

## CONCEPTUAL FRAMEWORK – Working Notes

ADHOC PN-III-P1-1.1-TE-2016-0892 | Contract number: 35/2018

### 1 Concepts

| 1.1 Normative (regulatory or political)   | 1.2 Theoretical   | 1.3 Methodological  |
|---|---|---|
| Free speech<br>Hate speech<br>Vulnerable group<br>Protected group<br>Hate crime<br>Bias against a member of community<br>Bias motivated crime<br>Prejudice against a member of community<br>Incitement against community<br>Right-wing extremism<br>Intolerance<br>Discrimination<br>Harassment<br>Solidarity and cohesion in society | Speech act<br>Antagonistic speech<br>Offensive speech<br>Dangerous speech<br>Social group<br>Stereotype<br>Prejudice<br>Insult<br>Slur<br>Verbal aggression:<br>Cyberbullying | Niche dictionary<br>Semantic network<br>Network agenda setting<br>Co-occurrence<br>Community<br>Discursive patterns |

### 1.4 Definitions

For legal definitions see document on International Regulatory Framework

Gagliardone et al. 2015

Facebook's Community Standards [summarized in Intersection paper]

Facebook defines hate speech with respect to `protected characteristics`:

“We define hate speech as a direct attack on people based on what we call protected characteristics — race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. We also provide some protections for immigration status. We define attack as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation.”

“Content that describes or negatively targets people with slurs, where slurs are defined as words commonly used as insulting labels for the above-listed characteristics.”

The company's policy explicitly mentions that the above criteria apply to both verbal and visual content, and also defines special cases of admissibility such as raising awareness, education,

<http://adhoc.granturi.ubbcluj.ro/>

self-referentiality, empowering expressions, humor and social commentary with clearly identifiable intent.

### **1.5 Incitement to hatred:**

In the practice of the Hu Media Council: ““an extreme and forceful expression of antipathy or hatred directed towards the given community, and which is likely to arouse similar feelings in others and thus provoke the risk of violations of the rights of this community. Thus, incitement to hatred as defined in the law is literally the emotional preparation for violence, and means incitement to the violent resolution of conflicts” (Hungary : Responding to ‘ hate speech ’. 2018 Country Report, 2018, p. 40.)

Content that results in exclusion:

““attempts to reach – or argues in favour of – the isolation of a given community, its alienation from other segments of society, or its segregation. In practice this effect can be brought about if the content in question

reinforces misleading or stereotypical ideas or opinions in the target audience, or aims to do so”. (Hungary : Responding to ‘ hate speech ’. 2018 Country Report, 2018, p. 40.) This can be the case regardless whether the effect was actually achieved, what matters from the point of view of the offence is the intent (ibid.)

### **1.6 Typologies**

Facebook’s Community Standards [summarized in Intersection paper]

The categories of hate speech are defined under three tiers and a supplementary category:

“Tier 1 attacks, which target a person or group of people who share one of the above-listed characteristics or immigration status [...]

Tier 2 attacks, which target a person or group of people who share any of the above-listed characteristics [...]

Tier 3 attacks, which are calls to exclude or segregate a person or group of people based on the above-listed characteristics. We do allow criticism of immigration policies and arguments for restricting those policies.

<http://adhoc.granturi.ubbcluj.ro/>

Gagliardone et al. 2015

Offensive speech: Antagonistic statement on group criteria (other than political); does not incite the audience

Hate speech: Uses derogatory terms, insult, humiliate, boycott, discriminate against an individual/group (individual because they are part of a group); Incites / encourages the audience to use derogatory terms, insult, discriminate against the individual/group

Dangerous speech: calls to violent action against a group, riot, loot, beat, evict, kill, other forms of physical violence; Incites the audience

Institutul Național pentru Studierea Holocaustului din România “Elie Wiesel”, 2017

proposes an analysis of the types of antisemitism using the following dominant categories: religion, racism, conspiracy, economy, anti-Israeli, demonizing

stereotypes associated with the Roma community: inferior race, criminals, uneducated / uncivilized, demographic threat, cultural threat, social welfare recipients, prejudicing the image of Romania

### **1.7 Hate as a social media affordance?**

KhosraviNik, M. (2018). Social media techno-discursive design, affective communication, and contemporary politics. *Fudan Journal of the Humanities and Social Sciences*, 1–16.

“At the interface of participatory web affordance and the user, users have the (perception of) chance of being able to act completely individualistic and are encouraged to prioritize their affective moods, e.g., rage, fear, etc. while in the past these affective qualities of individuals would be filtered, controlled and moderated by various systems of mass media gate-keeping.”

### **1.8 Limitations of our research:**

the problem with studying “hate”, “offense” or “danger” as a matter of content

“Hate” is an affect, “offense” is an effect, “danger” is a contextual potential - all these pertain to speaker, recipient and context

When setting out the norms for assessing the potential to incitement of an utterance, the Rabat Plan of Action of the UN OHCHR designates the following factors to consider: context; identity

<http://adhoc.granturi.ubbcluj.ro/>

of speaker; intent of speaker; content; extent and magnitude; likelihood of harm - of these, we can only study content and extent, maybe expressions of intent

## 2 Context

As digital communication becomes a bigger part of our lives and both the real and virtual world become increasingly globalized and diverse, new issues such as studying online hate speech make their way onto the scientific and policy agenda. With new networked digital platforms, collapsed public, semi-public, and private contexts and a wealth of data in public online conversations, digital social science methodologies move into the area of computational approaches.

In the international and European socio-political context, with economic migrants, refugees and increasing waves of extremism and xenophobia, hate-speech is becoming an increasingly important topic. Where the fundamental human right to freedom of speech and expression collides with the increasing need for tolerance and mutual respect demanded by life in racially, ethnically, and religiously diverse, multi-cultural societies, hate-speech becomes an important preoccupation for researchers, law-makers, civil society and stake-holders in public mediated communication, irrespective of the communication medium.

The issue of online hate speech has risen in importance in global and European debates over the past few years. Although European laws regulate hate speech acts, computer mediated communication through digital platforms owned by businesses outside the users' country may be subject to different legislation. At the heart of the most heated debate is the social media giant, Facebook, with its platform being used by approximately 2.2 billion people globally. Governments and NGOs look towards this company for mechanisms that properly deal with antagonistic speech, in accordance with national policies. The case of Germany, one of the European countries with the strictest regulatory frameworks concerning antagonistic speech, is well known. On the backdrop of the European refugee crisis, Facebook was pressured to take action and announce an initiative to deal with racist content on its German website<sup>1</sup>. In early

---

<sup>1</sup> Merkel Confronts Facebook's Zuckerberg Over Policing Hate Posts. Bloomberg.com. <https://www.bloomberg.com/news/articles/2015-09-26/merkel-confronts-facebook-s-zuckerberg-over-policing-hate-posts> Accessed: 06-01-2018

<http://adhoc.granturi.ubbcluj.ro/>

2016, the company reacted to public criticism over its reluctance to deal with hate speech within EU and European national legal frameworks by outsourcing the moderation of racist posts.<sup>2</sup> At the end of 2016, social media activity and social or political effects associated with it have driven lawmakers in both Europe and the United States 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*’<sup>3</sup>.

Romania and Hungary provide interesting cases for comparative research on the issue of online hate speech in Central and Eastern Europe for several reasons. As neighboring countries, the two share history and culture, and throughout the past century, the two modern national states have had conflicting territorial claims. Transylvania, a region of Romania since 1918, is inhabited by a substantial Hungarian minority. The two countries also share a recent common experience of communist regimes and propaganda until '89, both being part of the block behind the Iron Curtain. The two languages are significantly different - Romanian is an Eastern Romance language, Hungarian is a Finno-Ugric language, hence well suited for exploring differences in methodological approaches to studying the linguistic aspects of hate speech. Lastly, recent social and political developments in the two countries - from use of Facebook in relation to political debate and participation, use of social media by media institutions, alternative media and activism groups, and content of media and political agendas, drive research interest for a comparative approach. Comparative research traditionally aims to highlight differences between cases worth comparing, as our two countries are. However, sometimes similarities are also interesting, as they may be indicative of trends and phenomena that transcend the national context or linguistic boundaries. Whether hate speech has such components, beyond the obviously context-specific ones, is an issue central to our paper and worth investigating further.

---

<sup>2</sup> Facebook outsources fight against racist posts in Germany. *Reuters*. <http://www.reuters.com/article/facebook-germany-idUSKCN0UT1GM> Accessed: 06-01-2018

<sup>3</sup> Facebook Runs Up Against German Hate Speech Laws. *The New York Times*. <http://www.nytimes.com/2016/11/28/technology/facebook-germany-hate-speech-fake-news.html> Accessed: 06-01-2018

<http://adhoc.granturi.ubbcluj.ro/>

### 3 Approaches to studying hate speech

A 2015 UNESCO study (Gagliardone et al., 2015) outlines the key issues relevant in countering online hate speech:

**Definition:** There are multiple, differing definitions of hate speech, some mixing concrete threats to the security of individuals and groups with expressions of frustration and anger. Digital media communication platforms such as Facebook, Twitter or Google each define their own policies towards admissible content published by their users. However, as recent tensions have shown, these often clash with national legislation and consensus seems unlikely.

**Jurisdiction:** 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.

**Comprehension:** 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.

**Intervention:** 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, popular online social network type platforms seem reluctant to publish aggregate results that would allow an overview of the phenomenon.

The academic approach towards studying hate speech defines the phenomenon as an act of communication. An overview of the issue in the Romanian national context (Angi and Bădescu, 2014) recommends focusing on: content (what is being said); emitters (who is communicating); targets (who is the message about); and context (including when the act takes place).

A similar point is made in the context of Hungarian legal case studies by Peter Smuk, who argues that hate speech, understood as speech that incites hatred against persons or social groups, can be defined in terms of “actors (orators), the contents, targets (victims) and social dangers posed.” (Smuk, 2015: 64.)

For the purposes of this research, the main focus will be studying the mentions of targets, defined here as vulnerable groups in each national case (as identified by previous scientific literature) and the context - virtual space, temporal coordinates and conversational themes.

<http://adhoc.granturi.ubbcluj.ro/>

### 3.1 Defining hate speech

For the purposes of this research, the issue of defining hate speech is the most important. According to (Gagliardone et al., 2015: 19) ‘ICCPR [International Covenant on Civil and Political Rights] is the legal instrument most commonly referred to in debates on hate speech and its regulation, although it does not explicitly use the term’. The problem of defining hate speech is approached by researchers in various fields. In the case of online hate speech, the issue is particularly linked to jurisdiction - although there seems to be a consensus that it targets disadvantaged social groups in potentially harmful ways. Definitions exist in different national contexts but may differ substantially from each-other or those used by social media platforms in their content policies and community guidelines.

Although Facebook has been under criticism since 2015 for not blocking some content, especially by institutions and policy groups in the EU, the company released its Community Standards on April 24 2018<sup>4</sup>, stating its policy rationale in blocking hate speech because it ‘creates an environment of intimidation and exclusion and in some cases may promote real-world violence’. Its choice of definitions and approach are discussed as early as June 2017.<sup>5</sup>

Facebook defines hate speech with respect to ‘protected characteristics’:

‘We define hate speech as a direct attack on people based on what we call protected characteristics — race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. We also provide some protections for immigration status. We define attack as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation.’<sup>6</sup>

The categories of hate speech are defined under three tiers and a supplementary category:

‘Tier 1 attacks, which target a person or group of people who share one of the above-listed characteristics or immigration status [...]

---

<sup>4</sup> Facebook reveals 25 pages of takedown rules for hate speech and more.

<https://techcrunch.com/2018/04/24/facebook-content-rules/> Accessed: 13-10-2018

<sup>5</sup> Hard Questions: Who Should Decide What Is Hate Speech in an Online Global Community?

<https://newsroom.fb.com/news/2017/06/hard-questions-hate-speech/> Accessed: 10-06-2018.

<sup>6</sup> *Community Standards*. Facebook [https://www.facebook.com/communitystandards/hate\\_speech](https://www.facebook.com/communitystandards/hate_speech) Accessed: 13-10-2018

<http://adhoc.granturi.ubbcluj.ro/>

Tier 2 attacks, which target a person or group of people who share any of the above-listed characteristics [...]

Tier 3 attacks, which are calls to exclude or segregate a person or group of people based on the above-listed characteristics. We do allow criticism of immigration policies and arguments for restricting those policies.

Content that describes or negatively targets people with slurs, where slurs are defined as words commonly used as insulting labels for the above-listed characteristics.<sup>7</sup>

The company's policy explicitly mentions that the above criteria apply to both verbal and visual content, and also defines special cases of admissibility such as raising awareness, education, self-referentiality, empowering expressions, humor and social commentary with clearly identifiable intent.

### **3.2 Studies of Online Hate Speech in Central, Southern and Eastern Europe**

Although still relatively scarce, scholarship on online hate speech in Central and Eastern Europe has been emerging at a fast pace in the past decade from both academics and NGOs.

The overview of the issue mentioned above (Angi and Bădescu, 2014) finds the most frequent targets of hate speech in the Romanian national context as being the Roma, Hungarian and Jewish groups and members of the LGBTQ+ sexual minorities. Similarly, in Hungary, the most frequently targeted groups are reported to be the Roma, the Jews, the LGBTQ community, and, in recent years, refugees and migrants (Article 19, 2018: 8.) In Hungary, the very definition of hate speech or of incitement to hatred has also been the topic of highly politicized debates, the overview of which is beyond the scope of this paper (see Boromisza-Habashi 2011; Pál 2015.) The NGO sector has taken an increasing interest over the past two years in analyzing and developing strategies for countering hate speech in the traditional and online media. Reports and academic works emanating from these initiatives are starting to shape an emerging scholarship on the issue (Răileanu et al., 2016; Hann & Róna, 2017).

Existing academic research and the numerous reports emanating from the NGO sector focus mainly on legislation, media self-regulation and intervention strategies, while the actual contents of hate speech acts in the online media, especially in social media are analyzed mostly through case study methodology, leading to potential hasty generalizations or possibly

---

<sup>7</sup> *Community Standards*. Facebook [https://www.facebook.com/communitystandards/hate\\_speech](https://www.facebook.com/communitystandards/hate_speech) Accessed: 13-10-2018



<http://adhoc.granturi.ubbcluj.ro/>

overlooking some targets, contexts or emitters. The issue of hate speech in Central and Eastern European context has been approached mostly from a regulatory or normative perspective in relation to Western Europe and the United States in comparative studies (Heinze, 2013). However, it is only very recently that academic researchers started investigating the niche topic of online hate speech, making use of computational approaches towards the collection and analysis of large datasets of comments on news web sites, blogs and especially social media (Meza, 2016).

Other recent multi-country initiatives investigate the issue of online hate speech in the South Eastern Europe region, in countries such as Albania, Bosnia and Herzegovina, Kosovo, Macedonia, Montenegro, Serbia and Turkey, following international standards in raising issues such as the broader socio-historical context of the expression, the identity and intent of the speaker, the content of the expression and the magnitude of distribution and likelihood of ensuing discrimination.<sup>8</sup>

The Report on hate speech against Jewish and Roma groups on social media proposes an analysis of the types of antisemitism using the following dominant categories: religion, racism, conspiracy, economy, anti-Israeli, demonizing.<sup>9</sup> For analyzing of the stereotypes associated with the Roma community, the authors appeal to different categories such as: inferior race, criminals, uneducated/uncivilized, demographic threat, cultural threat, social welfare recipients, prejudicing the image of Romania. In Hungary, sociological research into anti-semitic attitudes, although not investigating online discourses themselves, also point to the importance of the online environment in the rise of antisemitism after 2010, as it appears to enable the spread of conspiracy-theories and misinformation in an age of post-truth (Hann and Róna, 2017: 38.)

#### **4 Networked Agendas - The media, politicians and the networked public**

Over the last 50 years, the agenda setting evolved from an initial focus on media effects on the public's perception of the most important issues to a more complex, hierarchical approach of the communication effects.

---

<sup>8</sup> *Hate speech in online media in South East Europe*. Albanian Media Institute. Tirana. <http://www.institutemedia.org/Documents/PDF/Hate%20speech%20in%20online%20media%20in%20SEE.pdf> Accessed: 10-06-2018

<sup>9</sup> *Raport cu privire la discursul instigator la ură împotriva evreilor și romilor în social media* (Report on hate speech against Jews and Roma in social media). Institutul Național pentru Studiarea Holocaustului din România "Elie Wiesel" [http://www.inshr-ew.ro/ro/files/proiecte/DIU/DIU\\_social\\_media\\_1.pdf](http://www.inshr-ew.ro/ro/files/proiecte/DIU/DIU_social_media_1.pdf) Accessed: 10-10-2018.

<http://adhoc.granturi.ubbcluj.ro/>

The news media transfer the salience of relationships between sets of objects and attributes to the public. These sets of relationships between elements of the media and public agendas are the third level of agenda-setting (Guo, 2014). This perspective on the bundling of agenda elements – the third level of agenda-setting – tests an agenda-setting hypothesis that the salience of relationships on the media network issue agenda can be transferred to the public network issue agenda (McCombs et al., 2014)

The Network Agenda Setting Model borrows concepts from the associative network model of memory and asserts that individuals' cognitive representation of objects and attributes is presented as a network-like structure where any particular node will be connected to numerous other nodes. This recent theoretical approach asserts that in order to describe an individual a person generates a network-shaped picture composed of various attributes which are connected to each other in his/her mind (Guo et al., 2012).

In the context of this research, beyond identifying and quantifying the mentions of targets of hate speech in comments to Facebook posts by news media, political leaders and political parties, the analysis of co-occurrence networks between such mentions, negative qualifiers and institutions connected to recurrent themes in society may reveal directions to be explored further. Beyond target groups identified by previous researchers studying hate speech in the Romanian and Hungarian national contexts, the research will try to identify mentions of social groups such as refugees, welfare recipients and pensioners who were salient in the media and political agenda in the two countries in the time-frame of the analysis.

## **5 Terms in Context and Co-occurrence Analysis**

Although text mining and natural language processing tools are being increasingly used by social scientists to study digital documents, there is still a considerable gap between tools available for international languages such as English, French, Spanish, Italian or German and languages which are spoken only in national contexts such as Hungarian or Romanian. Although in the past years, resources for languages such as Romanian and Hungarian have been increasingly made available, and newer approaches based on machine learning applied to large enough corpora are more and more language independent, social investigations on online hate speech in the two national contexts mostly apply traditional qualitative and quantitative analysis methods.

<http://adhoc.granturi.ubbcluj.ro/>

The exploratory approach presented here is based on researcher-defined niche dictionaries (of targets/vulnerable groups, issues/concepts/institutions and qualifiers defined as semantic families) and descriptive statistics in relation to contextual variables (Facebook page source and category, timeframe of the comment thread). Furthermore, this research uses semi-automated coding based on the above-mentioned niche dictionaries (for targets and issues/concepts) to map co-occurrences between the two categories. This approach allows for the identification of contexts where antagonistic speech has the potential to appear. Large scale research studies such as this require more advanced natural language processing tools (and machine learning techniques) for Hungarian and Romanian to provide automated classification of content. It is worth noting that even Facebook relies on the decision-making ability of over 7000 content moderators classify and potentially block such content from the platform.

Co-occurrence analysis is used to identify relations between the target groups and social institutions, issue concepts or qualifiers relating to stereotypes (based on semantic families). This method combines quantitative content analysis approaches (code/term frequencies) with network analysis (relations based on co-occurrence of terms/codes in the same context - e.g. in the same comment) (Danowski, 1993). The merits of the method are particularly notable when analyzing content produced in computer mediated communication, especially in the case of short text messages/documents such as user comments where the significance of two terms co-occurring in the same text is higher. Furthermore, applying network analysis methods, groups of well-connected terms or concepts may be detected by using algorithms for community structure detection in graphs (Clauset et al., 2004). As edges are defined based on co-occurrence relationship (the two terms or concepts coded appear in the same message), edges connecting separate (or loosely connected) parts of the graph will have high betweenness scores (they will frequently be found on the shortest path connecting those parts of the graph). A hierarchy of well-connected modules can be established by identifying edges with high betweenness scores, eliminating them, and then reiterating multiple times. As a result, communities will emerge as dense, well-connected groups of nodes, or in this case terms or concepts coded from the comments corpus. This approach may reveal latent connections.

## **6 Populist Discourses**

“(Right-wing) populism [...] does not convey a coherent narrative and ideology, but rather proposes a mixed, often contradictory array of beliefs, stereotypes, attitudes and related

<http://adhoc.granturi.ubbcluj.ro/>

programmes which aim to address and mobilize a range of equally contradictory segments of the electorate” (Wodak 2015)

Right-wing populist parties instrumentalize some kind of ethnic/religious/linguistic/political minority as a **scapegoat** for most if not all current woes and subsequently construe the respective group as dangerous and a threat ‘to us’, to ‘our nation’ (Wodak 2015)

## 7 Tropes, memes, myths, stereotypes, Topoi

- **Tropes** – Hayden White discusses the tropes of discourse by applying literary criticism to the analysis of representations in historical writing. Main tropes: metaphor, metonymy, synecdoche, irony.
- **Memes** – Richard Dawkins coins the term meme to refer to “cultural genes” in the context of his gene-centric reinterpretation of evolutionary theory. The concept gains popularity with respect to internet culture – specifically when referring to **internet memes** or memes in digital culture defined by Limor Shifman (2014)
- **Myths** – Roland Barthes defines the concept of (contemporary) myth in his essay Myth Today as a second order meaning in the realm of culture (some prefer to clarify this concept as: ideological narrative / ideological implication)
- **Stereotypes** – stereotypical thinking and stereotypes emerge as a way of simplifying the demands on the perceiver, rely on previously stored knowledge, may be a response to environmental factors – different social roles, group conflicts, differences in power. They may be a way to justify the status quo or might serve as a response to a need for social identity. (Hilton and Hippel 1996)
- **Topoi** – of urgency, threat, the saviour, scapegoat, history (Wodak 2015)

## 8 Self-presentation of the government institutions on social media

The cultural notions about notoriety, celebrity, and fame appear to be expanding and inclusive, thanks to personalized broadcast channels and social media. Thus, public institutions compete with professional news outlets, with professional and amateur communicators, macro and microcelebrities (Senft, 2013). Therefore, government institutions, in order to reach the audiences, need to provide an attractive content on social media. Harcup and O’Neill (2017, p.13) propose an updated set of 15 contemporary news values that, in various combinations,

<http://adhoc.granturi.ubbcluj.ro/>

seem to be identifiable within published informational content. Authors found that potential news stories must generally satisfy one and preferably more of the following requirements to be selected: exclusivity (stories generated by the organization), bad news (death, injury, defeat, loss), conflict (controversies, arguments, splits, warfare), surprise (unexpected or unusual elements), audio-visuals (photo, video, audio, infographics), shareability (the potential to generate sharing and comments), entertainment (showbiz stories, human interest, light and humorous stories), drama (unfolding accidents, rescues, searches), follow-up stories, the power elite (individuals or organizations), relevance (stories about influential actors familiar to the audience), magnitude (stories with potential impact), celebrity (already famous people), good news (recoveries, breakthroughs, cures, wins and celebrations), organization's agenda (stories that fit the organization's own agenda).

Some of the aforementioned informational features could be identified in DePaula's coding scheme of symbolic and presentational purposes of the government organizations. The self-representation communicational purposes were coded as follows: Favorable presentation (seeks attribution of likability, competency or worthiness, reporting of positive activity performed by the department, with positive imagery or self-referential language of gratitude); Political positioning (taking a clear stance on a political issue); Symbolic act (expressing congratulations, gratitude, condolences, celebration of holidays or trivia questions; references to cultural symbols); Branding and marketing (elaborate presentation of features of item or service, including qualities of item with intention to attract individuals to acquire) (DePaula, et al., 2018, p. 5)

## **9 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).

<http://adhoc.granturi.ubbcluj.ro/>

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.

## 10 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.  $I \subseteq G \times M$ ).  $(g, m) \in I$  is read “object  $g$  has attribute  $m$ ”.

A formal concept of a formal context  $(G, M, I)$  is a pair  $(A, B)$  with  $A \subseteq G$ ,  $B \subseteq M$ ,  $A' = B$  and  $B' = A$ . The sets  $A$  and  $B$  are called the extent and the intent of the formal concept  $(A, B)$ , respectively. The subconcept superconcept relation is formalized by:

$$(A_1, B_1) \leq (A_2, B_2): \Leftrightarrow A_1 \subseteq A_2 (\Leftrightarrow B_1 \supseteq B_2).$$

The set of all formal concepts of a context  $K$  together with the order relation  $\leq$  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 superconcept), 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 sets and  $Y$  is a ternary relation between  $G$ ,  $M$  and  $B$ , i.e.  $Y \subseteq G \times M \times B$ ; the elements of  $G$ ,  $M$ , and  $B$  are called objects, attributes and conditions, respectively, and  $(g, m, b) \in Y$  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  $(A_1, A_2, A_3)$  with  $A_1 \times A_2 \times A_3 \subseteq Y$  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.