Chapter

Techniques and Tools for Information Extraction: Application to Social Media

Lossan Bonde and Severin Dembele

Abstract

This chapter presents a comprehensive and systematic review of the various techniques and tools used for information extraction (IE), focusing on social media. The aim of the review is triple: (1) identify the techniques and tools used for IE applied to social media, with a recommendation on when to use these techniques and tools; (2) determine what applications along with the social networks use IE from social media; and (3) identify the challenges associated with IE from social media. We combined Kitchenham and the PRISMA methodologies to conduct the systematic review. Guided by a review protocol and quality assessment, 58 papers were selected and analysed. This review highlights the potential of social media information extraction, emphasizing the central role of Machine Learning alongside emerging techniques like transformers. Applications span sentiment analysis, health, security, marketing, and misinformation detection. Twitter dominates research, though other platforms are studied. Key challenges include ethics, privacy, data scarcity, accuracy, and computational costs, necessitating refined methods and ethical focus.

Keywords: information extraction, social media content processing, transformer-based models, large language models, natural language processing, machine learning, deep learning

1. Introduction

Information extraction (IE) from text, image, audio, and video content is essential in nearly all areas of business, including marketing, business decision-making, industrial processes, education, and research. Natural language processing (NLP) is the field of study that deals with information extraction. It has gone through different stages from the early rule-based systems to modern machine learning and deep learning approaches, and is now transitioning to Generative Artificial Intelligence (Gen AI). Each stage has contributed to strengthening the techniques and tools, making the field increasingly mature.

The subject of IE applied to social media is fascinating, given the increasing number of social networks and the massive amount of data generated from them. Innovative applications in Public Health Surveillance, Crisis and Disaster Management, Brand and Sentiment Analysis, Misinformation and Fake News Detection, ... are impacting individual lives, businesses, and organizations as a result of using IE from social media. The purpose of this chapter is triple: (1) identify the techniques and tools used for IE applied to social media, with a recommendation on when to use these techniques and tools; (2) determine what applications along with the social networks use IE from social media; and (3) identify the challenges associated with IE from social media.

To achieve the objectives, we combined Kitchenham and the PRISMA methodologies to conduct a systematic review. The Kitchenham method was used to plan and conduct the review, while the PRISMA guidelines have served as the basis for analysis and reporting. In particular, we adopt this pairing to capitalize on their complementary strengths: Kitchenham provides rigor in protocol design, search strategy, and quality assessment, whereas PRISMA adds transparency and reproducibility in screening, inclusion/exclusion, and flow reporting. This integrated approach—still uncommon in IE reviews—both justifies and improves our methodological contribution by reducing selection bias, shortening screening cycles through a structured flow, and leaving an auditable trail that makes the review efficient to replicate and extend. Following a well-defined methodology, the research objectives were translated into five research questions, which guided the search for relevant sources. Guided by the review protocol and the quality assessment, 58 papers were selected and analyzed. The review confirms the immense potential of social media information extraction while acknowledging substantial obstacles. Future research must not only refine existing techniques but also proactively address ethical and practical dimensions, ensuring that the full benefits of this technology can be realized in an increasingly connected and data-driven world.

The remaining part of the chapter is organised into four sections. Section 2 describes in detail the methodology. Section 3 presents the data, data analysis, and the insights drawn from the review. Section 4 discusses the conclusions of the review. Finally, Section 5 concludes the chapter and gives some recommendations.

2. Background of the study

Information extraction has evolved significantly over time, moving from traditional basic keyword searches to sophisticated AI-powered techniques. Early techniques relied on simple text analysis and keyword matching. Advancements in areas like deep learning (DL), natural language processing (NLP), and machine learning (ML) have enabled more nuanced and accurate information extraction techniques and tools. Results of these advancements include identifying entities, relationships, sentiment, and even complex events from social media data.

Figure 1 reveals a significant growth in the number of publications over the years, reflecting the increasing interest in information extraction, particularly in the context of social media and applications of machine learning and deep learning. This upward trend highlights the growing importance and intense research activity in the field of information extraction, with a notable surge in interest starting in 2020. This increase is likely influenced by the acceleration of digital transformation and the rise in data generated through social media platforms.

This constant and rapid growth of research activities around the topic of information extraction from social media has motivated our interest in conducting a systematic literature review with the aim to answer to the following question: What are the most effective techniques and tools for information extraction from social media platforms, and what are their main challenges?

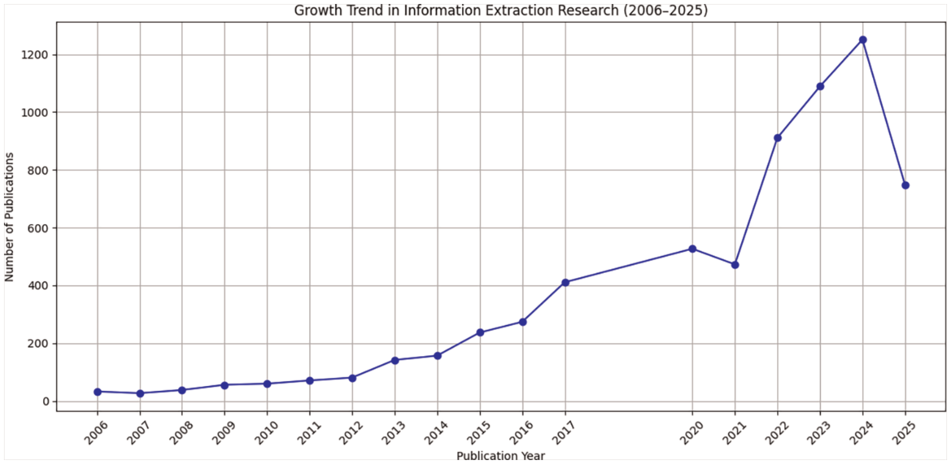

Figure 1.
Trend in information extraction research.

3. Methodology

The methodology of this review combines `Kitchenham` [1] and the `PRISMA` [2] methodologies. `Kitchenham` methodology was selected to conduct the review because it provides a structured and repeatable process for identifying, evaluating, and synthesizing research to answer specific research questions. The `PRISMA` methodology was used for the reporting of the findings. **Figure 2** presents the methodology of this study.

The review is a three-step process: (1) Planning the review, (2) Conducting the review, and (3) Reporting the results. Steps 1 and 2 are based on the `Kitchenkam`'s methodology, while step 3 follows the `PRISMA` guidelines. The latter is more robust in the analysis. The proposed methodology provides a comprehensive overview of existing research, identify key trends, evaluate current approaches, and formulate recommendations for future work.

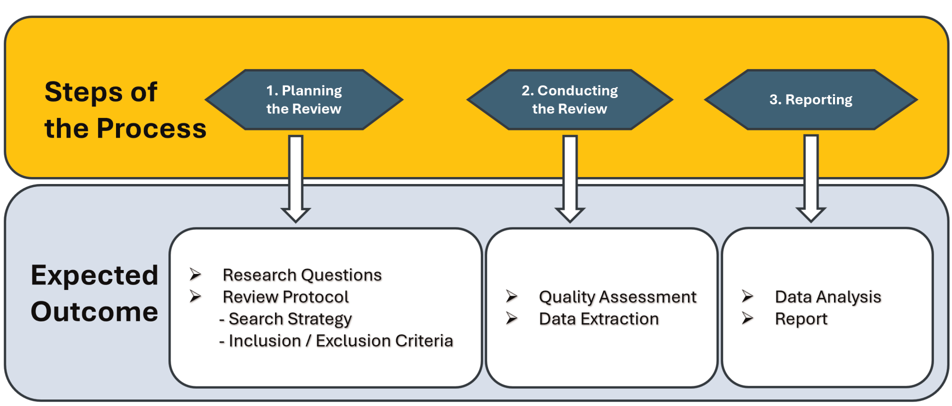

Figure 2.
Literature review methodology.

3.1 Planning the review

Two key activities are done at the planning stage: developing the research questions and developing a review protocol.

3.1.1 Research questions

1. *RQ1: What applications are making use of information extraction from social media*? Through this question, the study seeks to determine the main areas where information extraction from social media is used. Knowledge of the areas of application helps determine whether extracting information from social media is of importance for future research.

2. *RQ2: What social networks are generally used in information extraction literature?*

3. *RQ3: What are the techniques and tools used for information extraction from social media?* A survey of the techniques and tools is crucial as it informs on the effectiveness and efficiency of the current technologies related to information extraction from social media.

4. *RQ4: What techniques and tools are most appropriate for what applications and for what type of information?* The answer to this question guides users on what technique and tool to choose based on their application and the type of data they want to extract.

5. *RQ5: What are the challenges of information extraction from social media?* The knowledge of the current challenges helps in defining future research directions.

These questions contribute to developing the review protocol by informing the search strategy to be used. In particular, they serve as the basis for selecting the search keywords and terms.

3.1.1 Review protocol

The review protocol aims to establish how the study will identify and select relevant studies to include in the review. Developing a review protocol involves two key components: a search strategy and the inclusion and exclusion criteria.

Search strategy: Our search strategy relies on key thematic axes found in the topic. These axes are expanded into several search terms as formulated in **Table 1**.

IEEE Xplore was used to search the sources. The formulation of the search strings is as follows:

IEEE Xplore:

```
("Document Title":"social media" OR "Document
Title":"online platforms" OR "Document Title":"blog") AND
("Document Title":"Information Extraction" OR "Document
Title":"topic modeling") AND ("Document Title":"machine
learning" OR "Document Title":"deep learning")
```

Keyword-based search does not always guarantee complete coverage of all the relevant sources. Kitchenkam and Charters [ref] recommend extending the search

Thematic axes	Search terms
Social media	Social networks, Twitter, Facebook, Instagram,TikTok, YouTube, blogs, online platforms
Information extraction	Content extraction, text mining,opinion mining, sentiment analysis, event detection, topic modelling
Techniques and tools	Natural language processing (NLP), machine learning, deep learning, data mining, social network analysis

Table 1.
Thematic axes and search terms.

space. For this study the extension strategy involves selecting 10 sources randomly from those that have been cited at least 20 times and conducting a one-iteration backwards snowballing

Inclusion and exclusion criteria: The inclusion and exclusion criteria of this study are based on five factors:

- *Language*: The study considered only papers published in the English language.
- *Peer-review:* The study considered only peer-reviewed publications.
- *Originality*: The study considered only primary studies, excluding secondary and tertiary studies, including reviews, meta-analyses, and surveys.
- *Thematic*: The extension strategy may have included studies that treat information extraction but not on social media or vice versa. Sources included in the study must explicitly address information extraction in the specific context of social media.
- *Date of publication*: The year 2006 is important in the world of social media for two reasons: (1) It is the year `twitter`, now rebaptized `X`, was created, and (2) it is in 2006 that `Facebook` was made available to the public. Therefore, we consider 2006 as the starting reference year for our study.

With the considerations above, the inclusion and exclusion criteria are presented in **Table 2**.

Factors	Inclusion criteria	Exclusion criteria
Language	English	Not English
Peer-review	Peer-reviewed	Non peer-reviewed
Originality	Primary study	Secondary or tertiary study
Thematic	Addresses information extraction in the specific context of social media	Does not address both social media and information extraction
Date of Pub.	Published in 2006 or beyond	Published before 2006

Table 2.
Inclusion and exclusion criteria.

#	Quality criteria	Outcome
QC1	Is there a research methodology used?	Yes: The methodology is well described
		No: There is no or poor methodology
		Partially: The methodology is fairly described
QC2	Does it describe clearly, a technique or tool for IE	Yes: Description is detailed
		No: There is no meaningful description
		Partially: The description is missing details
QC3	Is the technique properly evaluated?	Yes: The evaluation is complete
		No: The evaluation is poor
		Partially: The evaluation is fair
QC4	Is there a specific tool used?	Yes: A tool is used with justification
		No: There is no tool used
		Partially: A tool is used, but the choice is not justified
QC5	Does the study include Challenges or Limitations?	Yes: Challenges/Limitations presented
		No: No clear indication of challenges
		Partially: The challenges/limitations are fair

Table 3.
Quality assessment criteria.

3.2 Conducting the review

The review is conducted in a three-step process, starting with the application of the review protocol described in Section 3.1.2, followed by the quality assessment of the sources, and concluding with data extraction.

3.2.1 Quality assessment

All the studies that passed the inclusion and exclusion criteria are subjected to quality assessment. The quality assessment requires an analysis of the full text of each study to determine its potential to provide meaningful insights for this review. We adopted a scoring system used in Ref. [1] to assign a score to each study. The scores are based on five quality criteria described in **Table 3**. The possible outcome for a study and for each quality criterion is either `Yes`, `No`, or `Partially`. The table describes how the outcome is determined for each criterion. Each possible outcome is assigned a value: 1 for "Yes," 0 for "No," and 0.5 for "Partially." With this scoring system, a study that fulfils all five criteria will score 5. Then, we set the acceptance score at 4 (i.e., 80% of the perfect score).

In summary, all the studies retained after applying the inclusion and exclusion criteria are scored using the five criteria, and the scoring system, and those scoring less than 4 are rejected.

3.2.2 Data extraction

For the studies that meet the quality assessment minimum score of 4, the following data were extracted:

- The full reference of the study.
- The application area of the study. The application areas include: Sentiment/Emotion Analysis and Opinion Mining, Event Detection and Crisis Monitoring, Security and Law Enforcement, Market and Business Intelligence, Political Analysis and Election Monitoring, Public Health Surveillance, Fake News and Misinformation Detection, Academic and Social Research, E-Commerce and Recommendation Systems, Smart Cities and Mobility Analysis, and Environmental Monitoring.
- The data sources: the list of social Networks involved.
- The techniques used.
- The tools used.
- The challenges/limitations.

The extracted data are tabulated and analysed following the PRISMA analysis guidelines. The actual analysis for this review will be presented in Section 4.

4. Results and discussions

Upon applying the search strategy outlined in Section 3.2, the steps and corresponding number of sources are depicted in **Figure 3**. An initial list of 7586

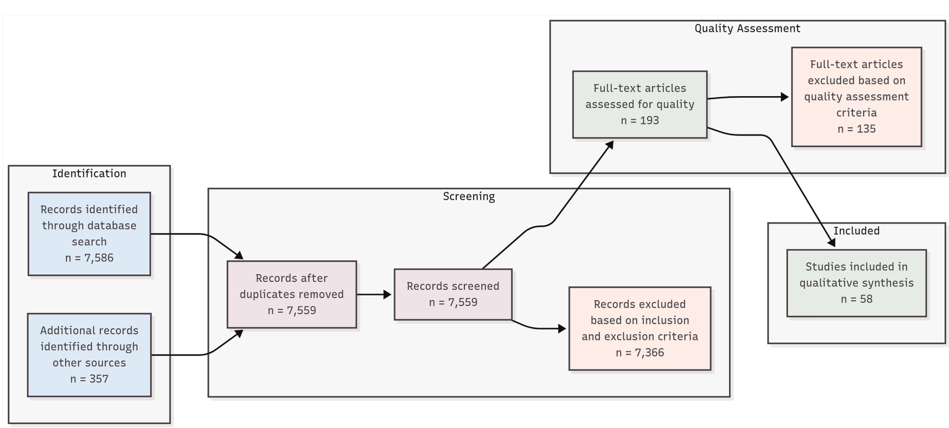

Figure 3.
Review protocol outcome.

papers was found. Using the backwards snowballing, 357 papers were added, making a total of 7943 papers. Then, removing the duplicates, we arrived at 7559. With the screening, 7366 were excluded. The remaining 193 were subjected to the quality assessment, out of which 135 papers were excluded. Finally, 58 papers were selected for the final review.

4.1 Data analysis

The data extracted from the 58 sources that were reviewed are summarized in **Tables 4–6**. From that summary, we draw the following insights:

- *Social networks*: **Table 4** shows that Twitter, Facebook, and Instagram are the most studied social networks; and Twitter is by far the most used.

- *Techniques and tools*: In **Table 5**, the techniques and tools have been classified into four groups: Machine Learning (ML), Deep Learning (DL), Transformer-based Models (TBM), and others. Traditional ML models are the most used, followed closely by DL.

- *Applications*: Among the multiple applications of information extraction from social media, **Table 6** shows that `sentiment Analysis` is the most frequent application, followed in order by `Public Health Surveillance`, `Security`, `Business Intelligence`, `Marketing`, and `Event Detection`.

Challenges and limitations The analysis of the extracted data uncovers several challenges, among which the most important are:

- Ethical, privacy, and security considerations,

- Generalization,

Social	Sources	# papers
Networks		
Twiter	[3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]	
	[16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]	46
	[27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38]	
	[25, 38, 39, 40, 41, 42, 43, 44, 45, 46]	
Facebook	[3, 4, 7, 8, 9, 11, 12, 13, 15, 17, 18, 19, 20, 23]	
	[21, 24, 25, 28, 31, 32, 33, 34, 35, 37, 38, 47]	30
	[40, 41, 43, 44]	
Instagram	[4, 11, 19, 23, 24, 28, 31, 32, 35, 38]	10
Others	[8, 9, 11, 14, 15, 16, 18, 19, 28, 34, 47]	
	[25, 41, 43, 48, 49, 50, 51]	18

Table 4.
Distributions of sources by social networks.

Tech/Tools	Sources	#papers
ML	[3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 32, 52]	
(Machine Learning)	[16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 53, 54, 55]	52
	[28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 47, 52, 56, 57]	
	[25, 38, 40, 41, 44, 45, 48, 49, 50]	
DL	[3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 32, 52]	
(Deep Learning)	[16, 18, 19, 20, 21, 22, 25, 26, 27, 28, 29, 53, 54, 55]	
	[30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 47, 52, 56, 57]	50
	[41, 42, 44, 45, 48, 49, 50]	
TBM	[3, 5, 6, 7, 9, 10, 13, 14, 17, 18, 19, 20, 22, 32, 52, 53]	
(Transformer-based Models)	[25, 26, 27, 28, 29, 30, 31, 34, 35, 47, 52, 54, 56]	38
	[36, 37, 38, 40, 41, 45, 48, 49, 50]	
Others	[3, 4, 6, 8, 11, 12, 13, 15, 16, 17, 18, 19, 20, 52]	
	[21, 23, 24, 25, 27, 30, 33, 34, 35, 40, 41, 47, 52, 54, 56]	35
	[25, 38, 42, 43, 49, 50]	

Table 5.
Distribution of papers by techniques/tools.

Applications	Sources	#papers
Fake news detection	[3, 8, 10, 11, 15, 16, 28, 37, 43, 50, 51, 52]	12
Business intelligence	[3, 4, 8, 9, 16, 19, 24]	
	[31, 38, 40, 44, 48, 55, 56]	14
Event detection	[3, 6, 7, 8, 11, 23, 25, 29, 31, 34, 35, 43, 48]	14
Sentiment analysis	[4, 11, 13, 14, 16, 18, 19, 20, 22, 27, 29, 30, 53]	
	[31, 33, 35, 36, 38, 40, 41, 48, 49, 56, 57]	
	[43, 45, 50]	28
Marketing	[4, 9, 14, 16, 24, 29, 30, 32, 47]	
	[38, 40, 41, 48, 57]	14
Security	[5, 8, 18, 21, 23, 26, 32, 52, 54]	
	[36, 39, 47, 49, 50]	15
Public health surveillance	[5, 7, 8, 11, 25, 29, 32, 34]	
	[3, 25, 41, 42, 45, 48, 49, 52]	16
Education	[9]	1
Recommendation systems	[11, 22, 24]	3
Intelligent vehicle networks	[43, 54, 55, 57]	4

Table 6.
Distribution of papers by application.

- Data scarcity,
- Real-time processing,
- Accuracy and computation cost,
- Noise in data,
- Processing multimodal and multilingual data.

4.2 Discussions

To discuss the results of the study, we consider each research question and interpret the related results.

1. RQ1: What applications are making use of information extraction from social media? Identifying sentiment analysis as the most frequent use of Information extraction from social media is not a surprising result. Sentiment analysis is not only an old explored technique, but also it is involved in many other applications, including Marketing, security, and health data analytics.
2. RQ2: What social networks are generally used in information extraction literature? Twitter (now X) is by far the most studied platform, which aligns with its microblogging nature and the ease of access to its data for research purposes.
3. RQ3: What are the techniques and tools used for information extraction from social media? Machine Learning (ML) and Deep Learning (DL) remain the solid foundation for information extraction. However, Transformer-based models (like BERT) are increasingly taking over.
4. RQ4: Which techniques and tools are most appropriate for which applications and types of information? This question requires a correlation between the identified applications (RQ1) and the techniques/tools (RQ3). This correlation inferred the following trends:

    - Sentiment analysis and fake news detection: NLP techniques and deep learning (especially recurrent neural networks such as GRU and LSTM, and Transformer-based models like BERT) are predominant. These methods excel at understanding linguistic nuances and classifying text.
    - Event detection and crisis monitoring: Topic modelling (LDA) and natural language processing techniques are often used to identify emerging themes and specific events in large volumes of textual data.
    - Security and malware detection: Machine Learning and Deep Learning are widely applied to identify abnormal patterns and classify malicious behaviors.
    - Recommendation systems and business intelligence: Machine Learning techniques, including topic models and Bayesian methods, are used to analyze user preferences and extract relevant information for decision-making.

- Specific applications (e.g., crop yield prediction, load forecasting): Deep Learning and reinforcement learning are increasingly adopted for their ability to handle complex data and optimize processes in dynamic environments.

- The choice of the most appropriate technique or tool strongly depends on the specific nature of the data (volume, velocity, variety, veracity), the available computational resources, and the performance requirements (accuracy, speed, interpretability) of the application.

5. RQ5: What are the challenges of information extraction from social media? The challenges identified by the study underscore the complexity of extracting information from social media, which is characterized by massive volume, high velocity, diverse formats, and often uncertain veracity. Future research must focus on developing more robust, efficient, and ethical methods to overcome these obstacles.

   In terms of techniques in **Table 7**, a wide range of approaches exist for information extraction. The most widely used approaches are based on Deep Learning (DL) and Machine Learning (ML). However, in our work we emphasize the use of emerging transformer-based techniques, which can provide better generalization capabilities for extracting data from any social media platform.

In terms of tools in **Table 8**, there are both general-purpose in **Table 9** solutions and platform-specific APIs. The strength of platform-specific APIs lies in their ability to provide highly detailed and reliable data for a single social media platform. By contrast, general-purpose tools offer flexibility by enabling information extraction across social media.

- Official APIs are the preferred option for large-scale social data extraction, offering stability and compliance with platform policies.

- General-purpose tools are useful when APIs are unavailable or when cross-platform data extraction is required.

- Professional tools and APIs usually require creating developer accounts with varying levels of access rights.

Category	Examples	Strengths	Limitations
Deep Learning	CNN, RNN, LSTM	High performance for targeted extraction tasks (e.g., NLP, image recognition).	Requires large datasets and significant computational resources.
Machine Learning	SVM, Random Forest, K-means	Lightweight models, easy to train, robust for simple tasks.	Less effective on large-scale or heterogeneous data.
Transformers	BERT, GPT, T5, LLaMA	State-of-the-art generalization across contexts and media; excellent performance in NLP.	High computational cost (GPU/TPU) and implementation complexity.

Table 7.
Comparison of extraction techniques.

Tool	Type	Strengths	Limitations
Zyte	Advanced scraping API	Handles IP rotation, bans, JavaScript rendering, structured extraction.	Paid service, requires configuration.
Bright Data	Web Unlocker	Anti-CAPTCHA, anti-ban, global access.	High cost.
ScraperAPI	Simple API	Returns raw HTML; supports JS rendering and geolocation.	Limited flexibility for complex extractions.
Apify	Actor-based platform	Ready-to-use scraping robots; accessible *via* REST and SDK.	Steeper learning curve.

Table 8.
Examples of general-purpose scraping tools.

Platform	API	Key features	Access conditions
X (Twitter)	X API v2/vNext	Posts, users, Spaces, DMs, analytics.	Free, Pro, or Enterprise tiers.
Facebook / Instagram	Meta Graph API (v22–v23, 2025)	Pages, content, insights, publishing.	Requires Meta Developer account.
YouTube	YouTube Data API v3	Videos, channels, playlists, search, upload.	Requires Google Cloud API key.
Reddit	Reddit API	Listings, OAuth, rate limits and rules.	OAuth authentication required.
TikTok	Business, Creator & Research APIs	Statistics, content, webhooks, posting.	Access depends on program (Business/Research).
LinkedIn	Marketing / Partner APIs	Recruitment, advertising, insights.	Restricted to approved partners.

Table 9.
Examples of platform-specific APIs.

5. Conclusion

The main objective of this study was to determine the techniques and tools used for information extraction from social media, as well as the challenges related to this field. This objective was broken into five research questions that directed the systematic review. Through the analysis of the data extracted from the selected papers, we were able to provide answers to the research questions.

5.1 Key findings

The findings of this systematic review confirm that information extraction from social media is a dynamic and rapidly expanding research area with the following:

- *Techniques and tools*: Machine Learning (ML) continues to play a fundamental role, while more specific techniques such as reinforcement learning and transformer-based models (e.g., BERT) are gaining importance.

- *Applications*: Applications are diverse, ranging from sentiment analysis, public health surveillance, security, business intelligence, event detection, marketing, misinformation detection, to more specialized uses across various domains.

- *Social networks*: Twitter remains the platform of choice for research, although other social media networks, including Facebook and Instagram, are also being explored.

- *Challenges*: Persistent challenges such as ethical, privacy, and security considerations, generalization, data scarcity, real-time processing, accuracy and computational cost, data noise, and processing multimodal and multilingual data require ongoing attention from the research community.

5.2 Implications and future directions

Addressing the challenges identified in this study calls for future research to focus on developing robust, efficient, and ethical methods. Innovative approaches in data preprocessing and data fusion are needed to overcome noise and heterogeneity. Privacy-preserving extraction techniques, potentially leveraging federated learning or homomorphic encryption, will be essential to address data confidentiality concerns. Moreover, detecting and mitigating biases in algorithms and datasets is crucial to ensuring fairness and reliability in information extraction systems.

Scalability remains a pressing requirement given the massive volume and velocity of social media data.

The integration of Explainable Artificial Intelligence (XAI) could enhance transparency and trust in extraction models, especially for critical applications such as public safety and healthcare.

While Twitter (now X) continues to dominate the research landscape, expanding studies to other platforms and considering their cultural and linguistic specificities will be essential for a more comprehensive and globally applicable understanding of information extraction.

Importantly, leveraging Transformer-based models, such as BERT and its successors present a significant opportunity to enhance information extraction capabilities. These models have demonstrated exceptional performance in handling context-rich, unstructured social media data, and can support more accurate, scalable, and multilingual information extraction.

Overall, this systematic review confirms the immense potential of social media information extraction while acknowledging the substantial obstacles. Future research must not only refine existing techniques but also proactively address ethical and practical dimensions, ensuring that the full benefits of this technology can be realized in an increasingly connected and data-driven world.

Author details

Lossan Bonde[1*] and Severin Dembele[2]

1 Adventist University of Africa, Nairobi, Kenya

2 United States International University, Nairobi, Kenya

*Address all correspondence to: bondel@aua.ac.ke

References

[1] Kitchenham B, Madeyski L, Budgen D. Segress: Software engineering guidelines for reporting secondary studies. IEEE Transactions on Software Engineering. 2023. ISSN 19393520.;**49**:1273-1298. DOI: 10.1109/TSE.2022.3174092

[2] Arya S, Kaji AH, Marja A. Boermeester. Prisma reporting guidelines for meta-analyses and systematic reviews. JAMA Surgery. 2021. ISSN 2168-6254;**156**: 789-790. DOI: 10.1001/JAMASURG.2021.0546. Available from: https://jamanetwork.com/journals/jamasurgery/fullarticle/2778468

[3] Mridha MF, Keya AJ, Hamid A, Monowar MM, Rahman S. A comprehensive review on fake news detection with deep learning. IEEE Access. 2021. ISSN 2169-3536;**9**: 156151-156170. DOI: 10.1109/ACCESS.2021.3129329 Available from: https://ieeexplore.ieee.org/document/9620068/

[4] Tam S, Said RB, Tanriover OO. A convbilstm deep learning model-based approach for twitter sentiment classification. IEEE Access. 2021. ISSN 2169-3536;**9**:41283-41293. DOI: 10.1109/ACCESS.2021.3064830. Available from: https://ieeexplore.ieee.org/document/9373371/

[5] Devika R, Vairavasundaram S, Mahenthar CSJ, Varadarajan V, Kotecha K. A deep learning model based on bert and sentence transformer for semantic keyphrase extraction on big social data. IEEE Access. 2021. ISSN 2169-3536;**9**:165252-165261. DOI: 10.1109/ACCESS.2021.3133651. Available from: https://ieeexplore.ieee.org/document/9641788/

[6] Fan C, Wu F, Mostafavi A. A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters. IEEE Access. 2020. ISSN 2169-3536;**8**:10478-10490. DOI: 10.1109/ACCESS.2020.2965550. Available from: https://ieeexplore.ieee.org/document/8955890/

[7] Karamat A, Imran M, Yaseen MU, Bukhsh R, Aslam S, Ashraf N. A hybrid transformer architecture for multiclass mental illness prediction using social media text. IEEE Access. 2025. ISSN 2169-3536;**13**:12148-12167. DOI: 10.1109/ACCESS.2024.3519308. Available from: https://ieeexplore.ieee.org/document/10804794/

[8] Rohera D, Shethna H, Patel K, Thakker U, Tanwar S, Gupta R, et al. A taxonomy of fake news classification techniques: Survey and implementation aspects. IEEE Access. 2022. ISSN 2169-3536;**10**:30367-30394. DOI: 10.1109/ACCESS.2022.3159651. Available from: https://ieeexplore.ieee.org/document/9734026/

[9] Bayazed AA, Almagrabi H, Alahmadi D, Alghamdi HS. Acom: Arabic comparative opinion mining in social media utilizing word embedding, deep learning model, & llm-gpt. IEEE Access. 2024. ISSN 2169-3536;**12**: 148741-148755. DOI: 10.1109/ACCESS.2024.3476336. Available from: https://ieeexplore.ieee.org/document/10707272/

[10] Hashmi E, Yayilgan SY, Yamin MM, Ali S, Abomhara M. Advancing fake news detection: Hybrid deep learning with fasttext and explainable ai. IEEE Access. 2024. ISSN 21693536;**12**: 44462-44480. DOI: 10.1109/ACCESS.2024.3381038

[11] Nedungadi P, Veena G, Tang K-Y, Menon RRK, Raman R. Ai techniques

and applications for online social networks and media: Insights from bertopic modeling. IEEE Access. 2025. ISSN 2169-3536;**13**:37389-37407. DOI: 10.1109/ACCESS.2025.3543795. Available from: https://ieeexplore.ieee.org/document/10892106/

[12] Latif ZS, Shafait F, Latif R. Analyzing lda and nmf topic models for urdu tweets via automatic labeling. IEEE Access. 2021. ISSN 2169-3536;**9**: 127531-127547. DOI: 10.1109/ACCESS.2021.3112620. Available from: https://ieeexplore.ieee.org/document/9536731/

[13] Alshahrani HJ, Hassan AQA, Almalki NS, Alnfiai MM, Salama AS, Hamza MA. Applied linguistics with red-tailed hawk optimizer-based ensemble learning strategy in natural language processing. IEEE Access. 2023. ISSN 2169-3536;**11**:132448-132456. DOI: 10.1109/ACCESS.2023.3331061. Available from: https://ieeexplore.ieee.org/document/10311595/

[14] Fadel AS, Saleh ME, Abulnaja OA. Arabic aspect extraction based on stacked contextualized embedding with deep learning. IEEE Access. 2022. ISSN 2169-3536;**10**:30526-30535. DOI: 10.1109/ACCESS.2022.3159252. Available from: https://ieeexplore.ieee.org/document/9733905/

[15] Bahurmuz NO, Amoudi GA, Baothman FA, Jamal AT, Alghamdi HS, Alhothali AM. Arabic rumor detection using contextual deep bidirectional language modeling. IEEE Access. 2022. ISSN 2169-3536;**10**:114907-114918. DOI: 10.1109/ACCESS.2022.3217522. Available from: https://ieeexplore.ieee.org/document/9931021/

[16] Abdelminaam DS, Neggaz N, Gomaa IAE, Ismail FH, Elsawy AA. Arabicdialects: An efficient framework for arabic dialects opinion mining on twitter using optimized deep neural networks. IEEE Access. 2021. ISSN 2169-3536;**9**:97079-97099. DOI: 10.1109/ACCESS.2021.3094173. Available from: https://ieeexplore.ieee.org/document/9471796/

[17] Sen O, Fuad M, Islam MN, Rabbi J, Masud M, Hasan MK, et al. Bangla natural language processing: A comprehensive analysis of classical, machine learning, and deep learning-based methods. IEEE Access. 2022. ISSN 2169-3536;**10**:38999-39044. DOI: 10.1109/ACCESS.2022.3165563. Available from: https://ieeexplore.ieee.org/document/9751052/

[18] Singh SS, Singh S, Singh K, Srivastava V, Shakya HK. Big data meets social networks: A survey of analytical strategies and research challenges. IEEE Access. 2025. ISSN 2169-3536;**13**: 98668-98698. DOI: 10.1109/ACCESS.2025.3576143. Available from: https://ieeexplore.ieee.org/document/11021641/

[19] Mathebula M, Modupe A, Marivate V. Chatgpt as a text annotation tool to evaluate sentiment analysis on south african financial institutions. IEEE Access. 2024. ISSN 2169-3536;**12**: 144017-144043. DOI: 10.1109/ACCESS.2024.3464374. Available from: https://ieeexplore.ieee.org/document/10684700/

[20] Anoop VS, Ajay Krishnan TK, Daud A, Banjar A, Bukhari A. Climate change sentiment analysis using domain specific bidirectional encoder representations from transformers. IEEE Access. 2024. ISSN 2169-3536;**12**: 114912-114922. DOI: 10.1109/ACCESS.2024.3441310. Available from: https://ieeexplore.ieee.org/document/10632142/

[21] Bilal M, Khan A, Jan S, Musa S. Context-aware deep learning model for

detection of roman urdu hate speech on social media platform. IEEE Access. 2022. ISSN 2169-3536;**10**:121133-121151. DOI: 10.1109/ACCESS.2022.3216375. Available from: https://ieeexplore.ieee.org/document/9926094/

[22] Eke CI, Norman AA, Shuib L. Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and bert model. IEEE Access. 2021. ISSN 2169-3536;**9**:48501-48518. DOI: 10.1109/ACCESS.2021.3068323. Available from: https://ieeexplore.ieee.org/document/9383219/

[23] Obaida MH, Elkaffas SM, Guirguis SK. Deep learning algorithms for cyber-bulling detection in social media platforms. IEEE Access. 2024. ISSN 2169-3536;**12**:76901-76908. DOI: 10.1109/ACCESS.2024.3406595. Available from: https://ieeexplore.ieee.org/document/10540101/

[24] Mameli M, Paolanti M, Pietrini R, Pazzaglia G, Frontoni E, Zingaretti P. Deep learning approaches for fashion knowledge extraction from social media: A review. IEEE Access. 2022. ISSN 2169-3536;**10**:1545-1576. DOI: 10.1109/ACCESS.2021.3137893. Available from: https://ieeexplore.ieee.org/document/9661397/

[25] Tahir WB, Khalid S, Almutairi S, Abohashrh M, Memon SA, Khan J. Depression detection in social media: A comprehensive review of machine learning and deep learning techniques. IEEE Access. 2025. ISSN 2169-3536;**13**:12789-12818. DOI: 10.1109/ACCESS.2025.3530862. Available from: https://ieeexplore.ieee.org/document/10843708/

[26] Alatawi HS, Alhothali AM, Moria KM. Detecting white supremacist hate speech using domain specific word embedding with deep learning and bert. IEEE Access. 2021. ISSN 2169-3536;**9**:106363-106374. DOI: 10.1109/ACCESS.2021.3100435. Available from: https://ieeexplore.ieee.org/document/9497095/

[27] Hassan ME, Hussain M, Maab I, Habib U, Khan MA, Masood A. Detection of sarcasm in urdu tweets using deep learning and transformer based hybrid approaches. IEEE Access. 2024. ISSN 2169-3536;**12**:61542-61555. DOI: 10.1109/ACCESS.2024.3393856. Available from: https://ieeexplore.ieee.org/document/10508575/

[28] Koru GK, Uluyol Ç. Detection of turkish fake news from tweets with bert models. IEEE Access. 2024. ISSN 2169-3536;**12**:14918-14931. DOI: 10.1109/ACCESS.2024.3354165. Available from: https://ieeexplore.ieee.org/document/10399808/

[29] Hossain MS, Hossain MM, Hossain MS, Mridha MF, Safran M, Alfarhood S. Emonet: Deep attentional recurrent cnn for x (formerly twitter) emotion classification. IEEE Access. 2025. ISSN 2169-3536;**13**:37591-37610. DOI: 10.1109/ACCESS.2025.3545863. Available from: https://ieeexplore.ieee.org/document/10904242/

[30] Nguyen A, Longa A, Luca M, Kaul J, Lopez G. Emotion analysis using multilayered networks for graphical representation of tweets. IEEE Access. 2022. ISSN 2169-3536;**10**:99467-99478. DOI: 10.1109/ACCESS.2022.3207161. Available from: https://ieeexplore.ieee.org/document/9893783/

[31] Chelloug SA, Muthanna MSA, Jamil F, Al-Gaashani MSAM, Alhelaly S, Aziz A, et al. Enhancing hajj and umrah services through predictive social media classification. IEEE Access. 2025. ISSN 2169-3536;**13**:67220-67238.

DOI: 10.1109/ACCESS.2025.3559204. Available from: https://ieeexplore.ieee.org/document/10960302/

[32] Essahraui S, Lamaakal I, Maleh Y, El Makkaoui K, Bouami MF, Ouahbi I, et al. Human behavior analysis: A comprehensive survey on techniques, applications, challenges, and future directions. IEEE Access. 2025. ISSN 2169-3536;**13**:128379-128419. DOI: 10.1109/ACCESS.2025.3589938. Available from: https://ieeexplore.ieee.org/document/11082162/

[33] Gasmi K, Ltaifa IB, Abdalrahman AE, Hamid O, Altaieb MO, Ali S, et al. Hybrid feature and optimized deep learning model fusion for detecting hateful arabic content. IEEE Access. 2025. ISSN 2169-3536;**13**:131411-131431. DOI: 10.1109/ACCESS.2025.3591673. Available from: https://ieeexplore.ieee.org/document/11088089/

[34] Xin C, Zakaria LQ. Integrating bert with cnn and bilstm for explainable detection of depression in social media contents. IEEE Access. 2024. ISSN 2169-3536;**12**:161203-161212. DOI: 10.1109/ACCESS.2024.3488081. Available from: https://ieeexplore.ieee.org/document/10738819/

[35] Zhou S, Mohd M. Mental health safety and depression detection in social media text data: A classification approach based on a deep learning model. IEEE Access. 2025. ISSN 2169-3536;**13**:63284-63297. DOI: 10.1109/ACCESS.2025.3559170. Available from: https://ieeexplore.ieee.org/document/10960428/

[36] Tam S, Tanriöver ÖÖ. Multimodal deep learning crime prediction using tweets. IEEE Access. 2023. ISSN 2169-3536;**11**:93204-93214. DOI: 10.1109/ACCESS.2023.3308967. Available from: https://ieeexplore.ieee.org/document/10231346/

[37] Ahammad M, Sani A, Rahman K, Masud MTIMMR, Hassan MM, Rony MAT, et al. Roberta-gcn: A novel approach for combating fake news in bangla using advanced language processing and graph convolutional networks. IEEE Access. 2024. ISSN 2169-3536;**12**:132644-132663. DOI: 10.1109/ACCESS.2024.3457860. Available from: https://ieeexplore.ieee.org/document/10677406/

[38] Tan KL, Lee CP, Lim KM, Anbananthen KSM. Sentiment analysis with ensemble hybrid deep learning model. IEEE Access. 2022. ISSN 2169-3536;**10**:103694-103704. DOI: 10.1109/ACCESS.2022.3210182. Available from: https://ieeexplore.ieee.org/document/9903622/

[39] Hina M, Ali M, Javed AR, Ghabban F, Khan LA, Jalil Z. Sefaced: Semantic-based forensic analysis and classification of e-mail data using deep learning. IEEE Access. 2021. ISSN 2169-3536;**9**:98398-98411. DOI: 10.1109/ACCESS.2021.3095730. Available from: https://ieeexplore.ieee.org/document/9477611/

[40] Es-Sabery F, Hair A, Qadir J, Sainz-De-Abajo B, Garcia-Zapirain B, Torre-Diez I. Sentence-level classification using parallel fuzzy deep learning classifier. IEEE Access. 2021. ISSN 2169-3536;**9**: 17943-17985. DOI: 10.1109/ACCESS.2021.3053917 Available from: https://ieeexplore.ieee.org/document/9333555/

[41] Albladi A, Islam M, Seals C. Sentiment analysis of twitter data using nlp models: A comprehensive review. IEEE Access. 2025. ISSN 2169-3536;**13**: 30444-30468. DOI: 10.1109/ACCESS.2025.3541494. Available from: https://ieeexplore.ieee.org/document/10883978/

[42] Olusegun R, Oladunni T, Halima A, Houkpati YAO, Bengesi S. Text mining and emotion classification on monkeypox twitter dataset: A deep learning-natural language processing (nlp) approach. IEEE Access. 2023. ISSN 2169-3536;**11**:49882-49894. DOI: 10.1109/ACCESS.2023.3277868. Available from: https://ieeexplore.ieee.org/document/10129946/

[43] Bahja M, Safdar GA. Unlink the link between covid-19 and 5g networks: An nlp and sna based approach. IEEE Access. 2020. ISSN 2169-3536;**8**: 209127-209137. DOI: 10.1109/ACCESS.2020.3039168. Available from: https://ieeexplore.ieee.org/document/9262907/

[44] Sansonetti G, Gasparetti F, D'aniello G, Micarelli A. Unreliable users detection in social media: Deep learning techniques for automatic detection. IEEE Access. 2020. ISSN 2169-3536;**8**: 213154-213167. DOI: 10.1109/ACCESS.2020.3040604. Available from: https://ieeexplore.ieee.org/document/9269985/

[45] Wang Q, Sailor HB, Lee KA, Ma K, Goh KH, Boh WF. Using twitter dataset for social listening in Singapore. IEEE Access. 2024. ISSN 2169-3536;**12**: 100015-100025. DOI: 10.1109/ACCESS.2024.3427760. Available from: https://ieeexplore.ieee.org/document/10597634/

[46] Endarnoto SK, Pradipta S, Nugroho AS, Purnama J. Traffic condition information extraction & visualization from social media twitter for android mobile application. In: Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011. Bandung, Indonesia: IEEE; 2011. DOI: 10.1109/ICEEI.2011.6021743

[47] Arya P, Pandey AK, Patro SGK, Tiwari K, Panigrahi N, Naveed QN, et al. Mscmgtb: A novel approach for multimodal social media content moderation using hybrid graph theory and bio-inspired optimization. IEEE Access. 2024. ISSN 2169-3536;**12**: 73700-73718. DOI: 10.1109/ACCESS.2024.3400815. Available from: https://ieeexplore.ieee.org/document/10530557/

[48] Jannani A, Bouhsissin S, Sael N, Benabbou F. Topic modeling and sentiment analysis of arabic news headlines for a societal well-being scoring and monitoring system: Moroccan use case. IEEE Access. 2025. ISSN 2169-3536;**13**:26345-26364. DOI: 10.1109/ACCESS.2025.3538888. Available from: https://ieeexplore.ieee.org/document/10872924/

[49] Kumi S, Snow C, Lomotey RK, Deters R. Uncovering concerns of citizens through machine learning and social network sentiment analysis. IEEE Access. 2024. ISSN 2169-3536;**12**: 94885-94913. DOI: 10.1109/ACCESS.2024.3426329. Available from: https://ieeexplore.ieee.org/document/10591750/

[50] Luqman M, Faheem M, Ramay WY, Saeed MK, Ahmad MB. Utilizing ensemble learning for detecting multi-modal fake news. IEEE Access. 2024. ISSN 2169-3536;**12**:15037-15049. DOI: 10.1109/ACCESS.2024.3357661. Available from: https://ieeexplore.ieee.org/document/10412049/

[51] Baashirah R. Zero-shot automated detection of fake news: An innovative approach (zs-fnd). IEEE Access. 2024. ISSN 2169-3536;**12**:182828-182840. DOI: 10.1109/ACCESS.2024.3462151. Available from: https://ieeexplore.ieee.org/document/10681065/

[52] Bashaddadh O, Omar N, Mohd M, Khalid MNA. Machine learning and deep learning approaches for fake news detection: A systematic review of techniques, challenges, and advancements. IEEE Access. 2025. ISSN 2169-3536;**13**:90433-90466. DOI: 10.1109/ACCESS.2025.3572051. Available from: https://ieeexplore.ieee.org/document/11008608/

[53] Soomro SA, Yuhaniz SS, Dootio MA, Mujtaba G, Siddiqui JA. Category-based sentiment analysis of sindhi news headlines using machine learning, deep learning, and transformer models. IEEE Access. 2025. ISSN 2169-3536;**13**: 99985-100001. DOI: 10.1109/ACCESS.2025.3576853. Available from: https://ieeexplore.ieee.org/document/11025811/

[54] Almehdhar M, Albaseer A, Khan MA, Abdallah M, Menouar H, Al-Kuwari S, et al. Deep learning in the fast lane: A survey on advanced intrusion detection systems for intelligent vehicle networks. IEEE Open Journal of Vehicular Technology. 2024. ISSN 2644-1330;**5**:869-906. DOI: 10.1109/OJVT.2024.3422253. Available from: https://ieeexplore.ieee.org/document/10582439/

[55] Matrone S, Pozzi A, Ogliari E, Leva S. Deep learning-based predictive control for optimal battery management in microgrids. IEEE Access. 2024. ISSN 2169-3536;**12**:141580-141593. DOI: 10.1109/ACCESS.2024.3458435. Available from:https://ieeexplore.ieee.org/document/10677472/

[56] Gholamiangonabadi D, Grolinger K. Federated learning for sentiment analysis in presence of non-iid data: Sensitivity of deep learning models. IEEE Access. 2024. ISSN 2169-3536;**12**: 128049-128060. DOI: 10.1109/ACCESS.2024.3453068. Available from: https://ieeexplore.ieee.org/document/10662960/

[57] Jin N, Wu J, Ma X, Yan K, Mo Y. Multi-task learning model based on multi-scale cnn and lstm for sentiment classification. IEEE Access. 2020. ISSN 2169-3536;**8**:77060-77072. DOI: 10.1109/ACCESS.2020.2989428. Available from: https://ieeexplore.ieee.org/document/9076160/