

Alian M. Awajan A., Al-Hasan A., Akuzhia R. 2019. Towards Building Arabic Paraphrasing Benchmark. Proceedings of DATA'19, December 2–5, 2019, Dubai, United Arab Emirates. Copyright 2019 Association for Computing Machinery. ACM ISBN 978-1-4503-7284-8/19/12. DOI: <http://doi.org/10.1145/3368691.3368708>.

## **Abstract**

This research describes a paraphrasing benchmark of Arabic sentences for evaluating algorithms developed to measure Semantic Similarity for sentences and Paraphrasing identification task. The sentences are formed based on a set of rules for Arabic paraphrasing. These sentences are constructed from words of different types of Arabic books; educational, semantic science, lexicons and AWSS dataset. The benchmark consists of 1011 sentence pairs and labeled with human ratings for semantic similarity between sentence pair.