# Embedding a Microblog Context in Ephemeral Queries for Document Retrieval

Shilpa Sethi

*Department of Computer Applications, J. C. Bose University of Science and Technology, Faridabad, Haryana, India*
*E-mail: munjal.shilpa@gmail.com*

## Abstract

With the proliferation of information globally, the search engine had become an indispensable tool that helps the user to search for information in a simple, easy and quick way. These search engines employ sophisticated document ranking algorithms based on query context, link structure and user behavior characterization. However, all these features keep changing in the real scenario. Ideally, ranking algorithms must be robust enough to time-sensitive queries. Microblog content is typically short-lived as it is often intended to provide quick updates or share brief information in a concise manner. The technique first determines if a query is currently in high demand, then it automatically appends a time-sensitive context to the query by mining those microblogs whose torrent matches with query-in-demand. The extracted contextual terms are further used in re-ranking the search results. The experimental results reveal the existence of a strong correlation between ephemeral search queries and microblog volumes. These volumes are analyzed to identify the temporal proximity of their torrents. It is observed that approximately 70% of search torrents occurred one day before or after blog torrents for lower threshold values. When the threshold is increased, the match ratio of

torrent is raised to ∼90%. In addition, the performance of the proposed model is analyzed for different combining principles namely, aggregate relevance (AR) and disjunctive relevance (DR). It is found that the DR variant of the proposed model outperforms the AR variant of the proposed model in terms of relevance and interest scores. Further, the proposed model's performance is compared with three categories of retrieval models: log-logistic model, sequential dependence model (SDM) and embedding based query expansion model (EQE1). The experimental results reveal the effectiveness of the proposed technique in terms of result relevancy and user satisfaction. There is a significant improvement of ∼25% in the result relevance score and ∼35% in the user satisfaction score compared to underlying retrieval models. The work can be expanded in many directions in the future as various researchers can combine these strategies to build a recommendation system, auto query reformulation system, Chatbot, and NLP professional toolkit.

**Keywords:** Page ranking, microblogs, temporal queries, query context, ephemeral information.

## 1 Introduction

Retrieval of information from search repositories is highly dependent on keywords supplied by the user in its query. The pages are retrieved by matching the query keywords with the term weight vector of the documents archived in the search engine's database [20]. The scheme performs outstandingly when the user transforms their information needs into a well-defined set of keywords. However, it has been observed in a study done on the Alta Vista query log that the average query length issued by the user is 2.31 words with less than 4% of the queries having six terms and nearly 70% of queries having only a single term [21]. So, speculating the information on such a condensed representation leads to the inclusion of the wrong document in search results.

Here, the concept of anticipating the context of user search may play an important role. Further, deducing the context of a user query is not an easy task. The reasons are synonymy and polysomy of words (keywords). Many keywords exist in literature having different meanings to different people in different contexts at different times. So, finding the context of the query is a subjective, intent-specific time-varying concept.

For instance, in today's scenario, if a user issues a query about CORONA, the users would be interested in knowing about the recent COVID-19 deadly virus and its variants. On the other hand, if no such event occurred in

2019–2021, the people may not be concerned about its cause, prevention and vaccination. Analogously, user interest sifted to Ukraine when Russia invaded and occupied part of Ukraine in 2022. If these events had not happened in recent years, people may not be bothered about COVID or the Russo-Ukraine war. These situations necessitate the need for the search engine to rearrange the search results as per the temporal interest of the user.

Thus, the work aims to devise a technique that can identify ephemeral popular queries, understand their search context and accordingly provide up-to-date information that best fits the user's interest.

Here, a query is said to be in the ephemeral popular state if some unusual event related to a query occurred in a short duration due to which the volume of query in search logs floods out. In such situations, if the search engine follows the traditional ranking algorithms and provides the search results based on query terms only, the user's information need will not be fulfilled. So, it becomes essential for the search engine to identify those social web resources (as social resources are a good reflector of human behavior and interest) that are also affected by the same cause.

Since microblogging is a popular activity where people share current events or time-sensitive information, it can serve the further purpose of this research. Some studies also show the credibility of information in time-sensitive social media like Twitter and Facebook [10]. In fact, work in [11] identified user temporal location using microblogging services [34]. Examples of such services also include predicting electoral results using Twitter [12], space-time analytics for crisis management [13] and real-time disaster detection by social sensors [15].

The objective of this work is to analytically observe the correlation between the search volumes and tweet volumes on a topic at varying times. The experimental results show that these two volumes are affected by the same cause for ephemeral popular queries. In addition, the contextual terms to query are determined by extracting neighboring words to a user query using microblogs. Finally, the retrieved documents from the search repository are rearranged based on contextual terms and the sorted list is presented back to the user. A survey conducted among the research scholars of the same domain revealed a significant improvement in satisfaction degree for the proposed system as compared to recent contextual models.

The major contributions of the proposed work are given below:

1. Develops a novel document retrieval model based on microblog context for ephemeral queries.

2. Exhaustive analysis is performed to discover the relationship between temporal search and Tweet content. It is analyzed that 70% of the temporal queries are submitted to a search engine one day before or after the Tweet about the event. The correlation becomes stronger on higher torrent points.
3. A significant improvement of $\sim$25% in result relevance score is recorded which leads to better search engine performance.
4. The use satisfaction score is raised by $\sim$35%.
5. Opens up a new research direction in the field of information retrieval, recommendation engines, NLP and Chatbots.

The rest of the paper is organized as follows: Section 2 covers some of the prevalent ranking algorithms. The proposed work is presented in Section 3. The significance of the proposed model over the baseline models is discussed in Section 4. The performance analysis is presented in Section 5. Section 6 concludes the work and provides future directions in the area.

## 2  Background

Sorting the search results in accordance with the user's needs is an inherent task that a search engine has to deal with. Various studies have shown that exploring context information embedded in search tasks and strengthening the retrieval models is effective [7, 8, 10]. A rich body of research has explored different forms of context related to search tasks and built predictive models over it. The related work may be categorized into two groups: query-driven and user-driven models.

**Query-driven models:** These models focus on deriving the contextual meaning of user queries and utilize them either at query submission or rank computation stage. Some of the early query models made use of the power of lexical databases such as Wordnet, PROBASE and Wikipedia categories for query disambiguation [6]. Shen et al. deduced search context by analysis of users' past queries from search logs and related click-through data and used the information to re-rank documents for future queries [5]. In [4], semantic change between consecutive queries and the relationship between the changed query and the clicked document is used to infer query context. In addition, query clustering [3], geographical location [15], and association rules [1] are some of the methods used by researchers for better information retrieval. However, we argued that these context extraction methods are confined by the capacity of their employed representation, which is hardly

generalizable and not optimal for retrieval tasks. Moreover, they are limited to how humans interact in the real world. Thus, they can't possess the proper granularity to represent human information needs.

**User-driven models:** These models are centered around the user's perspective. Cao et al. used the variable length hidden in Markov's model to infer user search intent and used it for query suggestion and document ranking [9]. Encouraged by the 'learn to rank' concept, various user-centric neural network models have been developed to personalize search results [16–19]. Recently, Chen et al. proposed a hierarchical neural session-based method to capture user behavior and search context [9]. Google also introduced an open-sourced NLP technique named 'BERT' to dynamically understand the context of a user query [20]. Many blogs are also available; however, the exact algorithm is not still disclosed [22]. In addition, several attempts have also been made by researchers to provide relevant information through microblogs. Yang et al. proposed a method to locate periods of news on popular subjects in microblogs and removed the noise to present messages that are of most interest to the user 114]. Ayan Bandyopadhyay used Google search API (GSA) to retrieve titles of web pages and used these titles to expand the queries for microblog matching [10].

However, it has been observed that existing solutions focus mostly on the static and regular information needs of users and none of these retrieval models paid attention to dependency structure embedded in ephemeral search tasks. In this paper, we proposed a novel contextual model for ephemeral queries by using real-time microblog data which substantially improves query suggestion and document ranking.

## 3 Proposed Methodology

The problem of ephemeral information retrieval converges to scoring a document set, $D = \{d1, d2, \ldots, dn\}$ with respect to temporal query set $Q = \{q1, q2, \ldots, qn\}$ based on context-matching of Q and D. So, we first need to determine those web sources which are rich in a temporal context.

### 3.1 Determining Context Location for Temporal Queries

Wu et al. [24] proposed a query-centric assumption that states that relevant information related to any query occurs in co-occurring terms around query terms in documents. This assumption is further validated in [26] and [27]

through user studies. The work basically evaluated the local context of static queries.

Extending the assumption to suit those temporal queries for which some sudden/interesting event has happened, we propose that query-related context can occur in co-occurring terms to query in microblogs. That is, if some sudden/interesting event related to a query takes place, then the context related to the query in microblog posting and web searching must be the same.

**Validation:** To validate our assumption, an experiment is conducted to analyze the ephemeral patterns in search volumes and microblog volumes. If the context of a query in search volumes and microblogs are affected by the same cause then ephemeral changes in both volumes will follow the same pattern, at least timing in their torrent must match.

For this purpose, ephemeral patterns in search volumes on Google and posting volumes on Twitter are observed for 5000 queries. These queries are randomly selected on the basis of their popularity from 2 million microblogs on Twitter crawled from 5 September 2022 to 5 February 2023. The context classifier defined in [24] is used for finding a statistical relationship between two volumes. It is represented by $\rho$ (rho) and computed by Equation (1).

$$\text{Statistical coefficient}(\rho) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\nu_{s,i} - \mu(\nu_s)}{\sigma(\nu_s)} \right) \left( \frac{\nu_{b,i} - \mu(\nu_b)}{\sigma(\nu_b)} \right) \quad (1)$$

where, *n* denotes number of days of observation, $\nu_{s,i}$ and $\nu_{b,i}$ represent search volumes and microblog volumes on the *i*th day, respectively. $\mu(\nu_s)$ and $\mu(\nu_b)$ denote the average volume of search and microblogs respectively. $\sigma_s$ and $\sigma_b$ represent standard deviation in two underlying volumes. Observed ephemeral patterns for sampled query 'Ukraine attack' and 'Hubble telescope' on Google and Twitter are shown in Figure 1(a) and 1(b) respectively.

It may be observed from Figure 1(a) that ephemeral patterns for query 'Ukraine attack' almost follow the same path in both media. In addition, torrent in both volumes occurs almost at the same time. The similarity coefficient for 'Ukraine attack' computed by Equation (1) is 0.964, which conveys that the two volumes are strongly affected by a common cause and hence share much context.

On the other hand, ephemeral patterns of 'Hubble telescope' on Google and Twitter are dissimilar so the context of 'Hubble telescope' in the two volumes may be distinct. This observation is verified through Equation (1).
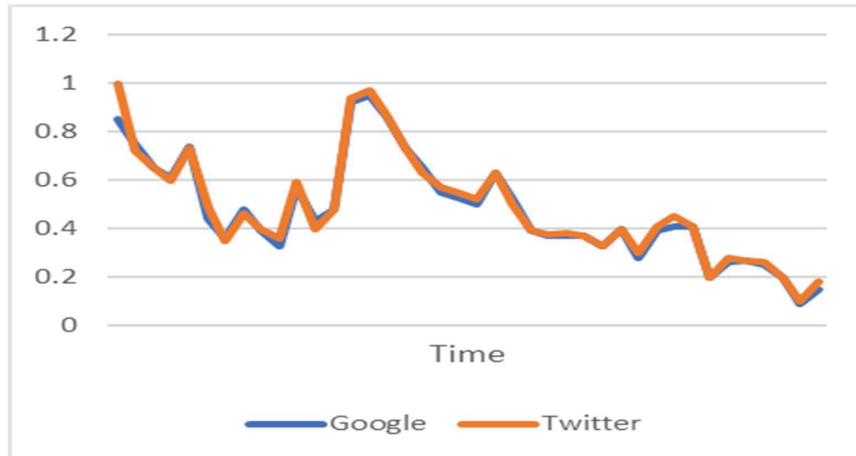
(a)
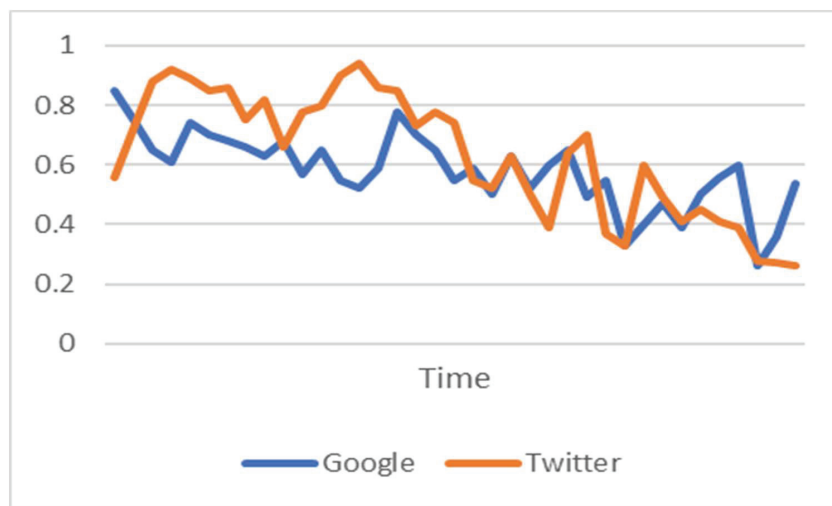


**(b)**

**Figure 1**    Ephemeral patterns of sampled queries on Google and Twitter.

The similarity coefficient for 'Hubble telescope' is 0.204. This implies a poor connection of two contexts.

In order to validate our assumption further, the experiment has to answer the following research questions (RQs).

### RQ1: What point will be considered the torrent point?

The point at which the daily search volume is higher than the average search volume by a threshold '$\tau$' is considered to be the torrent point.

### RQ2: How does the torrent point affect the correlation between search volume and posting volume?

To check the effect of torrent on correlation, the experiment is carried out with different $\tau$ ranging from $\mu + \sigma$ to $\mu + 5\sigma$ over those queries that encountered at least one torrent in search volumes during the observation period.

### RQ3: Is the timing for torrent matches in two volumes?

To compare torrents in two volumes, a test is conducted to check how many torrents in posting volumes timely match with those of search volumes.

The statistics of the constructed data set at different threshold points is given in Table 1. For instance, the values, say, in column 4 indicate that there exist 1087 queries of average length 2.86 words out of 5000 total queries and 98,021 tweets of average length 125 words out of 2 million for which search volumes in Google contain at least one torrent greater than $\mu + 3\sigma$ and the average similarity coefficient for this case is 0.36. If the torrent is increased to a point $> \mu + 5\sigma$, the average similarity coefficient between the two volumes is increased to 0.62.

So, our first observation reveals that search volume and posting volume are more contextually related with the existence of a sharper torrent.

Histograms are plotted for different torrents with similarity coefficients marked on the *X*-axis and number of queries on the *Y*-axis, as shown in Figure 2. It is noted that the similarity coefficient for the test dataset ranges between $[-1,1]$, indicating a strong contextual relationship for value '1' and vice versa. Note that value equals zero indicates that the two volumes are not related to each other.

**Table 1**   Statistics of constructed dataset at different thresholds

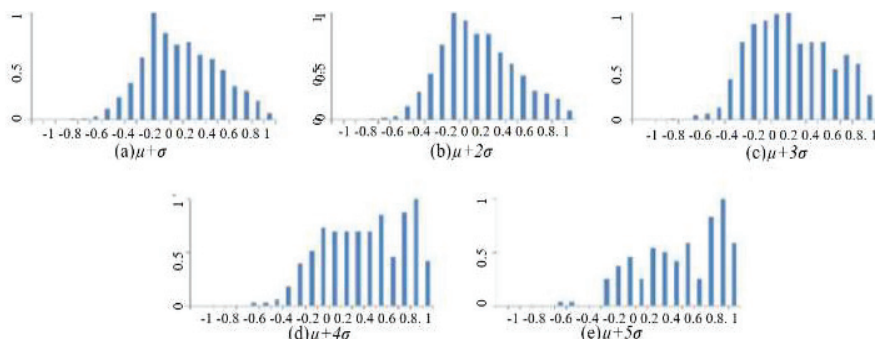| Dataset Spilt | Threshold ($\tau$) | | | | |
|---|---|---|---|---|---|
| | $> \mu + \sigma$ | $> \mu + 2\sigma$ | $> \mu + 3\sigma$ | $> \mu + 4\sigma$ | $> \mu + 5\sigma$ |
| *# Queries* | 3956 | 2013 | 1087 | 574 | 296 |
| *# Tweets* | 566,867 | 365,244 | 98,021 | 77,123 | 64,008 |
| *Avg. query length* | 2.86 | 2.86 | 2.86 | 2.86 | 2.86 |
| *Avg. Tweet length* | 107 | 115 | 125 | 125 | 138 |
| *Avg. sim. coefficient* ($\rho$) | **0.24** | **0.28** | **0.36** | **0.48** | **0.62** |

**Figure 2**   Histogram on similarity coefficient with different $\tau$.

From Figure 2(c) to 2(e), our second observation reveals that the proportion of queries having strong correlation between two volumes increases with increase in threshold. More than 50–70% queries possess a similarity coefficient greater than 0.4.

Further, histograms of both volumes are studied to identify the temporal proximity of their torrents. Our third observation reveals that approximately 70% of search torrents occurred one day before or after blog torrents for $\tau > \mu + 3\sigma$. When the threshold is increased to $\mu + 5\sigma$, the match ratio of torrent is also increased to $\sim$90%.

The analytical results validate our assumption that the context of searching and posting is the same if there exists at least one torrent in both media. It further implies that if the context of the temporal query is obtained by extracting co-occurring terms to query in microblogs and utilized in document ranking then more relevant results can be presented back to the user.

## 3.2  Determining Query State

When a query is submitted by a user to a search engine, first of all its popularity and ephemerality are checked. A query is said to be in a popular and ephemeral state if the current search volume of the query in the last one month is higher than $\mu + 3\sigma$. Here, the lower bound of the threshold is set to $\mu + 3\sigma$ because of two observations:

- More than 50% of popular queries possess a strong correlation at this point.
- More than 70% of torrents in both volumes are timely matched at this point.

So, the state of the query, *Q* on the *i*th day can mathematically be formulated as a unit step function given in Equation (2).

$$\text{State}(Q, i) = \begin{cases} \text{Popular and ephemeral} & \text{if } \nu_Q(i) \geq \mu_Q(i, i-30) \\ & \qquad\qquad + \sigma_Q(i, i-30) \\ \text{Not ephemeral} & \text{otherwise} \end{cases}$$

$$(2)$$

where $\nu_Q(i)$ denotes the search volume of *Q* on the *i*th day; $\mu_Q(i, i-30)$ and $\sigma_Q(i, i-30)$ represent average search volume and standard deviation of Q from the $i-30$th day to the *i*th day.

If an ephemeral query is in a popular state, the next step is to determine the interval for which it retains its popularity. The interval of popularity can be denoted by $\text{Popular}_Q[i, i+j]$ and includes all those days for which $\text{State}(Q, d) = \text{popular and ephemeral}$ and $d \in [i, i+1, \dots, i+j]$.

### 3.3 Determining the Temporal Context of a Query

Since the postings in microblogs reflect the ephemeral interest of the public, the next step is to extract all the Tweets posted in interval $\text{Popular}_Q[i, +j]$ from Twitter. The temporal context of query term $q_i$ inside microblog *B* is denoted by $C(q_i, B)$ and is determined by a window around each occurrence of $q_i$ inside *B*. Hence, a symmetric window of length *l*, around each occurrence of $q_i$ in *B* gives the temporal context to be use to find relevancy of document *D*

$$C(q_i, B) = [b_{j-l} \dots b_j = q_i, \dots, b_{j+l}]. \tag{3}$$

Thus, each temporal context of the query term has a length of $2l + 1$.

### 3.4 Determining the Document Similarity Score

The extracted context $C(q_i, B)$ can be used to identify more relevant documents from the result set. For this purpose, a semantic scoring function that satisfies the constraints defined on information retrieval model [29] for reasonable ranking is required. The log-logistic model is observed to be the best fit in semantic ranking [28]. Therefore, a scoring function can be derived by replacing the normalized frequency of query terms in a document (in formula of the log-logistic model) by semantic similarity of document terms with the terms in query context obtained from microblogs. Further, to account for terms that are semantically related to a query but occurring far

from query-centric in microblogs, it is proposed to extend the base formula of the log-logistic model to include the semantic difference between current query term $q_i$ and semantic similar terms from microblogs. The proposed scoring function is given in Equations (4) and (5).

$$\text{Score}(D, C(q_i, B) = \sum_{\substack{t \in C(q_i, B) \\ q_j \in B}} \log \left( \frac{\text{Sim}(D, C(q_i, B)) + \lambda_t}{\lambda_t} \right) \times \text{dis}(qi, qj)$$

(4)

$$\text{dis}(q_i, q_j) = 2 - \cos(q_i, q_j)$$

(5)

where $\lambda_t = N_t/N$, $N_t$ is the count of documents containing $q_t$ and $N$ is the total number of documents under consideration.

Further, word embedding is used to compute the value $\text{Sim}(D, C(q_i, B))$. Therefore, the similarity of $D$ with the temporal context of $Q$ is computed as per the indication function given in Equation (6).

$$\text{Sim}(D, C(q_i, B) = \sum_{\substack{t \in C(q_i, B) \\ d_j \in D}} \cos(d_j, t) \times f(d_j, t)$$

(6)

where the document vector and context vector denote embedding in continuous space. The cosine function is used to compute the similarity between these two vectors and its value lies between $[-1,1]$. Further, $f(dj, t)$ is an indication function that returns 1 if the cosine similarity between two terms is greater than the defined threshold, otherwise returns 0. Semantic distance between $q_i$ and $q_j$ is maximum when $\cos(qi, qj) = -1$.

### 3.5 Determining the Document Relevance Score

Having the semantic score of documents with a temporal context of each query term $q_i \in Q$, the next step is to aggregate these scores into a single value denoting document relevance score with respect to the entire $Q$. We performed our experiment using two distinct functions [30], namely the disjunctive relevance (DR) function given by Equation (7) and the aggregate relevance (AR) function given by Equation (8).

$$\text{Rel}(D, Q) = \max_{C_i \in \zeta(Q, B)} \text{Score}(D, C_i)$$

(7)

$$\text{Rel}'(D, Q) = \sum_{C_i \in \zeta(Q, B)} \text{Score}(D, C_i)$$

(8)

where $\zeta(Q, B) = \{C_1(q_1, B), C_2(q_2, B) \ldots C_n(q_n, B)\}$ denotes a set of all temporal contexts of a query in *B*. It is observed that retrieval performance is better with Equation (7). Therefore, the disjunctive relevance function is selected for the next computation.

## 3.6 Final Document Score

After obtaining a document relevance score, the next step is to combine the contextual relevance score with the Google page rank. Thus, three kinds of information are required:

- A. **Daily search volumes:** For identifying popular queries in the corpus, the daily search volumes for each query are recorded using the Google Insight Search Tool [23]. It provides relative search volumes of queries. Based on torrent point and search volumes, the seed set of queries are grouped in three categories listed below:

    - i. **Type A (highly ephemeral):** Queries currently in the popular state and have encountered more than one torrent point in the last one month.
    - ii. **Type B (ephemeral):** Queries currently not popular but that have encountered at least one torrent point in the last one month.
    - iii. **Type C (non-ephemeral):** Queries neither popular nor have encountered any torrent in the last one month.

- B. **Tweet volumes:** The second type of essential information required for setting up the experiment is the content of Tweets containing popular query keywords. For this purpose, the Spinn3r Tool [36] was used. Although it provides a wide range of information such as original source, user identity, creation time, sneak time, actual contents, etc., we required only Tweet creation time and its content. Out of millions of Tweets downloaded during the observation period, those having seed query terms are further processed to obtain named entities using [37]. These named entities serve the purpose of query context.
- C. **Google ranked list:** The third type of information that needs to be collected is the original ranked list related to a query. As it is widely accepted that a user tries to fulfil their information needs by scanning at most 10 web pages in a ranked list [32], so Google Developer API[3] is used to collect those pages. At last, the obtained list per query is rearranged based on query related context.

After obtaining all kinds of information, the final document score of all documents in the resulting list is obtained by evaluating the harmonic mean of the document score given by Equation (9) and the PageRank algorithm [33]. The final score of *D* w.r.t *Q* is thus given by Equation (9).

$$S(Q, D)_{\text{Harmonic}} = \frac{2 \times \text{Rel}(D, Q) \times \text{PR}(D, Q)}{\text{Rel}(D, Q) + \text{PR}(D, Q)} \qquad (9)$$

where $PR(D, Q)$ denotes the PageRank of document *D* with respect to *Q* and $\text{Rel}(D, Q)$ is the proposed page rank. The algorithm for proposed mechanism is given below:

**Proposed algorithm:**

**Step I:** Retrieve the search result set of query '*Q*' by search engine and mark it as '*D*'.

**Step II:** Check popularity of *Q* using Equation (2).

**Step III:** If popular $(Q, i) \neq$ ephemeral, go to step X.

**Step IV:** Determine the temporal context of *Q* using Equation (3).

**Step V:** Determine the document similarity score using Equation (4).

**Step VI:** Determine the document relevance score using Equation (7).

**Step VII:** Determine the final document score using Equation (9).

**Step VIII:** Reorder *D* as per scores obtained in step VII.

**Step IX:** Return updated *D*.

**Step X:** Return *D*.

## 4 Discussion

In order to get good performance from the IR system, Fang et al. proposed important constraints mandatory to satisfy for every retrieval model [29]. The proposed model satisfies these constraints while taking advantage of contextual similarity from microblogs for temporal information retrieval. According to [28], the log-logistic model satisfies these constraints in topic modeling for IR systems. Thus, the proposed model took the log-logistic model as a base and proposed modification to best fit for temporal retrieval while ensuring no violations in constraints laid down in [29].

To compute the contextual similarity between query terms and microblogs, the proposed model finds the exact position of query terms in microblogs followed by extracting co-occurring terms using a query-centric

window. In contrast to the generalized language model described in [25] that takes into account all the terms in documents for computing semantic similarity, the proposed model applies a more effective query-centric context from sources that map to human judgement. Hence, our approach is less expensive and closer to ephemeral retrieval needs.

The proposed approach not only reduces the cost but also leads to considerable reduction in time complexity over topic modeling systems [12]. Further, in comparison to the recently proposed embedding query expansion model (EQE1) that computes semantic similarity scores of query terms with all terms in vocabulary [31], the proposed model is optimal due to utilizing real time context from small size microblogs.

## 5 Performance Evaluation

The experiment is conducted to answer the following research questions:

RQ4: How many relevant results does our model provide compared to existing retrieval models?

RQ5: How many satisfactory results does our model provide to users as compared to existing models?

RQ6: What is the effect of including different combining principles aggregate relevance (AR) and disjunctive relevance in the proposed model?

**Experimental setup:** The experiment is set up to observe ephemeral patterns in search volumes of Google and posting volumes in Twitter over 5000 queries. These queries are randomly selected on the basis of their popularity from 2 million microblogs on Twitter crawled from 5 September 2022 to 5 February 2023. For identifying popular queries in the corpus, the daily search volumes for each query are recorded using the Google Insight Search Tool. The standard INQUERY [35] list is used to remove non-functional terms from the query. Further, the Spinn3r Tool is used to extract the content of Tweets containing popular query keywords. Next, Google Developer API is used to collect the top $N$ pages for each selected query.

**Baseline models:** The proposed model is compared with three categories of retrieval model: the log-logistic model that takes advantage of topic modeling, the sequential dependence model (SDM) that emphasizes query term dependencies in the language modeling framework and the embedding based query expansion model (EQE1).

**Performance metric:** Without loss of generality, four simple but effective performance metrics, MAP, P@10, nDCG@10 and GMAP, are used to evaluate the relevance and interest score of documents. Since the proposed approach is based on the current context of queries. A questionnaire survey is conducted among 30 Ph.D. and M.Tech. students of the information retrieval domain for obtaining reward values for the aforementioned metrics. We first considered the queries from the Type I group and applied the chosen baseline model along with the proposed model on the result list obtained from Google to answer RQ4 and RQ5. The obtained results are summarized in Table 2.

It may be observed from Table 2 that interest and relevance scores are increased significantly for ephemeral queries by applying temporal context from social media like Twitter. The obtained results for Type I category of queries are graphically presented in Figure 3.

Here, it is worth noting that the increase in the interest score is higher as compared to relevance, implying user satisfaction for the proposed results.

**Table 2**    Performance of the proposed model with a baseline model for the Type I category

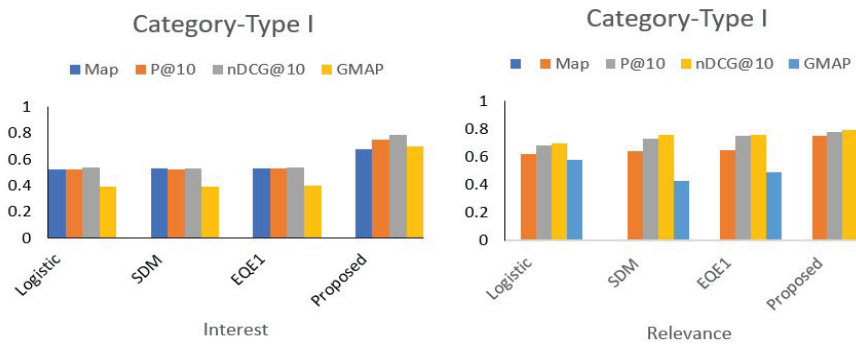| Parameter | Performance Metric | Logistic | SDM | EQE1 | Proposed |
|---|---|---|---|---|---|
| Interest score | MAP | 0.52 | 0.53 | 0.53 | 0.68 |
| | P@10 | 0.524 | 0.52 | 0.534 | 0.75 |
| | nDCG@10 | 0.54 | 0.534 | 0.54 | 0.79 |
| | GMAP | 0.39 | 0.39 | 0.4 | 0.7 |
| Relevance score | MAP | 0.62 | 0.645 | 0.65 | 0.75 |
| | P@10 | 0.68 | 0.73 | 0.75 | 0.78 |
| | nDCG@10 | 0.7 | 0.76 | 0.76 | 0.79 |
| | GMAP | 0.58 | 0.43 | 0.49 | 0.69 |



**Figure 3**    Precision of the proposed model compared to the baseline models.

**Table 3**   Performance of the proposed model compared with the baseline model for the Type II category

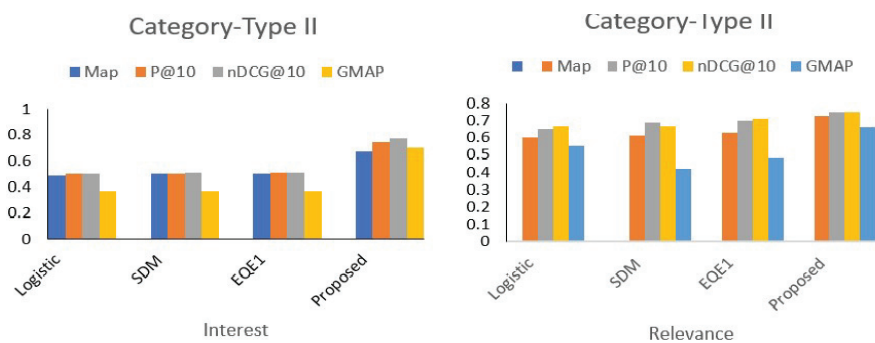| Parameter | Performance Metric | Logistic | SDM | EQE1 | Proposed |
|---|---|---|---|---|---|
| Interest score | MAP | 0.49 | 0.5 | 0.5 | 0.67 |
| | P@10 | 0.5 | 0.5 | 0.51 | 0.74 |
| | nDCG@10 | 0.5 | 0.51 | 0.51 | 0.77 |
| | GMAP | 0.372 | 0.37 | 0.37 | 0.7 |
| Relevance score | MAP | 0.6 | 0.61 | 0.63 | 0.73 |
| | P@10 | 0.65 | 0.69 | 0.7 | 0.75 |
| | nDCG@10 | 0.67 | 0.67 | 0.71 | 0.75 |
| | GMAP | 0.55 | 0.42 | 0.48 | 0.66 |



**Figure 4**   Precision of the proposed model compared to the baseline models for the Type II category.

Table 3 provides the comparison of relevance and interest scores obtained from the proposed and baseline methods for the Type II category.

The graph shown in Figure 4 reveals the important observation that the scores of relevance and interest are lower in the Type II category compared to Type I. This shows the effectiveness of the proposed approach in the integration of the temporal context for ephemeral information in retrieval models.

Table 4 contains the obtained results for various models for the Type III category queries. In Figure 5, the relevance scores of the proposed method indicate that applying the temporal context to non-ephemeral queries had an adverse effect whereas the interest score still signifies the satisfaction of the user.

**Role of different combining principles in the result relevance:** This section aims to find the answer to RQ6. The observed results of combining AR and

**Table 4** Performance of the proposed model compared with the baseline model for the Type III category

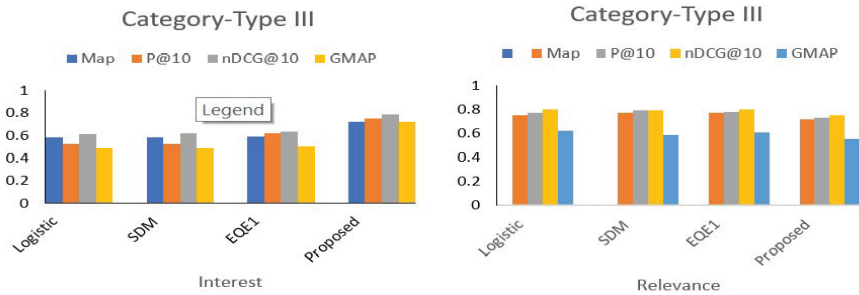| Parameter | Performance Metric | Logistic | SDM | EQE1 | Proposed |
|---|---|---|---|---|---|
| Interest score | MAP | 0.58 | 0.58 | 0.59 | 0.72 |
| | P@10 | 0.522 | 0.52 | 0.62 | 0.75 |
| | nDCG@10 | 0.61 | 0.62 | 0.63 | 0.79 |
| | GMAP | 0.49 | 0.49 | 0.5 | 0.72 |
| Relevance score | MAP | 0.75 | 0.77 | 0.77 | 0.72 |
| | P@10 | 0.77 | 0.79 | 0.78 | 0.73 |
| | nDCG@10 | 0.8 | 0.793 | 0.8 | 0.75 |
| | GMAP | 0.62 | 0.59 | 0.61 | 0.55 |



**Figure 5** Precision of the proposed model compared with the baseline models for the Type III category.

**Table 5** Performance of the proposed model with different principles

| Parameter | Performance Metric | AR-proposed | DR-proposed |
|---|---|---|---|
| Interest score | MAP | 0.682 | 0.70 |
| | P@10 | 0.74 | 0.75 |
| | nDCG@10 | 0.77 | 0.79 |
| | GMAP | 0.69 | 0.72 |
| Relevance score | MAP | 0.68 | 0.72 |
| | P@10 | 0.7 | 0.73 |
| | nDCG@10 | 0.74 | 0.75 |
| | GMAP | 0.654 | 0.66 |

DR within the proposed model is given in Table 5. It may be noted that the DR variant of the proposed model outperformed the AR model in combination with the proposed model.

The results reported in Tables 2–4 also include the DR variant of the proposed model.

# 6 Conclusion

In this work, a novel approach for dealing with ephemeral queries in information retrieval is proposed. It models temporal context by explicitly utilizing microblog content from social media. A query centric window is applied to extract co-occurring terms from microblogs and further used to re-rank the documents in the result set. Extensive experimentation demonstrates the correlation between search and Tweet volumes. Both volumes are studied to identify the temporal proximity of their torrents. It is observed that approximately 70% of search torrents occurred one day before or after blog torrents for lower threshold values. When the threshold is increased, the match ratio of torrent is even increased to ∼90%.

In addition, the experimental results also reveal the effectiveness of the proposed technique in terms of result relevancy and user satisfaction. Further, it may be noted that there is improvement of ∼25% in the result relevance score and ∼35% in the user satisfaction score for ephemeral as well as non-ephemeral queries.

Finally, the work opens up many interesting future directions for researchers. Firstly, the proposed work can be integrated into any baseline information retrieval models to deal with large volumes of temporal context. Secondly, the model can be expanded to incorporate deep neural retrieval models to extensively learn the query context from unbiased social media. Further, the proposed model is not only limited to the information retrieval domain, but in fact its utility can also be explored in recommendation systems, Chatbots, TARs (technical assistance reports) and microblogging.

# Acknowledgment

# References

[1] Xiang, Daxin Jiang, Jian Pei, Xiaohui Sun, Enhong Chen, and Hang Li. 2010. Context-aware ranking in web search. In Proceedings of the 33rd SIGIR. ACM, 451–458.

[2] Wu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. 2018. Query Suggestion with Feedback Memory Network. In Proceedings of the 2018 WWW. ACM, 1563–1571.

[3] N. Golbandi, L. Katzir, Y. Koren, et al., "Expediting search trend detection via prediction of query counts," Proceedings of the sixth ACM international conference on Web search and data mining, vol. 1, pp. 295–304, 2013.

[4] Ryen W White, Wei Chu, Ahmed Hassan, Xiaodong He, Yang Song, and Hongning Wang. 2013. Enhancing personalized search by mining and modeling task behavior. In Proceedings of the 22nd WWW. ACM, 1411–1420.

[5] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. A latent semantic model with convolution-pooling structure for information retrieval. In Proceedings of the 23rd CIKM. ACM, 101–110. https://doi.org/10.1145/2661829.2661935.

[6] Xuehua Shen, Bin Tan, and ChengXiang Zhai. 2005. Context-sensitive information retrieval using implicit feedback. In Proceedings of the 28th SIGIR. ACM.

[7] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q Zhu. 2012. Probase: A probabilistic taxonomy for text understanding. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 481–492.

[8] Rosie Jones and Kristina Lisa Klinkner 2008. "Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs", In Proceedings of the 17th CIKM. ACM, 699–708.

[9] Huanhuan Cao, Daxin Jiang, Jian Pei, Enhong Chen, and Hang Li. 2009. Towards context-aware search by learning a very large variable length hidden markov model from search logs. In Proceedings of the 18th WWW. ACM, 191–200.

[10] Zhen Liao, Yang Song, Li-wei He, and Yalou Huang, 2012. "Evaluating the effectiveness of search task trails", In Proceedings of the 21st WWW. ACM, 489–498.

[11] Phelan, O., McCarthy, K., and Smyth, B. (2009, October). Using twitter to recommend real-time topical news. In Proceedings of the third ACM conference on Recommender systems (pp. 385–388).

[12] CarlosCastillo, Marcelo Mendoza and BarbaraPoblete, 2013. "Predicting Information credibility in Time sensitive social media", Information Research, Emerald Group Publishing Limited, Vol. 23, No. 5, pp. 560–588, ISSN: 1066-2243.

[13] Anuj Jaiswal, Wei Peng and Tong Sun, 2014. "Predicting time sensitive user location from social media", 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 870–877, ISBN: 978-1-4503-2240-9.

[14] Yu, Y., Wan, X., and Zhou, X. (2016, August). User embedding for scholarly microblog recommendation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 449–453).

[15] Yogesh K. Dwived, Jyoti Prakash Singh, Nripendera P. Rana, Abhinav Kumar and Kawaljeet Kaur 2019. "Event classification and location prediction from tweets during disaster" Applications of OR in Disaster Relief Operations, Springer, pp. 737–757.

[16] Alexey Borisov, Martijn Wardenaar, Ilya Markov, and Maarten de Rijke. 2018. A Click Sequence Model for Web Search. In Proceedings of the 41st SIGIR. ACM, 45–54.

[17] Zhengdong Lu and Hang Li. 2013. A deep architecture for matching short texts. In NIPS. 1367–1375.

[18] Jiafeng Guo, Yixing Fan, Qingyao Ai, and W Bruce Croft. 2016. A deep relevance matching model for ad-hoc retrieval. In Proceedings of the 25th CIKM. ACM, 55–64.

[19] Dixit, P., Sethi, S., Sharma, A. K., and Dixit, A. (2012, November). Design of an automatic ontology construction mechanism using semantic analysis of the documents. In *2012 Fourth International Conference on Computational Intelligence and Communication Networks* (pp. 611–616). IEEE.

[20] Gupta, V., Dixit, A., and Sethi, S. 2022. A Comparative Analysis of Sentence Embedding Techniques for Document Ranking. *Journal of Web Engineering*, 2149–2186.

[21] Sethi, S. 2021. An optimized crawling technique for maintaining fresh repositories. *Multimedia Tools and Applications*, *80*(7), 11049–11077.

[22] https://searchengineland.com/welcome-bert-google-artificial-intelligence-for-understanding-search-queries-323976

[23] In Ho Kang and GilChang Kim, 2003. "Query type classification for web document retrieval," Proceedings of International ACM conference on research and development in information retrieval, vol. 1, pp. 64–71.

[24] Wu, H.C., Luk, R.W., Wong, K.F., Kwok, K, (2007) Word embedding based of a hybrid document-context based retrieval model. Inf. Process. Manag. 43(5), 1308–1331.

[25] Li, X., Liu, Y., Mao, J., He, Z., Zhang, M., Ma, S, (2018) Understanding reading attention distribution during relevance judgement. In: CIKM 2018, pp. 733–742.

[26] Li, X., Mao, J., Wang, C., Liu, Y., Zhang, M., Ma, S, (2019) Teach machine how to read: reading behavior inspired relevance estimation. In: SIGIR.

[27] Clinchant, S., Gaussier, E.: Information-based models for ad hoc IR. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 234–241. ACM (2010).

[28] Fang, H., Tao, T., Zhai, C.: Diagnostic evaluation of information retrieval models. ACM Trans. Inf. Syst. 29(2), 7:1–7:42 (2011).

[29] Kong, Y.K., Luk, R., Lam, W., Ho, K.S., Chung, F.L.: Passage-based retrieval based on parameterized fuzzy operators. In: The SIGIR 2004 Workshop on Mathematical/Formal Methods for Information Retrieval (2004).

[30] Zamani, H., Croft, W.B.: Embedding-based query language models. In: ICTIR 2016, pp. 147–156 (2016).

[31] Shilpa Sethi and Ashutosh Dixit 2017. "An automatic user interest mining technique for retrieving quality data" International journal of business analytics. Volume 4, pp. 62–79, ISSN: 2334-4547.

[32] L. Page, S. Brin, R. Motwani et al., "The PageRank citation ranking: bringing order to the Web," Stanford Digital Libraries Technologies Project, vol. 1, pp. 1–17, 1999.

[33] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In Proceedings of the 22nd CIKM. ACM, 2333–2338.

[34] Gan, C., Cao, X., and Zhu, Q. (2023). Microblog sentiment analysis via user representative relationship under multi-interaction hybrid neural networks. *Multimedia Systems*, 1–12.

[35] Broglio, J., Callan, J. P., Croft, W. B., and Nachbar, D. W. (1995). Document retrieval and routing using the INQUERY system. *NIST SPECIAL PUBLICATION SP*, 29–29.

[36] http://spinn3r.com

[37] https://www.analyticsvidhya.com/blog/2021/11/a-brief-guide-on-how-to-build-a-named-entity-extraction-ner-model-with-apache-opennlp-library/

## Biography



**Shilpa Sethi** received her Master of Computer Application from Kurukshetra University, Kurukshetra in 2005 and M. Tech. (CE) from MD University Rohtak in 2009. She completed her Ph.D. in Computer Engineering from YMCA University of Science & Technology, Faridabad in 2018. Currently she is serving as Associate Professor in the Department of Computer Applications, J.C. Bose University of Science & Technology, Faridabad, Haryana. She has published more than 40 research papers in various international journals and conferences. She has published 9 research papers in Scopus indexed journals, 2 in ESCI, 4 in SCI and more than 15 research papers in UGC approved journals. Her areas of research include internet technologies, web mining, information retrieval system, artificial intelligence and computer vision.