

# HOLIDAYS IN CULTURAL MEMORY: SOCIAL MEDIA MINING ACROSS GENERATIONS

**Tatiana Folomeeva, & Fedor Vinokurov**

*Faculty of Psychology, Lomonosov Moscow State University (Russia)*

## Abstract

We present the first results of an ongoing study of the construction and transmission of cultural memories about national holidays in Russia. Due to a widespread adoption of social media among multiple generations as a first- or second-choice way of communication and social interaction, we were able to retrieve a relatively large volume of public data for a representative sample of urban population of Russia older than 14 years. A sample of 2200 public profiles and 13800 of their publications was downloaded from Vk.com – the most popular social media website in Russia (43% of urban population older than 12 years use it daily according to Kantar TNS MediaScope). Research sample was constructed using a public Vk.com API and Web Census method. Our analysis involved only posts (not reposts) with textual content that were published in the interval of two days before and two days after the date of a national holiday in 2010-2018: January 1<sup>st</sup> (The New Year), February 23<sup>rd</sup> (Army Day or Men’s Day), March 8<sup>th</sup> (Women’s Day) and May 9<sup>th</sup> (Victory Day). Content analysis involved standard text-mining procedures of data cleansing, stemming and spam filtering (most frequent terms were reviewed by experts to create a list of stop-words). An explorative topic modelling with LDA algorithm for four separate text corpuses has shown that the publications were mostly connected with a consumer behavior on holidays, but one topic was found to be specific for authors older than 39 years – greeting cards. Greeting cards were the most accessible, affordable and familiar element of a festive ritual in Russia during 20<sup>th</sup> century. A demographic group also known as “Generation X”, those who have matured before public Internet and social media, are now transmitting their cultural memories digitally but with the most familiar and natural form of a pre-Internet illustrated message – a greeting card. Our research illustrates the advantages and limitations of automated analysis of big textual data in social psychology. We propose a novel procedure for automated longitudinal studies of cultural memory and social representations.

**Keywords:** *Cultural memory, holidays, social media, data mining.*

---

## 1. Introduction

Cultural memory research is a modern trend in social science, but the definition of “cultural memory” is still a point of discussion. The term is used as often as “collective memory” but the underlying theory is still under development (Wertsch, Roediger, 2008). Even those who are concerned with scientific usefulness of this term do not doubt the social determinants of memory and its dynamics. Employing Vygotsky's sociocultural theory we can describe human memories as result of communication, social rituals, signs and events involving others. The representations of the past are volatile and transform constantly while embedded in social change. Our research suggests that while social media stays the first- or second-choice way of communication and social interaction for the majority of country’s population, researchers could monitor public posts and extract the contents of cultural memory, which becomes a great source of insights about social structure and dynamics.

We discuss the construction and transmission of cultural memories about national holidays in Russia. Holidays play a major role in different cultures and are a well-researched phenomenon in anthropology, psychology, social and cultural studies. A. Comte describe holiday as a way of social control, E. Durkheim and M. Mead write about its symbolic meaning, J. Taylor and S. Turner discuss holiday myths and rituals, P. Berger and T. Luckmann view holiday as an opposite of everyday behavior. Based on the previous research of holiday rituals (Karpova, 2008), we propose that festive communication of multiple generations on social media would reflect differences in their cultural memory.

## 2. Objectives and methods

The goal of our research was to examine differences in social media communication on four national holidays among various demographics.

To gather a representative sample of festive communication about national holidays in Russia we applied social media mining techniques specific for the most popular social media platform in Russia – VK.com (also known as “VKontakte” or “In Contact”). According to TNS Gallup Mediascope – a leading media-measuring agency – there were 73% of monthly average internet users in Russia at the beginning of 2018 (Mediascope, 2018). A bit less than a half (43%) of population older than 12 years used VK.com website or mobile application daily. VK.com is more popular than Facebook, Instagram or any other social media among Russian population. Since it provides all the same features and content types as Facebook with a similar user interface, there are mostly technical differences between VK.com and Facebook as sources of social data.

At first, we have gathered a representative sample of Russian urban population consisting of 2200 public profiles (last active in March of 2018) using Web Census method (Chekmyshiev, Yashunskiy, 2014) which was proposed specifically for accomplishing this type of tasks with VK.com public API.

The second step of social media mining involved gathering public communication on national holidays: January 1<sup>st</sup> (The New Year), February 23<sup>rd</sup> (Army Day or Men’s Day), March 8<sup>th</sup> (Women’s Day) and May 9<sup>th</sup> (Victory Day). Publications by these 2 200 authors from January 2010 till March 2018 were downloaded for dates in the interval between two days before and two days after the date of a national holiday. A single publication in VK.com includes some textual data and optional media attachments (mostly images or videos). Each record in the resulting dataset of 13 800 publications was augmented with the information from author’s public profile: age, gender, city of residence, etc.

Profiles and publications data was sampled and retrieved with a public VK.com API. Text data preparation involved standard text-mining procedures of data cleansing, stemming, filtering of common words and spam messages. Experts (students in social psychology) reviewed most frequent terms to create a list of stop-words for a spam filter. Spam publications had low term variability, so we applied simple filtering rules based on keyword occurrence.

At the end of a data collection step, we had four separate text corpuses for each holiday prepared for explorative topic modelling with latent Dirichlet allocation (LDA) algorithm. LDA is a machine-learning model well suited for topic discovery tasks with “bag of words” approach (Nikolenko et al., 2015). LDA generates and continuously refines multiple hypothesis about implicit topics in a corpus of documents. A “topic” is a sample of words that frequently co-occur in different documents. In text classification tasks LDA calculates a probability that a document belongs to each of the previously learned topics. We performed both topic discovery and document classification for the current research.

The technological stack consisted of a Python notebook running on a Google Colab service, a PyMorph2 package for word lemmatization, pandas and scikit-learn packages for data analysis and modelling.

## 3. Results and discussion

For each of four holidays we have constructed a separate model of nine topics. The final number of modelled topics is a result of manual calibration for an optimal balance of topic interpretability and quantity.

Five to six topics in each model were linked to consumer behavior on holidays: buying flowers, pastries, various beauty services. One topic in all models was found to be specific for authors older than 39 years – greeting cards. This topic has slight variation in core terms depending on the holiday. These are top keywords sorted by weight (or importance): “The New Year” – greeting card, new year, picture, beautiful, greeting, here, watch, winter, hello, horse, happiness, rooster, goat, luck, dog, sheep, wish, success; “Victory Day” – victory, may, picture, postcard, day, beautiful, friend, great, sky, peaceful, head, answer, holiday; “Women’s Day” – postcard, march, picture, beautiful, women day, huge, holiday, sunshine, joy, colors, blossom; “Army Day or Men’s Day” – picture, postcard, February, beautiful, defender, fatherland, holiday, male.

Each of the discovered “greeting card” topics contains the word “picture” or “postcard” or both. The majority of posts classified with this topic also contain media attachments. A quick manual sampling of these images showed that these are indeed digital postcards – thematic images with holiday greetings. This topic is specific for publications authored by mature and senior demographic groups (39 years or older). Publications like these are almost non-existent among younger authors.

These results could illustrate a changing role of a greeting card as an element of festive rituals in Russian culture. From the beginning of mass production of paper postcards in the post-war USSR in 1945 till nowadays greeting cards are still the most affordable holiday gifts available to just about anyone. The evolution of telecommunication and digital services have transformed the look and feel of the modern greeting card but its importance for older generations preserved.

Our research shows that social media mining and topic modelling are suitable and versatile techniques to study large groups in social psychology. Explorative analysis could be performed on historical data (87 months in our case) or almost in real-time with online monitoring and online algorithms for machine learning. Unsupervised topic modelling with a raw social media data allowed us to grasp the festive traditions among different generations in Russia. These are only first steps in making big data analysis and text mining algorithms widely available for psychologists.

### References

- Karpova, G. (2008). *Prazdnik v kontekste sotsialnykh izmeneniy: tradicii i vlast [Holiday in context of social change: traditions and authority]*. Saratov: Nauchnaya kniga.
- Chekmyishev, O.A., Yashunskiy, A.D. (2014) Izvlechenie i ispolzovanie dannykh iz elektronnykh sotsialnykh setey [Extracting and using data from online social networks]. *Working Paper IPM im. M.V. Keldysha [Keldysh Institute of Applied Mathematics preprints], no 62*. Retrieved March 18, 2019, from: <http://library.keldysh.ru/preprint.asp?id=2014-62>
- Nikolenko, S., Koltcov, S., Koltsova, O. (2015) Topic modelling for qualitative studies. *Journal of Information Science 2015. Vol 43, Issue 1*. Pages 88-102.
- Wertsch, J.V., Roediger, H.L. (2008) Collective Memory: Conceptual foundation and theoretical approach. *Memory, 16 (3)*. Pages 318–326.
- Mediascope (July 2018) Mediascope Web Index. Retrieved April 3, 2018, from URL: <http://mediascope.net/services/media/media-audience/internet/information>.