Estimating Contemporary Relevance of Past News

Mari Sato
Kyoto University
Kyoto, Japan
msato@db.soc.i.kyoto-u.ac.jp

Adam Jatowt
University of Innsbruck
Innsbruck, Austria
adam.jatowt@uibk.ac.at

Yijun Duan
AIST
Tokyo, Japan
minsak1020@gmail.com

Ricardo Campos
LIAAD – INESCTEC, Polytechnic Institute of Tomar,
Ci2 - Smart Cities Research Center, Tomar, Portugal
ricardo.campos@ipt.pt

Masatoshi Yoshikawa
Kyoto University
Kyoto, Japan
yoshikawa@i.kyoto-u.ac.jp

***Abstract*—Our society generates massive amount of digital data, significant portion of which is being archived and made accessible to the public for the current and future use. In addition, historical born-analog documents are being increasingly digitized and included in document archives which are available online. Professionals who use document archives tend to know what they wish to search for. Yet, if the results are to be useful and attractive for ordinary users they need to contain content which is interesting and familiar. However, the state-of-the-art retrieval methods for document archives basically apply same techniques as search engines for synchronic document collections. In this paper, we introduce a novel concept of estimating the relation of archival documents to the present times, called *contemporary relevance*. Contemporary relevance can be used for improving access to archival document collections so that users have higher probability of finding interesting or useful content. We then propose an effective method for computing contemporary relevance degrees of news articles using Learning to Rank with a range of diverse features, and we successfully test it on the New York Times Annotated document collection. Our proposal offers a novel paradigm of information access to archival document collections by incorporating the context of contemporary time.**

***Index Terms*—news archives, information retrieval, contemporary relevance**

I. Introduction

Nowadays, many memory institutions provide access to online document archives which often span decades or centuries. The Times Digital Archive[1] or Chronicling America[2] are examples of such collections that contain millions of news articles published over multiple decades. Although, many included documents are obtained through scanning and optical character recognition, nowadays also born digital documents are being increasingly incorporated into document archives [1]. This continuous development of digital document archives allows current and future users to learn about historical events by searching and browsing primary sources. News articles constitute especially useful and attractive type of archived documents as they inform about the main events and concerns of our society in the past. News article archives allow users to retrieve detailed event-oriented information on the past from original sources, and could be useful for many purposes; for example, for verifying credibility of information provided by secondary sources.

The field of information retrieval has embraced many domain-specific tasks since its beginning (e.g., web search, legal document search, code retrieval, social media search, etc.) [2]. We believe it is also time for novel developments in the field of accessing archived heritage data. While text indexing and query suggestion methods have been already studied in the context of temporal search [3]–[8], relatively little has been done about accessing approaches to archival documents such as document ranking and retrieval. The common strategy is basically to apply the same access approaches as ones used for synchronic text collections. In this context, we argue that due to unique temporal characteristics of archives, the methods for effective document retrieval need to depart from standard IR solutions designed for synchronic document collections. Our work aims at providing useful signals for improving information access to document archives and, by this, the utilization of our heritage.

Professional users of archival collections such as historians or archivists typically know what they wish to find when searching in or interacting with archives. In other words, they tend to have rather precise search intents. Furthermore, such users usually have good skills in locating required documents. On the other hand, general users may have less defined search intents, oftentimes wishing to just sample few documents without a particular search intent in mind [9], or they may simply want to find educational or interesting content. We think that content related to the present, can be not only more pertinant and attractive to the users, but can also have increased chance to be of higher utility. For example, a journalist working on a story might be more interested in archival documents that are related to the present time, than the ones which have weak relation to the present.

We propose the concept of *contemporary relevance* of archival documents, which indicates their connection to the present time. In synchronic document collections, relevance is typically estimated as the match between query words and

[1] https://www.gale.com/intl/c/the-times-digital-archive
[2] https://chroniclingamerica.loc.gov/

document content. However, in diachronic document collections like long-term news archives, even if the returned documents are relevant to a user query, they may not necessarily correspond to the context and the circumstances familiar to the present day searchers. Archival documents are naturally detached from the present, especially, if they were published long time ago. We argue that the degree of the relation of documents to the current times or current context should be considered in order to return attractive and useful search results. Outputting documents that have certain relation to the present time should help users to find content potentially interesting to them and should even stimulate serendipitous discovery [10].

However, manually finding connecting points and determining the relevance of past documents to the present is not a trivial task. Our focus is then on automatically estimating this kind of temporal correspondence which we consider as a novel, additional signal of document usefulness besides traditional concept of keyword-based relevance. Note that contemporary relevance is actually a known concept in history science, which is used to represent and assess the significance of historical events and persons to the current times [11], [12]. Particularly, this concept plays an important role in history didactics by letting students understand the connection of history to the existing society, and by this to evoke their interests in history learning. Some scholars even went so far as to claim that ”Historical contents without contemporary relation are irrelevant for students and cannot be taught” [13].

In Fig. 1 we show two example news articles from late 1980s in order to demonstrate a possible way of how signals of contemporary relevance may look like. While both documents are about economic issues, the bottom article is also about ”Donald Trump.” Comparing these two documents, we can say that the bottom document has more chance to be regarded as relevant to the present times since Donald Trump was recently the US president and is still an important person in the present world. In another example, Fig. 2 shows two other document samples, this time from 2007. Both are related to ”France”, yet the bottom article describes citizen protests while the top one is about speed testing of TGV trains. The former document would be then considered as more related to the present times, since citizen protests in France have been recently quite common and violent, as well as they were frequently reported in various news channels worldwide[3][4]. Note that these two pairs of documents exemplify only few possible ways in which contemporary relevance can manifest itself in archival documents.

> Hale Stores, has moved to coordinate its major businesses by naming Ira Neimark, chairman and chief executive of its Bergdorf Goodman subsidiary, to the additional post of vice president of merchandise development for the Neiman-Marcus Group.

> When it comes to merchandising, Donald J. Trump and Mody Dioum could not be much farther apart. But they agree on one thing: The holiday season was hard on Fifth Avenue street peddlers. And as the avenue stepped back to normal yesterday, it appeared that the city’s recent crackdown on merchandise peddlers was still having an effect.

Fig. 1. Excerpt news articles from 1988 (top; available at https://www.nytimes.com/1988/04/12/business/credit-markets-neiman-shifts-key-executives.html) and 1987 (bottom; available at https://www.nytimes.com/1987/01/06/nyregion/anti-peddler-drive-pleases-fifth-ave-merchants.html).

> France’s high-speed TGV train broke the record for speed on a rail Tuesday in a much publicized test. The train reached a maximum of 574.8 kilometers an hour, or 357 miles an hour, but fell just short of the record for all types of train. That record is held by the magnetic levitation train of Japan...

> Violent protests against the election of Nicolas Sarkozy as president of France ended early Monday after hundreds of people were arrested, hundreds of cars gutted and hundreds of windows smashed in several cities.

Fig. 2. Excerpt news articles from 2007 (top; available at https://www.nytimes.com/2007/04/03/world/europe/03iht-train.4.5130569.html) and 2007 (bottom; available at https://www.nytimes.com/2007/05/07/world/europe/07iht-protest.4.5603711.html).

A. Proposed Approach & Contributions

So how can one find past documents which are related to the current times? We think that contemporary relevance is a complex construct that involves many aspects and is difficult to be captured based on a single signal or technique. We then propose a range of diverse yet intuitive features that have good probability to reflect the notion of document correspondence to the present context. We put a particular focus on named entities and event mentions in documents and we estimate their relevance to the present using an external knowledge base. We also consider content similarity to a reference collection of the present articles as well as we analyze temporal expressions embedded in the documents. Furthermore, as the semantics of terms used in the past may differ from their current meaning we employ semantic transformation technique for matching terms across time. We then train the learning to rank model based on crowdsourced annotated data and successfully evaluate the proposed approach. Finally, we also propose a weak supervision technique to further increase the ranking effectiveness and by this to alleviate the problem of producing costly annotations.

To sum up, we make the following contributions in this paper:

1) We introduce the concept of contemporary relevance of

[3] https://www.euronews.com/tag/protests-in-france

[4] https://en.wikipedia.org/wiki/Yellow_vests_movement

archival documents and propose a task of estimating it. By incorporating the notion of document relatedness to the searcher times into archival search mechanisms we aim to enhance the current access methods to heritage document collections.

2) We design learning to rank model equipped with a range of diverse yet intuitive features to solve this task.
3) We propose and apply a weak supervision idea according to which documents from the recent past are automatically assumed to be of higher value than documents from the distant past.
4) Finally, we successfully test our proposal on news articles annotated with their contemporary relevance.

The remainder of this paper is organized as follows. Section 2 discusses related works. Section 3 formally defines the notion of contemporary relevance. Section 4 describes the proposal of features used in our approach and their computation methods. Next, we describe the experimental setup in Section 5. Section 6 reports the evaluation results of the proposed approach together with the explanation of the two-stage learning process of our method. Finally, we conclude the paper and outline our future plans in Section 7.

II. Related Work

A. Temporal Information Retrieval

Temporal signals, both from queries and documents, have been increasingly utilized in the retrieval process giving rise to Temporal IR (T-IR), which aims to enhance information retrieval by combining the traditional document relevance with temporal relevance [14], [15].

Several research studies have then been proposed for ranking documents considering temporal aspects [3], [14]–[20]. Li and Croft (2003) [21] introduced a time-based language model that takes into account timestamp information of documents to favor recent documents. Metzler *et al.* (2009) [22] proposed a method that analyzes query frequencies over time to infer the implicit temporal information of queries and exploit this information for ranking results. Campos *et al.* (2016) [23] defines a similarity measure that makes use of co-occurrences of words and years based on corpus statistics and a classification methodology that is able to identify the set of top relevant dates for a given implicit time-sensitive query and as such improve the effectiveness of the ranked results. Arikan *et al.* (2009) [24] designed a temporal retrieval model that integrates temporal expressions of document content into query-likelihood language modeling. Berberich *et al.* (2010) [25] proposed a similar model but also considered uncertainty in temporal expressions. Kanhabua and Nørvåg (2010) [26] designed three different methods to determine the implicit temporal scopes of queries and exploited temporal information to improve the retrieval effectiveness by reranking documents. Campos *et al.* (2017) [27] made use of temporal signals extracted from document contents to associate relevant temporal expressions to implicit temporal queries.

Related to this research are also works on recency-oriented search and on estimating freshness of documents [28], [29] as well as those on detecting salient time periods of documents in historical archives [30] using diversification techniques. However, outputting recent or fresh documents is not applicable to archival collections. Diversifying search results with respect to their temporal distribution as proposed by Singh *et al.* [30] as well as by Berberich and Bedathur (2013) [31] cannot solve the issue of finding documents that have contemporary relevance, either.

The research on estimating the *focus time* of documents [32], [33], defined as the time period to which the document content refers to, is also related to this work, however, it is only applicable to the relation of documents to the past periods (e.g., documents referring to historical events) rather than to the correspondence of past documents to the present times.

To the best of our knowledge, no prior work has proposed the concept of computing contemporary relevance in long-term temporal document collections. In addition to this, our focus is on past articles rather than on recent articles, such as ones collected from online news streams. Hence, TDT (Topic Detection and Tracking) [34], [35] and other related initiatives based on story linking (e.g., the background linking task [36] in TREC news tracking[5]) are not applicable here.

B. Accessing Archival Documents

Thanks to the widespread digitization efforts, new interdisciplinary research areas started to emerge. Digital history [37] and archival informatics [38] are examples of novel disciplines for which digital document archives are of central importance. However, the majority of current efforts in these areas still go into digitizing, annotating, and organizing content as well as providing rudimentary processing and search techniques without significant advances in effective retrieval models. In general, memory institutions such as archives or libraries seem to lack deeper consideration of how to build effective and attractive services for users [39] and their strategies for user involvement are considered to be often ineffective [40].

We believe that this situation is likely one of the reasons why the level of usage of digital archival collections is generally still far from what could be expected. There is a need for more advanced search mechanisms to attract and engage users than just offering standard keyword matching based search. Several initial attempts aimed then at improving access to document archives. Berberich *et al.* [5] proposed *time-travel text search* over versioned document collection to effectively index and rank relevant documents considering the collection state at different times. Tran *et al.* [4] designed *time-aware re-contextualization* system, which automatically provides complementing information from Wikipedia to a text in archival document aiming at helping users understand the past content. Pasquali *et al.* [41] designed a temporal summarization method to query the Portuguese web Archive (Arquivo.pt) aiming at helping users to make sense of a given topic's evolution over time. Zhang *et al.* [42] and Duan *et al.* [43] have proposed mapping entities from the past with

[5] http://trec-news.org/guidelines-2020.pdf

corresponding present entities based on their descriptions in news archival collections. Finally, Jatowt *et al.* [44] proposed retrieving potentially interesting content from news archives based on modeling the feeling of surprise that past content may evoke in present-day readers. The underlying hypothesis is that the past content which is different from the contents common at present or which is unexpected could be attractive to current users.

Although, linking historical data has been already recognized as an important way to increase the value of archival content [8], [45], the idea of automatically connecting it with the present is still to be realized. Our proposal is then a unique endeavor within the recent synergy of history science and informatics; the endeavor that focuses on the needs of an end user – the average consumer of history-related content. Finding the actual correspondence of the historical content to the present should greatly increase the value and usefulness of historical data, and might lead to higher user engagement with heritage document collections and enhanced history education.

C. *Newsworthiness*

In journalism and media studies, an important discussion centers about which events should be selected and delivered to readers. Newsworthiness is the degree, based on a set of values, which determines whether a particular news is worth publishing. Gultung and Ruge published a theory of news selection and proposed 12 factors of newsworthiness [46]. They include (1) *frequency*: the time-span of the event to unfold itself, (2) *threshold*: the impact of the event, (3) *unambiguity*: the ambiguity degree of the event, and (4) *meaningfulness*: the relevance of the event in terms of cultural proximity. There have been also several attempts to assess newsworthiness by automatically analyzing the content of articles. For example, Di Buono and Snajder (2013) [47] evaluated correlations between newsworthiness and linguistic features of the headlines. De Nies *et al.* (2012) [48] proposed an approach to analyze the content of an article from six perspectives including similarity analysis, named entity recognition and topic detection.

Unlike the above-mentioned approaches, which assign time-agnostic scores to documents, we consider integrating the notion of newsworthiness into archival scenarios, and, we measure the relevance of documents to the time context of a user.

III. Problem Setting

We now provide a definition of the proposed task starting from a more general concept of temporal relevance.

Def. 1: Given a document collection $D = (d_1, ..., d_n)$ which contains documents created within a time period $T^D = [t_s^D, t_e^D]$ such that $t_s^D < t_e^D$, and given a pre-specified reference time period $T^{ref} = [t_s^{ref}, t_e^{ref}]$ such that $t_s^{ref} < t_e^{ref}$, the task is to estimate, for each document d_i $(i = 1, ..., n)$, the **temporal relevance** $R(d_i \mid T^{ref})$ which defines the degree of the relatedness of d_i to T^{ref}.

Def. 2: **Contemporary relevance** is the special case of temporal relevance, where $T^{ref} = T^{pre}$. The time period $T^{pre} = [t_s^{pre}, t_{now}]$ represents the abstract notion of the present/current times, and t_{now} is the current moment (i.e., the search time).
We note that, by definition, t_{now} as the end point of T^{pre} is assumed to be equal to now. However t_s^{pre} - the start point of T^{pre} - is not strictly defined and depends on the length of the archival collection and possibly other factors[6].

Finally, for the sake of completeness, we note that the reference time T^{ref} in Def. 1 could be set earlier or later in relation to T^D. In general, three basic cases of T^{ref} can be distinguished[7] (see Fig. 3 for visual representation): (1) T^{ref} is equal to the present time period T^{pre}, (2) T^{ref} is the past time period preceding T^D, and (3) T^{ref} is the time period between T^D and T^{pre}. In this work we focus on the first case as being most applicable for general users accessing archival collections[8].

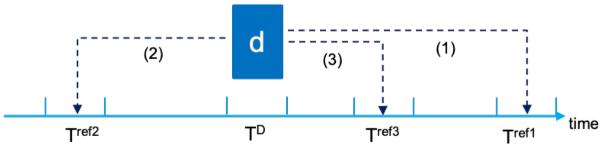

Fig. 3. Schematic visualization of the three possible cases of a reference time period in relation to a target document *d*. The reference time period denoted by T^{ref1} (and associated with relation (1)) represents the contemporary time (the right boundary of T^{ref1} is assumed to be the search time).

IV. Method

Fig. 4 shows the overview of our approach. We represent documents by a range of features that are likely to contribute to estimating their relatedness to the present. For this, we utilize an external knowledge base such as the Wikipedia and a reference collection of recent news. With such document representation, we then apply learning to rank models for ranking archival documents.

In the next sections, we will describe the proposed features along with the hypotheses and intuitions that support their choice. The features are grouped into three categories: *content-related features*, *entity-related features* and *event-related features*. All the features are normalized using min-max normalization.

A. *Content-related Features*

1) Similarity to present documents: We first introduce a measure of the similarity between a target past document and the representative collection of present documents (denoted here as D^{ref}) which is used for reflecting news stories common in the contemporary time. The objective is to determine

[6] In the experiments we test four different representations of T^{pre} based on the boundary set by t_s^{pre}: the last 1 month, the last 6 months, the last 1 year and the last 3 years.

[7] We exclude the overlap cases for the ease of exposition.

[8] The other two cases could have applications in specific situations, e.g., when historians search for connections and correspondences across different periods in the past.

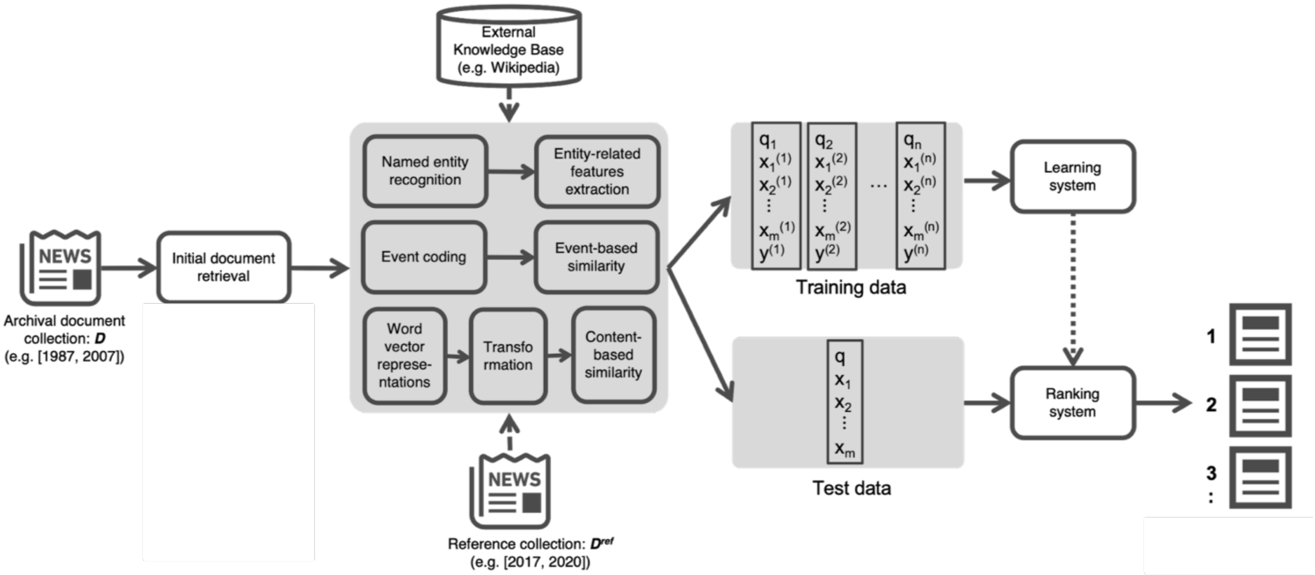
External Knowledge Base (e.g. Wikipedia)
NEWS
Archival document collection: D (e.g. [1987, 2007])
Initial document retrieval
Named entity recognition
Entity-related features extraction
Event coding
Event-based similarity
Word vector represe-ntations
Transfo rmation
Content-based similarity
NEWS
Reference collection: D^ref (e.g. [2017, 2020])
q1 x1(1) x2(1) xm(1) y(1)
q2 x1(2) x2(2) xm(2) y(2)
qn x1(n) x2(n) xm(n) y(n)
Training data
q x1 x2 xm
Test data
Learning system
Ranking system
1
2
3

Fig. 4. Overview of the proposed approach for determining contemporary relevance of archival documents (temporal expression related features are grouped together with the named entity related features).

the similarity of archival news articles to popular news stories in recent times.

For the comparison of semantics we use Word2Vec as a vectorization method with transformation between vector spaces. Vector space transformation is necessary due to semantic shifts of terms over time, especially, over longer time-spans common in archival scenarios. Even if two terms are semantically close, they may have quite low overlap between their contexts (reflected by surrounding terms) across time. For example, as demonstrated by Tahmasebi *et al.* (2012) [49], while it is reasonable to judge *iPod* as corresponding to *walkman* in the past (both are portable music devices), the set of the top co-occurring words with *iPod* in documents published in 2010s has little overlap with the set of the top co-occurring words with *walkman* that are extracted from documents in 1980s (only 1 out of 10 top co-occurring terms, which are not stopwords, is shared). For computing the transformation between the two different vector spaces we use orthogonal transformation with a weak supervision approach proposed by Zhang *et al.* (2015) [6] and also used in [8]. The similarity of a target document d and the reference collection D^{ref} is denoted as:

$$Sim(v_d, v^{ref}) = Sim_{cosine}(W \cdot v_d, v^{ref}) \tag{1}$$

where v^{ref} is the feature vector of the reference collection D^{ref} computed by averaging word embeddings after removing stopwords and W is the transformation matrix. v_d is obtained by calculating and averaging the vectors for words in d.

2) Temporal expressions: We consider that, if a document contains temporal expressions pointing to time periods in or near the present, it has high chance to be contemporary relevant. In practice, such documents may contain plans, predictions or expectations related to the future (future understood as time period after documents' publication dates).

We thus extract and normalize time expressions in each document. We use for this the HeidelTime temporal tagger[9] [50]. Each temporal expression is represented by a single year t_{te} denoting the mid year of that expression if the expression spans longer than a year. We next calculate the score τ_{te} based on the distance between the start of T^{pre} and t_{te} using the time-decaying function.

$$\tau_{te} = \alpha^{-\lambda(t_s^{pre} - t_{te})} \tag{2}$$

Finally, we take the average and the maximum values of τ_{te} over the set of all temporal expressions contained in a target document as two content features of that document.

B. Entity-related Features

The next feature group that we utilize relates to named entities. Named entities are central to news and should have special consideration. However, multiple entities may occur in a single news article, and the entities that are just passing mentions are of little significance. We then first select important entities in each document before computing entity-related features. We apply three approaches to select the most important entities in a document: $frequency$, $TextRank$ and $offset$. $Frequency$ highlights core entities repeated frequently in the document. $TextRank$ [51] selects important sentences which have high probability to capture the essence of a document, and the entities belonging to the top-scored sentences are deemed salient. $Offset$ is used to extract entities that appear in early parts of news articles such as in a title or a lead paragraph, as these are often indicative

[9] https://github.com/HeidelTime/heideltime

of the theme of a news article. We then select n ($n = 5$ by default) of the most important entities scored by each of the above-described measures for computing the entity-related features, which are (1) the current popularity of entities; (2) activity period of entities and (3) connectedness with present entities. For calculating these features, we use Wikipedia as our knowledge base[10]. We finally aggregate all the feature values of the selected important entities by max pooling and average pooling in order to construct the final set of entity-related features for each document.

1) Current popularity of entities: We consider that, when named entities central to a given document are currently not popular, a reader has lower probability to consider the document as related to the present. On the other hand, entities that are currently popular should be more relevant and interesting to the users. Based on this assumption, we exploit Wikipedia pageview statistics using Wikimedia REST API[11] to measure the popularity of entities. For each entity e_i we retrieve the pageview count, denoted as $pageview(e_i)$, of its corresponding Wikipedia article[12] over the last three years. We then compute the entity popularity $popularity(e_i)$ as $popularity(e_i) = \ln pageview(e_i)$.

2) Activity period of entities: When an article contains entities which are no longer active (especially ones which ceased to be active long time ago), a reader has lower probability to consider the article as present-relevant. Based on this hypothesis, we estimate the activity periods of entities. That is, for each entity, we retrieve the time intervals of its validity from the DBpedia Linked Data representation. First, we collect all the properties whose data type is date (e.g., $xsd : date$, $xsd : dateTime$, $xsd : gYear$[13]). Then, we extract temporal information associated with properties identifying time intervals (e.g., $birthDate/deathDate$ for persons, and $foundingYear/dissolutionYear$ for organizations). For querying, we use the DBpedia SPARQL endpoint[14]. For each document d along with the set of its important entities $E_d = \{e_1, e_2, \dots, e_l\}$ we compute the activity period of each entity $e_i \in E_d$ represented as $AP_{e_i} = \{t_s, t_e\}$. Then, the degree of the relevance of an entity to the present, denoted as r_{e_i}, is computed based on the distance between the entity's activity period and the present reference period T^{pre}:

$$r_{e_i} = \begin{cases} 1 & (if\ AP_{e_i} \cap T^{pre} \neq \emptyset) \\ \alpha^{-\lambda(t_s^{pre} - t_e)} & (if\ AP_{e_i} \cap T^{pre} = \emptyset) \end{cases} \tag{3}$$

where t_s^{pre} is the starting time point of T^{pre}.

3) Connectedness with present entities: When an entity has weak or no connections with entities that are active at present times, it should have smaller chance to be considered as related to the present. On the other hand, an entity that is related to many currently active or valid entities is likely to be relevant to the present. We then construct the entity graph $\boldsymbol{G(V, E)}$, where $\boldsymbol{V}$ is the set of nodes representing entities in the documents as well as the entities which are linked to them in the knowledge base, and $\boldsymbol{E}$ is the set of edges representing links between $\boldsymbol{V}$. We construct $\boldsymbol{G}$ using DBpedia Page Links dataset[15]. We next compute the connectedness of entities to the present using the biased random walk as described in Eq. 3. $\boldsymbol{R}$ is a vector containing node relatedness r_{v_i}, $\boldsymbol{M}$ is an aperiodic transition matrix, α is a decay factor ($\alpha = 0.85$), and $\boldsymbol{d}$ is the static score distribution vector summing up to one and which is bound to the distances of entity activity periods from the present (see Eqs. 2 and 4).

$$\boldsymbol{R} = (1 - \alpha)\boldsymbol{M} \times \boldsymbol{R} + \alpha \boldsymbol{d} \tag{4}$$

where

$$\boldsymbol{d} = r_{v_i} / \Sigma \boldsymbol{r_v}\ s.t.\ \Sigma_i d_i = 1 \tag{5}$$

C. Event-related Features

Besides entities, events are of course another important and useful signal for contemporary relevance. We hypothesize that, when a document contains mentions of events which are similar to the events common in the present time, a reader has a higher probability to consider the document to be related to the present.

We extract and structure event data (i.e., who did what to whom) from each document. Event data is assumed to consist of three codes, in our case: source actor, target actor, and action. To detect the event mentions, we use the popular event coding system called *Conflict and Mediation Event Observations* (CAMEO)[16] [55]. The code elements are classified into a number of categories, such as state actors, sub-state actor roles, regions, and ethnic groups. From each event mention, we obtain the following coded event vector representing the event:
$\{source\_actor\_state, source\_actor\_role, target\_actor\_state,$ $target\_actor\_role,\ action\_code\}$. For example, a coded event vector (USA, GOV, IRQ, None, 01) is obtained from the sentence *"Obama administration nearing decision on improving Iraqi training."*

We next compute the distance between event vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by using Hamming distance $dist(\boldsymbol{a}, \boldsymbol{b}) = m - \sum_{i=0}^{m} \delta(a_i, b_i)$, where m is the size of $\boldsymbol{a}$ and $\boldsymbol{b}$ ($m = 5$ in our case). The distances between all the possible pairs of events in D and ones in D^{ref} are then calculated. For each document $d \in D$, we finally use as features the average and minimum values of the distances computed between all the events mentioned in d and the events in D^{ref}.

[10] In prior research Wikipedia was found to be a convenient and effective resource for measuring global collective memory attention [52]–[54].

[11] https://wikimedia.org/api/rest_v1/

[12] Named entity recognition and linking is done using TextRazor tool: https://www.textrazor.com/

[13] https://www.w3.org/TR/xmlschema11-2/

[14] https://dbpedia.org/sparql

[15] https://wiki.dbpedia.org/downloads-2016-10#datasets

[16] One of the drawbacks of CAMEO coding is its strong focus on events involving pairs of actors. Events like natural disasters may be then reflected to lesser extent. Nevertheless given the simplicity and popularity of this coding scheme we have decided to use it in the current implementation.

D. Learning to Rank based on Weak Supervision

Using the features described in the previous sections, we next train the model to learn the contemporary relevance scores of documents. To gather enough training data we rely on weak supervision (the concept of which is schematically depicted in Fig. 5). In particular, we apply a pairwise model in which older documents are automatically considered as having lower scores than newer documents. This is based on the assumption that newer documents are on average more related to the contemporary times than much older documents. While this assumption is obviously not always true, it should hold for a large number of documents, especially if the compared documents were published in time periods separated by long time gaps. Intuitively, the recent past matters more and is remembered better than the distant past. This was also demonstrated on large size news collections from which past-referring temporal expressions were extracted and normalized, revealing an increased forgetting (with the shape similar to exponential function) of more distant years [56].

Thanks to this simple automatic labelling approach we can prepare a large training set without costly manual annotation. We use a pairwise approach for training the learning to rank model to determine which document is more relevant to the present. Each learning sample $s = (\boldsymbol{f_d}, r_d)$ consists of a feature vector $\boldsymbol{f_d}$ describing a document d and its assigned relatedness score r_d. We automatically assign the weak labels and use pairwise logistic loss to learn the ranking function.

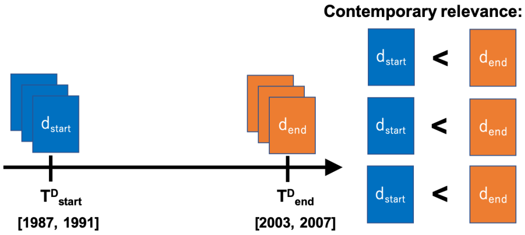

Fig. 5. Example documents used as weak supervision in Learning to Rank where documents from the distant past (ones published soon after the start date of the collection time span) are considered as less relevant to the present than more recent documents (published soon before the end date of the dataset). In the experiments, we assume the documents published in [2003,2007] to be more contemporary relevant than ones from [1987,1991].

In the experiments, we will use the model trained based on the above-outlined weak supervision trick to rank documents whose publication dates do not overlap with the publication time periods of the documents used for the training. In particular, we will train the model using the documents from two periods whose separating gap is sufficiently long to assume their different relation strengths with the present (12 years difference as shown in Fig. 5). Then we will evaluate the model using the documents from the middle of the gap period separating the two periods (i.e., from [1996,1998]). The next section provides more details about the experimental settings.

V. Experimental Setup

A. Document Collection

We use the New York Times Annotated Corpus (NYT) [57] as the underlying archival document collection, which we have indexed using Solr search engine. The collection contains over 1.8 million articles published from January 1987 to June 2007 and has been frequently used for Temporal Information Retrieval researches [14], [15]. We divide the dataset into three sub-collections for running the experiments. In particular, we use 860k articles published from 1987 to 1991, 1996 to 1998 and 2003 to 2007. Articles published from 1987 to 1991 and from 2003 to 2007 (650k in total) are used for training based on the weak supervision approach that was described in the previous section, while ones published from 1996 to 1998 (210k in total) are used as the document collection D for the evaluation.

To compute the entity-related features, we extracted and disambiguated named entities from each document using TextRazor API[17]. The API returns a Wikipedia article for each detected entity with a computed confidence score. We kept entities with the confidence score higher than 0.18, which was empirically found to perform best. To compute the content-related and event-related features, we also crawled 92k articles from the New York Times online archive portal that were published over the recent three years[18] to be used as a basis for our reference collection D^{ref}. In particular, we prepared 4 versions of D^{ref} based on four different representations of T^{pre}: the last 1 month, the last 6 months, the last 1 year and the last 3 years. Finally, for computing the event-related features, we extracted event data from each document using the pre-trained machine-coding tool, PETRARCH[19].

B. Ground-truth Dataset

In order to construct the ground-truth dataset, we used Figure Eight[20], which is a crowdsourcing platform specialized in machine learning annotation. We first collected *descriptor* fields of the New York Times Annotated Corpus in order to retrieve news articles for annotation. In particular, we used 42 such descriptors as broad queries to collect news articles from diverse domains and varying themes (e.g., art, elections, finances, sports, international relations, military forces, AIDS). For each query, we then selected the top 50 documents in the order of their Okapi BM25 scores. We asked annotators to give evaluation on the scale from 1 to 4 (1: not relevant, 2: weakly relevant, 3: moderately relevant and 4: strongly relevant) as for whether the document is *contemporary relevant* or not (1 and 2 are regarded as not relevant, and 3 and 4 are regarded as relevant). We recruited in total 277 annotators (3 annotators per article). To obtain high quality annotations, we allowed only users who are marked with the Type 3 contributor level denoting the group of users producing the highest quality

[17] https://www.textrazor.com
[18] We used articles published from the start of 2017 to the end of 2019
[19] https://petrarch2.readthedocs.io/en/latest/index.html
[20] https://www.figure-eight.com/

of annotations. Additionally, since many articles in the NYT collection focus on the events in North America, we invited contributors from only that region. Furthermore, we filtered annotators by using 16 test questions allowing them to provide answers only if they passed strict accuracy threshold. We also limited the maximum number of annotations made by one annotator to minimize chances of biased annotation. In total, we prepared 2,080 documents to be evaluated and we obtained 6,240 annotations.

C. Other Settings

For generating the ranking, we employed neural network based Learning to Rank pairwise method implemented in TensorFlow Ranking[21]. The method is proposed by Pasumarthi *et al.* [58] and was recently the state-of-the-art in Learning to Rank approaches. We set nDCG ($n = 1, 3, 5, 10$) as ranking metrics and pairwise logistic loss as a loss function. For computing the temporal similarity feature and activity period, we set $\lambda = -0.125$ and $\alpha = 0.5$ following [4]. To calculate the event-related features, we set $m = 5$ since we use the upper two elements of CAMEO code for keeping high generality of event types.

As evaluation metrics we used precision at rank 1, 3, 5, 10 (P@1, P@3, P@5, P@10, respectively) and mean average precision (MAP) at rank 10. As compared methods, we considered randomized ranking (Random), Okapi BM25 ranking (BM25), TensorFlow Ranking (TF-R) [58] with Word2Vec representation as input, and TF-R with Doc2Vec representation [59]. We also tested the proposed method without key feature groups. The Word2Vec model was trained on the New York Times Annotated Corpus. We have prepared 300 dimensions' long feature vectors of each document by averaging the vectors for their words (without stopwords). The Doc2Vec model was trained on the same corpus with 300 dimensions.

VI. Experimental Results

A. Selecting Reference Document Collection

First, we optimize the length of the present time period T^{pre} as we do not know the time span that would be most suitable to be used as the *present reference time*. As mentioned before, we set the last 1 month, last 6 months, last 1 year, and last 3 years as 4 candidates of T^{pre}, and we then compute all the features for each such span. Table I shows the performance obtained for weak-supervision based models trained using these time spans as T^{pre}. We can see that the approach with the time span of the last 1 year performs best. We then use this time frame as the reference time period T^{pre} in the subsequent analysis.

TABLE I
Ranking performance in precision and MAP measures for different lengths of reference time period.

	P@1	P@3	P@5	P@10	MAP
Last 1 month	0.548	0.532	0.557	0.564	0.546
Last 6 months	0.619	**0.603**	0.581	0.562	0.6
Last 1 year	**0.643**	0.563	**0.6**	**0.586**	**0.663**
Last 3 years	0.5	0.516	0.514	0.51	0.51

B. Ranking Performance

Next, we compare the performance of different ranking models and analyze the effectiveness of the proposed approaches. Table II shows the performance of each method. We can observe that our proposed method performs best over all the compared methods in terms of precision. For example, it outperforms BM25 up to 20% with the highest difference for P@5. Compared to TF-R with Word2Vec, our method has even higher gains with the highest difference observed for P@10. The low performance of TF-R indicates that content similarity without vector space transformation is not a good estimator of contemporary relevance. Finally, we see that our proposed method performs best in terms of MAP, which can be regarded as a summary of prediction quality.

TABLE II
Results for different ranking methods.

	P@1	P@3	P@5	P@10	MAP
Random	0.5	0.558	0.487	0.526	0.555
BM25	0.638	0.546	0.497	0.506	0.572
TF-R (Word2Vec)	0.462	0.359	0.338	0.277	0.404
TF-R (Doc2Vec)	0.571	**0.563**	0.576	0.550	0.633
Proposed Method	**0.643**	**0.563**	**0.6**	**0.586**	**0.663**

C. Effect of Feature Groups

We next conduct ablation study comparing the performance when a given feature group is removed (Entity, Event, Content feature groups). Table III shows that the removal of entity-related features results in the largest performance drop of precision, indicating that this feature group contributes most to the ranking performance. Content-related features seem also to be quite important followed by the Event-related features. Nevertheless, overall, we can conclude that all the feature groups contribute to the method effectiveness. -.5em

TABLE III
Results with a given feature group removed.

	P@1	P@3	P@5	P@10	MAP
Without Entity Features	0.488	**0.528**	**0.537**	**0.537**	**0.513**
Without Event Features	0.524	0.54	0.576	0.548	0.544
Without Content Features	**0.429**	0.587	0.548	0.545	0.517
All Features	0.643	0.563	0.6	0.586	0.663

D. Effectiveness of Using Weak Supervision

We next analyze the effectiveness of the training based on our idea of weak supervision and also propose its effective combination with the annotated data. In particular, we investigate the following methods:

Training with weak supervision only (Weak Supervision): This is the same model that uses the weak supervision as discussed and tested until now. Manual annotation data is then not used in the training stage (it is used only for testing).

[21] https://github.com/tensorflow/ranking

Training using 5-fold cross-validation on manually annotated data (Strong Supervision): We divided the manually annotated data into 5 groups and trained the model through 5-fold cross validation using only the manually annotated data.

Combining results from the models trained on the weakly annotated and on the manually annotated data by list merging (Merged Weak+Strong Supervision): We combined the prediction scores of the above two methods by averaging them. Each score was first normalized to fit between 0 and 1.

Training on weakly supervised data after filtering it using a classifier trained on manually annotated data (Enhanced Weak Supervision): We first automatically annotated the unlabelled documents using SVM classifier trained on the manually annotated data to filter low quality document pairs. In particular, we removed high-scored documents in the older part of the document collection and removed low-scored documents from the newer portion of the collection. The objective was to make the weak supervision approach more effective by pre-filtering its data. We filtered out 60% of documents using the threshold level determined in the validation experiments. After this filtering we have used the remaining documents for training in the same way as in the Weak Supervision approach.

Table IV shows that using the concept of weak supervision to gather large amount of training data is helpful and the proposed method performs in this case better than when trained only on the manually annotated data (i.e., Strong Supervision). This is likely due to relatively small amount of the ground truth data. The table also shows that our proposed combination of distantly supervised and manually annotated data through the above-described filtering process (i.e., the Enhanced Weak Supervision method) produces the best results, which are better than when the combination is done via a simple list merging (i.e., Merged Weak + Strong Supervision method).

Overall, the results indicate that it makes sense to use weakly supervised data and that automatically annotating it after the filtering step by a supportive classifier trained on a small but high quality data can further improve the results.

TABLE IV
RESULTS FOR DIFFERENT MODELS OF TRAINING WITH OUR METHOD.

	P@1	P@3	P@5	P@10	MAP
Weak Supervision	0.643	0.563	0.60	**0.586**	0.663
Strong Supervision	0.494	0.431	0.49	0.55	0.638
Merged Weak+ Strong Supervision	0.562	0.532	**0.513**	0.49	0.640
Enhanced Weak Supervision	**0.680**	**0.576**	0.502	0.463	**0.743**

VII. Conclusions & Future Work

In this paper, we focus on the question of correspondence of archival documents to the present. In particular, as the first contribution, we introduce a novel research task of estimating *contemporary relevance*. Contemporary relevance is considered as a novel criteria for effective search within long-span temporal document collections. We think that archival content which is not only relevant to the query in a traditional keyword-matching sense, but which has also some relation to the present circumstances can be attractive to users and have good chance to be perceived as of high utility. Our proposal is then a step towards more effective and citizen-friendly use of our heritage. An additional objective of this work is to trigger discussion on how to effectively access heritage data.

As the second contribution, we propose an effective method for estimating contemporary relevance of past documents using Learning to Rank and a range of dedicated features. The experimental evaluation demonstrates that our proposed approach performs satisfactory. We also prove that training with weak supervision is effective for document ranking. The proposed method can be used as an add-on to existing solutions for accessing document archives. We believe that this kind of search enhancement can also lead to serendipitous discovery of useful or interesting content, for example, by allowing otherwise buried documents to come to the top of ranked search results based on the current events or changes in the popularity or activity of the embedded entities.

Finally, our work opens several interesting avenues for further research. In the future, for example, we plan to design explanatory approaches to indicate why highly-ranked documents are judged as relevant to the present, and by this to provide users with explanations to help them better understand the returned results. We will also work on the combination of contemporary relevance with the traditional notion of relevance, and on testing or removing some assumptions we made in this work.

Acknowledgments. Ricardo Campos, one of the authors of this paper was financed by the ERDF – European Regional Development Fund through the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project PTDC/CCI-COM/31857/2017 (NORTE-01-0145-FEDER-03185). This funding fits under the research line of the Text2Story project.

References

[1] D. Gomes, D. Cruz, J. Miranda, M. Costa, and S. Fontes, “Search the past with the portuguese web archive,” in *Proceedings of the 22nd International Conference on World Wide Web*. Association for Computing Machinery, 2013, p. 321–324.

[2] W. B. Croft, D. Metzler, and T. Strohman, *Search engines: Information retrieval in practice*. Addison-Wesley Reading, 2010, vol. 520.

[3] H. Holzmann, W. Nejdl, and A. Anand, “Exploring Web Archives Through Temporal Anchor Texts,” *Proceedings of the 2017 ACM on Web Science Conference*, pp. 289–298, 2017.

[4] N. K. Tran, A. Ceroni, N. Kanhabua, and C. Niederée, “Back to the Past : Supporting Interpretations of Forgotten Stories by Time-aware Re-Contextualization,” *WSDM*, pp. 339–348, 2015.

[5] K. Berberich, S. Bedathur, T. Neumann, and G. Weikum, “A time machine for text search,” in *SIGIR*, ser. SIGIR ’07. New York, NY, USA: ACM, 2007, pp. 519–526.

[6] Y. Zhang, A. Jatowt, S. Bhowmick, and K. Tanaka, “Omnia mutantur, nihil interit: Connecting past with present by finding corresponding terms across time,” in *ACL*, vol. 1, 2015, pp. 645–655.

[7] T. Szymanski, “Temporal word analogies: Identifying lexical replacement with diachronic word embeddings,” in *ACL*, 2017, pp. 448–453.

[8] Y. Duan and A. Jatowt, “Across-time comparative summarization of news articles,” ser. WSDM ’19, 2019, p. 735–743.

[9] K. S. Calhoun, J. Cantrell, P. A. Gallagher, and D. Cellantani, “Online catalogs : what users and librarians want,” 2009.

[10] D. Kotkov, S. Wang, and J. Veijalainen, “A survey of serendipity in recommender systems,” *Knowledge-Based Systems*, vol. 111, pp. 180–192, 2016.

[11] E. Erdmann, "Contemporary relevance-a category of historical science and of the didactics of history and its consequences in teacher training," *Yesterday and Today*, no. 17, pp. 140–153, 2017.

[12] A. Keszei, "Memory and the contemporary relevance of the past," *The Hungarian Historical Review*, vol. 6, no. 4, pp. 804–824, 2017.

[13] P. Karsten, "Unterricht in geschichte-politik," *Geschichte, Politik und ihre Didaktik*, vol. 20, p. 14, 1992.

[14] R. Campos, G. Dias, A. M. Jorge, and A. Jatowt, "Survey of temporal information retrieval and related applications," *ACM Comput. Surv.*, vol. 47, no. 2, pp. 15:1–15:41, 2014.

[15] N. Kanhabua, R. Blanco, and K. Nørvåg, "Temporal information retrieval," *Foundations and Trends in Information Retrieval*, vol. 9, no. 2, pp. 91–208, 2015.

[16] O. Alonso, M. Gertz, and R. Baeza-Yates, "On the Value of Temporal Information in Information Retrieval," in *ACM SIGIR Forum*, vol. 41, no. 2. ACM, 2007, pp. 35–41.

[17] N. K. Tran, A. Ceroni, N. Kanhabua, and C. Niederée, "Time-travel translator: Automatically contextualizing news articles," in *Proceedings of the 24th Int. Conf. on World Wide Web*, ser. WWW '15 Companion. New York, NY, USA: Association for Computing Machinery, 2015, p. 247–250.

[18] T. N. Nguyen, N. Kanhabua, W. Nejdl, and C. Niederée, "Mining relevant time for query subtopics in web archives," in *Proceedings of WWW*, ser. WWW '15 Companion, 2015, p. 1357–1362.

[19] T. N. Nguyen, N. Kanhabua, C. Niederée, and X. Zhu, "A time-aware random walk model for finding important documents in web archives," in *Proceedings of SIGIR 2015*, ser. SIGIR '15, 2015, p. 915–918.

[20] Y. Duan and A. Jatowt, "Across-time comparative summarization of news articles," in *Proceedings of WSDM*, ser. WSDM '19, 2019, p. 735–743.

[21] X. Li and W. B. Croft, "Time-based language models," in *Proceedings of the twelfth international conference on Information and knowledge management*. ACM, 2003, pp. 469–475.

[22] D. Metzler, R. Jones, F. Peng, and R. Zhang, "Improving Search Relevance for Implicitly Temporal Queries," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. Citeseer, 2009, pp. 700–701.

[23] R. Campos, G. Dias, A. Jorge, and C. Nunes, "Gte-rank: a time-aware search engine to answer time-sensitive queries," *Information Processing & Management an International Journal*, vol. 2, no. 52, pp. 273–298, 2016.

[24] I. Arikan, S. Bedathur, and K. Berberich, "Time will tell: Leveraging temporal expressions in ir," in *In WSDM*, 2009.

[25] K. Berberich, S. Bedathur, O. Alonso, and G. Weikum, "A language modeling approach for temporal information needs," in *European Conference on Information Retrieval*. Springer, 2010, pp. 13–25.

[26] N. Kanhabua and K. Nørvåg, "Determining time of queries for re-ranking search results," in *International Conference on Theory and Practice of Digital Libraries*. Springer, 2010, pp. 261–272.

[27] R. Campos, G. Dias, A. Jorge, and C. Nunes, "Identifying top relevant dates for implicit time sensitive queries," *Information Retrieval Journal*, vol. 4, no. 20, pp. 363–398, 2017.

[28] N. Dai, B. D. Davison, and C. Sci, "Learning to Rank for Freshness and Relevance," *SIGIR*, 2011.

[29] A. Dong, Y. Chang, Z. Zheng, G. Mishne, J. Bai, R. Zhang, K. Buchner, C. Liao, and F. Diaz, "Towards recency ranking in web search," in *WSDM*, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 11–20.

[30] J. Singh, W. Nejdl, and A. Anand, "History by diversity: Helping historians search news archives," *Proceedings of the 2016 ACM Conference on Human Information Interaction and Retrieval, CHIIR 2016, Carrboro, North Carolina, USA, March 13-17, 2016*, pp. 183–192, 2016.

[31] K. Berberich and S. Bedathur, "Temporal diversification of search results," in *SIGIR workshop, TAIA*, 2013.

[32] C. Morbidoni, A. Cucchiarelli, and D. Ursino, "Leveraging linked entities to estimate focus time of short texts," in *Proceedings of the 22nd International Database Engineering & Applications Symposium*, 2018, pp. 282–286.

[33] A. Jatowt, C.-m. A. Yeung, and K. Tanaka, "Estimating Document Focus Time," *CIKM*, pp. 2273–2278, 2013.

[34] J. Allan, J. G. Carbonell, G. Doddington, J. Yamron, and Y. Yang, "Topic detection and tracking pilot study final report."

[35] J. Allan, R. Papka, and V. Lavrenko, "On-line new event detection and tracking," in *Proceedings of SIGIR*, 1998, pp. 37–45.

[36] I. Soboroff, S. Huang, and D. Harman, "Trec 2018 news track overview," in *The Twenty-Seventh Text RE-trieval Conference (TREC 2018) Proceedings*, 2018.

[37] C. Brennan, "Digital humanities, digital methods, digital history, and digital outputs: History writing and the digital revolution," *History Compass*, vol. 16, no. 10, p. e12492, 2018.

[38] A. Pederson, "Do real archivists need archives & museum informatics?" *The American Archivist*, vol. 53, no. 4, pp. 666–675, 1990.

[39] J. Stiller, "A framework for classifying interactions in cultural heritage information systems," *International Journal of Heritage in the Digital Era*, vol. 1, no. 1, pp. 141–146, 2012.

[40] C. Warwick, M. Terras, P. Huntington, and N. Pappa, "If you build it will they come? the lairah study: Quantifying the use of online resources in the arts and humanities through statistical analysis of user log data," *Literary and linguistic computing*, vol. 23, no. 1, pp. 85–102, 2007.

[41] A. Pasquali, V. Mangaravite, R. Campos, A. Jorge, and A. Jatowt, "Interactive system for automatically generating temporal narratives," in *Proceedings of ECIR*. Springer, 2019, p. 251–255.

[42] Y. Zhang, A. Jatowt, S. S. Bhowmick, and K. Tanaka, "The past is not a foreign country: Detecting semantically similar terms across time," *IEEE TKDE*, vol. 28, no. 10, pp. 2793–2807, 2016.

[43] Y. Duan, A. Jatowt, S. S. Bhowmick, and M. Yoshikawa, "Mapping entity sets in news archives across time," *Data Science and Engineering*, vol. 4, no. 3, pp. 208–222, 2019.

[44] A. Jatowt, I.-C. Hung, M. Färber, R. Campos, and M. Yoshikawa, "Exploding tv sets and disappointing laptops: Suggesting interesting content in news archives based on surprise estimation." in *ECIR (1)*, 2021, pp. 254–269.

[45] V. de Boer, J. Wielemaker, J. van Gent, M. Hildebrand, A. Isaac, J. van Ossenbruggen, and G. Schreiber, "Supporting linked data production for cultural heritage institutes: The amsterdam museum case study," in *ESWC*, ser. ESWC'12. Springer-Verlag, 2012, pp. 733–747.

[46] J. Galtung and M. H. Ruge, "The structure of foreign news: The presentation of the congo, cuba and cyprus crises in four norwegian newspapers," *Journal of Peace Research*, vol. 2, no. 1, pp. 64–90, 1965.

[47] M. Pia and J. Snajder, "Linguistic Features and Newsworthiness : An Analysis of News Style," *Proceedings of the Fourth Italian Conference on Computational Linguistics*, 2013.

[48] T. D. Nies, D. Evelien, S. Coppens, D. V. Deursen, S. Paulussen, and R. V. D. Walle, "Bringing Newsworthiness into the 21st Century," *Web of Linked Entities, Workshop proceedings*, 2012.

[49] N. Tahmasebi, G. Gossen, N. Kanhabua, H. Holzmann, and T. Risse, "Neer: An unsupervised method for named entity evolution recognition," in *Proceedings of COLING 2012*, 2012, pp. 2553–2568.

[50] J. Strötgen and M. Gertz, "Domain-sensitive temporal tagging," *Synthesis Lectures on Human Language Technologies*, vol. 9, no. 3, pp. 1–151, 2016.

[51] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," *EMNLP*, pp. 404–411, 2004.

[52] M. Ferron and P. Massa, "Collective memory building in wikipedia: The case of north african uprisings." Mountain View, California, USA: WikiSym '11, 2011, pp. 114–123.

[53] N. Kanhabua, T. N. Nguyen, and C. Niederée, "What triggers human remembering of events?: A large-scale analysis of catalysts for collective memory in wikipedia." JCDL '14, 2014, pp. 341–350.

[54] A. Jatowt, D. Kawai, and K. Tanaka, "Digital history meets wikipedia: Analyzing historical persons in wikipedia," in *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, 2016, p. 17–26.

[55] D. J. Gerner, P. A. Schrodt, Ö. Yilmaz, and R. Abu-Jabr, "The Creation of CAMEO (Conflict and Mediation Event Observations): An Event Data Framework for a Post Cold War World," *Annual meeting of the American Political Science Association*, 2002.

[56] C.-m. Au Yeung and A. Jatowt, "Studying how the past is remembered: towards computational history through large scale text mining," in *Proceedings of CIKM*. ACM, 2011, pp. 1231–1240.

[57] E. Sandhaus, "The New York Times Annotated Corpus," *Linguistic Data Consortium, Philadelphia*, vol. 6, no. 12, p. e26752, 2008.

[58] R. K. Pasumarthi, S. Bruch, X. Wang, C. Li, M. Bendersky, M. Najork, J. Pfeifer, N. Golbandi, R. Anil, and S. Wolf, "TF-Ranking: Scalable TensorFlow Library for Learning-to-Rank," *SIGKDD*, 2019.

[59] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*, 2014, pp. 1188–1196.