

令和 2 年 6 月 18 日現在

機関番号：11301

研究種目：基盤研究(B)（一般）

研究期間：2017～2019

課題番号：17H02065

研究課題名（和文）東日本大震災アーカイブから自然災害アーカイブへの転換に関する研究

研究課題名（英文）Conversion from Great East Japan Earthquake archive to natural disaster archive

研究代表者

柴山 明寛 (shibayama, akihiro)

東北大学・災害科学国際研究所・准教授

研究者番号：80455451

交付決定額（研究期間全体）：（直接経費） 13,000,000円

研究成果の概要（和文）：本研究では、東日本大震災アーカイブから自然災害アーカイブへの転換を行うために、歴史災害資料及び風水害などの自然災害記録の整理分類方法、自然災害全般を扱うためのメタデータスキーマ、大量の資料を整理分類するための自然災害資料の自動分類方法を明らかにした。さらに、自然災害デジタルアーカイブのパイロットシステムの構築を行い、資料を効率的に探すためのユーザインターフェース及び自然災害資料の自動分類の仕組みを構築した。

研究成果の学術的意義や社会的意義

本研究では、自然災害アーカイブの課題である風水害などの高頻度で発生する自然災害の資料増大に対して半自動分類化を可能にしたことで大幅な軽減にも繋がると考えている。最後に自然災害アーカイブの知識インフラが構築できることで、今まで見落としていた災害資料の発見に繋がりを、今後の防災・減災研究の深化と発展及び防災研究者の資料を探す手間の負担軽減に寄与することができると考えている。

研究成果の概要（英文）：Concerning the transition Great East Japan Earthquake Digital Archive” to the “Disaster Digital Archive,” this research analyzed and tested methods to organize and classify various records of natural hazard induced disasters such as storms and floods; developed metadata schemas to deal with general natural hazard induced disasters, and clarified techniques for automatic classification, categorization and organization of large quantities of materials on natural hazards. Moreover, a pilot system for the Disaster Digital Archive has been built, a user interface to enhance efficient search on disaster-related content and a system for an improved automatic categorization of data on natural hazard induced disasters have been created.

研究分野：災害情報学

キーワード：自然災害 デジタルアーカイブ

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

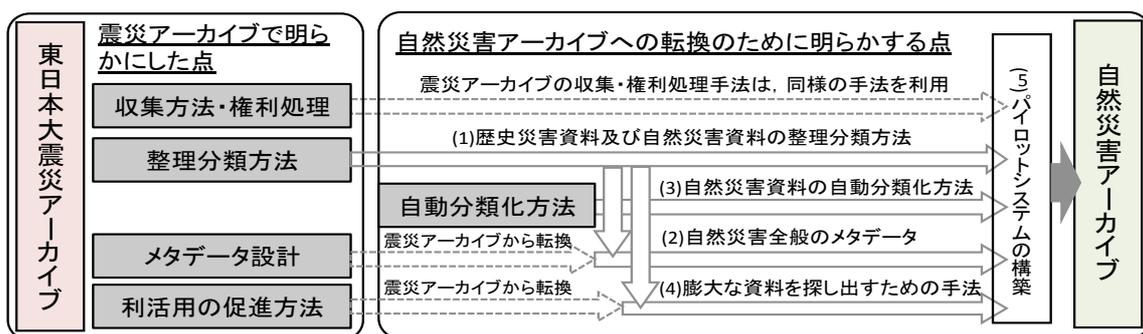
研究代表者は、東北大学の東日本大震災アーカイブプロジェクト「みちのく震録伝」や河北新報社、宮城県多賀城市「たがじょう見聞憶」、岩手県等のアーカイブ構築等の実践と研究、震災アーカイブに関わる研究活動、震災アーカイブの構築のためのガイドライン作成、公的機関での有識者会議での議論等から、東日本大震災アーカイブに内包される教訓や知見は、地震災害に限らず自然災害全般の対策・対応に対して有効に活用することができる。しかしながら、自然災害アーカイブとしては、個々の自然災害の対応・対策の記録及び歴史的な災害の教訓や知見などの記録資料などが欠如している。そこで過去の自然災害資料を震災アーカイブに加えることで、自然災害に対応した知識インフラの構築が可能になると考える。更に、複数の災害事象をアーカイブに加えることで、災害事象に囚われず横断的に比較検討や災害対応・対策の包括的な検討などに利用でき、防災研究の深化と進化に繋がるとの考えに至った。

研究代表者が執筆に関わった総務省「震災関連デジタルアーカイブ構築・運用のためのガイドライン」及び岩手県「震災津波関連資料の収集・活用等に係るガイドライン」において、震災記録の収集から権利処理、整理分類、利活用の方法までを明らかにした。しかしながら、自然災害アーカイブは、震災アーカイブの手法をそのまま活用することは困難である。例えば、東日本大震災のアーカイブのサイトの多くは、震災発生後からの記録が扱われることがほとんどであり、繰り返し起こる地震災害の記録や過去の様々な自然災害の記録が扱われていなく、過去の災害記録の整理分類方法が確立されていない問題がある。次に、災害記録の書誌情報となるメタデータは、震災アーカイブに特化した仕様となっており、自然災害に対応するためのメタデータが明らかになっていない。更に、東日本大震災アーカイブでは、数百万に及ぶ震災の記録を手作業で判断から整理分類、メタデータの付与までの作業を行い、構築までに多くの費用と時間が必要となっている。震災と比較して高頻度に発生する膨大な自然災害記録を常にアーカイブ化することは、費用面、作業人員に問題がある。最後に、震災アーカイブにおいても課題である膨大な記録資料を簡便に見つける技術が無いことである。現状、ユーザの知識による検索方法に頼っており、更に自然災害までが含まれることで複雑化し、より専門性の高い知識が必要となる。

これからわかる通り、東日本大震災の震災アーカイブからの自然災害アーカイブへの転換には、震災アーカイブの培われてきた方法論だけでは自然災害アーカイブへの転換は困難であり、新たな自然災害アーカイブ構築のための方法論が必要となる。

2. 研究の目的

本研究では、自然災害に関係する歴史災害資料及び近年の台風や豪雨災害、火山噴火などの自然災害記録の整理分類方法を明らかにする(図1の(1))とともに、震災アーカイブのメタデータを拡張した自然災害全般のメタデータの設計(図1の(2))を実施し、全世界に対応が可能なメタデータの標準化を目指す。そして、大量に記録の整理分類を自動化するために自然災害資料のテキストマイニング及び画像分類手法による自動分類化方法(図1の(3))を明らかにする。更に自然災害アーカイブを効率的かつ効果的に探し出すために時間軸を重視した資料関連性や資料同士の連携性などの新たな整理分類方法(図1の(4))を明らかにする。最後に、これらの方法を実証するために、自然災害アーカイブのパイロットシステムの構築(図1の(5))を行う。



3. 研究の方法

研究方法として、(1)自然災害に関係する歴史災害資料及び自然災害資料の整理分類方法、(2)自然災害全般のメタデータの設計、(3)自然災害資料の自動分類化方法、(4)膨大な自然災害資料を効率的かつ効果的に探し出すための手法、(5)自然災害アーカイブのパイロットシステムの構築、の手順で進める。

(1) 自然災害に関係する歴史災害資料及び自然災害資料の整理分類方法

既存の資料分類の方法としては、国立国会図書館分類表(NDLC)や日本十進分類法(NDC)、アメリカ議会図書館分類表などの図書館分類がある。図書館分類は多岐に渡り、災害関係の項目も多くあるものの、自然災害を取り扱うには項目が少ないことが言える。そのため、各自治体が策定している地域防災計画の項目及び各省庁が自然災害発生時に逐次報告される災害報の項目に着目し、整理分類のための項目の洗い出しを行う。

(2) 自然災害全般のメタデータの設計

既存の東日本大震災アーカイブの各団体が作成しているメタデータを収集し、メタデータの特徴を抽出する。2013年3月に国立国会図書館がメタデータスキーマを公開してから既に3年が経過しており、独自性を持った震災メタデータが存在するためである。次に、震災メタデータから自然災害メタデータに拡張を実施する。拡張に際しては、研究課題(1)で行った整理分類の項目及び自然災害資料の特徴を洗い出し、必要項目を付記する。

(3) 自然災害資料の自動分類化方法

自然災害資料には、映像、文章、画像、音声の4つの記録媒体が存在する。まず、文章記録においては、テキストマイニングにより解決を図る。次に、画像記録に関しては、画像認識技術を用いることで解決を図る。現在、画像認識技術の精度は飛躍的に向上しており、教師データを用いた機械学習法などを用いて検証を行う。映像及び音声の資料の自動化については、災害資料の絶対量が少ないことから研究課題から対象外とした。

(4) 膨大な自然災害資料を効率的かつ効果的に探し出すための手法

膨大な自然災害資料を検索するには、ある程度の防災知識を必要とする。そのため、知識が無くても必要な資料を探し出す手法が必要となる。そこで、災害事象の時間や空間での検索方法、語彙での検索方法などを検討する。

(5) 自然災害アーカイブのパイロットシステムの構築

上記の(1)から(4)を実装したパイロットシステムの構築を行う。構築には、既存のみちのく震録伝の震災アーカイブシステムを拡張し、開発期間と費用の軽減を図る。

4. 研究成果

(1) 自然災害に関する歴史災害資料及び自然災害資料の整理分類方法

自然災害資料を整理するため単位として、個別資料と資料群として分けることが必要であること、資料群は統一したキーワード付与し資料の関連性を持たせる必要性があることがわかった。国内外の自治体や防災関係機関が取り扱う自然災害関係資料は、年度単位や年単位、事業単位など個別の資料をまとめられていることが多い。例えば、復興計画資料では、復興計画を決めるための会議毎の資料、計画策定に必要なローデータ、現地写真などがある。個々の資料でも意味をなす資料もあるが、現地写真などはどのような目的で撮影されたものなのかがわからないことがある。そのため、資料群がわかる統一したキーワード付与、個々の資料内容のキーワードを合わせて付与することで、整理分類が容易になることがわかった。

資料分類方法について、既存の日本十進分類法(NDC)等の適用の検討を行ったが、分類項目に合わせて資料を正確に分類することは難しく、また、キュレータの自然災害関連の知識量によって分類精度が影響する。そのため、複数の分類項目に該当する場合も多く見られることから分類項目を設けず、キーワード付与を無制限にすることで対応することとした。

資料に対するキーワードは、風水害や地震災害に対応するために、近年の自然災害の各省庁から発行される被害報から語彙を抽出し、良く使用される約300語の語彙をまとめた。語彙は、シソーラスのようにまとめるのではなく、階層構造に再整理した。これは、資料分類項目と同じ役目を果たし、目的に合わせた語彙と階層構造を作成することにより分類できるようにした。

(2) 自然災害全般のメタデータの設計

震災デジタルアーカイブで基準となる国立国会図書館メタデータスキーマが存在する。国立国会図書館のメタデータ項目は、項目数で300を越え、その中には使用頻度が少ない項目もあり、さらに、使用していない項目もあることがわかった。その他の震災アーカイブの16サイトを調査し、必須項目数のばらつきが見られることがわかった。いくつかの団体に対してメタデータ入力に関するヒアリングも行ったが、人的リソースの問題で必須項目以外の項目まで、その他の項目を埋める余力が無いこともわかった。メタデータの項目数は、数が多いほど、検索の幅やコンテンツ同士の連携が可能になるが、項目数の入力限界は人的リソースに異存することもわかった。

これらの結果から自然災害のデジタルアーカイブのメタデータとしては、資料内容を示す項目(表題、要約、キーワード、作成日、場所、資料種別、資料群情報)、資料提供の項目(著作権者及び提供者の氏名・団体・連絡先等、著作権情報、提供日)、災害事象の項目(災害事象名、発生日)、管理項目(資料固有URI、変更履歴、言語、ファイル形式)を基本項目とする。災害事象は、正式名称や通称名称、国際名称など数多く存在するため、災害事象を管理する辞書を作成することが必要である。

(3) 自然災害資料の自動分類化方法

自然災害資料の文章記録と画像記録の2種類について自動分類化の研究を実施した。文章記録の自動分類化については、文章記録から文章の抽出し、(1)で作成した語彙集から該当する語彙を見つけ出しキーワード付けするプログラムの開発を行った。その結果、近年の各省庁の災害報に関しては、適切にキーワードを付与できることがわかった。しかしながら、災害毎に増える新たな事業名や行政組織内の部署名称の変更などには対応方法については、今後の検討課題とした。次に、画像記録の自動分類化には、2つの機械学習のシステムを利用した。一つ目は、無償のOSSライブラリKerasをGoogle社のTensorFlowをエンジンとして用い(以下、TensorFlow)、二つ目は、有償のIBM社のWatson Visual Recognitionを用いた(以下、WatsonVR)。学習用データは、近年の災害写真を用いて検証を行った。その結果、TensorFlowでは学習用データを数

多く用意できる場合には学習効果も高く、画像の自動判定精度が上がるということがわかった。しかしながら、災害写真が少ないものもあり、十分な学習ができないものもあった。WatsonVR は、学習用データの最小単位が 10 画像から学習が可能であり、学習用データが少なくてもある程度の判定精度が得られることがわかった。これらの結果から自動判定がある程度は可能であることがわかった。しかしながら、「陸閘」などの特殊な構造物の学習については、建物の画像を陸閘と誤認識するケースも見られ、学習データ数の不足の問題等があることもわかり、100%の精度での判定結果を得ることは現状では困難であることもわかった。そのため、パイロットシステムでは、判定結果の一覧を表示し、選択できるように改良を行った。

(4) 膨大な自然災害資料を効率的かつ効果的に探し出すための手法

膨大な自然災害資料を探し出すためには、防災や災害の知識が必要不可欠であり、知識量に応じて目的のものを見つけ出す確率が高くなる。特に日常的には使用されない語彙も多いことから、検索語彙を適切に選ぶことができない。そのため、上述の(1)で作成した 300 語の語彙をまとめ、階層構造化を行った。階層構造化は、地方自治体職員や防災関係機関職員の利用者を想定して、各省庁や県の被害報の項目を基本として作成を行った。今後は、学校の防災教育向けに語彙の構造化を図る。

(5) 自然災害アーカイブのパイロットシステムの構築

図 1, 2 に示す自然災害アーカイブのパイロットシステムの構築を行った。パイロットシステムの構築に当たり、既存のシステムを改良して構築した。図 1 左のトップページには、検索窓及びキーワードマップの 2 つの検索方法を選択できるようにした。キーワードマップを選択すると、図 1 右のように語彙の階層構造化したものをマインドマップ形式で表現した。マインドマップ上の語彙をクリックすると図 2 左のように検索結果が表示される。さらに絞り込みを行う場合は、左のメニューバーより絞り込みを行えるようにしている。



図 1 トップページ (左), キーワードマップ (語彙の階層構造化) (右)

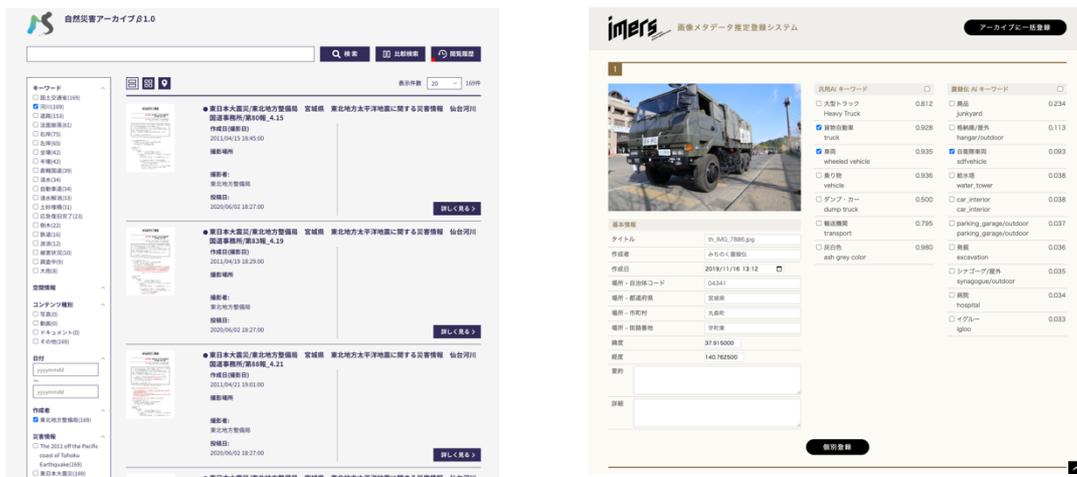


図 2 検索結果画面 (左), メタデータ推定登録システム (右)

次に、自然災害資料の自動登録方法として、図 2 右の「画像メタデータ推定登録システム」を Web システムとして作成した。登録したい画像をドラックアンドドロップすると、画像内の EXIF 情報が基本メタデータとして自動的に表示され、さらに、当該画像を自動分析し、キーワードの候補が自動表示される。画像のキーワードの候補は、WatsonVR と TensorFlow の 2 つの分析結果

を載せるようにした。これは、現状では自動判定の精度が 100%では無いこと、学習データが少ないものについては WatsonVR の結果も表示ができるようにしたためである。画像のキーワードの候補から該当する語彙を選択し、その他の基本項目を入力することでアーカイブへの登録が行われる。

(6) まとめ

本研究では、東日本大震災デジタルアーカイブから自然災害デジタルアーカイブに転換するために、震災デジタルアーカイブの既往研究を応用して資料分類方法及びメタデータ設計方法を明らかにし、自然災害資料の自動分類方法や新たな検索方法を組み入れたパイロットシステムを構築した。今後は、本研究成果を元に実際の自然災害デジタルアーカイブの構築現場での実践をしたいと考えている。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 2件）

1. 著者名 Boret Sebastien Penmellen, Shibayama Akihiro	4. 巻 29
2. 論文標題 The roles of monuments for the dead during the aftermath of the Great East Japan Earthquake	5. 発行年 2018年
3. 雑誌名 International Journal of Disaster Risk Reduction	6. 最初と最後の頁 55～62
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1016/j.ijdrr.2017.09.021	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 柴山 明寛、北村 美和子、ボレー セバスチャン、今村 文彦	4. 巻 1
2. 論文標題 [A02] 近年の震災アーカイブの変遷と今後の自然災害アーカイブのあり方について	5. 発行年 2017年
3. 雑誌名 デジタルアーカイブ学会誌	6. 最初と最後の頁 13～16
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.24506/jsda.1.Pre_13	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 北村美和子, 村尾修, 柴山明寛
2. 発表標題 東日本大震災後のコミュニティアーカイブの活動：仙台市荒浜地区を一例とした報告
3. 学会等名 デジタルアーカイブ学会第2回研究大会
4. 発表年 2018年

1. 発表者名 柴山明寛, ボレーセバスチャン, 北村美和子
2. 発表標題 震災アーカイブの利活用を促進するためのメタデータスキームについて
3. 学会等名 平成29年度東北地域災害科学研究集会
4. 発表年 2017年

〔図書〕 計1件

1. 著者名 今村文彦、鈴木親彦	4. 発行年 2019年
2. 出版社 勉誠出版	5. 総ページ数 208
3. 書名 災害記録を未来に活かす	

〔産業財産権〕

〔その他〕

図書の「災害記録を未来に活かす」は、第1部第1章の部分を柴山明寛が執筆を行った。
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	川島 秀一 (kawashima syuichi) (30639878)	東北大学・災害科学国際研究所・シニア研究員 (11301)	
研究分担者	蝦名 裕一 (ebina yuichi) (70585869)	東北大学・災害科学国際研究所・准教授 (11301)	
研究分担者	ボレー セバスチャン (boret sebastien) (70751676)	東北大学・災害科学国際研究所・准教授 (11301)	

6. 研究組織 (つづき)

	氏名 (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	有働 恵子 (udo keiko) (80371780)	東北大学・災害科学国際研究所・准教授 (11301)	