RE-Miner 2.0: A Holistic Framework for Mining Mobile Application Reviews

Max Tiessler[1], Quim Motger[1,*]

[1]Department of Service and Information System Engineering, Universitat Politècnica de Catalunya

Abstract

In the domain of application stores and marketplaces, user reviews are crucial for supporting multiple requirements engineering tasks. Feature extraction, emotion classification, topic analysis, review type identification, and polarity analysis are key components in requirements prioritization, feedback gathering, and release planning. Empirical evaluation of these techniques is challenging due to data collection complexities and a lack of reproducible methods and available tools. Furthermore, existing studies often focus on isolated tasks, hindering a comprehensive analysis of user perceptions. This paper introduces RE-Miner 2.0, a work-in-progress tool that integrates multiple data extraction and analysis methods in a distributed environment (RE-Miner Ecosystem), enabling a multidimensional and detailed analysis of user feedback. It offers a web-based service for task integration and comparison, supported by persistent storage and a web application that allows analytical visualization of reviews. As a result, RE-Miner 2.0 provides a platform for task integration, replication, and comparison of review mining techniques. Bringing advancements in deep review analysis for requirements engineering. A demo of the tool is showcased here: https://www.youtube.com/watch?v=a11bHSCYqqM.

Keywords

mobile app reviews, feature extraction, emotion classification, polarity analysis, topic classification, type classification, feature clustering, natural language processing

1. Introduction

User reviews are critical for requirements engineering tasks like feature elicitation [1], bug detection [2], and release planning [3]. While natural language processing advancements (e.g., transformers, generative AI) have improved review analysis through automated requirement extraction [4], feature extraction [5], and emotion classification [6], existing tools remain fragmented. Recent solutions like knowledge graphs [7] and open-source miners [8] address isolated tasks but lack integration with complementary techniques such as polarity detection [9], review-type classification [10], and feature-based topic modeling. This fragmentation leads to incomplete analyses, added to the absence of open-source tools that integrate these multidimensional tasks. Furthermore, method evaluation is hindered by data collection complexities [11], replication barriers, and resource-intensive deployments [12]. Practitioners also struggle to select domain-appropriate approaches [13], often due to unavailable implementations [14].

Building on our prior work [15], we present **RE-Miner 2.0**, a holistic framework addressing these gaps with two contributions: *(1)* **RE-Miner Ecosystem**, a web-based distributed and flexible architecture designed to easily incorporate new extraction and classification methods and comparing multiple review mining tasks (i.e., feature extraction, polarity detection, type classification, topic classification, emotion classification, and clustering of feature taxonomies); and *(2)* **RE-Miner 2.0**, a web application for visualizing user reviews, statistical data generated from these reviews and feature clustering. This contribution extends the previous version by *(2.1)* introducing full integration with MApp-KG, an RDF-based knowledge graph integrating a catalogue of mobile applications and user review [16]; (2.2)

In: A. Hess, A. Susi, E. C. Groen, M. Ruiz, M. Abbas, F. B. Aydemir, M. Daneva, R. Guizzardi, J. Gulden, A. Herrmann, J. Horkoff, S. Kopczyńska, P. Mennig, M. Oriol Hilari, E. Paja, A. Perini, A. Rachmann, K. Schneider, L. Semini, P. Spoletini, A. Vogelsang. Joint Proceedings of REFSQ-2025 Workshops, Doctoral Symposium, Posters & Tools Track, and Education and Training Track. Co-located with REFSQ 2025. Barcelona, Spain, April 7, 2025.

*Corresponding author.
max.tiessler@upc.edu (M. Tiessler); joaquim.motger@upc.edu (Q. Motger)
0009-0002-6535-9666 (M. Tiessler); 0000-0002-4896-7515 (Q. Motger)

integration with multiple additional descriptor and analysis services, specifically those related to the previously mentioned tasks (i.e., polarity detection, type classification, topic classification and feature clustering services); *(2.3)* implementation of an experimental feature hierarchical clustering analysis along with its interactive visualization; and *(2.4)* multiple usability and user experience improvements.

2. Related Work

Natural Language Processing for Requirements Engineering (NLP4RE) is a research field focused on software-based solutions for automating language processing and detection techniques within the Requirements Engineering (RE) domain. Despite surveys identifying over 130 tools [13], only 13% are publicly available for replication, resulting in difficulties when conducting comparative studies or trying to implement one of those systems. Key NLP4RE tasks [14] in mobile app review analysis include:

- **Feature Extraction**, identifying features in reviews using methods like the SAFE approach [17], which employs syntactic pattern matching and semantic similarity.
- **Type Classification**, categorizing reviews by intent (e.g., *bug report*, *feature request*) [18].
- **Topic Classification**, using semantic analysis to identify review aspects such as *aesthetics*, *compatibility*, *cost*, *effectiveness*, *efficiency*, *enjoyability*, *learnability*, *security*, or *usability* [19].
- **Sentiment Analysis**, which involves evaluating and extracting information from text to understand attitudes and emotions, is divided into various subtasks within the field of opinion mining. In RE-Miner 2.0, tasks such as polarity analysis, which determines the negativity or positivity of a text [14], and emotion detection (e.g., identifying emotions like *anger* and *sadness*) using models based on Ekman's six universal emotions [20] are applied.

Despite the diversity of tools, several challenges apply across multiple review mining tasks. Furthermore, studies highlight the limited use of NLP techniques for multidimensional review analysis. Dąbrowski et al. emphasize the scarcity of tools integrating various analyses [21]. Tools like SAFE [17] and GuMa [22] (a syntactic-semantic approach combining feature extraction, polarity analysis, and topic modeling to infer features and sentiments), rely on syntactic patterns but struggle with the informality of reviews [23], a problem addressed by newer deep learning methods [5]. However, most tools remain unavailable or non-extensible [14]. Progress has also been made in other tasks, such as emotion classification, which has evolved from basic sentiment analysis to transformer-based models [24]. However, integrated solutions capable of handling multiple tasks remain scarce. While some tools attempt to address this gap, such as the Appsent tool [25], which extracts and visualizes end-user feedback through sentiment, emotion, and feature-issue analysis, holistic solutions are still limited. Additionally, less popular techniques, such as topic or type classification, face challenges due to a lack of available solutions and tool integration. The initial version of RE-Miner filled these gaps by integrating feature extraction and emotion classification in an open-source tool, laying the basis for RE-Miner 2.0, which expands it into a multidimensional solution by offering an open-source, extensible framework with packaged web services, detailed documentation, and a sample dataset, integrating state-of-the-art models for multi-task review analysis.

3. Tool Description

RE-Miner 2.0 is designed to be user-friendly, accessible, and reusable, making it suitable for both researchers and mobile application designers [26]. The software focuses on five core objectives: *(1)* **Reusability**: users can easily reproduce studies and analysis, by reusing the integrated models, services, and datasets. *(2)* **Shareability**: the tool supports link sharing and result export, enabling users to share data for collaborative research. *(3)* **Flexibility**: built on a micro-services architecture, the tool allows users to easily extend it by integrating new tasks or executing only the necessary micro-services (i.e., extraction or clustering tasks) without risking the software's integrity. *(4)* **Result comparability**: the tool provides comparative analysis, enabling users to evaluate multiple methods in aggregation

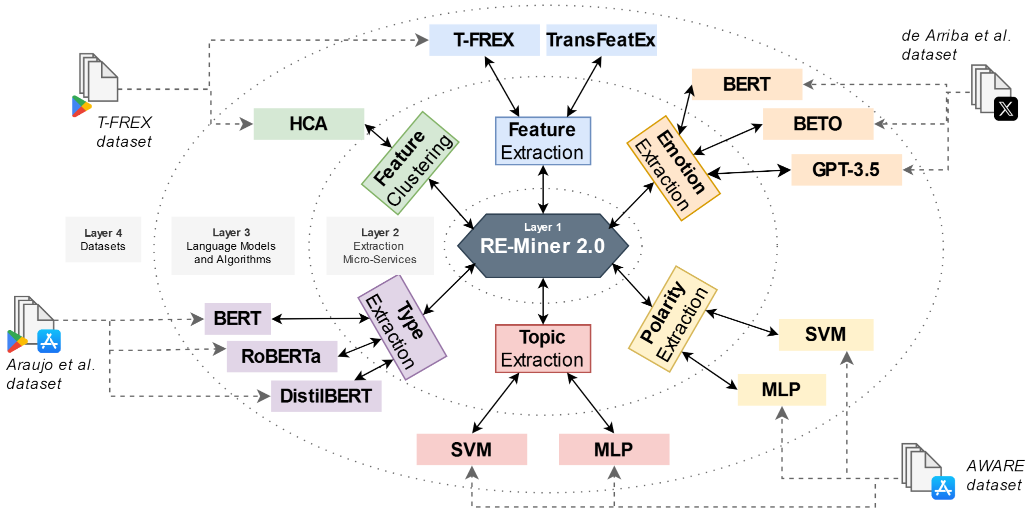


Figure 1: RE-Miner Ecosystem overview

or in parallel. *(5)* **Data privacy and security**: the tool protects sensitive information through review anonymization and access control mechanisms.

3.1. RE-Miner Ecosystem

The RE-Miner Ecosystem (Figure 1) is a distributed, microservice-based, four-layer architecture that is both interconnected and independent. Each layer can be used either in isolation or orchestrated through RE-Miner 2.0, aligning with the tool's flexibility objective. The techniques and models in Layer 3, as well as the datasets in Layer 4, are based in literature research and our own work. As a result, the tool reflects the state of the art and related work.

- **Layer 1 - Core Infrastructure (RE-Miner 2.0)**: The base of the ecosystem is the RE-Miner 2.0, the entry point for users to access the whole ecosystem via an user interface. It is detailed in Section 3.2.
- **Layer 2 - Extraction Micro-Services**: This layer contains all extraction and clustering services, organized as micro-services accessible via REST APIs. Its features include *(1)* compatibility with diverse environments, *(2)* adaptability to state-of-the-art NLP techniques (e.g., integrating new LLMs), and *(3)* scalability and customization, allowing developers to easily add or expand services by reusing components like data transfer objects (DTOs) or existing micro-service architectures. The micro-services in this layer support *(1)* type extraction, *(2)* topic extraction, *(3)* polarity extraction, *(4)* feature extraction, *(5)* emotion extraction, and *(6)* feature clustering.
- **Layer 3 - Language Models and Algorithms**: This layer provides the models and algorithms used by the services in Layer 2. Each task is supported by at least one model, which may be based on machine learning (e.g., Support Vector Machines (SVM)), deep learning architectures (e.g., T-FREX, Multilayer Perceptron (MLP)), transformer architectures (e.g., GPT-3.5, BERT), or custom-developed techniques for feature clustering (e.g., Hierarchical Clustering Analysis (HCA), which is an own work technique and an experimental work-in-progress feature). This layer is designed to be flexible and easily expandable, enabling developers to quickly integrate or update models to improve performance and expand scope. The methods and models that compose the current version of RE-Miner 2.0 are presented in Figure 1.
- **Layer 4 - Datasets**: The outermost layer contains datasets used for fine-tuning and validating the models in Layer 3. These datasets are diverse, with many derived from application reviews, though they also incorporate data from other domains. They are based on related research for training the models with relevant and accurate data. The new datasets introduced in this

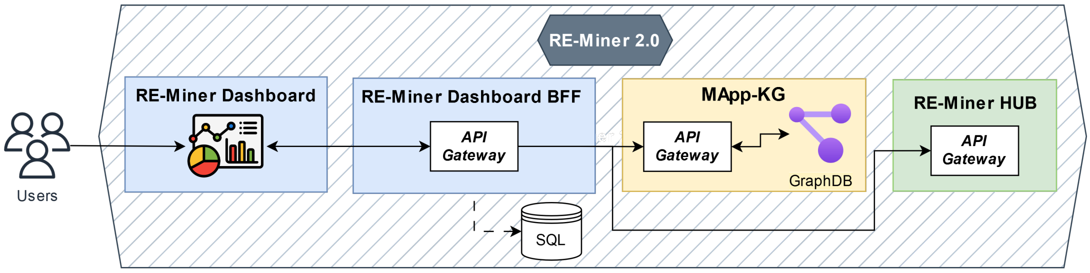


Figure 2: RE-Miner 2.0 core layer overview

version of the tool include: *(1)* **Polarity Analysis Dataset** [19]: annotated with 11,321 reviews, categorized into positive and negative polarities; *(2)* **Review Type Analysis Dataset** [18]: 3,691 annotated reviews, including types such as *Bug*, *Feature*, and *User Experience*; *(3)* **Review Topic Analysis Dataset**: AWARE dataset [19] with 11,321 annotated reviews, covering topics like Usability, Effectiveness, and Security; and *(4)* **T-FREX Dataset** [5]: 23,816 annotated reviews focused on feature-specific review extraction and analysis.

3.2. RE-Miner Core

The core of RE-Miner 2.0 acts as an entry point for users to access the system via an user interface (two view examples can be seen in Figure 4). It is composed of five components: *(1)* **RE-Miner Dashboard**, which offers a responsive and user-friendly interface, allowing researchers and developers to access the system, configure analysis tasks, manage reviews, and visualize results (as it can be seen in Figure 4); *(2)* **RE-Miner Dashboard BFF**, built on a Backend for Frontend (BFF) pattern, which handles the front-end business logic, API requests, data processing, and communication with other services within the core layer; *(3)* **SQL Database** that stores and manages user information, authentication data, and access permissions; *(4)* **RE-Miner HUB**, acting as a central integration point, connecting the core layer with the outer layers; and *(5)* **MApp-KG** [16], which organizes all of RE-Miner's data and relationships. An overview of these components and their interconnections is illustrated in Figure 2.

4. User Workflow

RE-Miner 2.0 expands from feature extraction and emotion classification to multidimensional analysis the two existing use cases (Single-review analysis, Batch-review analysis and Visual analytics) introduced in the first version [15] and introduces an additional use case (Feature Clustering Analysis) for review analysis. Users must complete sign-up, login, and upload their apps and reviews before using these features.

- **Single-Review Analysis:** Extended from the initial RE-Miner version [15]. This version introduces a new enhanced *Reviews Directory*, allowing users to search and filter reviews by application package, features, or descriptors (e.g., topic, type, polarity, or emotion). The review processing wizard now supports the selection of multiple tasks and methods for multidimensional review analysis, including new tasks such as topic classification, type classification, and feature clustering. Results can be viewed in the updated *Review Analyzer* by clicking the icon (see Figure 4), where new additional outputs such as detected polarities, topics, and types can be seen.
- **Batch-Review Analysis and Visual Analytics:** This version extends the original use case [15] by introducing improved analytical plots, including new charts such as the *Descriptor Polar Area* and a temporal *Descriptor Histogram*, improving the visualizations from the initial version. Existing plots have also been improved for a better user experience.
- **Feature Clustering Analysis:** RE-Miner 2.0 introduces a new hierarchical feature clustering view, designed to visualize results from the Hierarchical Clustering Analysis (HCA) technique

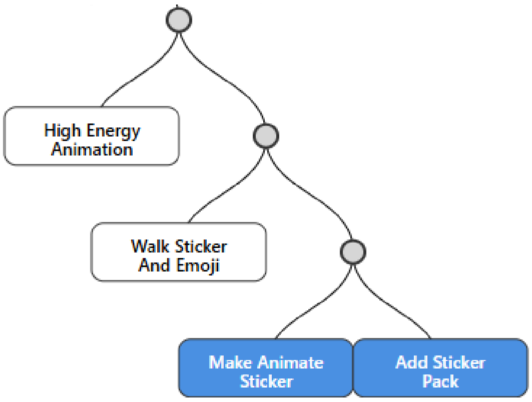
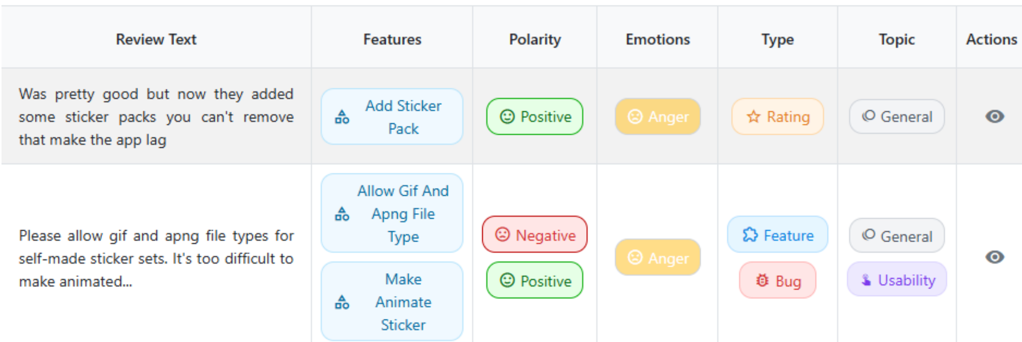


Figure 3: The Clustering View in RE-Miner 2.0. Selected features from a cluster family are shown. The derived result search is displayed in Figure 4.

introduced in Layer 3 (see Figure 1). This view is accessible through the *Tree Analyzer* tab, allowing users to explore and interact with the hierarchical clustering of features. In this view, shown in Figure 3, users can select an application package and a feature family from the generated clusters. An interactive visualization of the feature hierarchy then appears, enabling users to analyze and manipulate the clustered data. Users have the following options for further analysis and customization: *(1)* adjust the distance threshold between sibling nodes to refine cluster granularity according to their research or analysis needs; *(2)* select specific clusters or subclusters and download a JSON file containing the cluster data for reproducibility or sharing; *(3)* navigate directly to the *Reviews Directory* with pre-applied filters based on the selected application package and features, allowing for quick access to reviews linked to the chosen feature clusters (an example can be seen in Figure 4).



Review Text	Features	Polarity	Emotions	Type	Topic	Actions
Was pretty good but now they added some sticker packs you can't remove that make the app lag	Add Sticker Pack	Positive	Anger	Rating	General	
Please allow gif and apng file types for self-made sticker sets. It's too difficult to make animated...	Allow Gif And Apng File Type Make Animate Sticker	Negative Positive	Anger	Feature Bug	General Usability	

Figure 4: Improved Review Directory in RE-Miner 2.0.

5. Evaluation

5.1. Plan

Below we summarize the main steps of the planned (and ongoing) evaluation.

1. **Data collection and annotation.** We built the tool based on our previous work on mobile app repository mining for collecting multiple reviews using the mentioned datasets in Subsection 3.1. For the remaining tasks, more precisely in emotion classification, we are currently developing a dataset of tagged emotions through iterative annotation processes of subsets of reviews following structured guidelines. For future work, we plan to expand our datasets by including review topics

and types, using an iterative process that measures annotation agreement and applies strong evaluation criteria.

2. **Experimentation.** We plan to conduct an empirical evaluation of all tasks described in Layer 2 of Subsection 3.1. This will involve multiple cross-validation analyses on the full dataset annotated with features, emotions, types, topics, polarities, and clustered features. Our evaluation will include a quantitative ground-truth evaluation to assess and compare the effectiveness of each technique. Additionally, we will evaluate the tool's overall software product quality with a focus on two key aspects defined by ISO/IEC 25010 [27]: *(1)* **performance efficiency**, which will involve measuring the tool's response time and scalability under varying workloads, particularly for data upload and all descriptor extraction tasks, and *(2)* **usability**, which will be assessed through user studies evaluating the ease of interaction, explainability of results, and overall user satisfaction. Potential stakeholders will participate in these studies to provide feedback on the tool's usability.

5.2. Threats to validity

Based on the taxonomy proposed by Wohlin et al. [28], we identify key threats to the successful implementation and evaluation of RE-Miner 2.0 and the evaluation plan described in Section 5.1. We focus on internal and external validity, given their relevance to this study. Concerning *internal* validity, potential biases in data collection and model selection may affect our findings. The datasets used in RE-Miner 2.0 (Layer 4) originate from prior studies and public sources, which may introduce inconsistencies in annotation quality and domain specificity. Additionally, our evaluation relies on predefined tasks and configurations, which may not account for all real-world variations. To mitigate these risks, we plan to employ cross-validation and benchmark against multiple state-of-the-art approaches. Additionally, we facilitate the integration of new methods and algorithms for further evaluation.

Concerning *external* validity, generalization remains a key challenge, particularly regarding the usability, learnability, and functional suitability of RE-Miner 2.0. The tool's performance is highly dependent on the descriptors and services (Layer 2), models and algorithms (Layer 3), and datasets (Layer 4) selected. However, these choices are based on a comprehensive analysis of app review mining literature, ensuring broad coverage of effective methods. Furthermore, RE-Miner 2.0 is designed as a flexible framework rather than a gold-standard selection of services. Its modular architecture allows for easy customization, enabling users to adapt datasets, models, and descriptors to different contexts.

6. Conclusions and Future Work

RE-Miner 2.0 addresses the limitations of existing review mining tools by providing a flexible and holistic solution for comprehensive app review analysis. It also extends the features and usability of the first version. The tool introduces three main contributions: *(1)* **a holistic framework** for multidimensional analysis, integrating feature extraction and clustering, emotion classification, polarity detection, topic identification, and review-type classification to enhance understanding of user feedback and app requirements; *(2)* **improved reproducibility and extensibility** through a scalable, layered architecture (RE-Miner Ecosystem) that supports both pre-built and custom models, allowing continuous updates according with advances in NLP research; and *(3)* **a data-driven visualization, search, and feature clustering tool** (RE-Miner 2.0) for analyzing review data at both granular and aggregate levels, helping stakeholders and researchers in trend detection, requirement prioritization, and release planning. As future work, we plan to *(1)* implement advanced clustering for analyzing mobile apps and market segments and *(2)* expand the tool to support broader software ecosystems, including issue trackers and marketplaces.

Acknowledgments

With the support from the Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia and the European Social Fund. This paper has been funded by the Spanish Ministerio de Ciencia e Innovación under project / funding scheme PID2020-117191RB-I00 / AEI/10.13039/501100011033.

References

[1] C. Iacob, R. Harrison, Retrieving and analyzing mobile apps feature requests from online reviews, in: Proceedings of the 10th Working Conference on Mining Software Repositories, 2013. doi:10.1109/MSR.2013.6624001.

[2] W. Maalej, H. Nabil, Bug report, feature request, or simply praise? on automatically classifying app reviews, in: Int. Requirements Engineering Conference, 2015. doi:10.1109/RE.2015.7320414.

[3] L. Villarroel, et al., Release planning of mobile apps based on user reviews, in: Proceedings - Int. Conference on Software Engineering, 2016. doi:10.1145/2884781.2884818.

[4] A. Nguyen-Duc, B. Cabrero-Daniel, A. Przybyłek, C. Arora, D. Khanna, T. Herda, U. Rafiq, J. Melegati, E. Guerra, K.-K. Kemell, M. Saari, Z. Zhang, H. Le, T. Quan, P. Abrahamsson, Generative artificial intelligence for software engineering - a research agenda, SSRN Electronic Journal (2023). doi:10.2139/ssrn.4622517.

[5] Q. Motger, A. Miaschi, F. Dell'Orletta, X. Franch, J. Marco, T-frex: A transformer-based feature extraction method from mobile app reviews, in: Proceedings of the 2024 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), 2024. doi:10.1109/SANER60148.2024.00030.

[6] S. L. Ramaswamy, J. Chinnappan, Recognet-lstm+cnn: a hybrid network with attention mechanism for aspect categorization and sentiment classification, Journal of Intelligent Information Systems (2022). doi:10.1007/s10844-021-00692-3.

[7] D. Dave, A. Sharma, S. M. Abdulhamid, A. Ahmed, A. Akhunzada, R. Amin, Sappkg: Mobile app recommendation using knowledge graph and side information - a secure framework, IEEE Access (2023). doi:10.1109/ACCESS.2023.3296466.

[8] J. Wei, A.-L. Courbis, T. Lambolais, B. Xu, P. L. Bernard, G. Dray, Zero-shot bilingual app reviews mining with large language models, in: Proc. 35th IEEE Int. Conf. on Tools with Artificial Intelligence (ICTAI), IEEE, 2023. doi:10.1109/ICTAI59109.2023.00135.

[9] J. A. Roberto, M. Salamó, M. A. Martí, Genre-based stages classification for polarity analysis, in: Proc. 28th Int. Florida AI Research Soc. Conf. (FLAIRS-28), AAAI Press, 2015.

[10] A. Qazi, R. G. Raj, G. Hardaker, C. Standing, A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges, Internet Research (2017). doi:10.1108/IntR-04-2016-0086.

[11] F. Palomba, M. Linares-Vásquez, G. Bavota, R. Oliveto, M. D. Penta, D. Poshyvanyk, A. D. Lucia, Crowdsourcing user reviews to support the evolution of mobile apps, Journal of Systems and Software (2018). doi:https://doi.org/10.1016/j.jss.2017.11.043.

[12] R. Jongeling, P. Sarkar, S. Datta, A. Serebrenik, On negative results when using sentiment analysis tools for software engineering research, Empirical Software Engineering (2017). doi:10.1007/s10664-016-9493-x.

[13] L. Zhao, et al., Natural language processing for requirements engineering: A systematic mapping study, ACM Comput. Surv. (2021). doi:10.1145/3444689.

[14] J. Dąbrowski, E. Letier, A. Perini, A. Susi, Mining and searching app reviews for requirements engineering: Evaluation and replication studies, Information Systems 114 (2023). doi:10.1016/j.is.2023.102181.

[15] Q. Motger, M. Tiessler, M. Oriol, I. Bertolín, Re-miner: Mining mobile user reviews with fea-

ture extraction and emotion classification, in: Joint proceedings of REFSQ, 2024. doi:10.1007/978-3-031-19759-9_9.
[16] Q. Motger, X. Franch, J. Marco, Mapp-kg: Mobile app knowledge graph for document-based feature knowledge generation, in: Intelligent Information Systems, Springer Nature Switzerland, 2024. doi:10.1007/978-3-031-61000-4_15.
[17] T. Johann, C. Stanik, A. MollaAlizadeh Bahnemiri, W. Maalej, Safe: A simple approach for feature extraction from app descriptions and reviews, in: Proc. 25th IEEE Int. Requirements Engineering Conf. (RE), 2017. doi:10.1109/RE.2017.71.
[18] A. F. Araujo, M. P. S. Gôlo, R. M. Marcacini, Opinion mining for app reviews: an analysis of textual representation and predictive models, Automated Software Engineering (2022). doi:10.1007/s10515-021-00301-1.
[19] N. Alturaief, H. Aljamaan, M. Baslyman, Aware: Aspect-based sentiment analysis dataset of apps reviews for requirements elicitation, in: Proc. 36th IEEE/ACM Int. Conf. on Automated Software Engineering Workshops (ASEW), 2021. doi:10.1109/ASEW52652.2021.00049.
[20] P. Ekman, Basic Emotions, 1999. doi:10.1002/0470013494.ch3.
[21] J. Dąbrowski, et al., Analysing app reviews for software engineering: a systematic literature review, Empirical Software Engineering (2022). doi:10.1007/s10664-021-10065-7.
[22] E. Guzman, W. Maalej, How do users like this feature? a fine grained sentiment analysis of app reviews, in: 2014 IEEE 22nd International Requirements Engineering Conference (RE), 2014. doi:10.1109/RE.2014.6912257.
[23] F. A. Shah, K. Sirts, D. Pfahl, Is the safe approach too simple for app feature extraction? a replication study, in: Requirements Engineering: Foundation for Software Quality, Springer International Publishing, 2019, pp. 21–36.
[24] D. Carneros-Prado, et al., Comparative study of large language models for emotion and sentiment analysis: Gpt vs. ibm watson, in: Proc. 15th Int. Conf. on Ubiquitous Computing & Ambient Intelligence (UCAmI 2023), Springer, 2023.
[25] S. Malgaonkar, C. W. Lee, S. A. Licorish, B. T. R. Savarimuthu, A. Tahir, Appsent: A tool that analyzes app reviews, arXiv (2019). URL: https://doi.org/10.48550/arXiv.1907.10191.
[26] Q. Motger, M. Tiessler, et al., Gessi - nlp4se github, 2024. URL: https://github.com/gessi-chatbots.
[27] ISO/IEC, System and Software Quality Models (ISO/IEC 25010), Technical Report, 2011. URL: https://iso25000.com/en/iso-25010.
[28] C. Wohlin, P. Runeson, M. Hst, M. C. Ohlsson, B. Regnell, A. Wessln, Experimentation in Software Engineering, Springer Publishing Company, Incorporated, 2012.