

社会運動のウェブアーカイブス構築に向けた試論

SNS 運動の何を選び残そうとするのか

立命館大学先端総合学術研究科

山口 和紀

要旨：

本稿は SNS 運動のアーカイブス構築に向けた試論である。これまでウェブアーカイブスがどのようになされてきたのかを整理した。そのうえで、ハッシュタグ運動の先行研究をレビューし、そこで用いられている手法から、SNS 運動のアーカイブス構築においてどのようなデータが収集されるべきかを検討した。投稿データが取得されるだけでは不十分であり、投稿主のデータ等といったその投稿がどのようなネットワークのもとで投稿されたものなのかを分析しうる情報が残ることが好ましいことを論じた。最後に、SNS 運動のアーカイブスを今後検討していくにあたってのどのようなことがなされるべきかを検討した。

キーワード：

ウェブアーカイブ、ハッシュタグ運動、アーカイブス、SNS

1. はじめに

1-1 ネット上の「運動」の跡を残す

本稿の目的は、インターネット上で展開される社会運動のアーカイブス構築を検討するための足掛かりとして、SNS における投稿データの「収集」の方法についてを検討することにある。

これまで社会運動のアーカイブスで残されてきたのは「ビラ」や「機関紙（誌）」などを中心とする紙媒体の資料である。それはアーカイブスの対象とされてきた諸運動が、インターネットが普及する以前の運動であったがゆえである。

では、近年興隆した「ハッシュタグ運動」などのインターネット上の運動のアーカイブス★01はどのように構築すればよいのだろうか。インターネット上の情報資源を「記録化」することは「ウェブ・アーカイビング（以降、WA）」と呼ばれる★02。ウェブ上の情報は「紙」媒体の情報よりもなくなってしまいやすいものである。「ビラ」や「機関紙（誌）」は一たび発行され頒布されれば物理的に残るが、ウェブの情報は実体のある「もの」としては残らない。それはウェブにある情報が消えやすいということを直接には意味しないが、ここには利用可能な状態で保存するという観点から、大きな差がある。

例えば、ブログ機能を提供するサービスが終了した際、そのブログサービスを利用していたブログのほとんどが消えてしまうことがある。社会運動に関することと言えば、活動の記録が「ビラ」や「機関紙」などで残っていればそれらは「もの」として残り、後世にも参照される形で残りやすい。しかし、ブログで活動の記録を残している組織がいれば、多くの場合サービスの終了とともにその記録は消えてしまう。ブログに投稿されたデータを移行するのは時間や手間のかかる場合が多いからである。それはブログサービスだけでなく、SNS や HP にも同様のことがいえる。営利企業が提供しているサービスであれば、そのサービスが終了し、そこにあるデータがすべて消えてしまうということは当然起こり得る。「ビラ」や「機関紙」という「もの」として残ることが多かったものも、インターネットの普及によって消えやすくなってしまっていることは指摘できるだろう。

ゆえにインターネット上で展開される社会運動の記録をどのように残すかが問題となる。社会運動のウェブ・アーカイブスが、現在までに構築されてきたとは言い難い。

後述するようにウェブアーカイビングで多く用いられている手法は「バルク収集」である。これはつまり、データを選択せずにすべて集めるということである。反対に「選択的収集」と言って、なにかのトピックに対して、それを選択的に収集するということも行われているが、管見の限り「社会運動」というトピックから選択的なウェブ・アーカイブスを構築しようとする試みは世界的にもなされていないようである。つまり「社会運動」のウェブアーカイブスは、バルク収集で集められたデータのなかに、たまたま「社会運動」に関わるデータが入っているかもしれないという不安定で不透明な状況と言いうる。

ウェブアーカイビングのライフサイクルについて前田[2013]は次のように図解している。

「選定」→「収集」→「組織化」→「保存」→「公開」→（「選定」に戻る）

SNS 運動のアーカイブス構築については、各ライフサイクルで検討を進めていく必要があると考える。まず本稿で着手したいのは「収集」の段階についての検討である。そこで本稿では「社会運動のウェブアーカイブス」構築のための足掛かりとして、SNS 運動のアーカイブス構築のためにどのようなデータが収集されるべきなのかを検討する。

1-2 先行研究の整理

ウェブにおける社会運動としてまず注目できるのは「ハッシュタグ運動」に代表されるような SNS 上の社会運動だろう。SNS のアーカイビングについての研究に着目すると、国内の研究では先行研究が存在しないようである。国外では韓国の大統領府の SNS 記録のアーカイブに関する研究(DW Choi; SJ Lee; EH Youn; HJ Oh[2018])や、世界で行われているウェブアーカイビングのレビューを行った研究(Feng[2014])などがある。Vlassenroot ら[2021]はソーシャルメディアのアーカイビングを教育的に基礎づけるために、関連する文献のレビューを行っている。

直接的に SNS のアーカイブの方法を論じているのは DW Choi; SJ Lee; EH Youn; HJ Oh[2018]、Littman ら[2018]などがある。Littman ら[2018]はソーシャルメディアを研究利用しようとする

研究者のためのオープンソースアプリケーションである Social Feed Maneger の設計について論じている。

SNS 運動のアーカイピングに関しては国内外併せても、管見の限り研究が存在しない。

1-3 社会運動とウェブの関わり

便宜的に分ければ、社会運動とウェブの関係は2つある。一つは、現実空間で行われる社会運動がウェブに記録されることである。例えば、デモを行ったときにその写真を撮り、SNS にアップロードしたとする。あるいは現実空間におけるデモの呼びかけをウェブで行うということもあるだろう。これは現実空間の情報がウェブ上に記録されたということである。

もう一つは、情報の記録がウェブ上で完結しているものである。例えば、「#Metoo」運動に参加するために、そのハッシュタグを付けて賛意を示す投稿をする。これは現実空間における行動を記録したものではなく、ウェブ上で完結している——これは現実空間に影響を及ぼさないという意味ではない。

①現実空間からウェブへという流れと、②ウェブだけで完結するという2つがある。社会運動のウェブアーカイブスの構築という観点からは、どちらのアーカイブスも重要である。情報の流れの特性から、大きく2つに分けたとき何が言いうるだろうか。

現実空間で行われた運動の記録がデジタル化され、ウェブに残ることはアーカイブスという観点から有意義である。例えば、ある社会運動組織が、自らの活動の記録をサイト上に掲載し、情報を保全したとする。これはその社会運動組織が作るデジタルアーカイブスであると言える。

だが、そうしてサイトに載せているだけでは、消えてしまうことがある。ブログサービスや SNS に載せていた場合には、そのサービスの終了によって、その組織のアーカイブスは消えてしまうことになる。これは明らかに「ビラ」や「機関紙」に活動の記録を残すこととは異なる。組織であっても個人であってもこのことは言いうる。

それゆえそうした記録を集約してアーカイブスにする必要出てくる。現実の記録がデジタル化されたものは、どこかに集積されていたとしても消えやすいものである。それをラストリゾートへ移そうとすること、つまりは永続的に残そうとすることはなされてよい。この2つの段階がある。

このことは現実空間の出来事がデジタル化された（現実空間からウェブへ）情報ではなく、ウェブで完結する記録においても言いうる。「ハッシュタグ」運動へ誰かが参加した記録（個々の投稿データ）は、投稿された時点である程度消えにくいものになっている。それはある程度認められる。しかし、前述するようにサービスの終了と共に消えてしまう可能性がある。それを集めて、整理し、永続的な保存を目指すことに意義はある。

上記を整理する。社会運動のアーカイブスを考えるとき、①現実空間からウェブ（デジタル）の記録へと変換すること、②ウェブの記録をさらに消えにくい場所へと移し整理することの2つが重要であると考えられる。前者はなにも名づけられていないかあるいはデジタルアーカイ

ブと呼ばれ、後者はウェブアーカイブと呼ばれる。そのどちらもが重要であるが、本稿が議論の対象とするのは後者のウェブアーカイブであり、前者でない。

2. いまはどのようなになっているのか

2-1 SNS サービスの提供者によるデータの保存

SNS 運動において主たるプラットフォームとなってきたのはツイッターである。先行研究の多くがツイッターを対象として研究している。まずは議論をツイッターに限定する。重要なこととしてツイッター社は過去の投稿データを保存し続けている。ツイッター社がなくなる限りにおいて、それは続けられるだろう。しかし反対に言えば、ツイッター社がなくなればその膨大なデータは消えてしまうことになる。

そこで必要となるのがアーカイビングである。アーカイビングという目的においてデータを利用できるのかは定かでないが、ツイッター社は研究者に対しては過去の利用可能な投稿データすべてにアクセスできる権限を提供している。手続き上は、研究機関に所属していることが証明でき、適切な研究計画とデータ利用についての説明の提出が求められる。これは「the full-archive search endpoint」と呼ばれている。研究者であればこのエンドポイントを利用して、過去のすべての利用可能な投稿データにアクセスし、それを取得することができる。

留意しなくてはならないのは、このエンドポイントで取得できるデータは「公開されている」アカウントの投稿であり、なおかつ「削除されていない」ものに限るということである。いわゆる「鍵垢——投稿を限られたアカウントに限定しているアカウント」の投稿やすでに削除されてしまった投稿へはアクセスできないことになる。この点をどのように考えるのかは別稿に譲る。

これから次のことが言いうる。サービスを提供する側がデータを保管している。しかし、アーカイブとしてはこれで十分とはいえない。サービスが終了すれば、その保存されたデータもなくなる可能性が高いからである。また、そうした莫大なデータを保管し続ける費用を一営利企業が拠出しつづけるかどうかという問題もある。ゆえにサービス提供者が行っている保存をラストリゾート (last resort) として位置づけることはできない。

サービスの提供者が提供しているサービス提供者は投稿データを容易に利用できる形で（すくなくとも研究者には）開いている。これを用いてアーカイブを構築することが可能かどうかは検討される必要がある。現時点でアクセスしうるのは「公開された」「削除されていない」投稿データである。したがって、すぐにでも削除されたり、削除したりするなどして、データが消失する可能性もある。おそらく実際にそれは起きている。こうしたデータをいかにラストリゾートへ移すのかということが検討される必要がある。

ツイッターにおいて、公開されていない、あるいは削除された（されるべき）データを第三者が取得し保存することは、現時点ではいかなる方法でもできない。今後もそれは許されない可能

性が高い。投稿者のプライバシー権のためである。

2-2 NDLによるアーカイビング

国立国会図書館は2002年から日本国内のウェブサイトを保存する「国立国会図書館インターネット資料収集保存事業(Web Archiving Project(WARP))」を行っている。公開URLは<https://warp.ndl.go.jp/>である。WARPは「国立国会図書館インターネット資源選択的蓄積実験事業(Web Archiving Project)」として2002年に開始された。国会国立図書館インターネット資料収集保存事業[2022]に略歴が掲載されている。それによると以下のようなになる。日本国内のいくつかの公的機関と民間のウェブサイトについて、許諾を得て収集・保存・提供を行う実験的な事業だった。2006年には、公的機関の収集対象がすべての政府関係機関等に拡大され、「国立国会図書館インターネット情報選択的蓄積事業」として本格的に事業化された。2009年に国立国会図書館法が改正され、国・地方公共団体等の公的機関のウェブサイトを網羅的に収集・保存することが可能になった。翌年2010年の同法改正施行に伴い、制度収集が開始された。日本語名称が現在の「国立国会図書館インターネット資料収集保存事業」に変更された。つまり、公的機関のウェブ上の情報についての収集に限定されている。

そのうえで、WARPが提供する検索機能からURLに「twitter.com」を含む保存データを検索すると、26,704件がヒットする。Twitter.comに関するアーカイブデータ(その多くはTwitterのタイムラインのスクリーンショットなどである)の中で、もっとも古い保存データは2011年6月20日に保存された「厚生労働省」のアカウントページである。WARP事業では2011年以前のTwitterの投稿データは取得していないようである。その後、2022年までに最大で26,704件のTwitter.com内のページデータを収集していることになる。ただし、URLにTwitter.com以外のページも含まれてしまっているので、実際のアーカイブデータとしてはそれよりも少なくなる。公的機関についてはこうしたアーカイビングの試みがあり、すでに事業がなされてきた。その内容の検討については、本稿では行わない。

3. SNSが持つ特性

3-1 データを3つに分けて考える

SNSの投稿についてのデータとしては、性質ごとに大きく3つに分けて考えることができる。あくまでも便宜的な分類である。

α. 投稿内容——投稿されたテキストあるいは画像そのもの

β. 投稿者——どのような者が投稿したのか

γ. メタデータ——いつ、誰が、どこで、どのような端末から投稿したのか★03

この分類をもとに、実際に行われた研究をいくつか見ていく。そのことによって、どのようなデータが社会運動の研究の中で利用されているのかを検討する。

3-2 誰によって何が投稿されたのか

Alvin B;Tillery Jr[2019]は「What Kind of Movement is Black Lives Matter? The View from Twitter」(「ブラック・ライブズ・マターとはどのような運動なのか?—ツイッターの視点から」=筆者訳)という論文である。

扱われているデータは6つのBlack Lives Matter Movement(BLM)に関わった社会運動組織のツイートデータ(the public Twitter feeds)である。6つの組織のツイートデータを並べ、それぞれの投稿ごとに「資源動員」「情報提供」「表現」にコーディングし分析している。

ここで扱われているのは「投稿者」とその「投稿内容」のデータである。当然のようなことではあるが、誰がどのような内容のことを言ったのかということは残されるべきであろう。

3-3 α/β だけでは足りない

Modrek S;Chakalov B[2019]に「The# MeToo movement in the United States: text analysis of early twitter conversations」(「米国における#MeToo運動:運動形成期のTwitterにおける会話のテキスト分析」=筆者訳)という論文において、アメリカの#MeToo運動の分析をしている。

扱われたデータは2017年10月24日から21日に米国において投稿されたツイートのうち「Metoo」が本文に入っているツイート(12,337件)である★04。つまり「MeToo」という単語が含まれているかどうかを見るために「投稿内容」が扱われ、さらにいくつかの「メタデータ」が用いられている。

メタデータでは、ツイートをMeToo運動の形成期に限定するために投稿された「日時(timestamp)」と、米国内の投稿に絞るための「ジオタグ(geotag)」のデータが利用されている。

これをもとにすると単に「投稿内容」が保存されているだけでは、将来の利用には耐えない可能性が高い。すくなくとも、投稿日時やジオタグなどの基本的なメタデータがセットで保存されている必要がある。

3-4 自己紹介文/リツイート/いいね

鳥海はソーシャルメディアにおいて特定の話題が大きく取り上げられる「バースト現象」について、その正の側面として「バズり」があり、負の側面として「炎上」があるとする。ハッシュタグ運動を中心とするSNS運動もこうした「バースト現象」に包摂される概念だと言える。

榊・鳥海[2017]は「バースト現象」の分析技法を提案している。SNS運動のウェブアーカイビング手法を検討するうえで重要な研究だろう。榊・鳥海[2017]の手法について概観する。

榊・鳥海[2017]はまずツイートの分類を行っている。ツイートの分類について榊・鳥海[2017]が採用したのは「興味を示したユーザの類似性」によってツイートを分類する手法である。鳥海[2017:1289]によれば「ある 2 つのツイートを同時にリツイートしたユーザが複数人いた場合、2 つのツイートは共通した内容を有していると考えられる。そこで、リツイートしたユーザの重複度から類似したユーザを見つけることが可能である」。

ここで重要なのは榊・鳥海[2017]はツイートの分類に、単に投稿された本文データの自然言語処理的なクラスタリングを行っていないことである。

ツイートのクラスタリングの後、ツイートをしたユーザのコミュニティを抽出している。ある一定の期間内に、相互にメンションしあっている 2 ユーザをペアとみなし、リンクを張る。これによってコミュニケーションに基づくユーザネットワークが形成される。その後、そのネットワークに対して、密接しているリンクの集合を抽出する技法（榊・鳥海[2017]が利用したのは重みなし Louvain 法）によって、ユーザネットワークのクラスタリングを行う。

ここで生成されたユーザクラスタの特徴を理解するために、榊・鳥海[2017]はあるクラスターに属するユーザの「自己紹介文」を結合し 1 文書とした。この文書結合を異なるクラスタ同士で比較するために、文書に特徴的な単語を抽出する方法（榊・鳥海[2017]が利用したのは tf-idf 法）を用いて、特徴語を上位 50 語まで抽出する。この 50 語を Wikipedia の全記事データをインポートした検索エンジンにかけ、関連度の高い記事のタイトルをコミュニティのラベルとして用いた。

そのうえで榊・鳥海[2017]はこれらの分類手法について検証し、その有効性が高いことを示した。

榊・鳥海[2017]の提案するバースト現象の分析手法をもとに SNS 運動のアーカイビングについて検討する。当然ながら、ツイートの本文データとそれを投稿したユーザ情報が用いられる。ネットワークのクラスタリングに特徴的なのは、本文データに含まれる他ユーザへのメンション関係をノードとするグラフを作成することである。ただ、このメンションは本文データに含まれるものであるため、これは本文データに内包する情報と考えて良いため、前項までに論じたことと変わりはない。

榊・鳥海[2017]の手法でクラスタリングに用いられているのはメンションの他に「リツイート」がある。リツイート関係も本文データの一部であると見なすことができる。したがって、これもメンションと同様に本文データと考えて良い。

ユーザクラスタの生成のため、榊・鳥海[2017]の手法では投稿したユーザの「自己紹介文」も利用している。アーカイビングにおいては「自己紹介文」を保存することも有用であることが示唆できる。特にこうした投稿したユーザをクラスタリングすることは、特定のバースト現象を分析するうえで欠かすことのできない程度に重要なことである。それは SNS 運動の分析においても言うことであり、そのことから「自己紹介文」の保存も重要であると言える。

3-5 小括

ここまで、3つの論文で使われた、あるいは提案された手法について SNS 運動のアーカイビングという観点から検討を行った。ツイートの投稿本文だけの保存ではならず、そのほかのメタデータや投稿者の「自己紹介文」などの収集・保存も重要であることを示唆した。どのようなデータを収集し、保存すべきなのかについては、今後さらなる検討がなされる必要がある。本稿の役目はあくまでも試論であって、ここまでに留める。

4. 考察——ハッシュタグ運動のアーカイブス構築に向けた試論

4-1 方法論——すべて集める/選んで集める

ウェブアーカイビングには2つのアプローチが存在する。「選択的収集」と「バルク収集」である(廣瀬[2002])。選択的収集というのは、ウェブ上の情報資源の一部を選びながら収集することである。反対に「バルク収集」は非選択的にウェブ上の情報資源を収集することを言う。

廣瀬[2002]は「選択的収集」は人的にコストがかかることがデメリットであるが、書誌的なメタデータの作成によってきめ細かいアーカイブが可能になるとする。反対に「バルク収集」は収集作業のほとんどを自動化できるため、低コストで大規模なアーカイブを作成できるものの、均質性を欠く玉石混交のアーカイブになってしまうと指摘する。

これをハッシュタグ運動に援用して考えると、例えば「#MeToo」に限定して、その近傍の投稿のみを集めていくことが「選択的収集」と言える。反対に、ツイッターというプラットフォーム全体の投稿を非選択的に収集しておくことが「バルク収集」と言える。

4-2 すべて集める

すべて集めることができる、つまり利用されている SNS 全体を「バルク収集」によって収集できるとする。この場合、「収集」という限局的な段階においては、それ以上にできることはないということになる。

これは原理的には可能なことであろう。SNS を提供しているサービスはデータベースに「全体」のデータを入れているのであるから、その複製を作るということである。すくなくとも技術的には可能なことであろうし、サービスの提供元もバックアップのためにデータベースのスナップショットは取っているのだから、そのスナップショットの利用目的が異なるだけにすぎない。

例えば、国立国会図書館が SNS の提供元のデータベースのスナップショットを保存するという事業を行えば SNS 全体を収集することは可能であるように思われる。

本稿は「収集」段階について検討することが目的であるから、ひとまずは以上に留める。他の段階も視野に入れるのだとすれば、バルク収集によって「全体」を収集したとしても、それを組織化するためには結局は「選定」をすることにはなるということがある。この点をどのように考えるかは本稿では議論できない。

4-3 選んで集める

すべてを集めることができないのであれば、選択的に収集せざるを得ない。データの内容について、どのようなものが集まれば良いのかの基礎的な部分については前述したが、実際の選択的な収集においては、なにをどのような条件で集めるのかを明確にする必要がある。

例えば、特定のハッシュタグをすべて集めるという方法がある。#MeToo というハッシュタグを含んだツイートデータのみを選択的に収集することは技術的には容易である。おそらく「ハッシュタグ運動」においては適合的な手法になると考えられる。本稿で論じることができないが、ハッシュタグで選択的に収集することが、ハッシュタグ運動のアーカイビングにおいてどの程度効果的なものであり、適合しない部分がどこにあるのかは検討される必要がある。

ここで言うのはハッシュタグを含んでいるかどうかで選択的に収集を行った場合、その運動に言及しているツイートのデータをすべて取得することはできないということである。#MeToo 運動に言及していたとしても、必ずしも#MeToo をツイートに含むとは限らないからである。ハッシュタグ運動は、ハッシュタグを付けることでその運動の主意への賛同を示すというものである。必ずしもそうであるとは限らないだろうが、ハッシュタグを含むツイートは賛同的意見を含んでいる者が多いと考えられる。アーカイブスの構築という観点からは、賛意を示している者だけではなく、批判的意見もデータとして取得することが好ましいだろう。ハッシュタグで選択的に収集するという方法は、そうした偏りが生まれやすい可能性がある。これも検討されることが望ましいだろう。

このように「選んで集める」ことの技術や、そもそも何を集めるのか、集めることができるのかなど課題は山積する。今後の検討が必要である。

4-4 誰ならば何を集めうるのか

紙媒体の資料のアーカイブスは、比較的規模が大きな組織によってなされてきた側面がある。他方で、デジタルアーカイブスは原理的には一人でもできる。換言すれば、社会運動のウェブアーカイブスは大規模な拠点があり、そこに集積されていくことによって構築される必要はない。技術的に収集の技法が整備されれば（あるいは既存の手法を使えるならば）、データを集めること自体は誰にでもできる。しかも、大規模に集めることができる。データの整理も機械的に行うことができるだろう。紙媒体であれば、保存し管理するために「場所」が必要である。デジタルデータはそうした意味では「場所」を取らない。

社会運動のウェブアーカイブスがなんらかの「組織」によって、紙媒体のアーカイブスが大学や図書館/公文書館によって行われてきたのと違って、大きな規模で行われる必要はない。この前提のもとで考える必要がある。

そのうえで、誰ならば何を集めうるのかという問題は残る。国の機関が集め、その責任において保存するということと、個人が集めるということはまずことなる。

これは法的な問題と重なる。ウェブアーカイブスに関する法律が整備されれば、高度な個人情報や著作権の問題で個人には難しい部分を国がアーカイブスの中で扱うことができるようになる可能性もある。

ゆえに「何を集めるのか」という問いは、「誰が集めるのか」という問いと密接に関わりを持つ。この点からの検討もなされる必要がある。

5. 結語

本稿は SNS 運動のアーカイブス構築に向けた試論である。「収集」について論じるという限定を付けた上でも、きわめて限局的なことについてののみしか論じることができていない。本稿はどのようなことがなされるべきなのか/なしうるかを理念的に論じたのみであって、技術的・法的な課題についてはまったく触れていない。また、SNS と言ってもひとくくりにできるものではなく、ツイッターのようなテキスト主体のサービスと、インスタグラムのような画像主体のものではまったく異なるだろう。そうした議論について、今後も検討されていく必要がある。

■註

★01 電子記録のアーカイブやインターネット上の情報のアーカイブはいくつかの用語がある。ここで整理する。

デジタルアーカイブ：これは電子記録としてアーカイブすることを指している。例えば、紙媒体をスキャンして、電子記録として保存することが入る。

ウェブアーカイブ：ウェブ上の情報のアーカイブ。最初から電子記録として生み出された情報を、電子記録として保存することを指している。

インターネットアーカイブ：ウェブアーカイブと同様の意味で用いられる。ただし、クローラなどを用いて、自動で収集したウェブアーカイブのことを指すことが多いようである——有名なものでは「ウェイバックマシン」など。

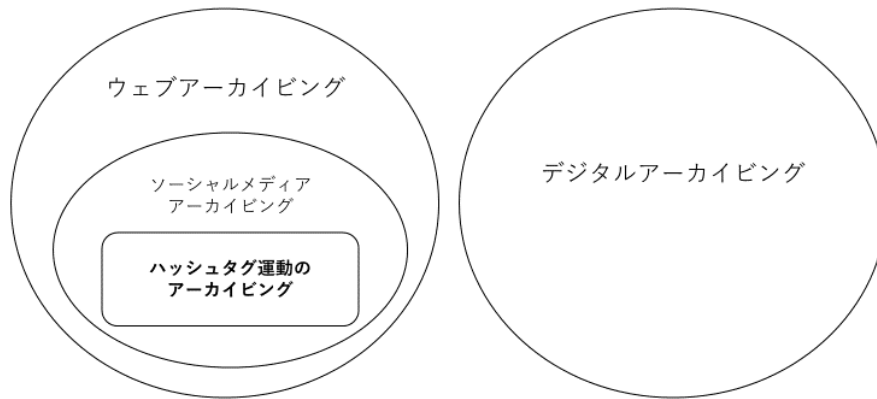


図 1. デジタル/ウェブ上のアーカイビングに関する定義の概念図

★02 ウェブアーカイビングについての定義として「ウェブ・アーカイビングとは、ウェブ上の情報資源を『記録化』し、その情報の内容と存在を空間的、時間的に安定化させることによって、インターネットのラスト・リゾートを構築しようとする試みである」(廣瀬[2003:3])がある。

★03 Twitter 社 が 提 供 し て い る API の 仕 様 書 (URL : <https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet>) によれば、Tweet Object に含まれる基本属性データとして次が挙げられている。

- ◇ツイートの作成日
- ◇ツイートの ID(整数型/文字列型)
- ◇投稿内容のテキスト
- ◇ツイートの URL
- ◇140 字の超過の真偽(ブーリアン)
- ◇ツイートが返信であれば返信先のツイート ID が整数型/文字列型で格納される。
- ◇ツイートが返信であれば返信先のユーザーのツイート ID が整数型/文字列型で格納される
- ◇ツイートを投稿したユーザーの情報
- ◇投稿時の地域データ
- ◇関連付けられた場所
- ◇引用リツイートである場合には、引用元のツイートの ID が整数型/文字列型で格納される
- ◇引用リツイートなのかどうかの真偽 (ブーリアン)
- ◇リツイートされたツイートであるかどうかの真偽 (ブーリアン)
- ◇引用リツイート数

◇返信の数

◇リツイートの数

◇いいねの数

◇センシティブな投稿かどうか

◇フィルターレベル

◇言語

これらのデータを取得することが可能である。

★04 ここで対象となっているのは「削除されていないパブリックツイート(all non-deleted public tweets)」である。反対に言えば「一度投稿はされたが削除された」ツイートや「投稿されたものではあるが、投稿者が”非公開”にしている」ものは入っていない。

この「削除」や「非公開」についてはアーカイブス構築という観点から検討されるべき事柄であるが、別稿に譲る。

■文献リスト

国立国会図書館インターネット資料収集保存事業 2022 「2022 年 8 月特集 WARP の 20 年を振り返る」, URL:<https://warp.ndl.go.jp/contents/special/special202208.html>.

廣瀬信己 2002 「消えゆくウェブを救え！——動き出すウェブ・アーカイビング」, 日本データベース協会編 『データベース』(21),9-12, URL:<http://www.asahi-net.or.jp/~ax2s-kmttn/internet/dina.html>(2022/08/06 閲覧).

廣瀬信己 2003 「国立国会図書館におけるウェブ・アーカイビングの実践と課題」, 情報処理学会 学会報告.

鳥海不二夫, & 榊剛史 2017 「バースト現象におけるトピック分析」, 情報処理学会論文誌, 58(6), 1287-1299.

前田直俊 20130518 インターネット資料収集保存事業(WARP)の10年とこれから, 日本図書館研究会情報組織化研究グループ月例研究会, <https://warp.ndl.go.jp/warp10years.pdf>.

Choi, D. W; Lee, S. J; Youn, E. H; Oh, H. J 2018 「A study on a presidential sns records management method」, Journal of Korean Society of Archives and Records Management, 18(2), 29-59.

Littman, J; Chudnov, D; Kerchner, D; Peterson, C; Tan, Y; Trent, R; Wrubel, L 2018 「API-based social media collecting as a form of web archiving」, International Journal on Digital Libraries, 19(1), 21-38.

Modrek, S; & Chakalov, B. 2019 「The# MeToo movement in the United States: text analysis of early twitter conversations」, Journal of medical Internet research, 21(9), e13837.

- Toyoda, M; & Kitsuregawa, M. 2012 「The history of web archiving. Proceedings of the IEEE」 , 100(Special Centennial Issue), 1441-1443.
- Tillery, A. B. 2019 「What kind of movement is Black Lives Matter? The view from Twitter」 , Journal of Race, Ethnicity, and Politics, 4(2), 297-323.
- Vlassenroot, E; Chambers, S; Lieber, S; Michel, A; Geeraert, F; Pranger, J; Mechant, P. 2021 「Web-archiving and social media: an exploratory analysis.」 , International Journal of Digital Humanities, 2(1), 107-128.
- Xiangjun, F. 2014 「Review of Foreign Web Archive Research」 , Information and Documentation Services, 35(6), 55-60.

A Trial Theory for the Construction of Web Archives of Social Movements

What to archive about the SNS movement?

Kazunori Yamaguchi

Abstract:

This paper is a trial argument for the construction of an archives of the SNS movement. The paper summarizes how web archives have been conducted so far. Then, we reviewed previous studies of the hashtag movement and examined what kind of data should be collected in the construction of an archive of the SNS movement based on the methods used in those studies. This paper argues that it is not sufficient to simply collect data on posts, but that it is preferable to retain information such as the data of the posters, which enables analysis of the networks under which the posts were made. Finally, the paper discusses what should be done in the future when considering the archives of the SNS movement.

Keyword:

Web Archives, Hashtag Movement, Archives, SNS