UNIVERSITY OF SOUTH FLORIDA

Defense of a Master's Thesis

Turkic Interlingua: A Case Study of Machine Translation in Low-Resource Languages

by

Jamshidbek Mirzakhalov

For the MSCS degree in Computer Science

Machine Translation (MT) has the potential to bridge the gap between the developed world and the marginalized communities by making information more accessible in real-time. While there are over 7000 spoken languages in the world, only about a hundred have access to high-quality MT systems and even fewer enjoy the benefits of more advanced language technologies. Unfortunately, resource scarcity and the lack of digital infrastructure are only some of the many challenges associated with globalizing NLP. Many large-scale multilingual studies and corpora often get little to no feedback from native speakers or linguistic experts of the languages involved, leading to serious problems of data quality and potential biases. In this thesis, we present a case study of participatory research in 22 Turkic languages involving native speakers, language technologists, researchers, linguists, commercial entities, and more. As accomplished through the thesis, we outline the curation and release of public datasets, development of machine translation technologies, and their deployment in real-world scenarios. In addition, we discuss the lessons learned through this case study, its applications, and limitations, as well as implications for future projects.

Friday, May 21st, 2021

Time: 10:00 am

Online (MS Teams)

Please email for more information

mirzakhalov@usf.edu

THE PUBLIC IS INVITED

Examining Committee

Sriram Chellappan, Ph.D., Co-Major Professor
John Licato, Ph.D., Co-Major Professor
Marvin Andujar, Ph.D.

Robert Bishop, Ph.D.
Dean, College of Engineering

Dwayne Smith, Ph.D.
Dean, Office of Graduate Studies

Disability Accommodations:

If you require a reasonable accommodation to participate, please contact the Office of Diversity & Equal Opportunity at 813-974-4373 at least five (5) working days prior to the event.