A Triadic Formal Concept Analysis Approach to Analyzing Online Hate Speech in Facebook Comments

Radu Mihai Meza
Babeș-Bolyai University, Cluj-Napoca, Romania
Strada Universității 7-9, Cluj-Napoca 400084
Phone: 0264 405 300
meza@fspac.ro

Șerban Nicolae Meza
Technical University Cluj-Napoca, Cluj-Napoca, Romania
Strada Memorandumului 28, Cluj-Napoca 400114
Phone: 0264 401 200
Serban.Meza@com.utcluj.ro

Abstract

This paper outlines computational thinking, language independent methodology for identifying and analyzing the contexts, targets and contents of online hate speech manifested in Facebook comments on popular Fan Pages or open groups. The three-step process involves data collection via API tools, a preliminary co-occurrence analysis of user-defined semantic field codes and clustering and visual analysis using triadic formal concept analysis navigation tools.

Keywords: Hate Speech; Co-Occurrence Analysis; Triadic Formal Concept Analysis; Unstructured Data; Digital Media Analysis.

1. Hate Speech in the Digital Media

The issue of online hate speech is the subject of heated policy debate in Europe and globally over the past few years. Although hate speech acts have been regulated by European laws, online communication through platforms owned by businesses located outside the users' country and subject to different legislation raises new issues. As Facebook is presently used by approximately 2.2 billion people globally, both governments and NGOs look towards the social media giant to ensure sound and effective mechanisms of dealing with hate speech acts per each country's policy. At the forefront of the group of European countries pressuring Facebook to come up with improved methods is Germany – which has some of the strictest regulatory frame-works concerning hate speech. In September 2015, after increasing pressure (Donahue, 2015) on the backdrop of the refugee crisis, Facebook announces an initiative to increase its efforts to tackle racist content on its German website. Furthermore, in early 2016, Facebook outsources the monitoring and control of racist posts over increased public criticism of the company's reluctance to deal with hate speech in accordance with the European regulatory frameworks ("Facebook outsources fight against racist posts in Germany," 2016). However, Facebook activity by users in both Europe and the United States at the end of 2016 has driven lawmakers to further increase pressure on Facebook to "clamp down on hate speech, fake news and other misinformation shared online, or face new laws, fines or other legal actions"(Scott & Eddy, 2016). There is clearly a job market opportunity for computer scientists and digital media specialists who are equipped with the conceptual and technical skills to tackle this complex problem. They will need to use computational approaches to deal with hate speech in social media streams, to sift through large amounts of data to identify, analyze, and classify potentially dangerous speech, hate speech and offensive speech and prepare intervention strategies.

A Brief Conceptualization of Hate Speech

The key issues in countering online hate speech are outlined in the 2015 UNESCO study ”Countering Online Hate Speech” (Gagliardone, Gal, Alves, & Martinez, 2015): definition, jurisdiction, comprehension and intervention.

There are different definitions of ”hate speech”, which frequently mix concrete threats to the security of individuals and groups with expressions of frustration and anger. Also, online media communication platforms such as Facebook, Twitter or Google define their own policies towards admissible content published by users. However, as recent tensions have shown, these often clash with national legislation and consensus seems unlikely.

Online networked communication platforms have given private spaces of expression a public function and the combined speed and reach of Internet communication raise new issues for governments trying to enforce national legislation in the virtual public sphere, often in contexts managed by companies located in other states.

There seems to be a lack of comprehension of the relation between online hate speech phenomena and offline speech and action or more precisely, violent action. In (Gagliardone et al., 2015), the authors highlight the lack of studies examining the links between hate speech online and other social phenomena.

Different contexts for online communication have led to different intervention strategies – from user flagging, reporting or ranking to monitoring, editorializing and counter-speaking. However, popular online social network type platforms seem reluctant to publish aggregate results that would allow an overview of the phenomenon.

Media platforms define their own policies towards admissible content published by users, governments find it hard to enforce national legislation in the virtual public sphere, there is little comprehension of the relation between online hate speech phenomena and violent action and different contexts for online communication have given birth to different intervention strategies – from user flagging, reporting or ranking to monitoring, editorializing and counter-speaking. However, all these four key issues are strongly related to the identification and analysis of hate speech from semi-structured or unstructured data such as Facebook comments.

Reports and academic works emanating from NGO initiatives are starting to shape an emerging scholarship on the issue. When analyzing hate speech as an act of communication, overviews of the issue (Angi & Bădescu, 2014) recommend focusing on:

- Content (what is being said);
- Emitters (who is communicating);
- Targets (who is the message about);
- Context (including when the action takes place).

There clearly lacks a large scale online data-driven study of contexts, emitters, contents and targets for hate speech in social media with generalizable results and evidence that could drive policy in the matter.

2. Computational Approaches and Co-occurrence Analysis

It is only very recently that digital social science academic research into the niche topic of online hate speech has emerged, using computational approaches towards collection and analysis of large datasets of comments on news sites, blogs and especially social media (Meza, 2016).

From a methodological standpoint, detecting violent, obscene or hate speech is a problem for both media researchers and content managers or digital platform owners. Natural language processing is a complex task and there is a scarcity of tools available for most languages.

Computational thinking was popularized a decade ago as ”a fundamental skill used by everyone in the world by the middle of the 21st century”(Wing, 2006). The concept developed and is still developing as it is adopted in education, but problem-solving via computational thinking may be defined by abstraction, automation and analysis (Lee et al., 2011).

Co-occurrence analysis is widespread in communication and information sciences, especially in library science, but also in machine translation or natural language processing. Recent efforts in computational linguistics applied to hate-speech use machine learning techniques similar to sentiment analysis in correlation with techniques for detecting terms used to reference racial, ethnic or religious groups (Gitari, Zuping, Damien, & Long, 2015).

Digital media analysis may make use of API interaction tools for data collection from social media, computational linguistics tools that allow the exploration of word or concept co-occurrence networks or user-friendly drag-and-drop visual environments for analysis of large data sets such as Tableau as research shows students in the Web 2.0 age prefer efficient, easy-to-use, accessible applications.

3. Dyadic and Triadic Formal Concept Analysis (FCA) Preliminaries and Tools

Formal Concept Analysis (FCA) is a method of knowledge representation introduced in the 1980s by Rudolf Wille, rooted in the pragmatic philosophy of Charles Sanders Peirce, based on a binary incidence relation, and building on applied lattice and order theory. It has applications in various fields and its advantage lies in the possibility to visualize and explore formal concepts in a formal context (a data table that represents binary relations between items in a set of objects and items in a set of attributes) as representations of complete lattices. The mathematical foundations are described as follows (Ganter & Wille, 2012):

> A formal context is a triple K := (G;M; I), where G is a set whose elements are called objects, M is a set whose elements are called attributes, and I is a binary relation between G and M (i.e. I ⊆ X) . (g, m) ∈ I is read "object g has attribute m".
>
> A formal concept of a formal context (G, M, I) is a pair (A, B) with A ⊆ G, B ⊆ M, A' = B and B' = A. The sets A and B are called the extent and the intent of the formal concept (A, B), respectively. The subconcept super concept relation is formalized by:

$$(A1, B1) \leq (A2, B2): \Leftrightarrow A1 \subseteq A2 \ (\Leftrightarrow B1 \supseteq B2).$$

The set of all formal concepts of a context K together with the order relation ≤ is always a complete lattice (i.e. for each subset of concepts, there is always a unique greatest common subconcept and a unique least common super concept), called the concept lattice of K, also called conceptual hierarchy. In a line diagram (in FCA, the term line diagram is used for the Hasse diagram of a lattice) each node represents a formal concept.

Triadic Formal Concept Analysis (3FCA) (Lehmann & Wille, 1995) was introduced to model relations between three sets:

A **triadic context** is defined as a quadruple K := (G;M;B; Y), where G, M, and B are set and Y is a ternary relation between G, M and B, i.e. Y ⊆G×M×B; the elements of G, M, and B are called objects, attributes and conditions, respectively, and (g,m,b) ∈ Y is read: the object g has the attribute m under the condition b.

A **triadic concept** of triadic context (G;M;B; Y) is defined as a triple (A1, A2, A3) with A1×A2×A3 ⊆ Y which is maximal with respect to component-wise inclusion.

Recent work on triadic conceptual navigation (Kis, Sacarea, & Troanca, 2015; Rudolph, Săcărea, & Troancă, 2015) has provided graphical navigation tools such as FCA Tools Bundle which use a local navigation paradigm to make 3FCA visualizations intuitive and applicable.

4. Proposed Analysis Method

The three-step computational thinking, language independent methodology can be summarized as follows:

4.1. Data collection via API interrogation tools or Web scraping

Comment threads associated with Facebook Fan Page posts can be obtained via API interrogation with various tools such as Facepager (Keyling & Jünger, 2013). Other contexts such as news websites comment threads, internet forums or message boards can be scraped with generic Web scraping tools.

4.2. Co-occurrence analysis of terms referring to potential targets of hate speech with codes for hate speech semantic fields

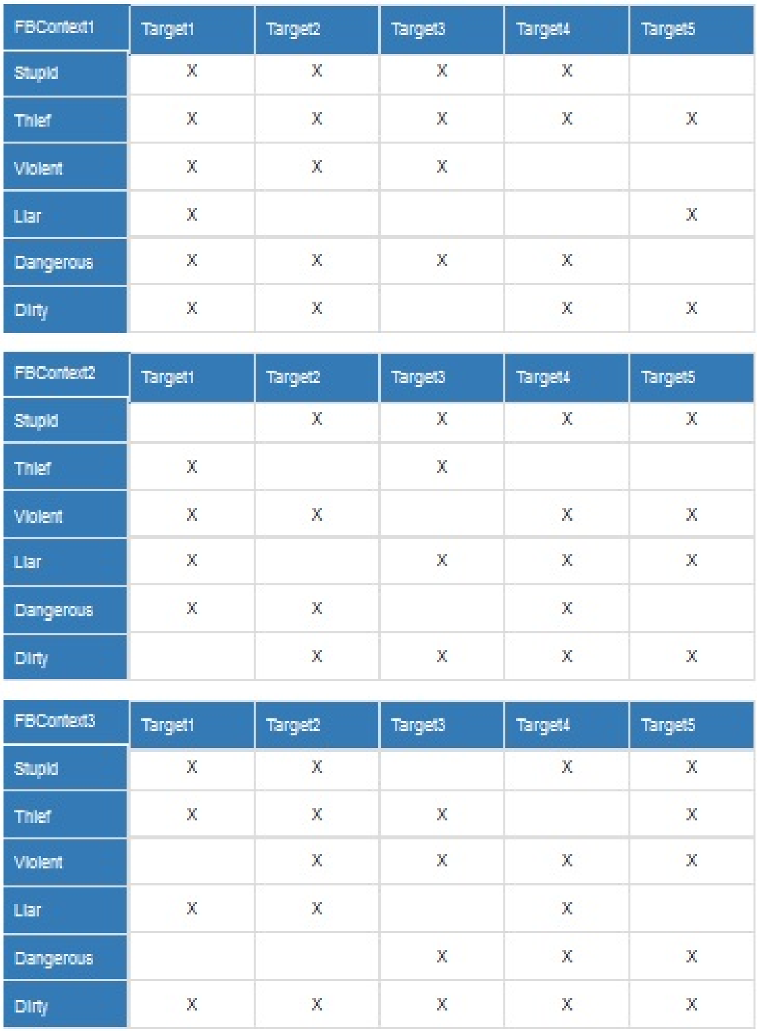
Co-occurrence analysis is deemed an appropriate method for studying large data sets of short texts such as comments or Tweets. For the specific purposes of analyzing hate speech, codes may be defined to identify:

- Key words, terms or expressions referring to targets of hate speech acts (e.g. immigrants, different racial or ethnic groups);
- Key words, terms or expressions that constitute dangerous, hate, offensive or violent speech (e.g. invectives, insults, swear words, negative stereotypes, calls to violent action).

Such a method is used in (Meza, 2016), yielding results in the form of co-occurrence frequency tables indicating the number of occurrences in the same comment (in the context of one or several Facebook Fan Pages) of the two types of codes. Visual representations are possible through tools like KH Coder (Higuchi, 2001), but lack navigation and concept discovery.

4.3. 3FCA visualization and navigation

By establishing threshold values (with respect to the size of the dataset), a co-occurrence matrix may be converted into a formal context. 3FCA allows the definition of objects (targets), attributes (hate codes over content) and conditions (Facebook Fan Pages).



FBContext1	Target1	Target2	Target3	Target4	Target5
Stupid	X	X	X	X	
Thief	X	X	X	X	X
Violent	X	X	X		
Liar	X				X
Dangerous	X	X	X	X	
Dirty	X	X		X	X

FBContext2	Target1	Target2	Target3	Target4	Target5
Stupid		X	X	X	X
Thief	X		X		
Violent	X	X		X	X
Liar	X		X	X	X
Dangerous	X	X		X	
Dirty		X	X	X	X

FBContext3	Target1	Target2	Target3	Target4	Target5
Stupid	X	X		X	X
Thief	X	X	X		X
Violent		X	X	X	X
Liar	X	X		X	
Dangerous			X	X	X
Dirty	X	X	X	X	X

Figure 1. Triadic formal context example

Figure 1 shows a triadic formal context built with FCA Tools Bundle (Meza, 2016), illustrating how co-occurrence analysis data from large datasets of millions of comments can be converted into formal contexts. The objects are defined as the codes for the likely targets of hate speech. Attributes are defined as codes that cover semantic fields of hate speech such as common negative stereotypes. The conditions are defined as the public Facebook Fan Pages where the collected comments originate from.

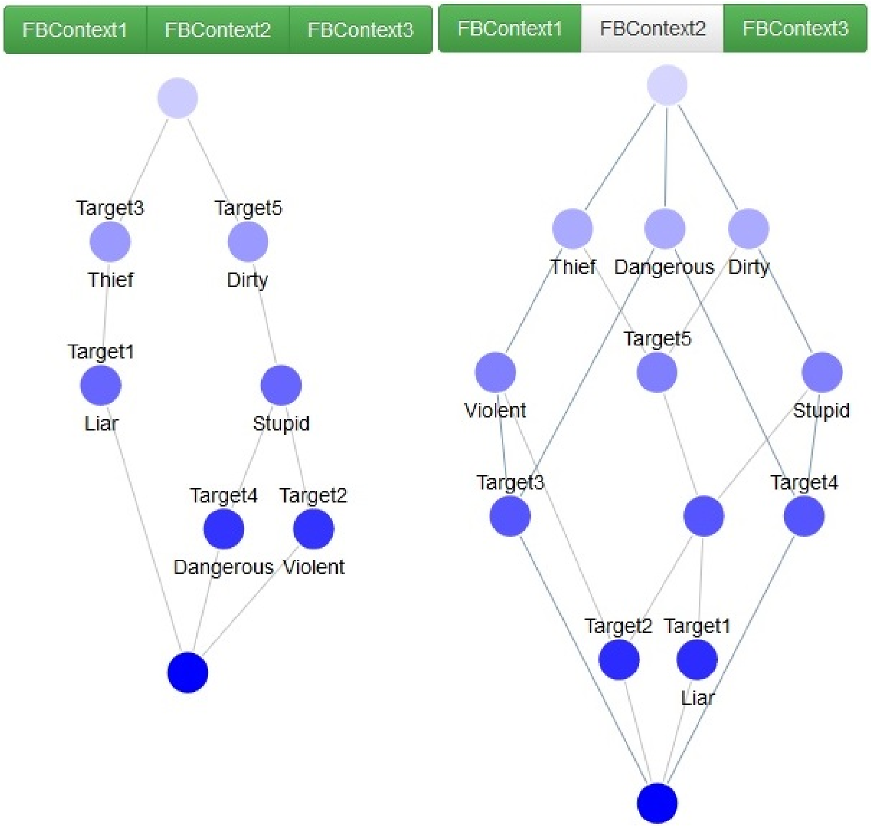


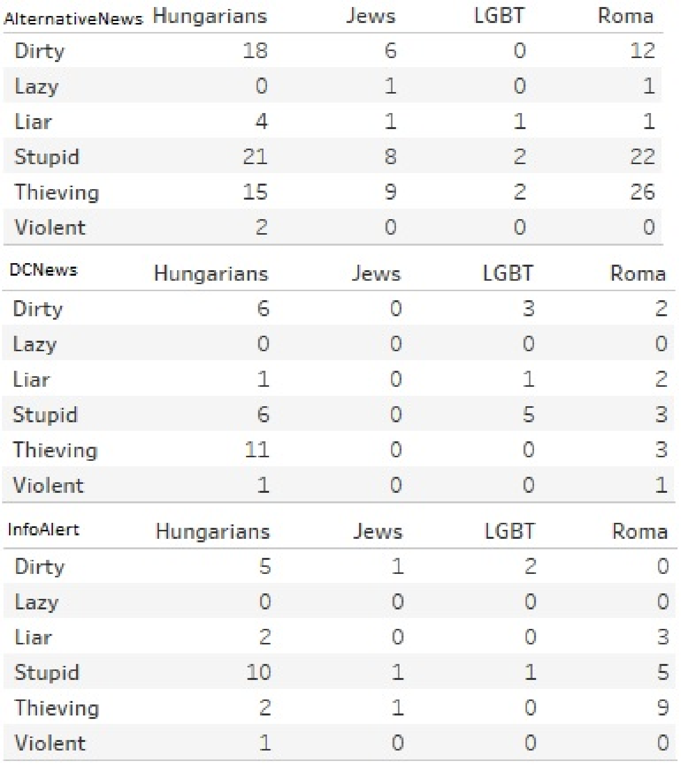
Figure 2. Local navigation of with 3 locked conditions (left) and 2 locked conditions (right)

Line diagram visualizations of 3FCA local navigation (Figure 2) with locked conditions can be used to identify what is being said about which targets in which contexts. For example, Target5 is associated with the attribute *Dirty* in all contexts, whereas in two of them it is also associated with *Thief*. Target4 and Target2 are represented as *Dangerous* and *Violent*, both being associated with *Stupid* and *Dirty*.

5. Case Study: Comments on Facebook Pages of Alternative Online News Outlets

In order to illustrate and evaluate the proposed methodology, a case study was designed. A dataset of 416.554 comments was collected from all the Facebook posts created in January-March 2017 of the Fan Pages of 23 online Romanian language alternative online news outlets. Data collection was done by interrogating the Facebook Open Graph API.

Keyword search was used to identify comments mentioning four groups who are often targets of hate speech: Roma, Hungarians, Jews and LGBT. For each of the groups, variants of the of the popular Romanian terms, both neutral and pejorative, were used. A total of 3620 comments contained references to the four target groups.



AlternativeNews	Hungarians	Jews	LGBT	Roma
Dirty	18	6	0	12
Lazy	0	1	0	1
Liar	4	1	1	1
Stupid	21	8	2	22
Thieving	15	9	2	26
Violent	2	0	0	0

DCNews	Hungarians	Jews	LGBT	Roma
Dirty	6	0	3	2
Lazy	0	0	0	0
Liar	1	0	1	2
Stupid	6	0	5	3
Thieving	11	0	0	3
Violent	1	0	0	1

InfoAlert	Hungarians	Jews	LGBT	Roma
Dirty	5	1	2	0
Lazy	0	0	0	0
Liar	2	0	0	3
Stupid	10	1	1	5
Thieving	2	1	0	9
Violent	1	0	0	0

Figure 3. Co-occurrence matrix of target groups and attributes for comments dataset

Keyword search was also used to detect terms referring to negative attributes such as *Dirty*, *Lazy*, *Liar*, *Stupid*, *Thieving*, *Violent* by several term stems from the appropriate semantic families. A total of 28.914 records in the dataset were found.

Co-occurrence analysis produced the matrix in Figure 3, with 506 comments containing both terms referring to the target groups and the attributes searched. Three Facebook pages which contained the highest number of occurrences were chosen.

The co-occurrence matrix is easily converted into a formal context by establishing a value threshold. In this case, the value threshold was set at a minimum (1).

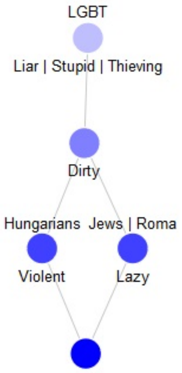


Figure 4. NewsFBPage1

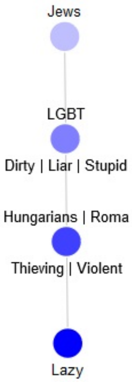


Figure 5. NewsFBPage2

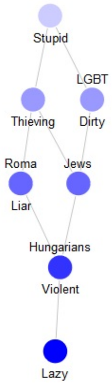


Figure 6. NewsFBPage3

Once FCA is applied using FCA Tools Bundle (Meza, 2016) and the formal concepts detected, each Facebook page context can be visualized as a line diagram as shown in figures 4,5 and 6. This allows comparative analysis of the concept hierarchy and identification of similarities between formal concepts.

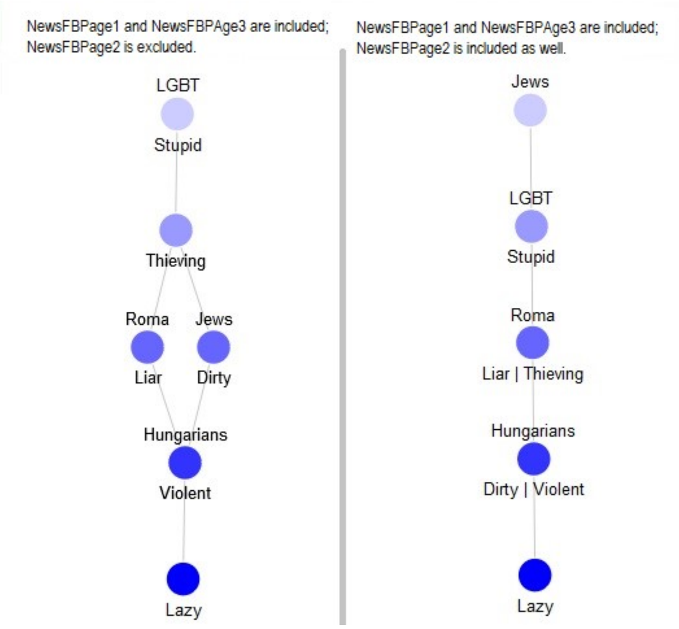


Figure 7. Triadic FCA application

Triadic FCA allows analysts to distinguish between discourse patterns in different contexts. For example, in Figure 7, we can see how the inclusion of one of the contexts may change the position of a target group in the concept hierarchy.

6. Conclusion

The proposed methodology improves on previous approaches by using 3FCA over formal contexts derived from preliminary co-occurrence analysis of content, emitters and targets in one or several contexts - comments to posts on popular Facebook Fan Pages.

The advantages of this methodology are that it is scalable (co-occurrence analysis of codes in short texts is not a very complex task for large datasets), language independent, and visually intuitive, hence easy to interpret for digital media specialists, linguists or social scientists. Introducing computational thinking in problem-solving like hate-speech analysis and making use of freely available software tools referenced here may provide new opportunities for digital social researchers, journalists, professional communicators and digital media specialists. Integrating API interrogation, Web scraping, co-occurrence analysis and 3FCA in future tools may prove useful in unstructured data analysis education.

Acknowledgements

This work was supported by a grant of Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2016-0892, within PNCDI III.

References

Angi, D., & Bădescu, G. (2014). *Discursul instigator la ură în România*. Fundația pentru Dezvoltarea Societății Civile. Retrieved from http://www.fdsc.ro/library/files/studiul_diu_integral.pdf.

Donahue, P. (2015, September 26). Merkel Confronts Facebook’s Zuckerberg Over Policing Hate Posts. *Bloomberg.com*. Retrieved from https://www.bloomberg.com/news/articles/2015-09-26/merkel-confronts-facebook-s-zuckerberg-over-policing-hate-posts.

Facebook outsources fight against racist posts in Germany. (2016, January 15). *Reuters*. Retrieved from http://www.reuters.com/article/facebook-germany-idUSKCN0UT1GM.

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech*. UNESCO Publishing. Retrieved from https://www.google.com/books?hl=en&lr=&id=WAVgCgAAQBAJ&oi=fnd&pg=PA3&dq=online+hate+speech+unesco&ots=TaamaoJQVB&sig=xUFAShQSkdkdHtMImSPL50myDRE.

Ganter, B., & Wille, R. (2012). *Formal concept analysis: mathematical foundations*. Springer Science & Business Media. Retrieved from https://www.google.com/books?hl=en&lr=&id=hNwqBAAAQBAJ&oi=fnd&pg=PA1&dq=Ganter,+B.,+Wille,+R.:+Formal+Concept+Analysis+-+Mathematical+Foundations.&ots=0cRL5XEe3p&sig=P4IzDkOx5d0I4IeSnYB6_5QVRJM

Gitari, N. D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(4), 215–230.

Higuchi, K. (2001). Kh coder. *A Free Software for Quantitative Content Analysis or Text Mining, Available at: Http://khc. Sourceforge. Net/en*.

Keyling, T., & Jünger, J. (2013). *Facepager. An application for generic data retrieval through APIs*. Source code available at https:// github. com/ strohne/ Facepager.

Kis, L. L., Sacarea, C., & Troanca, D. (2015). FCA Tools Bundle-a Tool that Enables Dyadic and Triadic Conceptual Navigation. *Proc. of FCA4AI*. Retrieved from https://hal.univ-lorraine.fr/hal-01425672/document#page=45

Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., … Werner, L. (2011). Computational thinking for youth in practice. *Acm Inroads*, *2*(1), 32–37.

Lehmann, F., & Wille, R. (1995). A triadic approach to formal concept analysis. In *International Conference on Conceptual Structures* (pp. 32–43). Springer. Retrieved from http://link.springer.com/chapter/10.1007/3-540-60161-9_27

Meza, R. (2016). Hate Speech in the Romanian Online Media. *Journal of Media Research*, *9*(3(26)), 55–77.

Rudolph, S., Săcărea, C., & Troancă, D. (2015). Towards a navigation paradigm for triadic concepts. In *International Conference on Formal Concept Analysis* (pp. 252–267). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-19545-2_16

Scott, M., & Eddy, M. (2016, November 28). Facebook Runs Up Against German Hate Speech Laws. *The New York Times*. Retrieved from http://www.nytimes.com/2016/11/28/technology/facebook-germany-hate-speech-fake-news.html

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35.

Radu Mihai Meza (b. November 15, 1985) received his BSc in Computer Science (2008), Bachelor of Communication Sciences - Journalism (2008), MA in Media Communication (2009) and Ph.D. in Sociology (2012) from ”Babeș-Bolyai” University, Cluj-Napoca. He is associate professor in the Department of Journalism, College of Political, Administrative and Communication Sciences at ”Babeș-Bolyai” University.

Șerban Nicolae Meza (b. May 7, 1983) received his Bachelor of Economics and Business Management (2006) from ”Babeș-Bolyai” University, Cluj-Napoca, B.Eng. in Telecomunications (2007), M.Eng. in Signal and Image Processing (2008), and Ph.D. in Electronics and Telecomunications from the Technical University of Cluj-Napoca. He is lecturer in the Faculty of Electronics, Telecommunications and Information Technology at the Technical University of Cluj-Napoca.