

Анализа информација из инфокутија српских књижевника на Википедији на српском језику

Марина Нешовић

Трећа београдска гимназија, Његошева 15, Београд

e-mail: marina82nesovic@gmail.com

проф. др Душан Тошић

Математички факултет, Студентски трг 16, Београд

e-mail: dtosic@matf.bg.ac.rs

Ђорђе Стакић

Математички факултет, Студентски трг 16, Београд

e-mail: djordjes@matf.bg.ac.rs

Апстракт. Википедија на српском језику је основана 2003. године, и у септембру 2015. године има нешто преко 324 хиљаде чланака. Она је једна од око 290 различитих језичких Википедија, по броју чланака се налази на 28. месту. По броју одредница и учесника ово је највећи енциклопедијски пројекат на српском језику. Значајан део сваке енциклопедије, па и Википедије на српском језику, чине биографски чланци. У биографским чланцима често су присутне инфокутије (infobox) које садрже основне податке о тој особи. Информације инфокутије су организоване у форми парова: атрибут – вредност. У раду су анализирани информације прикупљене из инфокутија српских књижевника прикупљене на Википедији на српском језику. Представљени су подаци који су добијени на тај начин, анализирани су по више критеријума. Сви подаци су добијени применом програмских скрипти написаних у језику Пајтон. Ресурси добијени на овај начин могу представљати добру полазну основу за семантички екстраховане информације. Методологија примењена у анализи информација из биографија српских књижевника на основу инфокутија је таква да се може применити и на друге категорије биографија са Википедије.

Кључне речи: Википедија, биографије, семантички веб, екстракција информација