

Every 60 seconds, over a million Facebook logins occur and over 80,000 tweets are posted. The vast amount of information gathered from this activity serves many business entities in their ongoing decision making, helping them reach more customers and improve their service to existing clientele.

As part of building the capabilities of social organizations to leverage data, we have chosen to delve into, study and disseminate knowledge concerning analysis of the discourse on social media networks (network research). The aim is to explore: How could analysis of such discourse serve social purposes? How is it conducted? Using what tools? And many more questions.

If we know how to analyze the vast information found online, we can learn about:

- Actions that should be taken to advance our social goals.
- Changes in the social discourse following actions we take.

Social media contents can supplement insights based on surveys and provide more information about trends, attitudes, behaviors and feelings. However, when taking this path, consideration should be given to its limitations, such as reliability of the information and its analysis while privacy is maintained, and all activities are in accordance with the law.

Characteristics of Social Media Content:

Before we get to the data analysis, the following are a few characteristics of contents found on social networks:

- **Scope:** The analysis of information from social networks enables gathering large amounts of collective information. Such information may not have been available otherwise or it might have taken a great deal of effort to gather it.
- **Spontaneity:** The content is usually spontaneous and frequent; thus, the information is up-to-date, dynamic and fast-flowing.
- **Masking:** Writing messages anonymously, with the writer concealed behind screen and keyboard, allows for the expression of positions that may be extreme and do not necessarily provide an accurate reflection.
- **Location:** For many messages, geographic analysis is relevant. Content on social media often contains location scores and can provide a solution to this issue. However, some users choose not to share information about their location resulting in possibly biased analysis.

On the difference between online discourse analysis and surveys

Surveys and network analysis are tools often used concurrently. Surveys based on a representative sample have an advantage in terms of data reliability, as samples are an approximate representation of the population that the survey aims to examine. Conversely, the discourse on networks is shaped only by those who choose to respond, so by definition this it is more extreme. Also, the discourse on social networks is continuous, open and allows for multiple opinions and positions, which a survey with pre-defined questions might miss.

We found two main uses for analyzing discourse on social networks:

Detection of Distress

Real-time discourse analysis can alert of people in a state of mental distress and allow for the rapid intervention of expert entities.

“Sahar” and “Eran” NGOs employ “Makam” technology for the purpose of real-time detection of people in distress.

According to web mapping of predefined search keywords, the software floats relevant posts from websites, blogs, Twitter and Facebook (open groups and profiles only).

Followed alert based retrieval, the posts are forwarded to the NGOs that work with content sites to reach the relevant person and offer them first aid and referral to relevant professional entities.

Operating NGOs



Technology provider



Position Analysis

Analyzing a discourse of attitudes toward a particular population or social issue can help a social organization fine-tune the messages it wishes to convey, measure the impact of the activities it performs, and make decisions about required actions - all in accordance with the issues that arise in the online discourse.

Berl Katznelson Center’s “Hate Report”, in collaboration with Vigo Research Firm, monitors more than half a million posts at any given time.

The report’s purpose is to provide a reflection indicative of the level of violent discourse online and of expressions of hatred and incitement. In addition, the center publishes reports revolving around certain events or figures, such as, in the run-up to elections, hate speech against the Druze community and the like.

<http://hasata.berl.org.il/>

Operating NGOs



Technology provider



How is discourse analysis actually performed?

Analysis of discourse on social networks consists of two tasks:

1. Collection of all the information we want to analyze, found on social networks (for example: posts, talkbacks, tweets, forums, videos on YouTube).
2. Analysis of gathered information by means of quantitative or qualitative research.

Prior to conducting the analysis, it must be determined whether we want to work independently, using free tools, or enlist the help of an external company, for a fee.

Quantitative Research versus Qualitative Research on Social Networks

1. Quantitative research begins with determining a list of keywords relevant to a specific research topic. Then a query is defined that automatically checks the frequency in which these keywords appear in the determined digital space per category. The research is automatically performed by computerized collection and analysis tools.
2. Qualitative research involves the work of analysts, who read in-depth and classify selected contents on a sample basis. Analysis of the space in which the discourse is conducted is used to build a sample and define the research questions. The analyst's work includes cataloging and classifying the post according to various categories.

	Quantitative	Qualitative
Type of Analysis	Keyword based queries	Sampling and analysts' work, including reading posts and cataloging
How up-to-date is the information	The research is conducted by an automated tool so the data can be updated continuously, even on the same day.	The research requires analysts' work and will therefore be conducted less frequently and will be less up-to-date in respect to a given point in time.
Accuracy / quality level	Accuracy level is mediocre and vulnerable to biases. The analysis is based on keyword search, but it is possible that within a given context their meaning is different. Sometimes the keyword pool must be reduced due to certain biases (for example, a word that has several meanings will be removed).	Information analysis level is high because the analysis considers the context as characterized by a human being. However, qualitative research might have difficulty reflecting the overall online discourse as it is based on sample, which is often small.
Dynamism	Quantitative research can be updated relatively easily by adding to / modifying the pool of sampled keywords.	Modifying qualitative research requires more significant work, at times to the point of forming a new research.
Cost	Low	High

Free tools for discourse analysis

There are free or relatively low-cost tools on the market that enable independent analysis of online discourse. Sometimes, the collection tools and analysis tools are separate, and in many cases the collection tools themselves are separate for different sources of data (a tool for collection from Facebook, another tool for collection from Twitter, etc.).

The major social networks provide built-in analysis tools, such as: Facebook Insights, Twitter Analytics, Instagram Insights, but these analyze the performance of posts / campaigns and do not monitor the discourse itself.



Comment
Exporter.com

A tool that enables exporting comments to a post on Facebook, Instagram or Twitter. Once the post URL is entered on the website, its comments are exported to Excel.



mention

A tool that collects and analyzes data from Instagram, Facebook, websites, blogs and the news. Offers not only online discourse monitoring, but also tools to increase the organization's responsiveness to what is happening on examined network, for example, by identifying network influencers.



awario

A tool with similar capabilities to those of "Mention" with the addition of a user-friendly interface, including: a virtual training center, analysis management in dedicated folders and analysis-based report generation at the click of a button.

Points to consider regarding use of free tools:

- Operating the free tools is not always simple and sometimes requires a technical-technological orientation.
- The market is flooded with software programs that specialize in this field, many of which focus on one capability (collection or analysis) or one type of network (Twitter, Facebook, etc.). Usually several tools would be needed to get a complete and comprehensive picture.
- There are tools with a free version and there are software programs that have various capabilities and a different pricing model.
- Social network - first and foremost Facebook - frequently change their permissions model, and that affects the relevance of different solutions. This is one of the reasons why this market is very dynamic and the tools available today may be obsolete before long.
- Automation tools, such as Zapier, can be used to perform repetitive collection and analysis

Working with expert suppliers

In case ability and/or possibility to perform the analysis independently are lacking, the services of providers who specialize in the field can be enlisted.

Such providers operate primarily in the business sector, assisting in analyzing various marketing questions related to brand status and crisis management.

In the social sector, quite a few studies have been conducted by Vigo, Buzzilla and Makam.

With the help of Vigo, Hamleh NGO analyzes violent / inciting speech against Arabs and Palestinians posted in Hebrew. The research monitors events, people (politicians and celebrities) as well as revolving around unique factors (such as the law of nationality). It also divides the discourse into categories, to allow for long-term follow-up. Reports generated by the research make it possible to monitor a variety of phenomena classified as different types of violence and over time.

The research, which is based on keywords, has gone almost unchanged over the years, to allow comparison over time.

Technology provider



Operating NGOs



How much does it cost?*

The cost of the research depends on whether it is qualitative or quantitative, as well as its scope and sampling type.

Qualitative research requires analysts to read texts and catalog them. The larger the scope of texts and the more challenging it is to comprehend them (e.g., analyzing a discourse on attitudes toward vaccines), the higher the cost of research will be.

From the data we have, it appears that qualitative research in the social world currently costs about NIS 15,000.

Quantitative research is cheaper, as it is based on automatic tools and can be obtained for a retainer of a few thousand NIS a month.



An application based on discourse analysis technologies

4Girls' MoodKnight is an artificial intelligence platform, based on machine learning and natural language processing (NLP), for the detection of mental distress and suicidal tendencies in digital content. 4Girls began as a social network dedicated to girls. Based on the knowledge accumulated by it, MoodKnight was developed, with the aim to detect in contents on the network situations of distress, risk and emergency. The system makes it possible to identify not only that it is a risk situation, but also to classify it according to its type of distress and level of severity.

What is unique to 4Girls is that distress is detected not only by keywords or combinations of words, but by various aspects of text style determined to be relevant. For example, the word 'I' when disproportionately repeated in the text, may be a sign of distress.

To date, the system operates on two large content sites. Additionally, the process of its implementation in a parental control app that monitors the behavior of children and adolescents on social networks is currently underway.

* Costs stated above are based on information from NGOs and service providers, and are to be considered merely as an estimate.

What are the limitations of discourse analysis?

There are several built-in limitations of analyzing social media discourse that affect the use of the tool:

- **Extreme discourse:** Discourse on social networks is extreme by its very nature; not only in terms of the way people express themselves, but also in the type of content they seek to upload to the networks. Often, it is extreme cases and intense emotions that are posted online.
- **Definition of sample and sampling scope:** The question of which groups and pages to collect information from is critical. It is important that they are selected in accordance with the research questions.
- **Privacy Restrictions:** Privacy restrictions prohibit access to and analysis of private profiles.

It is important that each study begins with an initial examination: Where does the discourse take place? In what pages / groups? What is the scope of the discourse? Following the initial examination, the type of research should be adjusted to the required sampling scope, needed percentage of manual sorting and the depth of analysis required.

Network research (discourse analysis on social networks) is a wide world of possibilities and gives authentic expression, free of any prior directing (surveys) to public notions not adequately expressed in public media.

There is a consensus that network research (discourse analysis) may shed further light beyond that obtained from surveys and their analysis. Network research could provide answers to questions such as: What expressions are used to describe a certain phenomenon? and deduction can be made as to how to best convey messages. Such research is better suited to answer open-ended questions. However, the reliability of such research is limited and therefore proper definition of the research question is of immense importance. Generally, the recommendation is to build a combined study comprised of network research and surveys and/or even analysis of media coverage in traditional public media, if needed. This way a more comprehensive picture can be obtained and congruency or gaps between public opinion, online discourse and media coverage on a given topic or towards a certain group can be examined.

Thanks to the NGOs who have devoted time to contribute to this research:

Eran, Hamleh, The Association for LGBTQ Equality in Israel, Sahar, 4Girls, Berl Katznelson Center and Panim Organization.

Thanks to the technology providers:

Vigo, Buzzilla and Makam for letting us into their professional realm and sharing their knowledge with us.

A special thanks to The Jewish Federation of New York (UJA) for its full support and funding of this research.