

Information Consolidation on Users of Social Networks to Determine Their Credibility

Oleksandr Markovets^a, Solomiia Albota^a and Oksana Horpyniuk^b

^a Lviv Polytechnic National University, Bandery Str. 12, 79013 Lviv, Ukraine

^b Lviv State University of Internal Affairs, Gorodotskaya Str. 26, 79007 Lviv, Ukraine

Abstract

This work is aimed at defining general concepts about the features of Internet users, the relevance of the research topic within the analysis of social networks and the main objectives to be achieved in the process of getting results. For determining social networking user credibility, data flow diagram has been provided. Figures of decomposition of different levels have been introduced in order to describe social networking processes within Facebook. As a means of characterizing data models, entity-relationship model has been provided to define the credibility of social network users as well. The obtained study results of the Facebook user credibility indicators have been illustrated in the table with users' rating. A combination of linguistic and statistical approaches is of importance regarding social networking user credibility as well as good quality and reliability results.

Keywords ¹

SMM, social network, promotion, consolidation, credibility, community, model, indicators, rating

1. Introduction

Social networks serve to be a means of destroying communication barriers, so that even strangers can find common ground. The process of ruining these barriers is also due to the openness of the information exchanged by users. It makes it easier for different companies to find and influence the appropriate audience for further collaboration.

The relevance of the topic is substantiated by the necessity to increase the range of development for organizations of different levels beyond routine life, in other words, in using social networks to improve awareness of the company about its customers and their interest in the company's activities.

The research object is the official page of the company on the Facebook social network.

The study subject - the level of interest and usefulness of users of the Facebook social network from the point of view of the company.

2. Studies referring to the issue

Web 2.0 is the result of constant current progress, logical improvement (and not just new trends) of a new stage of evolution in the Internet [1-3], which has not appeared suddenly, replacing outdated sites.

A web service is a software system designed to communicate between computers over the Internet. There are a lot of web services today, however, in the terms of Web 3.0, they require more attention. By

COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine
EMAIL: oleksandr.v.markovets@lpnu.ua (O. Markovets); Solomiia.M.Albota@lpnu.ua (S. Albota); Olepisevych@gmail.com (O. Horpyniuk)
ORCID: 0000-0001-8737-5929 (O. Markovets); 0000-0002-0845-5304 (S. Albota); 0000-0003-3110-6564 (O. Horpyniuk)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

combining semantic marking and web services, Web 3.0 increases the capabilities of applications, which, in turn, can provide more information retrieval through simple interfaces.

Along with the development of information technology, the Internet is constantly evolving. The modern version of the Internet is the so-called Web 3.0, which is a stage of "reading-writing/recording-performing" [4,5]. It is necessary to allocate two points which actually will be a basis of a semantic marking of Web 3.0 web services. Content marking refers to the process of communication between an Internet user, i.e., a person, and computer programs.

One of the most significant organizational problems of information representation on the Internet is the lack of providing context for data, which makes it difficult to determine the relevance and irrelevance of information, by web applications. The term performance refers to the formatting of data so that they are "clear" to software agents, web services [6].

SMM is a powerful tool for working with the company's image, increasing the qualitative and quantitative totality of all information about it, and the loyalty of target groups. This is not an alternative to search engine optimization (SEO) or contextual advertising, but can also give good results when used with other tools in combination.

Social Media Marketing (SMM) can provide the following opportunities:

- creating and promoting operative information source for users of social networks and potential clients;
- promoting sales;
- attracting the attention of a large number of customers to the proposed brand or topic to increase the interest of the audience, regardless of its age, gender, interests, status, education, etc.;
- receiving feedback from the target audience and direct communication with potential customers to increase customer loyalty;
- reducing the amount of negative information about the brand or topic (or its complete neutralization) in various Internet resources;
- increasing the quality of traffic for the corporate site.

Social network - a social structure of individuals or organizations, designed to reflect the various relationships between them through different levels of social relationships (from casual acquaintances to close family ties). A social network is an interactive, dynamic website created by network members.

In general, social network is a community of people connected by common interests, common cause, or people with other purposes to communicate with each other [7-9].

The social network on the Internet is a software service, a platform for interaction of people in a group (groups). Theoretically, a social network can include any online community whose members participate, for example, in joint discussions. The social network is also formed by readers of the thematic community based on any blog service.

The main feature of the new generation of services is the tools to find the right contacts and establish connections between people. Using the tools of social network, each user can create their own virtual portrait – to create a profile that will provide information about the person, his or her work experience, hobbies, interests and goals. Having a profile allows to use search mechanisms for like-minded people, fellow believers, colleagues, people with whom one need to communicate at work or at school.

Modern social network provides the following set of standard services:

- saved personal card with contact details;
- online address book;
- organizer, which is accessible from any computer;
- storage of user multimedia data;
- the ability to restrict communication with unwanted users, etc.

Among others, mostly impersonal forms of online communication, where people often use fictional nicknames and avatars, the feature of most social networks is the maximum reality of communication for the Internet. That is, in social networks, people are under their real names and show their real photos. At the same time, their communication is established with the same real people.

There are the following types of social networks:

- integrated networks;
- ego networks;
- incomplete networks.

Integrated networks are special social networks with established limits of permitted relationships between people. For instance, such networks used in online research are: e-mail, mailing lists, social networks of the Internet space (MySpace), people working in one office, who are connected by local networks. Integrated networks are the most frequently used objects in the analysis of social networks. However, these objects may not always be appropriate. They may vary depending on the research objectives set before; collecting information about, for example, an office network is not a particularly difficult task, but obtaining a valid list of all users of the MySpace network is virtually impossible, as this list will change during data collection. In the process of studying integrated networks, the researcher must determine the data on the structure of the group, specific types of network users, the most popular users. Currently, electronic methods allow to collect only rather moderate data from integral networks.

Ego networks are usually represented by a spontaneous sample of users. Moreover, the researcher always strives for the ostentation of such a sample. Each element of the sample in such an analysis is referred to as "ego", and nodes associated with "ego" are referred to as "others". The researcher can collect data on star-shaped network (ego-node and its connection with other nodes), or a complete ego-network (which includes the connection of other nodes with each other) [10-13].

Incomplete networks are, in a certain way, a sample of relative data created by the snowball method. Incomplete networks are the medium between a covered whole network and the fact that some whole networks are just too massive to cover them completely. The researcher can start the analysis from a single web page or several pages (so-called "sowing"), then go on to search for pages related to this "sowing", and other pages, which, in turn, are related to these pages. The process of organizing the sample ends when either a sufficient number of pages are collected, or all possible pages are collected, or when the sample meets certain criteria [14].

There are three types of connections on social networks.

The first type is friendly relations between acquaintances. People who know each other in real life add each other as friends, and this information is displayed in their profile. Thus, for any person-node of the network from the total faceless mass of other nodes of the network, there is a group of nodes along with which he or she receives additional software communication. This gives him or her new opportunities, the most important of which for this topic – the ability to send messages and mass inviting friends to the group.

In addition to the list of his or her direct friends, any user also has access to the lists of friends of his or her friends, which leads to the ability to see network nodes that are not directly related to him or her, but are associated with his or her friends.

The next type - the connections between group members. In virtual social networks, human nodes can not only communicate with each other privately, but also unite in groups on any topic. A separate profile similar to the user profile is created for the group. When people join into a group, all members of the group see the message of one group member in their group profile. Thus, the profile of the group becomes some analogue of the Internet forum.

Another type - connections between people who have particular information in the profile (strangers). The unification of social network profiles makes it a huge database of people with a large variety of information about them, and this information is strictly structured. This gives a great opportunity to search the social network for an interested person to find people with certain data in the profile, and such data act as search criteria.

Types of social networks:

- general format social networks. Such networks were created primarily to achieve communicative goals - that is, to communicate with friends, relatives, acquaintances and colleagues. The purpose of such social networks is to give people the opportunity to share their photos, videos, music and more (online "VK", "Moi Mir", "Facebook", "Bebo");

- professional social networks. They are created to find employees or work, establish working contacts with representatives of other companies, post corporate news or press releases ("LinkedIn" and "XING");
- social networks by interests. They are designed to communicate with like-minded people who share common interests or preferences concerning a particular activity ("Last.fm" and "MySpace").

3. The research purpose and objectives

The purpose of the study is to determine the level of interest of the target audience by processing the information interaction with the company and expand the impact on interested Internet users.

To achieve this goal, it is necessary to solve the following tasks:

- to reflect the model of the network society;
- to establish the level of interest of the target audience in the interaction within the network;
- to analyze the pages of groups created by different companies on the Facebook social network;
- to provide suggestions for improving the functioning of the company's official Facebook page.

Increasing the level of popularity in social networks has one feature that is both a plus and a minus. SMM is an advertisement that is represented by customer feedback. Such reviews can be both positive and negative, and attempts to influence them can be even more detrimental to the company's reputation and site. There are frequent situations when companies hire people who would make positive posts in the comments about a particular product or service, advising everyone to use this particular service or buy that particular product. Despite this, such advertisement has the highest degree of trust among users. That is, if a well-known blogger, a friend from a social network or just someone on a forum posts a positive review about a service or company, the average user will treat this message as valuable information, not advertisement.

4. Problem statement

There are two groups of promotion methods in social media. These are the so-called "gray methods", i.e., illegal, and "white methods", i.e., legal. Let us focus on the second group of methods. These include the following types:

1. creating a brand community;
2. using targeted advertising;
3. quality content, various promotions, competitions.

By creating a group, one can promote completely different objects:

- place: country, city or individual landmark. Such PR can be engaged, for example, in travel companies;
- event: for instance, a specific mass event. This way, groups involving music festivals, concerts, city days, etc. are created;
- music bands. Beginners and not so well-known groups use this PR opportunity actively (although this method is quite common among well-known bands that want to communicate more closely with all fans in general). And such PR is really effective;
- political parties and individual politicians.

PR in social networks can be summarized as the use of social networking services to promote and sell a product or service, as well as to obtain important information, by involving an existing audience in the communication process regarding the product or service.

The main forms of PR-activities in social services include personal PR, which is implemented through the creation of personal profile on social networks, PR through the creation of virtual communities and PR through the distribution of invitations to events.

The official representative of one of the SEO-companies, identified 4 benefits of marketing on social networks:

1. Social media is an ideal platform for brand development and establishing long-term relationships with regular customers. Due to the indexing of social network profiles by search engines, the brand's presence in the network is growing. And using groups of social networks, one can create additional space for direct communication and interaction with the target audience (customers);
2. profile in a social network acts as a so-called "filter", which attracts the necessary target audience to the official website or blog of the company. Increasing site traffic by people leads to increased opportunities for advertisers to make the brand more popular;
3. An increase in the number of users who have shared information from the advertiser's site leads to an increase in the significance of the site in terms of search engines. And regarding the SEO, such social signals significantly contribute to the promotion (increasing the position of the site in the ranking results). From a marketing point of view, the greater the number of users who share information from the site, the more potential customers will see it;
4. PR. Social media has become a revolutionary means of exchanging information. Thanks to social networks, brands have a chance to tell their audiences and media about themselves in real time. One can contact journalists directly by giving them quick access to materials on their page or in private messages.

A detailed list of problems that arise for network analysts due to the lack of automatic data collection includes [15-19]:

- data privacy – often access to user data is allowed only for registered and authorized network members, which requires support for emulating a user session using specific records (accounts);
- poor data structure – in many cases, the program interfaces (APIs) of social networks have limited functionality, which requires support for obtaining, using a user interface, static copies of HTML-pages, correct processing of their dynamic part (including execution of asynchronous requests to social network server), retrieving the required data using an algorithm and / or template and building their structured view, convenient for further automatic processing;
- restrictions on access and blocking – in order to prevent unauthorized automatic data collection and limit the load on the social network service infrastructure, service owners often impose explicit or implicit restrictions on the number of requests from a single user account and / or IP address per unit of time required for the number of requests to which one refers, as well as support for the dynamic rotation of accounts and IP addresses used to collect user data;
- the dimensionality of data is due to the need for a parallel method of data collection, as well as methods for obtaining a representative sample of social network users (sampling).
- Information analysis of the content of social networks during the preparation of materials should take into account their specific features. These include the following:
- social networks are ahead of the media in their information function, being more efficient and faster in sharing information;
- the feedback function provides an opportunity to obtain public opinion on a significant issue, which is implemented by studying the discussion of topical issues of political, socio-economic life, projects, reforms, laws, activities of state and political figures, etc .;
- social networks provide an opportunity to actively form a certain point of view on the issue under discussion, through the sharing of views of more credible users, the so-called leaders of public opinion, around which the relevant information fields are formed;
- social networks form an informational daily schedule, raising certain topics, issues that are discussed by a wide range of people, attracting public attention. The analysis of the shared information and reaction to it by the network community, gives the chance to define relevant problems for the public or a certain social group, to find out the relation to this problem and vision of ways of its decision, to predict the further public reaction or development which can demand managerial response.

Marketers consider social signals to be one of the most important factors in the successful promotion of resources. Sharing, comments and "likes" on social networks really help to increase audience reach, measure the effectiveness of marketing campaigns, determine the compliance of published content with the needs and expectations of the audience. A user of the Facebook social network, adding another user as a

friend receives certain privileges, namely: to read user posts, to study the content published on the page. At the same time, users are provided with a service that helps to notify status changes, publication of new content and other actions that are defined in the individual user settings.

Mutual adding to friends allows to read posts that can only be seen on the page by users who are added to friends. At the same time, the user determines who can leave comments on his posts – all registered users, or only friends, or any reader who visited the user's page and is registered on Facebook. Facebook can be described as a closed social network, because each user needs to create an account in order to leave a comment on the post or create their own post. For the commercial sector, this is, to some extent, not very convenient, as it significantly limits the opportunities for unregistered users who do not have access to content located on Facebook.[20]

Facebook profiles have a certain single format, which gives the user limited ability to modify the settings according to their own preferences (for example, to create a page design).

No less significant disadvantage of Facebook is the advertisement, which is becoming increasingly intrusive. One of its annoying drawbacks is the monotony that arises through the use of a single advertising format: the demand plus the transition to the site or page of the advertiser.

A significant advantage of advertising on Facebook is its verification, so that the user can already be sure that by following the link of the advertising banner, he will not get a "virus" on his PC, as it is possible on other, untested sites.

Facebook offers users a standard set of features and settings for all pages, expanded with additional options.

Facebook options include:

- different types of posts and opportunities to comment on them;
- indication of extended user information;
- friends and friends' thread;
- pictures of users (userpic, avatar);
- account security features.

In addition to individual pages, the so-called communities have become widespread on Facebook. Communities are collective pages in which different users have the opportunity to write comments on a community topic and post content that may be of interest to the group. Facebook groups are very widespread in this social network as they allow to expand the circle of communication through common interests, without adding extra "friends", whose activity outside the general topic is not of interest to the user.

Social networks, forums, news and entertainment portals, blogs share a lot of valuable material about the benefits and features of people and companies [21].

For such "sharing" of data, first of all, it is necessary to identify the client in each source, which is allowed not by not all resources – there are resources without mandatory registration, or with a small amount of input identifying data. In some places, to identify the client, useful additional information about him or her among the large array of information is not enough. The most convenient and appropriate source for obtaining / sharing data about a person or organization is social networks. They contain information to identify the client, and additional information about preferences, marital status, education, social circles, etc.

Social networks, like other open Internet data sources, contain huge amounts of information about people - both current and potential customers of organizations. Most of this information people post online on their own: information about the place of birth and residence, places of study and work, family and close friends, interests, hobbies, and so on. Some of the information can be obtained as a result of analyzing the client's behavior: since when he or she has been registered on social networks and how often he or she visits them, how many friends he or she has and how often he or she communicates with them, from which devices the client enters the network, etc. Finally, useful information can be obtained from the analysis of the client's surrounding clarifying who his or her friends are, what their level of education and social status is, whether it corresponds to the level of the client.

The easiest way to collect data is to use the services of specialized companies that collect and constantly update data from many sources. The advantage of this method is the speed of obtaining information, which is beneficial for large volumes of customer base and the use of various social networks. Disadvantage – paid subscription to update data [22].

Another way is to use the software interfaces provided by many popular social networks. Different networks provide different available data, set their own restrictions on publications and, accordingly, set different prices for services. For example, if with the help of the VK network interface one can get complete information about the user, Facebook provides such information, which can be called almost "zero". The disadvantages of this method include the limitation on the number of simultaneous requests and the number of requests that the application can make per unit time. In addition, it is necessary to constantly monitor changes within the data and update the application for data collection. Moreover, in some social networks, the provision of important data about the user is a paid service. The advantages of this method are the ability to obtain customer data in a structured form, as well as the simplicity of integrating data issues into the downloaded application.

Another way is to manually parse social networking web pages, as well as to use ready-made crawlers to collect data followed by manual parse. In this case, you can access all open data and, in principle, to collect them quickly. The disadvantages of this method include the complexity of implementation – the web page of each social network is unique, so each time you need to develop new rules of analysis. The disadvantage is also the complexity of maintenance and the need for large computing resources, however, in the end, the process is well distributed [23].

Customer identification is to reveal all possible profiles representing a particular person on social networks. The source data for the search refer to passport data, but additional information will be useful. Such information as the name of the company where person works, his or her phone number, e-mail address, place of study, friends list, etc. can narrow the circle and help find a person.

The easiest way to identify a user is to search for an exact match of all known data about him or her. However, it should be taken in consideration that the relevant characteristics in social networks cannot be considered completely reliable – they can be absent and untrue or otherwise interpreted. Therefore, before identification, it is necessary to clean and normalize the known data, as well as to check the correctness of the parameters specified in the profile – for example, the user's city can be specified based on the analysis of his or her subscriptions, posts and statuses.

Targeting on social networks corresponds in most cases to the selection of the so-called advertising platforms. This is the most popular type of targeting. It is carried out by selecting advertising spaces so that their visitors (viewers) correspond to the target audience.

To collect data about networks on the Internet, one must either use an existing archive or collect new data using scrapers (parsers) and spiders. Scrapers are automated computer programs that take a web page and parse its content. Thus, this content becomes suitable for usage as received data. Spider is a special class of programs. They analyze connections and collect information. For spiders, "initial sowing" is often used - a database of specially selected pages [24].

Each characteristic used in the identification has a certain significance – the sum of all values at the coincidence of all parameters must be equal to one. In this way, surname, name and gender are one of the most important parameters during identification, and if this data is specified incorrectly, it will be impossible to identify this user as accurately as possible. Data on date of birth are of secondary importance. These data are recoverable; however, it should be noted that without them the chance of successful identification is very low. City and year of birth have the lowest significance, but these parameters are best recovered based on other data.

In addition to the data that network users voluntarily provide in their profiles, one can learn enough from the analysis of publications, community lists and photos. At the same time, additional facts that can be obtained from this unstructured information are also important. For example, if most user "wall" posts are about impressions of movies, one can determine user's interest in the movie.

5. Methods for determining the credibility of Facebook users

Linguistic tools must be used for automatic text analysis. In addition, statistical methods, machine learning technologies and in-depth data mining will be useful in the analysis. Usually, when performing research on statistics and working with natural language there are some inaccuracies – in statistics there are always some new assumptions that are not always fully implemented, and in natural language there is always the possibility of ambiguous interpretation of statements and conclusions. Therefore, a combination of linguistic and statistical approaches should be used to achieve a high level of reliability of the result and its good quality.

One user, with a certain set of identification data, can have a lot of different friends on social networks, which, in their turn, have a fairly high level of authenticity of their identification data.

In this case, there is a need to consolidate the data of multiple users into a single client profile. How exactly this data aggregation will take place depends on the goal – for example, to create a common list of interests, one can select and combine only data on interests, provided that each user possesses them.

Approximate factors that can be used as criteria for searching on the Facebook social network, were identified by an SEO-expert from the western market segment Glenn Gabe. He tried to analyze current ranking factors and made a list of 11 items that may affect or will affect the search on the social network soon.

1. The total number of "likes" to the content. This can be compared to the number of references per page.
2. The number of "likes" to the entire site or community. Something like the overall credibility of the project.
3. The level of preferred information. Facebook already has some developed metrics - People Talking About This, Engaged Users.
4. The level of coverage of certain regions. The so-called regional search.
5. The level of interest of certain categories of users.
6. Reactions of friends. A page that is "liked" by friends will have more significance in the ranking for other friends. This is about the same analogue in Google with its +1 button.
7. The "friends of your friends" factor. Comments and likes of friends of friends will be more important than other users' comments.
8. Amount of sharing.
9. Brand page. It can also affect the search: the more it is "promoted", the better.
10. On-Page optimization. In addition to optimizing the content itself, Facebook also analyzes the text of comments.
11. Influence of authoritative users. Here, as with Page Rank from Google, the more authoritative the user makes references to the page or puts a "like", the more significance he gives to this page.

6. Constructing formal models of the process of determining the user credibility

To build a functional model, the DFD data flow diagram was chosen (Fig. 1).

The diagram is based on the main purpose of the work, namely: on the basis of certain consolidated information, to determine the credibility of users of social networks, in terms of brand page. In the course of work, three main entities were identified: Administrator, Communities, User.

The administrator enters the processed data used in social networks into the system. This data includes information that generates activity statistics and user ratings, which are provided to communities and users, respectively. In turn, communities and users provide certain publications (content) and personal data, which, after being processed by the administrator, form the general statistics of the social network page.

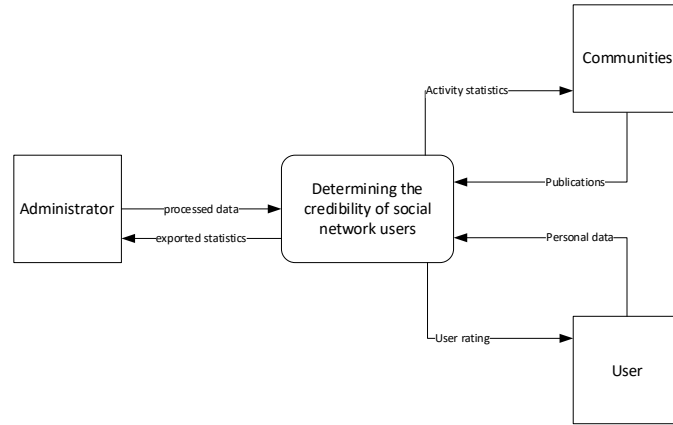


Figure 1: Contextual diagram of data flows

The decomposition of the first level, i.e., the division of the main process into subprocesses is shown in Fig.2.

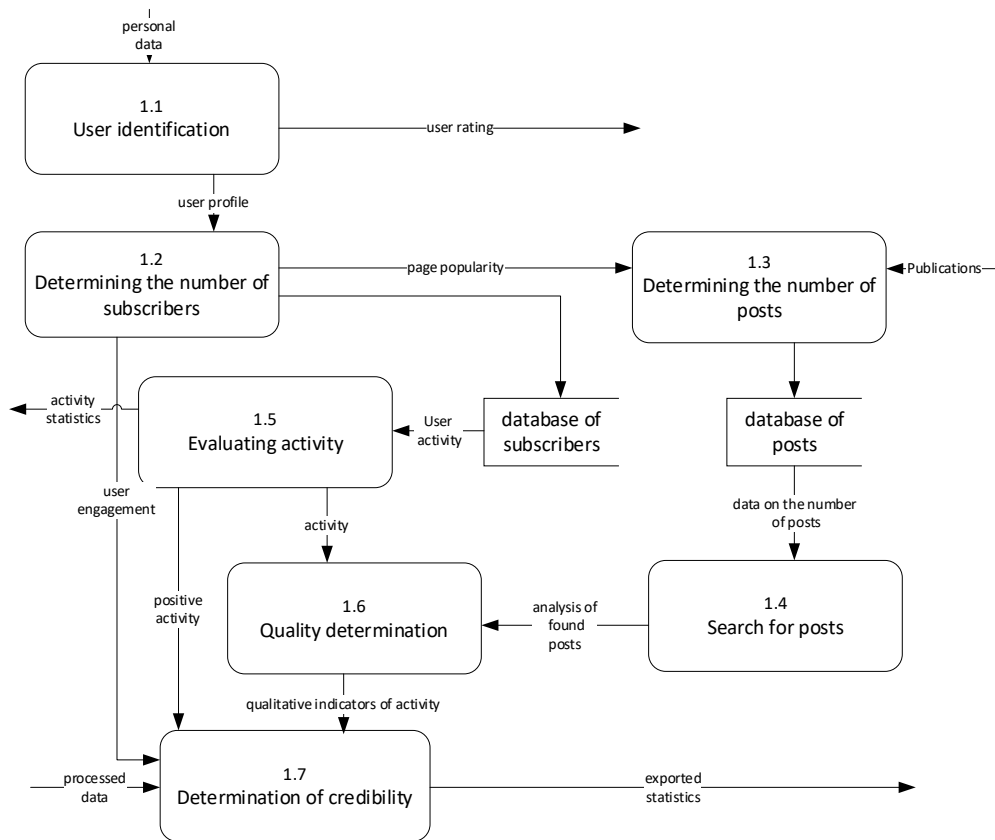


Figure 2: Detailing of the 1st level

As a result of the decomposition, the main process was divided into seven subprocesses: User Identification, Determining the Number of Subscribers, Determining the Number of Posts, Searching for Posts, Evaluating Activity, Quality Determination, and Determination of credibility.

First of all, to determine the credibility of social network users, it is necessary to identify the user of the social network, resulting in a rating of users.

The next step is to determine the number of subscribers to the brand page, and to move this data to a separate database. This data is used in the fifth stage along with the data on the posts to determine user activity. The third stage involves determining the number of posts on the page, which are similarly stored in a separate database.

The fourth stage deals with searching the posts with a high level of interest, which are taken from the database, where they were entered in the previous stage, and for which, in the fifth and sixth stages, respectively, user activity is evaluated and the quality of posting is determined (for which type of users and how far interesting the post is).

The seventh stage summarizes all the previous stages, and displays the overall credibility of users.

In Fig. 3 the decomposition of the second level of subprocess 1.7 Determination of credibility is shown.

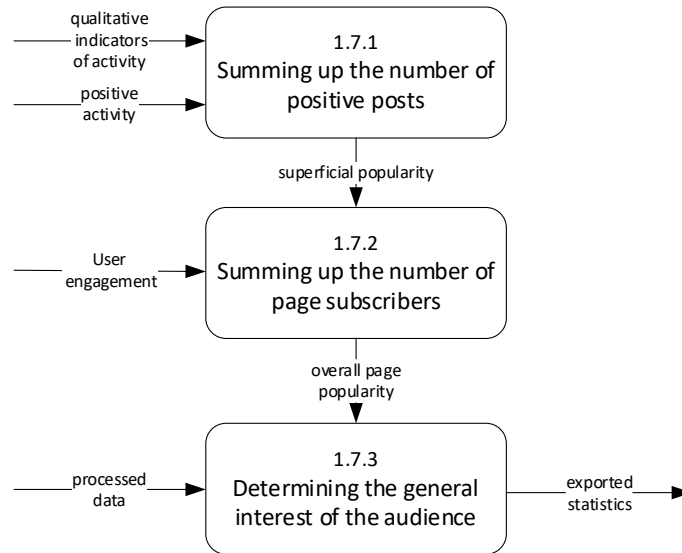


Figure 3: Decomposition of the 2nd level

In the sub-process Determination of credibility there are three additional sub-processes: Summarizing the number of interesting publications, Summarizing the number of subscribers to the page and Determining the general interest of the audience.

Entity-relationship model (ER-model) or entity-relationship diagram is a data model that allows to describe conceptual diagrams using generalized block designs. ER-model is a meta-model of data, i.e., a means of describing data models.

In this work, the diagram in the notation of Peter Chen was chosen (Fig. 4). Entities are displayed in the form of rectangles, connections in the form of diamonds. If entities are involved in a relationship, they are connected by a line. If the relationship is not mandatory, the line is dotted. Attributes are denoted as ovals and are associated with a single entity or relationship.

There are four entities in this diagram: Author of publication, Administrator, Author's post and Posts by other users.

The author of publication has the following attributes: Nickname of the author, Author ID, Surname and name, Reference. The administrator has the following attributes: Admin ID, Surname and name, Privileges, Virtual Address.

Author's post and Posts by other users have the following attributes: Author of publication, Publication ID, Reference.

Each author of the publication has the opportunity to subscribe to the same authors of publication and, accordingly, other authors of publication have the opportunity to subscribe to him or her and read his or her publications.

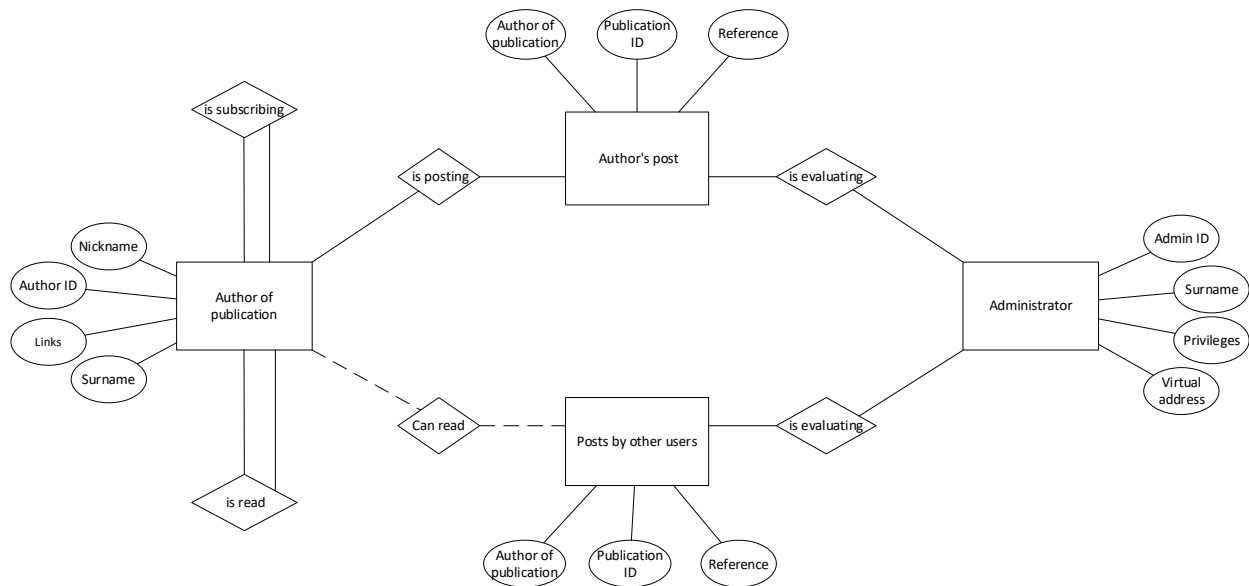


Figure 4: ER-diagram to determine the credibility of social network users

Each author of publication must publish the Post, but not necessarily read the publications of other authors.

In his turn, Administrator, to determine the credibility of the Authors of publication must evaluate the publications of one particular author and all others.

7. The results of research on the credibility of Facebook users

To determine the credibility of users of the Facebook social network, one should take into account two points. The general credibility of the user includes internal (credibility of the user as a subscriber of the page) and external (credibility of the user without taking into account the activity as a subscriber of the page) credibility. Depending on what goals need to be achieved as a result of the analysis of activity and credibility of the user, the proportions of credibility are correlated as 0.1 to 0.9 (a higher percentage is given to a more significant credibility). Therefore, if, as a result of the analysis, it is necessary to determine how much the user is interested in the activities of the page under analysis, and to plan further activities to increase this interest, internal authority will be more important. Conversely, if it is necessary to determine how active and credible the user is within the social network in general, whether other users listen to his advice and whether he has an influence on them, the external authority will be more important.

External and internal credibility are not always equal. The user may be a very active subscriber to the page, comment on each post, share it on the page, but, for example, have a small number of friends, insufficient response to these posts from other users, or even have a newly created profile or outdated profile which is not updated. Without any doubt, it is possible that the user has a very active profile, posts a lot every day and this is noticeable to both his friends and other users of the network, however, he does not take any action on the posts posted on the page. These two points must be considered to determine a user's credibility.

Therefore, you can calculate the overall credibility of the user by the following conditional formula:

$$C_{ext} + C_{int} = C_{total} \quad (1)$$

where C - credibility

To calculate the external credibility, it is necessary:

- to check whether the user, the subscriber of this page, is a member of pre-selected communities related to the subject of this page;

- to give the user a score from 1 to 10, depending on the membership / non-membership of a community (if the user is a member of all selected communities, he or she is assigned a rating of 10, if none - 1);
- to check user activity (profile support, registration date, last post date, number of posts) and provide a score from 1 to 10;
- to check the number of friends and give a score from 1 to 10;
- to summarize the received scores and provide a rating of the most authoritative users-subscribers

To calculate the internal authority of the user, it is necessary:

- to calculate the number of user comments on posts on the page, and provide a score from 1 to 10 (the more comments, the higher the score);
- to calculate the number of users sharing of the page's publications and provide a score from 1 to 10;
- to calculate the number of likes from the user and give a rating from 1 to 10;
- to summarize the received scores and provide a rating of the most credible users-subscribers.

After calculating external as well as internal credibility, they must be summed up, taking into account additional coefficients of significance, i.e., the formula will take one of two forms:

$$0,9 \times C_{ext} + 0,1 \times C_{int} = C_{total} \quad (2)$$

or

$$0,1 \times C_{ext} + 0,9 \times C_{int} = C_{total} \quad (3)$$

The user page analysis is mostly done manually, and all scores provided to users for meeting certain criteria can be considered subjective. Depending on who is conducting the analysis, the ratings may differ. When it comes to fulfilling user membership search queries in selected communities, not all data can be obtained using Graphic Search. For instance, if a user hides information about his or her membership in certain communities, the result of the request will not give anything, which, in some way, makes it impossible to properly assess the user.

As a result of all points of the analysis, the table in Excel with scores, which can be edited arbitrarily, was provided (table 1).

Table 1

Table with user rating

Name and surname	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14
Community membership	10	10	10	10	10	10	10	10	10	10	10	8	10	10
Profile activity	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Number of friends	10	10	10	10	10	10	10	10	10	10	10	10	8	8
General external credibility	30	30	30	30	30	30	30	30	30	30	30	28	28	28
Comments	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Sharing	1	1	1	1	1	1	1	1	1	1	1	9	1	1
likes	8	7	6	5	5	1	1	1	1	1	1	10	9	9
General internal credibility	10	9	8	7	7	3	3	3	3	3	3	20	11	11
C_{total} formula (2)	28	27,9	27,8	27,7	27,7	27,3	27,3	27,3	27,3	27,3	27,3	27,2	26,3	26,3
C_{total} formula (3)	12	11,1	10,2	9,3	9,3	5,7	5,7	5,7	5,7	5,7	5,7	20,8	12,7	12,7

When calculating the authority of the user according to formula (2), the most authoritative user-subscriber of the page is a user named «User 1».

When calculating the authority of the user according to formula (3), the most authoritative user-subscriber of the page is the user named «User 12».

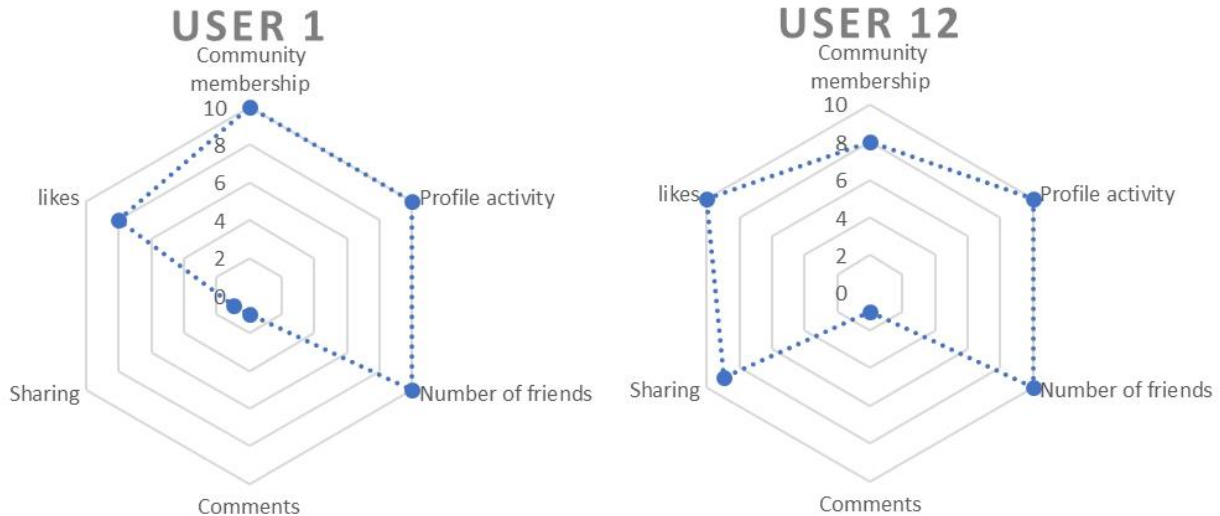


Figure 5: Radar chart of values of user activity indicators

A comparison of the two most reputable users according to formulas (2) and (3) is shown in Figure 5. It also showed which parameters of the user defined him/her as the most authoritative one. In each case, when determining the authority, you need to specify which of the parameters are important for the calculation: internal or external authority.

8. Conclusion

The main methods and tools for analyzing user data from social networks were considered. The tasks, methods and programs of network and text data analysis were described, as well as approaches to obtaining initial data for analysis: collection of real data by accessing the web interfaces of social services through the social search of the social network Facebook.

Based on theoretical background of social networking, the structure of formal model to determine the credibility of users was provided. Data Flow Diagram was used for this purpose. Social networking processes were introduced involving decomposition of both first and second levels. Determining the credibility of social network users was provided with entity-relationship model as a means of describing data models. To sum up the research results of the Facebook user credibility indicators, the table with users' rating was introduced. Combining linguistic and statistical approaches help determine social network user credibility.

In perspective, the extended formal model pattern of different social networking approaches may be considered and used as a basis for further studies in the field of global social networking.

9. References

- [1] M. Thai, N. Sheeran, D. J. Cummings, We're all in this together: The impact of Facebook groups on social connectedness and other outcomes in higher education, *The Internet and Higher Education*, (2019), pp. 44–49. doi:10.1016/j.iheduc.2018.10.001

- [2] K. Osaka, F. Toriumi, T. Sugawara, Effect of direct reciprocity and network structure on continuing prosperity of social networking services, *Computational social networks* 4(1):2 (2017). doi: 10.1186/s40649-017-0038-2
- [3] A. Richter, M. Koch, Functions of social networking services, in: *Proceedings of the International Conference on the Design of Cooperative Systems (COOP 2008)*, Carryle-Rouet, Provence, France, 2008, pp.87–98.
- [4] O. Markovets, N. Vovk, Y. Turchyn, O. Smirnov, Model of informational support for social network administrators' content creation, in: *Proceedings of the 2st International workshop on control, optimisation and analytical processing of social networks, COAPSN-2020*, volume 2616, Lviv, 2020, pp.125–136. URL: <http://ceur-ws.org/Vol-2616/paper11.pdf>.
- [5] S. Fedushko, Y. Syerov, R. Korzh, Validation of the User Accounts Personal Data of Online Academic Community, in: *Proceedings of the 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (Tcset)*, 2016, pp. 863–866.
- [6] A. Peleshchyshyn, N. Vovk, The Problem's Formation of the Entire Informational Environment at the Long Existed Universities, *Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (Csit 2017)*, volume 1, 2017, pp. 139–142.
- [7] S. Albota, Resolving conflict situations in reddit community driven discussion platform, *Proceedings of the 4th International conference on computational linguistics and intelligent systems (COLINS 2020)*, Lviv, Ukraine, 2020, pp. 215–226.
- [8] A. Peleshchyshyn, S. Albota, Contradictory statement as a basis for conflict resolution strategies, Methods and algorithms for performing separate operational tasks for the protection of the state information space In: *Proceedings of the international workshop on conflict management in global information networks (CMiGIN 2019)*, volume 2588, Lviv, 2019, pp. 336–345.
- [9] A. Peleshchyshyn, V. Vus, O. Markovets, R. Pazderska, Methods and algorithms for performing separate operational tasks for the protection of the state information space In: *Proceedings of the international workshop on conflict management in global information networks (CMiGIN 2019)*, volume 2588, Lviv, 2019, pp. 392-403.
- [10] A. Petrushka, M. Komova, I. Demydov, Social networks as an alternative environment for the implementation of scientific communication. In: *Proceedings of the 1st International workshop on control, optimisation and analytical processing of social networks, COAPSN-2019*, volume 2392, Lviv, 2019, pp. 79-94. URL: <http://ceur-ws.org/Vol-2392/paper7.pdf>
- [11] V. Vysotska, A. Sachenko, V. Lytvyn, Y. Burov, Information resources processing using linguistic analysis of textual content, in: *Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Bucharest, 2017, pp. 573–578.
- [12] I. Khomytska, V. Teslyuk, A. Holovatyy, O. Morushko, Development of methods, models, and means for the author attribution of a text. In: *EasternEuropean Journal of Enterprise Technologies*, volume 3, issue 2 (93), 2018, pp. 41–46.
- [13] J. Masip, M. Bethencourt, G. Lucas, M. Sanchez-San Segundo, C. Herrero, Deception Detection from Written Accounts. *Scandinavian Journal of Psychology*, 53 (2), 2012, pp. 103–111.
- [14] J. Pennebaker, M. E.Francis, R. J. Booth, *Linguistic Inquiry and Word Count: A computerized text analysis program*, Mahwah, NJ, Erlbaum, 2001.
- [15] J. Pennebaker, M. R.Mehl, K. G. Niederhoffer, Psychological aspects of natural language use: Our words, our selves, *Annual Review of Psychology*, 54(1):547-77, 2003. doi: 10.1146/annurev.psych.54.101601.145041
- [16] T. Wenginger, An exploration of submissions and discussions in social news: mining collective intelligence of reddit, *Social Network Analysis and Mining*, 4(1), 2014.
- [17] R. Kumar, M. Mahdian, M. McGlohon, Dynamics of conversations. *ACM KDD*, 2010.
- [18] F. Benevenuto, T. Rodrigues, F. Magno, V. Almeida, Characterizing user behavior in online social networks, *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, 2009, pp. 49–62.

- [19] K. Molodetska, Y. Tymonin, O. Markovets, A. Melnychyn, Phenomenological model of information operation in social networking services, *Indonesian Journal of Electrical Engineering and Computer Science*, volume 19 (2), 2020, pp. 1078–1087
- [20] M. Arzubov, N. Shakhovska, P. Lipinski, Analyzing ways of building user profile based on web surf history, *Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT 2017)*, volume 1, 2017, pp. 377–380. doi:10.1109/STC-CSIT.2017.8098809
- [21] M. Risius, O. Aydingül, Facebook User Segmentation to Enable Targeted Social Advertisement, *Proceedings of the 39th International Conference on Information Systems*, At San Francisco, USA, 2018, pp. 1–8.
- [22] M.E. Marrakchi, H. Bensaid, M. Bellafkih, E-Reputation prediction model in online social networks, *International Journal of Intelligent Systems and Applications*, 9(11), 2017, pp. 17–25. doi: 10.5815/ijisa.2017.11.03
- [23] B. Schivinski, D. Dąbrowski, The effect of social-media communication on consumer perceptions of brands, *Journal of Marketing Communications*, 22(2), 2014, pp. 189–214.
- [24] J. Luo, X. Pan, S. Wang, Y. Huang, Identifying target audience on enterprise social network, *Industrial Management & Data Systems*, volume 119 (1), 2019, pp. 111–128, doi: 10.1108/IMDS-01-2018-0007