.268	0.971	23.264	0.020	5.085	0.023	3.871	33.502
13		0.431	0.377	22.979	0.016	4.625	0.021	3.548	31.997
14		0.004	0.859	20.452	0.027	4.371	0.004	6.186	31.903
15		0.942	0.511	28.219	0.029	3.298	0.006	4.043	37.048

表 2 生成ファイル数

実行回数	ファイル数
1	820
2	793
3	778
4	713
5	800
6	690
7	756
8	740
9	838
10	512
平均	744

マイドライブ > edinet ▾

× 1個選択中 👤 ↓ 📄 🗑️

名前 ↑

PDF	S100RB4M有価証券報告書.pdf
PDF	S100RBKT四半期報告書.pdf
PDF	S100RCCH有価証券報告書.pdf
PDF	S100RCE3四半期報告書.pdf
PDF	S100RCEB四半期報告書.pdf
PDF	S100RCGA四半期報告書.pdf
PDF	S100RCIR四半期報告書.pdf
PDF	S100RCIS四半期報告書.pdf
PDF	S100RCNH四半期報告書.pdf
PDF	S100RCPH四半期報告書.pdf
PDF	S100RCS9四半期報告書.pdf
PDF	S100RCSM四半期報告書.pdf
PDF	S100RCXI四半期報告書.pdf
PDF	S100RDQS有価証券報告書.pdf

図 5 Google Drive への保存結果