Iztok Kosem

TRENDI – A MONITOR CORPUS OF SLOVENE

Abstract In this paper we present Trendi, a monitor corpus of written Slovene, which has been compiled recently as part of the SLED (Monitor corpus and related resources) project. The methodology and the contents of the corpus are presented, as well as the findings of the survey that aimed to identify the needs of potential users related to topical language use. The Trendi corpus currently contains news articles and other web content from 110 different sources, with the texts being collected and linguistically annotated on a daily basis. The corpus complements Gigafida 2.0, a 1.13-billion-word reference corpus of standard written Slovene. Also discussed are the ways in which the corpus will be integrated into various lexicographic projects, helping not only in the identification of neologisms but also in monitoring changes in already identified language phenomena.

Abstract Monitor corpus; language use; trends; Slovene; neologisms; lexicography; newsfeed

Contact information

Iztok Kosem
Jožef Stefan Institute & Faculty of Arts, University of Ljubljana
iztok.kosem@ijs.si

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.