

Farasa: A Fast and Accurate Text Processing Toolkit for the Arabic Language

AUTOMATING ARABIC TRANSLATION

Farasa is a state-of-the-art full-stack package to process Arabic Language to enable downstream applications.

Farasa includes: word segmentation, lemmatization, part-of-speech tagging, named entity recognition, syntactic and dependency parsing, and diacritic recovery.

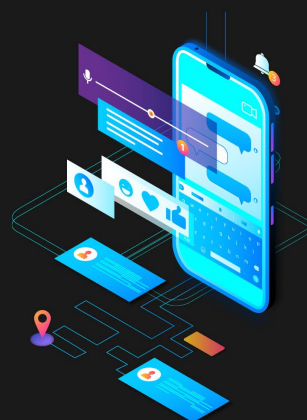
Such fundamental processing modules are essential for a myriad of applications such as search, machine translation, question answering, information extraction, dialogue systems, text classification, text-to-speech, and many others.

All Farasa components achieve state-of-the-art results while being highly efficient and fast.

A COMPREHENSIVE SOFTWARE SOLUTION

Farasa is an HBKU Technology that is available in the form of Java jars that are completely self-contained with no external dependencies and in the form of RESTful web API that can be accessed from any programming language.

Internally, Farasa utilizes a variety of classification techniques including support vector machines, conditional random fields, recurrent neural network, and sequence-to-sequence transformer models.



APPLICATIONS

- Arabic translation
- Text processing research
- Automated machine translation and information retrieval



VALUE PROPOSITIONS

Fast: Processes 1 billion words in less than five hours— at least one order of magnitude faster than leading translation tools

Accurate: Outperforms or equals state-of-the-art Arabic segmenters

Comprehensive: Includes features that account for likelihoods of stems, prefixes, suffixes, and their combinations; presence in lexicons containing valid stems and named entities; and underlying stem templates

Sophisticated: Offers advancements in speed and accuracy over other contemporary automatic translation systems, while maintaining the significant cultural traditions of Arabic translation

Adaptable: Has the potential to translate other complex languages



PATENT STATUS

A copyright exists for this technology.



LICENSING OPPORTUNITIES

Hamad Bin Khalifa University is offering this technology for license. For more information, please contact: innovation@hbku.edu.qa