

Table 5 summarizes the KPIs values for the Articles Analyzing Process. This process aims to analyze the aggregated news articles in order to classify them accordingly based on the news categories (as shown in figure 4), locations, interrelated sets of articles, breaking news, and relevant to a particular user's inquiries. This set of KPIs measures the accuracy of classifying and clustering the fetched news. Results show that the percentage of articles that clustered in a wrong cluster/category is about 30%. The percentage of combined sub-clusters or sub-topics in a cluster/topic is about 16%. This means that the total number of the clustered topics could be increased by 16%. This could be achieved by clustering the sub-clusters that found in a cluster to separate additional clusters. For example, in technology news, a cluster about "New features" is found. However, analysis showed that this cluster could be sub-classified to sub-clusters, where one individual cluster for the new features news for each app or software. Results also show that there are about 30% of news are located to a false or not-accurate nearby location. It is important to mention that the topic modelling and clustering algorithms are still under research and development in TAKA system. Applying the proposed performance evaluation system enabled us to emphasis the shortage in this process performance. However, TAKA technical teams are targeting improving the performance of this process very shortly.

Table 5. The KPIs values for the Articles Analyzing Process.

Measures	Value
The percentage of articles that clustered in a wrong cluster/category	30%
The percentage of combined sub-clusters or sub-topics in a cluster/topic	16%
False positive related news articles	19%
False positive nearby news articles	30%
Repeated top news	08%
Redundant top news	17%

4.3. Discussion

Generally, the evaluation results show that TAKA system performs well in most functions. However, during the study, we did observe number of strengths, weaknesses and challenges, which are discussed as below.

Applying the proposed performance evaluation system enabled us to identify and detect the performance shortages in each of the four processes. Furthermore, it enabled the traceability of those issues. Root cause analysis is conducted to investigate the reasons behind those shortages. Then, a set of corrective and proactive actions are taken to handle those issues. TAKA technical teams are targeting to automate the traceability system to be able to detect, analyze, investigate and take actions for any of those issues could exist in TAKA system.

Moreover, applying the proposed performance evaluation system emphasized the challenges of analyzing Arabic texts in comparison with other languages like, i. e., English. This issue is one of our main concerns because we target the Middle East market. The Arabic language, as a natural language, is more challenging in clustering and topic modelling rather than other languages. This could be explained due to the different writing styles and the lack of standards.

It is important to mention that results, in particularly Figure 5, show that media focuses a lot in politics and less about health, environmental and technology. This could raise different research questions about the interrelationship between the media main concerns and the development of our life.

5. Conclusion

In this paper, an ontology-based performance evaluation system for TAKA system is developed. The enterprise engineering approach DEMO is applied to describe, understand and analyze the main processes in TAKA system. Then, a performance evaluation system is proposed for TAKA system by developing a set of KPIs for each process. The paper contribution is to present a performance evaluation system for the entire news monitoring system from user, enterprise and systems engineering point of views. On contrast, studies in the literature consider only the users' perspectives. This study enabled us to identify and detect the performance shortages in each of the four processes. Furthermore, it enabled the traceability of those issues. Moreover, it emphasized the challenges of analyzing Arabic texts. It is also showed that media focuses a lot about politics and less about technology. Future work may focus on tackling those issues. Next, we may analyse the media main focus in each country/region and investigate the interrelationship between the media main concerns and the development.

Acknowledgements

The authors wish to acknowledge the University of Johannesburg for providing financial support. The authors would like also to acknowledge SMAgroup teams and, in particular, TAKA development and technical teams for the direct technical support and the rich discussions.

References

- Bokhari, M.U. and Adhami, M.K., A new criterion for evaluating news search systems, *Communications*, vol. 2, no. 7, pp. 28-35, 2015.
- Bokhari, M.U. and Adhami, M.K., How well they retrieve fresh news items: News search engine perspective, *Perspectives in Science*, vol. 8, pp. 469-471, 2016.
- Curtiss, M., Bharat, K. and Schmitt, M., Systems and methods for improving the ranking of news articles, U.S. Patent 7,577,655, Google Inc, 2009.
- Dietz, J.L., *What is Enterprise Ontology?* Springer Berlin Heidelberg, pp. 7-13, 2006.
- Doğan, G., From Newspapers to News Search Systems: Turning Libraries “Inside-Out”, *Proceedings of the 21st International BOBCATSSS Conference*, Ankara, Turkey, January 23-25, 2013.
- Gulli, A., The anatomy of a news search engine, *In Special interest tracks and posters of the 14th ACM international conference on World Wide Web*, Chiba, Japan, May 10-14, 2005.
- Liu, K.L., Meng, W., Qiu, J., Yu, C., Raghavan, V., Wu, Z., Lu, Y., He, H. and Zhao, H., AllInOneNews: development and evaluation of a large-scale news metasearch engine, *In Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, Beijing, China, June 11-14, 2007.
- Öcalan, H.Ç., *Bilkent News Portal: A system with new event detection and tracking capabilities*, master’s thesis, Computer Engineering Department, Bilkent University, 2009.
- Uyar, E., *Near-duplicate news detection using named entities*, master’s thesis, Department of Computer Engineering, Bilkent University, 2009.

Biographies

Tarek Fatyani is a Researcher and a Business Analyst. He received his PhD in Industrial Engineering and Management from Tokyo Institute of Technology 2016. His research was about business process analysis and how is it related to software development. He established his company in Japan SMAGROUP LLC in 2015 while he was a PhD candidate. His research and experience are related to virtual collaboration tools, innovation management, project management, enterprise engineering, modelling and software development. Currently, he is focusing on utilising machine learning and data analysis to advance the software that he is leading.

Zakaria Yahia received the M.Sc. and Ph.D. degrees in Industrial Engineering from Cairo University, Giza, Egypt, in 2012 and Egypt-Japan University of Science and Technology (E-JUST), Alexandria, Egypt, in 2015, respectively. As a visiting Ph.D. student, he spent one academic year at the Tokyo Institute of Technology (TITECH), Tokyo, Japan, working on the research project “Developing a Design and Engineering Methodology for Organization (DEMO)-based simulation model for surgery room system”. From 2015 to 2017, he was an Assistant Professor with the Department of Mechanical Engineering, Fayoum University, Fayoum, Egypt. Currently, he is a Post-Doctoral Researcher with the Department of Quality and Operations Management, University of Johannesburg, South Africa. His research interests include the areas of Applied Operations Research & Simulation, Scheduling, Healthcare Management, Smart Grid Management and DEMO-Enterprise Ontology.