NetAC, An Automatic Classifier of Online Hate Speech Comments

Constança Elias, Jorge Brandão Gonçalves, Maria Araújo, Pedro Pinheiro, Cristiana Araújo(✉), and Pedro Rangel Henriques

Universidade do Minho, Braga, Portugal
pg44421@alunos.uminho.pt

Abstract. Nowadays in many linguistic and social areas researchers collect the reaction of people to someone's statements (newspaper articles or social network posts) with the intention of analyzing the speech style. From that analysis different conclusions can be inferred giving rise to a large number of social impact attitudes. However it is not enough to create a huge corpus of texts. It is necessary to process the collected statements and comments and resort to appropriate tools to extract the relevant terms from the texts and analyze their occurrences. This paper is about a statistical framework, NetAC, built in the context of NetLang Project to study prejudice discourse aiming at individual or group discrimination. Given a categorization table the tools included in NetAC search for frequency of occurrence of the keywords in each category and, based on the greatest frequency, propose a classification for each comment and for the overall text. Besides the main classifier, other features will be presented.

Keywords: Linguistic analysis support tools · Data analysis · Hate speech · Aggressive language · Social media

1 Introduction

Every day we realize that people often use social networks, online newspaper sites, online platforms such as YouTube, forums, among others [10,14]... On these platforms, millions of people express their opinions, beliefs, lifestyles, etc., through comments (in the comment sections) on someone else's statements. In many cases, in these comments people often resort to the use of Aggressive Language and even Hate Speech [3,7,15].

Aggressive language is often characterized by personal attacks and uncivil language, such as provocative, sarcastic, insulting, cynical or negative comments [1,2,8,10,13,15].

Hate Speech is defined as "Hatred or disqualification of an individual or a group based on their race, skin colour, ethnicity, sex, disability, religion or sexual orientation" [12]; "A mechanism of subordination for generating an atmosphere of fear, intimidation, harassment and discrimination" [11]. Fortuna and Nunes

Á. Rocha et al. (Eds.): WorldCIST 2021, AISC 1367, pp. 494–505, 2021.
https://doi.org/10.1007/978-3-030-72660-7_47

define Hate Speech as being "Language that attacks or diminishes, that incites violence or hate against groups, based on specific characteristics such as physical appearance, religion, descent, national or ethnic origin, sexual orientation, gender identity or other, and it can occur with different linguistic styles, even in subtle forms or when humour is used" [5]. Although the definitions of Hate Speech are slightly different, they often include hate speech or violence towards an individual or group based on characteristics such as: nationality or ethnic origin, sexual orientation, gender identity, religion, physical appearance, among others...

Both, Hate Speech and Aggressive Language, are subclasses of Socially Inappropriate Discourse (SID). SID "covers all forms of aggressive, confrontational, offensive and obscene language, as well as discourse in which hate takes on a more militant nature and is directed at specific groups" [4].

1.1 NetLang Project

This problem served as motivation for the NetLang project, which targets the analysis of the comments present in online newspaper sites and online social platforms, aiming at identifying words, idioms, recurring constructions, pragmatic interaction patterns, among others, that can serve as the basis for detecting hate speech and aggressive language. The project aims at building and annotate a comparable corpus of online texts on a variety of similar topics in Portuguese and English [6].

In a first phase of the project, a table of keywords was built that includes the following elements: different types of prejudice; the corresponding sociolinguistic variables for which they are intended (specifying Hyperonym and Hyponym); and the various keywords and expressions (in English and Portuguese) that are considered typical lexical signs of expression of prejudice. This table aims at assisting in the search for texts to be analyzed by Linguists or other Experts in social science areas; it also will assist in the classification of those same texts by sociolinguistic variable [6]. In Fig. 1 a fragment of the keywords table is displayed. To see the full table of keywords please access: https://bit.ly/netlang-keywords-table.

TYPES OF PREJUDICE	SOCIOLINGUISTIC VARIABLES		KEYWORDS (English)	KEYWORDS (Portuguese)
	Hyperonym	Hyponym		
HOMOPHOBIA	Sexual Identity	General	Homophobia, Gay, Queer, LGBT, Homophobic, Homoerotic, Homosexual	Homofobia, Gay, LGBT, Homoerótico, Homofóbico, Homossexual
		Male homosexuality	Gay, Queer, Drag queen, LGBT, - Beta male, Effeminate, Fag, Fagel, Faggot, Fruit, Fruitcake, Fudgepacker, Girlie man, Neck beard, Nelly, nellie, Pansy, Poof, Sissy, Twink	Abichanado, Amaneirado, Bicha, Bichona, Efeminado, Larilas, Maricas, Maricão, Mariconço, Pan*sca, Paneleiro, Rabeta, Roto, -Bicharia, Maricote, Maricotes, Paneleiragem, Pega de empurão, -Aberração Veado, Boiola, Baitola, Queima rosca, Meio afeminado, Coisa de boiola, Parece uma bixa, Jeitinho gay, Jeito gay
		Female homosexuality	Lesbian, - Battle-axe, Butch woman, Dike, Tomboy	Fufa, Machona, Maria-rapaz, Matrafona, Lambe c*nas
		Transexuality	Ladyboy, LGBT, Transsexual, Transgender, Transition, Transvestite, Tranny, Drag queen	Transsexual, Transgênero, Transição, Travesti Cara de traveco, Voz de traveco

Fig. 1. Keywords table (a fragment)

These keywords are used as key concepts in the search for each online platform to find publications that contain these keywords. After finding them, we start the extraction. Figure 2[1] shows the NetLang Architecture.

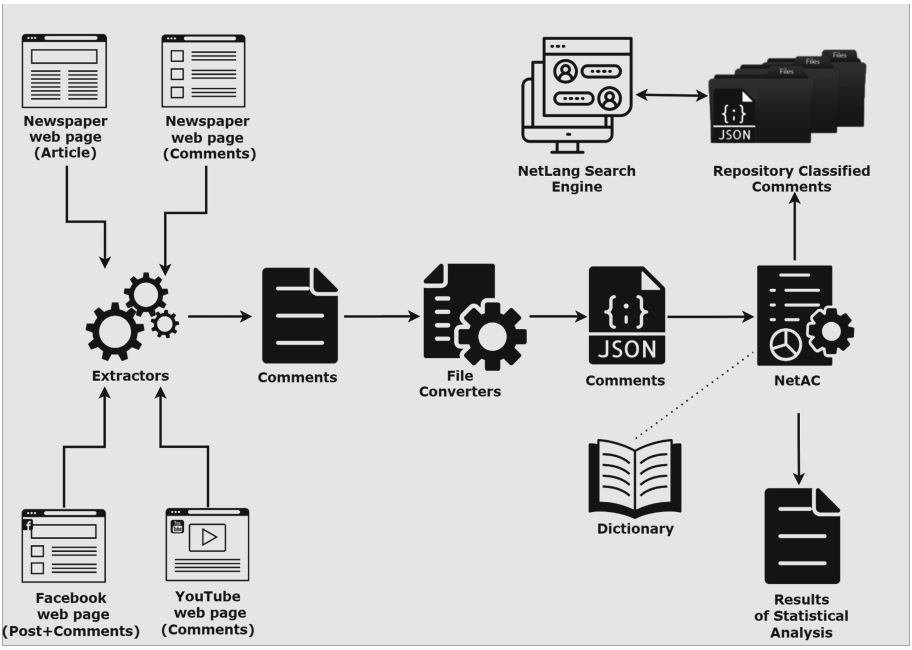


Fig. 2. NetLang architecture

Figure 2 presents, in a simplified way, all the processes from extraction to searching the NetLang platform. The `Extractors` will scrape web pages (Newspaper, YouTube, etc.) and as a result we obtain a file, which can be in a variety of formats, such as HTML, JSON, CSV, etc., depending on each extractor. Subsequently, the extracted files will be converted into a standard format using the `File Converters`. In this case, we choose JSON as the standard format. After the extracted files are converted to the standard format, they are analyzed by the `NetAC` tool. The `NetAC`'s main objective is to search for the frequency of occurrence of keywords in each category (based on the Keywords Table – `Dictionary`). Based on the highest frequency, it classifies each comment and the text in general. As a result of this analysis, a document is generated that presents the classification per comment and per input text (`Results of Statistical Analysis`); the original JSON document is also classified with the Sociolinguistic Variables and keywords found in the comments are annotated. These JSON documents

[1] All the icons of Figs. 2, 3 and 4 were taken from the website: "The Noun Project" (https://thenounproject.com/). Accessed: 2020-11-10.

are stored in a `Repository of Classified Comments` – Corpus[2] and can be searched, on the `NetLang Search Engine`, by Type of Prejudice and Sociolinguistic Variable, in Portuguese or English, as shortly stated above. According to the project proposal, the extracted corpus should be stored on to the NetLang Server and made accessible to the international researchers community through an Web Interface linked to the project homepage. In this context, this paper aims at presenting and discussing the `NetAC` (NetLang Analyzer and Classifier). In the following sections, we will explain all of these processes in more detail.

1.2 Paper Organization

Section 2 discusses the core of the NetAC for comments classification. Section 3 presents complementary features provided by the NetAC, such as: adding or removing keywords; Dictionary lookup; upload files to analyze, etc. Section 4 presents a case study that illustrates the functionalities offered by the NetAC. Finally, Sect. 5 summarizes the paper and its contributions and presents suggestions for future work.

2 NetAC

As previously mentioned, `NetAC` – NetLang Analyzer and Classifier has as main objective to analyze and classify files with comments (`Analyzer and Classifier`). A variant of the main module (`Analyzer and Classifier`) is to search the text for a specific sociolinguistic variables (`Analyzer by Sociolinguistic Variable`). This feature does not classify the JSON file, it only generates the results of the statistical analysis. However, in addition to this functionality, it also has other complementary features such as: add or remove keywords to the dictionary (`Update Dictionary`), consult the dictionary by language (`Consult the Dictionary`), upload JSON files for analysis (`Upload JSON Files`), add new languages in addition to Portuguese and English (`Add Language`). Figure 3 shows the general architecture of `NetAC`.

In the next section, the `Analyzer and Classifier` will be presented and discussed in detail. Complementary features will be presented in more detail in Sect. 3.

2.1 Analyzer and Classifier – Main Module

The `Analyzer and Classifier`, discussed along this section, is a chief component inside the `NetAC` introduced in the previous section as one of the main blocks belonging to the *NetLang Architecture* (Fig. 2). This component receives a list of comments (pre-processed after extraction and converted into JSON format) and

[2] Corpus is a set of linguistic data pertaining to the oral or written use of the language (example of speeches: debates in digital media, historical texts, etc.) and which can be processed by computer [9].

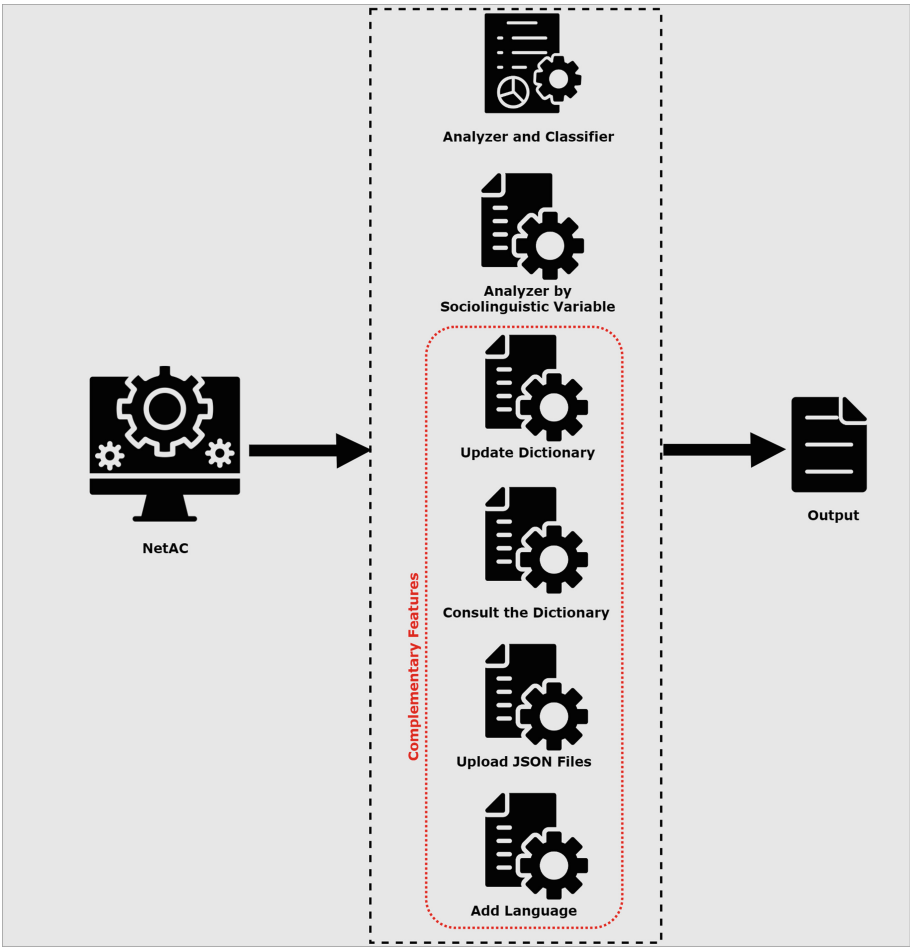


Fig. 3. `NetAC` architecture

receives the NetLang Keywords Table also introduced above and produces a list of comments classified in a Sociolinguistic category according to the frequency of occurrence of the keywords in the comments. For the classification purpose, the Keywords Table is split into different tables (one for each language) which will be called Dictionaries in the sequel.

The `Analyzer and Classifier` module is responsible for detecting the presence of keywords in the comment texts; there after the frequency of the keywords will be used to propose a classification taking into account the Sociolinguistic variable (Addiction Shaming, Racism, Sexism, and so on) to which each keyword found belongs (of course the type of prejudice can also be deducted from the Sociolinguistic variable).

Let's now take a closer look inside the engineering behind this sociolinguistic Analyzer and Classifier. Figure 4 details the architecture of that component. As it is shown, there are two different inputs, and two components that produce the result. This tool receives: the working language identifier to get the appropriate `Dictionary`; and a set of `JSON files` (or just a single one) to analyze. At the end of the analysis, a PDF file is created with the results obtained (`Results of Statistical Analysis`). This PDF file becomes then available to the user.

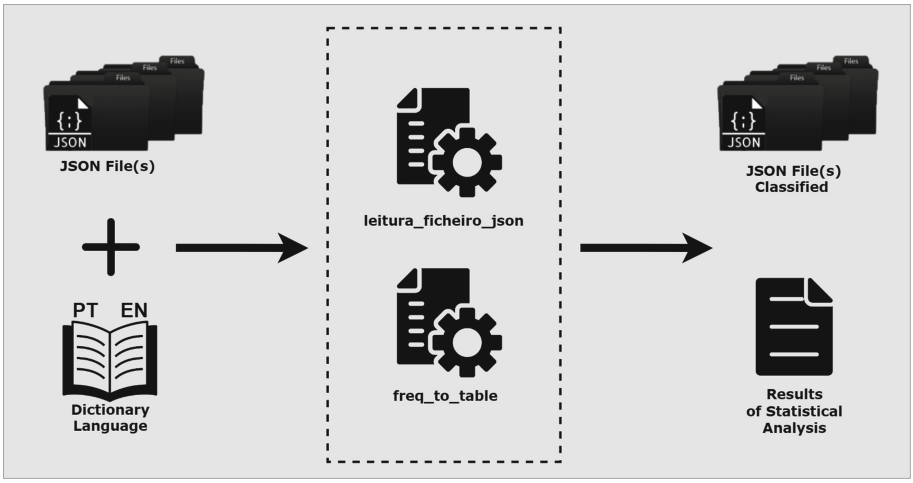


Fig. 4. `Analyzer and Classifier` Architecture

The core of the Analyzer and Classifier tool is the JSON file reading function (named as `leitura_ficheiro_json.py` in the architecture); that function is responsible for uploading the dictionary to be used, and the JSON file to classify. First of all, the comments from the JSON file are selected and then a pattern matching algorithm is used to find the occurrence of the dictionary words. For this, the Python module goes through the keywords in the dictionary to check, for each one, if they are present in the file. When a keyword is found, the relevant information (as the number of occurrences and the sociolinguistic variable to which it belongs) is saved to enable later the production of the results. This function is also responsible for calling another python script, the second component of this tool named as `freq_to_table.py`, which generates a PDF document with the following structure: a table containing the analysis made by comment; and another table with the analysis by sociolinguistic variable. In Sect. 4, this working process will be illustrated.

Another variant of the `Analyzer and Classifier` is to search for hate speech words associated with a specific sociolinguistic variable (`Analyzer by Sociolinguistic Variable`). This feature differs from the previous one because it receives a sociolinguistic variable as an input. This will define the keywords

that will be searched for throughout the text. At the end we obtain a list of the keywords found and an analysis of their frequency. In this variant the JSON file is not classified.

2.2 Dictionaries

As previously mentioned, dictionaries play a role of uttermost importance in that system. There is one for each language. The dictionary keys are tuples (*"type", "sociolinguistic variable"*). A sociolinguistic variable corresponds to a *hyperonym - hyponym* pair. Each key is associated to a list of keywords, as illustrated by the following example:

$$(Ageism, Age - General) \mapsto [Age, Ageing, Ageism, Ageist]$$

To guarantee the correct performance of the `Analyzer and Classifier` tool at every moment, the dictionaries are dynamic and can be updated at any time by the project management team.

So, when a new keyword is added, there is a script responsible for generating, in LaTeX, a table with the `Dictionary` new version. This table is accessible at any time on the web page and facilitates the visualization of keywords. Figure 5 shows a fragment of this table.

Types of Prejudice	Sociolinguistic variables (Hiper - Hipo)	Keywords
Addiction Shaming	Behavioural Addiction - Alcohol	Drunkie, alcohol
	Behavioural Addiction - Drugs	Junkie, Druggie, Drug addict
	Behavioural Addiction - General	
	Behavioural Addiction - Pornography	
	Behavioural Addiction - Sex	
Ageism	Age - General	Age, Ageing, Ageism, Ageist
	Age - Over 65s	Bed-blocker, Elder, Elderly, OAP, Old, Pensioner, Senior, Wrinklies
	Age - Youngsters	Brat, Half-pint, Insect, Lightweight, Morsel, Nobody, Nonentity, Nothing, Twerp, Whippersnapper, Zero, Zilch
Anti-Clericalism	Religious Identity - Christian	Fundie, Bible thumper, Bible beater, God botherer
	Religious Identity - General	
	Religious Identity - Jewish	Abbie, Abe, Abie, Christ-killer, Ikey, Ike, Iky, Kike, Heeb, Hebe, Hymie, Ikey-mo, Ikeymo, Mocky, moky, moxy, mockey, mockie, mocky, Oven, dodger, Sheeny, Shapeshifter, Shlomo, Shylock, Yid
	Religious Identity - Muslim	Asslifter, Carpet kisser, Goatfucker, Kebab, Koran basher, Koranimal, Mussie, Quran, Koran thumper, Hajji, Haji, Hodgie, Mudslum, Muslimtard, Musloid, Muzrat, Pisslamist, Terrorist, Raghead, Towelhead
	Religious Identity - Protestant	Bible basher, Holly roller, Orangie, Snout, Prod, Proddy dog, Proddywoddy, Proddywhoddy, Russellite, Spike, Soup taker
	Religious Identity - Roman Catholic	Creeping Jesus, Left-footer, Mackerel Snapper, Mick, Papist, Taig
	Religious Identity - Sikh	Raghead, Towelhead
Body Shaming	Physical Identity - General	Body Shaming
	Physical Identity - Physical (and Mental) Impairments	Ableism, Eugenics, Blind, Deaf, Dumb, Stutterer, Dwarf, Hunchback, Lame, Limp, Midget, Cripple, Disabled, Freak, Handicapped, Wheelchair bound, Bonkers, Crazy, Dim, Feeble-minded, Lunatic, Loony, Mental, Mong, Mongoloid, Nutter, Psycho, Retard, Spazz, Spazzie, Spastic, Weird, W... Window licker
	Physical Identity - Physical Features	Fat, Fatso, Fatty, Butterball, Chubby, Lardass, Lump, Obese, Porker, Thin, Bag-of-bones, Beanpole, Beanstalk, Bony, Skinny, Skin-and-bones, Scr... Scraggy, Twiggy, Double eyes, Double chin, Bald

Fig. 5. Dictionary used by the analyzer and classifier for the English version (a fragment)

The next section presents some features, already referred, that were added to the project in order to make it more proactive and versatile.

3 Complementary Features

After the description of the `Analyzer and Classifier` main module, the present section is devoted to introduce the complementary modules that provide features to maintain the information system updated and flexible.

3.1 Adding/Removing New Keywords

In order to facilitate the utilization of our tool, we decided to implement the possibility of adding and removing new Keywords in already existing dictionaries by accessing an area reserved for administrators that requires a password. Taking this into account, we created a module that receives the action intended (be it adding or removing a keyword), data about the keyword (Type of Prejudice and Sociolinguistic Variable), and lastly the module adds the keyword in the intended place in the dictionary, or removes it entirely.

3.2 Dictionary Consulting

With the goal of facilitating the data analysis, it was implemented the possibility of consulting every dictionary available. This functionality, despite its easy implementation, will greatly help the users with the analysis of the data that our tool generated. This feature lets everyone consult the keywords registered as hate speech and also lets everyone verify, for example, if there are any words wrongfully added with the help of the previous spoken feature.

3.3 Adding New Languages

Taking into account the implementation of the previous feature, the next step is to allow more languages to be added to the tool, since originally it was only supposed to work with the Portuguese and English languages. In order to do so, a function which its input is a string that represents the language the user wants to add, was created. But before adding this new language, the function checks to see if the language is already registered on the platform. If it isn't, we create another dictionary, respective to that language. But if it is already on the platform, there's no need to create a dictionary, because there already is one, so the function stops and doesn't do anything else. This new dictionary can be filled with new keywords using the Adding New Keywords feature, previously talked about.

3.4 Uploading Files

With the intent of, once again, facilitating the usage of this tool, we implemented a feature that helps the user analyze the JSON files he wants. For this, we created a module that lets the user upload files, directly from their computer, to the server, since the tool can only analyze files that are in a certain directory in the server.

4 A Case Study

Consider a case in which the End-user wants to study the hate speech emanating from a set of short sentences (comments) collected from some social network. The purpose of such study is to understand what kind of hate speech is the most prominent in that conversation aiming at fighting against that overwhelming problem. Our tool helps dealing with this problem by analyzing the referred document (although it must be converted firstly into a JSON file).

To analyze that file[3] it is only necessary to access the tool at: http://netlang-corpus.ilch.uminho.pt:10100/. The Application homepage shown in Fig. 6 exhibits the menu with the features available.

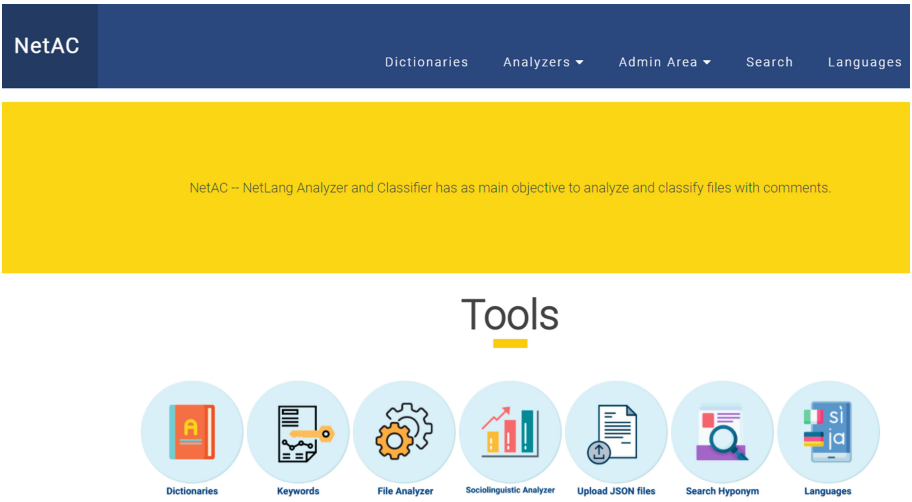


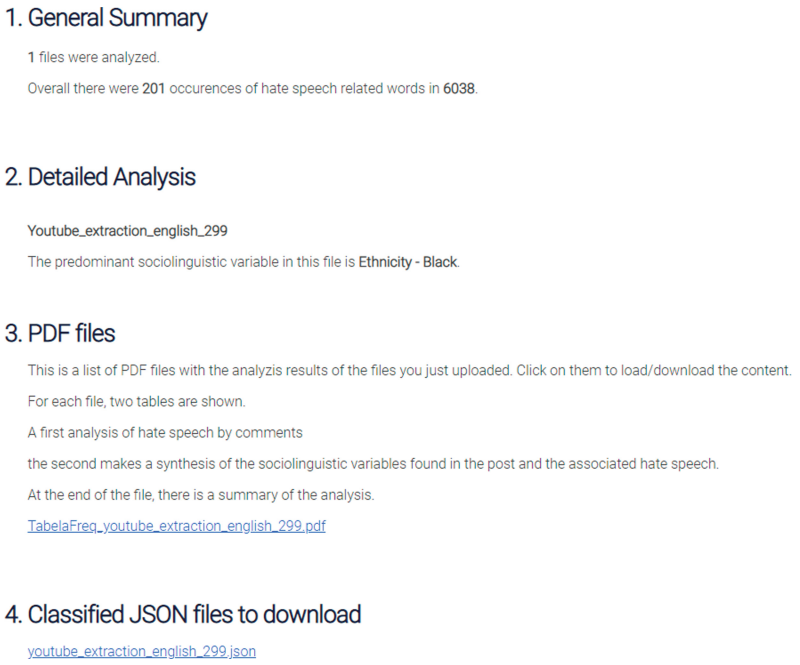
Fig. 6. NetAC – homepage

First of all the user needs to select the appropriate menu option to upload the desired file. Then the user can either choose to do a full analysis, computing the occurrences of all the keywords in a kind non-oriented approach, or a specific one, oriented to a sociolinguistic variable.

After process the file, a brief summary of the analysis (see Fig. 7) and a PDF file (see Figs. 8 and 9) are generated and can be downloaded for further studies. Moreover, it is also possible to download the original JSON file with the comment fields enriched with annotations derived from keywords found during the analysis.

At the end of the document, a detailed summary of the analysis performed is presented, as shown in (Fig. 7). This summary contains the total number of

[3] Notice that the User can process more than one JSON files at the same time if he wishes to.



1. General Summary

1 files were analyzed.

Overall there were **201** occurences of hate speech related words in **6038**.

2. Detailed Analysis

Youtube_extraction_english_299

The predominant sociolinguistic variable in this file is **Ethnicity - Black**.

3. PDF files

This is a list of PDF files with the analyzis results of the files you just uploaded. Click on them to load/download the content.

For each file, two tables are shown.

A first analysis of hate speech by comments

the second makes a synthesis of the sociolinguistic variables found in the post and the associated hate speech.

At the end of the file, there is a summary of the analysis.

TabelaFreq_youtube_extraction_english_299.pdf

4. Classified JSON files to download

youtube_extraction_english_299.json

Fig. 7. System output: analysis summary for a given file

Comment	KeyWords	Sociolinguistic variables (Hiper - Hipo)	Hate Speech Frequency	Hate Speech Frequency(%)
All **racism** is evil! That includes The Root when you publish articles like 'the 10 laziest types of white people ranked' **black** and white supremacists will share the same corner of hell, while all of us good people will party together regardless of **race** laughing at your **racist** asses.	Black, Race, Racism, Racist	Ethnicity - Black, Ethnicity - General	4/49	8.163
I wonder what a **black** America would look like? Oh yeah probably just like every other **black** 3rd world shit hole country.	Black	Ethnicity - Black	2/22	9.091
You **racist** scum are all the same, just with different colors. Learn how to understand equality. This is one of the reasons why I hate YouTube. Allowing such deplorables to make heinous content, but always turning a **blind** eye to that and mess with content creators who haven't been anything wrong.	Blind, Racist	Ethnicity - General, Physical Identity - Physical (and Mental) Impairments	2/51	3.922
Amusing how you idiots on "The Root"are incapable of going after the actual white supremacists. White people are SO **RACIST** in your eyes. I feel sorry for anyone who takes you seriously. **Black** supremacists like you love to be retarded, I guess.	Black, Racist	Ethnicity - Black, Ethnicity - General	2/43	4.651

Fig. 8. System output: the PDF document – a summary of the results per comment

offensive words detected, the percentage of hate speech present in the document, the sociolinguistic variables associated with the keywords found and the identification of the predominant variable.

Sociolinguistic variables (Hiper - Hipo)	KeyWords	Number of occurrences	Frequency	Frequency(%)
Ethnicity - General	Race, Racism, Racist	52	52/6038	0.86
Ethnicity - Black	Black, Niggah, African	97	97/6038	1.6099999999999999
Physical Identity - Physical (and Mental) Impairments	Blind, Dumb, Crazy, Mental	8	8/6038	0.13
Religious Identity - Muslim	Terrorist	2	2/6038	0.03
Physical Identity - Physical Features	Fatty, Lardass, Fat, Thin	7	7/6038	0.12
Age - Youngsters	Nothing, Nobody, Zero	11	11/6038	0.18
Gender - General	Gender, Woman	2	2/6038	0.03
Age - Over 65s	Old	2	2/6038	0.03
Gender - Female age and physical appearance	Witch	9	9/6038	0.15
Ideological and Political Identity - General	Nazi	3	3/6038	0.05
Ethnicity - Asian (East- China, Japan, Korea, Philippines, Vietnam)	Yellow	1	1/6038	0.02
Nationality - Chinese	Yellow	1	1/6038	0.02
Nationality - Japanese	Yellow	1	1/6038	0.02
Nationality - General	Nation	5	5/6038	0.08

Fig. 9. System output: the PDF document – a summary of the results per sociolinguistic variable

5 Conclusion

Along the paper, we introduced a Web application called NetAC that works on a set of short texts, extracted from online social networks or online newspapers, that comments on toxic posts or news articles and analyzes the hate speech springing out from those comments.

The tool's purpose is to classify each comment and the overall document into a predefined category of Socially Inappropriate Discourse (Hate or Aggressive speech) according to the most used set of keywords.

The system is accessible at http://netlang-corpus.ilch.uminho.pt:10100/ and is being tested successfully with the NetLang Corpus extracted from Youtube and popular English and Portuguese Newspapers. After the analysis, the results can be consulted in the PDF documents output or searched by Sociolinguistic Variables using the NetLang Search Engine, also accessible via the same Web interface. It is worthwhile to emphasize that the source language can be easily adapted to other languages (as long as they are based on the our alphabet). For that is enough to create a new dictionary for that languages keywords, and than set it up as the working one.

At present we are designing an experiment to be conducted with a Team of Linguists and Social Science Researches for validating the automatic classification proposed by our tool.

In terms of future work, a significant tool improvement can be done. Using artificial intelligence approaches to refine the linguistic analysis of the comments, some flaws of the present implementation, like the problem of false positive identification, can be overcome.

Acknowledgment. This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope: PTDC/LLT-LIN/29304/2017.

References

1. Anderson, A.A., Huntington, H.E.: Social media, science, and attack discourse: how twitter discussions of climate change use sarcasm and incivility. Sci. Commun. **39**(5), 598–620 (2017)

2. Blom, R., Carpenter, S., Bowe, B.J., Lange, R.: Frequent contributors within U.S. newspaper comment forums: an examination of their civility and information value. Am. Behav. Sci. **58**(10), 1314–1328 (2014)
3. D'Errico, F., Poggi, I., Corriero, R.: Aggressive language and insults in digital political participation. WBC **2014**, 105–114 (2014)
4. Ermida, I.: (Re)Defining Hate Speech. Among other Forms of "Socially Inappropriate Discourse". Unpublished Presentation (2020)
5. Fortuna, P., Nunes, S.: A Survey on Automatic Detection of Hate Speech in Text
6. Henriques, P.R., Araújo, C., Ermida, I., Dias, I.: Scraping news sites and social networks for prejudice term analysis. In: Weghorn, H., Rodrigues, L. (eds.) Proceedings of the 16th International Conference on APPLIED COMPUTING 2019, Cagliari, Italy, pp. 179–189, November 2019
7. König, L., Jucks, R.: Hot topics in science communication: aggressive language decreases trustworthiness and credibility in scientific debates. Public Underst. Sci. **28**(4), 401–416 (2019)
8. Lapidot-Lefler, N., Barak, A.: Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. Comput. Hum. Behav. **28**(2), 434–443 (2012)
9. Mendes, A.: Linguística de corpus e outros usos dos corpora em linguística. Manual de linguística portuguesa **16**, 224 (2016)
10. Moor, P.J., Heuvelman, A., Verleur, R.: Flaming on YouTube. Comput. Hum. Behav. **26**(6), 1536–1546 (2010)
11. Nielsen, L.B.: Subtle, pervasive, harmful: racist and sexist remarks in public as hate speech. J. Soc. Issues **58**(2), 265–280 (2002)
12. Nockleby, J.T.: Encyclopedia of the American Constitution, 2nd edn. Macmillan Reference, New York (2000). ch. Hate Speech
13. Pfeffer, J., Zorbach, T., Carley, K.M.: Understanding online firestorms: negative word-of-mouth dynamics in social media networks. J. Market. Commun. **20**(1–2), 117–128 (2014)
14. Rowe, I.: Civility 2.0: a comparative analysis of incivility in online political discussion. Inform. Commun. Soc. **18**(2), 121–138 (2015)
15. Rösner, L., KrÄmer, N.C.: Verbal venting in the social web: effects of anonymity and group norms on aggressive language use in online comments. Soc. Media Society **2**(3), 1–13 (2016)