

# The study of the relationship between publications in social networks communities via formal concept analysis

Kristina Pakhomova and Alina Belova

Siberian Federal University, Kraskoyarsk, Russia  
kpahomova@yandex.ru, alina.belova94@mail.ru

**Abstract.** Nowadays the users generate considerable amount of information in the Internet. Therefore, they have a deal with the issue of how to retrieve the required information perhaps much more relevant than he/she supposed. Moreover, the mix of different data sources usually muddles users in the case of searching for the instructive information. The authors of this paper will introduce the approach based on text processing and formal concept analysis in order to structure the information from a variety of sources, particularly, social media communities. Additionally, they will clarify the relations between community posts with the same topic, where these relations become a recommendation tool for the user's decision making. In conclusion, the authors will build a diagram that will be a convenient visualization tool in an effort to structure information that was obtained from various sources.

**Keywords:** Formal Concept Analysis · Semantic Analysis · Social Network · Data Mining · Community topic.

## 1 Introduction

Due to a rapidly increasing amount of data which is generated by users of the Internet, how to deal with this data becomes the most important issue. For instance, in 2019, 4.39 billion people were registered in services provided by social networks, which is 366 million (9%) more than in January 2018 (information provided by 'We Are Social agency' and 'Hootsuite' service [9]). Social networking services contain information that includes a variety of data, for instance, text and media. Obviously, text information is publications in social networking communities, each of those are described by a wide range of topics. In this case, users need to quickly filter and analyze a large amount of information. Currently, the filtering may be carried out in two ways: automatically, by using data mining methods, and manually by the user. However, the relevant result of the user request is an extremely important issue, but also the speed of request and visualization of the result has value. According to the wide range of data mining methods, the authors will explain a mathematical approach - formal concept analysis (FCA) that satisfies the above criteria.

Copyright ©2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Current research that based on the analysis of social network data is presented in the following works [5,6,7,8]. It is worth noting that the most significant work is the one [8], in which scientists deal with social media data via FCA.

This paper includes four-section: an introduction, approach explanation, experiment computation, and conclusion. Firstly, the authors will explain briefly the theory of FCA and basics semantic analysis methods. Secondly, the authors propose a solution based on methods of semantic data analysis and FCA. Finally, they will explain an experiment implementation based on the social network dataset and also they will propose to visualize the results of computation by using concepts lattice as a diagram, in which each circle will be marked in a certain color.

## 2 Computation approach

### 2.1 Brief review about Semantic Analysis

In order to study the text data ,in particular, the meaning of the text obviously the authors deal with semantic analysis [11,12,13]. According to its theory in general, the text should be subjected to the basic text manipulation methods such as tokenize, lemmatize, and etc. We apply semantic analysis in order to compute the set of keywords that will be related to the posts with the common topic. Where set of posts  $G = \{g_1, g_2, \dots, g_n\}$ ,  $N = 0, \dots, n$ , then the set of words and symbols is  $W = \{w_1, w_2, \dots, w_p\}$ ,  $P = 0, \dots, p$  when the word includes in text  $W \subseteq G$  for  $g \in G$  and  $w \in W$ . We deal with tokenize, lemmatizer, removing stop words and parts of speech other than nouns in order to compute keywords. So the final set of keywords is  $M$  which satisfy  $M \subseteq W$ , where  $M = \{m_1, m_2, \dots, m_l\}$ ,  $L = 0, \dots, l$ .

### 2.2 Brief review of Formal Concept Analysis

FCA was invented by Germany mathematic Rudolf Wille in 1981 [2]. According to Wille paper, after 1996 Wille and Ganter clarify the theory of formal concept [3].

According to Wille and Ganter paper that introduce to formal context is  $(G, M, I)$ , where  $G$  - the set of objects,  $M$  - set of attributes, binary relationship between objects and attributes explains  $I \subseteq G \times M$  for  $g \in G$ ,  $m \in M$  when  $gIm$ , if object  $g$  has an attribute  $m$ .

Formal concept is pair  $(A, B) : A \subseteq G, B \subseteq M$  satisfy the Galois connection between  $(2^G, \subseteq)$  and  $(2^M, \subseteq)$ ,  $A' = B$  and  $B' = A$ . Where  $A$  and  $B$  - extent and intent of formal concept  $(A, B)$ . The ordered set of all formal concept forms is called the concept lattice  $\underline{\mathbf{B}}(G, M, I)$  [1].

### 2.3 Concept lattice building

In a previous subsection, the authors have defined the main mathematical method founded on formal concept computation. Therefore, the ordered concepts build the concept lattice, which is visualised by diagram, Hasse. Every set of formal concepts has a great common subconcept as supremum. Its extent consists of those objects that are common to all extents of the set. Every set of formal concepts has a least common superconcept, the intent of which comprises all attributes which all objects of that set of concepts have. Additionally, ordered this way the set should satisfy the axioms defining a lattice, there are commutative, associative, absorption laws. Thus, a complete lattice is an ordered poset in which all subsets have both a supremum and an infimum [4]. This diagram consists of main elements such as circles are set of formal concepts, lines explain the relation between formal concepts and labels. Notably, an attribute can be reached from an object via an ascending path according to subconcept-superconcept hierarchy. Additionally, it satisfies if and only if the object has the attribute.

Each post includes specific metadata that describe users' personal interest in the particular post, they are likes and reposts. Concerning of those measures we want to compute for every concept the average values of likes and reposts, and after we are going to clustering the ordered set of concepts by using k-means (KMeans) [14]. This also will assist to fix the color of each circle of the diagram in order to visualize the clusters of formal concepts.

## 3 Computation experiment

The authors have investigated the social network dataset concerning the common topic that takes place in the social network communities. Thus, this dataset includes information about Id post, the text of the post, value of users' attitudes, and value of reposts (see Tab. 1). It was obtained from 'Vk' social network, It community, which text of the posts in Russian and English [10].

**Table 1.** Experimental dataset

Id Post	Text	Value of likes	Value of reposts
1297322	<i>g1</i>	6	0
1297419	<i>g2</i>	41	5

In order to compute keywords, the authors used semantic analysis which was explained before. We computed the formal context according to set of key-words will be presented as attributes of a formal context and the set of objects - a number of posts (see Tab. 2).

**Table 2.** Example of formal context

Id post	microsoft	httpamp	remote	...	javascript
1297322	1	1	0	...	0
...	...	...	...	...	...
1297419	0	1	0	...	1

According to FCA theory the set of formal concepts was computed. Although the set of posts includes no more than 100 items, the number of obtained formal concepts is quite huge, approximately 800 items. Moreover, for each formal concept the average of users' likes and average of reposts were computed (see Tab. 3).

**Table 3.** Example of formal concepts

Extent	Intent	color like	color repost
1297430, 1297330, 1297313,			
1297235, 1296857	javascript, httpamp	blue	green
1296857, 1296779	javascript, help	blue	green
...	...	...	...
1297419, 1297266	javascript, developer, company	green	red
1297419, 1296779	javascript, developer, middle	blue	green
1297266, 1296779	javascript, developer, senior	green	red

However, the great number of ordered formal concepts build a huge diagram, for this reason, the authors have explained only the set of concepts and their clustering. According to Tab. 3 in which an object '*javascript*' takes place by following the next set of formal concepts. We deal with diagram in order to concentrate on users preferences, for instance, the user who takes an interest in hiking a job or career development and he \ she takes an interest on '*javascript*' may be recommended the next set of posts: 1297419, 1297266, 1296779 and etc. This argument makes sense due to lattice lines properties that explain the relation between formal concepts.

By using the measures (likes and reposts) we computed three concept clusters, where centroids according to likes are (*color cluster : value*) : (*red* : 108.333, *blue* : 55.3985, *green* : 24.227), and reposts are (*green* : 9.1345, *red* : 3.254). Additionally, each circle of the diagram has its own color that explains a cluster number. This opportunity allows users to visualize their searching and to rank posts according to measures. For instance, the user concentrates on '*javascript*' therefore high priority has 1296779 and after 1296779, 1297419 as stated by post likes. However, measure repost supports a few users so its values explain only two clusters but another hand this measure makes more sense than likes measure in the opinion of an issue of ranking.

## 4 Conclusions and future work

This paper has discussed the approach which manipulates the social network dataset by using FCA, particularly, it assists the user of social media to deal with communities' posts. The authors take into account the specific framework of a dataset that is satisfied with a variety of social services. Moreover, the authors concentrated on such criteria as accessible, high quality, immediate, and relevant information that can be provided to the user by his/her request. Although this approach partially satisfies this number of criteria, so it will be tried to improve in the future by using FCA advantages.

## References

1. Ferré,S.,Huchard, M.,Kaytoue,M., Kuznetsov, S. O.,Napoli A.: Formal Concept Analysis: From Knowledge Discovery to Knowledge Processing. A Guided Tour of Artificial Intelligence Research, pp. 411–445, Springer Nature Switzerland (2020).
2. Wille, R.: Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. Ordered Sets, pp. 445–470, Springer Netherlands (1982).
3. Ganter, B., Wille, R.: Formal Concept Analysis. Berlin.Heidelberg: Springer-Verlag (1999).
4. Davey, B.A., Priestley, H.A.: Introduction to Lattices and Order. Cambridge: Cambridge University Press (2002).
5. Ignatov, D.I., Kuznetsov, S.O.: Concept-based Recommendations for Internet Advertisement. Palacky University, vol. 433, pp. 157–166, Olomouc (2008).
6. Medina, J., Pakhomova, K., Ramirez-Poussa, E.: Recommendation Solution for a Locate-Based Social Network via Formal Concept Analysis. Trends in Mathematics and Computational Intelligence. Studies in Computational Intelligence, vol. 796, pp. 131–138, Springer, Cham (2019).
7. Cordero, P. et al.: Knowledge discovery in social networks by using a logic-based treatment of implications. Knowledge-Based Syst, Elsevier B.V., vol. 87, pp. 16–25, (2015).
8. Missaoui, R., Kuznetsov, S., Obiedkov, S.: Formal Concept Analysis of Social Networks. Springer International Publishing (2017).
9. LNCS Homepage, <https://www.web-canape.ru/business/vsya-statistika-internetana-2019-god-v-mire-i-v-rossii/>. Last accessed 19 Jun 2020
10. LNCS Homepage, <https://vk.com/habr>. Last accessed 19 Jun 2020
11. Bird,S.,Klein, E.,Loper, E.: Natural Language Processing with Python. O'Reilly Media, Inc.,Gravenstein Highway North, Sebastopol (2009).
12. Manning,C.,Schütze,H.: Foundations of Statistical Natural Language Processing. MIT Press. Cambridge (1999).
13. Russell,S.,Norvig,P.: Artificial Intelligence: A Modern Approach, 3rd Edition. Prentice Hall (2010).
14. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In Proc. 5th Berkeley Symp. on Math. Statistics and Probability, pp 281–297,(1967).

