Robustness of Cross-Lingually Trained Dense Retrievers Against Domain Shifts on Monolingual WebFAQ Subsets

Assignee Research

June 11, 2026

Abstract

Cross-lingual representations of words enable us to reason about word meaning in multilingual contexts and are a key facilitator of cross-lingual transfer when developing natural language processing models for low-resource languages. In this survey, we provide a comprehensive typology of cross-lingual word embedding models. We compare their data requirements and objective functions. The recurring theme of the survey is that many of the models presented in the literature optimize for the same objectives, and that seemingly different models are often equivalent, modulo optimization strategies, h

1 Introduction

This paper examines: A Survey of Cross-lingual Word Embedding Models. Research question: What is the robustness of cross-lingually trained dense retrievers against domain shifts when evaluated on specific monolingual subsets of the WebFAQ dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.8/10.

3 Results

11 papers retrieved. 6 claims extracted; 6 independently verified. Quality review score: 8.8/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Cross-lingual representations of words enable reasoning about word meaning in multilingual contexts.	✓	0.31
Cross-lingual representations are a key facilitator of cross-lingual transfer when developing natural language processin	✓	0.43
Many cross-lingual word embedding models optimize for the same objectives.	✓	0.36
Seemingly different cross-lingual word embedding models are often equivalent, modulo optimization strategies, hyper-para	✓	0.47
The survey discusses different ways cross-lingual word embeddings are evaluated.	✓	0.32
The survey highlights future challenges and research horizons in cross-lingual word embeddings.	✓	0.30

References

- https://doi.org/10.1613/jair.1.11640
- https://doi.org/10.1145/3123266.3123326
- https://doi.org/10.17863/cam.30462