The presented code is part of a Master's thesis project for the Social Sciences and Computing programme at the University of Belgrade. The research description can be found below.

Research description:

Headline: Automatic Answer Assessment Using Latent Semantic Analysis within Digital Flashcards

Implementing technology in a modern-day classroom is an ongoing challenge. In this paper, we created a system for an automatic assessment of student answers using Latent Semantic Analysis (LSA) – a method with an underlying assumption that words with similar meanings will appear in the same contexts. The system will be used within digital lexical flashcards for L2 vocabulary acquisition in a CLIL classroom. Results presented in this paper indicate that while LSA does well in creating semantic spaces for longer texts, it fell somewhat short of detecting topics in answers and word definitions. The answers were classified using KNN, for both binary and multinomial classification. The results of KNN classification are as follows - 73% (0.73) precision, 100% (1) recall, 0.85 F1 score for binary classification, and 50% (0.50), 47% (0.47), 0.46 precision, recall, and F1 score for the multinomial classifier, respectfully. The results are to be taken with a grain of salt, due to a small test and training dataset.