

過去を参照するツイートの検索

澄川 靖信[†] ヤフトアダム^{††}

[†] 首都大学東京大学教育センター 〒 192-0364 東京都八王子市南大沢 1-1

^{††} 京都大学大学院情報学研究科 〒 606-8501 京都府京都市左京区吉田本町

E-mail: [†]sumikawa-yasunobu@tmu.ac.jp, ^{††}adam@dl.kuis.kyoto-u.ac.jp

あらまし 歴史を理解することによって、現代社会がどのように形成されたのかの理解、実験的に評価を行うことが難しい社会的な方法論に関する議論、といった事が可能になることが知られている。Twitter 上では多くの人に知られている過去や個人史が共有されているので、各利用者の関心に沿った歴史の提示や、同じ出来事に関する様々な理解の共有、などの歴史との対話を実現することによって、歴史への理解や関心の促進が期待できる。しかし、歴史学者ではない一般の人が気軽に利用できる対話システムは実現されていない。本研究では Wikipedia から収集した過去の出来事をつぶやくチャットボットを実現する。本チャットボットは、同じ日に起きた過去の出来事を定期的につぶやくだけでなく、利用者からのメンションに応じて適切な出来事をリプライする。また、利用者からのツイートが過去の出来事に関するとき、そのツイートをリツイートして拡散する。本稿では、チャットボットの実現方法と、リツイートのために利用する分類器の訓練方法について述べる。訓練した分類器は再現率・適合率・F 値のすべてにおいて 98%を超えることを確認した。

キーワード デジタルヒストリー、パブリックヒストリー、Twitter、分類器

1 はじめに

歴史をよく知り、深く理解することは、現代社会の形成課程のより良い理解や、様々な地域や時代で生じた事柄を類推して現代社会の諸問題について考察するための足場かけとなる効果があるので [10] [1]、社会的に重要な役割を果たしている。実際、歴史の授業は多くの国で小学校から開講されている基礎的な科目の 1 つであることや、歴史的類推を促すための学習デザインに関する研究 [27] が行われていることから、歴史の重要性は広く認識されている。特に歴史学の中の研究領域であるパブリックヒストリーは「歴史と研究者以外の人との対話」と定義され [4]、その実現方法について議論されている。

ソーシャルメディアは現代で生じている事柄を報じるニュース等の情報収集のために用いられるだけでなく、過去に関する情報収集やその共有のためにも利用されている。この特徴を利用して、Twitter に投稿されたパブリックな態度を利用したアメリカ大統領選挙に関する予測 [25] といった研究だけでなく、第一次世界大戦に関するコモレーションの分析 [5] や、大規模な過去に関するツイートデータセットを構築し、そのデータセットではどのような歴史が参照されるのか、なぜそのような過去を利用者は思い出したのか、を分析した研究は行われている [23]。

本研究では、利用者との対話を促すためのチャットボットを実現する。このチャットボットは Web 上に蓄積されている歴史を収集・分析し、利用者へ共有する一方通行なコミュニケーションではなく、利用者からの問い合わせに応じた返答を行うことを目的とする。この目的を実現するために、本研究では以下の方法でのコミュニケーションが可能なチャットボット

トを実現する。

- 過去に起きた同日の出来事の共有
- 利用者が指定した日に起きた過去の出来事の共有
- 利用者が指定した固有表現（出来事・人物・組織）に関する過去の出来事の共有
- 利用者からの過去に関するツイートの拡散

利用者からのツイートが過去に関するものかどうかを判別するために、本稿では歴史に関するツイートかどうかを判別する分類器を実現する。

本稿の構成は以下の通りである。第 2 節では関連研究と本研究を比較する。第 3 節で分類器を訓練するために本研究が利用するデータセットについて述べた後、分類器の訓練方法について第 4 節で述べる。本手法の効果を第 5 節で評価した後、第 6 節でまとめる。

2 関連研究

2.1 デジタルヒューマニティ

電子化された過去の出来事に関するデータを分析する研究はデジタルヒューマニティとして行われている。特に、集合的記憶に関する研究として、一人一人のアカウントデータや政治的・文化的な組織ごとの活動記録を分析した研究 [12] や、20 世紀に発行された新聞記事を用いて時間の経過と名声の失墜の関係性を調査したもの [7]、90 年分の英語版の新聞記事 [2] や Wikipedia [9] [11]、Twitter [23] をデータセットとして、どのように過去が記憶され、どのように忘れられていくのかを調査した結果が報告されている。

2.2 短文分類

過去に起きた出来事は簡潔な記述によって文章中で言及されていることがある。例えば、地震が起きたことを報じる新聞記事が、これから取るべき対策について過去に生じた地震の結果を踏まえて議論することや、歴史上の都市に関する文章において、過去にその場所でどのような出来事が起きたのかを述べることもある。なお、このような過去の出来事は、上記のような長文の中で言及されるだけでなく Wikipedia の Current Portal¹ のような時系列に列挙されることや、関連する出来事をまとめてリスト化されている場合もある。したがって、出来事に関する文章分類の研究では、新聞記事のような十分な記述量のある文章を対象としている研究 [14] だけでなく、上述したような副次的に言及される場合や時系列に列挙される出来事はわずかな分量で簡潔に記述されている傾向があるので、効果的な特徴ベクトルを設計することによって精度を向上させる研究が行われている [20][21]。

これまでに行われた短文分類に関する研究は、文脈情報を用いるものと外部の情報源を用いるものの 2 種類に分けられる。前者の研究としては、ツイートの分類のために Twitter ユーザのプロフィールやツイートに含まれる URL とハッシュタグを用いた手法 [19]、Q&A 文章のために画像や動画を利用した単純ベイズ分類器を利用したもの [16]、Web 検索におけるユーザの目的を明らかにするために各ユーザがクリックした情報を利用したクエリ分類 [15] といった研究が行われている。一方、外部の情報源を用いた後者の研究として、Wikipedia 上でモデルを構築した LSA を適用した特徴ベクトルを基に分類器を訓練した手法 [26]、LSA [8]、や LDA [3] といった潜在意味解析を用いた分類器の訓練法を一般化した枠組みの提案 [17]、が行われている。また、明示的意味解析 (Explicit Semantic Analysis, ESA) を用いて短文に対応する Wikipedia 記事を取得し、その文章を基に分類器を訓練する手法も提案されている [24]。

2.3 パブリックヒストリー

歴史家による非研究者と歴史の対話を促す試みは古くから行われている。イタリアの歴史家である Guiccardini による市民への歴史の紹介 [13] など、15 世紀にはパブリックな役割を担う歴史家による活動が行われているが、歴史家とパブリックの間の関係性に影響を及ぼすような歴史の専門性は 19 世紀末から始まった [4]。また、Rebecca Conard は、教育以外の活動のために歴史が利用されていることを指摘しながら、現代と関係する歴史を作る価値を 20 世紀初頭に提案した [6]。歴史家の Raphael Samuel は、Ruskin 大学で歴史のワークショップを開催し、ワークショップ開催者が地域史やオーラル口述歴史の新たな実践を推進したと述べている [18]。

近年では、コンピュータを用いて、単に過去の出来事を暗記するだけでなく、その知識を活用できるように支援する研究が行われている。池尻・澄川は現代社会の問題のカテゴリを入力すると、同じカテゴリを多く持つ過去の因果関係を入力する検

索エンジンを利用した高等学校の授業デザインを提案した [27]。この研究で利用している検索エンジンは、予め研究者や教師が適切なカテゴリを過去の因果関係に付与したデータベースを作成していることを仮定している。Sumikawa and Jatowt は、カテゴリを入力すると、分類器によって付与されたカテゴリとの類似度を基に過去の出来事を出力する検索エンジンを提案し、拡張性の問題を解決している [22]。これらの検索エンジンを用いた研究は、利用者が起点となって利用者や歴史のコミュニケーションを行っていると考えられるが、歴史が起点となる両者のコミュニケーションは難しい。本研究で実現するチャットボットは、定期的に過去の出来事をつぶやくこと、利用者からのリプライに応じること、によって、両者が起点となる双方向のコミュニケーションが可能である。

表 1 データセットの統計情報

歴史ツイート数	960,540
非歴史ツイート数	960,540
歴史ツイートの平均単語数	15.46
非歴史ツイートの平均単語数	14.603
歴史ツイートの収集期間	2016/3/8 ~ 2018/7/2
非歴史ツイート数の収集期間	2019/12/6 ~ 2019/12/8
歴史カテゴリ数	6
過去の出来事数	71,374
言語数	7

3 データ収集

本節で本稿で実現するチャットボットを作成するために利用しているデータの収集方法を述べる。表 1 に収集したデータの統計情報を示す。

3.1 過去の出来事データの収集

本チャットボットが多くの人に利用されるシステムとなるように、チャットボットがつぶやく過去の出来事は読みやすさを重視するために短文で記述され、重要なものを選出する。データベースの拡張性を考慮すると、オンラインで公開されている無料の文章データセットを利用して自動的にデータを収集することを仮定する。特に、データ収集中のエラーを防ぐために、本稿では、Wikipedia に記録されている出来事を集めた。Wikipedia は、年や月日をタイトルとする記事に、各年²や月日³に生じた過去の出来事を記録している。収集データの期間は 1~2019 年、1 月 1 日 ~ 12 月 31 日までの各年・各日である。収集データは英語・日本語・ポーランド語・中国語・ドイツ語・フランス語・スペイン語の各言語版の Wikipedia に記録されている。これらのデータを収集した結果、71,734 個の出来事を収集した。

図 1 に 1862 年⁴に起きた出来事の例を示す。本稿のデータ

2 : 例 : <https://ja.wikipedia.org/wiki/1998年> or <https://ja.wikipedia.org/wiki/138年>

3 : 例 : <https://ja.wikipedia.org/wiki/2月25日> or <https://ja.wikipedia.org/wiki/3月19日>

4 : <https://ja.wikipedia.org/wiki/1862年>

1 : https://en.wikipedia.org/wiki/Portal:Current_events

ベースは、「できごと」節に記録されている記述を BeautifulSoup⁵ を用いて収集したデータから作成した。

また、出来事の潜在意味解析を実現するために、収集した出来事の記述からリンクされている Wikipedia 記事のタイトルとカテゴリも収集した。

できごと [編集]

1月 [編集]

- 1月6日 - メキシコ出兵: フランス、スペインおよび英国軍がベラクルスに至る
- 1月21日 - 文久遣欧使節出発
- 1月24日 - ブカレストがルーマニア公国の首都となる
- 1月30日 - 米海軍初の装甲艦「モニター」進水
- 1月31日 - アルヴァン・グラハム・クラークがシリウス伴星 (シリウスB) を発見

図1 Wikipedia の出来事記述の例

3.2 歴史ツイート収集

歴史に関するツイートとして、[23] で収集したツイートをを用いる。このデータセットは、ハッシュタグに基づくクローリングによって収集した歴史的なイベントや固有表現に強く関連するツイートを含んでいる。具体的な収集手順は、まず、歴史に関連するハッシュタグを研究している専門家によって選別されたハッシュタグ (**#history**、**#HistoryTeacher**、**#WmnHist** など) を収集した⁶。また、過去を参照する際に広く使用されている7個のハッシュタグ (**#throwbackthursday**、**#historyrepeating**、**#historicalevent**、**#thisdayinhistory**、**#otd**、**#onthisday**、**#time-travel**) も収集した。これらのハッシュタグを検索ワードとして Twitter 社が提供している公式 API⁷ を用いてツイートを収集した。これらの手続きは、2016年3月8日から2018年7月2日まで実施した。またツイートの収集と同時に、シードハッシュタグの数を増やすために、シードハッシュタグと共によく使用され、未だシードハッシュタグとして扱われていない歴史的なハッシュタグをブートストラップ法によって検索し、専門家による検査の後にシードハッシュタグとして追加した。最終的に、本稿で用いるデータセットは130個のシードハッシュタグと2,370,252件のツイートデータを含む。このデータセットに含まれるツイートの言語分布の上位5件は英語 (82.4%)、フランス語 (3.0%)、スペイン語 (1.7%)、ドイツ語 (1.3%)、イタリア語 (0.6%) である。

3.3 非歴史ツイート収集

本稿では Twitter API で取得可能なトレンドを含むツイートを非歴史ツイートとみなす。表2に本稿で使用したトレンド単語の一部を示す。トレンドの中には歴史に関するツイートを含む可能性はあるものの、それらの多くは歴史とは関係しないと考えられるので本稿では非歴史ツイートとして扱う。また、歴史ツイートの多くは英語なので、収集したトレンド情報は、アメリカ・イギリス・カナダの三ヶ国に限定した。

表2 トrend単語のツイート数上位12件

Marriage Story	#adoreyou	#manchesterderby
Spotify	Ruiz	Fury
BRING ANNE HOME	#RuizJoshua2	Juice Wrld
#fridayfeeling	#romance	Clemson

3.4 歴史ツイートカテゴリ

本稿では、[23] で提案された歴史に関するハッシュタグのために定義されたカテゴリを、ツイートに関するものとみなして利用する。

(1) 一般的な歴史 (General History、**GH**): **#history** や **#historicalcontext** といった、一般的な歴史を参照するツイートで使用されているハッシュタグを含む。

(2) 一国史・地域史 (National or Regional History、**NH**): **#canadianhistory** や **#ushistory** のような、ある特定の国や地域に関する歴史について言及しているツイートで使用されているハッシュタグを含む。

(3) テーマ史 (Facet-focused History、**FH**): 美術史やスポーツ史といった、歴史の特定の側面に関連するハッシュタグを含む。

(4) コメモレーション (General Commemoration、**GC**): このカテゴリは、**#onthisday**、**#otd**、**#todayweremember**、**#4yearsago** のような、ある特定の日や期間に生じた事柄への記憶を表すハッシュタグを含む。

(5) 歴史的なイベント (Historical Events、**Hv**): 過去の特定のイベントに関するハッシュタグを含む。例えば、**#wwI**、**#SevenYearsWar** などが含まれる。

(6) 歴史的な固有表現 (Historical Entities、**Hn**): 人や組織や物などの固有表現 (ただし、国や地域、土地は除く) に関するハッシュタグを含む。

表3に本稿で収集したツイートに含まれる歴史ハッシュタグの分布を示す。

表3 カテゴリ分布

カテゴリ	歴史	非歴史
GH	208,590	83
NH	65,328	34
FH	108,819	10
GC	349,707	241
Hv	180,472	22
Hn	57,576	5

4 提案手法

図2に本稿で実現するチャットボットのシステム全体像を示す。本チャットボットシステムと利用者との対話は Twitter 上のみで行う。以降では、本チャットボットと利用者との対話を実現するために重要な機能を述べる。これらの機能では、複数の出来事が出力候補となったとき、それぞれに対してスコアを求め、最もスコアが高いものを出力する。

5 : <https://www.crummy.com/software/BeautifulSoup/>

6 : <http://blog.historians.org/2013/08/history-hashtags-exploring-a-visual-network-of-twitterstorians/>

7 : <https://dev.twitter.com/rest/public>

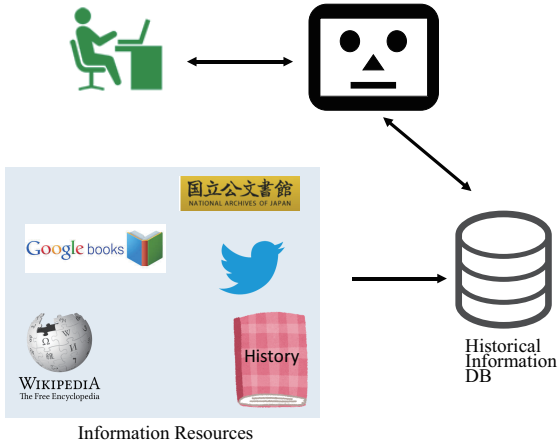


図2 チャットボットシステムの全体像

$$Importance(evt) = \sum_{ett \in E(evt)} Link(ett) \quad (1)$$

$$result(E) = \arg \max_{evt \in E} Importance(evt) \quad (2)$$

上記の式は、出来事の記述中の固有表現の重要度の総和からその出来事の重要度を近似し、その値でランキングを求めている。本稿では、固有表現は Wikipedia の記事として存在し、Wikipedia 内での参照数によって固有表現の重要度を測定できると仮定している。 evt は出来事、 ett は固有表現、関数 $Link(ett)$ は引数で与えられた ett の被参照数を求める。

4.1 過去の出来事に関するツイート

本チャットボットの最も基本的な機能は、過去のその日に起きた出来事を、定期的につぶやくものである。この機能はタ刊と同様に、過去の同じ日に何が起きたのか、新聞で報道された内容についてつぶやく。

本機能は毎日、過去のその日に起きた出来事を出力するが、以下の式によって出力する出来事を求める。

$$result(d, E) = \arg \max_{evt \in E} \{Importance(evt) \mid date(evt) == d\} \quad (3)$$

d は月日で表された日付情報、 $date(evt)$ は引数で与えられた出来事が起きた日を返す関数である。

この機能は、特定の期間で定期的につぶやくように設定しており、本チャットボットで唯一のチャットボットから利用者への方向でのコミュニケーションを実現している。

4.2 コメモレーションに関するリプライ

利用者が特定の年月日に起きた出来事を知りたいとき、その日付をチャットボットにツイートすると、その日に起きた過去の出来事をチャットボットが返答する。英語では mm/dd/yyyy、日本語では yyyy 年 mm 月 dd 日の形式でチャットボットにツイートすると本機能が実行される。なお、年や月日のいずれかは省力できる。

本機能は式 3 を利用し、利用者が入力した日付 d に起きた出来事の中から重要なものを出力する。

4.3 文章類似度に基づくリプライ

もしチャットボットに文章を入力すると、類似する過去の出来事を得られる。類似度を求めるために、本チャットボットはツイートや過去の出来事の記述、収集した Wikipedia 記事のタイトルとカテゴリから単語を取り出して特徴ベクトルを作成する。次に、ツイートの特徴ベクトルと過去の出来事の特徴ベクトルとの類似度をコサイン類似度で求める。最後に最もコサイン類似度が高い出来事をチャットボットは出力する。以下に本機能を求める式を示す。

$$score(m, e) = \alpha CosSim(m, e) + (1 - \alpha) Importance(e) \quad (4)$$

$$result(m, E) = \arg \max_{evt \in E} score(m, evt) \quad (5)$$

ここで、 m は利用者からチャットボットへのツイート、 e は出来事、 α は $CosSim$ と $Importance$ の二つの値を調整するためのハイパーパラメータを表す。本稿ではこのハイパーパラメータを 0.5 とする。

4.4 固有表現フィルタを用いたリプライ

特定の固有表現に関連した出来事を出力できるように、本稿ではハッシュタグのテキストを固有表現とみなす⁸。

次に、本システムのデータベースの中から、すべての固有表現を含む出来事を探し、リプライする候補を選ぶ。もしすべての固有表現を含むものが無い場合、少なくとも一つを含む出来事を出力候補とする。チャットボットは式 2 に従って上記の候補の中から最もスコアの高いものを出力する。なお、もし利用者のツイートにハッシュタグが含まれ無ければ式 5 を適用して出力する出来事を決定する。

4.5 歴史関係ツイートの検出

最後に、もし利用者が歴史に関する文章を入力し、本システムのデータベースに記録されていない場合、他の利用者に情報共有するために利用者からのツイートをリツイートする⁹。

このリツイートを実現するために、本稿では TF-IDF を用いた特徴ベクトルで訓練された二クラス分類器を用いる。分類器の訓練のために、第 3 節で述べた歴史関係ツイートとトレンドツイートに対して特徴ベクトルを作成する。

5 実験

本稿では訓練した分類器の精度について評価する。

評価方法. 2 クラス分類器の評価としてマイクロ平均適合率、マイクロ平均再現率、マイクロ平均 F 値の 3 つの指標を用いる。収集したデータセットに対して分割数を 10 とした交差検定を適用し、各評価結果の平均値を示す。

比較手法. 本稿では以下の分類器を評価する。

(1) SVM. サポートベクターマシン (Support Vector Machine, SVM) は広く使われている分類器である。本稿では線形カーネルを用いて SVM を訓練した。

⁸: 現在のシステムはハッシュタグのみを固有表現として利用できるが、今後は自動的に固有表現を抽出できるように NER といった手法を利用する計画である。

⁹: 現在のシステムでは英語に限定しているが、他の言語にも順次対応する計画である

(2) RFs. ランダムフォレスト (Random Forests, RFs) も比較対象としてよく使われる多クラス分類器である。

(3) NB. 単純ベイズ (Naive Bayes, NB) も比較対象としてよく使われる多クラス分類器である。

表 4 精 度.			
	適合率	再現率	F 値
SVM	94.7%	94.7%	94.7%
RFs	98.9%	98.9%	98.9%
NB	97.6%	97.6%	97.6%

5.1 結 果

表 4 に各手法のマイクロ平均の適合率、再現率、F 値を示す。いずれの分類器も高い精度が得られているが、RFs が最も高い精度で、3 種類全ての評価指標で 98%を超えている。

5.2 デモンストレーション

最後に、前節で述べた各機能がどのように稼働するのかを示す。図 3 はメモレーションに関するリプライの例を示す。この図では”HistoChatbot Developer”が本チャットボットの利用者として、この図は利用者がチャットボットへのツイートとして”3/19”を入力していることを示す。この内容は第 4.2 節で述べた mm/dd 形式に該当するので、データベースの中から 3 月 19 日に起きた過去の出来事を探し、式 1 に従って最もスコアの高いものを出力する。



図 3 ”3/19”ツイートへのメモレーションに関するリプライ例

図 4 は文章類似度に基づくリプライの例を示す。チャットボットは式 5 に従って、この文章に最も類似するものを出力する。



図 4 ”supreme court japan”ツイートに類似する出来事を検索する文章類似度に基づくリプライの例

図 5 は固有表現フィルタを用いたリプライの例を示す。まず、チャットボットは利用者のツイートに含まれるすべてのハッシュタグのテキストを収集する。次に、式 5 によって、出力候補の出来事の中から利用者の文章との類似度が最も高いものを出力する。



図 5 ”1988 #japan #china”ツイートへの固有表現フィルタを用いたリプライの例

6 おわりに

本稿では過去の出来事を定期的につぶやくだけでなく、年月日や固有表現を指定されるとそれらに関する出来事をリプライするチャットボットを実現した。

今後の課題としては次の 2 つが考えられる。1 つ目は、多くの人にとって有益と思われる過去に関するツイートの発見と共有を実現することである。2 つ目は、過去の出来事を翻訳することである。これによって、ある出来事が他の国や地域ではどのように記録されているのかを知ることが可能になる。

謝辞 本研究の一部は科研 (#17H01828 and #17K12792) と Microsoft Research Asia Collaborative Research Grant 2018 の助成を受けたものである。

文 献

- [1] Abelson, R.P., Levi, A.: Decision making and decision theory, handbook of social psychology. pp. 231–309 (1985)
- [2] Au Yeung, C.m., Jatowt, A.: Studying how the past is remembered: Towards computational history through large scale text mining. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11, pp. 1231–1240. ACM, New York, NY, USA (2011)
- [3] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
- [4] Cauvin, T.: The rise of public history: An international perspective. Historia Crítica No.40 **68**, 3–26 (2018)
- [5] Clavert, F., Majerus, B., Beaupre, N.: #ww1. twitter, the centenary of the first world war and the historian
- [6] Conard, R.: The pragmatic roots of public history education in the united states. The Public Historian **37**(1), 105–120 (2015)
- [7] Cook, J., Das Sarma, A., Fabrikant, A., Tomkins, A.: Your two weeks of fame and your grandmother’s. In: Proceedings of the 21st International Conference on World Wide Web, WWW '12, pp. 919–928. ACM, New York, NY, USA (2012)
- [8] Deerwester, S., T. Dumais, S., W. Furnas, G., Thomas K., L., Harshman, R.: Indexing by latent semantic analysis. J. Amer. Soc. Inform. Sci. **41**(6), 391–407 (1990)
- [9] Ferron, M., Massa, P.: Collective memory building in wikipedia: The case of north african uprisings. In: Proceedings of the 7th International Symposium on Wikis and Open Collaboration, WikiSym '11,

- pp. 114–123. ACM, New York, NY, USA (2011)
- [10] Gilovich, T.: Seeing the past in the present: The effect of associations to familiar events on judgments and decisions. *Journal of Personality and Social Psychology* **40**(5), 797 (1981)
- [11] Kanhabua, N., Nguyen, T.N., Niederée, C.: What triggers human remembering of events?: A large-scale analysis of catalysts for collective memory in wikipedia. In: *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '14*, pp. 341–350. IEEE Press, Piscataway, NJ, USA (2014)
- [12] Kanhabua, N., Niederee, C., Siberski, W.: Towards concise preservation by managed forgetting: Research issues and case study. In: *Proceedings of the 10th International Conference on Preservation of Digital Objects (iPres)* (2013)
- [13] Knevel, P.: Public history: the european reception of an american idea? *Levend Erfgoed* **6**, 4–8 (2009)
- [14] Košmerlj, A., Belyaeva, E., Leban, G., Grobelnik, M., Fortuna, B.: Towards a complete event type taxonomy. In: *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pp. 899–902. ACM, New York, NY, USA (2015). DOI 10.1145/2740908.2742005
- [15] Lee, U., Liu, Z., Cho, J.: Automatic identification of user goals in web search. *WWW '05*, pp. 391–400. ACM, New York, NY, USA (2005)
- [16] Nie, L., Wang, M., Zha, Z., Li, G., Chua, T.S.: Multimedia answering: Enriching text qa with media information. *SIGIR '11*, pp. 695–704. ACM, New York, NY, USA (2011)
- [17] Phan, X.H., Nguyen, L.M., Horiguchi, S.: Learning to classify short and sparse text & web with hidden topics from large-scale data collections. *WWW '08*, pp. 91–100. ACM, New York, NY, USA (2008)
- [18] Samuel, R.: 'editorial introduction', in samuel, ed., *history workshop: A collectanea 1967–1991* (oxford: Hw 25, 1991), iii; colin ward, 'fringe benefits', *new society* (1989)
- [19] Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. *SIGIR '10*, pp. 841–842. ACM, New York, NY, USA (2010)
- [20] Sumikawa, Y., Ikejiri, R.: Multi-label classification for past events. In: *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 562–567 (2018)
- [21] Sumikawa, Y., Jatowt, A.: Classifying short descriptions of past events. In: *Advances in Information Retrieval, ECIR '18*, pp. 729–736. Springer International Publishing (2018)
- [22] Sumikawa, Y., Jatowt, A.: System for category-driven retrieval of historical events. In: *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, JCDL '18*, pp. 413–414. ACM, New York, NY, USA (2018)
- [23] Sumikawa, Y., Jatowt, A., Düring, M.: Digital history meets microblogging: Analyzing collective memories in twitter. *JCDL '18*, pp. 213–222. ACM, New York, NY, USA (2018)
- [24] Sun, X., Wang, H., Yu, Y.: Towards effective short text deep classification. *SIGIR '11*, pp. 1143–1144. ACM, New York, NY, USA (2011)
- [25] Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welp, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. **10**, 178–185
- [26] Zelikovitz, S., Marquez, F.: Transductive learning for short-text classification problems using latent semantic indexing. *International Journal of Pattern Recognition and Artificial Intelligence* **19**(2), 146–163 (2005)
- [27] 池尻良平, 澄川靖信: 真正な社会参画を促す世界史の授業開発 – その日のニュースと関連した歴史を検索できるシステムを用いて-. *社会科学研究* **84**, 37–48 (2016)