

**MINISTRY OF HIGHER EDUCATION, SCIENCE AND
INNOVATIONS OF THE REPUBLIC OF UZBEKISTAN**

**NATIONAL UNIVERSITY OF UZBEKISTAN NAMED AFTER
MIRZO ULUGBEK**



RAQAMLI TEXNOLOGIYALAR VA
SUN'IY INTELLEKTNI RIVOJLANTIRISH
ILMIY-TADQIQOT INSTITUTI



INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE

**PROCEEDINGS
of materials on topic:**

**«CONTEMPORARY TECHNOLOGIES OF
COMPUTATIONAL LINGUISTICS - CTCL.2024»**



Crossref
we use DOIs



Google
scholar

OPEN ACCESS



TASHKENT - 2024

**“KOMPYUTER LINGVISTIKASINING ZAMONAVIY TEXNOLOGIYALARI – CTCL.2024”
NOMLI XALQARO ILMIY-AMALIY KONFERENSIYASINING ILMIY
MAQOLALARI VA TEZISLAR TO‘PLAMI. T.: O‘zMU, 2024.
2024-yil 22-aprel**

TASHKILIY QO‘MITA TARKIBI:

N.Z.Abduraxmonova.	O‘zMU Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası mudiri, f.f.d., professor.
Z. T.Taxirov.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası professori, f.f.n.
A.R.Iskandarova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası dotsenti, PhD.
D.A.Fattaxova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi, PhD.
U.R.Xalmuxamedova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi.
N.Sh.Inogamova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi.
F.R.Topildiyeva.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi.
D.K.Xudoyberganova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi, PhD.
Z.M.Kasimova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi, PhD.
S.M.Po‘latova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi, PhD.
G.G‘.Xoliyorova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
O.O.Yulbarsov.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
F.M.Aliyev.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
M.S.Jo‘raqulova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
Z.G.Qodirova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
G.I.Kuvondikova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
M.F.Abdunazarova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
D.Xolbekova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
D.A.Mirvaliyeva.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
M.H.Eshmatova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
Q.R.Ibodullayeva.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
U.M.Askarova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
S.A.Abduraxmonova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
N.O‘rinboyeva.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası magistranti.
M.Xudoyberganova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası magistranti.

Mazkur xalqaro ilmiy-amaliy konferensiya to'plami kompyuter lingvistikasining ayrim yo'nalishlari, jumladan, korpus lingvistikasi, til ta'limida kompyuter texnologiyalari, NLPda til va nutq tahlili (lingvistik analizator: morfologik, sintaktik va semantik tahlil, nutq analizi va sintezi), mashina tarjimasining lingvistik va dasturiy ta'minoti, kompyuter leksikografiyasi: ontologiya, tezaurus, shuningdek, zamonaviy tilshunoslik masalalari kabi mavzular tahlilga tortilgan maqola va tezislardan tashkil topgan.

Ushbu ilmiy maqola va tezislar to'plami kompyuter lingvistikasi, kompyuter texnologiyalari, amaliy filologiya kabi yo'nalishlarda ilmiy izlanishlar olib borayotgan professor-o'qituvchilar, doktorantlar, mustaqil izlanuvchilar, magistr va talabalar uchun mo'ljallangan.

To'plamga kiritilgan maqolalar va tezislar mazmuni hamda sifatiga mualliflar mas'uldir.

MAS'UL MUHARRIR:

prof., f.f.d. N.Z.Abduraxmonova

TAHRIR HAY'ATI:

f.f.n., prof. Z. T.Taxirov

f.f.n., dots. A.R.Iskandarova

ped.f.n., dots. D.A.Fattaxova

PhD D.K.Xudoyberganova

PhD Z.M.Kasimova

Ushbu xalqaro ilmiy-amaliy anjuman to'plami Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti Ilmiy-texnik kengashi tomonidan (2024-yil 27-martdagi 3-sonli majlis bayonnomasi) nashrga tavsiya qilingan.

**СБОРНИК НАУЧНЫХ СТАТЕЙ И ТЕЗИСОВ
МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ
“СОВРЕМЕННЫЕ ТЕХНОЛОГИИ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ – CTCL.2024”.**
Т.: НУУз, 2024.
22 апреля 2024 г.

СОСТАВ ОРГАНИЗАЦИОННОГО КОМИТЕТА:

<i>Н.З. Абдурахмонова.</i>	Заведующий кафедрой компьютерной и прикладной лингвистики НУУз, д.ф.н., профессор.
<i>З. Т. Тахиров.</i>	Профессор кафедры компьютерной и прикладной лингвистики, к.ф.н.
<i>А.Р.Искандарова.</i>	Доцент кафедры компьютерной и прикладной лингвистики, к.ф.н.
<i>Д.А. Фаттахова.</i>	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.пед.н.
<i>У.Р. Халмухамедова.</i>	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
<i>Н. Ш. Иногамова.</i>	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
<i>Ф.Р. Топилдиева.</i>	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
<i>Д.К. Худойберганава.</i>	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
<i>З.М. Касимова.</i>	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
<i>С. Пулатова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
<i>Г.Г. Холёрова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>О.О. Юлбарсов.</i>	Преподаватель кафедры компьютерной и прикладной лингвистики.
<i>Ф.М. Алиев.</i>	Преподаватель кафедры компьютерной и прикладной лингвистики.
<i>М.С. Джуракулова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>З.Г. Кадирова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>Г. И. Кувондикова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>М.Ф. Абдуназарова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>Д. Холбекова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>Д.А. Мирвалиева.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>М.Х. Эшмаматова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>К.Р. Ибодуллаева.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>У.М. Аскарова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>С.А. Абдурахмонова.</i>	Преподавательница кафедры компьютерной и прикладной лингвистики.
<i>Н.З. Уринбоева.</i>	Магистрант кафедрой компьютерной и прикладной лингвистики.
<i>М. Худойберганава.</i>	Магистрант кафедрой компьютерной и прикладной лингвистики.

Материалы международной научно-практической конференции, включающие статьи и тезисы, посвящены некоторым направлениям компьютерной лингвистики, в том числе корпусной лингвистике, компьютерным технологиям в языковом образовании, анализу языка и речи в NLP (Лингвистический анализатор: морфологический, синтаксический и семантический анализ, анализ речи и синтез), Лингвистические базы данных и программное обеспечение машинного перевода, Компьютерная лексикография: онтология, тезаурус, а также проблемы современной лингвистики.

Сборник научных статей и диссертаций предназначен для профессоров и преподавателей, докторантов, независимых исследователей, магистров и студентов, ведущих научные исследования в таких областях, как компьютерная лингвистика, компьютерные технологии и прикладная лингвистика.

Авторы несут ответственность за содержание и качество статей и тезисов, включенных в книгу.

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

д.ф.н., проф. Н.Абдурахмонова

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

PhD, проф. З.Тахиров

PhD, доц. А.Искандарова

PhD З.Касимова

PhD Д.Фаттахова

PhD Д.Худойбергана

Сборник международной научно-практической конференции рекомендован к публикации Научно-технической комиссией Национального университета Узбекистана имени Мирзо Улугбека (протокол № 3 от 27 марта 2024 года).

PROCEEDINGS OF ARTICLES AND THESIS
OF THE INTERNATIONAL SCIENTIFIC -PRACTICAL CONFERENCE TITLED
«CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS-CTCL.2024»
T.: NUUZ, 2024.
April 22, 2024.

ORGANIZING COMMITTEE:

<i>N. Z. Abdurakhmonova.</i>	Head of Department of Computational and Applied Linguistics of UzMU, DSc, Professor.
<i>Z. T. Takhirov.</i>	Professor of the Department of Computational and Applied Linguistics, PhD.
<i>A.R. Iskandarov.</i>	Associate Professor of the Department of Computational and Applied Linguistics, PhD.
<i>D.A. Fattakhova.</i>	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
<i>U.R. Khalmukhamedova.</i>	Senior teacher of the Department of Computational and Applied Linguistics.
<i>N. Sh. Inogamova.</i>	Senior teacher of the Department of Computational and Applied Linguistics.
<i>F.R. Topildieva.</i>	Senior teacher of the Department of Computational and Applied Linguistics.
<i>D.K. Khudoyberganova.</i>	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
<i>Z.M. Kasimova.</i>	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
<i>S.M. Polatova.</i>	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
<i>G.G. Kholiyorova</i>	Senior teacher of the Department of Computational and Applied Linguistics.
<i>O. O. Yulbarsov.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>F.M. Aliyev.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>M.S. Jurakulova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>Z.G. Kadirova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>G. I. Kuvondikova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>M.F. Abdunazarova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>D. Kholbekova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>D.A. Mirvaliyeva.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>M.H. Eshmatova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>Q.R. Ibodullayeva.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>U.M. Askarova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>S.A. Abdurakhmonova.</i>	Teacher of the Department of Computational and Applied Linguistics
<i>N.Z. Urinboeva.</i>	Master student of Computational and Applied Linguistics.
<i>M. Khudoyberganova.</i>	Master student of Computational and Applied Linguistics.

This proceedings of international scientific-practical conference including articles and theses deals with some directions of Computational Linguistics, including Corpus Linguistics, Computer Technologies in language education, Language and speech analysis in NLP (Linguistic analyzer: morphological, syntactic and semantic analysis, speech analysis and synthesis), Linguistic database and software of Machine Translation, Computer Lexicography: ontology, thesaurus, as well as Modern Linguistic Issues.

The proceedings of scientific articles and theses is intended for professors and teachers, doctoral students, independent researchers, masters and students who are conducting scientific research in such areas as Computational Linguistics, Computer Technologies, and Applied Linguistics.

The authors are responsible for the content and quality of the articles and theses included in the book.

RESPONSIBLE EDITOR:

DSc, prof. N.Abdurakhmonova

EDITORIAL BOARD:

PhD, prof. Z.Taxirov

PhD, dots. A.Iskandarova

PhD Z.Kasimova

PhD D.Fattaxova

PhD D.Xudoyberganova

This proceeding of international scientific - practical conference was recommended for publication by the Scientific and technical commission of the National University of Uzbekistan named after Mirzo Ulugbek (protocol № 3 dated March 27, 2024).

MUNDARIJA / TABLE OF CONTENTS

I SHO'BA KORPUSLINGVISTIKASINING DOLZARB MASALALARI

SECTION 1 CURRENT ISSUES OF CORPUS LINGUISTICS

Г. МАДИЕВА, К. ПИРМАНОВА, А. СЕРИКБАЕВА. ИННОВАЦИИ В ИЗУЧЕНИИ ПОЭЗИИ АБАЯ: РАЗРАБОТКА КАЗАХСКОГО ПОЭТИЧЕСКОГО ПОДКОРПУСА	15
E.G' XONNAZAROV. ZAMON MA'NOSINING TUR, YUR, YOT, O'TIR FE'LLARIDA VOQELANISHINI KORPUSDA TEGGLASH MASALASI	22
Д. ТОҚМЫРЗАЕВ, Е. ЖҮБАНОВА, А. ПИРМАНОВА. СОЗДАНИЕ ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКОЙ МОДЕЛИ КАЗАХСКОГО ТЕКСТА	27
S.R. BAXODIROV. MATN TOZALASH TIZIMI DASTURLARINING UMUMIY TAVSIFI. AFZALLIKLARI VA KAMCHILIKLARI	34
I.Z. BEKMURADOVA. SLENGLARNI AVTOMATIK TEGGLASHDA SLENG BIRLIKLARI LUG'ATINING AHAMIYATI.....	40
A. SULTANOVA. QORAQALPOQ TA'LIMY KORPUSINI YARATISHNING NAZARIY ASOSLARI	44
M. ESHMAMATOVA. TURKIY TILLARDA KORPUSLARNING KENG KO'LAMI VA ULARNING RIVOJLANISHI.....	48
D.B. AXMEDOVA. O'ZBEK TILI KORPUSLARIDA O'Z QATLAM LEKSIKASI ETIMOLOGIYASINI SEMANTIK KENGAYTMA SIFATIDA IFODALASH MUAMMOLARI	53
G.I. KUVONDIKOVA, N. URINBOEVA. CREATION OF THE EDUCATIONAL CORPUS OF STUDENT SPEECH IS AN EXAMPLE OF WORLD EXPERIENCE.....	56
K.CH. RAJABOVA. KORPUSGA ASOSLANUVCHI GRAMMATIK VA LEKSIKOGRAFIK TADQIQOTLAR	63
M. NORMAMATOVA, N. ABDURAXMONOVA. BOLALAR NUTQI KORPUSI SHAKLLANISH BOSQICHLARI.....	68
A. VOSILJONOV. JAHON KORPUS LINGVISTIKASIDA DIALEKTAL KORPUSLAR.....	72
M. SIROJIDDINOVA. KORPUSGA ASOSLANGAN MASHINA TARJIMASI UCHUN IKKI TILLI FRAZEOLGIK LUG'AT YARATISHNING LINGVISTIK AHAMIYATI	76
O.O. YULBARSOV, YU.S. ABDIRAHIMOVA. MATN TAHLILIDA TERMINOLOGIYA MASALASI	80
S. OTAMURODOVA. IDENTIFICATION OF SEMANTIC FEATURES OF PHRASAL VERBS IN A PARALLEL CORPUS.....	84
A. АЙТБЕККЫЗЫ, А. ШОРМАКОВА. ОБЗОР И РАЗРАБОТКА ПРОГРАММЫ ПО СБОРУ ЭЛЕКТРОННЫХ РЕСУРСОВ НА КАЗАХСКОМ ЯЗЫКЕ ПО ЗАКОНОДАТЕЛЬСТВУ РЕСПУБЛИКИ КАЗАХСТАН	89
R.T. O'RINBOYEVA, I.J. ASQAROV. O'ZBEK KORPUSSHUNOSLIGIDA DIALEKTAL TA'MINOT YARATISH MASALALARI.....	95
M. SULTONOVA. DEVELOPING PRE-SERVICE STUDENTS' CORPUS LITERACY SKILLS IN TEACHING ENGLISH	99

Г. ХИДИРОВА. ЛИНГВОМЕТОДИЧЕСКИЙ ПОТЕНЦИАЛ ПАРАЛЛЕЛЬНОГО КОРПУСА...	102
R.T. MUFTILLAYEVA. NUROTA TUMAN LINGVISTIK TAMINOTI (TUZILISHI VA TAVSI..	119
N. URINBOEVA, G.I. KUVONDIKOVA. CORPUS ANALYSIS OF SPEECH UNITS IN SOCIAL NETWORKS.....	123
P.V. ORIPOVA, G.I. TOIROVA. MINNATDORCHILIK – MULOQOT MANBAI SIFATIDA.....	127
E. NAZIROVA, N. ABDURAKHMONOVA, K. USMONOVA. CHALLENGES IN CORPUS LINGUISTICS AND MACHINE TRANSLATION	131
SH. GULYAMOVA, M.V. NURBOYEVA. FRAZEOLOGIK POLISEMIYA VA UNI KORPUS TIZIMIDA MODELLASHTIRISH MUAMMOLARI.....	135
D.A. FATTAXOVA, U.T. TURAYEVA. TURKCHA - O‘ZBEKCHA PARALLEL KORPUSI UCHUN IBORALARNING SEMANTIK TAHLILI	140
N.B. MURADOVA. MUALLIFLIK KORPUSINING AHAMIYATI.....	145
D.O. O‘ROQOVA. KORPUSDAN QIDIRUV TIZIMI SIFATIDA FOYDALANISH	149
K. VASLIDDINOVA. THE INSTRUMENTAL OBJECTIVE ROLE OF CORPUS IN RESEARCHES OF VARIOUS FIELDS	155
A. QARSHIYEV, S.A. KARIMOV, M.S. TURSUNOV. O‘ZBEKCHA SO‘ZLARNING GRAMMATIK TAHLIL QILISH ALGORITMLARI	160
M.F. XOMIDOVA. KORPUS LINGVISTIKASIDA INTERTEKSTUALLIKNING ROLI.....	165
R.T. O‘RINBOYEVA, D.G. ALLABERDIYEVA. CHILDES KORPUSLAR TIZIMIDA O‘ZBEK TILI UCHUN MA‘LUMOTLAR BAZASI YARATISH.....	169
M. URAZALIYEVA. KOMPYUTER LINGVISTIKASIDA AUDIOKORPUS MASALALARI	173
G. ERGASHEVA, N. INOGAMOVA. O‘ZBEK TILINING TA‘LIMIY KORPUSI UCHUN BOLALAR ADABIYOTINI YARATISHNING DIDAKTIK ASOSLARI YUZASIDAN AYRIM MULOHAZALAR.....	176

II SHO‘BA

TIL TA'LIMIDA KOMPYUTER TEXNOLOGIYALARI

SECTION 2

COMPUTER TECHNOLOGIES IN LANGUAGE EDUCATION

A.P. ГАТИАТУЛЛИН, Н. АБДУРАХМОНОВА. БАЗА ЗНАНИЙ ПОРТАЛА «ТЮРКСКАЯ МОРФЕМА» КАК ЛИНГВИСТИЧЕСКИЙ РЕСУРС ДЛЯ СОЗДАНИЯ ИНСТРУМЕНТОВ ОБУЧЕНИЯ ТЮРКСКИМ ЯЗЫКАМ.....	180
K. MAVLONOVA. O‘QUVCHILARDA MEDIA SAVODXONLIK KO‘NIKMASINI SHAKLLANTIRISH	188
A.A. JUMABOEV, A. SAITKHANOVA. COMPUTER TECHNOLOGIES IN LANGUAGE EDUCATION	195
D.A. FATTAXOVA, D. JUMANAZAROVA. O‘ZBEK TILI MORFOLOGIK LUG‘ATINING LINGVISTIK TA‘MINOTI	199
J.B. BERDIYEV. BERT MODELINING QO‘LLANILISHI VA VAZIFALARI	203
M. AXMEDOVA, Д. ФАТТАХОВА. РОЛЬ И ФУНКЦИЯ КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В ЯЗЫКОВОМ ОБРАЗОВАНИИ.....	210

K.T. ODILOVA. ПОВЫШЕНИЕ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ С ПОМОЩЬЮ МУЛЬТИМОДАЛЬНЫХ ТЕКСТОВ: НА ПРИМЕРЕ УЗБЕКИСТАНА	215
SH. ASADOVA, H. ABDURAHMONOVA. HORIJY TILLARNI O'RGANISHDA KOMPYUTER TEXNOLOGIYALARINING O'RNI.....	219
Д. ЖУПАРБЕК. ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ВЕБ-КВЕСТОВ ДЛЯ ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ НА РАЗЛИЧНЫХ УРОВНЯХ	224
M.G'. HASANOVA, N. ABDURAXMONOVA. TIL TA'LIMIDA KOMPYUTER LINGVISTIKASI IMKONIYATLARIDAN FOYDALANISH: MUAMMO VA YECHIMLAR.....	233
C. ИБРАГИМОВА, Г. МИРЗАЕВ. ЛИНГВИСТИЧЕСКИЙ ТРЕНАЖЕР ОБУЧАЮЩИЙ УЗБЕКСКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ НА БАЗЕ АНАЛИЗАТОРА РЕЧИ	237
A.N. AKHMEDOVA. THE IMPORTANCE OF CORPORA IN TRANSLATION TEACHING AN..	241

III SHO'BA

NLPDA TIL VA NUTQ TAHLILI

(LINGVISTIK ANALIZATOR: MORFOLOGIK, SINTAKTIK VA SEMANTIK TAHLIL, NUTQ ANALIZI VA SINTEZI)

SECTION 3

LANGUAGE AND SPEECH ANALYSIS IN NLP (MORPHOLOGICAL, SYNTACTIC AND SEMANTIC ANALYSIS; SPEECH ANALYSIS AND SYNTHESIS)

L. BOIZOU, M. HOUSOUBA. MAKING USEFUL TOOLS FOR LANGUAGE COMMUNITIES WITHOUT BIG CORPORA: A VERSATILE MORPHOLOGICAL ANALYSER AND SPELLCHECKER FOR SONGHAY BASED ON FREE SOFTWARE	244
E. NAZIROVA, K. USMONOVA. ADVANCED NLP AI TOOLS IMPORTANCE IN MACHINE TRANSLATION FROM UZBEK TO ENGLISH.....	255
V. KARYUKIN, N. ABDURAKHMONOVA. THE DEVELOPMENT OF AN APPLICATION FOR COLLECTING AND FORMING CORPORA OF THE KAZAKH LANGUAGE	260
Э.Ш. НАЗИРОВА, К.А. УСМАНОВА. ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ЯЗЫКОВОМ ПЕРЕВОДЕ	266
Z. ISMADIYOROV. SEMANTIK TEGLAHNING FE'L SO'Z TURKUMIGA OID GURUHLAR	270
M.T. UZOQOVA. SINTAKTIK SINONIMIYA VA UNI ANIQLASH XUSUSIDA	273
SH. OTAJONOVA, N. INOGAMOVA. NATURAL LANGUAGE PROCESSING (NLP) APPLICATIONS IN LANGUAGE TEACHING.....	278
M.S. JO'RAQULOVA, Z.G. QODIROVA. STRUKTUR TAHLIL METODLARI ASOSIDA YARATILGAN SINTAKTIK NAZARIYALAR.....	284
A. МИНДУБАЕВ. НЕЧЕТКАЯ ЛОГИКА НА БАЗЕ ЯЗЫКОВЫХ МОДЕЛЕЙ: ПОСТРОЕНИЕ ГИБРИДНОЙ СИСТЕМЫ КЛАССИФИКАЦИИ ДИАЛЕКТОВ И ОПРЕДЕЛЕНИЯ	293
M. SUYUNOVA. NATURAL LANGUAGE PROCESSING AND ITS MAIN APPROACHES	304
E. MALIKOV. BIR SO'Z TURKUMIDAGI OMONIM SO'ZLARNI ANIQLASH USULLARI	310
М.И. ИСМОИЛОВА. ЎЗБЕК ТИЛИДА ҚАРАШЛИЛИК/ТЕГИШЛИЛИК МАЪНОСИНИНГ СЎЗ СЕМАНТИКАСИ ТАЪСИРИДА ЯШИРИН ИФОДАЛАНИШИ УСУЛИ	317
SH. ABDISALOMOVA. LINGVISTIK CHEKLOVLAR – KOREFERENTLIKNI AVTOMATIK HAL ETISHNING ASOSI.....	324

A. OTEMISOV, SH. SHARBAEV. MORFOANALIZATOR BOSQICHI UCHUN QORAQALPOQ TILIDAGI FE'L SO'Z TURKUMINI FORMALLASHITISH MODULI	329
G.E. UMURZAKOVA. UNDOV SO'ZLAR PRESUPPOZITSIYA SIGNALIZATORI SIFATIDA ...	334
Z.I. FAYZULLAYEVA, N.N. KARIMOV. NLPDA MATNLARNI UMUMIYLASHTIRISH VA LEKSIK TAHLIL	338
B. ELOV, A. ABDULLAYEV, N. XUDAYBERGANOV. O'ZBEK TILI KORPUSI MATNLARI ASOSIDA TIL MODELLARINI YARATISH	344
B.B. БАРАХНИН, Д.Б. МЕНГЛИЕВ. АЛГОРИТМИЧЕСКИЙ ПОДХОД К АНАЛИЗУ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ И ПОЭТИЧНОСТИ УЗБЕКСКИХ ТЕКСТОВ	354
U. YAXSHIBOYEV. SIYOSIY MATNLARNI O'RGANISHDA NUTQIY YONDASHUV	359
U. TULIEV, N. ABDURAKHMONOVA. METHOD OF DEVELOPING ARTIFICIAL PATTERNS TO ANALYZING SEMANTIC SIMILARITY OF DOCUMENTS.....	362

IV SHO'BA

MASHINA TARJIMASINING LINGVISTIK VA DASTURIY TA'MINOTI

SECTION 4

LINGUISTIC DATABASE AND SOFTWARE OF MACHINE TRANSLATION

R. ALAYEV, G. MAXMUDJONOVA. COUNTVECTORIZER YORDAMIDA SO'ZLAR STATISTIKASINI ANIQLASH.....	366
N.S. SOTVOLDIYEVA. MASHINA TARJIMASIDA SINONIMLAR BILAN BOG'LIQ MUAMMOLAR	371
U. ABDURAXMONOV. INGLIZCHA-O'ZBEKCHA MASHINA TARJIMASI UCHUN FRAZEOLGIK BIRLIKLAR BAZASINI YARATISH TAVSIFI.....	376
M.S. SHARIPOV, O.O. SOBIROV, G.M. SHARIPOVA. MASHINALI TARJIMA TIZIMLARI UCHUN O'ZBEK TILI SO'ZLARINI LEMMALASH VA GAP STRUKTURALARINI TAHLIL QILISH.....	379
M. XUDOYBERGANOVA. MASHINA TARJIMASIDA SINTAKTIK QOLIPLARNING QIYOSIY TAHLILI (QORAQALPOQ TILI MISOLIDA).....	384
A.Z. OTEMISOV, A. ABDULLAEVA. THE ROLE OF LINGUISTICS IN MACHINE TRANSLATION SOFTWARE DEVELOPMENT	388
M.A. JO'RAQULOVA. PARALLEL KORPUS MASHINA TARJIMASINING O'RGANISH OBYEKTI SIFATIDA	392
R. YULDASHEVA. A LINGUISTIC STUDY OF COLLOCATION IN ENGLISH.....	396
B. KHOLBEKOVA. THE ESSENCE OF THE CHARMING POETRY OF ZAHIRUDDIN MUHAMMAD BABUR.....	400
G. SHAMSIYEVA. MASHINA TARJIMASI UCHUN GRAMMATIK SATHDA MODELLASHTIRISH BOSQICHLARI	404
N. SHAMIEVA. BREAKING DOWN LANGUAGE BARRIERS: THE POWER AND PROGRESS OF MACHINE TRANSLATION	411
H. ZOKIROVA. TERMINOLOGY AS A LEARNING OBJECT IN MACHINE TRANSLATION....	415
G.A. SHIKHNAZAROVA, D.R. GULYAMOVA, N.R. SAMANDAROVA. MATHEMATICAL MODELING OF MACHINE LEARNING ALGORITHMS	418

V SHO'BA

KOMPYUTER LEKSIKOGRAFIYA: ONTOLOGIYA, TEZAVURUS

SECTION 5

COMPUTER LEXICOGRAPHY: ONTOLOGY, THESAURUS

N. ABDURAXMONOVA, U. ASKAROVA, YU. YULBARSOV. KORPUSNING LINGVISTIK VA TA'LIMIY AHAMIYATI.....	421
G. PO'LATOVA. O'ZBEK TILI TEZAVURUSI UCHUN SIFAT SO'Z TURKUMI LINGVISTIK BAZASINI YARATISH.....	429
Z. UBAYDULLAYEVA. TEZAVURUS TUSHUNCHASI VA ILMIY YONDASHUV	435
M. SAPAROVA. ISSUES, THEORIES, AND APPROACHES IN THE PRESENT-DAY LEXICOGRAPHIC SEMANTIC STUDIES	438
Z.G. QODIROVA, M.S. JO'RAQULOVA. TERMINOLOGIK ONTOLOGIK BILIMLAR BAZASINI MODELLASHTIRISH BOSQICHLARI	443
L. BABAJANOVA. BILINGVAL TEZAVURUS YARATISHDA O 'ZBEK VA INGLIZ TILLARINING VALENTLIK MUNOSABATI.....	454
G'.G'. ABDUVAXOBV. O'ZBEK TILIDA ELEKTRON LUG'ATLAR YARATISH ZARURATI	459
N. ABDURAXMONOVA, G. XOLIYOROVA. O'ZBEK TILI TEZAVURUSINI YARATISHDA STANDART ME'YORLAR.....	464
G. HASANOVA. THE PROBLEMS OF REPRESENTING NEAR-SYNONYMS IN A COMPUTATIONAL LEXICON	470
S.O. OYBEKOVA. TEZAVURUSLAR VA ULARNI TAHLIL QILISH TAMOYILLARI	475

VI SHO'BA

ZAMONAVIY TILSHUNOSLIK MASALALARI

SECTION 6

ISSUES OF MODERN LINGUISTICS

N.I. ADIZOVA. SAYOHAT TURIZMIGA OID MATNLARDA MILLIY-MADANIY BIRLIKLARNING LEKSIK-PRAGMATIK XUSUSIYATLARI	479
D.B. JUMAYEVA, S.M. USMONOVA. BOSHLANG'ICH SINIF O'QUVCHILARINI ORFOEPIK ME'YORLARGA O'RGATISH VA UNING AMALIY AHAMIYATI	485
N.M. MIROBIDOVA. O'ZBEK AYOLLAR NUTQIDA KONNOTATSIYA VA PRAGMATIK T...488	
S.D. NURMONOVA. XALQONA QASAMLARDA FONATSION VOSITALAR	493
D.X. XOLBEKOVA, K. IBODULLAYEVA. TILSHUNOSLIKDA LINGVOSTATISTIK TAHLIL MUAMMOLARINING O'RGANILISHI	496
SAJID V. VALAPPIL, N. ABDURAKHMONOVA. ISSUES OF MODERN LINGUISTICS: - INHIBITING LINGUISTIC CREATIVITY AMONG YOUTH ON SOCIAL MEDIA BY USING EMOJI AND INTERNET SLANG.....	505
S.A. ABDURAXMONOVA. O'ZBEK BOLALAR NUTQINING KORPUS ASOSIDA PSIXOLINGVISTIK TADQIQI	510
I.E. UMIROV. SEMANTOGRAFIK TARJIMA YOZUVINING LEKSIK, GRAMMATIK VA STILISTIK XUSUSIYATLARI	514

Н. АБЛАЕВА, И. ДЖУМАНИЯЗОВА. ВОЗМОЖНОСТИ КОРПУСНОЙ ЛИНГВИСТИКИ В ИЗУЧЕНИИ РУССКОГО СЛОВООБРАЗОВАНИЯ	520
A.Q. ABDULLAYEV, B.B. ELOV, M.T. SAMATBOYEVA. O‘ZBEK TILIDAGI MATNLARDA ATOQLI OTLARNI AVTOMATIK ANIQLASH NE VA NER TUSHUNCHASI.....	525
M.F. ABDUNAZAROVA. GENDEROLINGVISTIKA ZAMONAVIY TILSHUNOSLIK SOHASI..	529
D. XUDAYBERGANOVA, Z. KASIMOVA. SOTSIOLINGVISTIKADA O‘RGANILADIGAN	535
J. MAHMUDOVA. AN ANALYSIS OF PRISON ARGOT AND ITS TYPES	540
K. IBODULLAYEVA, D. XOLBEKOVA. SCIENTIFIC DISCOURSE AS AN OBJECT OF STUDY OF SCIENTIFIC RESERCH	543
A.R. ISKANDAROVA. IJTIMOIIY MAQOM, IJTIMOIIY ROL, LISONIY ISHTIMOIIYLASHUV	548
A.R. ISKANDAROVA, G. JABBOROVA. TOHIR MALIKNING “ALVIDO BOLALIK” QISSASIDA QO‘LLANGAN SOTSIOLEKTLARNING SINTAKTIK KONSTRUKSIYALARI VA MA‘NOVIY	552
F. KULTAYEVA. ERGONIMLARNI MA‘NOVIY VA VAZIFAVIY KLASSIFIKATSIYALASH ..	556
S. BOYSARIYEVA. MORFEM LUG‘ATLAR KLASSIFIKATSIYASI VA TABIATI.....	561
M.B. BOTIROVA. O‘ZBEK TILIDA REKLAMALARNING SINTAKTIK STRUKTURASI	565
M.N. XUSANOVA. TARIXIY DIGNITONIM VA FALERONIMLARNING LISONIY TADQIQI ...	568
F.R. TOPILDIYEVA. FARIDA AFRO‘Z SHE‘RIYATINING LINGVOPOETIK TADQIQI VA O‘ZIGA XOS JIHATLARI.....	571
Z.T. TAXIROV. TIL, USLUBIYAT, NUTQ MADANIYATI VA ADABIY TAHRIR.....	575
D.B. JUMAYEVA, N.O. NOROVA. BOSHLANG‘ICH SINFLARDA IMLO QOIDALARI ASOSIDA TOG‘RI YOZISHNI SHAKLLANTIRISHGA O‘RGATISH	582
N. RASULOV, A. AKBAROVA. O‘ZBEK TILIDAGI HAKAKAT MAZMUNLI SLENGLAR	586
S. PO‘LATOVA. IZOGLOSSA	590
S.T. ISRAILOVA. TEZAURUSLARDA LEKSIK-SEMANTIK MOHIYATNING IFODASI.....	592
A.T. KAZAXBAEVA. T.JUMAMURATOV ASARLARI TILIDAGI PRAGMATIK BELGILAR.....	598
X.K. RAXIMOV. SENTIMENT TAHLIL DOIRASIDA OLIB BORILGAN ILMIY QARASHLAR TADQIQI.....	603
D. SULAYMANOVA. QO‘SHMA GAPLARDA TRANSFORMATSIYA: NOMINATIV BIRLIK (CHEGARALI QURILMA)LARDA POG‘ONAVIYLIK	608
M.A. ABJALOVA, G.SH. SAFAROVA, N. RAVSHANOVA. ZAMONAVIYLASHGAN VA “TIRILGAN” ESKI SO‘ZLAR.....	614
U.R. HALMUKHAMEDOVA. THE SIGNIFICANCE OF TRANSLATION TEXTS IN TEACHING THE UZBEK LANGUAGE.....	618
M.Q. ABDULLAYEVA. TILSHUNOSLIKDA MEDIAMATNNING O‘RNI.....	622
M.R. SIROJIDDINOVA, B.Sh. MIRZARAHIMOV. LINGVISTIK TADQIQOTLARDA STATISTIK METODNING O‘RNI.....	626
U. QO‘ZIYEV. O‘ZBEK TILI KORPUSINI YARATISHDA O‘ZLASHMA GRAMMATIK MODELLARNI AJRATISH MASALASI.....	630

M.A. ABJALOVA, M.D. TURDIALIYEVA. KORPUS MATERIALLARINI SEMANTIK TEGLAH PRINSIPLARI.....	634
M. MAMITOVA. O‘ZBEK TALABA YOSHLARI NUTQIDA SLENGLAR VA UNING ADABIY TILGA TA’SIRI.....	640
F.R. TOPILDIYEVA, D.R. MURODOVA. TIL VA MADANIYAT TADQIQI MASALALARI.....	645

O'ZBEK TILI KORPUSI MATNLARI ASOSIDA TIL MODELLARINI YARATISH

Elov Botir

Texnika fanlari falsafa doktori, dotsent.

Alisher Navoiy nomidagi

Toshkent davlat o'zbek tili va adabiyoti universiteti.

E-mail: elov@navoiy-uni.uz

ORCID: 0000-0001-5032-6648

Abdullayev Abdulla

O'qituvchi. "Urganch innovatsion universiteti"

NTM ta'limni kredit tizimini boshqarish bo'lim boshlig'i.

E-mail: abdulla_abdullayev9270@mail.ru

Xudayberganov Nizomaddin

O'qituvchi. Alisher Navoiy nomidagi

Toshkent davlat o'zbek tili va adabiyoti universiteti.

E-mail: nizomaddin@navoiy-uni.uz

ORCID: 0000-0002-6213-3015

Annotatsiya: Nutqni tanib olish, mashina tarjimasini, POS teglash, matnni intellektual tahlil qilish, yangi matnni yaratish va so'zlarni bashorat qilish kabi NLP vazifalarini hal qilish uchun til modellaridan foydalaniladi. Til modellari har qanday NLP vazifasining muhim komponent hisoblanadi. Ushbu maqolada til modellarini ishlab chiqish bosqichlari va o'zbek tili korpusi matnlari asosida til modellarini yaratishning n-gramm metodi keltiriladi.

Kalit so'zlar: til modellari, NLP, til korpusi, UzbNlp paketi, unigram, bigram, trigram, n-gramm, so'zni bashorat qilish.

Kirish

Tabiiy tilni qayta ishlash (Natural Language Processing, NLP) dagi til modeli – oldingi so'zlar asosida gapda berilgan so'zlar ketma-ketligining hosil bo'lish ehtimolini aniqlaydigan statistik model. Ushbu model asosida matn konteksti asosida mos so'zni bashorat qilish imkonini beradi. Til modellaridan nutqni aniqlash, mashina tarjimasini, imloni tuzatish va boshqa NLP vazifalarida keng foydalaniladi. Til modelini ishlab chiqish odatda katta hajmdagi namunaviy gaplarini mashinali o'qitish (ML) qadamidan boshlanadi. O'qitish jarayonida model so'zlar ketma-ketligi bo'yicha ehtimollik taqsimotini hisoblaydi. Kontekstdagi so'zni qo'yilgan NLP vazifasiga muvofiq bashorat qilish uchun *unigram*, *bigram*, *trigram* yoki *n-gram* usullaridan foydalanish mumkin.

Til modellari

Til modellari tabiiy tilni qayta ishlashning asosini tashkil qiladi. Til modeli matn haqidagi ma'lumotni mashinalar tushunadigan miqdoriy ma'lumotga aylantiradi [1,2]. Bugungi kunda til modellari asosida moliya, sog'liqni saqlash, harbiy va shu kabi turli sohalaridagi NLP ilovalari ishlab chiqilmoqda [3]. Kundalik hayotda mobil telefonda bashoratli matn kiritish yoki Google

qidiruv tizimida til modellaridan foydalaniladi. Demak, til modellari har qanday tabiiy tilni qayta ishlash dasturining ajralmas qismini tashkil qiladi. Ushbu maqolada til korpusidagi strukturlanmagan matn asosida unigram, bigram va trigram til modellarini yaratish va ular yordamida keyingi soʻzlarni bashorat qilish usullari keltiriladi [1,2,4].

Til korpusidan matnni o‘qish

Til modelini ishlab chiqish uchun o‘zbek tili milliy korpusidagi 2024 yil mart oyida kun.uz va qalampir.uz onlayn nashrlaridagi post/maqolalarni yuklab olamiz. Strukturlanmagan matnlar yuklab olinganidan so‘ng, matn faylidagi belgilarning umumiy sonini aniqlash lozim.

```
file = open("UzbCorpus.txt", "r")
```

```
rawReadCorpus = file.read()
```

```
print ("O'qitiladigan ma'lumotlar to'plamidagi belgilar soni: {}".format(len(rawReadCorpus)))
```

Toshkent davlat o'zbek tili va adabiyoti universiteti kompyuter lingvistikasi va raqamli texnologiyalar kafedrasida ilmiy tadqiqotchilari tomonidan ishlab chiqilgan UzbNlp paketida o'zbek tilidagi matnlarni tahlil qiluvchi bir qator funktsiya va metodlar mavjud [5]. Ushbu paketni import qilish lozim:

```
import UzbNlp
```

```
from UzbNlp.tokenize import uzb_word_tokenize, uzb_sent_tokenize
```

Strukturlanmagan matnni boshlang'ich qayta ishlash

Birinchi navbatda, matnlardagi barcha ortiqcha qatorlarni va maxsus belgilarni olib tashlash kerak [6]. Buni quyidagi kod bilan amalga oshiramiz:

```
import string
```

```
string.punctuation = string.punctuation + '“”‘’_!@!@!_'
```

```
string.punctuation = string.punctuation.replace('.', '')
```

```
file = open('UzbCorpus.txt').read()
```

```
# yangi qatorlar va maxsus belgilarni olib tashlash uchun ma'lumotlarni boshlang'ich qayta  
ishlash
```

```
file new = ""
```

```
for line in file:
```

```
line_new = line.replace("n", " ")
```

```
file new += line new
```

```
preprocessedCorpus = "".join([char for char in file_new if char not in string.punctuation])
```

Yangi qatorlar va maxsus belgilarni olib tashlaganimizdan so'ng, **UzbNlp.tokenize** dan **uzb_sent_tokenize** va **uzb_word_tokenize** metodlari yordamida so'zlar va gaplarni ajratib olish uchun korpus matnlarini parchalash mumkin [7,8]. Quyidagi dastur kodida korpusdan olingan dastlabki 5 ta gaplarni va birinchi 5 ta so'z chop qilingan:

```
sentences = sent_tokenize(preprocessedCorpus)
```

```
print("Korpusdagi birinchi 5 ta gap: ")
```

```
print(sentences[0:5])
```

```
words = word_tokenize(preprocessedCorpus)
```

```
print("Qayta ishlangan korpusdagi birinchi 5 ta soʻz/token: ")
```

```
print(words[0:5])
```

[Bilgil va ogoh bo'lg' ilki dunyoda ilm o'qimoq va ilm istamoqdin g'araz-u maqsud ikki dunyoning saodatiga yetmoqdur.]

[Lekin qancha mulla va qancha olim bo'lsang ham dilingdan axloqi zamima (buzuq xulqlar) va yomon fe'llarni chiqarib ul dilingni axloqi hamida va axloqi pisandidalar (yaxshi va maqtovg'a sazovor xulqlar) bilan ziynat-u oro bermasang hanuz ikki dunyoni saodatiga yeta olmassen, nechuk?]

[Zero ki ko'ngul bir podshoh va o'zga a'zolar fuqaro manzilindadur.]

[Agar ko'ngul durust bo'lsa, boshqa a'zolar ham durust bo'lur.]

[Va agar ko'ngul buzuq bo'lsa, o'zga a'zolar ham buzuq bo'lur.]

[Bilgil], [va], [ogoh], [bo'lg' ilki], [dunyoda]

Keyingi qadamda, korpusdagi **nomuhim so'zlar (stopwords)**ni olib tashlash kerak. Stopwords – bu gapga hech qanday maxsus ma'no yoki ahamiyatga ega bo'lmagan **"va"**, **"biroq"**, **"ammo"** kabi keng qo'llaniladigan so'zlar. O'zbek tili uchun nomuhim so'zlar ro'yxatiga namuna quyida keltirilgan:

```
UzbNlp.download('uzb_stopwords')
from UzbNlp.corpus import stopwords
stop_words = set(stopwords.words('uzbek'))
filtered_tokens = [w for w in words if not w.lower() in stop_words]
```

Unigram, bigram va trigram til modellarini yaratish

O'zbek tilidagi matnlarga mos n-gramm qiymatlarni UzbNlp paketidagi metodlar yordamida yaratish mumkin [5]. N-grammalar korpusda uchraydigan n ta ketma-ket so'zdan iborat ketma-ketlikdir. Masalan, **"Men kitob o'qishni yaxshi ko'raman"** gapida **"men"**, **"kitob"**, **"o'qishni"**, **"yaxshi"** va **"ko'raman"**, so'zlari unigramlar, **"men kitob"**, **"kitob o'qishni"**, **"o'qishni yaxshi"** va **"yaxshi ko'raman"** esa bigrammalardir. Quyidagi kod yordamida korpusdagi unigramlar, bigramlar va trigramlar aniqlanadi:

```
from collections import Counter
from UzbNlp.util import ngrams
unigrams=[]
bigrams=[]
trigrams=[]
for content in (sentences):
    content = content.lower()
    content = word_tokenize(content)
    for word in content:
        if (word == '.'):
            content.remove(word)
        else:
            unigrams.append(word)
            bigrams.extend(ngrams(content,2))
            trigrams.extend(ngrams(content,3))
```

```
print ("n-gramlar:n" + "-----")
print ("--> UNIGRAMLAR: n" + str(unigrams[:5]) + " ...n")
print ("--> BIGRAMLAR: n" + str(bigrams[:5]) + " ...n")
print ("--> TRIGRAMLAR: n" + str(trigrams[:5]) + " ...n")
```

UNIGRAMLAR:

[*'Bilgil', 'va', 'ogoh', 'bo'lg'ilki', 'dunyoda', 'ilm', 'o'qimoq', 'va', 'ilm', 'istamoqdin', 'g'araz-u', 'maqsud', 'ikki', 'dunyoning', 'saodatiga', 'yetmoqdur'*]

BIGRAMLAR:

[(*'Bilgil', 'va'*), (*'va', 'ogoh'*), (*'ogoh', 'bo'lg'ilki'*), (*'bo'lg'ilki', 'dunyoda'*), (*'dunyoda', 'ilm'*), (*'ilm', 'o'qimoq'*), (*'o'qimoq', 'va'*), (*'va', 'ilm'*), (*'ilm', 'istamoqdin'*), (*'istamoqdin', 'g'araz-u'*), (*'g'araz-u', 'maqsud'*), (*'maqsud', 'ikki'*), (*'ikki', 'dunyoning'*), (*'dunyoning', 'saodatiga'*), (*'saodatiga', 'yetmoqdur'*)]

TRIGRAMLAR:

[(*'Bilgil', 'va', 'ogoh'*), (*'va', 'ogoh', 'bo'lg'ilki'*), ... (*'dunyoning', 'saodatiga', 'yetmoqdur'*)]

Keyingi qadamda, korpusdan **ko'makchi**, **bog'lovchi** yoki **olmosh** kabi nomuhim so'zlarga ega bo'lmagan unigramma, bigramma va trigrammalarni shakllantirish lozim. N-gramlardan nomuhim so'zlarni olib tashlash uchun quyidagi koddan foydalanamiz.

```
def stopwords_removal(n, a):
```

```
    b = []
    if n == 1:
        for word in a:
            count = 0
            if word in stop_words:
                count = 0
            else:
                count = 1
            if (count==1):
                b.append(word)
        return(b)
    else:
        for pair in a:
            count = 0
            for word in pair:
                if word in stop_words:
                    count = count or 0
                else:
                    count = count or 1
            if (count==1):
                b.append(pair)
        return(b)
```

```
unigrams_Processed = stopwords_removal(1,unigrams)
```

```
bigrams_Processed = stopwords_removal(2,bigrams)
trigrams_Processed = stopwords_removal(3,trigrams)
print ("Korpusni boshlang`ich qayta ishlangandan keyingi n-gramlar ro`yxati:")
print ("--> UNIGRAMLAR: n" + str(unigrams_Processed[:5]) + " ...n")
print ("--> BIGRAMLAR: n" + str(bigrams_Processed[:5]) + " ...n")
print ("--> TRIGRAMLAR: n" + str(trigrams_Processed[:5]) + " ...n")
```

UNIGRAMLAR:

[‘Bilgil’, ‘ogoh’, ‘bo‘lg‘ilki’, ‘dunyoda’, ‘ilm’, ‘o‘qimoq’, ‘ilm’, ‘istamoqdin’, ‘g‘araz-u’, ‘maqsud’, ‘ikki’, ‘dunyoning’, ‘saodatiga’, ‘yetmoqdur’]

BIGRAMLAR:

[(‘Bilgil’, ‘ogoh’), (‘ogoh’, ‘bo‘lg‘ilki’), (‘bo‘lg‘ilki’, ‘dunyoda’), (‘dunyoda’, ‘ilm’), (‘ilm’, ‘o‘qimoq’), (‘o‘qimoq’, ‘ilm’), (‘ilm’, ‘istamoqdin’), (‘istamoqdin’, ‘g‘araz-u’), (‘g‘araz-u’, ‘maqsud’), (‘maqsud’, ‘ikki’), (‘ikki’, ‘dunyoning’), (‘dunyoning’, ‘saodatiga’), (‘saodatiga’, ‘yetmoqdur’)]

TRIGRAMLAR:

[(‘Bilgil’, ‘ogoh’, ‘bo‘lg‘ilki’), ... (‘dunyoning’, ‘saodatiga’, ‘yetmoqdur’)]

Yuqoridagi natijalar asosida korpusda hosil bo‘ladigan har bir n-grammning sonini yoki chastotasini aniqlash mumkin. Ushbu natija n-grammlar asosida kontekstdagi keyingi mumkin bo‘lgan so‘zning ehtimolini hisoblash uchun kerak bo‘ladi. Quyida keltirilgan **get_ngrams_freq_dist** funksiyasi orqali har bir n-gramga mos keladigan chastota aniqlanadi [9]. Ushbu funksiya orqali korpusdagi barcha unigramlar, bigramlar va trigramlarning chastotalari hisoblanadi.

```
def get_ngrams_freq_dist(n, ngramList):
    ngram_freq_dict = {}
    for ngram in ngramList:
        if ngram in ngram_freq_dict:
            ngram_freq_dict[ngram] += 1
        else:
            ngram_freq_dict[ngram] = 1
    return ngram_freq_dict
unigrams_freqDist = get_ngrams_freqDist(1, unigrams)
unigrams_Processed_freqDist = get_ngrams_freqDist(1, unigrams_Processed)
bigrams_freqDist = get_ngrams_freqDist(2, bigrams)
bigrams_Processed_freqDist = get_ngrams_freqDist(2, bigrams_Processed)
trigrams_freqDist = get_ngrams_freqDist(3, trigrams)
trigrams_Processed_freqDist = get_ngrams_freqDist(3, trigrams_Processed)
```

Bigram va trigram til modellari yordamida keyingi uchta so‘zni bashorat qilish

Til modelida gapning ehtimolini hisoblash uchun zanjir qoidasi qo‘llaniladi. $w_1w_2...w_n$ gap, undagi w_1 , w_2 , w_n alohida so‘zlar/tokenlar bo‘lsin. Ushbu gapning hosil bo‘lish ehtimoli quyidagi formula bilan aniqlanadi:

$$P(w_1, w_2 \dots w_n) = \prod_i P(w_i | w_1, w_2 \dots w_{i-1})$$

Masalan, "Men kitob o`qishni yaxshi ko'raman" gapining ehtimoli quyidagicha aniqlanadi:

$$\begin{aligned} P(\text{"Men kitob o`qishni yaxshi ko'raman"}) &= \\ P(\text{"men"})P(\text{"kitob"}|\text{"men"})P(\text{"o`qishni"}|\text{"men kitob"}) & \\ P(\text{"yaxshi"}|\text{"men kitob o`qishni"}) & \\ P(\text{"ko'raman"}|\text{"men kitob o`qishni yaxshi"}) & \end{aligned}$$

Yuqoridagi formuladagi individual ehtimolliklar quyidagi usul orqali aniqlanadi:

$$\begin{aligned} P(\text{"men"}) &= \text{Count}(\text{"men"}) / \text{umumiy so`zlar soni} \\ P(\text{"kitob"}|\text{"men"}) &= \text{Count}(\text{"men kitob"}) / \text{Count}(\text{"men"}) \\ P(\text{"o`qishni"}|\text{"men kitob"}) &= \text{Count}(\text{"men kitob o`qishni"}) / \text{Count}(\text{"men kitob"}) \\ P(\text{"yaxshi"}|\text{"men kitob o`qishni"}) &= \text{Count}(\text{"men kitob o`qishni yaxshi"}) / \text{Count}(\text{"men kitob o`qishni"}) \\ P(\text{"ko'raman"}|\text{"men kitob o`qishni yaxshi"}) &= \text{Count}(\text{"men kitob o`qishni yaxshi ko'raman"}) / \text{Count}(\text{"men kitob o`qishni yaxshi"}) \end{aligned}$$

Bunda,

$\text{Count}(\text{"men"}), \text{Count}(\text{"men kitob"}),$

$\text{Count}(\text{"men kitob o`qishni"}),$

$\text{Count}(\text{"men kitob o`qishni yaxshi"})$  va  $\text{Count}(\text{"men kitob o`qishni yaxshi ko'raman"})$

qiymatlar yuqoridagi `get_ngrams_freq_dist(n, ngramList)` funksiyasidan foydalangan holda hisoblagan tegishli unigram, bigram va trigramma chastotalaridir.

Yuqordagi formulalardagi ehtimolliklarni hisoblash uchun bigram modelidan foydalanganda, har bir yangi so'zning ehtimolligi faqat oldingi so'zga bog'liqligini qayd etish mumkin. Ya'ni, oldingi misol uchun, gapning ehtimoli quyidagicha hisoblanadi:

$$\begin{aligned} P(\text{"Men kitob o`qishni yaxshi ko'raman"}) &= \\ P(\text{"men"})P(\text{"kitob"}|\text{"men"})P(\text{"o`qishni"}|\text{"kitob"}) & \\ P(\text{"yaxshi"}|\text{"o`qishni"})P(\text{"ko'raman"}|\text{"yaxshi"}) & \end{aligned}$$

Xuddi shunday, trigramm modeli uchun ehtimollik quyidagicha hisoblanadi:

$$\begin{aligned} P(\text{"Men kitob o`qishni yaxshi ko'raman"}) &= \\ P(\text{"men"})P(\text{"kitob"}|\text{"men"})P(\text{"o`qishni"}|\text{"men kitob"}) & \\ P(\text{"yaxshi"}|\text{"kitob o`qishni"}) & \\ P(\text{"ko'raman"}|\text{"o`qishni yaxshi"}) & \end{aligned}$$

Trigramm modellashirishni quyidagi diagramma bilan yaxshiroq tushuntirish mumkin:

Berilgan gap	3-gramm model
$P(\text{"Men kitob o`qishni yaxshi ko'raman"}) =$ $P(\text{"men"})$ <ul style="list-style-type: none"> <li><math>P(\text{"kitob"} \text{"men"})</math></li> <li><math>P(\text{"o`qishni"} \text{"men kitob"})</math> $P(\text{"yaxshi"} \text{"men kitob o`qishni"})</math></li> <li><math>P(\text{"ko'raman"} \text{"men kitob o`qishni yaxshi"})$	$P(\text{"Men kitob o`qishni yaxshi ko'raman"}) =$ $P(\text{"men"})$ <ul style="list-style-type: none"> <li><math>P(\text{"kitob"} \text{"men"})</math></li> <li><math>P(\text{"o`qishni"} \text{"men kitob"})</math> $P(\text{"yaxshi"} \text{"men kitob o`qishni"})</math></li> <li><math>P(\text{"ko'raman"} \text{"men kitob o`qishni yaxshi"})$

Biroq, bunday modellashirishning o`ziga xos jihati bor. O`quv namunalarida (dataset) mavjud bo`lmagan, ammo joriy matn/korpusda mavjud qandaydir bigramma bor. Yuqoridagi

hisob-kitoblar asosida ushbu bigrammaga **0** ehtimol belgilanadi va gapdagi umumiy ehtimollik qiymati **0** ga teng bo‘ladi. Ushbu muammoni bartaraf etish uchun “**silliqlash**” (**smoothing**) amalga oshiriladi. Ba‘zi ehtimollik qiymatlarini hisobga olimagan (ko‘rib chiqilmagan) hodisalarga qayta belgilash uchun parametrlar tartibga solinadi. Bunday yumshatish usullaridan biri **Add-one** yoki **Laplace smoothing** bo‘lib, yuqoridagi til modelini ishlab chiqishga joriy qilamiz ushbu maqolada foydalanamiz. Buning uchun barcha bigramm qiymatlariga **1** va barcha unigramma qiymatlarga **V** (korpusdagi unikal so‘zlar soni)ni qo‘shish lozim.

Avvalgi baho:

$$P_{MLE}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i)}{c(w_{i-1})}$$

Add-one baho:

$$P_{Add-1}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) + 1}{c(w_{i-1}) + V}$$

Silliqlangan bigram va trigram modellari haqida ma’lumot hosil bo‘lganidan so‘ng, modelda mos o‘zgarishlarni amalga oshirish kerak. Bashorat qilish uchun qayta ishlanmagan bigramm va trigram elementlaridan (nomuhim so‘zlarga ega matn) foydalanamiz.

```
smoothed_bigrams_probDist = {}
```

```
V = len(unigrams_freqDist)
```

```
for i in bigrams_freqDist:
```

```
    smoothed_bigrams_probDist[i] = (bigrams_freqDist[i] + 1)/(unigrams_freqDist[i[0]]+V)
```

```
smoothed_trigrams_probDist = {}
```

```
for i in trigrams_freqDist:
```

```
    smoothed_trigrams_probDist[i] = (trigrams_freqDist[i] + 1)/(bigrams_freqDist[i[0:2]]+V)
```

Keyingi qadamda, hisoblangan bigram va trigram tili modellari yordamida keyingi uchta so‘z bashorat qilinadi.

```
testSent1 = "Qolaversa, durustgina "
```

```
testSent2 = "Har bir odam uchun"
```

```
testSent3 = "Aytmat bobom qo‘li gul usta edi"
```

Berilgan namunaviy gaplarni tokenlarga ajratilib, ulardagi oxirgi unigramlar va bigramlar aniqlanadi.

```
token_1 = word_tokenize(testSent1)
```

```
token_2 = word_tokenize(testSent2)
```

```
token_3 = word_tokenize(testSent3)
```

```
ngram_1 = {1:[], 2:[]}
```

```
ngram_2 = {1:[], 2:[]}
```

```
ngram_3 = {1:[], 2:[]}
```

```
for i in range(2):
```

```
    ngram_1[i+1] = list(ngrams(token_1, i+1))[-1]
```

```
    ngram_2[i+1] = list(ngrams(token_2, i+1))[-1]
```

```
    ngram_3[i+1] = list(ngrams(token_3, i+1))[-1]
```

```
print("1-gap: ", ngram_1, "2-gap: ", ngram_2, "3-gap: ", ngram_3)
```

So'ngra, yangilangan bigram modelidan foydalanib, berilgan namunaviy gaplarga mos keyingi 3 ta so'zni bashorat qilish funksiyalarini shakllantiramiz.

```
def predict_next_word(last_word,probDist):
    next_word = {}
    for k in probDist:
        if k[0] == last_word[0]:
            next_word[k[1]] = probDist[k]
    k = Counter(next_word)
    high = k.most_common(1)
    return high[0]

def predict_next_3_words(token,probDist):
    pred1 = []
    pred2 = []
    next_word = {}
    for i in probDist:
        if i[0] == token:
            next_word[i[1]] = probDist[i]
    k = Counter(next_word)
    high = k.most_common(2)
    w1a = high[0]
    w1b = high[1]
    w2a = predict_next_word(w1a,probDist)
    w3a = predict_next_word(w2a,probDist)
    w2b = predict_next_word(w1b,probDist)
    w3b = predict_next_word(w2b,probDist)
    pred1.append(w1a)
    pred1.append(w2a)
    pred1.append(w3a)
    pred2.append(w1b)
    pred2.append(w2b)
    pred2.append(w3b)
    return pred1,pred2

print("Silliqlashtirilgan bigram modeli yordamida keyingi 3 ta mumkin bo'lgan so'zlar ketma-
ketligini bashorat qilish:")
pred1,pred2 = predict_next_3_words(ngram_1[1][0],smoothed_bigrams_probDist)
print("1a)" + testSent1 + " " + '33[1m' + pred1[0][0] + " " + pred1[1][0] + " " + pred1[2][0] + '33[0m')
print("1b)" + testSent1 + " " + '33[1m' + pred2[0][0] + " " + pred2[1][0] + " " + pred2[2][0] + '33[0m')
pred1,pred2 = predict_next_3_words(ngram_2[1][0],smoothed_bigrams_probDist)
print("2a)" + testSent2 + " " + '33[1m' + pred1[0][0] + " " + pred1[1][0] + " " + pred1[2][0] + '33[0m')
print("2b)" + testSent2 + " " + '33[1m' + pred2[0][0] + " " + pred2[1][0] + " " + pred2[2][0] + '33[0m')
pred1,pred2 = predict_next_3_words(ngram_3[1][0],smoothed_bigrams_probDist)
print("3a)" + testSent3 + " " + '33[1m' + pred1[0][0] + " " + pred1[1][0] + " " + pred1[2][0] + '33[0m')
print("3b)" + testSent3 + " " + '33[1m' + pred2[0][0] + " " + pred2[1][0] + " " + pred2[2][0] + '33[0m')
```

Yangilangan bigram modelining bashoratlari:

- 1a) Qolaversa, durustgina baxshi ham edi
- 1b) Qolaversa, durustgina rassom ham edi
- 2a) Har bir odam uchun dunyo ostonadan boshlanadi
- 2b) Har bir odam uchun hayot juda go`zal
- 3a) Ahmad bobom qo`li gul usta edi
- 3b) Ahmad bobom qo`li juda chiroyli edi

Xuddi shunday uslubda yangilangan trigram modelidan bashorat olamiz.

`def predict_next_word(last_word,probDist):`

```
next_word = {}
for k in probDist:
    if k[0:2] == last_word:
        next_word[k[2]] = probDist[k]
k = Counter(next_word)
high = k.most_common(1)
return high[0]
```

`def predict_next_3_words(token,probDist):`

```
pred = []
next_word = {}
for i in probDist:
    if i[0:2] == token:
        next_word[i[2]] = probDist[i]
k = Counter(next_word)
high = k.most_common(2)
w1a = high[0]
tup = (token[1],w1a[0])
w2a = predict_next_word(tup,probDist)
tup = (w1a[0],w2a[0])
w3a = predict_next_word(tup,probDist)
pred.append(w1a)
pred.append(w2a)
pred.append(w3a)
return pred
```

`print("Silliqlashtirilgan trigram modeli yordamida keyingi 3 ta mumkin bo`lgan so`zlar ketma-ketligini bashorat qilish:")`

```
pred = predict_next_3_words(ngram_1[2],smoothed_trigrams_probDist)
print("1)" +testSent1 + " " + '33[1m' + pred[0][0]+ " " +pred[1][0]+ " " +pred[2][0] + '33[0m')
pred = predict_next_3_words(ngram_2[2],smoothed_trigrams_probDist)
print("2)" +testSent2 + " " + '33[1m' + pred[0][0]+ " " +pred[1][0]+ " " +pred[2][0] + '33[0m')
pred = predict_next_3_words(ngram_3[2],smoothed_trigrams_probDist)
print("3)" +testSent3 + " " + '33[1m' + pred[0][0]+ " " +pred[1][0]+ " " +pred[2][0] + '33[0m')
```

- 1) Qolaversa, durustgina omadli biznesmenga aylanishdi
- 2) Har bir odam uchun eng yaxshi kun

3) Ahmad bobom qo'li yarador holda qaytdi

Xulosa.

Tilni modellashtirish (Language modeling, LM) kontekstda so'zlarni bashorat qilish uchun korpus matnlarini tahlil qiladi va tabiiy til modelini shakllantiradi. Ushbu modellar gapda ma'lum bir so'z ketma-ketligining yuzaga kelish ehtimolini aniqlash uchun statistik va ehtimollik usullaridan foydalanadi. NLP kontseptsiyalariga asoslangan ilovalar va dasturlar audiodan matnga o'tkazish, hissiyotlarni tahlil qilish, nutqni aniqlash va imlo tuzatish kabi vazifalarda til modellaridan foydalaniladi. Til modellari matn ma'lumotlarining tahlil qilingan qismidagi so'z ehtimolini aniqlash orqali ishlaydi. Ma'lumotlar ushbu tabiiy tilda kontekst qoidalarini qidiradigan mashinali o'rganish algoritmgiga berilgandan so'ng qo'llaniladi. Ushbu maqolada o'zbek tili kospusi matnlarini n-gramm usuliga asoslangan til modeli ishlab chiqildi va namunaviy gaplarga qo'llash usullari keltirildi.

Foydalanilgan adabiyotlar:

1. Jurafsky, D., & Martin, J. H. (2019). Chapter 3: N-Gram Language Models N-Gram Language Models. *Speech and Language Processing*.
2. Chen, M., Suresh, A. T., Mathews, R., Wong, A., Allauzen, C., Beaufays, F., & Riley, M. (2019). Federated learning of N-gram language models. *CoNLL 2019 - 23rd Conference on Computational Natural Language Learning, Proceedings of the Conference*. <https://doi.org/10.18653/v1/k19-1012>
3. Republic, C., & Mikolov, T. (2012). Statistical Language Models Based on Neural Networks. *Wall Street Journal, April*. <https://doi.org/10.1016/j.csl.2015.07.001>
4. P.~Brown, V.~Della Pietra, de Souza, P., J.~Lai, & R.~Mercer. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, 18.
5. Boltayevich, E. B., Mirdjonovna, H. S., & Ilxomovna, A. X. (2023). Methods for Creating a Morphological Analyzer. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13741 LNCS. https://doi.org/10.1007/978-3-031-27199-1_4
6. Bommasani, R., Liang, P., & Lee, T. (2023). Holistic Evaluation of Language Models. *Annals of the New York Academy of Sciences*, 1525(1). <https://doi.org/10.1111/nyas.15007>
7. Konstantopoulos, S. (2010). Learning language identification models: A comparative analysis of the distinctive features of names and common words. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*.
8. Wallace, E., Gardner, M., & Singh, S. (2020). *Interpreting Predictions of NLP Models*. <https://doi.org/10.18653/v1/2020.emnlp-tutorials.3>
9. Roh, J., Park, S., Kim, B. K., Oh, S. H., & Lee, S. Y. (2021). Unsupervised multi-sense language models for natural language processing tasks. *Neural Networks*, 142. <https://doi.org/10.1016/j.neunet.2021.05.023>