

# ETL Design Toward Social Network Opinion Analysis

Afef Walha, Faiza Ghozzi and Faïez Gargouri

**Abstract** Now-a-days, social networking sites have been created lot of buzz in technology world. They are considered as a rich source of information because people share and discuss their opinions about a certain topic freely. Sentiment analysis or opinion mining is used for knowing voice or response of crowd for products, services, organizations, individuals, events, etc. Due to the importance of user's opinions in decisional systems, several Data Warehouse approaches integrate them through a cleaning and transformation processes. However, there is a clear lack of a standard model that can be used to represent the ETL processes. We propose an ETL design approach integrating user's opinion analysis, expressed on the popular social network Facebook. It consists in the extraction of opinion data on Facebook pages (e.g. comments), its pre-processing, sentiment analysis and classification, reformatting and loading into the Data WeBhouse (DWB).

**Keywords** ETL · Opinion analysis · Social network

## 1 Introduction

The Web has dramatically changed the way that people express their views and opinions. They can now post reviews of products at merchant sites and express their views on almost anything in Internet forums, and social networking sites (e.g. Facebook, twitter), which are collectively called user-generated content. This online behavior represents new and measurable sources of information to an organization. For a company, it may no longer be necessary to conduct surveys, organize focus

---

A. Walha (✉)

Multimedia, Information systems and Advanced Computing Laboratory,  
MIR@CL, University of Sfax, Sfax, Tunisia  
e-mail: afef\_walha@yahoo.fr

F. Ghozzi · F. Gargouri

Institute of Computer Science and Multimedia, University of Sfax, Sfax, Tunisia  
e-mail: jedidi.faiza@gmail.com

F. Gargouri

e-mail: faiez.gargouri@gmail.com

groups or employ external consultants in order to find consumer opinions about its products and those of its competitors because the user-generated content on the Web can already give them such information.

With the growing popularity of social networks, millions of users interact frequently and share variety of digital content with each other. They express their feelings and opinions on every topic of interest. These opinions carry import value for personal, academic and commercial applications. Social network sites contain a lot of customers' opinions on certain products that are helpful for decision making. In spite of this importance, there is a clear lack of a standard model that can be used to represent the ETL processes (Extraction, Transformation and Loading) of social networking sites integrating opinion data. In this paper, we propose to design ETL processes for Facebook page interactions.

This paper is organized as follow: Sect. 2 presents a brief review on ETL design and sentiment analysis approaches. Then, we detail our proposed ETL design processes applied on Facebook pages. Finally, we conclude and present some perspectives in Sect. 4.

## 2 Background

This section deals with two main aspects in the literature: ETL design processes and opinion analysis methods and techniques.

### 2.1 *ETL Modeling Approaches*

ETL processes design is a crucial task in DW development due to its complexity and its time consuming. Works dealing with this task [2, 3, 12, 14, 16, 17] can be classified into two main groups: Specific ETL modeling and Standard ETL modeling. The first group [3, 16] offers specific notations and concepts to give rise for new specialized modeling languages. Extraction, transformation and loading processes proposed in [16] are limited to typical activities (e.g. join, filter). El-Sappagh et al. [3] extends these proposals by modeling advanced operations, like user define functions and conversion into structure, etc. In order to design complex ETL scenario, specific modeling approaches propose conceptual and formal models. However, the standardization is an essential asset in modeling. The goal of the second group is to overcome this problem by using modeling languages such as UML and BPMN. Trujillo and Luján-Mora [14] and Muñoz et al. [12] use UML class diagram to represent ETL processes statically or dynamically by using UML activity diagram. Wilkinson et al. [17] and Akkaoui et al. [2] use BPMN standard where ETL processes can be a particular type of business.

Even though ETL processes modeling approaches succeeded in providing interesting several modeling languages, they don't cover opinion data sources available on Web resources like social networks, blogs, reviews, etc.

## 2.2 *Opinion Analysis Approaches*

Opinions are usually subjective expressions that describe people's sentiments, appraisals or feelings toward entities, events and their properties. The concept of opinion is very broad. In this paper, we focus only on opinion expressions that convey people's positive and negative sentiment.

Integrating opinion data is nowadays a hot topic for many researchers. The common goal of opinion analysis approaches is to detect text polarity: positive, negative or neutral. In Medhat et al. [9], categorize sentiment analysis approaches into machine learning and lexicon approaches. Machine learning approaches [1, 18] use classification techniques to classify text (e.g. Naive Bayes (NB), maximum entropy (ME), and Support Vector Machines (SVM)). Lexicon approaches [5, 6, 8, 11, 13] rely on a sentiment lexicon, a collection of known and precompiled sentiment terms. They use sentiment dictionaries with opinion words and match them with the data to determine text polarity. They assign sentiment scores to opinion words according to positive or negative words contained in the dictionary. Lexicon-based approaches are divided into dictionary-based approaches and corpus-based approaches.

A Dictionary-based approach [7, 11] begins with a predefined dictionary of positive and negative words, and then uses word counts or other measures of word incidence and frequency to score all the opinions in the data. The idea of these approaches is to firstly collect manually a small set of opinion words with known orientations (seed list), and then to grow this set by searching in a known lexical DB (e.g. WordNet dictionary) for their synonyms and antonyms. The newly found words are added to the seed list [8]. Opinion words share the same orientation as their synonyms and opposite orientations as their antonyms. In [5, 13], authors use this technique to find semantic orientation for adjectives. Qiu et al. [13] worked on web forums to identify sentiment sentences in contextual advertising. They used syntactic parsing and sentiment dictionary and proposed a rule-based approach to tackle topic word extraction and consumers' attitude identification in advertising keyword extraction.

Corpus based techniques rely on syntactic patterns in large corpora. Corpus-based method can produce opinion words with relatively high accuracy. A corpus-based method needs very large labeled training data. Jiao and Zhou [6] used the Conditional Random Fields (CRFs) methods in order to discriminate sentiment polarity by multi-string pattern matching algorithm applied on Chinese online reviews in order to identify sentiment polarity. They established emotional and opinion words dictionaries.

Machine learning and lexicon approaches use opinion words and classification techniques to determine text polarity. In addition to the use of opinion words to analyze sentiments, emoticons decorating a text can give a correct insight of the sentence or text. For example, the emoticon “☺” expressing “happiness” means positive opinion. Further researchers take care of the increasing using of these typographical symbols for sentiment classification [4, 15]. Vashisht and Thakur [15] identify the possible set of emoticons majorly used by people on Facebook and use them to classify text polarity. Then, they used a finite state machine to find out the polarity of the sentence or paragraph. The problem with this approach is performing sentiment analysis on text-based status updates and comments, disregarding all verbal information and using only emoticons to detect both positive and negative opinions. Hogenboom et al. [4] propose a framework for automated sentiment analysis, which takes into account information conveyed by emoticons. The goal of this framework is to detect emoticons, determine their sentiment, and assign the associated sentiment to the affected text in order to correctly classify the polarity of natural language text as either positive or negative.

Existing ETL design approaches model various web sources without considering user’s opinions available on social networks, reviews, blogs, forums or emails, etc. In the past few years, many researchers have shown interest to opinions expressed by people on any topic. They proposed sentiment analysis methods and techniques to determine text polarity. Some approaches apply classification algorithms and use linguistic features (machine learning approaches). Others use sentiment dictionaries with opinion words and match them with data sources to determine polarity (lexicon-based approaches). These approaches assign sentiment scores to opinion words according to positive or negative words contained in the dictionary. Others researchers use emoticons to disambiguate sentiment when sentiment is not conveyed by any clearly positive or negative words in a text segment.

Sentiment analysis approaches presented in the literature are very helpful and interesting in order to classify a text (positive or negative polarity). In spite of the importance of sentiment classification approaches, we note that few of them employ the coupling between opinion analysis and ETL processes in order to enhance semantic orientation to multidimensional design.

In the current work, we define a new approach of ETL processes design integrating people’s opinions exchanged on Facebook social network. Facebook users express their opinions about any topic freely through opinion words and emoticons. Sentiment analysis is required to classify user opinion. For that, we adopt a lexicon approach based on dictionaries used as lexical DBs in our ETL processes design. We are based on the modeling standard BPMN 2.0 to design Extraction, Transformation and Loading processes because of its completed graphical notation in modeling business processes understandable by all business categories of users [2].

### 3 Proposed ETL Processes Modeling

DWB (Data WeBhouse) sources may include several data types, such as geographic DBs, web sites DBs, web logs, language recognition systems and social networking sites, etc. In order to enrich ETL processes design with semantic orientations, we are interested to opinion data shared and discussed freely on the popular social network Facebook.

Our ETL design approach provides to company's ETL designers a framework integrating costumers' opinions about their products or services. User actions (comments, messages, posts and likes) exchanged within Facebook pages are pertinent for marketing and advertising industry to gather opinions about a particular product. The goal of our ETL design approach is to analyze user actions on product features in order to classify his opinion (positive or negative). To assume this analysis, we are based on verbal cues: opinion words and graphical cues: emoticons. For that, we identify two dictionaries: opinion dictionary composed of opinion words (e.g. best, good) and emoticons dictionary (e.g. :, :(, ;), etc.).

#### 3.1 Lexical DB Description

Opinion and emoticons dictionaries serve as lexical DB in our ETL design approach. Opinion dictionary is composed of opinion words that express desirable (e.g. great, amazing, etc.) or undesirable (e.g. bad, poor, etc.) states. Emoticons dictionary contains positive (e.g. ☺) and negative (e.g. ☹) emoticons majorly used by Facebook users.

Figure 1 illustrates the process of defining our lexical DB. To identify opinion dictionary, we follow a dictionary-based method [8]. Its main idea is to manually collect a small set of terms (seed words), and then search in the well known corpora WordNet [10] of their synonyms and antonyms to enrich them. Then, a manual inspection is carried out to remove or correct errors existing in opinion dictionary. In some texts, opinion word can be related to a modifier term that changes its sentiment polarity (e.g. in the sentence "it is not beautiful", the modifier term "not" changes the sentiment polarity of the opinion word "beautiful"). Also, amplifier terms can increase or decrease the polarity of the affected opinion word (e.g. the word "very" in the sentence "it's very big" increase the polarity of the opinion word "big"). For that, we classify opinion words to two types: modifier terms (like "not" and "very") and carrying-sentiment terms (such as "big", "beautiful").

With the increasing use of emoticons, it is of utmost importance to consider these typographical symbols to discriminate sentiment polarity. So, we collect a set of emoticons majorly used by people on Facebook including positive and negative emoticons defined in [15].

The final step in lexical DB definition process (Fig. 1) is to associate polarity score to each opinion dictionary term and emoticon already defined in opinion and

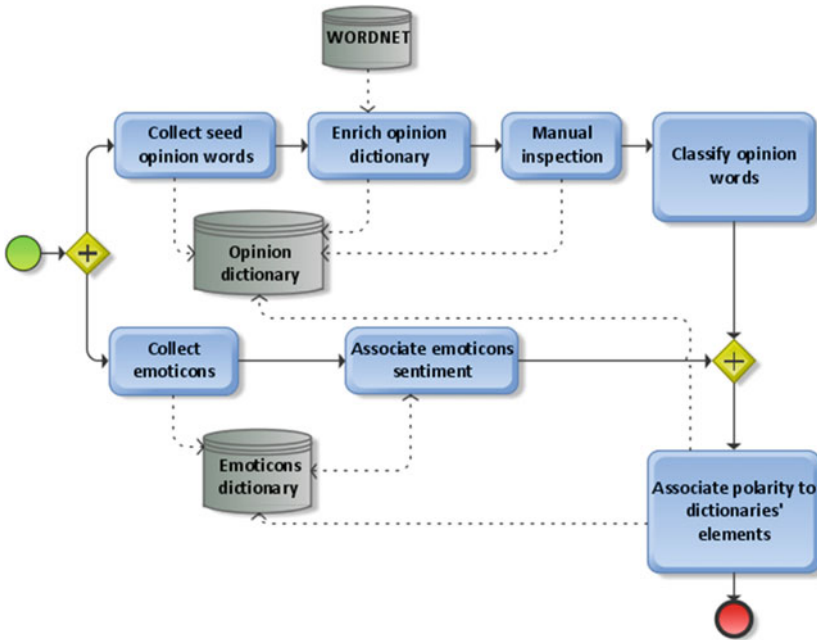


Fig. 1 Lexical DB definition process

emoticons dictionaries. This score has positive (between (0) and (1)) or negative (between (-1) and (0)) value. This real value is determined by linguistic experts according to their sentiment classification. The positive polarity (0.8) is then associated to the opinion word “enjoy” expressing “Happiness” sentiment. Tables 1 and 2 detail examples of carrying-sentiment words and modifiers defined in opinion word dictionary. Moreover, Table 3 shows examples of emoticons and their associated polarities.

Dictionaries defined in this process aims to determine the sentiment polarity of opinions expressed on product features in Facebook pages. Emoticons and opinion dictionaries are used in our ETL processes design to analyze user actions in order to be transformed to DWB model.

### 3.2 ETL Processes Design

Our ETL scenario aims to capture Facebook data through Facebook API graph explorer, bring it to an adapted format and feed the transformed data into the target DWB.

Figure 2 is an overview of the proposed ETL processes: Extraction, Transformation and Loading. These processes are based on the lexical DB (opinion and

**Table 1** Examples of sentiment-carrying words, their associated sentiments and polarity scores

Sentiment-carrying word				Sentiment classification	Polarity score	
Love	Lovely	Loved	Loving	Heart	0.9	
Like					0.8	
Hate		Dislike			-0.8	
Curious					0.5	
Fantastic		Perfect			Quality	1
Fabulous	Best		Excellent	0.9		
Better				0.8		
Good				0.7		
Bad		Badly		-0.8		
Worst				-0.9		
Happy				Happiness		0.8
Excited						0.95
Unhappy				Sadness		-0.7
Depressed					-1	
Amazing				Amused	0.8	

**Table 2** Examples of modifiers and their associated polarity scores

Modifier term				Modifier Polarity
So				1
Totally				0.9
Very				0.8
Small				0.5
Little			Few	0.4
Not	Don't		Didn't	-1

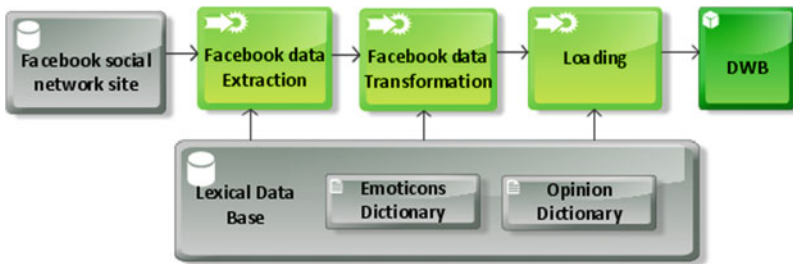
emoticons dictionaries) to analyze user actions expressed on products features within Facebook pages. The result of this analysis is to determine polarity score reflecting user's opinion.

### 3.3 Extraction Step

Extraction step is responsible for capturing data from different sources. According to DWB multidimensional schema (presented in Fig. 8), we aim to analyze user actions associated to posts shared on Facebook pages. A post is an individual entry of a user, page, or group. A list of available actions (comments and likes) is associated to each post. These actions can help to gather people's opinions related to a post.

**Table 3** Examples of emoticons and their associated polarity scores

Emoticon	Sentiment classification	Polarity
:-)      =)      :]	Happiness/smile	0.6
:)		0.7
:-(	Sadness	-0.6
:(		-0.7
:*(		-0.9
:D	Amused	0.9
=D		0.8
<3	Love/heart	1
>:(	Anger	-0.64
(y)	Thumbs up	0.84



**Fig. 2** ETL processes modeling framework

Figure 3 details extraction process. It starts by collecting general information about each Facebook page (page name, website, description, category, etc.). Then, it extracts posts shared on this page. The next step consists in extracting post information including source, message, picture, description, link, and created-time. Finally, this step collects actions (user likes and comments) associated to each post.

Figure 4 illustrates an example of post shared on “Sephora” Facebook page Comments associated to this post are shown in Fig. 5.

### 3.4 Transformation Step

Transformation step tends to make cleaning and conforming on DWB sources (Facebook page actions) to gain correct, complete, consistent, and unambiguous data.

Transformation step is organized in three main steps: pre-processing, analysis and mapping (see Fig. 6).



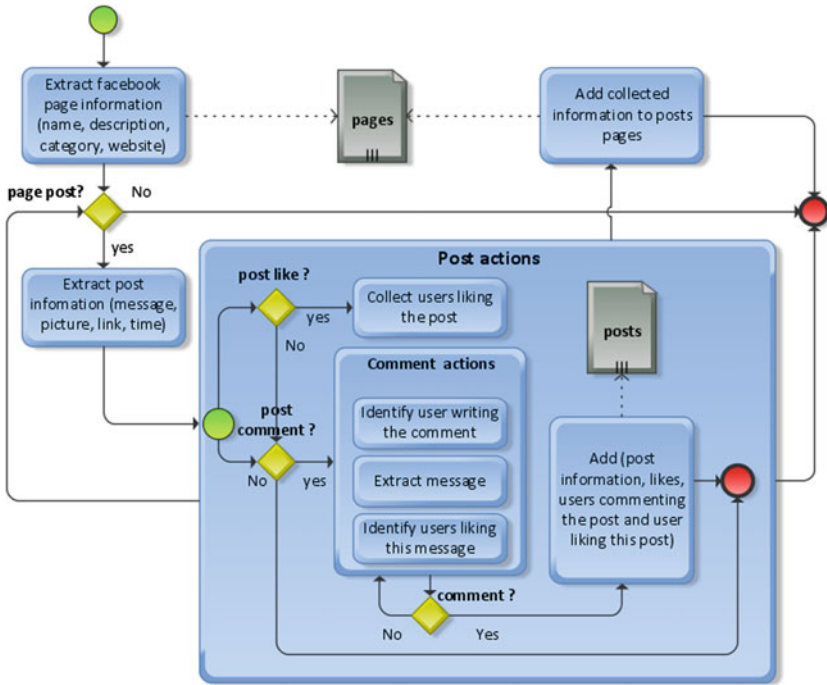


Fig. 3 Facebook page extraction process

```

message = "Milks nourish and hydrate beautifully—and are especially great during drying,
cold weather. Are you a fan?"
created_time = "2015-01-29T02:00:01+0000"
type = "photo"
link = "https://www.facebook.com/Sephora/photos/...../?type=1"
    
```

Fig. 4 Post (P) shared on the page “Sephora”

Pre-processing starts by comments cleaning which replaces all capital letters with small letters and removes diacritics. For example, in comment (1) (Fig. 5), the term “fabuloüs” is replaced by “fabulous”. Then, it identifies each comment word POS (Part-Of-Speech) and its type, i.e., sentiment-carrying or modifying terms [16]. The latter change the sentiment of corresponding opinion word(s) such as negations that change the sentiment sign (e.g. the modifier “not”, used in comment (6), change the sentiment polarity of the opinion term “good”). Also, amplifiers increase the sentiment of the affected sentiment words (e.g. the amplifier “very” in comment (2) modifies the sentiment of the opinion word “good”).

Analysis is the main step of transformation process. It aims to calculate sentiment score of a post (P), i.e.,  $Sent(P_U)$ . This score is equal to the average of comments’ sentiment scores associated to the post (P), as in (1), i.e.

U <sub>1</sub> => C <sub>1</sub> : "I shop Sephora frequently!! Love !!! Love !!! Love !!!"
U <sub>2</sub> => C <sub>2</sub> : "Fabuloüs stuff!! fantastic, I love it."
U <sub>3</sub> => C <sub>3</sub> : "I'm curious to try it"
U <sub>2</sub> => C <sub>4</sub> : "Can't wait to try it! It is very good <3"
U <sub>4</sub> => C <sub>5</sub> : "I'm :( badly product. not good"
U <sub>5</sub> => C <sub>6</sub> : "Excellent sephora!"
U <sub>6</sub> => C <sub>7</sub> : "I LOVE SEPHORA. My husband surprised me with this today when I came home. So excited! And it really smells"
U <sub>5</sub> => C <sub>8</sub> : "(y) Amazing milk!"
U <sub>7</sub> => C <sub>9</sub> : "I am using this product. I'm unsatisfied !!!"
U <sub>4</sub> => C <sub>10</sub> : "I don't like this cosmetic product >:( "
U <sub>3</sub> => C <sub>11</sub> : "I bought It. perfect !!!"
U <sub>2</sub> => C <sub>12</sub> : "LOVELY !!!"

**Fig. 5** Examples of comments associated to the post (P)

$$\text{Sent}(P_U) = \frac{\sum_{j=1}^N \text{Sent}(C_i)}{N} \quad (1)$$

With  $N$  the number of comments ( $C_i$ ) associated to the post ( $P$ ) published by the user ( $U$ ).

To compute sentiment score of the comment message ( $C_i$ ), we propose a lexicon-based method. Its goal is to associate sentiment score to each comment ( $\text{Sent}(C_i)$ ). The principle of this method is the following: if the comment ( $C_i$ ) contains opinion words and emoticons,  $\text{Sent}(C_i)$  is computed as the average of all emoticons' sentiment polarities ( $\text{Sent}(e_{ij})$ ) and polarities of sentiment-carrying words ( $w_{ij}$ ) and their modifiers ( $m_{ij}$ ). Otherwise, if the comment ( $C_i$ ) contains opinion words without visual cues (emoticons),  $\text{Sent}(C_i)$  is calculated as the average of sentiment-carrying words ( $w_{ij}$ ) and their modifiers ( $m_{ij}$ ) polarities (if any,  $\text{Sent}(m_{ij})$  defaults to 0). The sentiment score equation of the  $i$ th comment ( $C_i$ ) is then defined in (2), i.e.,

$$\text{Sent}(C_i) = \begin{cases} \frac{\sum_{j=1}^{v_i} \text{Sent}(e_{ij}) + \sum_{j=1}^{t_i} \frac{|\text{Sent}(m_{ij})| + \text{Sent}(w_{ij})}{2} \times S(\text{Sent}(m_{ij}))}{v_i + t_i} & \text{if } v_i > 0 \\ \frac{\sum_{j=1}^{t_i} \frac{|\text{Sent}(m_{ij})| + \text{Sent}(w_{ij})}{2} \times S(\text{Sent}(m_{ij}))}{t_i} & \text{else,} \end{cases} \quad (2)$$

With  $v_i$  and  $t_i$  correspond respectively to the number of emoticons and the number of sentiment-carrying words used in the comment ( $C_i$ ).  $S(\text{Sent}(m_{ij}))$  depends on the polarity (+/-) of the modifier ( $m_{ij}$ ) related to opinion word. We assign the value (1) to  $S(\text{Sent}(m_{ij}))$  if the modifier polarity is positive ( $\text{Sent}(m_{ij}) > 0$ ). Otherwise, if the ( $m_{ij}$ ) has a negative polarity,  $S(\text{Sent}(m_{ij}))$  is equal to (-1).

Comment sentiment analysis process, described in Fig. 7, details steps to determine comment's sentiment score ( $\text{Sent}$ ). This process starts by computing the number of opinion words ( $t_i$ ) and emoticons ( $v_i$ ) used in the comment ( $C_i$ ), and initializing ( $\text{Sent}$ ) to the value (0). For each opinion word ( $w_{ij}$ ) used in ( $C_i$ ), it searches the mod-

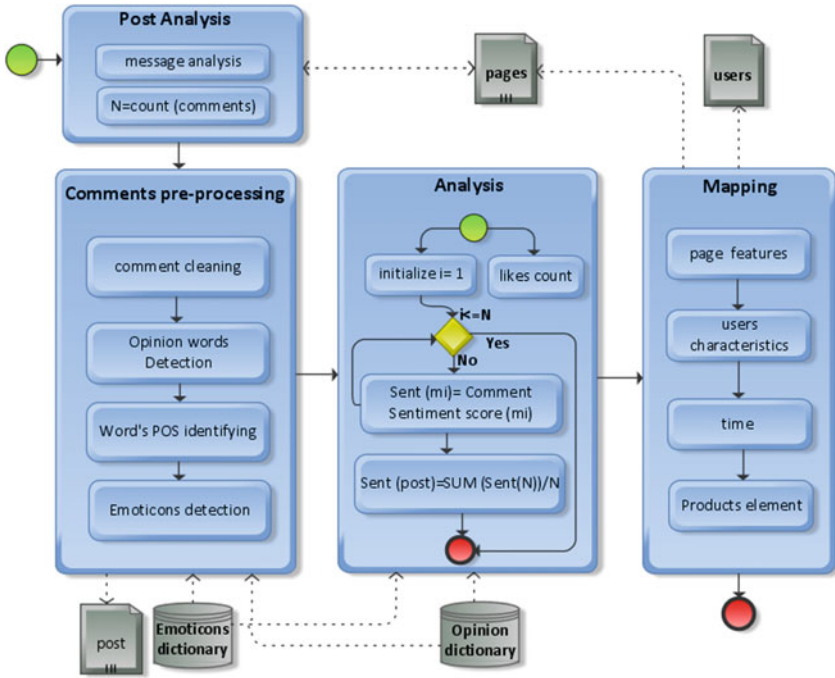


Fig. 6 Facebook page transformation process

ifier ( $m_{ij}$ ) related to this word. If ( $m_{ij}$ ) exists, it recuperates its sentiment score ( $Sent(m_{ij})$ ) defined in opinion dictionary. The absolute value of this score ( $Sent(m_{ij})$ ) is added to ( $w_{ij}$ ) polarity score ( $Sent(w_{w_{ij}})$ ), divided by (2), then multiplied by the modifier score polarity (S), i.e. (1) or (-1) and added to (Sent). Next, his process follows by determining sentiment polarity ( $Sent(e_{ij})$ ) of each emoticon ( $e_{ij}$ ) exploited in the comment ( $C_i$ ) and add it to (Sent). The final step is to determine the final value of Sent, i.e. the average of opinion words and emoticons scores as defined in (2).

To determine users' opinions corresponding to the post (P), we apply "Comment sentiment score" process (Fig. 7) on a set of comments (Fig. 5) associated to this post (Fig. 4). Results are depicted in Table 4. 0Post's sentiment score ( $Sent(P_U)$ ) is computed according to (1).

The final step in transformation process (Fig. 6) is the mapping. Its role is the matching between the source (concepts of "Facebook" model) and the target (DWB multidimensional elements). For example, the attribute "Category" of the class PAGE (source model) corresponds to the parameter "categoryPP" of the dimension FACEBOOK POSTS (DWB multidimensional schema presented in Fig. 8).

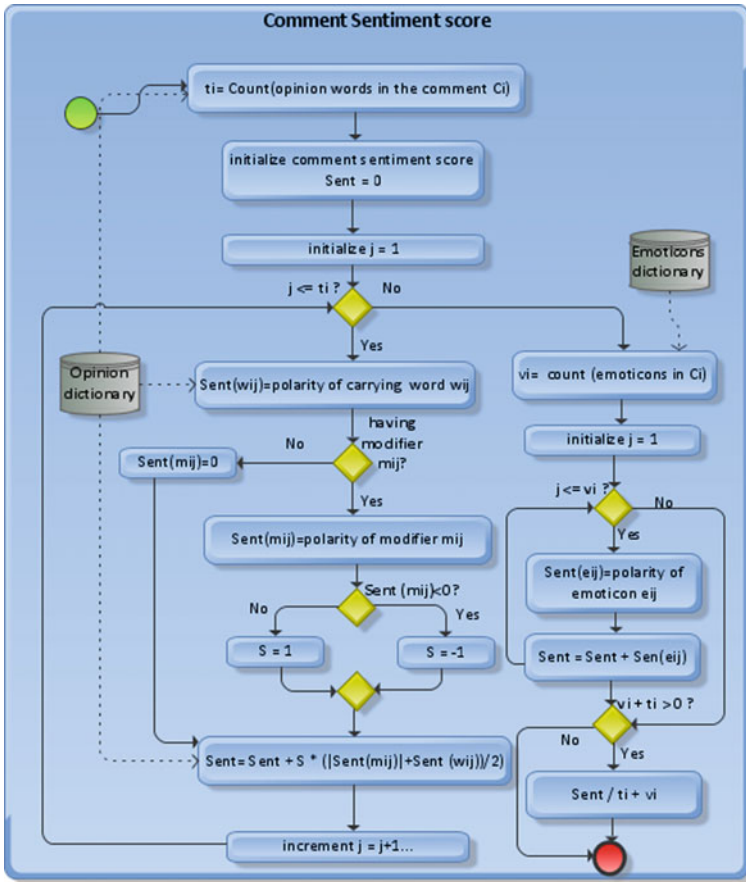


Fig. 7 Comment sentiment analysis process

Table 4 Comments' sentiment polarity

User U	U <sub>1</sub>	U <sub>2</sub>	U <sub>3</sub>	U <sub>4</sub>	U <sub>5</sub>	U <sub>6</sub>	U <sub>7</sub>					
Comment C <sub>i</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>4</sub>	C <sub>12</sub>	C <sub>3</sub>	C <sub>11</sub>	C <sub>5</sub>	C <sub>10</sub>	C <sub>6</sub>	C <sub>8</sub>	C <sub>7</sub>	C <sub>9</sub>
Sent (C <sub>i</sub> )	0.9	0.93	0.87	0.9	0.5	1	-0.78	-0.77	0.9	0.8	0.93	-0.75
Sent (P <sub>U</sub> )	0.9	0.9			0.75		-0.77		0.85		0.93	-0.75

### 3.5 Loading Step

The goal of loading process is to feed the DWB with data resulted from transformation step. It consists in loading data into DWB multidimensional elements including dimensions, measures, facts, attributes and parameters. These elements are illustrated in Fig. 8. The fact POST\_ACTION analyzes user actions (comments and likes)

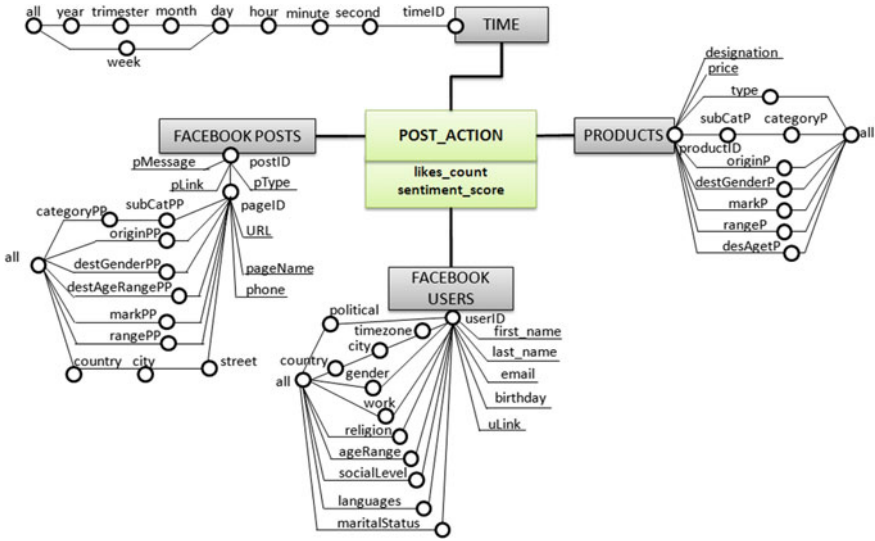
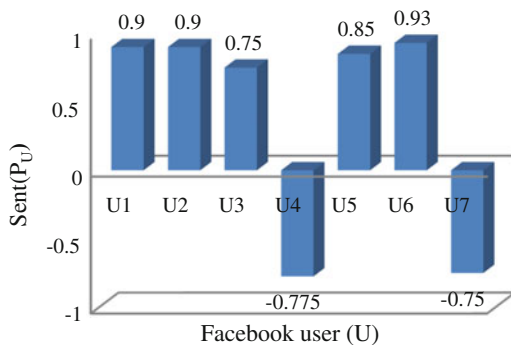
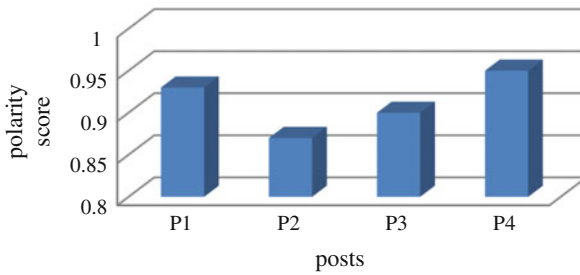


Fig. 8 DWB star schema

associated to a post commented by users on Facebook pages. Decisional makers can then analyze likes\_count and sentiment\_score according to TIME, PRODUCT, FACEBOOK POSTS and FACEBOOK USERS dimensions. For examples, Fig. 9 provides them with sentiment scores resulted from the analysis of users’ comments associated to the post (P) corresponding to the product described in (P) on “Valentine’s” day. Manager can notice that users (U<sub>4</sub>) and (U<sub>7</sub>) have negative opinions. So, he can define user profile interested to this product. Figure 10 shows also analysis results of comments shared by the user U<sub>2</sub> related to four products presented respectively in posts (P1), (P2), (P3) and (P4) during “February”.

Fig. 9 Sentiment Polarity Scores associated to the post (P)





**Fig. 10** Sentiment Analysis of the user “U2” on “February”

## 4 Conclusion and Future Works

Opinions are usually subjective expressions that describe people sentiments and appraisals. Social networks are platforms where millions of users interact frequently and express opinions on every topic of interest. Due to the importance of user opinions to decisional systems, we worked on integrating them DWB design.

We present in this paper a new ETL processes modeling approach using BPMN standard. This approach integrates user opinions expressed by comments shared on the social network Facebook. Its goal is to detect both positive and negative comment polarity. We associate for that a sentiment score depending on comments opinion terms and emoticons. This sentiment analysis is a lexicon method. This analysis is based on opinion and emoticon dictionaries to classify comment polarity.

As future works, we will evaluate our sentiment analysis process on a large test collection of user actions and enrich our lexical DB in order to adapt context-specific opinion analysis. Also, we will extend our ETL processes design approach by integrating more opinion web sources available on web logs, web sites and other social networks.

## References

1. Abbasi, A., Chen, H., Salem, A.: Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.* **26**(3), 1–34 (2008)
2. Akkaoui, Z.E., Mazón, J., Vaisman, A.A., Zimányi, E.: BPMN-Based conceptual modeling of ETL processes. In: 14th International Conference on Data Warehousing and Knowledge Discovery (DaWaK), pp. 1–14 (2012)
3. El-Sappagh, S., Hendawi, H., Bastawissy, A.H.: A proposed model for data warehouse ETL processes. *J. King Saud Univ. Comput. Inf. Sci.* **23**(2), 91–104 (2011)
4. Hogenboom, A., Bal, D., Frasinca, F.: Exploiting emoticons in sentiment analysis. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC), pp. 703–710 (2013)
5. Hu, Y., Li, W.: Document sentiment classification by exploring description model of topical terms. *Comput. Speech Lang. J.* **25**(2), 386–403 (2011)

6. Jiao, J., Zhou, Y.: Sentiment polarity analysis based multi dictionary. In: International Conference on Physics Science and Technology (ICPST) (2011)
7. Kim, S., Hovy, E.: Determining the sentiment of opinions. In: Proceedings of the 20th International Conference on Computational Linguistics (COLING'04) (2004)
8. Liu, B.: *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, 2nd edn. Springer, New York (2011)
9. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng. J.* **5**(4), 1093–1113 (2014)
10. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: WordNet: an on line lexical database. In: *International Journal of Lexicography*, vol. 3. Oxford University Press (1990)
11. Minging, H., Bing, L.: Mining and summarizing customer reviews. In: *Proceeding of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*, pp. 168–177 (2004)
12. Muñoz, L., Mazón, J.N., Trujillo, J.: A family of experiments to validate measures for UML activity diagrams of ETL processes in data warehouse. *Inf. Softw. Technol.* **52**(11), 1188–1203 (2010)
13. Qiu, G., He, X., Zhang, F., Shi, Y., Bu, J., Chen, C.: DASA: dissatisfaction-oriented advertising based on sentiment analysis. *Expert Syst. Appl. J.* **37**(9), 6182–6191 (2010)
14. Trujillo, J., Luján-Mora, S.: A UML based approach for modeling ETL processes in data warehouses. In: *22nd International Conference on Conceptual Modeling—ER. Lecture Notes in Computer Science*, vol. 2813, pp. 307–320 (2003)
15. Vashisht, S., Thakur, S.: Facebook as a corpus for emoticons-based sentiment analysis. *Int. J. Emerg. Technol. Adv. Eng. J. (IJETAJ)* **4**(5), 904–908 (2014)
16. Vassiliadis, P.: A survey of extract–transform–load technology. *Int. J. Data Wareh. Min. (IJDWM)*, **5**(3), 1–27 (2009)
17. Wilkinson, K., Simitsis, A., Dayal, U., Castellanos, M.: Leveraging business process models for ETL design. In: *ER 2010: 29th International Conference on Conceptual Modeling*, November 2010. *Lecture Notes in Computer Science*, vol. 6412, pp. 15–30 (2010)
18. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceeding of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics*, pp. 347–354 (2005)