

Documents

Bakar, N.S.A.A.

The Development of an Integrated Corpus for Malay Language

(2020) *Lecture Notes in Electrical Engineering*, 603, pp. 425-433.

DOI: 10.1007/978-981-15-0058-9_41

International Islamic University Malaysia, Kuala Lumpur, Malaysia

Abstract

Generally, a corpus serves as the source of data for various types of research. As such, there are several Malay corpora being developed to support the needs of the researchers. However, the various corpora of Malay text are distributed and not integrated, where some words are not included or missing in some corpora. The focus of this paper is to develop an integrated corpus that will combine four most comprehensive Malay corpora. The intention is to provide comprehensive coverage of Malay corpora which would be beneficial for any relevant work. © 2020, Springer Nature Singapore Pte Ltd.

Author Keywords

Malay corpora; Mobile application; Sentiment analysis; Web crawler

Index Keywords

Sentiment analysis; Malay corpora, Malay languages, Malay texts, Mobile applications; Web crawler

Funding details

RIGS-17-156-0731

Funding details

This research is fully funded by the International Islamic University's RIGS grant, ID: RIGS-17-156-0731.

Correspondence Address

Bakar N.S.A.A.; International Islamic University MalaysiaMalaysia; email: nsham@iium.edu.my

Editors: Alfred R., Lim Y., Haviluddin H., On C.

Publisher: Springer Verlag

ISSN: 18761100

ISBN: 9789811500572

Language of Original Document: English

Abbreviated Source Title: Lect. Notes Electr. Eng.

2-s2.0-85072977332

Document Type: Conference Paper

Publication Stage: Final

Source: Scopus