Machine Learning and Hebrew NLP for Automated Assessment of Open-Ended Questions in Biology

Tanya Nazaretsky, Moriah Ariely, *& Giora Alexandron*
Weizmann Institute of Science, Israel

Machine learning algorithms that automatically score open-ended questions can be used to measure students' conceptual understanding, identify gaps in their reasoning, and provide them with timely and individualized feedback. This talk will present the results of a study that uses Hebrew NLP to automatically score students' open-ended questions in Biology. The experimental results show that our algorithms achieve a high-level of agreement with human experts, on par with related work in English, in which this area is well-established. The contribution is twofold. First, we present a conceptual framework for constructing grading rubrics that are designed to support automated guidance and are geared towards machine learning-powered automated assessment. Second, we use this approach to build an NLP-pipeline for a new context – Hebrew, which belongs to a group of languages known as Morphologically-Rich. In languages of this group, among them also Arabic and Turkish, each input token may consist of multiple lexical and functional units, making them particularly challenging for NLP. This is the first study on automatic assessment of open-ended questions in Hebrew, and among the firsts to do so in Morphologically-Rich Languages.