Implementing Accuracy Quality for Responsible AI in Newsrooms

Roxana L.Q. Portugal[1,2], Juliana Delle Ville[3], Leandro Antonelli[3,4]

[1]Universidad Nacional de San Antonio Abad del Cusco, Cusco, Perú
[2]Ludwig Maximilians Universität, München, Alemania
[3]Lifia, Fac. de Informática, UNLP, La Plata, Bs As, Argentina
[4]UAI CAETI – Facultad de Tecnología Informática- Universidad Abierta Interamericana

roxana.quintanilla@unsaac.edu.pe
{juliana.delleville, leandro.antonelli}
@lifia.info.unlp.edu.ar

Abstract. This paper explores the intersection between software development and journalism, highlighting the fundamental importance of implementing non-functional requirements to achieve a balance between immediacy and accuracy. In software development, requirements encompass user needs and demand precise and continuous updates to meet time-to-market demands. In journalism, audience engagement drives the need for precise news coverage, especially with the growth of artificial intelligence (AI). Non-functional requirements (NFRs) have gained relevance, emphasizing the need for effective balance. This work investigates the integration of technologies such as Named Entity Recognition (NER) and topic modeling into news updating processes as means to enhance both efficiency and precision. Additionally, this strategy, beneficial in requirements management and applicable across domains, is explored. The article is structured to delve into the imperatives of news writing, the proposed strategy, potential applications, and directions for future research.

Keywords: Requirements engineering, Artificial Intelligence (AI), Natural Language Processing (NLP), Journalism, Non-Functional Requirements (NFRs).

1. Introduction

In the realm of software development, requirements are paramount for ensuring the quality and success of a product. These requirements manifest in various forms, encompassing the needs and expectations of users, the necessity for reengineering or refactoring existing software applications, and compliance with laws and regulations.

In journalism, audience engagement with published news performance holds significant weight in deciding upon updates, achieved through the acquisition of information from reliable sources. Consequently, newsrooms strive for heightened accuracy in their coverage, particularly with the advent of artificial intelligence (AI), which has escalated the demand for producing transparent, precise, and dependable news for the audience. This evolution in software requirements has underscored the

significance of Non-Functional Requirements (NFRs) and the need to balance them effectively [2].

The research challenge arises when exploring how technologies such as Named Entity Recognition (NER) and topic modeling can be effectively integrated into the news updating process within media outlets, aiming to balance both efficiency and precision in identifying relevant information and adapting news content based on emerging sources and discoveries. It's worth noting that this strategy is also beneficial in requirements management by incorporating information obtained from big data, thus improving both knowledge and the requirements themselves [3]. Additionally, we find that the contribution of this strategy is useful in other areas such as the legal field, where it seeks jurisprudence on specific facts [4].

The remainder of the document is organized as follows: Section 2 delves into the imperative for precision and immediacy within newsrooms. Section 3 outlines the strategy devised to fulfill this imperative. Section 4 examines potential applications of this strategy in diverse domains. Lastly, we conclude by suggesting potential avenues for future research.

2. Background

The problem we are addressing falls within the realm of AI-driven news curation [5], which involves using AI technology to assist in tasks such as gathering, organizing, filtering, and presenting news. In this context, there are Non-Functional Requirements (NFRs) that influence or constrain these tasks, with three key metrics regarding audience interest: timeliness, relevance, and diversity of stories [6]. In this work, we aim to balance the NFR of recency/immediacy with the precision/accuracy demanded by journalistic rigor.

The first author's experience with the journalism industry[1] confirms this demand by eliciting requirements for responsible fact-checking with AI, using an ad hoc process [7]. The Soft Goals Interdependence Graph (SIG)[8] shown in Figure 1 indicates that *immediacy* negatively impacts *accuracy*. To address this discrepancy, the journalists involved in the project have worked on operationalizations that can help balance *accuracy* with *immediacy*. Figure 2 shows four operationalizations that help improve *accuracy*: i) the four-eyes principle, ii) preventing factual errors, iii) error-free text, and iv) not forgetting news updates. The latter is divided into two operationalizations: a) correction updates and b) coverage updates. This work implements the latter operationalization.

In addition to addressing the *recency/immediacy* of news, our proposal seeks operationalizations to improve the *transparency* of these automations, facilitating their

[1] www.algorithmic.news

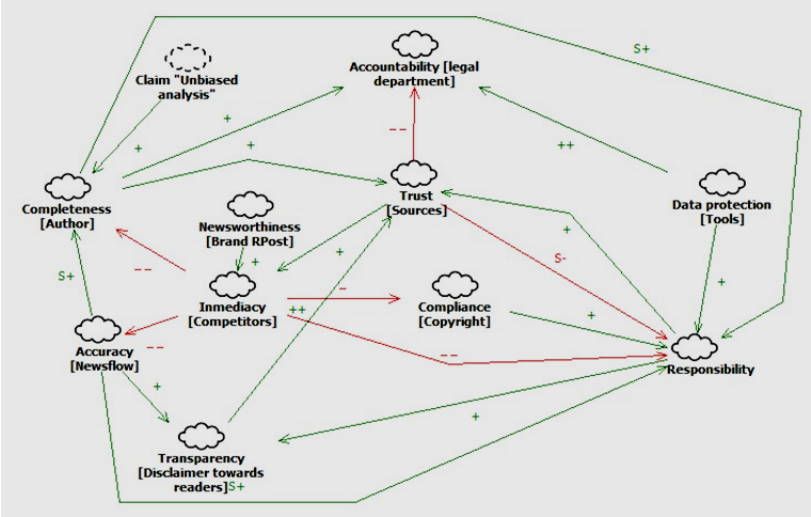

Figure 1. SIG for Responsible Fact Checking

use, understanding, and intervention by journalists when necessary. In this sense, we aim to overcome the limitations of cutting-edge algorithmic techniques that, by operating opaquely, do not contribute to the goals shown in Figure 1. To the best of our knowledge, this problem is being addressed through Machine Learning methods [Wang.] mainly focused on speeding up news production. However, the issue of *responsibility* [1], as often demanded by journalists, is a less explored path, especially in terms of a *transparent* implementation [10].

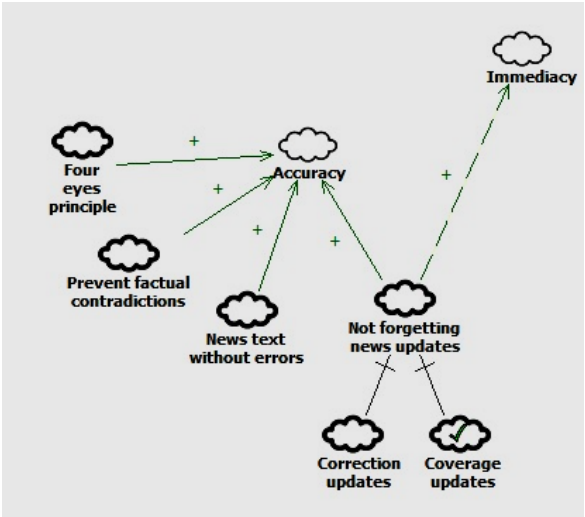

Figure 2. Operationalization Contributing to Accuracy in Responsible Fact-Checking

3. Strategy implementation

In text analysis, it is essential to think about reducing the complexity of automatically synthesizing documents, which would allow efficient comparison of news to obtain greater coverage. In this sense, key elements, such as verbs and nouns, help identify concepts related to people, companies, organizations, locations, dates, and more. These elements are commonly known as entities, which Artificial Intelligence frameworks identify using modules known as Named Entity Recognition (NER). NER comprises trained models that can identify specific concepts such as people and organizations, using predefined rules to recognize dates.

Another valuable technique for extracting high-level concepts from text is topic modeling, which encapsulates the main themes discussed in a given text. Various algorithms, including both supervised and unsupervised machine learning models, are employed to conduct topic modeling. For instance, Sajid et al. [11] utilize topic modeling to summarize short texts, Kim et al. [12] leverage it to uncover new topics for text mining, and Li et al. [13] employ it for tagging purposes. One widely used model for topic extraction is Latent Dirichlet Allocation (LDA), an unsupervised model that identifies topics based on the text corpus.

To achieve a balanced implementation that satisfices[2] *accuracy*, *transparency*, and *efficiency* in identifying news articles that enhance coverage of a pivot news item, we evaluate the similarity of texts. The greater the similarity of a document to the pivot news, the less additional value it contributes to the overall coverage. We combine named entity recognition (NER) techniques with latent Dirichlet allocation (LDA) to improve *accuracy* and *efficiency*, although these techniques cannot guarantee *transparency*. Our methodology streamlines the document comparison process through a structured pipeline. Initially, the texts undergo preprocessing, which involves converting the text to lowercase, removing stop-words to eliminate irrelevant words, and stemming to unify conjugated forms. Following preprocessing, named entity recognition (NER) is applied to construct a set of entities for each document. Subsequently, latent Dirichlet allocation (LDA) is utilized on these entities to generate a set of topics for each document. Finally, Jaccard's similarity measure is employed to compare the sets of topics across documents. To evaluate the proposed strategy, 15 news articles were tested, including reports on incidents such as MH370, AF447, and Nepal, all related to flights and climbing.

News	Similarity Index
MH370 plane crash	0.8571428571428571
NEPAL plane crash	0.7916666666666666
NEPAL plane crash	0.7391304347826086
MH370 plane crash	0.72
top 10 plane crashes	0.6521739130434783
MH370 plane crash	0.64
EVEREST climb	0.64
MH370 plane crash	0.6363636363636364
EVEREST climb	0.625
EVEREST climb	0.5833333333333334
AF447 plane crash	0.5769230769230769
AF447 plane crash	0.56
MH370 plane crash	0.5555555555555556
MH370 plane crash	0.5517241379310345
MH370 plane crash	0.48148148148148145

Figure 3. Similarity index of NER + LDA + Jaccard strategy

2

The results in Figure 3 indicate a significant deviation in *accuracy*, particularly in texts with high similarity indices that are not related to flight MH370 (hereinafter referred to as false positive texts). Although syntactic processing suggests similarities in shared thematic content, nuanced distinctions emerge when examining the specificity of the flights boarded. To address these distinctions, we evaluated two individual strategies: NER and LDA. Figure 4 presents the results of all tested strategies. Additionally, to enhance *transparency*, we processed our texts using Term Frequency-Inverse Document Frequency (TF-IDF), which can provide more *transparency* and *accountability* compared to the other strategies. However, TF-IDF yielded the worst results among all the methods tested.

N	News_ID	NER	LDA	NER+LDA	TF-IDF	Gold Standard measured with Likerd scale (1-5)
1	MH370 plane crash (1)	0,13	0,11	0,64	0,58	1
2	MH370 plane crash (2)	0,20	0,25	0,56	0,00	1
3	MH370 plane crash (3)	0,50	0,18	0,64	0,00	2
4	MH370 plane crash (4)	0,17	0,54	0,86	0,00	5
5	MH370 plane crash (5)	0,18	0,18	0,55	0,00	4
6	MH370 plane crash (6)	0,17	0,11	0,48	0,00	3
7	MH370 plane crash (7)	0,40	0,33	0,72	0,00	3
8	AF447 plane crash (1)	0,05	0,11	0,58	0,00	1
9	AF447 plane crash (2)	0,11	0,11	0,56	0,00	1
10	Top 10 plane crashes	0,03	0,11	0,65	0,00	1
11	NEPAL plane crash (1)	0,03	0,11	0,74	0,00	1
12	NEPAL plane crash (2)	0,07	0,05	0,79	0,00	1
13	EVEREST climb (1)	0,04	0,00	0,63	0,06	1
14	EVEREST climb (2)	0,03	0,00	0,64	0,06	1
15	EVEREST climb(3)	0,00	0,00	0,58	0,00	1

Color legend
True positive (TP)
TP: is close to the equivalence parameters (+/- 0.1)
TP: is acceptable if we evaluate its rank among the values of the strategy to which it belongs. For example, 0.54 in LDA represents the highest value within the group.
False positive (FP)
Those that give more coverage to the pivot news

Equivalence for assessment	
result index	Likert scale
0 and 0.2	1
0.2 and 0.4	2
0.4 and 0.6	3
0.6 and 0.8	4
0.8 and 1	5

Figure 4. Comparison of results of all tested strategies

Based on our current understanding, we are addressing a problem with diverse approaches when considering Non-Functional Requirements (NFRs) as primary requirements. Employing Recall, Precision, and F1-score as metrics to *satisfice* accuracy, the Latent Dirichlet Allocation (LDA) strategy yields the best performance, particularly when evaluating the harmonic mean F1-score (Fig. 5).

A detailed and granular analysis of each strategy reveals that NER achieves the highest precision at 100%. Conversely, the combined NER + LDA strategy attains the highest Recall at 80%, positioning it as the optimal strategy, especially when considering the significance of Recall in Natural Language Processing (NLP) tasks for Requirements Engineering. As cited by D. Berry [14], “A tool that falls short of close to 100% recall, applied to the development of a high-dependability system, may even

be useless, because to find the missing information, a human has to do the entire task manually anyway."

To this point, the strategies barely satisfice *accuracy*. Nevertheless, *efficiency* is achieved across all strategies given the limited corpus of 15 texts. However, none of the strategies can satisfy the criterion of *transparency*, as both NER and LDA algorithms are often perceived as black boxes due to the inherent complexity in interpreting their results.

NER	Prediction Relevant	Prediction Non-Relevant
Actual-Relevant	1	4
Actual-Non-Relevant	0	10

LDA	Prediction Relevant	Prediction Non-Relevant
Actual-Relevant	3	2
Actual-Non-Relevant	1	9

NER+LDA	Prediction Relevant	Prediction Non-Relevant
Actual-Relevant	4	1
Actual-Non-Relevant	10	0

TF-IDF	Prediction Relevant	Prediction Non-Relevant
Actual-Relevant	0	5
Actual-Non-Relevant	1	9

	NER	LDA	NER-LDA	TF-IDF
Precision	100%	75%	28,60%	0
Recall	20%	60%	80%	0
F1-Score	33,3%	66,7%	42,1%	0

Figure 5. Confusion matrix of all tested strategies

The proposed strategy is implemented through a Python script utilizing NLP libraries such as Spacy [15] and NLTK [16], along with the Keras [17] library for creating and training an unsupervised LDA model. This algorithm code is available at GitHub.

4. Applications beyond journalism

The described text analysis workflow, which comprises techniques such as entity extraction, topic modeling, and result ranking, extends its relevance beyond journalism into diverse domains, including the legal sector. In legal contexts, this approach streamlines document review processes, aiding in the analysis of legal documents, contracts, and court rulings. By extracting entities such as legal terms, entities, and dates, legal professionals can efficiently identify relevant information, track case precedents, and conduct comprehensive legal research. Moreover, it facilitates *compliance* monitoring by automatically analyzing regulatory documents and identifying compliance risks [18]. The ability to uncover patterns, trends, and relationships within textual data empowers legal professionals to make informed decisions, streamline workflows, and ensure regulatory adherence. Thus, the application of this text analysis workflow in the legal domain can enhances *efficiency*, *accuracy*, and *compliance*, offering significant benefits to legal practitioners and organizations alike.

5. Conclusion

Our proposal distinguishes itself from the state of the art by not relying on predefined models that could overlook relevant information if it does not fit the parameters set in the algorithm. This limitation could lead to a biased or incomplete presentation of information, thus affecting the quality and objectivity of journalistic reports. In contrast, as seen in Figure 4, our methodology is based on domain-independent and more syntactic strategies. These strategies are designed to meet the demands of journalists. Furthermore, by gradually refining a corpus, each step of the process can be intervened and verified by the user, contributing to the *transparency* and *accuracy* required.

It is worth noting that this strategy can be useful not only in the journalistic field but also for requirements management. Identifying similar requirements, merging them, or enriching them can help create products that rely on time to market [19] for their success. The ability to continuously adapt and improve requirements throughout the product lifecycle is crucial to ensuring competitiveness and relevance in a constantly changing business environment.

As for future work, we aim to implement a minimum viable product that can be used in newsrooms for validation. Additionally, we seek to explore further the implementation of Non-Functional Requirements (NFRs) identified in the journalistic environment, as well as within the framework of the mentioned project.

Acknowledgment

We extend our sincere gratitude to the Volkswagen Foundation for sharing data, and Portugal appreciates the support of the research project "Towards Responsible AI in Local Journalism" with grant number 9B761 and 9B390.

References

1. Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62. https://doi.org/10.1145/2844110
2. Portugal, R. L. Q. (2020). Speeding-Up Non-Functional Requirements Elicitation.
3. Henriksson, A., Zdravkovic, J. Holistic data-driven requirements elicitation in the big data era. *Softw Syst Model* **21**, 1389–1410 (2022). https://doi.org/10.1007/s10270-021-00926-6.
4. Mentzingen, H., António, N., & Bacao, F. (2023). Automation of legal precedents retrieval: findings from a literature review. *International Journal of Intelligent Systems*, *2023.*
5. Manisha, R., & Acharya, K. (2023). The Impact of Artificial Intelligence on News Curation and Distribution: A Review Literature. Journal of Communication and Management, 2(01), 23-26.
6. Chakraborty, A., Luqman, M., Satapathy, S., & Ganguly, N. (2018, April). Editorial algorithms: Optimizing recency, relevance and diversity for automated news curation. In Companion Proceedings of the The Web Conference 2018 (pp. 77-78).
7. Portugal, R.L.Q., Wilczek, B., Eder, M., Thurman, N., & Haim, M. (2023). Design Thinking for Journalism in the AI age: Towards an Innovation Process for Responsible AI

Applications. In *The Joint Computation+ Journalism European Data & Computational Journalism Conference* (pp. 22-24).
8. Chung, L., Nixon, B. A., Yu, E., & Mylopoulos, J. (2012). *Non-functional requirements in software engineering* (Vol. 5). Springer Science & Business Media.
9. Wang, H. C., Chen, C. C., & Li, T. W. (2022). Automatic content curation of news events. *Multimedia Tools and Applications*, *81*(8), 10445-10467.
10. Portugal, R. L. Q., Engiel, P., Roque, H., & do Prado Leite, J. C. S. (2017, September). Is there a demand of software transparency?. In *Proceedings of the XXXI Brazilian Symposium on Software Engineering* (pp. 204-213).
11. A. Sajid, S. Jan and I. A. Shah, "Automatic Topic Modeling for Single Document Short Texts," 2017 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 2017, pp. 70-75, doi: 10.1109/FIT.2017.00020. keywords: {Text mining;Tools;Tagging;Natural language processing;Social network services;Speech;Data Mining;Text mining;Topic Modeling},
12. Hyun Duk Kim, Malu Castellanos, Meichun Hsu, ChengXiang Zhai, Thomas Rietz, and Daniel Diermeier. 2013. Mining causal topics in text data: iterative topic modeling with time series feedback. In Proceedings of the 22nd ACM international conference on Information & Knowledge Management (CIKM '13). Association for Computing Machinery, New York, NY, USA, 885–890. https://doi.org/10.1145/2505515.2505612
13. F. Li, H. Shen and T. He, "Tag-topic model for semantic knowledge acquisition from blogs," 2011 7th International Conference on Natural Language Processing and Knowledge Engineering, Tokushima, Japan, 2011, pp. 221-226, doi: 10.1109/NLPKE.2011.6138198.
14. Berry, D. M. (2017). Evaluation of tools for hairy requirements engineering and software engineering tasks. *School of Computer Science, University of Waterloo, Tech. Rep*.
15. Spacy, Industrial-Strength Natural Language Processing, https://spacy.io/, last accessed March 30 2024.
16. NLTK, Natural Language Toolkit, https://www.nltk.org/, last accessed March 30 2024.
17. Keras, https://keras.io/, last accessed March 30 2024
18. Engiel, P., Leite, J. C. S. D. P., & Mylopoulos, J. (2017, May). A tool-supported compliance process for software systems. In *2017 11th International Conference on Research Challenges in Information Science (RCIS)* (pp. 66-76). IEEE.
19. Portugal, R. L. Q., do Prado Leite, J. C. S., & Almentero, E. (2015, August). Time-constrained requirements elicitation: reusing GitHub content. In 2015 IEEE Workshop on Just-In-Time Requirements Engineering (JITRE) (pp. 5-8). IEEE.