Wiki Memory Culture

Cintia P. Dewi and Farhad Ahmad
RWTH University Aachen, Germany

Abstract. People tend to remember distinct events that left a remarkable impact on the society. This phenomenon is also known as memory culture. In this digitized era, Wikipedia has been treated as a global memory place where people search for basic information. This paper focuses on investigating online memory culture trends occurring on a yearly basis. Our investigation is done based on Wikipedia page view analysis which involves multiple algorithms to detect such yearly events. The resulting data is visualization of the yearly events in the form of histogram.

Keywords: Memory Culture, Wikipedia, Page View.

1 Introduction

Major events are collectively remembered by the society. These events can be divided into different categories, such as historical, religious, national, science/nature, technology, and birth/death day of significant person. Such collective memories that play big roles in the society have a higher chance of being remembered. People are likely to find information related to particular events, which they remember, during the time when such events happened in the past. This phenomenon is known as a memory culture. However, the events that built such a memory culture are not the same for every society (e.g. people in America remember different events compared to people in Germany).

In this modern era of digitization, people depend a lot on online media as source of information. Wikipedia has been treated as a global memory place where

people search for basic information regarding specific events. The use of such online encyclopedia has encouraged the trend of online memory culture. With the availability of online data, the memory culture behavior could be analyzed to determine the events that each community remember.

The different memory culture behavior in each community motivates us to analyze events that are relevant to that community and decide how well they remember those events. Our approach to determine the events in memory culture is by focusing on the page view analysis of each Wikipedia article. The details of our approach are discussed in the next section.

2 Detection of Yearly Events

To start investigating the memory culture, we chose English Wikipedia articles as the main language. Our investigation is based on the analysis of the yearly events which currently can be seen manually on Wikipedia Pageview Analysis Tool [1] as shown in Fig. 1.

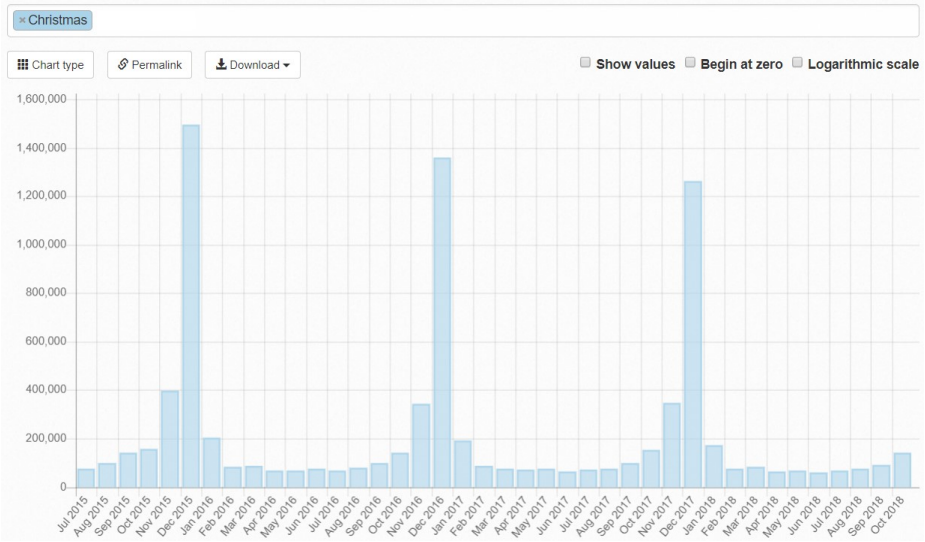

Fig. 1. Wiki Pageview Analysis Tool

Based on the data visualized on Pageview Analysis tool, a yearly event can be detected by the peak views which occur once in every year, mostly in the same month. We implemented several algorithms, in Python programming language, to automate the detection of a yearly event.

First, in order to refine the list of Wikipedia articles, we first extract the relevant articles. This is done by using an API provided by Mediawiki [2] to get the most visited articles of each month during 40 months (July 2015 until October 2018). Total of relevant articles without duplicates are more than 10.000 articles.

The next step is detecting whether the articles obtained are a yearly event or not. This detection is done by adopting another provided Mediawiki API to fetch the number of page views for each article on a monthly basis. Using these monthly pageviews, yearly peak detection can be done by running multiple algorithms namely:

- Standard Deviation based peak detection
- Relative maxima based peak detection
- Nearest Neighbors based peak detection

Multiple algorithms were required to decide the most efficient approach for peak detection. The following sub-sections describe all three algorithms in detail.

2.1 Standard Deviation based peak detection

This algorithm is based on a peak detection approach which utilizes standard deviation as a means of detecting a peak. The average number of page views in a month and the standard deviation for a specific article is calculated first. Based on these values, a factor that determines the distance of the monthly page

views from the average in terms of standard deviation is calculated. For example, a factor of 2 indicates a distance of 2 standard deviations between the average monthly page view and the actual number of page views in a month. This value is used as a determining factor for peak detection in this algorithm.

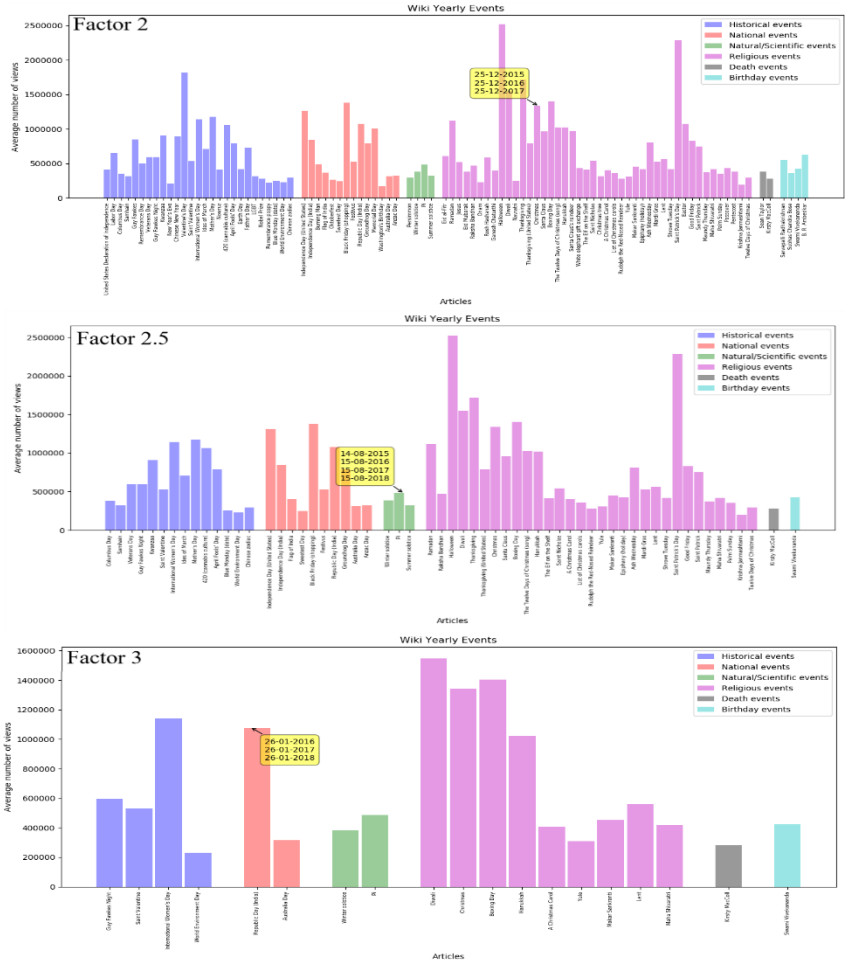

Fig. 2. Result of Standard Deviation based peak detection

To get an accurate factor value, we executed the algorithm using multiple values for it. The factor values that the algorithm was executed are 2, 2.5 and 3 with results as shown in Fig. 2. From analyzing result which shows the detected

events, we decided that a factor value of 2 is the most efficient for peak detection.

After the peak detection is applied on all 40 months for an article, the distance between two corresponding peaks is calculated to check that corresponding peaks have a distance of 11 to 13 months. This ensures that the given article is associated to a yearly event.

2.2 Relative Maxima – based peak detection

The relative maxima peak detection implements the Argrelmax method of the library Scipy [3]. This algorithm uses a set of datapoints and calculates the maxima among them. In order to detect the yearly events, the number of minimum datapoint chosen were equals to 10.

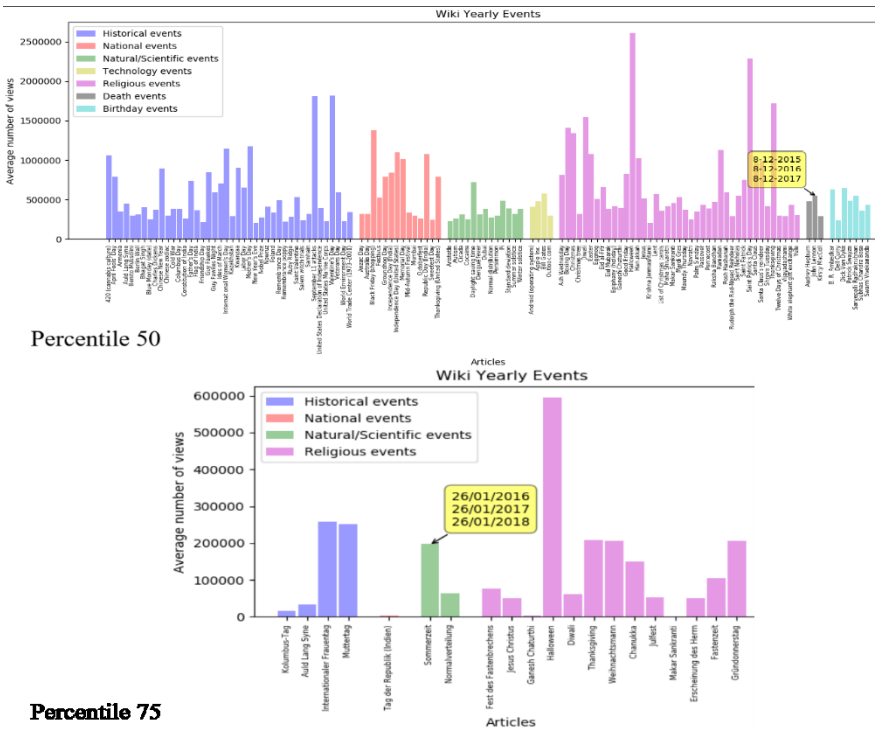

Fig. 3. Results of Relative Maxima-based peak detection

Since the maxima can also include relatively low peaks, therefore a way of ensuring such undesired detected peaks is required. For this purpose, we use an additional percentile-based logic by checking whether the peak detected lies within given percentiles. Different set of percentiles were use to obtain the accurate results of peak detection. The percentile values were 50 and 75 with results as shown in Fig. 3. Comparing both results, percentile 75 produced a very few peaks and highly missed most of the possible yearly events. Hence, percentile 50 was chosen to determine the peak values.

2.3 Nearest Neighbors-based peak detection

This peak detection algorithm works based on the detection on local peaks between nearest neighbors and find the significant peak among local peaks [4]. For this work, the method of getting the local peaks and the significant peaks is done automatically by a module in SciPy library [3], which is called *scipy.signal.find_peaks*.

In order to identify the significant peaks, an additional percentile value is required. To decide the accurate percentile value, we compared the result of peak detection using two different percentile values of 50 and 75. Result of the peak detection is as shown in Fig. 4. Based on the results given from different percentile values, we concluded that percentile value of 50 is the most suited value to detect the significant peaks.

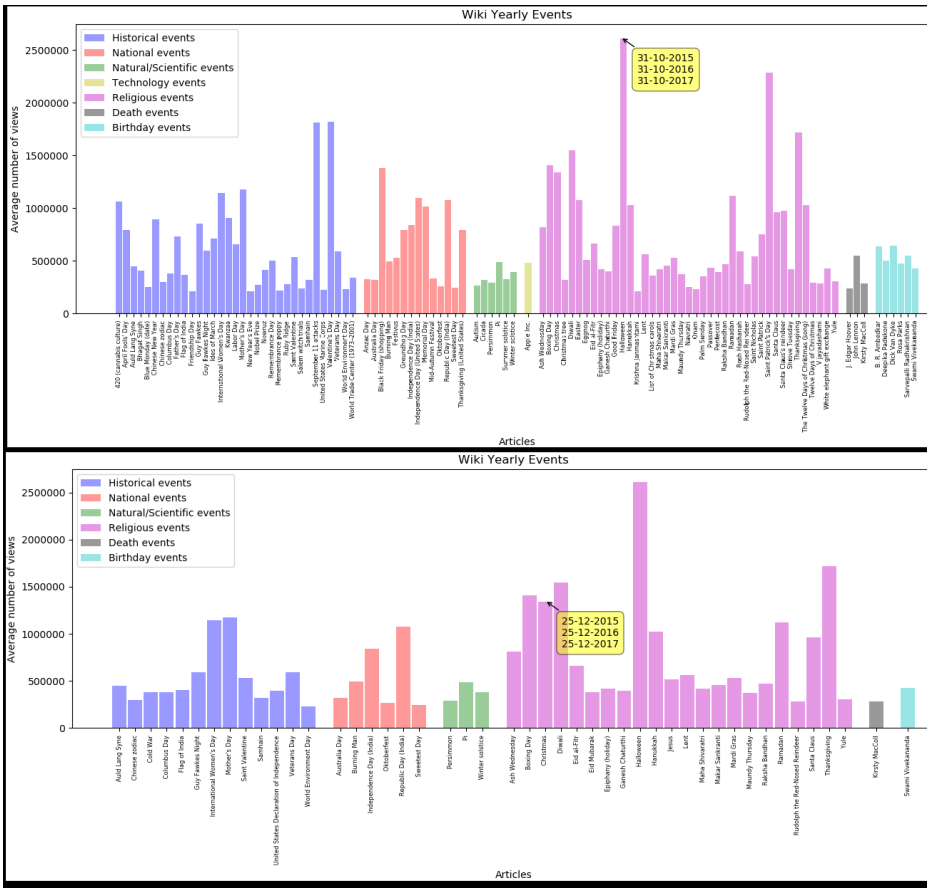

Fig. 4. Results of Nearest Neighbors based peak detection

3 Categorization and Visualization of Yearly Events

Continuing the memory culture investigation, the yearly events detected by the algorithms are later categorized into the following categories:

- Historical Events: Events that have a major significance in history.
- National Events: Events that concern a specific country.
- Religious Events: Events that have religious significance.
- Natural/Scientific Events: Events that are related to natural phenomenon or scientific inventions.

- Death Events: Death day of a significant person.
- Birth Events: Birthday of a significant person.

The categorized events are then visualized in the form of a histogram. Events are differentiated in the histogram based on their category. Visualization is done for the results of each algorithm as shown in the previous section. The histogram is created with the events on the horizontal axis and the average number of peak views on the vertical axis. This gives an idea regarding the nature of the event as well as how well it is remembered during its occurrence.

4 Comparison of Different Cultures

After completing our analysis on English Wikipedia articles, our next aim was to compare different cultures based on the Wikipedia articles in their local languages. The languages that we chose were German and Bahasa Indonesia. Two different approaches were used for this purpose.

The first approach used the same list of articles as the one used for English Wikipedia. In this approach, the list of articles was first translated into the respective languages. This was done using another Mediawiki API [2] that returns the name of the article in all available languages. All three algorithms were then used on the translated articles. The results of detected yearly events all three languages are compared based on the event categories and the number of articles in each category. This gives an overview of how people in different cultures remember different types of events. The comparison result is as shown in Fig. 5.

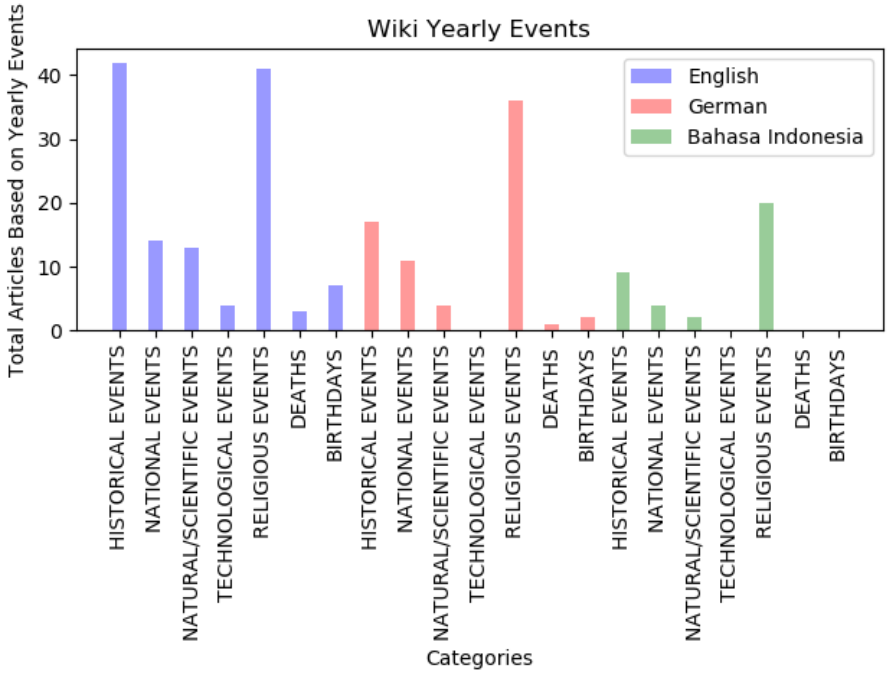

Fig. 5. Comparisons of Different Languages from same list of articles

The second approach used different list of articles of Bahasa Indonesia, which is obtained by utilizing the same Mediawiki API to get the top 1000 articles of each month from Bahasa Indonesia Wikipedia. Same algorithms were used to detect the yearly events. The results of the yearly events were categorized and compared to the English articles. Comparison result, which is depicted in Fig. 6, also shows such different culture of remembering different types of events between two languages.

While first approach, which used the same list of articles as the English ones, produced a low count of events in almost all categories, the second approach produced a much more significant result. The second approach showed the actual trend of the respective society based on the language used.

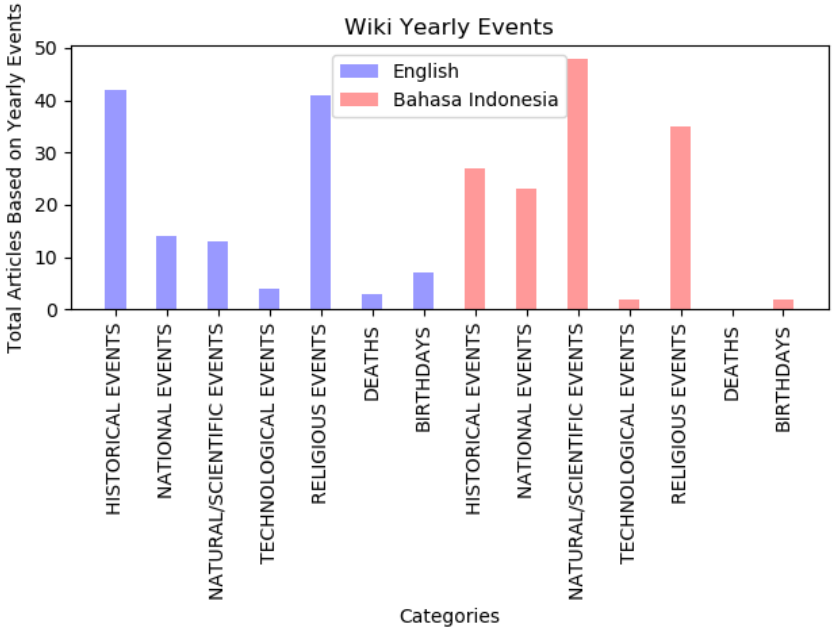

Fig. 6. Comparison result using new list of Bahasa Indonesia articles

5 Conclusion

Based on our work in investigating Wikipedia memory culture, it can be concluded that events which are yearly celebrated or left remarkable impact in history are remembered the most by people. The trend that we analyzed proved that people tend to search for information related to the event during occurrence of the events every year.

Comparing the different languages of Wikipedia articles, it is easy to analyze that different societies remember different type of events relevant to them. However, some events are remembered almost the same way in different cultures and society (e.g. religious events).

References

1. https://tools.wmflabs.org/pageviews/
2. https://www.mediawiki.org/wiki/API:Main_page
3. https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.signal.argrelmax.html
4. G. Palshikar, et al., Simple algorithms for peak detection in time-series, in: Proc. 1st Int. Conf. Advanced Data Analysis,Business Analytics and Intelligence, 2009, pp. 1–13
5. Pentzold, C. (2009). Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. Memory Studies, 2(2), 255–272.