Don’t Forget This: Augmenting Results with Event-Aware Search

Hugo Sousa[1, 2] **Austin R. Ward**[3] **Omar Alonso**[3]
[1]University of Porto [2]INESC TEC [3]Amazon

*When do users buy for **Valentine’s day**? What about for **Christmas**, **Halloween**, or even **Star Wars day**?*

This question can be extended to any event of interest to the user. In this project, we analyze user behavior around various events to understand how temporal proximity influences user intent. Our work results in a temporal model of events that can be applied to tasks such as recommender systems and autocomplete. To develop this model, we begin by compiling a comprehensive catalog of events.

Compiling the **Event Catalog**

For the event collection, we use two sources — **Wikidata** and **Time and Date** website — each with a specific processing pipelines. Figure1 illustrates the pipeline.

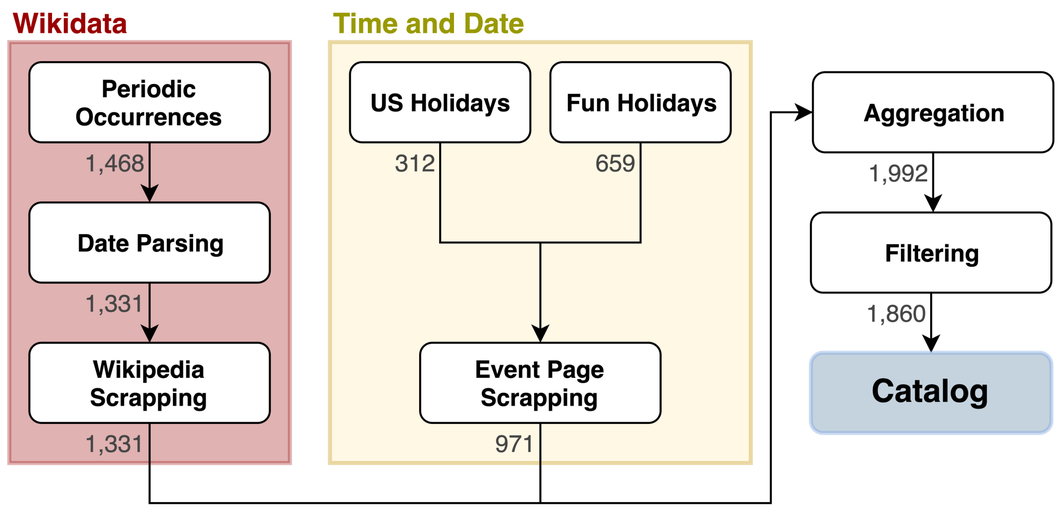

Figure 1. A graphical representation of the pipeline used to collect and process event data.

From Wikidata, we extracted recurring events with defined dates, focusing on those celebrated in the United States, resulting in 1,331 unique events. Additionally, we implemented a parser to handle various date formats and excluded events with invalid or unsupported dates. For the Time and Date website, we scraped both Fun Holidays and US Holidays, carefully extracting dates for events that vary annually, leading to 971 events. After aggregating the data from both sources, we compiled a comprehensive catalog of 1,992 events. In the final step of the pipeline, we applied an adult filter to exclude any events that might be offensive or inappropriate, refining the catalog to 1,860 events.

Building the **Temporal Model**

If we can classify whether search queries or products are related to a specific event, we can estimate the event's importance by counting the frequency of those queries in the days leading up to and following the event. Therefore, the key challenge lies in determining whether a query is related to that event.

• Find the **Relevant Queries**

To identify relevant queries, we begin by collecting all query and product pairs purchased on each day of the year. For each event, we get the data from 90 days prior to the event and 60 days following it. We then perform semantic matching between the event name and the product name using L2 distance, considering all products within a radius of 1.2 as relevant. The queries associated with these relevant products are then classified as relevant queries. As an example, Figure 3 illustrates the count of queries deemed relevant for Valentine's Day.

• Find the **Takeoff and Dropoff Dates**

To analyze when an event gains momentum (takeoff) and when it fades (dropoff), we use two moving averages: a long-term average to identify stable patterns and a fast-moving average to detect short-term changes. The long-term average is set to four times the event duration, estimated by the number of days the signal is above the mean plus one standard deviation. The fast-moving average is set to three days. The takeoff point is defined as the day when the fast average exceeds the slow average for more than half the estimated event duration consecutively, while the dropoff point occurs when the fast average remains below the slow average for the same duration. Figure 2 illustrates this approach for the Valentine’s Day and Christmas events.

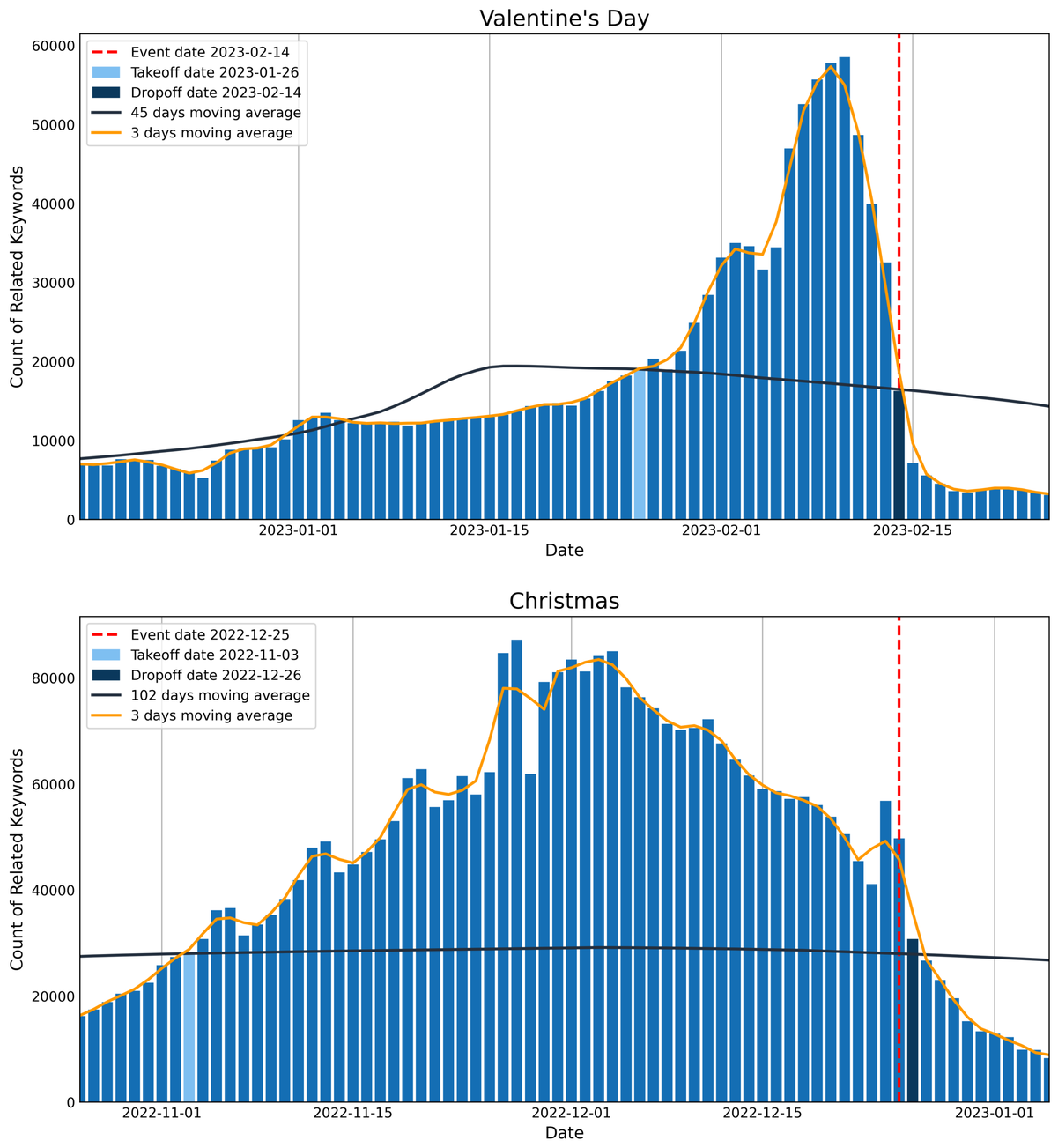

Figure 2. Histogramas showing the daily count of queries deemed relevant for Valentine’s Day and Christmas.

Applying to **Recommender System**

One potential application of the temporal model is in recommender systems, where the significance of an event on a given day can influence the results of a query. For example, when the query ”gift card" is made close to Valentine’s Day, users are likely seeking Valentine's-themed gift cards. Conversely, if the same query is made near Christmas, users may expect results more aligned with Christmas themes. This context-aware approach enhances the relevance of search results by aligning them with temporal event significance. Figure 3 illustrates the idea for the “gift card query”.

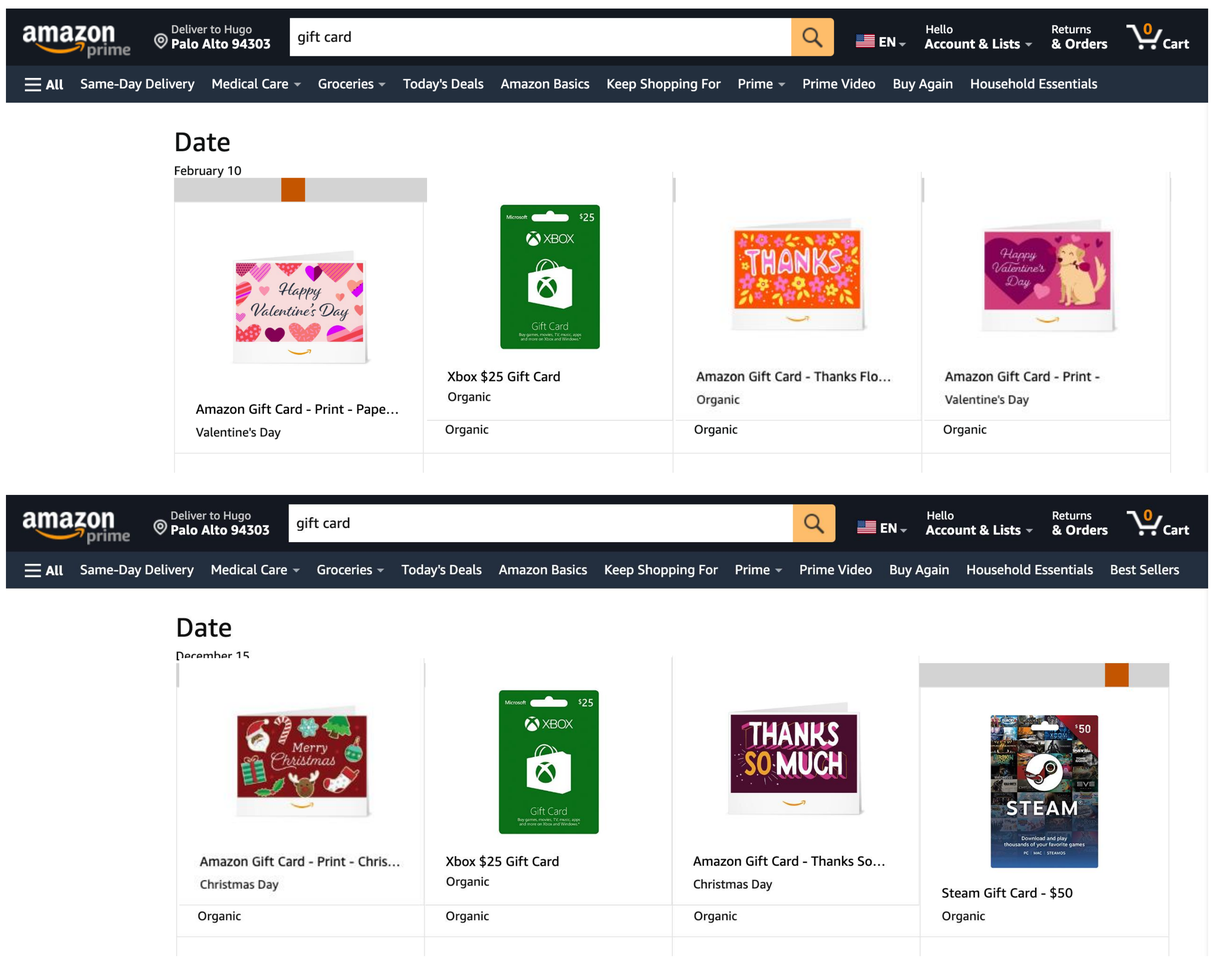

Figure 3. The influence of query date on "gift card" search results, comparing outcomes from February 10 and December 15. Each product is represented by its image, name, and source. "Organic" indicates that the product was retrieved from an organic search, while "Valentine's Day" means that the product was found using a query augmented for Valentine's Day.

Hugo Sousa is supported by national funding from the Portuguese Foundation for Science and Technology (FCT) under the project with DOI 10.54499/LA/P/0063/2020 and his PhD grant with ID 2022.14691.BD.