

シンプルなメタデータが付与された東日本大震災アーカイブの 写真資料のための時空間情報を利用したコンテンツ集約手法

横山 雄哉^{†1} 積 佑典^{†2} 三原 鉄也^{†3} 永森 光晴^{†3†4} 杉本 重雄^{†3}

筑波大学情報メディア創成学類^{†1} 筑波大学大学院図書館情報メディア研究科^{†2}

筑波大学図書館情報メディア系^{†3} 知的コミュニティ基盤研究センター^{†4}

1. はじめに

東日本大震災の発災後、写真を中心として行政文書や動画等、多種の資料を収集したデジタルアーカイブ(震災アーカイブと呼ぶ)が作られた。本研究では、写真資料へのアクセス性向上のために、資料の構成や内容に応じて作成したまとめ(資料の集約と呼ぶ)の作成に取り組んでおり、震災アーカイブの時空間情報を中心とするメタデータを利用した集約手法を提案している。本稿では、手法の概要と、青森震災アーカイブ[1]、久慈・野田・普代震災アーカイブ[2]、みちのく震録伝[3](それぞれ青森、久慈、震録伝と呼ぶ)の3アーカイブから集約を作成した。

2. 震災アーカイブの横断的利活用

震災アーカイブでは、写真を中心とした価値のある資料を検索することが可能である。多数の震災アーカイブを横断的に利用するためのポータルとして、国立国会図書館が運営するNDL東日本大震災アーカイブひなぎく[4](ひなぎくと呼ぶ)がある。

多くの震災アーカイブが採用しているひなぎくのメタデータスキーマ(NDLKN)[5]は多様かつ大量の資料に、低コストでメタデータを作成し、多数のアーカイブのメタデータを横断的に利用できるようにするために、Dublin Coreを基盤としたシンプルな構造となっている。しかし、シンプルなメタデータの記述能力には限りがあるため、主題等による絞り込みが難しく、もともとひとまとまりの資料であっても個別の資料単位でしか扱えないといった問題が生じる。

本研究では、地理情報や時間情報といった時空間情報は機械的に付与されていると推測し、ひなぎくから収集した3アーカイブの集約に利用出来ると考えた。

本稿では、対象とした3アーカイブの時間情報、地理情報を調査し、精度やばらつきといったデータの特性を利用して集約を作成することとし、震災アーカイブに適用できる集約手法を構築、そして実際の震災アーカイブの写真資料に適用し、作成した集約について述べる。

3. 3アーカイブのメタデータ

3アーカイブから得たメタデータの調査の結果、地理情報では地名だけのものや、緯度経度データが細かく与えられたもの、時間情報では「震災後」とだけ書かれたものや「2011年3月17日12:55:43」と詳細なものがあるなど、精度が様々であることが判明した。

また、震録伝では、1つの資料に付与されている緯度経度データは1つであるが、青森、久慈では提供者と写真が撮られた場所を示す2つの緯度経度データが付与されており、ひなぎくから得たデータでは、その緯度経度データがどちらのものか判断できなかった。それ以外にも、緯度経度データが1つ、または3つ以上付与されている資料も確認できた。

4. 3アーカイブのコンテンツ集約手法

調査の結果を踏まえて、i) 青森、久慈では2つの、震録伝は1つの緯度経度データが付いている、ii) 東半球、iii) 時間情報が秒単位まで付与されている、という条件を設定し、この条件を満たす資料を集約対象資料とした。これらに対して下記の手法で集約を行った。資料数を表1に示した。(1) **地理情報**: 本手法では緯度経度データに対してGeoHex V3[6]を利用し、場所による資料のばらつきを確認した。

GeoHexとは、六角形の領域(Hexと呼ぶ)を地図上に敷き詰めたものである。緯度経度データからその資料が属するHexを識別し、集約に利用する。本手法ではあるHexにおける資料数が10件を超えた場合、集約の大きさの均一化のため

“Aggregating Photographic Resources of the Great East Japan Earthquake Digital Archives by Geo-temporal Information in Simple Metadata”

†1 Yuya Yokoyama. College of Media Arts, Science and Technology, Univ. of Tsukuba.

†2 Yusuke Seki. Graduate School of Library, Information and Media Studies, Univ. of Tsukuba.

†3 Tetsuya Mihara. Mitsuharu Nagamori. Shigeo Sugimoto. Faculty of Library, Information and Media Science, Univ. of Tsukuba.

†4 Research Center for Knowledge Communities, Univ. of Tsukuba.

表 1 本手法で対象とする資料数

アーカイブ	青森	久慈	震録伝
資料数	52,698	78,171	124,435
集約対象資料数	26,759	49,144	96,441

めに、レベルを1段階上げてHexの指定領域を小さくすることで粒度を細かくする。

(2) **時間情報:**地理情報で集約を行った後、資料に付与された時間情報を利用して集約化を進める。特に、連続して撮影された資料の集まりを識別するため、一定の単位時間を設定し、その時間での資料の出現数を確認することによって、時間軸上でのかたまりを集約とした。

(3) **作成者情報:**(1)と(2)に加えて、時空間情報を利用して作成した集約内で、作成者が複数存在した場合、作成者別に集約化を進めることにする。ただし、作成者は名前の文字列一致のみで判別している。

5. コンテンツの集約結果

上記の手法を3アーカイブに適用した結果を表2に示した。その際、時間情報の単位時間は30分とした。また、青森、久慈では、2つの緯度経度データが同じ場合は、1つのデータとして扱った。

集約あたりの資料数の平均を確認すると、青森で約11件、久慈で約14件、震録伝では約12件となった一方で、作成した集約数の内訳を確認すると、青森、久慈、震録伝それぞれの全集約の内、約26%、約25%、約10%が集約に含まれる資料数(集約サイズと呼ぶ)が1となっている。これらのうち、地理情報での集約を作成した時点で集約サイズが1のものは少なく、多くは前後に撮影された資料がなく、単独で撮影されたため、時間情報を利用して集約を行った際に集約サイズが1の集約となったことが明らかになった。

一方、集約サイズが1000を超えるような集約は青森、久慈ともに3つ確認できた。これらは同じ緯度経度データを持ち、長時間にわたって同一の作成者が撮影したものであった。

アーカイブ別に比較すると、久慈と震録伝ではほぼ同数の緯度経度データであるにも関わらず、最大集約サイズや集約サイズ100以上の集約数で違いが生まれた。これは久慈は市町村を対象としており、東北地方広域を対象とする震録伝と比べて近い地点に緯度経度データが集中したためと考えられる。また、青森、久慈では、提供者の緯度経度データが提供者によって同じ場所を指している場合も多く、その結果集約サイズが100以上のものが多数出来たと考えられる。

表 2 資料の集約結果

アーカイブ	青森	久慈	震録伝
集約対象資料数	26,759	49,144	96,441
緯度経度データ	52,545	95,703	96,441
集約数	4,812	7,077	8,188
最大集約サイズ	1,562	1,585	487
集約サイズ			
100以上の集約数	71	141	59
集約サイズ			
1の集約	1,244	1,734	962

極端なサイズの集約を適切なサイズにするためには、キーワードやタイトルなど、他のメタデータ項目を利用して関連する集約を一つにする、もしくは集約を細かくするなど、時空間情報以外の別のアプローチからの手法が必要となってくる。また、それらはアーカイブごとの特徴を踏まえたものであることが望ましい。

6. おわりに

時空間情報を利用した集約手法を実際の震災アーカイブのデータに適用し、集約を作成することができた。作成した集約が広く利用されるために、集約に対するメタデータを付与することも必要となってくる。今後は、提案した集約手法をLinked Dataに基づく主題中心の集約手法と組み合わせることで、資料に対するアクセス性が向上していくものと考えている。

謝辞

本研究は JSPS 科研費 JP16H01754 の助成による。

参考文献

[1] 青森震災アーカイブ.
<http://archive.city.hachinohe.aomori.jp/> (参照日:2017/01/11)

[2] 久慈・野田・普代震災アーカイブ. <http://knf-archive.city.kuji.iwate.jp/> (参照日:2017/01/11)

[3] みちのく震録伝.
<http://shinrokuden.irides.tohoku.ac.jp/> (参照日:2017/01/11)

[4] NDL 東日本大震災アーカイブひなぎく.
<http://kn.ndl.go.jp/> (参照日:2017/01/11)

[5] 凡例(NDL 東日本大震災アーカイブメタデータスキーマ 2016年10月版).
http://kn.ndl.go.jp/static/files/ndlkn_schema_Ja201610.pdf (参照日:2017/01/11)

[6] GeoHex. <http://geogames.net/geohex/v3> (参照日:2017/01/11)