

# **Corpus-based topic modeling for the cognitive study of the 21st century sociocultural challenges**

**Vera Zabotkina**

Russian State University for the Humanities, Moscow, Russian Federation

**Boris M. Velichkovsky**

Kurchatov Institute, Moscow, Russian Federation

**Artemy Kotov**

Kurchatov Institute, Moscow, Russian Federation

**Dmitry Orlov**

Russian State University for the Humanities, Moscow, Russian Federation

**Alexander Piperski**

Russian State University for the Humanities, Moscow, Russian Federation

**ELENA POZDNYAKOVA**

Russian State University for the Humanities, Moscow, Russian Federation

## **Abstract**

The results were obtained in the course of a two-stage study. At the first stage (2018) linguists analyzed the conceptual domain sociocultural challenges on the basis of purposely elaborated Russian language THREAT-corpus (10.4 m words) and built a frame of the domain. At the second stage (2018-2019) the research was carried out with methods of automated topic modeling for two Russian language corpora: THREAT-corpus and alternative corpus collected using WebBootCaT tool in the SketchEngine corpus management system. Methods of topic modeling (PLSA, LDA, BigARTM et al.) allowed eliciting thematic profiles for texts of both corpora. Comparison of two datasets was carried out by applying set theory, graph theory, and probabilistic analysis. Combining topic modeling with linguistic frame analysis resulted in more precise configurations of cognitive models in the conceptual domain sociocultural challenges. Word frequency for lexemes manifesting sociocultural challenges proved to be an important factor of conceptual structures representation.