Extract-Transform-Load Process for Recognizing Sentiment from User-Generated Text on Social Media

Afef Walha[1,2], Faiza Ghozzi[1,3] and Faiez Gargouri[1,3]
[1]MIRACL Laboratory, Sfax, Tunisia
[2]Higher Institute of Information Science and Multimedia of Gabes (ISIMG), University of Gabes, Tunisia
[3]Higher Institute of Information Science and Multimedia of Sfax (ISIMS), University of Sfax, Tunisia

Keywords: Sentiment, Classification, BPMN, Polarity, ETL, Process, Formalization, Social Media.

Abstract: In today's world, business intelligence systems must incorporate opinion mining into their decision-making process. Sentiment analysis of user-generated content on social media has gained significant attention in recent years. This method collects user opinions, feelings, and attitudes toward a topic of interest and helps determine whether their sentiment is positive, neutral, or negative. This paper addresses text classification in sentiment analysis and presents a solution to the Extract-Transform-Load (ETL) process based on a lexicon approach. This process involves gathering media clips, converting them into sentiments, and loading them into a social data warehouse. We provide generic and customizable models to aid designers in integrating pre-processing techniques and sentiment analysis into the ETL process. By formalizing new ETL concepts, designers can create a reliable conceptual design for any ETL process related to opinion data integration from social media.

1 INTRODUCTION

Business Intelligence (BI) systems analyze data for decision-making, while social networks facilitate social interactions. Integrating social network data into BI requires considering user experiences and organizational goals. Businesses should evaluate the advantages and difficulties of utilizing social media data (Sinha et al., 2024). Social media platforms like Facebook, Twitter, and Instagram allow people to share their thoughts and interests through user-generated content (UGC). Companies use this data to improve marketing, customer service, and public relations. UGC, like tweets, has given rise to sentiment analysis (Wankhade et al., 2022). Because it deals with human-generated informal text, this field is complex. Most researchers focus on sentiment analysis, which involves using various cleaning techniques and polarity detection methods to identify opinions. Designers working on a data warehouse (DW) may require assistance integrating UGC from diverse social media sources into their Extract-Transform-Load process (Khan et al., 2024). Therefore, it is crucial to model the ETL process for integrating opinions, regardless of the social platform or topic of interest. This paper proposes ETL4Social-Process-Sentiment, a process for gathering UGC text and converting it into opinions for the Social DW. The paper focuses on modeling complex operations' control and data flows, emphasizing text pre-processing and polarity detection. It also suggests formalizing various ETL concepts, thus enabling designers to reuse and adjust current models or develop new ones that meet their specific requirements. The manuscript is divided into seven sections. Section 2 discusses opinion integration and notable works. Section 3 outlines the sentiment analysis method and proposed model for ETL4Social-Process-Sentiment. Section 4 highlights its complex operation models, while Section 5 formalizes concepts. Section 6 compares our contribution to existing works. Lastly, Section 7 concludes and provides a foundation for future research.

2 MOTIVATION AND RELATED WORK

Social DW (SDW) is a central storage for analyzing UGC from various social media platforms. We suggest a global schema for all media types, as shown in Figure 1. This model enables decision-makers to analyze opinions expressed by OpinionFact according to favorites_count, republished_count, and opin-

Walha, A., Ghozzi, F. and Gargouri, F.
Extract-Transform-Load Process for Recognizing Sentiment from User-Generated Text on Social Media.
DOI: 10.5220/0012706100003687

In *Proceedings of the 19th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2024)*, pages 641-648
ISBN: 978-989-758-696-5; ISSN: 2184-4895

ion_occurrence. The analysis can be done based on several dimensions, such as Media_ClipDim, DateDim, TimeDim, LocationDim, UserDim, SentimentDim, ContextDim, and TopicDim. This present paper focuses on the process that performs complex ETL steps to extract user-generated text, clean it, transform it into opinions, and load it into the SDW, especially to aliment the SentimentDim, which is the user's sentiment regarding a social event. It is associated with a ”positive”, ”negative”, or ”neutral” polarity. The polarity depends on the polarity value, a float on the range of [-1,1], computed based on a sentiment classification algorithm. The Sentiment attribute (belongs to SentimentDim) includes several variants of the polarity, such as ”high positive”, ”positive” and ”low positive” for a ”positive” polarity.

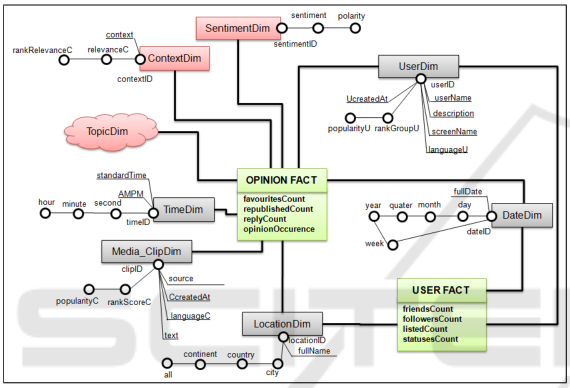

Figure 1: Multidimensional Schema of Social DW.

2.1 Challenges of Matching UGC-Text with Sentiment

Table 1 shows examples of tweets related to ”mobile technology”. We manually analyzed the human-generated text based on sentiment words, emoticons, and other indicators to determine whether the sentiment expressed in the tweets is positive, negative, or neutral. It's important to note that we only explored a small set of tweets. However, how can we proceed when there are hundreds or thousands of tweets to analyze?

2.2 Related Work

2.2.1 Sentiment Analysis Approaches

Sentiment analysis determines the polarity of a text. Methods include sentiment lexicon-based approaches, machine learning-based approaches, and deep learning techniques (Li et al., 2022). The sentiment lexicon-based approaches (e.g., (Darwich et al., 2019), (Ojeda-Hernández et al., 2023)) use a sentiment lexicon, such as dictionaries or corpus, to analyze text polarity. These approaches use annotations that describe how the text matches the lexicons. Machine learning (ML) approaches ((Silva et al., 2022), (Li et al., 2023)) classify data to predict emotional polarity. At the same time, deep learning techniques have been used to identify frequently used models in sentiment analysis research (e.g., (Alamoudi and Alghamdi, 2021), (Su and Shen, 2022)). Although analyzing social media data can provide valuable insights, managing this data still presents challenges. These issues include the need for a clear pattern in conceptual models and the conflicting goals of companies and researchers who use this data.

Table 1: Classification of tweets (”mobile technology”).

Tweet_text	Sentiment	Polarity
WOW the new #Google #Nexus is so beautiful !!! totally boost google's market share in the smartphone. :D	high positive	positive
Houston we have a problem !! My iPad has been restoring for 12+ hours after installing @apple IOS5. This can't be right	low negative	negative
@Apple needs to give me a contract deal i get them new customers all the time #teamiphone PrettyAmaazing	neutral	neutral

2.2.2 Social Media Data Integration Approaches

Several approaches propose frameworks that transform social media data into meaningful, valuable information to enable more effective decision-making. They incorporate these data into the existing multidimensional structures. Several studies from the past five years have recommended incorporating user opinions as an analysis dimension in the DW. (Walha et al., 2017) developed a Twitter DW model to evaluate opinions presented by *Tweet Fact*. It analyzes opinions across dimensions, including Tweet, Date, Time, Location, User, Sentiment, Context, and Topic. Sentiment is categorized as positive, negative, or neutral, and is determined using an opinion analysis algorithm called POLSentiment (Walha et al., 2016). In (Ben Kraiem et al., 2020), a model was developed to analyze tweet data using OLAP. The model examines user activities over time, connections between tweets and respondent users, and tweet sentiment data. The ”DTweet-Metadata” dimension of the model provides opinion data on user sentiment and tweet topics. Tweet-Sentiment data is obtained by counting the tweets for each sentiment category. (Valêncio et al., 2020) proposed a DW model for Facebook and Twitter using a normalized constellation schema to support opinion analysis. The approach involves the ETL stage, which eliminates redundant data to improve performance. The model helps in data acquisition, transformation, and loading and can extract valuable insights when human comprehension falls short. Recently, (Moalla et al., 2022) created a data mart to analyze user opinions on social media platforms such as Facebook, Twitter, and YouTube. They implemented an ETL process with three stages and used a supervised learning classification technique for sentiment analysis (Moalla et al., 2018). Despite consistent experimental results, the study's drawback was the need for further formalization and design of the ETL process steps. Detecting sentiments in UGC text is crucial. Effective preprocessing methods have been defined, but designers need to incorporate them into ETL modeling to align UGC with social DW.

3 SENTIMENT ANALYSIS AND ETL MODELING

We employ the POLSentiment lexicon-based approach to analyze sentiment indicators in text and evaluate polarity values. This method, detailed in (Walha et al., 2016), allows us to determine SentimentDim attributes, as defined in Figure 1.

3.1 Overview of POLSentiment

PolSentiment is sentiment analysis method that uses dictionaries to express positive or negative sentiments in user-generated texts. Opinion dictionaries are formed from sentiment indicators widely used on social media. These indicators can be verbal expressions known as opinion words or graphic symbols known as emoticons. POLSentiment algorithm is divided into three stages, as illustrated in Figure 2. Step (1) in analyzing informal UGC text is to perform a primary "Text Cleaning" phase that removes unknown characters, URLs, punctuations, and repetitive characters. After this, the text is segmented into Tok 1, Tok 2, ..., Tok n through "Text Tokenization". Step (2) identifies sentiment expressions in the UGC text, including emoticons or opinion words. It also considers the possibility of a modifier preceding an opinion word, such as "not," "little," "very," and so on, which can alter the word's meaning in terms of the expressed sentiment. A sentiment and a valence (a float value in [-1, 1]) describe each sentiment indicator. In step (3), called "Opinion analysis", the main goal is to determine the sentiment expressed in the UGC entry. This step is realized by computing the text's polarity based on the valences of the sentiment indicators extracted in Step (2). To achieve this, POLSentiment evaluates the valence of opinion words, their modifiers, and even emoticons to determine the text's polarity accurately. This algorithm outperforms other methods and is evaluated on a pre-annotated dataset. Our primary objective is to model the ETL process to match the UGC text with SentimentDim attributes. This paper proposes an ETL process called ETL4Social-Process-Sentiment that assists designers in integrating the main stages of POLSentiment in a generic and reusable manner.

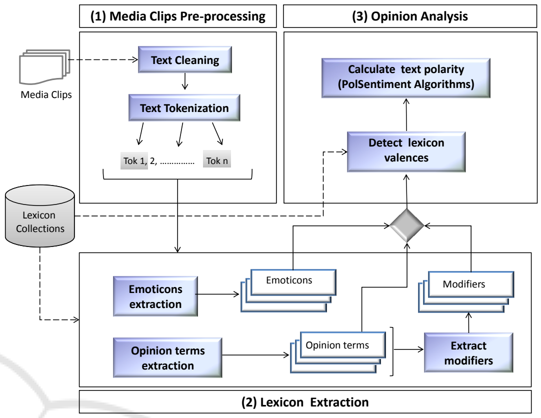

Figure 2: *POLSentiment* steps (Walha et al., 2016).

3.2 ETL4Social-Process-Sentiment Model

ETL4Social-Process-Sentiment is a process that extracts, transforms, and loads user-generated texts into SDW for easier matching with the SentimentDim dimension. Figure 3 shows a diagram of the BPMN[1] model used, including a sequence of ETL operations executed automatically in a specific order. The

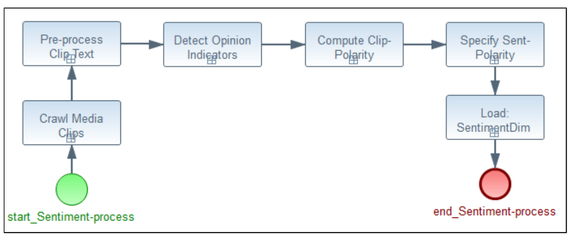

Figure 3: BPMN model of ETL4Social-Process-Sentiment.

Crawl Media Clips operation searches social media

[1] Business Process Modeling and Notation

UGCs, known as Media Clips, published on a specific topic. The SentimentDim dimension gathers data from informal texts, which must be cleaned before use. The Pre-process Clip-Text operation uses well-known cleaning techniques to clean this data. This operation includes a tokenization stage to facilitate sentiment classification, separating words from the clean text. Detecting Opinion Indicators is used to identify sentiment indicators in a given text, such as opinion words and emoticons. Based on these indicators, the Clip-Polarity Compute operation computes the text polarity, ranging from (-1) to (1). The Specify Sent-Polarity operation then uses this value to determine the overall sentiment, such as ”high positive” or ”low positive”, as well as the polarity of the media clip, which can be ”positive”, ”negative”, or ”neutral.” Finally, the Load SentimentDim operation loads the data in this dimension. ETL4Social-Process-Sentiment model is an abstract view of extracting UGCs, transforming them into attributes of the SentimentDim dimension, and loading them into the DW. It efficiently manages the execution order of all the operations involved in this process while each operation receives input data, requires intermediate data, and produces output data. However, it is essential to note that the model does not control the data flow needed to execute these operations. ETL4Social-Sentiment-Operation, on the other hand, focuses on this aspect of modeling the ETL process.

4 MODELING OF ETL4Social-SENTIMENT-OPERATION

The ETL4Social-Process-Sentiment model converts media clip texts into SentimentDim dimension data. ETL operations are specified with BPMN sub-processes, as shown in Figure 3. The ETL4Social-Sentiment-Operation models manage data flow between tasks and activities. These models are detailed in the following sections.

4.1 Pre-Process Clip-Text Operation Model

The data collected through the Media Clips Crawling operation is human-generated text. Therefore, it may contain informal messages with abbreviations, spelling errors, and symbols. This issue can present a significant challenge when performing sentiment analysis on such content. However, we have designed a model specifically for the Pre-process Clip-Text operation, as illustrated in Figure 4, which comprises a set of activities including Get_media_clips, Clean_texts, Tokenize_text, and Store_tokens, each of which is specified by either a BPMN ”Sub-Process” element for composite activities or a BPMN ”Task” item for atomic activities.

Opinion analysis involves various data-cleaning techniques. Figure 4 shows the ”Clean_texts” activity applied to each media clip. This activity is a well-defined sequence of BPMN tasks, including Clean_URLs, Clean_users, Clear_diacritics, Clean_repetitive_letters, and Clean_stop_words. The sole objective of these tasks is to eradicate any trace of URLs, email addresses, user mentions (@username), diacritical signs, repetitive letters (for example, ”veryyyy” being converted to ”very”), and stop words (such as ”the” and ”a”). As a result of this highly efficient process, we obtained a collection of cleaned texts. Next, the ”Tokenize_texts” activity segments the cleaned text by breaking it down into individual words or symbols known as tokens. These tokens are stored in a temporary data object (clip_c) and then used to identify opinion indicators in the clip text.

4.2 Detect Opinion Indicators Model

POLSentiment efficiently categorizes the sentiment of text-based content by analyzing opinions, modifiers, and emoticons. It identifies sentiment indicators among the tokens present in the text. The model that controls the data flows of the Detect Opinion Indicators operation is depicted in Figure 5. The first step is the ”Detect_indicators” activity, which helps determine whether a given token is an opinion word or an emoticon. It utilizes the ”Is_an_opinion_word” and ”Is_an_emoticon” tasks, which query the opinion dictionaries ”Dict_Opinion_Words” and ”dict_Emoticons” respectively.

”Is_a_modifier” activity thoroughly searches for a modifier before an opinion word within the text segments. If emoticons and opinions are present in the text, the ”Store_emoticon” and ”Store_modifier” activities will temporarily store them in the objects ”clip_emoticons” and ”clip_opinion_words”, which will eventually be stored in the DSA.

4.3 Compute Clip-Polarity Model

Based on (Walha et al., 2016), the Compute Clip-Polarity operation calculates text polarity (TPol) using valences of opinion words, modifiers, and emoticons. TPol is determined by initializing the polarities of polarities generated by opinion words (Tpol_W) and emoticons

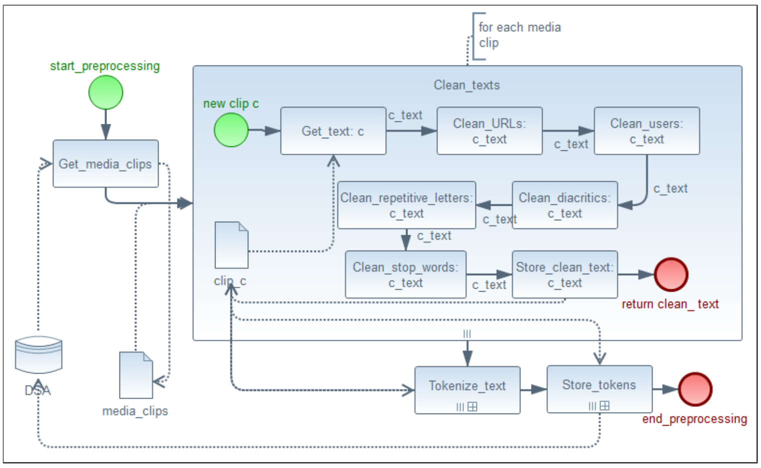

Figure 4: BPMN model of Pre-process Clip-Text operation.

(TPol_E) to 0, retrieving "clip_emoticons" and "clip_opinion_words," and running three activities in parallel: "Compute_opn_words_polarity", "Compute_emoticons_polarity", and "Count_opinion_indicators". The last activity counts opinion words and emoticons in the input text, represented by count_ind.

The TPol score considers the valence of emotions and opinions. The "Compute_emoticons_polarity" operation adds Tpol_E to TPol. "Compute_opn_words_polarity" returns Tpol_W, calculated based on opinion valences and modifiers. A BPMN gateway called "has modifier?" checks for a modifier (m) for the word (w). Val_w (the valence of w) is added to Tpol_W if there is no modifier. Otherwise, Val_w and Val_m are considered. If Val_w is negative, the opposite of Val_w and the absolute value of (Val_m) are added to Tpol_W. Otherwise, this step is completed by carrying out the "Average: Neg(Val_w), Abs(Val_m)" activity. If (Val_m) is negative, its opposite is added to Tpol_W. To determine a text's polarity, we add the polarity values of opinion words and emoticons and divide the total by the count_ind. Due to diverse social media platforms, designers of social data warehouses may need help with the ETL process. Identifying specific ETL concepts can help modify or create new models. Our method includes sentiment analysis and introduces new social ETL modeling concepts.

5 FORMALIZATION OF ETL4Social-SENTIMENT CONCEPTS

5.1 "O-Activity" Concept

"O-Activity" is the executable component in a social ETL operation.This concept has three variations: "Social Activity", "Standard Activity", and "Semantic Activity".

Definition 1. O-Activity
O-Activity (O_A) Is Defined with the n-uplet ($Name^{O\_A}$, $InpFlObj^{O\_A}$, $OutFlObj^{O\_A}$, $InpData^{O\_A}$, $OutData^{O\_A}$)**, Where:**

- $Name^{O\_A}$**: Is the Name of the Activity (O_A),**
- $O-Model\_Name$**: Is the Name of the ETL Operation Model Containing this Activity**
- $InpFlObj^{O\_A}$**: Is the Input Flow Object of O_A. this Object Might Be an Instance of an Activity (*O-Activity*), a Gateway, or an Event,**
- $OutFlObj^{O\_A}$**: Is the Output Flow Objects of O_A,**
- $InpData^{O\_A} = \{Din_1;...;Din_k\}$**: Is a Set of Input Data Objects (O-Data) Used to Execute O_A,**
- $OutData^{O\_A} = \{Dout_1;...;Dout_l\}$**: Is a Set of Data Resulting from O_A Execution.**

5.2 "Social Activity" Concept

"Social Activity" is a step of the ETL operation for UGC data mapping into social DW, and it can be ei-

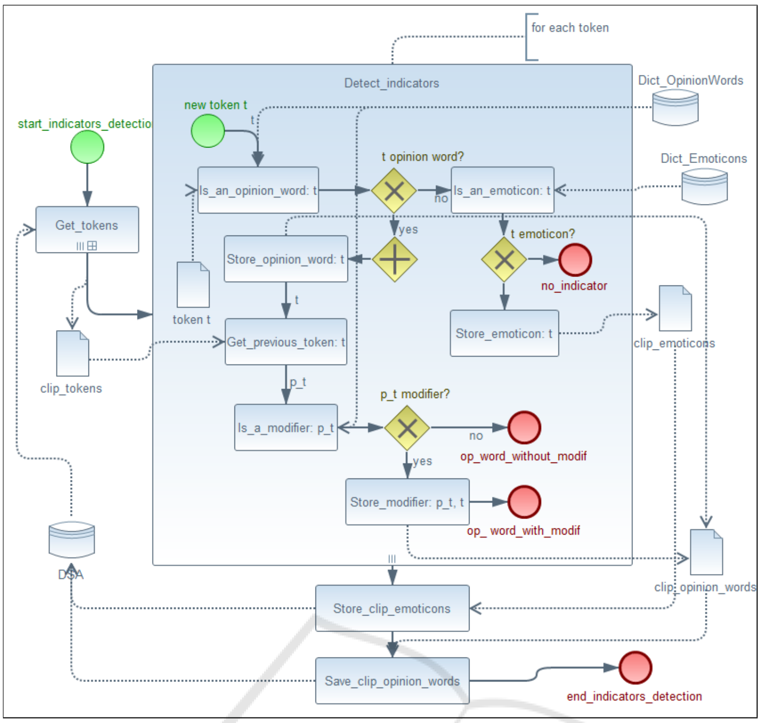

Figure 5: BPMN model of Detect Opinion Indicators operation.

ther standard or semantic.

Definition 2. Social Activity
A Social Activity (A_SO) Is an n-tuple ($Name^{A\_SO}$, $InpFlObj^{A\_SO}$, $OutFlObj^{A\_SO}$, $InpData^{A\_SO}$, $OutData^{A\_SO}$, $StdAct^{A\_SO}$, $SemAct^{A\_SO}$)**:**

- **the Parameters** $Name^{A\_SO}$, $InpFlObj^{A\_SO}$, $OutFlObj^{A\_SO}$, $InpData^{A\_SO}$, $OutData^{A\_SO}$ **Are Already Described in (Definition 1)**
- $StdAct^{A\_SO} = \{Ast_1;...;Ast_n\}$**: a Set of Standard Activities Parts of the Activity (A_SO),**
- $SemAct^{A\_SO} = \{Asm_1;...;Asm_m\}$**: a Set of Semantic Activities of A_SO.**

Example 1. "Detect_Indicators" is a social activity that identifies emoticons and opinion words in a UGC text. It involves standard tasks like "Get" and "Store" and semantic tasks, such as identifying opinion words and emoticons, under the "Detect_Opinion_Indicators" operation (Figure 5).
"Detect_Indicators" is defined as follows:

$A\_SO^{DetIndic}$ = ("Detect_Indicators", $InpFlObj^{DetIndic}$, $OutFlObj^{DetIndic}$, $InpData^{DetIndic}$, $OutData^{DetIndic}$)

with:

- $InpFlObj^{DetIndic} = \{Get\_tokens\}$,
- $OutFlObj^{DetIndic} = \{Store\_clip\_emoticons\}$,
- $InpData^{DetIndic} = \{Inpclip\_tokens\}$,
- $OutData^{DetIndic} = \{clip\_emoticons, clip\_opinion\_words\}$.
- $StdAct^{DetIndic} = \{Store\_opinion\_word, Get\_previous\_token, Store\_modifier, Store\_emoticon\}$.
- $SemAct^{DetIndic} = \{Is\_an\_opinion\_word, Is\_a\_modifier, Is\_an\_emoticon\}$.

5.3 "Standard Activity" Concept

ETL processes involve common activities like "Get", "Join", "Merge", and "Store" used to promote structured data processing.

Definition 3. Standard Activity
"Standard Activity (A_ST)" Is a Variant of the "O-Activity" Concept, Defined by the n-uplet ($Name^{A\_ST}$, $InpFlObj^{A\_ST}$, $OutFlObj^{A\_ST}$, $InpData^{A\_ST}$, $OutData^{A\_ST}$)***.***

Example 2. The activity "Get_previous_token" in the "Detect_Indicators" social activity, described in Example 1, is an instance of the "Standard Activity"

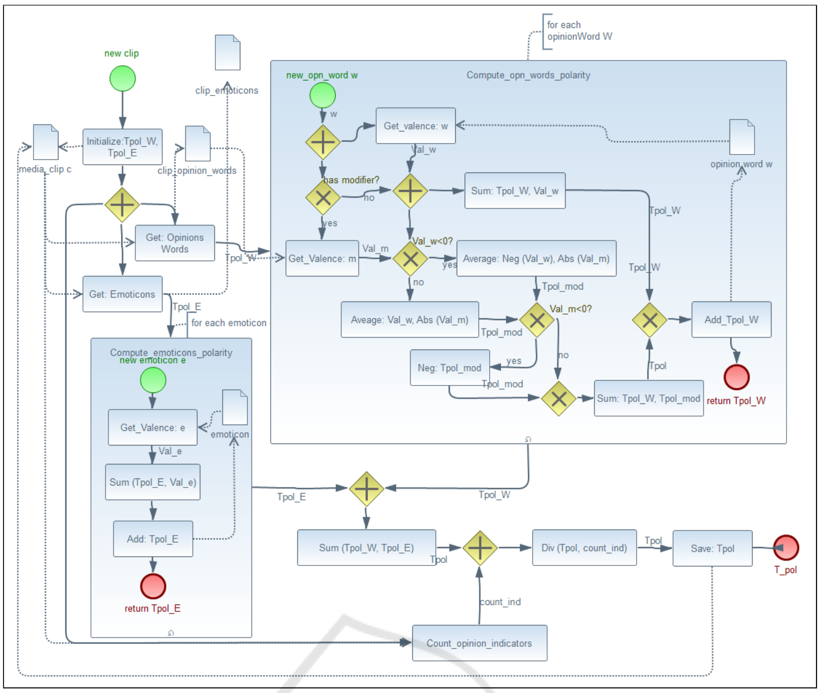

Figure 6: BPMN model of Compute Clip-Polarity operation.

concept. It is defined as:

$A\_ST^{GetPrevT} = (\text{"Get\_previous\_token"}, InpFlObj^{GetPrevT}, OutFlObj^{GetPrevT}, InpData^{GetPrevT}, OutData^{GetPrevT})$

with:

- $InpFlObj^{GetPrevT} = \{Store\_opinion\_word:\ t\}$,
- $OutFlObj^{GetPrevT} = \{Is\_a\_Modifier:\ p\_t\}$,
- $InpData^{GetPrevT} = \{t,\ clip\_tokens\}$,
- $OutData^{GetPrevT} = \{p\_t\}$.

5.4 "Semantic Activity" Concept

"Semantic Activity" transforms human-generated text into decisional information using lexical or semantic resources.

Definition 4. ***Standard Activity "Semantic Activity (A_SE)", a Variant of the "O-Activity", Is Defined by the n-uplet*** $(Name^{A\_SE}, InpFlObj^{A\_SE}, OutFlObj^{A\_SE}, InpData^{A\_SE}, OutData^{A\_SE})$.

Example 3. "Is_an_emoticon" is an activity under "Detect_Indicators", as defined in Example 1 and shown in Figure 5. This activity used for identifying emoticons in text is summarized as follows:

$A\_SE^{IsAnEmot} = (\text{"Is\_an\_emoticon"}, InpFlObj^{IsAnEmot}, OutFlObj^{IsAnEmot}, InpData^{IsAnEmot}, OutData^{IsAnEmot})$

with:

- $InpFlObj^{IsAnEmot} = \{opinion\ word?\}$,
- $OutFlObj^{IsAnEmot} = \{emoticon?\}$,
- $InpData^{IsAnEmot} = \{t,\ Dict\_Emot\}$,
- $OutData^{IsAnEmot} = \{resp\}$.

6 DISCUSSION

Table 2: ETL4Social-Sentiment Vs. Opinion Integration Approaches.

Approaches	Social Media	ETL Design	ETL concept	SA method
(Walha et al., 2017)	TW	yes	yes	no
(Ben Kraiem et al., 2020)	TW	no	no	no
(Valêncio et al., 2020)	FB, TW	no	no	no
(Moalla et al., 2022)	FB, TW YT	no	no	ML
ETL4Social-Sentiment	Global SDW	yes	yes	POLSent-iment

This paper presents a design solution for incorporating sentiment analysis of UGC into the SDW, reducing computational costs, and enabling opinion discovery. We compared our proposal with existing approaches in Section 2.2.2. Table 2 shows the results based on specific criteria.

- Social Media. It lists social media platforms such as Twitter (TW), Facebook (FB), and YouTube(YT) used by the DW and ETL models.

- ETL Design. It verifies if the approach proposes ETL process model to map UGC into sentiment.
- ETL concepts. It checks whether the approach defines or formalizes the ETL concepts.
- SA method. It determines if the sentiment is generated based on a valid sentiment analysis method.

Although some approaches have proposed solutions to integrate opinions from text UGCs, we note the need to model ETL processes to transform UGC text into DW. In this context, our contribution addresses this problem at a conceptual level. The models (cf. Sections 3.2 and 4) serve as design patterns for opinion data integration and simplify the ETL designer's task. Morover, the ETL4Social-Sentiment-Process and operation models apply to all social media types for sentiment analysis. It provides ETL designers with a standardized approach to optimizing social ETL processes using formalized concepts.

7 CONCLUSION

Integrating opinion data from unstructured text sources into a decisional system can be challenging when designing ETL processes. A social data warehouse can help with this. However, careful handling of user-generated content is required to identify sentiment. Our research aimed to develop practical approaches for sentiment analysis on social media. We proposed design models for the ETL4Social-Sentiment process and operations. These models handle activities and data to match UGC text with the SentimentDim dimension of the SDW. The models are generic and customizable based on the ETL-formulized concepts. Big data sources require powerful ETL tools that are efficient in execution cost, transformations, and parallel data processing. To improve our proposal, we must use MapReduce as a distributed execution framework to process big data in parallel, saving time and reducing the risk of errors.

REFERENCES

Alamoudi, E. S. and Alghamdi, N. S. (2021). Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings. *Journal of Decision Systems*, 30(2-3):259–281.

Ben Kraiem, M., Alqarni, M., Feki, J., and Ravat, F. (2020). Olap operators for social network analysis. *Cluster Computing*, 23:2347–2374.

Darwich, M., Mohd, S. A., Omar, N., and Osman, N. A. (2019). Corpus-based techniques for sentiment lexicon generation: A review. *J. Digit. Inf. Manag.*, 17(5):296.

Khan, B., Jan, S., Khan, W., and Chughtai, M. I. (2024). An overview of etl techniques, tools, processes and evaluations in data warehousing. *Journal on Big Data*, 6.

Li, J., Zhang, Y., Li, J., and Du, J. (2023). The role of sentiment tendency in affecting review helpfulness for durable products: nonlinearity and complementarity. *Information Systems Frontiers*, 25(4):1459–1477.

Li, S., Liu, F., Zhang, Y., Zhu, B., Zhu, H., and Yu, Z. (2022). Text mining of user-generated content (ugc) for business applications in e-commerce: A systematic review. *Mathematics*, 10(19):3554.

Moalla, I., Nabli, A., and Hammami, M. (2018). Towards opinions analysis method from social media for multidimensional analysis. In *MoMM*, pages 8–14.

Moalla, I., Nabli, A., and Hammami, M. (2022). Data warehouse building to support opinion analysis in social media. *Social Network Analysis and Mining*, 12(1):123.

Ojeda-Hernández, M., López-Rodríguez, D., and Mora, Á. (2023). Lexicon-based sentiment analysis in texts using formal concept analysis. *International Journal of Approximate Reasoning*, 155:104–112.

Silva, L. M. M., Valêncio, C. R., Zafalon, G. F. D., Columbini, A. C., Filipe, J., Smialek, M., Brodsky, A., and Hammoudi, S. (2022). Feature selection with hybrid bio-inspired approach for classifying multi-idiom social media sentiment analysis. In *ICEIS*, pages 297–307.

Sinha, S., Narayanan, R. S., and Rakila, R. (2024). Harnessing sentiment analysis methodologies for business intelligence enhancement and governance intelligence evaluation. *Journal of Intelligent Systems and Applications in Engineering*, 12(11s):166–176.

Su, Y. and Shen, Y. (2022). A deep learning-based sentiment classification model for real online consumption. *Frontiers in Psychology*, 13:886982.

Valêncio, C. R., Silva, L. M. M., Tenório, W., Zafalon, G. F. D., Colombini, A. C., and Fortes, M. Z. (2020). Data warehouse design to support social media analysis in a big data environment. *Journal of Computer Science*, pages 126–136.

Walha, A., Ghozzi, F., and Gargouri, F. (2016). A lexicon approach to multidimensional analysis of tweets opinion. In *AICCSA*, pages 1–8. IEEE.

Walha, A., Ghozzi, F., and Gargouri, F. (2017). Etl4social-data: Modeling approach for topic hierarchy. In *KEOD*, pages 107–118.

Wankhade, M., Rao, A. C. S., and Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7):5731–5780.