**Tyumen State University**

e.v.mikhalkova@utmn.ru

# Russian Jeopardy! Data Set for Question-Answering Systems

**Elena Mikhalkova, Alexander Khlyupin**

## Overview

Open-domain question answering is a branch of automatic question answering that focuses on systems that can answer questions on all kinds of topics. Even if these questions do not come in the form of an actual question. Quizzes often contain non-trivial questions and answers, and there are large collections of them on the Internet. As Jordan Boyd-Graber and Benjamin Börschinger suggest in their article "What Question Answering can Learn from Trivia Nerds", sadly, these collections are ignored, although they can provide interesting insights on open-domain QA. Among others, the database of the quiz "Own Game" (a Russian analogue of Jeopardy!) is a valuable resource of about 30,000 pairs of questions and answers. The existing open-source Russian QA data sets are more like trivia questions and answers resembling TREC. In the “Own Game” dataset nearly all the questions are in the form of statements and the object of interest, about which the question is asked, is often capitalized. Questions are short and fact-oriented as they are meant for single players competing against each other.

## Example Question

**Topic**: OST
**Question**: Van Gogh, Gauguin, and Toulouse-Lautrec belonged to THIS movement.
**Answer**: Post-impressionism

## Our Contribution

1. **Dataset of Russian Jeopardy! (Own Game) Q&A**
2. **Description of some of its linguistic features**
3. **Offline challenge and ML competition based on the dataset**
4. **Baseline, evaluation and sets for developers**

## Link to Dataset and Information on the Challenge

https://github.com/evrog/Russian-QA-Jeopardy