



Воронешки државен универзитет
Русија

Универзитет „Гоце Делчев“ – Штип,
С. Македонија

Воронежский государственный университет
Россия

Универзитет имени Гоце Делчева, г. Штип,
С. Македония

Voronezh State University
Russia

Goce Delcev University in Stip,
N. Macedonia

I

Шестая международная научная
конференция

Шеста меѓународна научна
конференција

Sixth International Scientific Conference

ФИЛКО FILKO

ФИЛОЛОГИЈА, КУЛТУРА И ОБРАЗОВАНИЕ

ФИЛОЛОГИЈА, КУЛТУРА И ОБРАЗОВАНИЕ

PHILOLOGY, CULTURE AND EDUCATION

ЗБОРНИК НА ТРУДОВИ
СБОРНИК СТАТЕЙ
CONFERENCE PROCEEDINGS

18-19 октјабр 2021 / 18-19 октомври 2021 / 18-19 October 2021



Воронешки државен универзитет
Русија
Воронежский государственный университет
Россия
Voronezh State University
Russia

Универзитет „Гоце Делчев“ – Штип,
С. Македонија
Универзитет имени Гоце Делчева, г. Штип,
С. Македонија
Goce Delcev University in Stip,
N. Macedonia

Шестая международная научная
конференция

Шеста меѓународна научна
конференција

Sixth International Scientific Conference

ФИЛКО FILKO

ФИЛОЛОГИЈА, КУЛТУРА И ОБРАЗОВАНИЕ

ФИЛОЛОГИЈА, КУЛТУРА И ОБРАЗОВАНИЕ

PHILOLOGY, CULTURE AND EDUCATION

ЗБОРНИК НА ТРУДОВИ СБОРНИК СТАТЕЙ CONFERENCE PROCEEDINGS

18-19 октjабрjа 2021 / 18-19 октомври 2021 / 18-19 October 2021

ЗБОРНИК НА ТРУДОВИ
СБОРНИК СТАТЕЙ
CONFERENCE PROCEEDINGS

ФИЛКО
FILKO

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

80(062)
821.163.3.09(062)

ЗБОРНИК на трудови [Електронски извор] / Шестата меѓународна научна конференција ФИЛКО филологија, култура, и образование, 18-19 октомври 2021 = Сборник статей / Шестая международная научная конференция ФИЛКО филология, культура и образование, 18-19 октября 2021 = Conference proceedings / Sixth International scientific conference FILKO philology, culture and education, 18-19 October 2021 ; [главен и одговорен уредник Драгана Кузмановска]. - Штип : Универзитет "Гоце Делчев", 2022

Начин на пристапување (URL):
<https://js.ugd.edu.mk/index.php/fe/issue/view/289>. - Текст во ПДФ формат, содржи 295 стр., илустр. - Наслов превземен од екранот. - Трудови на мак., рус. и англ. јазик. - Фусноти кон текстот. - Библиографија кон трудовите

ISBN 978-608-244-871-8

а) Филологија -- Собири б) Македонска книжевност -- Компаративни студии -- Собири в) Англиска книжевност -- Компаративни студии -- Собири г) Руска книжевност -- Компаративни студии -- Собири д) Високо образование -- Онлајн-настава -- Ковид -- Собири

COBISS.MK-ID 56582405

ОРГАНИЗАЦИСКИ КОМИТЕТ

Жана Грачева, Филолошки факултет при ВГУ
Драгана Кузмановска, Филолошки факултет при УГД
Татјана А. Тернова, Филолошки факултет при ВГУ
Весна Коцева, Филолошки факултет при УГД
Генадиј Ф. Коваљов, Филолошки факултет при ВГУ
Ана Витанова Рингачева, Филолошки факултет при УГД
Лариса В. Рибачева, Филолошки факултет при ВГУ
Костадин Голаков, Филолошки факултет при УГД

МЕЃУНАРОДЕН ПРОГРАМСКИ КОМИТЕТ

Драгана Кузмановска (С. Македонија)
Даниела Коцева (С. Македонија)
Светлана Јакимовска (С. Македонија)
Ева Ѓорѓиевска (С. Македонија)
Ана Витанова Рингачева (С. Македонија)
Силвана Симоска (С. Македонија)
Татјана Стојановска Иванова (С. Македонија)
Лариса В. Рибачева (Русија)
Софија Заболотнаја (Русија)
Татјана А. Тернова (Русија)
Татјана Атанасоска (Австрија)
Олег Н. Фенчук (Белорусија)
Јулиа Дончева (Бугарија)
Мадлен Данова (Бугарија)
Билјана Мариќ (Босна и Херцеговина)
Душко Певуља (Босна и Херцеговина)
Волф Ошлис (Германија)
Волфганг Моч (Германија)
Габриела Б. Клајн (Италија)
Михал Ванке (Полска)
Мајкл Рокланд (САД)
Даниела Костадиновиќ (Србија)
Милутин Ѓуричковиќ (Србија)
Селена Станковиќ (Србија)
Тамара Валчик-Булик (Србија)
Ахмед Ѓуншен (Турција)
Нецати Демир (Турција)
Шерифе Сехер Ерол Чал'шкан (Турција)
Карин Руке-Брутен (Франција)
Танван Тонтат (Франција)

Технички секретар

Ирина Аржанова
Наташа Сарафова

Главен и одговорен уредник

Драгана Кузмановска

Јазично уредување

Даница Атанасовска-Гаврилова (македонски јазик)
Марјана Розенфелд (руски јазик)
Биљана Иванова (англиски јазик)
Снежана Кирова (англиски јазик)
Татјана Уланска (англиски јазик)

Техничко уредување

Кире Зафиров

Адреса на организацискиот комитет:

Воронешки државен универзитет
Филолошки факултет
г. Воронеж, пл. Ленина, 10, корпус 2, к. 34, Русија

Универзитет „Гоце Делчев“ – Штип

Филолошки факултет
ул. „Крсте Мисирков“ бр. 10-А
Пош. фах 201, Штип - 2000, Р. Македонија

Е-пошта: filko.conference@gmail.com

Веб-страница: <http://js.ugd.edu.mk./index.php/fe>

Година на издавање и печатење - 2022

Место на издавање - Штип

РЕДАКЦИОННЫЙ СОВЕТ

Жанна Грачева, Филологический факультет при ВГУ
Драгана Кузмановска, Филологический факультет при УГД
Татьяна А. Тернова, Филологический факультет при ВГУ
Весна Коцева, Филологический факультет при УГД
Геннадий Ф. Ковалев, Филологический факультет при ВГУ
Ана Витанова-Рингачева, Филологический факультет при УГД
Лариса В. Рыбачева, Филологический факультет при ВГУ
Костадин Голаков, Филологический факультет при УГД

МЕЖДУНАРОДНАЯ РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Драгана Кузмановска (С. Македония)
Даниела Коцева (С. Македония)
Светлана Якимовска (С. Македония)
Ева Гёргиевска (С. Македония)
Ана Витанова-Рингачева (С. Македония)
Силвана Симоска (С. Македония)
Татьяна Стояновска-Иванова (С. Македония)
Лариса В. Рыбачева (Россия)
Софья Заболотная (Россия)
Татьяна А. Тернова (Россия)
Татяна Атанасоска (Австрия)
Олег Н. Фенчук (Беларусь)
Мадлен Данова (Болгария)
Юлиа Дончева (Болгария)
Биляна Марич (Босния и Херцеговина)
Душко Певуля (Босния и Херцеговина)
Вольф Ошлис (Германия)
Волфганг Моч (Германия)
Габриелла Б. Клейн (Италия)
Михал Ванке (Польша)
Майкл Рокланд (США)
Даниела Костадинович (Сербия)
Милутин Джуричкович (Сербия)
Селена Станкович (Сербия)
Тамара Валчич-Булич (Сербия)
Ахмед Гюншен (Турция)
Неджати Демир (Турция)
Шерифе Сехер Эрол Чал'шкан (Турция)
Карин Рукэ-Брутэн (Франция)
Танван Тонтат (Франция)

Ученый секретарь

Ирина Аржанова
Наташа Сарафова

Главный редактор

Драгана Кузмановска

Языковая редакция

Даница Атанасовска-Гаврилова (македонский язык)
Марьяна Розенфельд (русский язык)
Бильяна Иванова (английский язык)
Снежана Кирова (английский язык)
Татьяна Уланска (английский язык)

Техническое редактирование

Кире Зафиров

Адрес организационного комитета

Воронежский государственный университет
Филологический факультет
г. Воронеж, пл. Ленина, 10, корпус 2, к. 34, Россия

Университет им. Гоце Делчева – Штип

Филологический факультет
ул. „Крсте Мисирков“ д. 10-А
Пош. фах 201, Штип - 2000, Р. С. Македония

Э-почта: filko.conference@gmail.com

Веб-сайт: <http://js.ugd.edu.mk/index.php/fe>

Год издания и печати - 2022

Место публикации - Штип

EDITORIAL STAFF

Zhanna Gracheva, Faculty of Philology, VGU
Dragana Kuzmanovska, Faculty of Philology, UGD
Tatyana A. Ternova, Faculty of Philology, VGU
Vesna Koceva, Faculty of Philology, UGD
Genadiy F. Kovalyov, Faculty of Philology, VGU
Ana Vitanova Ringacheva, Faculty of Philology, UGD
Larisa V. Rybatcheva, Faculty of Philology, VGU
Kostadin Golakov, Faculty of Philology, UGD

INTERNATIONAL EDITORIAL BOARD

Dragana Kuzmanovska (N. Macedonia)
Daniela Koceva (N. Macedonia)
Svetlana Jakimovska (N. Macedonia)
Eva Gjorgjievska (Macedonia)
Ana Vitanova Ringacheva (N. Macedonia)
Silvana Simoska (N. Macedonia)
Tatjana Stojanovska-Ivanova (N. Macedonia)
Larisa V. Rybatcheva (Russia)
Sofya Zabolotnaya (Russia)
Tatyana A. Ternova (Russia)
Tatjana Atanasoska (Austria)
Oleg N. Fenchuk (Belarus)
Biljana Maric (Bosnia and Herzegovina)
Dushko Pevulja (Bosnia and Herzegovina)
Madlen Doncheva (Bulgaria)
Yulia Doncheva (Bulgaria)
Karine Rouquet-Brutin (France)
That Thanh-Vân Ton (France)
Wolf Ochlies (Germany)
Wolfgang Motch (Germany)
Gabriella B. Klein (Italy)
Michal Wanke (Poland)
Danijela Kostadinovic (Serbia)
Milutin Djurichkovic (Serbia)
Selena Stankovic (Serbia)
Tamara Valchic-Bulic (Serbia)
Ahmed Gunshen (Turkey)
Necati Demir (Turkey)
Şerife Seher Erol Çalışkan (Turkey)
Michael Rockland (USA)

Conference secretary

Irina Arzhanova
Natasha Sarafova

Editor in Chief

Dragana Kuzmanovska

Language editor

Danica Atanasovska-Gavrilova (Macedonian)
Maryana Rozenfeld (Russian)
Biljana Ivanova (English)
Snezana Kirova (English)
Tatjana Ulanska (English)

Technical editing

Kire Zafirov

Address of the Organizational Committee

Voronezh State University

Faculty of Philology

10 pl. Lenina, Voronezh, 394006, Russia

Goce Delcev University - Stip

Faulty of Philology

Krste Misirkov St. 10-A

PO Box 201, Stip - 2000, Republic of Macedonia

E-mail: filko.conference@gmail.com

Web-site: <http://js.ugd.edu.mk./index.php/fe>

Year of publication - 2022

Place of publication - Stip

СОДРЖИНА / СОДЕРЖАНИЕ / CONTENT

1. Adili Bujar, Petrovska Sonja and Xhambazi Gzim - GENDER AND LEVEL OF TEACHERS' KNOWLEDGE ON KEY FEATURES OF INTERCULTURAL EDUCATION.....	13
2. Ćeklić Nina S. - CONCEPTUAL POSITION AS A COMPOSITION PRINCIPLE IN REMNANTS OF THE WORLD BY IGOR MAROJEVIĆ.....	23
3. Gjorgjeva Ivana, Jankova Alagjovska Natka - ANGELA CARTER'S BLOODY CHAMBER IN RELATION TO THE SECOND WAVE FEMINISM	31
4. Пиев Крсте, Поп Зариева Наталија, Крстева Марија, Донеv Драган - THE NOTION OF BANISHMENT IN SHAKESPEARE'S PLAYS	39
5. Ivanova Biljana, Kuzmanovska Dragana, Kirova Snezana - FOREIGN LANGUAGE ACQUISITION IN SCHOOLS.....	45
6. Јованова-Митковска Снежана- КВАЛИТАТИВНАТА МЕТОДОЛОГИЈА И НЕЈЗИНАТА ПРИМЕНА ВО ОПШТЕСТВЕНИТЕ И ХУМАНИСТИЧКИТЕ НАУКИ	49
7. Крстев Марија, Донеv Драган, Поп Зариева Наталија, Илиев Крсте HEMINGWAY IN HEMINGWAY'S GIRL.....	59
8. Kujundžić Maja, Ćeklić Nina - SOCILINGUISTIC USE OF THE PASSIVE IN TWO 19TH CENTURY BRITISH NOVELS.....	65
9. Vangelov Nikola - INTERCULTURAL CHARACTERISTICS OF DIGITAL MARKETING	75
10. Антониевски Горан, Асани Агим - КАРАКТЕРИСТИКИ НА ГАСТРОНОМИЈАТА ВО ПОЛОШКИОТ РЕГИОН.....	83
11. Аржанова Ирина- СУЩНОСТЪ ЧЕЛОВЕКА И ИДЕАЛЫ ЖИЗНИ У СЛАВЯН.....	91
12. Арязмова Ольга Витальевна - ЛИТЕРАТУРНЫЙ ТРАВЕЛОГ В ЛИНГВОКУЛЬТУРНОМ ИЗМЕРЕНИИ (на материале «Африканской книги» А. М. Стесина).....	99
13. Bangoji Salija -TEACHER COMPETENCIES FOR THE KNOWLEDGE SOCIETY	105
14. Василевская Анастасия - РАБОТА С ХУДОЖЕСТВЕННОЙ ДЕТАЛЬЮ НА УРОКАХ ЛИТЕРАТУРЫ В СРЕДНЕЙ ШКОЛЕ.....	113
15. Витанова-Рингачева Ана, Мицева Сузана ЈУРУЦИТЕ – МАКЕДОНСКИТЕ АМИШИ.....	117
16. Гальцова Дарья - СИНОНИМИЧЕСКИЕ СООТВЕТСТВИЯ В НАИМЕНОВАНИЯХ ОБРАГОВ (НА МАТЕРИАЛЕ ВОРОНЕЖСКИХ ГОВОРОВ)	123
17. Гладышева Светлана - КНИГОИЗДАТЕЛЬСТВО «СЛОВО» В СОЦИОКУЛЬТУРНОМ ПРОСТРАНСТВЕ РУССКОГО ЗАРУБЕЖЬЯ.....	129
18. Голаков Костадин, Методијески Дејан - КАРАКТЕРИСТИКИ НА РУСКАТА ГАСТРОНОМИЈА.....	135
19. Грачева Жанна, Ганделидзе Диана - ЯЗЫКОВАЯ МОДА В РОССИИ НАЧАЛА XXI ВЕКА.....	143

20. Гркова-Беадер Марија - ПРАВОПИСОТ ВО НАСТАВАТА ПО МАКЕДОНСКИ ЈАЗИК ВО ОСНОВНОТО ОБРАЗОВАНИЕ	155
21. Денкова Јованка - БЕГСТВО ВО/ОД ВИРТУЕЛНАТА РЕАЛНОСТ НА СОНИШТАТА	165
22. Ђуричковић Милутин - РУСКА ДРАМА ЗА ДЕЦУ И МЛАДЕ НА СРПСКОМ ЈЕЗИКУ	175
23. Донев Драган, Поп Зариева Наталија, Костова Кристина, Илиев Крсте ИГИТАЛНИТЕ НОМАДИ КАКО КУЛТУРОЛОШКИ ФЕНОМЕН ВО СОВРЕМЕНОТО ОПШТЕСТВО.....	185
24. Иљина Татьяна, Моу Шужань - ОСОБЕННОСТИ ДИСТАНЦИОННОГО ОБУЧЕНИЯ В РОССИИ И КИТАЕ.....	191
25. Иванов Стефан, Мартиновска Банде Цвета, Ивановска Билјана ОБРАБОТКА НА ПРИРОДЕН ЈАЗИК СО ТЕХНИКИ НА МАШИНСКО УЧЕЊЕ	197
26. Ивановска Билјана, Џафери Гзим - ГОВОРНИТЕ ЧИНОВИ И НИВНАТА ПОВРЗАНОСТ СО УЧТИВОСТА ВО ВРЕМЕ НА КОРОНА КРИЗАТА	203
27. Карначук Ирина - КЛАССИЧЕСКАЯ ЛИТЕРАТУРА КАК СПОСОБ СОХРАНЕНИЯ НРАВСТВЕННОСТИ НА УРОКАХ ЛИТЕРАТУРЫ В СРЕДНЕЙ ШКОЛЕ	209
28. Ковилоски Славчо - ЗА ПОСТОЕЊЕТО НА МАКЕДОНСКАТА КНИЖЕВНА КРИТИКА ВО XIX ВЕК.....	213
29. Косяков Сергей - ИСТОРИЯ ПОПЕЧИТЕЛЬСТВА О СЛЕПЫХ И БИБЛИОТЕЧНОГО ОБСЛУЖИВАНИЯ НЕЗРЯЧИХ В ВОРОНЕЖСКОЙ ОБЛАСТИ	221
30. Котев Иван - ПРАВДАТА И НЕПРАВДАТА ВО ФOLKЛОРОТ (Во македонските народни песни, приказни и кратките жанри)	227
31. Мелкадзе Нанули - ЗАСТОЛЪЕ КАК ФРАГМЕНТ ГРУЗИНСКОЙ ЯЗЫКОВОЙ КАРТИНЫ МИРА	237
32. Младеновска-Ристовска Катерина - КУЛТОТ НА ВОДАТА ВО РЕЛИГИЈАТА НА АНТИЧКИТЕ МАКЕДОНЦИ.....	245
33. Нагина Ксения - ОБ УЧЕБНЫХ ПОСОБИЯХ В ВЫСШЕЙ ШКОЛЕ: ПРОБЛЕМЫ ИЗУЧЕНИЯ ТВОРЧЕСТВА Л. ТОЛСТОГО	253
34. Негриевска Надица - СИНТАКСИЧКИ ФУНКЦИИ НА ПРОСТИТЕ ПРЕДЛОЗИ ВО ИТАЛИЈАНСКИОТ ЈАЗИК И НИВНИТЕ МОЖНИ ЕКВИВАЛЕНТИ ВО МАКЕДОНСКИОТ ЈАЗИК.....	261
35. Пигунова Ксения - СПОСОБЫ ОРГАНИЗАЦИИ ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ КАРТИНЫ МИРА В РОМАНЕ ЕВГЕНИЯ ВОДОЛАЗКИНА «ЛАВР».....	271
36. Попов Сергей - ДИАЛЕКТНАЯ ГЕОГРАФИЧЕСКАЯ ЛЕКСИКА В РЕГИОНАЛЬНОЙ ТОПОНИМИИ РОССИИ	279
37. Рогич Петја -БУГАРСКИОТ, МАКЕДОНСКИОТ И СРПСКИОТ ЈАЗИК – СЛИЧНОСТИ И РАЗЛИКИ.....	283
38. Сарафова Наташа - РАДИКАЛНООСЛОБОДИТЕЛНИОТ ФЕМИНИЗАМ ВО ДЕЛАТА НА АНГЕЛА КАРТЕР	291
39. Саздова Викторија, Ивановска Билјана - ЈАЗИЧНИТЕ ЕДИНИЦИ СО ПОЗИТИВЕН ЕМОЦИОНАЛЕН ПРИЗНАК ВО ЈАЗИЧНИОТ ПАР ГЕРМАНСКИ-МАКЕДОНСКИ	299

40. Стојановска-Стефанова Анета, Рунчева-Тасев Христина, Магдинчева Шопова Марија - КОМУНИКАЦИСКИ ПРЕДИЗВИЦИ ВО МУЛТИКУЛТУРНИТЕ ОПШТЕСТВА.....	307
41. Тасевска Марица - ОНЛАЈН НАСТАВАТА ПО ГЕРМАНСКИ ЈАЗИК ВО ВРЕМЕ НА ПАНДЕМИЈА. ИСКУСТВА И ПРЕДИЗВИЦИ.....	315
42. Тернова Татјана, Фролова Анна - РЕПРЕЗЕНТАЦИЈА ПАМЈАТИ В ПОЕЗИИ ОКСАНЫ ВАСЈАКИНОЙ	321
43. Ткаченко Дмитрий - ИНКЛУЗИВНОЕ ОБРАЗОВАНИЕ В РОССИИ В КОНТЕКСТЕ ЕВРОПЕЙСКОГО ОПЫТА	327
44. Ткаченко Татјана - ОПЫТ ОБРАЩЕНИЯ К СЛАВЯНСКИМ ЯЗЫКАМ НА ЗАНЯТИЯХ ПО РУССКОМУ ЯЗЫКУ В СПОРТИВНОМ ВУЗЕ	333
45. Тодорова Марија, Уланска Татјана - ПЕРЦЕПЦИЈАТА НА СТУДЕНТИТЕ ЗА ОНЛАЈН НАСТАВАТА ПО ШПАНСКИ И АНГЛИСКИ ЈАЗИК ВО УСЛОВИ НА ПАНДЕМИЈА	339
46. Урумова-Марковска Славица - ТЕХНОЛОГИЈАТА ВО ОБРАЗОВАНИЕТО КАКО ПРЕДИЗВИК ЗА ЗДРАВ ДЕТСКИ РАЗВОЈ	347
47. Цацков Оливер - МОРАВСКАТА МИСИЈА НА КИРИЛ И МЕТОДИЈА ЗНАЧАЕН ФАКТОР ЗА СЛОВЕНСКАТА ПИСМЕНОСТ.....	355
48. Шутаров Васко - СПЕЦИФИКИ НА МАКЕДОНСКАТА КУЛТУРНА ДИПЛОМАТИЈА	363

ОБРАБОТКА НА ПРИРОДЕН ЈАЗИК СО ТЕХНИКИ НА МАШИНСКО УЧЕЊЕ

¹Стефан Иванов, ²Цвета Мартиновска Банде, ³Билјана Ивановска

¹Факултет за информатика, ²Филолошки факултет

Универзитет „Гоце Делчев“, Штип

E-mail: stefan.102187@student.ugd.edu.mk

E-mail: cveta.martinovska@ugd.edu.mk

E-mail: biljana.ivanovska@ugd.edu.mk

Апстракт

Во невропсихологијата, лингвистиката и филозофијата на јазикот, природен јазик е секој јазик што еволуирал природно кај луѓето преку употреба и повторување без свесно планирање или предумисла. Природните јазици можат да имаат различни форми, како што се говор или сигнализирање. Тие се разликуваат од конструираниите и формалните јазици, како што се оние што се користат за програмирање компјутери или за проучување на логиката.

Текстуалните податоци се најчести и опфаќаат повеќе од 50% од неструктурираните податоци. Неколку примери вклучуваат – твитови / објави на социјалните медиуми, разговори на чет, вести, блогови и написи, прегледи на производи или услуги, како и евиденција на пациентите во здравствениот сектор. Неколку понови апликации се ботови управувани од говор како Siri, Alexa, итн. Со цел да се произведат значајни и ефективни сознанија од текстуалните податоци, односно за да се искористи потенцијалот на текстуалните податоци, се користи обработка на природен јазик заснована на машинско учење и пред сè на длабоко учење.

Клучни зборови: машинско учење, обработка на текст, сумирање на текст, категоризација на текст

1. Вовед

Обработка на природниот јазик (NLP - natural language processing) е многу популарна област за истражување на машинското учење што, исто така, може да се користи пошироко во апликации кои се однесуваат на секојдневниот живот на луѓето од реалниот свет. Истражувачката податочна анализа е поврзана со подобластите на вештачката интелигенција кои се наречени *откривање на знаење* и *машинско учење*. (Рамакришнан Р. & Герке Ј., 2010: 890)

Во областа на обработката на македонскиот јазик, обработката на природниот јазик (NLP) во минатото имала некои значајни напори и потешкотии во истражувањето на текст-во-говор и говор-во-текст. Во базата на податоци “common voice” на Мозила се јавуваат примери од типот *Македонка*, *мајка* и сл, како македонски дел од говорот претставени преку платформа за собирање на податоци за текст-во-говор и говор-во-текст. Како и овие, постојат и македонски податоци во рамките на базата на податоци Татоеба и корпусот Оскар.

Моделите што се лесно достапни и едноставни за употреба на англиски или на германски јазик не постојеа или не беа достапни за анализа на истражувањата на македонски јазик. Секој систем што беше покомплексен подразбира кодирање од почеток на модели и алатки за сите на кои им се потребни, без разлика дали се студенти или професионалци.

2. Обработка на текст - токенизација и лематизација

Токенизацијата се однесува на разделување на текстот на минимални значајни единици. Постои токенизатор на реченици и токенизатор на зборови. Во примерот подолу се прикажува токенизатор на зборови, што е задолжителен чекор во претпроцесирањето на текстот за секаков вид анализа. Во следниот пример прикажана е токенизација на реченица напишана на македонски јазик. За да работи овој код потребно е да се инсталира библиотеката **spaCy**. Оваа библиотека нема функции за стеминг, туку само за лематизација. Во следната програма се врши токенизација на текстот.

```
sentence3 = nlp („Д-р Мите и неговата жена живеат во Америка.“)
```

```
In [9]: sentence3 = nlp("Д.р Мите и неговата жена живеат во Америка.")  
print(sentence3)
```

```
Д.р Мите и неговата жена живеат во Америка.
```

```
In [10]: for word in sentence3:  
print(word.text)
```

```
Д.р  
Мите  
и  
неговата  
жена  
живеат  
во  
Америка  
.
```

Лематизацијата во компјутерската лингвистика, е алгоритамски процес за одредување на лемата на зборот врз основа на наменетото значење во целиот текст. Лематизацијата често се користи кај јазиците кои се со богата морфологија во својата граматичка структура за да се именуваат проблеми кои се предизвикани од морфолошката сложеност на јазичниот систем, изведени преку леми кои се засноваат на граматиката. (Hämäläinen M.; Partanen N.; Rueter J. et.al., 2021) За разлика од стемингот, лематизацијата зависи од правилно идентификување на наменетиот дел од говорот и значењето на зборот во реченицата, како и во поголемиот контекст што ја опкружува таа реченица, како што се соседните реченици или дури и цел документ. „**SpaCy**“ библиотеката во која е имплементиран македонскиот јазик не поддржува стеминг, туку само лематизација и овозможува создавање на вештачка интелигенција која ќе биде функционална во македонскиот јазик (исто како и во англискиот јазик) и ќе овозможи создавање на модел кој ќе може лесно да го препознае македонскиот

јазик. Во следниот пример е прикажана лематизација на текст напишан на македонски јазик.

Се користи текстот `doc = nlp(«Јас читам книга.»)`

```
In [8]: lemmatizer = nlp.get_pipe("lemmatizer")
```

```
In [11]: doc = nlp("Јас читам книга.")
```

```
In [12]: print([token.lemma_ for token in doc])
          ['Јас', 'чита', 'книга', '.']
```

За извлекување на именки, придавки, глаголи и други зборовни групи на македонски јазик потребно е да се користи библиотеката **SpaCy** во која е внесен македонскиот јазик. Македонскиот SpaCy е објаснет во статијата дадена на следниов линк подолу, каде се прикажани и импликациите кои овозможуваат обработка на текстот (линк: <https://blog.netcetera.com/macedonian-spacy-f3c85484777f>).

Може да се ексцерпираат именки, придавки, заменки и други зборовни групи од текстот со едноставен код.

Пример: `print("Verb:", [token.lemma_ for token in doc if token.pos_ == "VERB"])`

Овде се проверува дали зборовите одговараат на тагот VERB и се печатат сите глаголи во реченицата. Само со замена на тагот, програмата може да издвои и други видови фрази и/или изрази и/или зборовни групи (отворени и затворени зборовни групи).

Во следниот пример прикажани се линии код кои влечат повеќе видови на фрази од реченицата. На секој дел од текстот може да му се додели специфичен таг (verb, modal verb, auxiliary verb, proper noun, noun, coordinating conjunction, subordinating conjunction и др.) врз база на веќе научен модел на македонски јазик.

```
In [28]: input_string = "Бојан е тој што сака да се врати во Македонија. Тој ќе се врати за 3 или 4 дена."
          doc = nlp(input_string)
```

```
In [29]: print("Verb:", [token.lemma_ for token in doc if token.pos_ == "VERB"])
          print("Proper noun:", [token.lemma_ for token in doc if token.pos_ == "PROPN"])
          print("Auxiliary verb:", [token.lemma_ for token in doc if token.pos_ == "AUX"])
          print("Pronoun:", [token.lemma_ for token in doc if token.pos_ == "PRON"])
          print("Noun:", [token.lemma_ for token in doc if token.pos_ == "NOUN"])
          print("Coordinating conjunction:", [token.lemma_ for token in doc if token.pos_ == "CCONJ"])
          print("Numeral:", [token.lemma_ for token in doc if token.pos_ == "NUM"])
          print("Subordinating conjunction:", [token.lemma_ for token in doc if token.pos_ == "SCONJ"])
```

Може да се извлекат именки, придавки и други зборови со едноставен код.

Пример: `print("Verb:", [token.lemma_ for token in doc if token.pos_ == "VERB"])`

Овде се проверува дали зборовите одговараат на тагот VERB (глагол) и се печатат сите глаголи во реченицата. Само со замена на тагот, програмата може да влече и други видови фрази или зборовни групи.

Во следниов пример прикажани се линии код кои влечат повеќе видови на фрази од реченицата (различни зборовни групи поткрепени со конкретни примери од анализираната реченица).

Резултатот од програмата е прикажан на следнава слика.

```
print("Subordinating conjunction:", [token.lemma_ for token
Verb: ['врати', 'врати']
Proper noun: ['Бојан', 'Македонија']
Auxiliary verb: ['е', 'се', 'се']
Pronoun: ['тој', 'тој']
Noun: ['ден']
Coordinating conjunction: []
Numeral: ['3', '4']
Subordinating conjunction: ['што', 'да']
```

Друг начин на извлекување на фрази е со следниот код:

`for entity in doc.ents: print(entity.text, entity.label_)`

Се користи следниот текст: «Бојан е тој што сака да се врати во Македонија. Тој ќе се врати за три или четири дена. Тој ќе патува за Австралија и ќе го посети својот другар Мартин и заедно ќе учат на ФИНКИ во Скопје.»

Резултатот е прикажан на следнава слика:

```
In [63]: for entity in doc.ents:
print(entity.text, entity.label_)

Бојан PERSON
Македонија GPE
Австралија GPE
Мартин PERSON
ФИНКИ ORG
Скопје GPE
```

3. Заклучок

Обработката на природен јазик (NLP) им помага на компјутерите да комуницираат со луѓето на нивниот јазик и ги разменуваат другите задачи поврзани со јазикот. На пример, NLP им овозможува на компјутерите да читаат текст, да слушаат говор, да го толкуваат, да го мерат чувството и да одредат кои делови од текстот се важни за одредена анализа.

Денешните машини можат да анализираат повеќе податоци базирани на јазик отколку луѓето, без замор и на доследен, непристрасен начин. Со оглед

на неверојатната количина на неструктурирани податоци што се генерираат секој ден, од медицински досиеја до социјални медиуми, автоматизацијата ќе биде клучна за целосно и ефикасно да се анализираат податоците за текстот и говорот.

Човечкиот јазик е неверојатно сложен и разновиден. Се изразуваме на многу начини, и говорно и писмено. Не само што има стотици јазици и дијалекти, туку во секој јазик има единствен сет на граматички и синтаксички правила, термини и различен сленг. Кога се пишува, честопати се пишува погрешно, се користат скратени зборови, или се испушта интерпункцијата. При разговор, постојат многу регионални акценти и сленгови и многу пати се позајмуваат термини од други јазици.

Иако надгледуваното учење, а особено длабокото учење, сега широко се користат за моделирање на човечкиот јазик, исто така се јавува потреба од синтаксичко и семантичко разбирање и експертиза на текстот во доменот на компјутерската лингвистика - современи и модерни аспекти и тенденции што се нужни и неопходни во пристапите за машинското учење. NLP е важен затоа што помага да се реши двосмисленоста на јазикот и дава корисна нумеричка структура на податоците за многу последователни апликации, како што се препознавање говор или анализа на текст.

Користена литература

А) Интернет линкови

1. <https://universaldependencies.org/u/pos/all.html#al-u-pos>
2. https://en.wikipedia.org/wiki/Machine_learning
3. https://en.wikipedia.org/wiki/Natural_language_processing
4. https://en.wikipedia.org/wiki/Document_classification
5. <https://www.kaggle.com/subhassing/exploring-consumer-complaint-data/data>
6. https://www.sas.com/en_in/insights/analytics/what-is-natural-language-processing-nlp.html
7. <https://blog.netcetera.com/macedonian-spacy-f3c85484777f>
8. <https://monkeylearn.com/what-is-text-classification/>
9. [https://searchenterpriseai.techtarget.com/definition/natural-language-processing-NLP#:~:text=Natural%20language%20processing%20\(NLP\)%20is,in%20the%20field%20of%20linguistics.](https://searchenterpriseai.techtarget.com/definition/natural-language-processing-NLP#:~:text=Natural%20language%20processing%20(NLP)%20is,in%20the%20field%20of%20linguistics.)
10. <https://www.ibm.com/cloud/learn/machine-learning#:~:text=Machine%20learning%20is%20a%20branch,learn%2C%20gradually%20improving%20its%20accuracy.&text=Machine%20learning%20is%20an%20important,growing%20field%20of%20data%20science.>
11. <https://spacy.io/usage/linguistic-features>

Б) Книги / статии на кирилица

12. Рамакришнан Р. Герке Ј. (2010). Систем за управување со бази на податоци . 3 издание. Скопје. Абакус (превод од англиски Кацарова Ј. И Солаков К.).

В) Книги/статии (на латиница)

13. Kulkarni A.& Shivananda A. (2019). Natural Language Processing Recipes. Apress, Berkeley, CA.

14. Härmäläinen M., Partanen N., Rueter N. & Alnajjar K. (2021). Neural Morphology Dataset and Models for Multiple Languages, from theLarge to the Endangered. University of Helsinki. Proceedings of the 23 Nordic Conference on Computational Linguistics (NoDaLiDa) <https://aclanthology.org/2021.nodalida-main.pdf>