

СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА INTERNATIONAL JOURNAL OF WORD ART

Гули Тоирова Ибрагимовна
филология фанлари бўйича фалсафа
доктори(PhD) доцент,
Бухоро давлат университети
E-mail: tugulijon@mail.ru

ТИЛ КОРПУСИЛАРИ ВА МИЛЛИЙ КОРПУСНИ ЯРАТИШ ХУСУСИДА

 <http://dx.doi.org/10.26739/2181-9297-2021-4-69>

АННОТАЦИЯ

Мақолада корпуснинг кўп қиррали лингвистик манба сифатидаги асосий аҳамияти, корпус асосан икки турдаги маълумотга эга эканлиги ҳамда унинг турлари таҳлил қилинган. Мақолада матн таҳлили жараёнига сарфланадиган вақтнинг тежалиши, тил бирликларининг нутқ таркибидаги хусусиятларини минглаб мисоллар билан кўрсата олиш корпуснинг асосий самарали имкониятлари сифатида эътироф этилган. Компьютер лингвистикаси соҳасида миллий корпус, таълимий корпус, параллел корпус ҳақида маълумот берилган. Уларни лингвистик, экстралингвистик теглаш, корпус тузиш алгоритмини ишлаб чиқиш, корпус учун лингвистик таъминот яратиш ижтимоий зарурат эканлиги таъкидланган. Унда ўзбек тили корпусларини яратишнинг асосларини ишлаб чиқиш, компьютер лингвистикаси йўналишида тадқиқотларни амалга ошириш илмий-назарий манба сифатида долзарблиги эътироф этилган.

Калит сўзлар: корпус, сунъий интеллект, лексик маълумот, морфологик белги, сўз, гап, миллий корпус, таълимий корпус, параллел корпус, абзац, матн, разметка, интерфейс

Гули Тойрова Ибрагимовна
Доктор философских наук, доцент
Бухарского государственного университета.
E-mail: tugulijon@mail.ru

ЯЗЫКОВЫЕ ЗДАНИЯ И СОЗДАНИЕ НАЦИОНАЛЬНЫХ ЗДАНИЙ

АННОТАЦИЯ

В статье анализируется основная роль корпуса как многогранного лингвистического источника, то обстоятельство, что корпус в основном содержит информацию двух типов, и ее типы. В статье признается, что экономия времени, затрачиваемого на процесс анализа текста, возможность иллюстрировать особенности языковых единиц в речи тысячами примеров, являются основными эффективными возможностями корпуса. В области компьютерной лингвистики дается информация о национальном корпусе, учебном корпусе, параллельном корпусе. Было подчеркнуто, что лингвистическая, экстралингвистическая их маркировка, разработка алгоритма создания корпуса, создание лингвистической поддержки корпуса является социальной необходимостью. Признается актуальность разработки основы для

создания корпуса узбекского языка, проведения исследований в области компьютерной лингвистики как научно-теоретического источника.

Ключевые слова: корпус, искусственный интеллект, лексическая информация, морфологический признак, слово, предложение, национальный корпус, учебный корпус, параллельный корпус, абзац, текст, знак, интерфейс.

Guli Toirova Ibragimovna
Doctor of Philosophy (PhD),
Associate Professor,
Bukhara State University
E-mail: tugulijon@mail.ru

LANGUAGE BUILDINGS AND CREATION OF NATIONAL BUILDING

ANNOTATION

The article analyzes the main role of the corpus as a multifaceted linguistic source, the fact that the corpus mainly has two types of information, and its types. The article recognizes that saving time spent on the text analysis process, being able to illustrate the features of language units in speech with thousands of examples, are the main effective capabilities of the corpus. In the field of computer linguistics, information is given about the national corpus, the educational corpus, the parallel corpus. It was emphasized that linguistic, extralinguistic tagging of them, development of corpus creation algorithm, creation of linguistic support for corpus is a social necessity. It recognizes the urgency of developing the basis for the creation of the Uzbek language corpus, conducting research in the field of computer linguistics as a scientific and theoretical source.

Keywords: corpus, artificial intelligence, lexical information, morphological sign, word, sentence, national corpus, educational corpus, parallel corpus, paragraph, text, mark, interface

Кириш. Компьютер технологиялари воситасида ахборотни қайта ишлаш, машина таржимаси, электрон луғатшуносликни ривожлантириш, тезауруслар тузиш, тил корпусини яратиш ишларини дунё лингвистикаси цивилизациясига эришган тиллар аллақачон амалга оширганлигини таъкидлаган эдик. Инглиз, рус, араб, француз, немис, испан, тожик каби тиллар шулар жумласидандир. Мазкур тилларда интернет тизимида тил корпусини яратиш масаласининг илмий-назарий жиҳатлари ҳам ишлаб чиқилганки, бу ўзбек тилини ҳам интернет “тушунадиган” тилга айлантириш (яъни тил корпусини яратиш) саъй-ҳаракатларини жадаллаштириш заруратини кун тартибига қўяди.

Сунъий интеллект орқали тилнинг имкониятларидан фойдаланиш борасида замонавий ахборот технологиялари бениҳоя кенг қулайликлар эшигини очди. У инсон онги бажариши мумкин бўлган бир қанча вазифаларнинг бажара олмоқда. Сунъий интеллект маҳсули бўлмиш электрон манбалар инсонга зарар етказмаслик ва инсонларнинг оғирини енгиллаштириш мақсадида яратилмоқда. Чунончи, ўзбек тилининг интернет ва электрон тилга айланиши, миллий тилнинг электрон ресурсларини (ўзбек тили корпуси, электрон луғатлари, интернет саҳифаларидаги матнлар) ва технологияларини такомиллаштириш долзарб вазифалардандир.

Методлар. Корпус – бу корпус лингвистикасининг предмети. Ушбу атама илмий адабиётларда турлича таърифланади. Чунончи, инглиз тилида linguistic corpus ёки text corpus каби атамалар билан қўлланилади[8]. Муайян тилнинг миллий корпусни яратиш, унинг аналитик технологиясини, корпус лингвистикаси соҳасини ривожланиш каби муаммоларни ҳал қилишда А.Н.Хомский, Г.Н.Луч, Ч.Ф.Меер, Ж.Синклер, М.З.Курди кабиларнинг илмий изланишларини эътироф этиш лозим [18,19,22,23].

Джон Синклеър “корпус” атамасини қуйидагича таърифлайди: “ корпус тил ёки лингвистик хилма-хилликни тадқиқ этиш, маълумот манбаи сифатида намойиш этиш учун кўринадиган мезонларга мувофиқ танланган электрон кўринишдаги матнлар фрагментида иборат” [7].

Рус корпус лингвистикасида катта массивли матнлар мажмуаси, корпус тузиш принциплари, лингвистик маълумотлар базаси В.Г.Бритвин, В.П.Захаров, И.А.Мелчук, А.Б.Кутузов, Р.Г.Котов, Л.И.Беляева, Е.В.Недошивина, В.В.Рыков, В.Плунгянларнинг мақсадли тадқиқотларида ўз аксини топган[1,2,3,4,5,6,8,9].

Рус олими В.П. Захаров “корпус” атамасини қуйидагича изоҳлайди: “корпус – оғзаки ҳамда ёзма матнлар асосида жамланган тил бирликларининг лингвистик маълумотлар мажмуи”[4].

Ўзбек тилшунослигида матнни лингвостатистик таҳлил этиш, унга лексикографик ишлов бериш, автоматик таҳрир қилувчи дастурнинг лингвистик таъминотини, таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари, миллий корпусининг синоним сўзлар базаси, муаллифлик корпусининг лингвистик асослари борасида Х.Исхакова, С.Мухаммедов, С.Ризаев, С.Мухаммедова, Б.Менглиев, Д.Ўринбоева, А.Пўлатов, У.Дысимова, Г.Валиева, Г.Жуманазарова, Н.Абдурахмонова, Ш.Ҳамроева, М.Абжалова, А.Эшмўминов, Ў.Холиёров, Р.Каримовларнинг ишлари диққатга сазовор. С.Каримов, С.Мухаммедова Ш.Ҳамроева каби олимларимиз корпус лингвистикасининг 10.00.01- ихтисослиги бўйича илмий тадқиқот ишлари олиб боришган.

Ўзбек тилшунос олимлари эса “корпус” атамасига қуйидагича изоҳ беради: Ўзбек тилшунос олимлари эса “корпус” атамасига қуйидагича изоҳ беради: “корпус – маълум мақсадда йиғилган матнлар мажмуини ташкил этувчи тил бирликларининг йиғиндиси” [14], “корпус – исталган табиий (реал) тилдаги электрон шаклда сақланадиган ёзма ёки оғзаки, компьютерлаштирилган қидирув тизимига жойлаштирилган матнлар йиғиндиси”[17] каби таърифларда ўз ифодасини топган. Ўзбек тилшунослигидаги тадқиқотларда корпуснинг моҳияти қуйидагича ёритилади: “Корпус – мавжуд маълумотларни матн ҳолида тақдим эта олиш; корпуснинг ҳажмига кўра имкон даражада кўпроқ ахборот беришнинг имконияти; бир марта яратилган корпус маълумотларидан такроран, турли масалаларни ҳал қилишда фойдаланиш имкониятидир” [13].

“Корпус – тил бирликларининг хусусиятларини аниқлаш мақсадида қидирув дастурига бўйсундирилган матнлар мажмуи, табиий тилдаги электрон шаклда сақланадиган ёзма ёки оғзаки, компьютерлаштирилган қидирув тизимига дастурий таъминот асосида жойлаштирилган on-line ёки off-line тизимда ишлайдиган матнлар жамланмаси”[12] – деб изоҳланади бошқа бир манбада.

Шунингдек, “таълимий корпус” ҳақида илмий тадқиқот олиб борган Ў.Холиёров ўз ишида қуйидаги фикрларни айтиб ўтади: “Таълимий корпус – материаллари тилни ўргатишга йўналтирилган, лингводидактик хусусиятга эга тил корпуси. Ўзбек тилининг таълимий корпуси ўзбек тили имкониятларини ўргатишга мўлжалланган, лингводидактик хусусият касб этган, электрон матнларни қамраб олган, махсус сайт кўринишида амал қиладиган корпусдир”[16].

Паралел корпус ҳақида Р.Каримов шундай фикрларни айтади: “таржима матнларининг параллел электрон аналог; бир қанча «оригинал матн ва уларнинг бир/бир неча таржимаси» блокларидан иборат”[10].

Тил корпусларини тузилиши, мақсади, турғун, ўзгарувчанлиги каби жиҳатдан турлича кўринишларга ажратиш мумкин. Масалан, В.П.Захаров қуйидаги кўринишларини санаб беради: “маълумотларни сақлаш шаклига кўра: овозли, ёзма, аралаш; матн тилига кўра: бир ва кўп тилли; жанрий мансублигига кўра: адабий, диалектал, оғзаки, публицистик, аралаш; корпусга кириш имкониятига кўра:эркин, тижорат корпуслари, ёпиқ; мақсадига кўра: тадқиқий, иллюстратив; ўзгарувчанлигига кўра: динамик ва турғун; қўшимча ахборотга (аннотацияланган) эгалигига кўра разметкаланган ва разметкаланмаган ”[4].

В.В. Рыков эса корпус турларини таснифлашда қуйидаги жиҳатларга эътиборни қаратади: “Маълумотларнинг даражаси ва қурилишига кўра, тилнинг хронологик белгиси(ҳолати)га кўра, қўлланилиш тилига кўра, қўлланилиш мақсадига кўра”[6].

Ш.Ҳароева ўз тақиқотида корпусни қуйидаги кўринишларга ажратади: “Тилнинг маълум даврига ёки унинг маълум бир воқеланиш тури (жанр, услуб, бирор ижтимоий ёки

ёшга оид гуруҳ, ёзувчи ёки олим тили)га кўра; лингвистик разметка турига кўра; нутқ турига кўра: ёзма, оғзаки, аралаш; мультимодал корпус, махсус матнлар корпуси каби кўринишлари мавжуд”[17].

Ў.Холиёров эса корпус турлари дея қуйидагиларни кўрсатади:” Ихтисослашган корпус: маълум бир турдаги матнлар тўплами: газета матни, илмий мақолалар; умумий корпус; киёсий корпус; параллел корпус; таълимий корпус; дидактик корпус”[16].

Натижалар. Ҳар бир олим ўз нуқтаи назаридан тил корпусларига таъриф берган ва уларни турли жиҳатларидан таснифлаган. Хўш, ўзбек тили корпуси ўзида қайси жиҳатларни акс эттириш ва бугун қандай корпусларни яратиш устида қизғин ишлар амалга оширилмоқда?

Ўзбек тили миллий корпусни яратиш ўзбек тилшунослигида ҳам, замонавий ахборот технологиялари тизимида ҳам нисбатан янги йўналиш ҳисобланади. Тил корпуси катта ҳажмдаги луғатларни тузиш учун асосий манба ва кучли ахборот ресурсидир. Тил корпуси компютер ёрдамида луғатларнинг тезлик билан тузилиши ва қайта ишланишини таъминлайди. Корпуснинг сўзшунослик соҳасидаги аҳамияти шундаки, сўзнинг қўлланиш даври ва частотасини аниқлашда ҳеч қандай восита корпусга тенглаша олмайди. Яқин орада тил ўрганаётган ўқувчи ёки унинг бирор жиҳатини тадқиқ этаётган тадқиқотчининг бугунги кунда луғатга бўлган эҳтиёжи корпусга қўчиши шубҳасиз.

Бугунги кунда Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетининг тилшунос олимлари томонидан “Таълимий корпус” деб номланган лойиҳа бўйича илмий ва амалий ишлар олиб борилмоқда. Ўзбек тилининг таълимий корпусини яратиш хорижий тажрибалардан келиб чиқиб, босқичма-босқич маълумотлар шакллантирилишига қаратилган ва ўзбек адабий тилининг бугунги сўзлигини қамраб олган электрон дарслик, кўп тилли сўзлашгичлар ва ўзбек тилининг бошқа тилларда эквиваленти бўлмаган, таржима қилинмайдиган лексик бирликлари – миллий-маданий сўзларнинг изоҳи берилган глоссарий, ўзбекча тўғри талаффуз кўникмаларини шакллантиришга қаратилган аудио ва видеоматериаллар ҳамда мобил илова қамраб олинган мультимедиа маҳсулотлари мажмуасини яратиш кўзда тутилган.

Таълимий корпус - ўзбек тилини давлат тили ва иккинчи тил, шунингдек хорижий тил сифатида чуқурроқ ўрганиш имконини беради. Таълимий корпусни ташкил этувчи электрон контент - аудио, видео, мултимедия иловалар, талаффуз ва имло қоидаларни ўргатувчи дастурлар, электрон ўқув луғатлари фойдаланувчиларга ўзбек тилини мустақил ҳолатда ҳам қулай ўрганишга шароит яратади. Амалдаги ўқув дастурлари мазмунидан фарқли равишда мазкур комплекс ўзбек тилини нотаниш вазиятларда ҳам қўллай олиш кўникмаларини шакллантиради. Ундан ўзбек тилини ўрганишни бошловчилар, ўқувчилар, ота-оналар, ўқитувчилар, талабалар кенг жамоатчилик фойдаланиши мумкин. Бу эса республикамизнинг иқтисодий ўсиши ва ижтимоий ривожланишини таъминлайдиган илмий ва технологик захираларни шакллантиришга хизмат қилади.

Ўзбек компютер лингвистикаси соҳасида ўзбек тилига замонавий ахборот технологиялари воситасида ишлов бериш борасидаги илмий изланишларга эътибор қаратилиши натижасида амалий ишлар сифатида миллий корпуснинг илк кўринишлари юзага келмоқда.

“Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетининг олимларининг ташаббуси билан 2021-йилнинг май ойида ўтказилган “Ўзбек миллий ва таълимий корпусларини яратишнинг назарий ва амалий масалалари” мавзусидаги халқаро илмий-амалий конференцияда нафақат юртимиздаги балки, халқаро миқёсдаги мутахассислар ҳам амалий таклифлари билан иштирок этди. Конференция якунида халқаро тажрибаларни ўзлаштирган ҳолда, мукамал миллий ва таълимий корпусларини яратиш борасидаги аниқ вазифалар белгилаб олинди.

Шунингдек, Самарқанд давлат университети олимлари томонидан “Ўзбек тилининг миллий корпусини лойиҳалаш ва дастурий мажмуа ишлаб чиқиш” деб номланган лойиҳа ҳамда Бухоро давлат университети олимлари томонидан ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалаларига бағишланган тадқиқот ишлари олиб борилмоқда.

Олиб борилаётган тадқиқотларнинг натижалари ўзбек тилининг халқаро миқёсдаги мақомини ошириш, уни жаҳон мулоқот тили даражасига кўтариш, ўзбек тилини чет элларда ўрганиш ва ўргатиш, миллий тилимизнинг имкониятларини кенгайтиришда катта аҳамиятга эга.

Ўзбек тили миллий корпусининг яратилиши ўзбек тилини “рақамлаштириш”, интернет тилига айлантириш имконини беради.

Таълим олиш самарадорлигини ошириш учун янги имкониятлар эшигини очиб беради. Корпус орқали жуда осонлик билан кам ишлатиладиган сўз, ибора ёки сўз бирикмасини топиш ёки унинг қўлланиши ва имлоси (орфографияси) билан боғлиқ муаммо жуда қисқа муддат ичида ҳал этилади. Бугун грамматика билан шуғулланувчи олимдан кўра оддий тадқиқотчи маълум бир сўз, ибора ёки конструкциянинг қўлланилиш ҳолати, даражаси, ким, қачон илк марта бу тузилмани қўллаганлиги, қандай услуб учун хосланишини билишга кўпроқ эҳтиёж сезади. Корпус эса мана шу каби муаммоларни ҳал қилишга йўналтирилгандир. Миллий корпус мавжуд тилнинг лексикаси ва грамматикасини ўрганиш учун зарур. Корпуснинг бошқа вазифаси эса кўрсатиб ўтилган соҳалар бўйича (лексика, грамматика, акцентология, тил тарихи соҳалари бўйича) тегишли маълумотларни етказиб беришдир. Миллий корпуси – нафақат тилшунослар, балки ўзбек тилидан фойдаланувчи барча кишилар: турли соҳа мутахассислари, олимлар, сиёсатчилар, луғат дизайнерлари, тадқиқотчилар ва турли мақсадларда ишлатилиши мумкин бўлган кенг қамровли универсал ахборот-қидирув тизими.

Таҳлиллар. Ўзбек тили миллий корпуси ўзбек тилида мавжуд бўлган лексик бирликларнинг, чунончи, ўзлашма сўзлар, синоним, антоним, омоним, сўзларнинг даражаланиши; сўзнинг морфологик таркиби, сўзнинг ясалиши, сўзларнинг маъноси, унинг морфологик хусусиятларини автоматик таҳлил қилиб бера олиши керак. Яъни корпусни тузиш, леммалаш, разметкалаш жараёнида корпус таркибига кирган шундай сўзларни бирма-бир қидирув асосида матнлар ичидан топиш ва уларни махсус изоҳлаш керак бўлади. Бунинг учун эса алгоритм, лингвистик молеллаштириш ишлари амалга оширилиш лозим.



Alisher Navoiy
1441-1501

ROMANLARI
HIKYOYALARI
PYESA
PORTRET
FELYETON

O'ZBEK TILI MILLIY KORPUSI

• Alisher Navoiy hayoti va ijodi
• O'zbek mumtoz adabiyoti

FOTO ALBOM
+ Buyuk siymolar

TAQDIMOT
+ Taqdimot

Korpus nima...

Korpus — til birliklarining xususiyatlarini aniqlash maqsadida qidiruv dasturiga bo'yisundirilgan matnlar majmui, tabiiy tildagi elektron shaklda saqlanadigan yozma yoki og'zaki, kompyuterlashtirilgan qidiruv tizimiga dasturiy ta'minot asosida joylashtirilgan on-line yoki off-line tizimda ishlaydigan matnlar jamlanmasi[1]. U elektron shaklda mavjud bo'lgan matnlarni, hujjatlarni ma'lumotlarni qayta ishlash, ularni avtomatik tahlil, ya'ni morfologik sintaktik va semantik tahlil qilish, morfologik analiz va sintez qilish bilan birga, yarim avtomatik rejimda ma'noni buzmasdan ishonchli nutq materialini moslashtirish darajasini tekshirish ushbu sohada tizil bilan qiziqishni oshirishga xizmat qiladi. Korpus - tilni tadqiq etish, til o'rganish, lug'at tuzishda eng zamonaviy, keng imkoniyatli dasturlashtirilgan tizimdir. Korpus - millionlab so'z kontekstlariga havola qiluvchi, maxsus qidiruv tizimi asosida ishlaydigan elektron matnlar yig'indisi. U internet tizimida elektron kutubxona, lug'at va lingvistik grammatikadir. Korpus - tabiiy (real) tilning elektron shaklda bo'lib, uni to'ldirish, tuzatish, tahrir qilish imkoniyatiga ega[2]. Korpus bu so'zlar, so'z birikmalari, grammatik shakllari ma'nosini ma'lum bir izlash tizimi orqali topishning elektron shakldagi matnlar to'plamidir. Korpuslarning har xil turlari mavjud. Masalan, bir muallif korpusi[3], bir kitob korpusi (jumladan dastlabki korpuslar "Bibliya" uchun qilingandir). Ma'lum bir til Milliy korpusi shu tilning xayotining barcha qirralarini, janrlarini, usullarini, xududiy va ijtimoiy variantlarini o'zida ifoda etadi. Masalan, o'zbek tili Milliy korpusini Internetda amalga qo'yilsa, 140 mln so'z ishlatmalaridan iborat shu tildagi barcha turdagi matnlarni o'zida mujassamlashtira oladi. Kelajakda o'zbek tili Milliy korpusi o'zida 300 mln so'z ishlatmalarini jamlashi lozim bo'ladi. Boshqa milliy tillar korpuslari kabi o'zbek tili Milliy korpusi ham ikkita muxim xususiyatga egadir. Birinchidan u ancha salobatli va barcha yo'nalishlar uchun (turli badiiy janrlar: publitsistik, o'quv, ilmiy, ish yuritish, nutk - so'zlashuv, shevalar va boshqalar uchun) bir xilda muvofiqlashtiradigan. Bu matnlar o'z davrida tildagi ta'sir kuchiga qarab proporsional ravishda korpusda jamlangan bo'ladi. Ikkinchidan korpus matnlarga xos bulgan alohida jihatlarni qo'shimcha ma'lumot sifatida o'zida aks ettiradi (masalan ma'lum bir jixatdan belgilanishlar va annotatsiyalar). So'zlar va boshqa birikmalar belgilanishi - korpusning bosh xarakteristikasidir (tasnifidir); U korpusni hozirda Internetni to'ldirib yuborgan boshqa matnlar kolleksiyalari va kutubxona matnlaridan farqlaydi. Matnlar belgilanishi qanchall boy va turli tuman bo'lsa korpusning ilmiy ahamiyati shuncha yuqori bo'ladi. O'zbek tili Milliy korpusini belgilanishlarning uch turi e'tiborga olinish kerak bo'ladi: metamatnli (u matnni muallif qarashi, janri va boshqa xususiyatlari bo'yicha to'liq ifodalaydi), morfologik va semantik belgilanishlar (morfologik va semantik belgilanishlar matni emas balki alohida bir so'zni tasniflab beradi). ADABIYOTLAR: 1. Grudeva V.A. Korpusnaya lingvistika. Uchebnoe posobie. 2-e izd., stereotip. - M.: Flinta, 2012. - 165 s. 2. Zakharov V.P. Korpusnaya lingvistika: Uchebno-metod. posobie. - SPb., 2005. - 48 s. 3. Zakharov V.P., Bogdanova S.Yu. Korpusnaya lingvistika. - Irkutsk: IGLU, 2011. - 161 c. 4. Toirova G.I. Importance of Interface in Creating Corpus. International Journal of Recent

SO'Z QIDIRISH

Hikoya nomi:
Kartina

Qidiruv

LEKSIK-GRAMMATIK QIDIRUV

☒ So'z
☐ Grammatik holati
☐ Omonim
☐ Qo'llanilishi
☐ Paronim
☐ Varianti
☐ Sinonimi
☐ Antonimi
☐ Uslubi
☐ Qo'llanilish davri

Qidiruv

SINTAKTIK QIDIRUV

☒ Gapning tuzilishiga ko'ra turlari
☐ Gapning maqsadiga ko'ra turlari

Qidiruv

MUALLIFLAR

Mualliflar

DASTUR HAQIDA

Ushbu korpus qidiruv natijasi shu hiquoyat matni asosida amalga oshiriladi. Korpus

144

Миллий корпуснинг интерфейси турли дизайн, тузилишга эга бўлиб, унинг мукамаллиги корпус яратувчи муаллифнинг зиммасига юкланади. Чунки интерфейс корпус ҳақида илк таассурот қолдирувчи, ўзига жалб этувчи умумий кўринишдир. Интерфейс яратишда миллий коллоритни акс эттирувчи безаклар ҳамда мумтоз ёки замонавийликни акс эттирувчи белгилар эътиборга олиниши лозим.

Демак, миллий корпусдан фойдаланишда унинг қулай ва энг самарали имкониятларининг намоён бўлишида интерфейснинг мукамал ҳамда тизимли равишда ишлаб чиқилганлиги муҳим саналади. Шу боис интерфейсни замонавий дастурий дизайн талабларига жавоб берадиган, фойдаланувга тушунарли, ишлаш учун қулай шаклда яратилади.

Хулоса. Корпус лингвистикаси - тилшуносликнинг энг ривожланган соҳаси, корпус эса тилшуносларнинг зарурий иш қуроли; оғзаки, ёзма ёдгорликлар, миллий-маданий меросни акс эттирувчи ахборот манбаидир. Корпус- кидирув дастурига бўйсундирилган матнлар йиғиндиси, мукамал разметкага эга корпус лингвистик тадқиқотлар самарадорлигини таъминлашда барқарор лингвистик база вазифасини бажаради. Лингвистик корпус сунъий интеллект маҳсули сифатида электрон луғат, таржима портали, терминологик маълумотлар банки, вертуал (электрон) кутубхона, электрон ҳукумат, электрон нашр, электрон дарслик ва қўлланмалар қаторида туради. Ўзбек тили миллий корпусининг умумий кўриниши бир неча ойналарга ҳамда ўнг ва чап устунчаларга ажратилган. Унда қуйидаги ойналар мавжуд бўлади: «Лексик кидирув», «Морфологик кидирув», «Синтактик кидирув». Ундан сўз ва гаплар бир неча сония вақт ичида автоматик таҳлил қилиб берилади.

Фойдаланилган адабиётлар:

1. Бритвин В.Г. Прикладное моделирование синтагматической семантики научно-технического текста (на примере автоматического индексирования). КД.- М.: МГУ, 1983.
2. Блумфилд Л. Язык. – М.: «Прогресс», 1968. – М., 1967.
3. Беляева Л.И., Чижаковский В.А. Тезаурус в системах автоматической переработки текста. – Кишинев, 1983.
4. Захаров В.П. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск, 2011. – 161 с.
5. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург. – 2006. 26 с.
6. Рыков В.В. Курс лекций по корпусной лингвистике. URL: <http://rykov-cl.narod.ru/c.html>; Плунгян В. Зачем мы делаем Национальный корпус русского языка? [Электрон ресурс] «Отечественные записки» 2005, –№2. http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html
7. Синклер Д. Предисловие к книге «Как использовать корпуса в преподавании иностранного языка»/ Д.Синклер [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/corpora-infro.html>, свободный;
8. Кутузов А.Б. Корпусная лингвистика. – [Электрон ресурс]: Лицензия Creative commons Attribution Share-Alike 3.0 Unported [Электрон ресурс] – [//lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf](http://lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf);
9. Котов Р.Г. Лингвистические аспекты автоматизированных систем управления. – Москва: Наука, 1977.;
10. Каримов Р. Ўзбек ва инглиз тили параллел корпусини тузишнинг лингвистик ва дастурий масалалари Филол.фан.бўйича фалсафа доктори (PhD)...дис. автореф. –Бухоро, 2021. – 151 б.
11. Мельчук И.А. Порядок слов при автоматическом синтезе русского слова (предварительные сообщения) // Научно –техническая информация. 1985, №12. – С.12-36.

12. Менглиев Б. Ўзбек тили корпуси яратиларми? “Маърифат” газетаси. 2018 йил, 3 апрель . http://marifat.uz/marifat/v_pomosh_uchitelu-marifat/savol/1142.htm
13. Менглиев Б., Бобожонов С., Ҳамроева Ш. Ўзбек тили миллий корпуси. 2018 йил, 26 апрель, <http://marifat.uz/marifat/ruknlar/fan/1241.htm>
14. Пўлатов, А. Қ. Компьютер лингвистикаси /Масъул муҳаррирлар: А.А.Абдуазизов, М.М.Орипов. – Т.: Akademnashr, 2011. – 520 б. –Б. 7.
15. Эшмуминов А. Ўзбек тили миллий корпусининг синоним сўзлар базаси. Филол.фан.бўйича фалсафа доктори (PhD)...дис.. – Тошкент, 2019.–Б.50-51.
16. Холиёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари. Филол.фан.бўйича фалсафа доктори (PhD)...дис. автореф. –Термиз, 2021. – 52 б.
17. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, 2018.
18. Bongers H. The history and principles of Vocabulary control. – Woerden: WOCOPI, 1947
19. Mohamed Zakaria Kurdi. Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax, Great Britain, USA: Wiley-ISTE, 2016, 300 p.
20. Toirova G. About the technological process of creating a national corps. // Foreign languages in Uzbekistan. Electronic scientific-methodical journal. - Tashkent. 2020, № 2 (31), –Б.57– 64. <https://journal.fledu.uz/uz/2-31-2020>
21. Toirova G. The importance of the interface in the creation of the corpus. International Scientific Journal «Internauka», // Mejdunarodnyy nauchnyy zhurnal «Internauka». - 2020. - №7. Online magazine. <https://doi.org/10.25313/2520-2057-2020-7-5944>
22. 13. Toirova G. The Role Of Setting In Linguistic Modeling. // International Multilingual Journal of Science and Technology. ISSN: 2528-9810 Vol. 4 Issue 9, September - 2019, - P.722-723 <http://imjst.org/index.php/vol-4-issue-9-september-2019/>
23. Charlez Meyer English corpus linguistics: An introduction. Cambridge University Press, 2004. 168 p.
24. Chomsky N., The logical basis for linguistic theory, Proc. IXth Int. Cong. of Linguists, 1962.
25. Leech G. The State of Art in Corpus Linguistics // English Corpus Linguistics / Aimer K., Altenberg K.(eds.) – London, 1991. – P. 8-29.
26. Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – L.,1969.