



ZVEZA BIBLIOTEKARSKIH DRUŠTEV
SLOVENIJE

Turjaška ulica 1, 1000 Ljubljana
tel.: 01 2001-176, faks: 01 4257-293, e-pošta: info@zbds-zveza.si

Kongres ZBDS 2019
Knjižnice – obvladovalke podatkov?
(25.–27. september 2019, IZUM, Maribor)

**RAZVOJ METODOLOGIJE ZA SAMODEJNO VSEBINSKO OZNAČEVANJE
ELEKTRONSKIH PUBLIKACIJ Z VRSTILCI UNIVERZALNE DECIMALNE
KLASIFIKACIJE (UDK)**

**DEVELOPMENT OF THE METHODOLOGY FOR AUTOMATIC CONTENT
INDEXING OF ELECTRONIC PUBLICATIONS WITH THE UNIVERSAL
DECIMAL CLASSIFICATION (UDC)**

Matjaž Kragelj
(Narodna in univerzitetna knjižnica)

Knjižnično gradivo predstavlja enega od temeljev naše kulturne dediščine, je dokaz in merilo našega ustvarjanja, priča o našem razvoju, delovanju, bivanju in ustvarjanju v nekem času in prostoru. Najbogatejši in najpopolnejši vir digitalizirane in digitalne "Slovenike" na spletu predstavlja portal Digitalna knjižnica Slovenije (www.dlib.si), ki vsebuje več kot petsto tisoč elektronskih publikacij.

Težava, ki smo ji priča, je plod dejstva, da za več sto ali tisoč člankov oz. izvodov revij obstaja zgolj po en zapis v knjižničnem katalogu COBIB. Kot primer lahko navedemo serijsko publikacijo »Laibacher Zeitung« (več kot 58.000 števil in zgolj en zapis v knjižničnem katalogu), »Ljubljanski Zvon« (več kot 11.000 člankov – en zapis v knjižničnem katalogu) ter »Dom in svet« (več kot 16.000 člankov, ki delijo enako usodo).

Vsi članki serijske publikacije »Dom in svet«, ki je izhajala med leti 1888 in 1944, so umeščeni v enotno klasifikacijo UDK in sicer v podskupini Slovenska književnost, ter Kultura (821.163.6 in 008). To pomeni, da je tudi vsak članek znotraj revije »Dom in svet« okarakteriziran kot "Slovenska književnost" in "Kultura" in nič več kot to.

V raziskavi, ki smo jo opravili in katere rezultate bom predstavil na posvetovanju, je poskus avtomatske klasifikacije poljubne publikacije (v našem primeru člankov iz Digitalne knjižnice Slovenije) v Univerzalno decimalno klasifikacijo – UDK.

S pomočjo umetne inteligence, natančneje postopkov NLP (Natural Language Processing), bomo prikazali, kako se s problemom klasifikacije spopadajo algoritmi strojnega učenja. Uporabili smo pristop nadzorovanega in nenadzorovanega strojnega učenja.