SOCIAL NETWORKS AS CHALLENGE FOR MARKETING INTELLIGENCE

Professor **SANJA BIJAKŠIĆ**, Ph.D.
Faculty of Economics University of Mostaru,
e-mail: sanja.bijaksic@sve-mo.ba

Professor **BRANO MARKIĆ**, Ph.D.
Faculty of Economics University of Mostaru,
e-mail: brano.markic@sve-mo.ba[2]

Professor **ARNELA BEVANDA**, Ph.D.
Faculty of Economics University of Mostaru,
e-mail: arnela.budimir@sve-mo.ba

ABSTRACT

Social networks are changing the way of connection and communication between people by increasing the amount of publicly available information and knowledge. People of similar professional backgrounds and occupations link to online communities to share information. This has a direct impact on what is one of the most difficult aspects of marketing intelligence "efficient and rapid collection and sharing of data and information". The aim of marketing intelligence is not only access data but manage them, analyze them and based on the analysis to make the right decisions related to customers, products, price, promotion, sale. Therefore, a large number of companies today are looking for solutions by marketing intelligence that will enable access to text data, analyze them and improve the quality of marketing decisions.
The paper raises the hypothesis that it is possible to build a system for marketing intelligence that collects and analyzes data from social networks and uses the analysis results (information) to make precise, concise and accurate marketing decisions. In the paper is used the R programming language for marketing intelligence system and the R language demonstrated satisfactory simplicity and application power.

KEYWORDS: marketing intelligence, text mining, social network, sentiment analysis, marketing decision making.

1 INTRODUCTION

Marketing intelligence is a systematic approach to gathering data from internal and external sources of companies with the ultimate goal of generating reliable, usable and timely

_____

information for making marketing decisions. Crowley, Ed., [1] analysis marketing intelligence by three interconnected pyramids: information, communication and people / processes. Information pyramid as component of marketing intelligence has three levels: data, analysis and insight, communication pyramid has again three levels: gathering, dissemination and dialoque and people& processes have research, reporting and capability. Marketing intelligence for Crowley is decomposing in competitive insight (competitor intelligence and product intelligence) and market insight (market understanding and customer insight). Marketing intelligence on social networks is focusing on customers, their preferences, satisfaction rates, customer opinion and attitudes. Social networks are today a big data sources available to marketing intelligence. The real power of the social networks for marketing intelligence are feedbacks of customers about products, promotion, prices, employees, brand awareness, brand preferences, loyality.

Market intelligence is closely related to marketing research. But there are some differences because marketing research involves solving a specific marketing problem at a specific point in time, market intelligence involves gathering information on a regular, ongoing basis to stay in touch with what's happening in the marketplace [8].According to Kotler [4] marketing intelligence provides current information about marketing environment to help managers to organize and prepare marketing plans. Marketing intelligence increasing attention from internal data such as transaction histories, databases, data warehouses, data marts, loyalty cards and so to social networks. Social media monitoring is a type of marketing intelligence that overlaps with market research. But the social networks analysis is not specific marketing problem because today it involves gathering and analysis on a regule basis.

Social networks are becoming a safe source of quality data about the products, the mood of the market to the products but also the image of the company as a whole, as a system on the market. Without proper software tools and information technology support is not possible to collect and analyze huge amounts of data in the form of messages and comments on social networks.

However, it is necessary to be cautious in the use of data from social networks. The reliability of conclusions and information obtained by analysis is particularly important. The representativeness of data on social networks depends on the wide of their use. If social network has more users then the data representativeness for marketing intelligence is higher (properly represent the views of the whole market and exclude dominant influence of opinion leaders).

2 MARKETING INTELLIGENCE AND THE SOURCE OF DATA

Marketing intelligence follows the logic of any scientific research. The research is based on a preset reasonable assumption about the relationship of two or more research variables. Such logical and meaningful assumptions in terms of analysis are called hypotheses.

Figure 1. Marketing intelligence dynamics
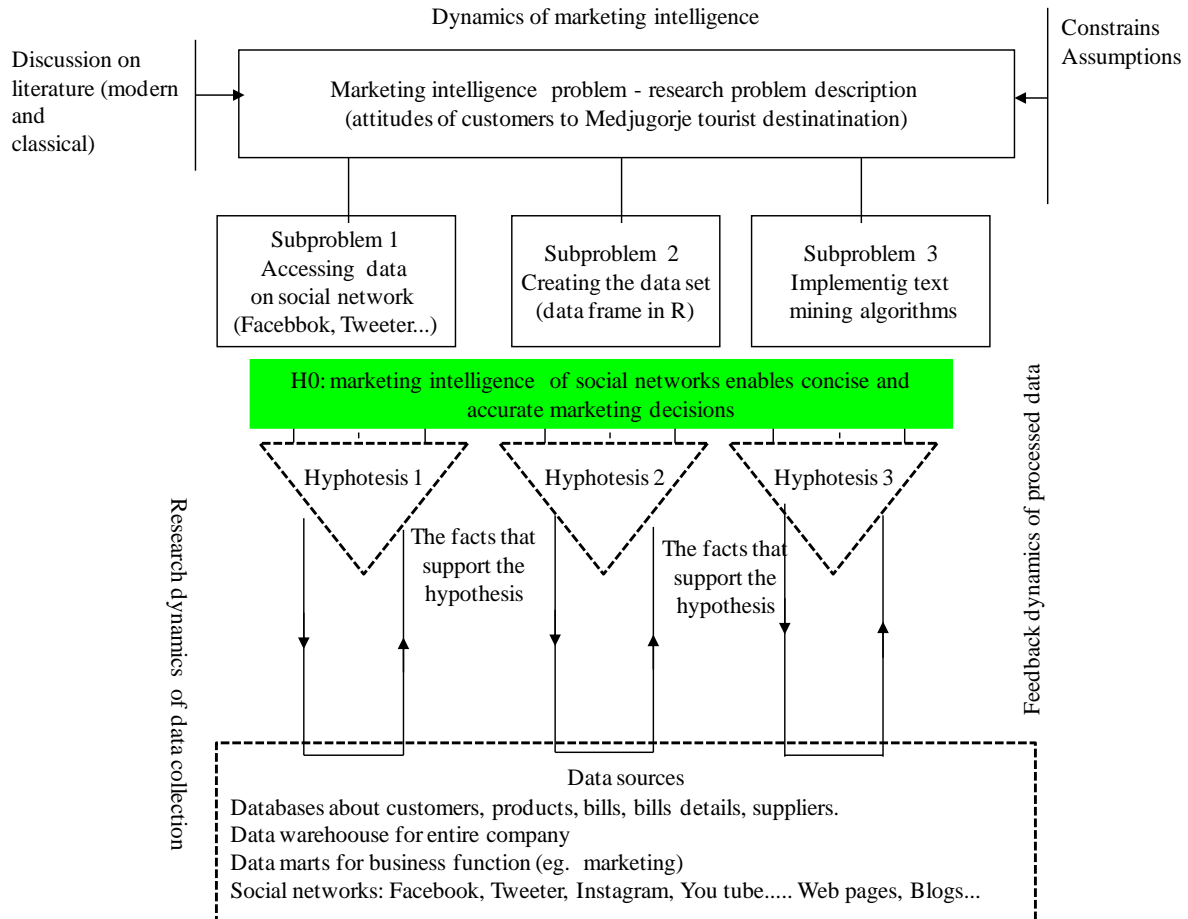
Research dynamics within the marketing intelligence presupposes the existence of data and appropriate algorithms (methods) of data analysis in order to refute or accept the proposed hypotheses. Data from social networks contributes to shape a complete view of the market, the market intelligence view. Today, research is based in principle on the huge amounts of data and uncovering hidden regularities that are established among them. Such regularities are not trivial nor simple. Information technology is almost at all stages of marketing intelligence from collecting, storing data to their analysis and visualization of the analysis results.

The corporate database relational or object oriented is first and indispensable data source for marketing intelligence. In transactional database are stored data about all daily business transaction. This database records data about customers, suppliers, order details, products, orders, inventories etc. Semantically rich data model implemented in the form of a database enables marketing experts analysis and this analysis ends with information for better decision making [7]. However, transactional database hiding most information about customers, their behavior, correctness ways of pricing, the results of promotional campaigns.

_____

The growth of transactional databases is very fast and they are not adequit for analytical purposes. Therefore, medium and large enterprises develop specific databases for analytical purposes. These are called data warehouses. Foundation of the data warehouse concept is a separation of day-to-day transactions from analyzing and reporting that make marketing analysts or entrepreneurs.

According to Inmon [2] the key objective of data warehouse is to place end users directly in touch with the data they need to make better decisions. Kimball [3] states data staging process called ETT (Extraction, Transformation, and Transportation) as the key part of the data warehouse project.

However, databases and data warehouses are not the only sources of data for marketing intelligence. Thus, the views and opinions of the customers can be analyzed by using a questionnaire as an instrument of data collection or data stored in social networks.
Corporate website may contain valuable data about who is looking for its products or services. Finding hidden patterns and correlations within the data from a large dataset is a challenging data mining or text mining.

## 3 SOCIAL NETWORKS AS DATA SOURCE FOR MARKETING INTELLIGENCE

Social networks such as Twitter, Facebook, LinkedIn, blog, online forums have important roles for people, business, political parties, science because enable communication and sharing information with others. Internet marketing implements various social networks in order to achieve integrated marketing communication by supporting activities of social sharing of content, images, and video. Social networks are important element in marketing plans. Namely, internet marketing using social networks may offer valuable information that customers will find interesting, may create a variety of content by implementing videos,images, graphics in addition to classic text-based content. Using social networks improves business sharing of original content in order to gain followers and fans. Sharing content and linking to outside sources stimulates customers to improve trust and reliability in products and corporate.

The Facebook social network is accessible online communication service used for business purposes. Allows you to connect with users. They save money, they cost almost nothing and are made very quickly. Simply, invitations sent via social networks save you money and time. Marketing and promotion are particularly valuable for social networks and vice versa. Measuring the success of promotional activities and marketing decisions is carried out by analyzing the content on social networks.

_____

The idea of Facebook social network is based on a personal profile and communication among friends but the profile can not be used for commercial purposes. For the promotion activities customers can access the group. In the group are linked profiles with some common characteristics so they can interchange information on familiar topics (eg. from a same professional area as is marketing analytics).

The fan page can promote a product, service, tourist destination. Friends (fans) read messages, informing about products, events, services, prices. Commitment is to shape the page so it is interesting to the users and they want to follow. It becomes a channel for two-way communication and social network channel for the promotion of ideas, products, services. Companies that constantly innovate existing products or create new products and services need to communicate with customers but the customers have also a need to monitor their news, ask them questions and seek answers. In Figure 2. is given one example of brand page on Facebook:

Figure 2: Brand Page on Facebook social network
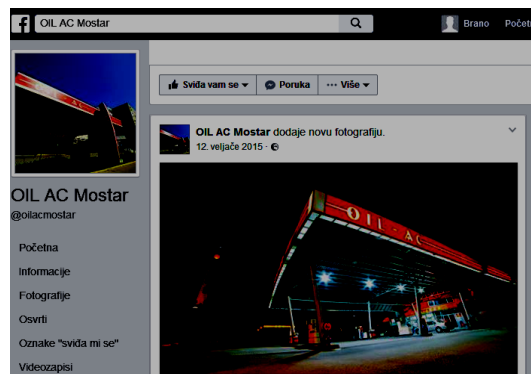
Facebook becomes a channel of communication with the market. Companies have at disposal a social network as an indispensable channel of communication with consumers and it's environment and often, on a daily basis, inform customers, publish sweepstakes, undertaken action, price changes, provide new products or services. Number of fans becomes the main measure of the relevance of the page as well as the product or service. The obligation of the management is to recognize and exploit the potential of information technology.

On social networks are huge amounts of data that are often do not even analyze. There are many reasons but the the two most significant. The first is the lack of time and insufficient level of training and knowledge management. These reasons may be called subjective. Other reasons can be attributed to complexity of the data analysis on social networks in terms of building the appropriate software tools and recognition of algorithms that enable analysis [6].
For marketing intelligence the most important are comments posted by people, customers where they express their attitudes, emotions and opinions via texts.

_____

4 RESEARCH METHODOLOGY

The research framework in this paper consists of four main phases which are definition of research problem, data gathering and building the data set for analysis, text mining and result analysis.

4.1.DATA COLLECTION

The first step will be to take data from from Medjugorje Facebook pages. For access any data from facebook pages, user required to login to Facebook (the assumption isthat there is a profile on Facebook). A precondition for data collection from Facebook's is to access URL address: https://developers.facebook.com, register and fill required information on Facebook developer page (form).

After register and login to Facebook the next step for the user (marketing intelligence) is getting the basic application settings: (the name of application is BrAnSa) application ID and secret code.

Data from Facebook will be collected as well as analysied using R programming language[3]. The packages devtools and Rfacebook have to be install using the statements[4]:
>install.packages("devtools")
>library(devtools)
>install_github("Rfacebook", "pablobarbera", subdir="Rfacebook")

The package „Rfacebook" provides a few functions that allow R users to get information about Facebook users, pages, comments and posts.

After installing these packages need to connect R session with test application and authenticate it to Facebook Profile for Data Mining. The package Rfacebook offers a very simple function for that. All you need is to copy *app id* and *app secret* from *app settings* on the Facebook developer page:

```
>require("Rfacebook")
> library(httr)
> fb_oauth <- fbOAuth(app_id=" app id from our app settings ", app_secret =" app secret
oour app settings ", extended_permissions = TRUE).
```

---

[3] R is an interpreted computer programming language where most user-visible functions are written in R itself. R is made up of: operators (+ - <- * %*% …) for calculations, collection of functions for making quality graphics and sets of functions (packages). R is simple and also suitable for data analysis.

[4] These packages are installed from github as this is currently the most recent version.

_____

The function fbOAuth creates an access token that enables R to make authenticated calls to the Facebook API[5].

The package httr is organised around four most popular verbs of http (hypertext transfer protocol): GET (*fetch* an existing resource), POST (*create* a new resource), PUT (*update* an existing resource) and DELETE (*delete* an existing resource).

After dialog with R language using the fbOauth() function:

> fb_oauth <- fbOAuth(app_id=" *app id from our app settings* ", app_secret =" *app secret oour app settings* ", extended_permissions = TRUE)
Copy and paste into Site URL on Facebook App Settings: http://localhost:1410/
When done, press any key to continue...

The main hypotheses of this paper is that it is possible to build a system for marketing intelligence that collects and analyzes data from social networks in order to make accurate marketing decisions.

Therefore will be collected data in the form of comments for the pages with keyword Medjugorje on social network Facebook. The packet Rfacebook includes the function searchPages() whichs retrieves public pages that mention a given keyword.[6]

>pages<-searchPages(string="Medjugorje", token= fb_oauth, n = 20)
The keyword is Medjugorje, token is assigned fb_oauth and the number of pages is 20.
The application goal is to extract list of posts from a public Facebook page Medjugorje. The function getPage() with three parameters is applied:

>pageMedjugorje <- getPage("Medjugorje", token= fb_oauth, n = 2000)
where the first parameter *„Medjugorje"* is page name, the second parameter *token= fb_oauth* is the OAuth token created with fbOAuth and the third parameter *n=2000* is number of posts of page to return. We were requested 2000 posts and API returns all of them:

>pageMedjugorje <- getPage("Medjugorje", token= fb_oauth, n = 2000)
>25 posts 50 posts 75 posts 100 posts..... 1796 posts 1821 posts 1823 posts
The total number of messages that contain the name of Medjugorje destination is 2000 so we in the R session did not get any warning message (we defined in the previous statement 2000 posts). The first three amomg 1823 posts are:
>pageMedjugorje[1:2,]

_____

[5] OAuth is a protocol that provides to clients a "secure delegated access" to server resources on behalf of a resource owner.
[6] This function has three parameters searchPages(string, token, n = 100), where parameters have the next meaning: string: containing keywords to search, token: either a temporary access token created at https://developers.facebook. com/tools/explorer or the OAuth token created with fbOAuth, n: mximum number of pages to return. Authors of this function are Pablo Barbera and Joel Gombin.

_____

The returned data frame basically contains the fields from_name, from_id for every single like of the post, message, created_time, type, likes_count, comments_count and shares_count. For the first row in data frame pageMedjugorje result is displayed in the next table:

Table 1. Data frame of posts for pages Medjugorje on Facebook

	from_id	from_name	message	created_time	type
1	169173713621	Medjugorje	nella tempesta, fidiamoci di Gesu...	2016-08-18T14:02:54+0000	link

id		likes_count	comments_count	shares_count
1	http://www.amicidilazzaro.it/index.php/la-fiducia-in-dio-forza-nelle-tempeste-raniero-cantalamessa	166	5	15

The last three columns in data frame pageMedjugorje contain information about the number of likes, comments and shares. This information can be used to get the popularity of a post in a page. The most messages are written in italien language[7].

Marketing intelligence is concetrated only on text data stored in column message of data frame pageMedjugorje. This is the third column in data frame and extraction of all messages (comments) enables the function:

>fb_wcloud=pageMedjugorje[, 3]

The first eight messages will display the next statemenf od R languages:

> comMed=pageMedjugorje[1:3,3]

> comMed

[1] "nella tempesta, fidiamoci di Gesù..."

[2] "Affidarsi a Dio....\n\n(in English : Act fo trust ---> http://goo.gl/NKWocM)"

[3] "\"Aiutaci a portare il fardello delle tribolazioni quotidiane, non con l'anima dei disperati, ma con la serenità di chi sa di essere custodito nel cavo della mano di Dio...   If you followed all the previous steps of the algorithm for extracting data from Facebook pages and R language commands data in the form of text will be stored in variable fb_wcloud. This data can be analyzed using text mining algorithms and in theb paper will be implemented a word cloud. It is an acceptable form of textual data analysis because in graphical form shows the keywords most frequently-occurring in the text.

_____

[7] Now we could analyse the gender od users by implementation the R functions which are based on the column from_id. The user insights can get with the getUsers() function. This function extracts the user_id which is the field from_id. From the returning user data is possible extract the gender and save it to a temporary gender_frame. The complete code and solution is at the URL https://www.r-bloggers.com/gender-analysis-of-facebook-post-likes/.

_____

4.2. TERM DOCUMENT MATRIX

Initial data sets fb_wcloud consisted of textual messages and comments. Term document matrix reprensets the frequency of each term in document. If the D_i denotes i-th document (message) and T_j denotes j-th term then f_{ij} is frequency of j-th term i j-th document. Often used in information retrieval and text mining is tf-idf parameter. This parameter stands for *term frequency - inverse document frequency and* evaluates how important a word is to a document in a corpus (collection). The package tm is a framerk for text mining in R language and need to be load first.

The main structure for handling documents in the package ***tm*** makes the so-called Corpus and it represents a set of documents [9]. You must enter the following sequence of commands:[8]
> library(tm)
>fb_Corpus<-Corpus(VectorSource(fb_wcloud))

The package tm provides a set of predefined sources (DirSource, VectorSource, DataframeSource). VectorSource interpreting each component as document and only accepts (character) vectors. After being "caught" a collection of documents (fb_Corpus) we need to modify this collection so that will eliminate the blank spaces in the documents (messages), signs of punctuation and the like. The logic is to eliminate all signs and words that do not carry information relevant to the semantics of the text. The aim is to translate the documents into a simpler form so that they are suitable for parsing. It is sufficient in the loop *for(i in seq(fb_Corpus))* add commands that from the collection of documents extract characters or words that will be replaced by spaces. It displays the following loop[9]

```
for(i  in seq(fb_Corpus))
  {
  fb_Corpus [[i]] <- gsub("del", " ", fb_Corpus [[i]])
  fb_Corpus [[i]] <- gsub("#", " ",  fb_Corpus [[i]])
  fb_Corpus [[i]]  <- gsub("\\|", " ",fb_Corpus [[i]])
  fb_Corpus [[i]]  <- gsub("con", " ",fb_Corpus [[i]])
  ............................................................
  fb_Corpus [[i]]  <- gsub("que", " ",fb_Corpus [[i]])
```

It is very useful implement the transformation of corpus to be sure all of data is in PlainTextDocument:

_____

[8] To understand thecommands, packages and functions in the R language is helpful to visit websites that give detailed explanations of the use of the. For package tm useful site is:
 http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf.

[9] Most of the text in the collection of documents fb_Corpus is written in Italian. Therefore, from the text are eliminated the Italian words that not have a special semantic value.

_____

_____

>fbCorpus<- tm_map(fb_Corpus, PlainTextDocument)[10]

The next step is to create a matrix for learning which includes frequency of terms. A document term matrix or term-document matrix is a mathematical matrix that describes the frequency of terms that occur in a collection of messages from Facebook pages Medjugorje as tourist destination. In a document term matrix, rows correspond to documents in the collection fb_wcloud and columns correspond to terms. DocumentTermMatrix is a central part of text mining and in the *tm* package the functions DocumentTermMatrix or TermDocumentMatrix (depending on whether you want documents as rows and terms as columns, or vice versa) transforms corpora into matrix. This is very important step because the analysis and text mining is based on the quality of document term matrix. The function DocumentTermMatrix()in R language has the following syntax:

TermDocumentMatrix(x, control = list())

where is: *x* - a corpus for the constructors; *control* - a named list of control options

In our example control includes in term document matrix only the words with minimum length of five characters:

```
> fb_tdm<- TermDocumentMatrix(fbCorpus, control = list(minWordLength = 5))
> fb_tdm
<<TermDocumentMatrix (terms: 9910, documents: 1823)>>
```

The number of terms in given example is 9910 and documents 1823. The terms (terms) are chosen based on the frequency of the document (Document Frequency-DF). Words that exceed a certain threshold (minimum length of three characters) will form a list of index terms.

4.3. RESULT ANALYSIS

The next step is to create a term matrix that contains frequencies of terms for learning. In our example, the number of terms is 9910 and the number of documents 1802. On term document matrix can be applied a huge amount of R functions for data mining like classifications, prediction, clustering etc.

First will fb_tdm transform in matrix y function as.matrix(). It is generic function and for data frames will return a character matrix:

```
>m_dtm<- as.matrix(fb_tdm)
```

Ne next very useful step is to sum the frequency each one term (row of term document matrix) and sort the frequency in decreasing order.

```
>m_s<- sort(rowSums(m_dtm),decreasing=TRUE)
```

_____

[10]**Plain text** is file contents that represent only characters but not its graphical representation nor other object. That is done by fbCorpus<- tm_map(fb_Corpus, PlainTextDocument) and now should made DocumentTermMatrix.

_____

Than will be created the data frame m_df with two columns: word and freq (frequency):
>m_df<- data.frame(word = names(m_s),freq=m_s)
Ten words with the highest frequency shows the function head():
> head(m_df,3)

Table 2. The most frequency terms in term document matrix of colelction fb_corpus

word	freq
medjugorje	805
dio □	185
reina	173

Now we can start asking questions like: what are the most association with word „dio" where corellation is greater than 0.28? Answer is very simple by implementation the function findAssocs():

Table 3. Associations in fb_corpus with word „dio"

```
> findAssocs(m_dtm, "dio", corlimit=0.28)
$dio
    dio:   liberaci,   suppli, disprezzare, protezione,  prova,  rifugio,  santa,  benedetta,
    0.31     0.31      0.31       0.30        0.30       0.30     0.30     0.30     0.29
    icolo,   ogni,    tua,   cerchiamo
    0.29     0.29     0.29      0.28
```

The primary goal of the paper is to present the analysis result in visual form which is especially acceptable in business environment. There are a few reasons why is word cloud used to represent the research information, First, word cloud discovers the essential and an overall sense of comments and messages about Medjugorje as destination. It may stimulate more questions and answers, but that it is a good entry point to take marketing and promotional actions. The marketing intelligence using word cloud may give a fast insights in products, services, customer satisfactions, perception etc. Marketing expert does not need to understood all technological and sofware aspects of research results but can share back results from reserach very quickly. To display the results of research it is necessary to load the wordcloud package in the R session:
>library(wordcloud)
The function wordcloud plot a word cloud[11]:

_____

[11] The first argument m_df$word takes the terms from data frame m_df, the second argument m_df$freq is the words frequency, the third argument scale=c(4,.5) scale a vector of length 2 indicating the range of the size of the words, min.freq denotes that words with frequency below min.freq will not be plotted, the argument max.words is maximum number of words to be plotted,, the argument random.order plot words in random order (false means they will be plotted in decreasing frequency and the last argument colors choose colors of the word cloud.

_____

>wordcloud(m_df$word, m_df$freq, scale=c(4,.5), min.freq=40, max.words=Inf, random.order=FALSE, rot.per=.15, colors= brewer.pal(8,"Dark2"))
The result of reserach is visible at the next Figure:

Figure 3. Word cloud for page Medjugorje on social network Facebook

The most frequent words are Medjugorje, message, reina, dio, preghiera, madre, regina, papa, vaticano and so on. The most messages and key words in word cloud are in Italian. The marketing intelligence expert could conclude that Medjugorje is a place of prayer (preghiera) in which respects the Queen of Peace (regina) and about Medjugorje care the pope and Vatican. The same algorithm and analysis could be implemented to other destinations, brands, events, people and so on. The marketing intelligence get a very fast, accurate and reliable insight into these objects of analysis.

5 CONCLUSION

Systematic view decomposing marketing intelligence in competitive insight (competitor intelligence and product intelligence) and market insight (market understanding and customer insight). Social networks are today a big data sources available to marketing intelligence and the real power of the social networks for marketing intelligence are feedbacks of customers about products, promotion, prices, employees, brand awareness, brand preferences, loyality. This paper shows in details how is possible build the application for marketing intelligence which gathers data from Facebook as one of the most important social networks, analyze this data and get the complete insight about customers opinion about tourist destination. Except the marketing knowledge it was necessary to know the way of functioning of social networks but also knowledge of software tools, packages and their features that allow the construction of such a system for marketing intelligence. The programming language R showed a strong

_____

application and development power. The research results are visualized in the form of a word cloud and are simple to use in marketing.

This system is open to new extensions in terms of collecting data from social networks Facebook and analyze them by other complex algorithms such as are algorithms of clustering (k-means or fuzzy c-means), classification (naive Bayes classifier, logistic regression, Support Vector Machines) and prediction.

REFERENCES

[1]   Crowley, Ed., (2004), *Market intelligence versus marketing research*, Quirk's Marketing Research Review.

[2]   Inmon W.H.,Welch J.D., Glassey K.L.,(1997), *Managing the Data Warehouse*, New York: John Wiley & Sons.

[3]   Kimball R.,Reeves L.,Ross M.,Thornthwaite W.,(1998), *The data warehouse Lifecycle Toolkit*, New York: John Wiley & Sons.

[4]   Kotler, P., (1996), *Principles of Marketing*. The European Edition, USA: Prentice Hall.

[5]   Markić, B., Bijakšić, S., Bevanda, (2016), *Sentiment analysis of social networks as a challenge to the digital marketing*, Ekonomski vjesnik/ECONVIEWS, God.XXIX, BR. 1/2016, str. 95-107.

[6]   Markić, B., Bijakšić, S., Bevanda(2016), *Customer Intelligence Analytics on Social Networks*, Expert Journal of Marketing, 4 (1), pp. 24-30.

[7]   Markić, B., Bijakšić, S., Bevanda, A., (2016), *Data driven marketing as a challenge for entreprenurship,* REDETE 2016 Conference; Economic development and entrepreneurship in transition economies: Assessment of the last 25 years, going beyond the „transition"; Belgrade, Serbia, October, 2016.

[8]   Tanner, John F. , Raymond, M., A., (2010), *Principles of Marketing 2.0*, University of Minnesota Libraries Publishing.

[9]   Wu, H., Luk, R., Wong, K., Kwok, K., (2008), *Interpreting TF-IDF term weights as making relevance decisions*", ACM Transactions on Information Systems, 26 (3).