

デジタルアーカイブ学会 第3回研究大会

我が国における地方紙のデジタル化と
活用の促進に向けた課題抽出

～法制度的・倫理的、社会的、技術的、経済的・制度的な課題について～

2019年3月16日

東京大学大学院情報学環

DNP学術電子コンテンツ研究寄付講座

平野 桃子



研究の概要

■ 研究の目的：

- 地方新聞記事のデジタル化・活用促進の障害となる法制度的・経済的・社会的・技術的な課題の整理と解決の方向性の提示
- 課題解決の具体例となる、地方紙デジタル化記事の新たな活用策モデルの提示

■ 研究体制：「地方紙デジタル化・活用研究会」

■ 検討期間： 2018年6月～

■ 研究会構成

- | | |
|--------------------|--------------------------------|
| ・植村八潮（専修大学教授） | ・松岡資明（学習院大学客員教授） |
| ・柴野京子（上智大学准教授） | ・美馬秀樹（東京大学准教授） |
| ・数藤雅彦（弁護士） | ・宮本聖二（立教大学教授・ヤフー(株)チーフプロデューサー） |
| ・時実象一（東京大学高等客員研究員） | ・柳与志夫（東京大学特任教授：座長） |
| ・丹羽美之（東京大学准教授） | ・平野桃子（東京大学客員研究員） |
| ・東由美子（東京大学特任講師） | |

問題意識

- 新聞記事のデジタル化とネット上の公開が進むにつれて、全国紙の記事と同じラインで記事が並び、全国からの地方紙記事へのアクセスが増えるという状況が生まれている。地方紙記事が大量にデジタル化され、検索されるようになれば、情報価値の高いビッグデータとして大きな利用可能性が見込まれる。
- そのような環境変化にもかかわらず、我が国では全国紙は別として、デジタル編集導入以前の過去に発行された地方紙のデジタルデータの公開と活用は進んでいない。しかも、昔の新聞原紙の劣化は日々進行しており、地方紙の将来的な活用に向けての取り組みが早急に求められている。
- 今回の発表では、2017年に地方新聞社に対して行ったデジタル化状況の調査結果及び「地方紙デジタル化・活用研究会」における議論から、地方紙のデジタル化とその活用促進に向けた課題を提示する。

調査概要

2017年2月から5月にかけて、日本新聞協会の協力を得てアンケート調査を行なった。

日本新聞協会に加盟している県紙相当の地方新聞社 全73社の発行する73紙 を調査対象とし、2017年2月に日本新聞協会からメールで各社の担当部署宛に調査票を送付した結果、送付した73社のうち47社から回答を得た。

日本新聞協会に加盟する県紙相当の地方紙	73
アンケート回答紙	47
アンケート未回答紙	26
回答率(%)	64.4

なお、アンケート調査における質問項目の大枠は、以下の4つである。

- (1)各社における原紙、縮刷版、マイクロフィルム、DVD、CD等による保存状況
- (2)オンライン上で公開されたデジタルデータに関する状況
- (3)非公開の形でのデジタルデータ化に関する状況
- (4)デジタルデータの公開基準（フリーアンサー形式）

調査結果

- 過去の新聞記事のデジタル化およびテキスト化については、想定していた通り進んでいないことが改めてわかった。
 - しかし、イメージデータ化に限定すれば、非公開の形ながらも、予想以上に進んでいたことも判明した。
 - また、多額の予算をかけてデジタルデータ化を進めたとしても、利用者をどの程度確保できるのかの見通しがつかないため、採算性の確保が地方新聞社にとって重い課題となっていることも明らかとなった。
 - そして、データの公開や保存、維持については、メタデータの付与やシステム構築の問題、人手と予算の確保に関する問題のほかに、著作権の問題など、技術的、経済的、法的対応も急務となっていることがわかった。
- ref. 東由美子,時実象一,平野桃子,柳与志夫. 我が国における地方紙のデジタル化状況に関する調査報告：デジタルアーカイブ学会誌. 2019, vol.3, No.1, p.35-40.

4つの課題

調査結果を参考に、地方紙のデジタル化とその活用を進めていくための課題・問題点を4つの項目に整理した。

- ①法制度的・倫理的課題
- ②技術的課題
- ③経済的課題
- ④社会的・制度的課題

①法制度的・倫理的課題

(1) 法制度的課題

- 地方紙の記事や広告、写真などの公開に関する、著作権やプライバシー権等の権利をいかにクリアするか。
- 権利処理費用の基準が不明確、統一化されていない。
- 通信社などの他社の提供記事の権利処理。
- 広告の取扱い。新聞社への権利譲渡を行っていないため、広告主が不明になっていることにより権利処理が難しい。

(2) 倫理的課題

- 過去の犯罪報道や前科報道等の取扱い。プライバシー権の侵害。
- 検索キーワードのマスキングにより、適切な記事を発見できない。

②技術的課題

(1) 文字認識

①OCR：

- ルビの認識については、漢字とルビが重なってしまっている場合は認識することができない。また、一般と異なった読み方をしているものの対処法がない。(例：「政府」のルビが「おかみ」)

②ディープラーニング：

- ディープラーニングによる分析では、何をコンピュータが学習しているのかがわからない。→ 原文と意味が異なってしまう可能性や、見出しのフォントや大きさが勝手に変えられてしまう恐れがあり、最終的に人手で確認作業を行わなければならないのが課題である。

(2) 構造化

- 新聞のレイアウトは新聞社によって異なり、過去の新聞紙ではページがとんでいるケースが少なくない。
- 広告の扱い方。著作権の関係で、広告を排除するような仕組みが必要。
- 記事コンテンツデータの構造化がなされなければ、全国の地方紙を横断的に効果的・効率的に利用することは難しい。→ 費用のかかる人手ではなく、自動的に行えるための技術開発が必要。

③経済的課題

- 実際のデジタル化の費用に明確な基準がない。
- 過去のデジタル化の実績がほとんど公開されていない。
- 原紙を保存するコストがかかってしまう。
- 海外の事例のように公的資金の援助も必要。

④社会的・制度的課題

- 紙の売上を重視している新聞社が多い。
- 紙面掲載・ウェブ掲載（無料）・ウェブ掲載（有料）の3パターンに分けて記事を作成しなければならなくなっている。
- 新聞の記事内容の価値判断というジャーナリズム教育や、リテラシー的な観点でのウェブニュースの読み方を教える必要がある。
- 地方人材育成や地域振興における役割を果たすことを求められている地方大学との連携。
- 過去のデジタル化記事の活用が、どのように社会的に有益・有効なものであるかの社会的共通認識が得られていない、各々の記事の重要度がわからない。
- デジタル化が進んでいないために、海外に日本の情報が伝わらず、日本が立ち遅れてしまう。
- 現代の若者にとって情報のほとんどがボーンデジタルであり、古い紙形態の記事をデジタル化すること自体を必要と認識されていない。
- 商業的に利用するために、記事コンテンツを再編集して付加価値をつけていくことが必要。
- デジタル化新聞記事の利用実態を調査・分析したものがほとんどなく、どのような利用のニーズがあるか、把握できていない。
- 新聞界全体としてのデジタル化記事利用の提供・課金共通プラットフォームが形成されていない。 9

今後の課題

- 地方新聞記事のデジタル化・活用促進の障害となる
法制度的・経済的・社会的・技術的な課題の解決の
方向性の提示
- 課題解決の具体例となる活用モデルの提示