What makes a query temporally sensitive?

No Author Given

No Institute Given

Abstract. Keywords: We would like to encourage you to list your keywords within the abstract section

1 Introduction

A basic intuition in temporal information retrieval research is that time should be modeled explicitly when scoring and ranking documents with respect to users' queries. Time-related criteria such as recency, currency, and freshness have long been recognized as factors in studies of relevance[2]. Based on these criteria, information retrieval researchers have explored a variety of temporal retrieval models [11, 5, 4] that explicitly incorporate time into the document ranking. These researchers refer to general classes of "temporal queries" or "temporal information needs." Models have been proposed and evaluated for "recency queries" [11, 5], "time-sensitive queries" [4], "implicitly temporal queries" [12], and "temporally biased queries" [7]. For evaluation, these studies rely on manual classifications of topics into temporal categories.

In this paper, we take a deeper look into these manually classified topics to develop a clearer understanding of *what makes a query temporally sensitive*? While many researchers have relied on these manually classified topics, the methods used for classification are not clearly explained. Researchers often refer to vague concepts of "newsworthiness" or a "bona fide temporal dimension" of topics without clear explanation of the criteria or processes used in classification. As such, using these manually classified topics for evaluation is of limited value, since we cannot be clear on the process that is being modeled. The study presented in this paper is a necessary step in temporal information retrieval research.

To address our question, we analyze 660 topics from the Text Retrieval Conference (TREC) previously used in the experimental evaluation of temporal retrieval models. We employ qualitative techniques to identify characteristics of topics that might affect manual assessment of "temporality." The resulting coded topics are used in a set of regression analyses to determine the specific relationships between these characteristics and manually assigned categories. Finally, we use the coded topics to predict when temporally-sensitive retrieval models might be effective.

This paper is structured as follows....

2   Background

2.1   What is a TREC topic?

The test collections developed for TREC generally consist of a collection of documents, a set of topics, and a set of relevance judgments. Although the methods of creation have changed over time, TREC topics are generally intended to mimic real users' information needs. Topics are contained in a set of structured text files associated with the test collection and generally consist of a title and description. Some collections include more detailed narratives, while others may have only a short title or query string without a detailed description of the information need.

The concept of an *information need* has been extensively discussed in the information science literature and is widely used in the information retrieval community. Here we define an *information need* as the unobservable motivation behind individual users' information seeking. Information needs reflect the individual user's current state of mind and context, including social and cultural environments. While information needs themselves are not observable, user's information seeking behaviors – including queries and document relevance judgements – are incomplete but observable evidence of the underlying need. We believe that this definition captures the spirit of the concept of *information need* as commonly used in information retrieval research, particularly in the case of "topics" in TREC.

2.2   Time and relevance

There are numerous notions of temporality in information retrieval research, each of which requires different methodologies for analysis. In this study, we are primarily concerned with what we term as *temporal relevance*. We define this as the condition where information needs are satisfied by documents published at particular points in time. We distinguish this from *temporal topicality* which refers to information needs that are satisfied by documents about certain periods in time. Of course, an information need may combine the two conditions. Examples of studies concerned with temporal topicality include [3, 8].

There are, of course, other notions of temporality in information retrieval research. For example, *temporal relevance dynamics* is concerned with how the relevance of a document changes with respect to an information need over time [13]; *temporal query dynamics* is concerned with the study of how queries are issued over time [16, ?,14, 10]; and *temporal content dynamics* is concerned with the effects of content changes over time [9].

In the next section we review examples of studies focused on *temporal relevance*.

2.3   Time-sensitive queries

Temporal relevance is concerned with conditions where documents published at particular points in time are considered more or less relevant than those pub-

lished at other times. This condition is studied independent of or in conjunction with traditional topical relevance. In this section, we review four studies addressing temporal relevance.

Li and Croft's study of "recency queries" [11] was the first in this line of research. They hypothesize that some queries are "recency queries" where the most recently published documents are more likely to be relevant. They analyze queries associated with TREC topics 301-400 from TREC disks 4 and 5. Through the direct analysis of the temporal distribution of judged relevant documents, they classify 36 of the queries as recency queries because they have "more relevant documents in the recent past."

Jones and Diaz [7] study the temporal characteristics of queries with the goal of query classification through the analysis of three TREC news collections and a web search engine log. They use the Associated Press (1988, 1989) and Wall Street Journal (1990-1992) collections from TREC disks 1 and 2 as well as the AQUAINT corpus with the 2003 Novelty track topics. They define three classes of queries based on manual analysis of topics: temporally ambiguous (requesting multiple events), temporally unambiguous (requesting a single event), and atemporal (having no preference). They employed annotators to manually classify 100 TREC queries based only on the topic title, description, and narrative. Specific criteria for classification are not given. For the AP and WSJ collections, they found that all of the queries were either atemporal or temporally ambiguous. As a result, they incorporated the 2003 Novelty track because it includes topics classified as "event" or "opinion," which the authors found to correspond to the "temporally unambiguous" and "atemporal" categories.

Expanding on the work of Li and Croft, Dakka et al [4] investigate a broader class of queries which they refer to as "time-sensitive." They hypothesize that there are queries for which more relevant documents are found at specific points in time, not just recently. They evaluate their models using a subset of TREC disks 4 and 5, manually identifying a subset of topics 301-450 that they consider to be "time-sensitive." To do so, they manually examine the title, description and narrative of each topic and identify queries associated with specific news events. If the topic information is insufficient to make a decision, they analyze the distribution of relevant documents. This resulted in a collection of 86 temporally sensitive queries. The document collection used for evaluation was restricted to the Financial Times (1991-1994) and Los Angeles Times (1989-1990) subcollections.

Also expanding on the work of Li and Croft, Efron and Golovchinsky [5] investigate additional models for recency queries. They use subsets of several TREC ad-hoc collections including the Associated Press documents from disks 1 and 2 with topics 101-200; Los Angeles Times and Financial Times documents from disks 4 and 5 with topics 301-450. Topics were classified as "recency" if at least 2/3 of the relevant documents occurred after the median document time and the topic had a "bona fide" temporal dimension based on manual review. The specific criteria for manual review are not specified.

Finally, Peetz, Meij, and Rijke [15] investigate the effect of temporal bursts in estimating query models. Building on the earlier studies, they evaluate their models using the above test collections as well as a new collection based on TREC Blog06. As in the previous studies, the authors construct a subset of "temporal" queries through manual evaluation of topic descriptions and relevant document distributions. No specific criteria for classification are given.

2.4   What makes a query temporally sensitive?

We return to the motivating question: what makes a query temporally sensitive? Dakka et al [4] present a compelling definition. A query is "time sensitive" if "the relevant documents for the query are not spread uniformly over time, but rather tend to be concentrated at restricted intervals." In other words, a query is temporally sensitive if relevant documents are more likely to occur at some points in time more than others. This is an essential point, since many temporal retrieval models rely on the temporal distribution of results in document scoring. However, the distribution of relevant documents alone is not sufficient to determine true temporality. To address this, researchers rely on common-sense notions of temporality based on the topic itself considered independent of the distribution of relevant documents. Dakka et al refer to "newsworthiness"; Efron and Golovchinksy to a "bona fide temporal dimension"; and Jones and Diaz to "events." A goal of the current study is to look deeper into these common-sense criteria.

In the next section we report the methods used in our analysis of 660 TREC topics used in temporal information retrieval research.

3   Methods

In the studies reviewed above, researchers rely on existing test collections, such as those available through TREC, to evaluate temporal retrieval models. In each study, topics are manually categorized as temporal or non-temporal to assess model performance. The purpose of this study is to further investigate the characteristics of topics deemed temporal. To achieve this, we use a combination of qualitative content analysis and regression analysis, as described below.

3.1   Qualitative coding

We use content analysis [?] to identify characteristics of TREC topics potentially associated with temporal sensitivity. 660 topics were selected from the TREC Ad-hoc, Novelty, Blog, and Microblog tracks. These topics were selected because they have previously been used by researchers to evaluate temporal retrieval models and have associated manual classifications or are from collections with known temporal characteristics (i.e., Microblog). The complete set of topics used in this study are listed in Table 1 along with the temporal constraints of each collection or sub collection.

Topics	Collections
51-200	TREC Disks 1-2 AP (88-89); WSJ (87-92)
301-450	TREC Disks 4-5 FT (91-94); LA Times (88-89)
N1-100	AQUAINT Xinhua (1996-2000); NYT (1999-2000); AP (1999-2000)
851-1050	Blog06 (Dec 6, 2005 - Feb 21, 2006)
MB1-110	Tweets 2011 (Jan 24, 2011 - Feb 8th, 2011)

Table 1: TREC topics and collections used in this study

Two of the authors participated in the development of the codebook and subsequent coding of topics. Codebook development began with a preliminary reading of all topic titles, descriptions and narratives. Codes were defined based on characteristics of topics expected to be related to temporal sensitivity, informed by the literature. Of the 660 topics, 330 were coded by both coders. During this process, code definitions were refined and clarified. In the final coding, only topic title and description were used. The final codebook is presented in Table 3 in the appendix. Coding was completed using the Dedoose[1] service.

An example of a coded topic from the 2004 Novelty test collection is presented in Figure 1. This topic refers to a specific event and contains place entities as well as an explicit date. Topic N57 is categorized as an "event" by the TREC topic creator and is therefore an unambiguous temporal topic as defined by Jones and Diaz.

$$\textbf{Title:} \left[(\text{East Timor})_{PlaceEntity} \text{Independence} \right]_{SpecificEvent}$$

$$\textbf{Description:} \left[(\text{East Timor})_{PlaceEntity} \text{ vote for independence from} \right.$$
$$\left. (\text{Indonesia})_{PlaceName} \text{ in } (\text{August 1999})_{ExplicitDate} \right]_{SpecificEvent}$$

Fig. 1: TREC Novelty 2004 topic N57 example annotation

In addition to coding the topics based on the defined codes, the coders assigned a temporal designation to the distribution of relevant documents for each topic. Graphs were generated for each topic with more than 20 relevant documents. A non-parametric density is fit to the relevant document distribution. Each coder reviewed the relevant document distribution along with the total number of relevant documents for each topic and assigned one of four values: too few observations (-1), low or no temporality (0), moderate temporality (1), and high temporality (2). These designations are included as features in the regression analysis and compared to the ACF and DPS values used as proxies of the relevant document distribution.

---

[1] http://www.dedoose.com

After coding all 660 topics, the topic/code matrix was exported for subsequent reliability and regression analysis, as described in the following sections.

3.2   Reliability analysis

For this study, coding reliability is measuring using Cohen's κ for classification of the relevant document. A variation of percent overlap is used for the broader qualitative coding, since coding is performed on arbitrary segments of text in each topic. We define the *percent overlap* as:

$$overlap = \frac{m}{m + u_1 + u_2}$$

Where m is the number of excerpts assigned the same code by both coders, u_1 is the number of codes assigned to excerpts only by coder 1 and u_2 is the number of codes assigned to excerpts only by coder 2. If both coders assign no codes to a topic, this is considered perfect agreement. We report the macro overlap calculated over all topics, the micro overlap calculated as a per-topic average, and per-code overlaps to understand coder agreement within each category.

For the final designation of topic temporality, coder agreement is measured by Cohen's κ.

3.3   Relevant document distributions

In each of the four prior studies, the authors acknowledge using the distribution of judged-relevant or pseudo-relevant documents in determining topic temporality. For this study, we use two different measures to represent these distributions: the first-order time series autocorrelation (ACF) and the dominant power spectrum (DPS).

Jones and Diaz [7] use the ACF created by the temporal distribution of pseudo relevant documents for a query as a predictor of query temporality. They note that queries with strong inter-day dependencies will have high ACF values, indicating predictability in the time series. The first-order autocorrelation of the time series is defined as:

$$r_1 = \frac{\sum_{i=1}^{N-1}(x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\left[\sum_{i=1}^{N-1}(x_t - \bar{x}_{(1)})^2\right]^{1/2}\left[\sum_{i=1}^{N-1}(x_{t+1} - \bar{x}_{(2)})^2\right]^{1/2}} \tag{1}$$

He, Chang, and Lim [?] use the DPS as a predictor of the "burstiness" of temporal features for event detection. The DPS is the highest power spectrum, estimated using the periodogram. The periodogram is the sequence of the squared magnitude of the Fourier coefficients $\|X_k\|^2$ indicating the signal power at frequency k/T in the spectrum. The discrete Fourier transform is defined as:

$$X_k = \sum_{t=1}^{T} y_f(t)e^{\frac{2\pi i}{T}(k-1)t}, t = 1, 2, ...T \tag{2}$$

In this study, both of these measures are used to represent the distribution of judged-relevant or pseudo-relevant documents for the regression analysis, as described in the next section.

3.4 Regression analysis

A primary goal of this study is to determine the characteristics that contribute to the judgment of topic temporality. We use logistic regression using the generalized linear model (GLM) implementation in R. The predictors are binary presence indicators for each of the qualitative codes along with the first-order autocorrelation and dominant power spectra of the temporal distribution of true-relevant documents. The response variables are the binary temporal/non-temporal indicators manually assigned in the four studies. Model variables are selected using standard step-wise procedures based on the Akaike information criterion (AIC). Coefficients are reported using the log-odds and model fit is assessed using pseudo R^2.

3.5 Predicting temporal model effectiveness

Next, we develop standard linear regression models to predict whether to use a temporal retrieval model for each topic. We use the standard query likelihood [?] and kernel density estimate (KDE) temporal mode [6] to measure query temporality. The response variable is the difference in average precision (AP) between KDE and QL for each topic. The predictors are the assigned codes represented by binary indicators and the ACF and DPS of the temporal distribution of the top 1000 pseudo-relevant documents using the QL score.

4 Results

What are the characteristics of topics that affect the manual assessment of topic temporality?

4.1 Codes

Our qualitative analysis suggests three broad classes: events, named entities, and explicit dates. A basic intuition is that topics focused on a specific and important events will have a higher degree of temporal relevance. Following the TDT definition, seminal events happen at specific times in specific places, often to individuals or other named entities (e.g., organizations). Perhaps the most essential code is the "SpecificEvent" – something important that happens at a particular time and place. Related to SpecificEvent is the PeriodicEvent, which refers to an event that recurs periodically, such as the Super Bowl, World Cup, or Olympics. Jones and Diaz [7] note that many of the early ad-hoc queries were temporally ambiguous, referring to multiple events. We incorporate this concept through the "GenericEvent" code, which captures topics concerned with a class

of specific events, such as earthquakes, elections, or strikes. While analyzing topics, it became apparent that some topics were likely to be inspired by a specific event, but the event is not referenced in the topic description. This concept is captured through the "IndirectEventReference" code. The remaining codes are concerned with the identification of specific types of named entities, which are expected to have some association with topic temporality.

4.2   Code distributions

Figure 2 summarizes the percent of topics in each test collection with each code assigned. From these results, we can see that the Novelty and Microblog collections have a higher percentage of specific events than the Blog and ad-hoc collections. The ad-hoc collections have a higher number of generic events, which supports the findings of Jones and Diaz [7]. The Blog, Novelty, and Microblog test collections each have larger numbers of named entities in the topic titles and descriptions.
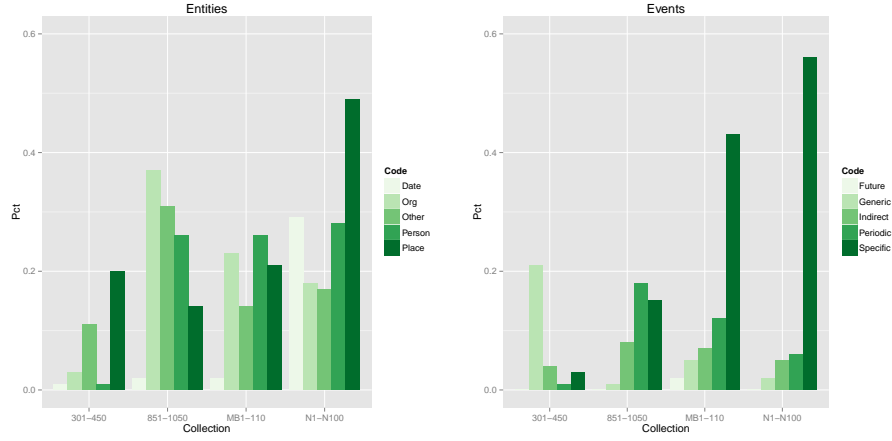

Fig. 2: Percent of topics in each collection with codes assigned from the (left) entity code group and (right) events code group

4.3   Reliability

To assess coding reliability, a total of 1244 codes were assigned to 330 topics by the two coders. The macro percent overlap is 0.71 and micro percent overlap is 0.83. The per-code overlap is reported in Figure 3. Higher overlap indicates greater agreement between coders. As expected, some codes have higher agreement than others. Specifically, personal names (0.94), locations (0.91), and explicit dates (0.89) have very high agreement whereas indirect event references (0.19) and generic events (0.45) have lower agreement.
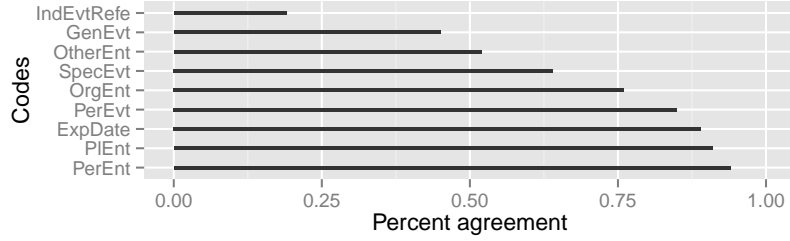
Fig. 3: Percent overlap by code

4.4 Regression analysis

In this section, were report the results of a logistic regression analysis, predicting the manually assigned categories for each test collection. The resulting models are reported in Table 2.

Name	Model	AIC	R^2
Novelty	$-3.767 + 5.848 * SpecEvt + 2.523 * Other$	52	0.669
Novelty (Rel)	$-3.539 + 7.006 * SpecEvt + 2.530 * Other - 7.343 * ACF$	49	0.706
Dakka	$0.134 + 0.878 * Place$	205	0.020
Dakka (Rel)	$-0.917 + 0.393 * DPS_{FT}$	155	0.263
Efron	$-1.765 + 2.353 * Place + 1.410 * Other$	150	0.181
Efron (Rel)	$-2.727 + 1.965 * Place + 1.787 * Other + 0.163 * DPS_{FT}$	118	0.377
Peetz	$-0.336 + 1.682 * SpecEvt + 0.982 * PerEvt + 0.672 * Person - 0.6175 * OrgEnt$	192	0.127
Peetz (Rel)	$-1.245 + 1.218 * SpecEvt + 0.797 * Person + 2.835 * ACF + 0.002 * DPS$	171	0.223

Table 2: Logistic regression models for each test collection with and without (Rel) ACF/DPS variables. Model fit reported based pseudo-R^2 after stepwise variable selection based on AIC.

For the 2003-2004 Novelty collection, the response variable is the manually assigned "opinion" (0) or "event" (1) categories. Following Jones and Diaz [7], we adopt "event" as the temporal category. Logistic regression analysis is performed with and without the ACF and DPS variables. SpecificEvent and OtherEntity are significant predictors of the "event" category (p ¡ 0.01), with a pseudo-R^2 of 0.669. Including the ACF of the true-relevant distribution is significant, with a minor improvement in model fit. The high pseudo-R^2 is unsurprising in this case, since the SpecificEvent code corresponds to the Novelty "event" category. It does, however, confirm our code definition.

Dakka et al manually classified "time-sensitive queries" for TREC topics 301-450. As reported in Table 2, only the PlaceEntity code is a significant predictor of the manual classification. However, the pseudo-R^2 is very low (0.02). Dakka et al acknowledge examining the relevant document distributions for LA Times

and Financial Times sub collections. Including the DPS of the true-relevant document distribution in Financial Times (DPS_{FT}) increases the pseudo-R^2 to 0.263, suggesting that the relevant document distribution played a significant role in the manual classification.

Efron and Golovchinsky also classified topics 301-450, in this case focusing on the identification of "recency" queries. As reported in 2, both PlaceEntity and OtherEntity are useful predictors of the temporal response. As with Dakka, including DPS_{FT} increases pseudo-R^2 from 0.181 to 0.377. This again suggests that the distribution of relevant documents played an important role in the determination of topic temporality.

Finally, we look at Peetz et al's classification of the Blog06-08 topics 850-1050. In this case, the SpecificEvent, PeriodicEvent, Person and Organization entities are useful predictors of the temporal category (pseudo-R^2=0.127). Including DPS improves model fit (pseudo-R^2=0.223), again suggesting that the distribution of relevant documents played a role in manual classification.

4.5   Relevant document distributions

In this section, we report the results of the manual classification of the temporal distribution of true-relevant documents for each collection. All topics were coded by both coders into the four categories described in section X. The weighted Cohen's κ is used for inter-coder reliability. AP=0.554, Blog06=0.857, Tweets=0.482 indicating moderate to high agreement.

5   Conclusions

In this study, we have tried to identify the factors or characteristics of information needs that can be used to predict manual classifications of temporal queries. We performed a qualitative analysis of over 660 topics used in temporal information retrieval research, identifying candidate predictors and performing a series of regression analyses to quantify our ability to explain previous manual classifications of topics.

We were successful in modeling the "event" category in the 2003-2004 Novelty test collection, primarily through the identification of specific events in the topic titles and descriptions. The resulting logistic regression model was able to explain 70% of the variance in the event/opinion categorization. In the three other collections created by Dakka, Efron, and Peetz respectively, our qualitative codes were able to explain 2%, 18%, and 13% of the variation in manual temporal classification respectively. Incorporating the temporal distribution of judged-relevant documents, this increases to 26%, 38%, and 22% respectively. This indicates that the distribution of relevant documents plays a central role in the judgment of temporal sensitivity in these collections.

While these results are promising, we were unable to identify characteristics of these topics that fully explain the manual classification decisions. If we cannot explain the process the determines the classifications, it raises questions about

the value of these test collections for evaluation. Specifically, how can we be clear that the queries previously identified as "temporally sensitive" are indeed so?

(More needed)

6   Acknowledgments

This section is optional

References

1. G. Amodeo, G. Amati, and G. Gambosi. On relevance, time and query expansion. *Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11*, page 1973, 2011.
2. C. Barry and L. Schamber. Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, 34(2):219–236, 1998.
3. K. Berberich, S. Bedathur, O. Alonso, and G. Weikum. A Language Modeling Approach for Temporal Information Needs. In *Proceedings of the 32Nd European Conference on Advances in Information Retrieval*, pages 13–25, 2010.
4. W. Dakka, L. Gravano, and P. Ipeirotis. Answering General Time-Sensitive Queries. *IEEE Transactions on Knowledge and Data Engineering*, 24(2):220–235, Feb. 2012.
5. M. Efron and G. Golovchinsky. Estimation methods for ranking recent information. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information - SIGIR '11*, page 495, 2011.
6. M. Efron, J. Lin, J. He, and A. D. Vries. Temporal Feedback for Tweet Search with Non-Parametric Density Estimation. *umiacs.umd.edu*, pages 33–42, 2014.
7. R. Jones and F. Diaz. Temporal profiles of queries, July 2007.
8. N. Kanhabua and K. Nø rvåg. A comparison of time-aware ranking methods. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information - SIGIR '11*, page 1257, 2011.
9. A. Kulkarni, J. Teevan, K. M. Svore, and S. T. Dumais. Understanding temporal query dynamics. *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*, page 167, 2011.
10. V. Lavrenko, M. Schmill, and D. Lawrie. Mining of concurrent text and time series. *. . . on Text Mining*, pages 2–9, 2000.
11. X. Li and W. B. Croft. Time-based language models. *Proceedings of the twelfth international conference on Information and knowledge management - CIKM '03*, page 469, 2003.
12. D. Metzler, R. Jones, F. Peng, and R. Zhang. Improving search relevance for implicitly temporal queries. *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '09*, (1):700, 2009.
13. S. Mizzaro. Relevance: The whole history. *Journal of the American Society for Information Science*, 48(9):810–832, 1997.
14. N. Parikh and N. Sundaresan. Scalable and near real-time burst detection from eCommerce queries. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, page 972, 2008.

15. M. Peetz and M. D. Rijke. Cognitive temporal document priors. *Advances in Information Retrieval*, 2013.
16. M. Shokouhi. Detecting Seasonal Queries by Time-Series Analysis Categories and Subject Descriptors. In *SIGIR 11*, pages 1171–1172, 2011.

Appendix: Codebook

Code	Description	Examples
Specific Event	Something significant that happens at a specific time and place. Code title and description in concert, even if title does not contain event specifics.	Mount Pinatubo erruption on June 15, 1991; 2008 State of the Union; Hurricane Hugo
Generic Event	Use this code when the topic refers to more than one specific event or a class or type of event. Only use this code if every instance of the event type would be newsworthy (i.e., a specific event) and the central topic of a news article.	Earthquakes, volcano erruptions, elections, disputes, strikes
Indirect Event Reference	Apply this code to indicate when the topic might be indirectly referring to a *specific* event. Use only if you need to turn to external information to identify potential specific events (e.g., your personal knowledge, wikipedia). Do not use if specific event information is contained in the description.	Legally assisted suicide, related to Kevorkian controversy. Partial birth abortion ban, related to partial birth abortion ban legislation. Surrogacy related to Baby M.
Periodic Event	Apply this code to indicate when an event is periodic, recurring at regular, predictable intervals. Never double-code as SpecificEvent or as an entity even though periodic events are often named entities.	Super bowl, Nobel awards, Oscars, State of the Union
FutureEvent	Apply this code to indicate when a topic refers to a future predicted specific event. Never double-code as SpecificEvent.	2020 Fifa, 2016 Summer Olympics
Person Entity	Apply this code to identify personal names in topics.	President Bush; Sasha Cohen;
Place Entity	Apply this code to identify places in topics. Limit to proper names. Also apply to references to nations, governments or government bodies.	Peru; Africa; African; European; Atlanta
Organization Entity	Apply this code to identify organizations in topics. Limit to proper names. Do not apply to references to governments (e.g., United States), use PlaceEntity instead.	Hitachi Data Systems; U.S. Congress;
Other Entity	Apply this code to named entities that are not people, places or organizations. Limit to proper names. Includes movies, books., etc.	Hubble Telescope; The Avengers; Euro
Explicit Date	Apply this code to identify explicit dates.	1988; June 15, 1991; October 2007; Monday

Table 3: Codebook