

✓

**O‘ZBEKISTON RESPUBLIKASI**  
**OLIV TA‘LIM, FAN VA INNOVATSIYALAR VAZIRLIGI**  
**TOSHKENT DAVLAT SHARQSHUNOSLIK UNIVERSITETI**



**KOMPYUTER LINGVISTIKASI**  
**FANINING O‘QUV DASTURI**

**Bilim sohasi:** 200 000 - San‘at va gumanitar fanlar  
**Ta‘lim sohasi:** 230 000 – Tillar  
**Ta‘lim yo‘nalishi:** 60230400 Kompyuter lingvistikasi

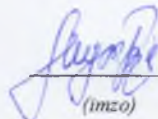
Sharq xalqlari tillari va adabiyoti  
instituti direktori:



X.V. Mirzaxmedova

(imzo)

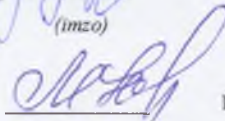
Tarjimashunoslik, tilshunoslik va  
xalqaro jurnalistika oliy maktabi boshlig'i:



S.T. Mustafayeva

(imzo)

Arm boshlig'i:



M. Yuldasheva

(imzo)

|                                   |  |   |                     |                                     |                               |                            |
|-----------------------------------|--|---|---------------------|-------------------------------------|-------------------------------|----------------------------|
| <b>Fan/modul kodi</b><br>KL 15-08 |  | <b>O'quv yili</b><br>2027/2028          | <b>Semestr</b><br>5 | <b>YeCTS - Kreditlar</b><br>8       |                               |                            |
| <b>Fan/modul turi</b><br>Majburiy |  | <b>Ta'lim tili</b><br>O'zbek            |                     | <b>Haftadagi dars soatlari</b><br>8 |                               |                            |
| 1.                                | <b>Fanning nomi</b>  | <b>Auditoriya mashg'ulotlari (soat)</b> | <b>Ma'ruza</b>      | <b>Amaliy</b>                       | <b>Mustaqil ta'lim (soat)</b> | <b>Jami yuklama (soat)</b> |
|                                   | Kompyuter lingvistikasi  | 96                                      | 48                  | 48                                  | 144                           | 240                        |
| 2.                                | <p align="center"><b>I. Fanning mazmuni:</b></p> <p><b>Fanning maqsadi:</b><br/>Kompyuter lingvistikasi fanining asosiy maqsadi — tabiiy tilni kompyuterlar yordamida qayta ishlash va tahlil qilishni o'rganish, shuningdek, inson tilining turli nuqtai nazarlarini kompyuterlar uchun tushunarli va ishlanadigan shaklda ifodalash imkoniyatlarini yaratishdir. Bu soha tilshunoslik va kompyuter fanlarining o'zaro aloqasini ta'minlab, kompyuterlarga tilni tushunish, qayta ishlash va o'rganish imkonini beradi.</p> <p><b>Fan vazifasi:</b><br/>Matnlarni morfologik, sintaktik va semantik tahlil qilish. So'zlarning shaklini aniqlash (lematizatsiya, stemming). Jumlar tuzilishini va grammatik tuzilmalarni aniqlash (sintaktik parsing). Tabiiy tilni kompyuter yordamida tushunish va unga javob berish tizimlarini ishlab chiqish. Ovozli yordamchilar, chatbotlar va suhbat robotlarini yaratish. Tabiiy tilni bir tilga tarjima qilishda kompyuter lingvistikasi metodlaridan foydalanish. Mashina o'qitish va sun'iy intellekt asosida tabiiy tilni qayta ishlashni o'rganish. Ma'lumotlarni tasniflash, klasterlash, sentiment tahlili kabi vazifalarni bajarish. Yirik matnlar bazasini (korpus) yaratish va ularni tahlil qilish orqali tilning umumiy qonuniyatlarini aniqlash. Tilshunoslik ma'lumotlarini to'plash va ularni kompyuterda ishlashga yaroqli shaklga keltirish. Matnning ma'nosini tushunish va uning semantik tuzilishini aniqlash. Semantik tahlil va yashirin semantik modellar (LSA) kabi metodlardan foydalanish.</p> <p>Mazkur fan dasturi xalqaro tan olingan reytinglarda birinchi top 300 talik ro'yxatga kiruvchi Berlin Erkin Universiteti (Freie Universität Berlin 104 THE) xorijiy tajribasini inobatga olgan holda takomillashtirildi.</p> <p><a href="https://www.fu-berlin.de/studium/studienangebot/grundstaendige/informatik_kombi_lo/index.html">https://www.fu-berlin.de/studium/studienangebot/grundstaendige/informatik_kombi_lo/index.html</a></p> |   |                     |                                     |                               |                            |

## II. Asosiy nazariy qism (ma'ruza mashg'ulotlari)

### II.1. Fan tarkibiga quyidagi mavzular kiradi:

#### 1-mavzu. Kompyuter lingvistikasi: Kirish va asosiy tushunchalar

fanining qisqacha izohida, bu mavzu tabiiy tilni kompyuterlar yordamida qayta ishlashning asoslarini o'rgatadi. Kompyuter lingvistikasi, tilshunoslik va kompyuter fanlari o'rtasidagi kesishgan soha bo'lib, tabiiy tilni tahlil qilish, ma'lumotlarni qayta ishlash, matnlarni tushunish va avtomatik tarjima qilish kabi amaliy sohalarida qo'llaniladi.

Tabiiy tilni qayta ishlash (NLP). So'z shakllarining tahlili va ularni lemmatizatsiya qilish. So'zlar va iboralar o'rtasidagi grammatik aloqalarni aniqlash. Matnning ma'nosini tushunish va semantik tahlil. Yirik matnlar to'plamlarini yig'ish, saqlash va tahlil qilish. kompyuter lingvistikasi sohasining muhim tamoyillari, asosiy metodlari va amaliy qo'llanilishi haqida umumiy tushuncha beradi.

#### 2- mavzu. Tabiiy tilni qayta ishlash (NLP) asoslari

Tabiiy tilni qayta ishlash (NLP) asoslari fani tabiiy tilni kompyuterlar yordamida tahlil qilish va qayta ishlashning asosiy tamoyillari va metodologiyasini o'rgatadi. NLP so'zi tabiiy tilni tushunish, qayta ishlash va yaratish jarayonlarini o'z ichiga oladi va kompyuter lingvistikasi, sun'iy intellekt va ma'lumotlar tahlili sohalarini birlashtiradi.

NLP asoslari quyidagi muhim tushunchalarni o'z ichiga oladi:

Tokenizatsiya. Lemmatizatsiya va Stemming. Sintaktik tahlil. Semantik tahlil Ma'lumotlarni tasniflash. Qayta ishlash va tarjima qilish.

Bu mavzu NLP ning asosi bo'lib, talabalarga tabiiy til bilan ishlashning zamonaviy metodlarini o'rgatish va ularni amaliyotda qo'llashni o'rganishni maqsad qiladi.

#### 3- mavzu. Morfologik tahlil Sintaktik tahlil: Grammatika va parsing

**Morfologik tahlil** fani tabiiy tilni qayta ishlashda (NLP) so'zlarning tuzilishini va ularning grammatik xususiyatlarini tahlil qilishni o'rgatadi. Morfologiya so'zlarning tarkibiy qismlarini, ya'ni ildiz, qo'shimchalar va oxirlarini aniqlashga qaratilgan. Morfologik tahlil yordamida til birligining turli shakllari, ularning ma'nolari va grammatik funksiyalari o'rganiladi.

Morfologik tahlil quyidagi asosiy tushunchalarni o'z ichiga oladi:

Lemmatizatsiya, stemming, morfemalar, shakl va funksiya.

**Sintaktik tahlil** (grammatika tahlili) va **parsing** — kompyuter lingvistikasi va tabiiy tilni qayta ishlashda matnning sintaktik tuzilishini aniqlash va tahlil qilish jarayonlaridir.

**Sintaktik tahlil** tilning grammatik qoidalariga asoslanib, gapdagi so'zlar va iboralarning tuzilishini o'rganadi. Bu jarayon grammatik xatolarni aniqlash va gapni tahlil qilishda yordam beradi.

**Parsing** — sintaktik tahlilning amaliyoti bo'lib, bu jarayonda matn sintaktik daraxtga (parse tree) aylantiriladi. Har bir tugun (node) biror sintaktik birlikni bildiradi, va bog'lanishlar (edges) o'zaro aloqani ko'rsatadi.

Ikkala jarayon ham tabiiy tilni tushunish, avtomatik tarjima, matn tahlili kabi sohalarida keng qo'llaniladi.

#### **4- mavzu. Semantik tahlil Morfologiya va lemmatizatsiya**

**Semantik tahlil** — bu tabiiy tilni qayta ishlashda so'zlar va jumlar ma'nosini tushunish va tahlil qilish jarayonidir. U tilning grammatik tuzilishini emas, balki so'zlar orasidagi ma'no aloqalarini aniqlashga qaratilgan. Semantik tahlilning asosiy maqsadi — so'zlarning to'g'ri ma'nosini aniqlash, noaniqliklarni bartaraf etish va so'zlar o'rtasidagi munosabatlarni tushunishdir. Bu jarayon avtomatik tarjima, matn tahlil qilish va savol-javob tizimlarida qo'llaniladi. Morfologiya va lemmatizatsiya — bu tilshunoslik va kompyuter lingvistikasi sohalarida so'zlarning tuzilishi, ularning qismlari va asosiy shakllariga qaytarilishi jarayonlarini o'rgatadigan fan.

Morfologiya so'zlarning ichki tuzilishini tahlil qiladi. Bu jarayonda so'zlarning ildizlari, qo'shimchalari, o'zgartirishlari va grammatik xususiyatlari o'rganiladi. Masalan, bir so'zning turli shakllari (son, shaxs, zamon) va ularning o'zaro aloqalari aniqlanadi. Lemmatizatsiya esa so'zlarni ularning asosiy shakliga keltirish jarayonidir. Bu jarayon orqali, masalan, "yurdi", "yurish" kabi so'zlar "yur" shakliga qaytariladi. Ushbu mavzu talabalarni tilshunoslik va kompyuter lingvistikasi sohalarida tilni tahlil qilish, avtomatik matn qayta ishlash va mashina o'rganish tizimlarida ishlashga tayyorlaydi.

#### **5- mavzu. So'zlararo aloqalar, semantik tarmoqlar va leksik ontologiyalar**

So'zlararo aloqalar va semantik tarmoqlar — bu tabiiy tilni qayta ishlash va kompyuter lingvistikasi sohalarida so'zlar orasidagi semantik munosabatlarni o'rganadigan fan. Ushbu mavzu talabalarni tilning semantik tuzilishini va so'zlar o'rtasidagi aloqalarni tushunishga o'rgatadi.

So'zlararo aloqalar so'zlarning ma'no jihatidan bir-biri bilan qanday bog'langanini o'rganadi. Bu jarayonda sinonimlar, antonimlar, giponimlar va hiperonimlar kabi so'zlar orasidagi munosabatlar tahlil qilinadi.

Semantik tarmoqlar so'zlar va ularning ma'nolarini bog'lab, tarmoq shaklida ifodalaydi. Har bir tugun (node) so'zning ma'nosini, bog'lanishlar esa ularning o'zaro aloqalarini ko'rsatadi. Bu semantik tarmoqlar, tilni tushunish va ma'lumotlarni qidirishda muhim vosita bo'lib xizmat qiladi.

Mavzu davomida talabalar so'zlararo aloqalarni tahlil qilish, semantik tarmoqlarni qurish va ularni turli tahlil va ilovalar uchun qo'llash ko'nikmalarini egallaydi.

#### **6- mavzu. Korpus lingvistikasi, uning ahamiyati va til korpuslarini yaratish hamda ulardan foydalanish**

Korpus lingvistikasi va uning ahamiyati — bu tilshunoslikning bir sohasi bo'lib, til ma'lumotlarini to'plab, tahlil qilish va o'rganishga qaratilgan. Korpus lingvistikasi, tilni empirik (tajribaga asoslangan) yondashuv bilan o'rganishni ta'minlaydi va tabiiy tilni qayta ishlashda muhim vosita hisoblanadi. Korpus lingvistikasi tilning real ishlatishdagi namunalarini (korpus) to'plash, tahlil qilish va o'rganish bilan shug'ullanadi. Korpus — bu tilning katta miqdordagi yozma yoki og'zaki materiallardan tashkil topgan ma'lumotlar bazasidir. U turli tillarda so'zlar, frazalar

va grammatik strukturalarning qanday ishlatilishini aniqlash imkonini beradi. Korpus lingvistikasi tilni yaxshiroq tushunish, yangi til modellari yaratish, grammatik qoidalar va soʻz boyligini oʻrganish, avtomatik tarjima va boshqa tabiiy tilni qayta ishlash tizimlarida ishlatiladi. U sunʼiy intellekt va mashina oʻrganish tizimlarining rivojlanishiga katta hissa qoʻshadi. Mavzu davomida talabalar tilning empirik tahlilini oʻrganib, korpusni yaratish va undan samarali foydalanish koʻnikmalarini egallaydi.

#### **7- mavzu. Maʼlumotlarni tasniflash, guruhlash va segmentatsiya qilish**

Maʼlumotlarni tasniflash — bu statistik va mashina oʻrganish usullaridan foydalanib, maʼlumotlarni belgilangan sinflarga ajratish jarayonidir. Ushbu fan talabalarni maʼlumotlarni toʻgʻri tahlil qilish, ularni mos kategoriyalarga ajratish va tasniflash uchun zarur boʻlgan texnikalar bilan tanishtiradi. Maʼlumotlarni tasniflash — bu maʼlumotlarni bir nechta sinf yoki kategoriya boʻyicha tasniflashni oʻz ichiga oladi. Masalan, elektron pochta xabarlarini spam yoki spam emas deb ajratish, yoki tibbiy tasvirlarni kasallik turlariga qarab tasniflash. Mavzuda mashina oʻrganishning asosiy algoritmlari, masalan, qaror daraxtlari, Naive Bayes, k-yaqin qoʻshni (k-NN), sunʼiy neyron tarmoqlari kabi usullar oʻrganiladi. Bu algoritmlar maʼlumotlarni sinflarga ajratish uchun ishlatiladi. Maʼlumotlarni tasniflash texnologiyalari turli sohalarda, jumladan, marketing, tibbiyot, moliya va tabiiy tilni qayta ishlashda keng qoʻllaniladi. Mavzu davomida talabalar tasniflash modellari va metodlarini amaliyotda qoʻllashni oʻrganadilar.

#### **8- mavzu. Tabiiy tilni avtomatik tarjima qilish**

Tabiiy tilni avtomatik tarjima qilish — bu kompyuter lingvistikasi va sunʼiy intellekt sohasidagi fan boʻlib, inson tilini avtomatik tarzda boshqa tilga tarjima qilishni oʻrganadi. Ushbu fan talabalarni tabiiy tilni qayta ishlashning asosiy texnologiyalari va usullari bilan tanishtiradi. Avtomatik tarjima — bu mashina yordamida bir tildagi matn yoki nutqni boshqa tildagi ekvivalentga aylantirish jarayonidir. Bu jarayonda tilning grammatik, sintaktik va semantik tuzilmalari hisobga olinadi. Darsda ikki asosiy avtomatik tarjima usuli oʻrganiladi:

Statistik tarjima (Statistical Machine Translation, SMT) — tilni statistik model va korpus asosida tarjima qilish.

Neural Machine Translation (NMT) — neyron tarmoqlari yordamida tarjima qilish, bu soʻnggi yillarda samarali natijalar berayotgan usul.

Tabiiy tilni avtomatik tarjima qilish texnologiyalari global kommunikatsiyani osonlashtiradi, xalqaro biznes va ilm-fan sohaslarida til toʻsiqlarini bartaraf etadi. Dars davomida talabalar tarjima tizimlarini yaratish, ularning ishlash prinsiplarini oʻrganish va amaliyotda qoʻllash koʻnikmalarini egallaydi.

#### **9- mavzu. Sentiment tahlili**

Sentiment tahlili — bu tabiiy tilni qayta ishlash va kompyuter lingvistikasi sohasida, matn yoki nutqdagi hissiy holat (ijtimoiy, pozitiv, negativ yoki neytral)ni aniqlash jarayonidir. Ushbu fan talabalarni matn orqali inson fikrini, kayfiyatini va hissiyotlarini avtomatik tarzda tahlil qilishga oʻrgatadi. Sentiment tahlili — matn yoki ijtimoiy media postlarini, sharhlarni va boshqa turdagi matnlarni tahlil qilish orqali

muallifning hissiy holatini aniqlashga yordam beradi. Masalan, mahsulot yoki xizmat haqida yozilgan ijobiy yoki salbiy fikrlarni aniqlash. Darsda sentiment tahlilini amalga oshirish uchun turli texnikalar, masalan, soʻzlar asosida tahlil qilish, mashina oʻrganish algoritmlari (k-nn, qaror daraxtlari) va neyron tarmoqlari yordamida tahlil qilish oʻrganiladi. Sentiment tahlili reklama, marketing, ijtimoiy tarmoqlar va mijozlar fikrini oʻrganishda keng qoʻllaniladi. Dars davomida talabalar matn tahlil qilish, hissiy holatlarni aniqlash va amaliyotda qoʻllash koʻnikmalarini egallaydi.

#### **10- mavzu. Statistika lingvistika**

Statistik lingvistika — bu tilshunoslik va statistik analizni birlashtirgan fan boʻlib, tilning strukturasi va xususiyatlarini statistik metodlar yordamida oʻrganishga qaratilgan. Ushbu fan talabalarni til maʼlumotlarini statistik tahlil qilish, til modellari yaratish va tilga oid empirik maʼlumotlardan foydalanishni oʻrgatadi. Statistik lingvistika tildagi soʻzlar, frazalar, grammatik tuzilmalar va boshqa til elementlari orasidagi statistik aloqalarni oʻrganadi. Masalan, soʻzlarning tez-tez takrorlanishi, soʻzlarning bir-biri bilan bogʻlanish ehtimolliklari va tilning strukturalarini tahlil qilish. Darsda turli statistik usullar, masalan, ehtimollar nazariyasi, regressiya tahlili, klasterlash va matn tasniflash kabi metodlar oʻrganiladi. Bu usullar yordamida tilni modellashtirish va tahlil qilish mumkin. Statistik lingvistika tabiiy tilni qayta ishlash, mashina oʻrganish, tilni oʻrganish va koʻplab amaliyotlarda (avtomatik tarjima, sentiment tahlili, soʻz tanish) qoʻllaniladi. Dars davomida talabalar tilni statistik tahlil qilish, model yaratish va amaliyotda foydalanish koʻnikmalarini egallaydi.

#### **11- mavzu. N-gramma modellar**

N-gramma modellar — bu tabiiy tilni qayta ishlashda soʻzlar yoki belgilar ketma-ketligini oʻrganish va tahlil qilish uchun ishlatiladigan statistik modellaridir. Ushbu fan talabalarni N-gramma modellarini yaratish va ularni tilni modellashtirish, tahlil qilish va bashorat qilishda qoʻllashga oʻrgatadi. N-gramma modeli — bu tilning soʻzlar yoki harflar ketma-ketliklarini statistika asosida tahlil qilishni taʼminlaydi. "N" soni ketma-ketlikdagi soʻzlar yoki belgilar sonini bildiradi (masalan, unigramma — bitta soʻz, bigramma — ikki soʻz, trigramma — uch soʻz va hokazo). Darsda N-gramma modellarini tilni bashorat qilish, matn tasniflash, soʻzlarni aniqlash, avtomatik tarjima va boshqa tabiiy tilni qayta ishlash vazifalarida qanday ishlatish oʻrganiladi. N-gramma modellar tilni qayta ishlash va mashina oʻrganish tizimlarida keng qoʻllaniladi, chunki ular tilning statistik xususiyatlarini oʻrganish va soʻzlar orasidagi aloqalarni aniqlashda samarali vositadir. Dars davomida talabalar N-gramma modellarini yaratish va amaliyotda qoʻllash koʻnikmalarini egallaydi.

#### **12- mavzu. Markov zanjirlari va ularning qoʻllanilishi**

Markov zanjirlari va ularning qoʻllanilishi — bu matematik model va algoritmlar yordamida tizimlar va jarayonlarning ehtimoliy xususiyatlarini oʻrganadigan fan. Markov zanjirlari, asosan, tizimlarning kelajak holati faqat hozirgi holatga bogʻliq boʻlganda ishlatiladi, yaʼni ularning xotirasi yoʻq (Markov xususiyati). Markov zanjiri — bu hodisalar yoki holatlar ketma-ketligidan iborat boʻlib, har bir keyingi holat faqat avvalgi holatga bogʻliq boʻladi. Ushbu model ehtimollar yordamida tizimning

kelajakdagi holatlarini bashorat qilishni ta'minlaydi. Markov zanjirlari tabiiy tilni qayta ishlashda, masalan, so'zlar yoki frazalarning ketma-ketligini bashorat qilishda, shuningdek, mashina o'rganish, genomika, iqtisodiyot, o'yin nazariyasi va boshqalarda keng qo'llaniladi. Markov zanjirlari tizimlar va jarayonlarning harakatini va rivojlanishini tushunishga yordam beradi. Dars davomida talabalar Markov zanjirlarini yaratish, ularni tahlil qilish va amaliyotda qo'llash ko'nikmalarini egallaydi.

### **13- mavzu. Fuzzy logic va tilshunoslikda qo'llanilish**

Fuzzy logic va tilshunoslikda qo'llanilishi — bu noaniqlik va aniqlikni hisobga olgan holda qarorlar qabul qilish va tizimlarni modellashtirish jarayonlarini o'rgatadigan fan. Fuzzy logic (noaniq mantiq) — bu an'anaviy mantiqdan farqli o'laroq, qiymatlarni faqat "ha" yoki "yo'q" bilan ifodalash o'rniga, ularni o'rtacha qiymatlar bilan ifodalashni ta'minlaydi. Fuzzy logic — bu tizimlar va jarayonlarni matematik jihatdan modellashtirishda noaniqliklarni hisobga olish uchun ishlatiladigan mantiqiy tizimdir. U klassik mantiqdagi aniq "ha" yoki "yo'q" qarorlar o'rniga, so'zlar va tushunchalar o'rtasidagi noaniqlikni (masalan, "issiq", "salqin", "yaxshi", "yomon") aniqlashga imkon beradi. Tilshunoslikda qo'llanilishi: Fuzzy logic tilshunoslikda semantik tahlil, so'zlarning noaniq ma'nolarini aniqlash, va tabiiy tilni qayta ishlashda noaniqliklarni hisobga olishda qo'llaniladi. Masalan, so'zlarni ma'no jihatidan tasniflashda yoki kontekstga asoslangan ma'no topishda fuzzy logic tizimlari yordam beradi. Fuzzy logic tilshunoslikda noaniqlikni hal qilish va tabiiy tilni qayta ishlashni yanada samarali qilish uchun muhim vosita bo'lib xizmat qiladi. Dars davomida talabalar fuzzy logic tizimlarini yaratish va ularni tilshunoslikda amaliyotda qo'llash ko'nikmalarini egallaydi.

### **14- mavzu. Ovozli yordamchilar va dialog tizimlari**

Ovozli yordamchilar va dialog tizimlari — bu tabiiy tilni qayta ishlash va sun'iy intellekt sohasiga oid fan bo'lib, kompyuterlar bilan insonlar o'rtasida ovozli muloqotni tashkil etish va avtomatik ravishda suhbatlashadigan tizimlarni yaratishni o'rgatadi. Ovozli yordamchilar — bu foydalanuvchining ovozli buyruqlariga javob beradigan va turli vazifalarni bajaradigan tizimlardir. Misollar sifatida Siri, Alexa, Google Assistant kabi tizimlar keltirilishi mumkin. Bu tizimlar foydalanuvchining talablarini tushunib, ma'lumot taqdim etish, xabarlar yuborish, musiqalar o'ynatish kabi vazifalarni bajaradi. Dialog tizimlari — bu foydalanuvchilar bilan ikki tomonlama muloqotni amalga oshiradigan tizimlardir. Ular ovozli yoki matnli suhbatni amalga oshirish, so'rovlarni tushunish va tegishli javoblarni berish orqali ishlaydi. Dialog tizimlari sun'iy intellekt va tabiiy tilni qayta ishlash usullaridan foydalanadi. Ovozli yordamchilar va dialog tizimlari kundalik hayotda, biznes, mijozlarga xizmat ko'rsatish, ta'lim va boshqa sohalarda keng qo'llaniladi. Dars davomida talabalar ovozli yordamchilarni yaratish, dialog tizimlarining ishlash prinsiplarini tushunish va ularni amaliyotda qo'llash ko'nikmalarini egallaydi.

### **15- mavzu. Mashina o'qitishning asoslari**

Mashina o'qitishning asoslari — bu sun'iy intellekt va kompyuter fanlari sohasida, kompyuterlarga ma'lumotlardan o'rganish va qarorlar qabul qilishni o'rgatadigan fan. Ushbu mavzu talabalarni mashina o'qitishning asosiy tushunchalari, algoritmlari va amaliyotlarini o'rganishga yo'naltiradi.

Mashina o'qitish — bu kompyuterlarga ma'lumotlar asosida o'zini-o'zi takomillashtirish va mustaqil qarorlar qabul qilish imkonini beradigan texnologiyadir. Bu jarayonda tizimlar turli algoritmlar yordamida o'rgatilib, yangi ma'lumotlarga asoslangan bashoratlar qilishga qodir bo'ladi. Dars davomida talabalar nazariy va amaliy jihatdan mashina o'qitishning turli metodlarini o'rganadilar. Mashina o'qitish texnologiyalari turli sohalarda, masalan, tibbiyot, moliya, sanoat, tabiiy tilni qayta ishlash va ko'plab boshqa sohalarda qo'llaniladi. Dars davomida talabalar mashina o'qitishning asosiy texnikalari va ularni turli vazifalarda qo'llash ko'nikmalarini egallaydi.

### **16- mavzu. Tabiiy tilni qayta ishlashda algoritmik yondashuvlar**

Tabiiy tilni qayta ishlashda algoritmik yondashuvlar — bu tabiiy tilni qayta ishlash (NLP) sohasidagi algoritmlar va ularning tilshunoslik muammolarini hal qilishdagi qo'llanilishini o'rganadigan fan. Ushbu mavzu talabalarni tabiiy tilni avtomatik tushunish va qayta ishlashda ishlatiladigan asosiy algoritmik usullar bilan tanishtiradi. Algoritmik yondashuvlar — bu tabiiy tilni qayta ishlashda tilni tahlil qilish, ma'lumotlarni chiqarish, matni tasniflash, tarjima qilish, so'zlar va frazalarning ma'nolarini tushunish kabi vazifalarni hal qilish uchun ishlatiladigan algoritmlar yig'indisidir. Ushbu algoritmlar kompyuterga tilni tushunish va insonlar bilan samarali muloqot qilish imkonini beradi.

Darsda quyidagi asosiy algoritmik yondashuvlar o'rganiladi:

Matni tahlil qilish va segmentatsiya (so'zlarni ajratish va tuzish).

Sintaktik tahlil (grammatik strukturalarni aniqlash).

Semantik tahlil (so'zlar va jumlar ma'nosini tushunish).

Mashina tarjimasi va sentiment tahlili kabi ilovalar uchun algoritmlar.

Tabiiy tilni qayta ishlashda algoritmik yondashuvlar avtomatik tarjima, matni tasniflash, sentiment tahlili va boshqa NLP vazifalarida keng qo'llaniladi. Dars davomida talabalar NLP muammolarini hal qilishda samarali algoritmlar va metodlarni qo'llash ko'nikmalarini egallaydi.

### **17- mavzu. Kengaytirilgan til modellari**

Kengaytirilgan til modellari — bu tabiiy tilni qayta ishlash (NLP) sohasidagi ilg'or yondashuvlarni o'rgatadigan fan bo'lib, tilning murakkab tuzilmalarini tahlil qilish va ishlash uchun zamonaviy modellarni qo'llashni o'rganadi. Ushbu mavzu talabalarni tilni avtomatik qayta ishlashda so'nggi yutuqlar va metodlar bilan tanishtiradi. Kengaytirilgan til modellari — bu tilni tushunish va ishlash chiqishdagi ilg'or yondashuvlar bo'lib, ularni sun'iy neyron tarmoqlari, transformatik modellar va boshqa chuqur o'rganish usullari orqali amalga oshirish mumkin. Bunday modellar tilning murakkab tuzilmalari, konteksti, semantikasini va sintaksisini aniqroq tushunishga imkon beradi.

Asosiy modellar va usullar:

Transformatorlar (masalan, BERT, GPT) — bu modellar tilning kontekstual tuzilishini o'rganishga imkon beradi va ko'plab NLP vazifalarida yuqori natijalar beradi.

Pre-trening va finetuning — modelni avvaldan o'rganish va maxsus vazifalarga moslashtirish jarayonlari.

Chuqur o'rganish va neyron tarmoqlari — tilni murakkab holatlarda tahlil qilish uchun ishlatiladigan ilg'or usullar.

Kengaytirilgan til modellari tabiiy tilni qayta ishlashda yanada aniq va samarali natijalarga erishish imkonini beradi. Dars davomida talabalar bu modellarni yaratish, o'rganish va amaliyotda qo'llash ko'nikmalarini egallaydi.

### **18- mavzu. Leksikal ma'lumotlarni qayta ishlash**

Leksikal ma'lumotlarni qayta ishlash — bu tabiiy tilni qayta ishlash (NLP) sohasidagi fan bo'lib, so'zlar va ularning ma'nolarini avtomatik tahlil qilish, strukturaviy va semantik jihatlarni aniqlash jarayonlarini o'rgatadi. Ushbu mavzu talabalarni leksikal ma'lumotlar bilan ishlash va ularni turli tilshunoslik vazifalarida samarali qo'llashga o'rgatadi. Semantik tahlil — so'zlarning ma'nolarini va ular orasidagi munosabatlarni aniqlash. So'zlarning vektorizatsiyasi — so'zlar yoki iboralarni matematik ko'rinishda ifodalash (masalan, Word2Vec, GloVe, FastText). Leksikal ma'lumotlarni qayta ishlash so'zlarni aniq va samarali tushunish va ishlashni ta'minlaydi. Bu texnologiyalar avtomatik tarjima, sentiment tahlili, so'z tanish, ma'lumotlarni chiqarish kabi turli NLP ilovalarida qo'llaniladi. Dars davomida talabalar leksikal ma'lumotlarni qayta ishlash usullarini o'rganib, ularni amaliyotda qo'llash ko'nikmalarini egallaydi.

### **19- mavzu. Chastota va Zipf qonuni**

Chastota va Zipf qonuni — bu tabiiy tilni qayta ishlash (NLP) va lingvistika sohasida so'zlar va belgilarning chastotasini o'rganish hamda Zipf qonunining amaliy qo'llanilishini tushuntiruvchi fan. Ushbu mavzu so'zlar va ularning tahlilini statistik asosda o'rganishga yo'naltirilgan bo'lib, tilni qayta ishlashda statistik metodlar va qonuniyatlarni qo'llashni o'rgatadi. Chastota — bu so'z yoki belgi qanday tez-tez uchrashini o'lchashni anglatadi. Tabiiy til matnlarida ba'zi so'zlar juda tez-tez, boshqalari esa kamroq uchraydi. So'zlarning chastotasini o'rganish, ularning tildagi roli va ahamiyatini tushunishda muhimdir. Zipf qonuni — bu statistik qonuniyat bo'lib, so'zlarning chastotasi va ularning rangida ko'rsatilgan o'rni o'rtasida muayyan tenglama mavjudligini ta'kidlaydi. Zipf qonuniga ko'ra, bir tilning eng ko'p ishlatiladigan so'zi boshqa so'zlarga nisbatan eng yuqori chastotaga ega bo'ladi va bu tartib sekin-asta kamayadi. Zipf qonuni va so'z chastotasi tilni tahlil qilishda, masalan, so'zlarni tasniflash, ma'lumotlarni chiqarish, klasterlash va til modellari yaratishda keng qo'llaniladi. Dars davomida talabalar so'z chastotasini o'rganish, Zipf qonunini tushunish va uni amaliyotda qo'llash ko'nikmalarini egallaydi.

## **20- mavzu. Matnlarni klasterlash va guruhlash**

Matnlarni klasterlash va guruhlash — bu tabiiy tilni qayta ishlash (NLP) sohasida matnlarni o'rganish va tasniflashga asoslangan fan bo'lib, ma'lumotlar orasidagi o'xshashliklar asosida ularni guruhlariga ajratish metodlarini o'rgatadi. Ushbu mavzu matnlarni avtomatik tahlil qilish va ma'lumotlarni samarali qayta ishlash uchun turli algoritmlarni qo'llashni o'rganadi. Matnlarni klasterlash — bu ma'lumotlar to'plamini (masalan, hujjatlar yoki xabarlar) o'xshashliklariga qarab guruhlash jarayonidir. Klasterlashda matnlar orasidagi o'xshashliklar va farqlar hisobga olinadi, va natijada bir-biriga o'xshash matnlar bir guruhga kiritiladi.

Guruhlash usullari: Matnlarni klasterlashda turli algoritmlar qo'llaniladi:

K-means klasterlash — eng keng tarqalgan klasterlash algoritmlaridan biri bo'lib, matnlarni K guruhga ajratadi.

Hierarchical clustering — matnlarni ierarxik tarzda guruhlash va turli darajalarda klasterlar yaratish.

DBSCAN — zichlikka asoslangan klasterlash algoritmi.

Ahamiyati: Matnlarni klasterlash va guruhlash texnologiyalari ma'lumotlarni tahlil qilish, hujjatlarni tasniflash, sentiment tahlili, tavsiya tizimlari yaratish va ko'plab boshqa NLP ilovalarida qo'llaniladi. Dars davomida talabalar matnlarni klasterlash va guruhlash algoritmlarini o'rganib, amaliyotda qo'llash ko'nikmalarini egallaydi.

## **21- mavzu. Avtomatik til identifikatsiyasi Semantik tarmoqlar va ontologiyalar**

"Avtomatik til identifikatsiyasi" (Avtomatik nutqni tanib olish) fanining maqsadi — nutqni avtomatik ravishda tanish va nutq egasini aniqlash texnologiyalari va metodlari bilan tanishtirishdir. Talabalar nutq signalini qayta ishlash va tahlil qilish asoslari, shuningdek, nutqni tanish va identifikatsiya qilish uchun ishlatiladigan algoritmlar va tizimlar bilan tanishadilar.

Mavzu davomida nutqni qayta ishlashning nazariy asoslari, nutqni tanish va identifikatsiya qilish uchun mashinani o'rganish metodlari, shuningdek, bu texnologiyalarning xavfsizlik tizimlari, sog'liqni saqlash, mijozlarga xizmat ko'rsatish va boshqa sohalarda qo'llanilishi kabi amaliy jihatlari ko'rib chiqiladi. Mavzu davomida talabalar semantik tarmoqlar va ontologiyalarni yaratish va ulardan foydalanishning nazariy va amaliy jihatlari o'rganadilar.

Semantik tarmoqlar — bu bilimlarni tasvirlash va ularga bog'liq munosabatlarni ifodalashda ishlatiladigan graf asosidagi modeldir. Ontologiyalar esa, ma'lumotlarni tasvirlash uchun bilimlar tizimi, ob'ektlar, ularning xususiyatlari va ular orasidagi munosabatlarni aniq belgilaydigan tuzilmalardir.

Mavzu da ontologiyalarni yaratish, ular bilan ishlash va semantik tarmoqlar yordamida ma'lumotlarni tahlil qilish va izlash metodlari o'rgatiladi. Bu texnologiyalar sun'iy intellekt, ma'lumotlar bazalari, ma'lumotlarni boshqarish va boshqa sohalarda keng qo'llaniladi.

## **22- mavzu. Suhbat robotlari va dialog tizimlari Ko'p tilli tizimlar**

"Suhbat robotlari va dialog tizimlari" mavzusi suhbat robotlari (chatbotlar) va interaktiv dialog tizimlarini yaratish, ularning ishlash tamoyillari hamda amaliy qo'llanilishini o'rganishga qaratilgan. Mavzu davomida talabalar tabiiy tilni qayta

ishlash (NLP), mashinani o'rganish, sun'iy intellekt va boshqa texnologiyalar yordamida suhbat robotlarini yaratish usullarini o'rganadilar.

Mavzuda dialog tizimlarining tuzilishi, foydalanuvchilar bilan muloqot qilish uchun algoritmlar, suhbat robotlarini trening qilish, ulardan ma'lumot olish va ularga javob berish usullari kabi mavzular ko'rib chiqiladi. Suhbat robotlari va dialog tizimlari mijozlarga xizmat ko'rsatish, elektron tijorat, ta'lim va boshqa ko'plab sohalarda keng qo'llaniladi. Mavzu davomida talabalar ko'p tilli tizimlar, ularning qurilishi, tilni qayta ishlash (NLP), tarjima, va semantik tahlil kabi sohalarda qo'llaniladigan metodlar bilan tanishadilar. Mavzuda ko'p tilli tizimlar uchun ma'lumotlarni qayta ishlash, tilni avtomatik tarjima qilish, til modeli yaratish va tillararo muloqotni samarali tashkil etish usullari ko'rib chiqiladi. Ushbu texnologiyalar xalqaro biznes, ta'lim, madaniyat, va boshqa sohalarda muhim ahamiyatga ega.

### **23- mavzu. Ovozli tilni qayta ishlash Tiling tahlilida sun'iy intellekt**

"Ovozli tilni qayta ishlash" mavzu nutq va ovozli ma'lumotlarni qayta ishlash texnologiyalarini o'rganishga qaratilgan. Mavzu davomida talabalar ovozli ma'lumotlarni tahlil qilish, ovozni aniqlash, nutqni matnga aylantirish (speech-to-text), matnni ovozga aylantirish (text-to-speech), nutqni tanish va boshqa ovozli tizimlar yaratish bo'yicha nazariy va amaliy bilimlarga ega bo'ladilar.

Mavzuda nutqni qayta ishlashning asosiy usullari, ovozli interfeyslar, sun'iy intellekt yordamida ovozli tizimlarni yaratish, shuningdek, ovozli tizimlarning amaliy qo'llanilishi, masalan, mijozlarga xizmat ko'rsatish, xavfsizlik va sog'liqni saqlash sohasidagi yondashuvlar ko'rib chiqiladi. Mavzu davomida talabalar tilni tahlil qilishda ishlatiladigan sun'iy intellekt metodlari, shu jumladan mashinani o'rganish, chuqur o'rganish (deep learning) va tabiiy tilni qayta ishlashning ilg'or usullari bilan tanishadilar.

Mavzuda tilni semantik, sintaktik va pragmatik tahlil qilish, til modellari yaratish, ma'lumotlarni avtomatik tahlil qilish, tarjima qilish va tilni avtomatik ravishda tushunish kabi mavzular ko'rib chiqiladi. Sun'iy intellekt yordamida tilni tahlil qilish texnologiyalari ilmiy izlanishlar, biznes, ta'lim, sog'liqni saqlash va boshqa sohalarda keng qo'llaniladi.

### **24- mavzu. Korpusni tahlil qilish va til modellari yaratish Matematik modellash va tilshunoslik**

"Korpusni tahlil qilish va til modellari yaratish" mavsuvi tilshunoslik va sun'iy intellektning integratsiyasini o'rganishga qaratilgan. Mavzu davomida talabalar katta hajmdagi matnlar (korpuslar)ni tahlil qilish, ularni o'rganish va til modellari yaratish metodlarini o'rganadilar.

Mavzuda tilni statistik va ma'lumotlarga asoslangan tahlil qilish, korpusdan olingan ma'lumotlar yordamida til modellari (masalan, so'z birikmalari, grammatika, sintaksis va semantika)ni yaratish, hamda tabiiy tilni qayta ishlashda (NLP) qo'llaniladigan metodlar, algoritmlar va texnologiyalar ko'rib chiqiladi. Korpusni tahlil qilish va til modellari yaratish mashinani o'rganish, tarjima tizimlari, avtomatik matn tahlili, va boshqa sohalarda keng qo'llaniladi. "Matematik modellash va tilshunoslik" mavsuvi tilshunoslik va matematik modellashning o'zaro aloqasini

o'rganishga qaratilgan. Mavzu davomida talabalar tilni matematik jihatdan tasvirlash, tilshunoslikdagi masalalarni matematik modellar yordamida yechish va tilni tahlil qilishda matematik yondashuvlarni qo'llashni o'rganadilar.

Mavzuda tilning sintaktik, semantik va pragmatik aspektlarini matematik modellar yordamida ifodalash, til modellari yaratish, algoritmlar va optimizatsiya usullarini qo'llash, shuningdek, tabiiy tilni qayta ishlash (NLP)da matematik metodlardan foydalanish kabi mavzular ko'rib chiqiladi. Bu mavzu sun'iy intellekt, kompyuter lingvistika, ma'lumotlar tahlili va boshqa sohalarda keng qo'llaniladi.

## II.2. Amaliy mashg'ulotlari bo'yicha ko'rsatma va tavsiyalar

*"Kompyuter lingvistikasi" fani bo'yicha amaliy mashg'ulotlar uchun quyidagi mavzular tavsiya etiladi:*

### 1-mavzu. Kompyuter lingvistikasi: Kirish va asosiy tushunchalar

fanining qisqacha izohida, bu mavzu tabiiy tilni kompyuterlar yordamida qayta ishlashning asoslarini o'rgatadi. Kompyuter lingvistikasi, tilshunoslik va kompyuter fanlari o'rtasidagi kesishgan soha bo'lib, tabiiy tilni tahlil qilish, ma'lumotlarni qayta ishlash, matnlarni tushunish va avtomatik tarjima qilish kabi amaliy sohalarda qo'llaniladi.

Tabiiy tilni qayta ishlash (NLP). So'z shakllarining tahlili va ularni lemmatizatsiya qilish. So'zlar va iboralar o'rtasidagi grammatik aloqalarni aniqlash. Matnning ma'nosini tushunish va semantik tahlil. Yirik matnlar to'plamlarini yig'ish, saqlash va tahlil qilish. kompyuter lingvistikasi sohasining muhim tamoyillari, asosiy metodlari va amaliy qo'llanilishi haqida umumiy tushuncha beradi.

### 2- mavzu. Tabiiy tilni qayta ishlash (NLP) asoslari

Tabiiy tilni qayta ishlash (NLP) asoslari fani tabiiy tilni kompyuterlar yordamida tahlil qilish va qayta ishlashning asosiy tamoyillari va metodologiyasini o'rgatadi. NLP so'zi tabiiy tilni tushunish, qayta ishlash va yaratish jarayonlarini o'z ichiga oladi va kompyuter lingvistikasi, sun'iy intellekt va ma'lumotlar tahlili sohasini birlashtiradi.

NLP asoslari quyidagi muhim tushunchalarni o'z ichiga oladi:

Tokenizatsiya. Lemmatizatsiya va Stemming. Sintaktik tahlil. Semantik tahlil. Ma'lumotlarni tasniflash. Qayta ishlash va tarjima qilish.

Bu mavzu NLP ning asosi bo'lib, talabalarga tabiiy til bilan ishlashning zamonaviy metodlarini o'rgatish va ularni amaliyotda qo'llashni o'rganishni maqsad qiladi.

### 3- mavzu. Morfologik tahlil Sintaktik tahlil: Grammatika va parsing

**Morfologik tahlil** fani tabiiy tilni qayta ishlashda (NLP) so'zlarning tuzilishini va ularning grammatik xususiyatlarini tahlil qilishni o'rgatadi. Morfologiya so'zlarning tarkibiy qismlarini, ya'ni ildiz, qo'shimchalar va oxirlarini aniqlashga qaratilgan. Morfologik tahlil yordamida til birligining turli shakllari, ularning ma'nolari va grammatik funksiyalari o'rganiladi.

Morfologik tahlil quyidagi asosiy tushunchalarni o'z ichiga oladi:

Lemmatizatsiya, stemming, morfemalar, shakl va funksiya.

**Sintaktik tahlil** (grammatika tahlili) va **parsing** — kompyuter lingvistikasi va tabiiy tilni qayta ishlashda matnning sintaktik tuzilishini aniqlash va tahlil qilish jarayonlaridir.

**Sintaktik tahlil** tilning grammatik qoidalariga asoslanib, gapdagi soʻzlar va iboralarning tuzilishini oʻrganadi. Bu jarayon grammatik xatolarni aniqlash va gapni tahlil qilishda yordam beradi.

**Parsing** — sintaktik tahlilning amaliyoti boʻlib, bu jarayonda matn sintaktik daraxtga (parse tree) aylantiriladi. Har bir tugun (node) biror sintaktik birlikni bildiradi, va bogʻlanishlar (edges) oʻzaro aloqani koʻrsatadi.

Ikkala jarayon ham tabiiy tilni tushunish, avtomatik tarjima, matn tahlili kabi sohalarda keng qoʻllaniladi.

#### **4- mavzu. Semantik tahlil Morfologiya va lemmatizatsiya**

**Semantik tahlil** — bu tabiiy tilni qayta ishlashda soʻzlar va jumlar maʼnosini tushunish va tahlil qilish jarayonidir. U tilning grammatik tuzilishini emas, balki soʻzlar orasidagi maʼno aloqalarini aniqlashga qaratilgan. Semantik tahlilning asosiy maqsadi — soʻzlarning toʻgʻri maʼnosini aniqlash, noaniqliklarni bartaraf etish va soʻzlar oʻrtasidagi munosabatlarni tushunishdir. Bu jarayon avtomatik tarjima, matn tahlil qilish va savol-javob tizimlarida qoʻllaniladi. Morfologiya va lemmatizatsiya — bu tilshunoslik va kompyuter lingvistikasi sohalarida soʻzlarning tuzilishi, ularning qismlari va asosiy shakllariga qaytarilishi jarayonlarini oʻrgatadigan fan.

Morfologiya soʻzlarning ichki tuzilishini tahlil qiladi. Bu jarayonda soʻzlarning ildizlari, qoʻshimchalari, oʻzgartirishlari va grammatik xususiyatlari oʻrganiladi. Masalan, bir soʻzning turli shakllari (son, shaxs, zamon) va ularning oʻzaro aloqalari aniqlanadi.

Lemmatizatsiya esa soʻzlarni ularning asosiy shakliga keltirish jarayonidir. Bu jarayon orqali, masalan, "yurdi", "yurish" kabi soʻzlar "yur" shakliga qaytariladi.

Ushbu fan talabalarni tilshunoslik va kompyuter lingvistikasi sohalarida tilni tahlil qilish, avtomatik matn qayta ishlash va mashina oʻrganish tizimlarida ishlashga tayyorlaydi.

#### **5- mavzu. Soʻzlararo aloqalar, semantik tarmoqlar va leksik ontologiyalar**

Soʻzlararo aloqalar va semantik tarmoqlar — bu tabiiy tilni qayta ishlash va kompyuter lingvistikasi sohalarida soʻzlar orasidagi semantik munosabatlarni oʻrganadigan fan. Ushbu mavzu talabalarni tilning semantik tuzilishini va soʻzlar oʻrtasidagi aloqalarni tushunishga oʻrgatadi.

Soʻzlararo aloqalar soʻzlarning maʼno jihatidan bir-biri bilan qanday bogʻlanganini oʻrganadi. Bu jarayonda sinonimlar, antonimlar, giponimlar va hiperonimlar kabi soʻzlar orasidagi munosabatlar tahlil qilinadi.

Semantik tarmoqlar soʻzlar va ularning maʼnolarini bogʻlab, tarmoq shaklida ifodalaydi. Har bir tugun (node) soʻzning maʼnosini, bogʻlanishlar esa ularning oʻzaro aloqalarini koʻrsatadi. Bu semantik tarmoqlar, tilni tushunish va maʼlumotlarni qidirishda muhim vosita boʻlib xizmat qiladi.

Mavzu davomida talabalar soʻzlararo aloqalarni tahlil qilish, semantik tarmoqlarni qurish va ularni turli tahlil va ilovalar uchun qoʻllash koʻnikmalarini egallaydi.

#### **6- mavzu. Korpus lingvistikasi, uning ahamiyati va til korpuslarini yaratish hamda ulardan foydalanish**

Korpus lingvistikasi va uning ahamiyati — bu tilshunoslikning bir sohasi bo‘lib, til ma'lumotlarini to‘plab, tahlil qilish va o‘rganishga qaratilgan. Korpus lingvistikasi, tilni empirik (tajribaga asoslangan) yondashuv bilan o‘rganishni ta'minlaydi va tabiiy tilni qayta ishlashda muhim vosita hisoblanadi. Korpus lingvistikasi tilning real ishlashdagi namunalarini (korpus) to‘plash, tahlil qilish va o‘rganish bilan shug‘ullanadi. Korpus — bu tilning katta miqdordagi yozma yoki og‘zaki materiallardan tashkil topgan ma'lumotlar bazasidir. U turli tillarda so‘zlar, frazalar va grammatik strukturalarning qanday ishlatilishini aniqlash imkonini beradi. Korpus lingvistikasi tilni yaxshiroq tushunish, yangi til modellari yaratish, grammatik qoidalar va so‘z boyligini o‘rganish, avtomatik tarjima va boshqa tabiiy tilni qayta ishlash tizimlarida ishlatiladi. U sun'iy intellekt va mashina o‘rganish tizimlarining rivojlanishiga katta hissa qo‘shadi. Mavzu davomida talabalar tilning empirik tahlilini o‘rganib, korpusni yaratish va undan samarali foydalanish ko‘nikmalarini egallaydi.

#### **7- mavzu. Ma'lumotlarni tasniflash, guruhlash va segmentatsiya qilish**

Ma'lumotlarni tasniflash — bu statistik va mashina o‘rganish usullaridan foydalanib, ma'lumotlarni belgilangan sinflarga ajratish jarayonidir. Ushbu fan talabalarni ma'lumotlarni to‘g‘ri tahlil qilish, ularni mos kategoriyalarga ajratish va tasniflash uchun zarur bo‘lgan texnikalar bilan tanishtiradi. Ma'lumotlarni tasniflash — bu ma'lumotlarni bir nechta sinf yoki kategoriya bo‘yicha tasniflashni o‘z ichiga oladi. Masalan, elektron pochta xabarlarini spam yoki spam emas deb ajratish, yoki tibbiy tasvirlarni kasallik turlariga qarab tasniflash. Mavzuda mashina o‘rganishning asosiy algoritmlari, masalan, qaror daraxtlari, Naive Bayes, k-yaqin qo‘shni (k-NN), sun'iy neyron tarmoqlari kabi usullar o‘rganiladi. Bu algoritmlar ma'lumotlarni sinflarga ajratish uchun ishlatiladi. Ma'lumotlarni tasniflash texnologiyalari turli sohalarda, jumladan, marketing, tibbiyot, moliya va tabiiy tilni qayta ishlashda keng qo‘llaniladi. Mavzu davomida talabalar tasniflash modellari va metodlarini amaliyotda qo‘llashni o‘rganadilar.

#### **8- mavzu. Tabiiy tilni avtomatik tarjima qilish**

Tabiiy tilni avtomatik tarjima qilish — bu kompyuter lingvistikasi va sun'iy intellekt sohasidagi fan bo‘lib, inson tilini avtomatik tarzda boshqa tilga tarjima qilishni o‘rganadi. Ushbu fan talabalarni tabiiy tilni qayta ishlashning asosiy texnologiyalari va usullari bilan tanishtiradi. Avtomatik tarjima — bu mashina yordamida bir tildagi matn yoki nutqni boshqa tildagi ekvivalentga aylantirish jarayonidir. Bu jarayonda tilning grammatik, sintaktik va semantik tuzilmalari hisobga olinadi. Darsda ikki asosiy avtomatik tarjima usuli o‘rganiladi:

Statistik tarjima (Statistical Machine Translation, SMT) — tilni statistik model va korpus asosida tarjima qilish.

Neural Machine Translation (NMT) — neyron tarmoqlari yordamida tarjima qilish, bu so‘nggi yillarda samarali natijalar berayotgan usul.

Tabiiy tilni avtomatik tarjima qilish texnologiyalari global kommunikatsiyani osonlashtiradi, xalqaro biznes va ilm-fan sohaslarida til to'vsiqlarini bartaraf etadi. Dars davomida talabalar tarjima tizimlarini yaratish, ularning ishlash prinsiplarini o'rganish va amaliyotda qo'llash ko'nikmalarini egallaydi.

#### **9- mavzu. Sentiment tahlili**

Sentiment tahlili — bu tabiiy tilni qayta ishlash va kompyuter lingvistikasi sohasida, matn yoki nutqdagi hissiy holat (ijtimoiy, pozitiv, negativ yoki neytral)ni aniqlash jarayonidir. Ushbu fan talabalarni matn orqali inson fikrini, kayfiyatini va hissiyotlarini avtomatik tarzda tahlil qilishga o'rgatadi. Sentiment tahlili — matn yoki ijtimoiy media postlarini, sharhlarni va boshqa turdagi matnlarni tahlil qilish orqali muallifning hissiy holatini aniqlashga yordam beradi. Masalan, mahsulot yoki xizmat haqida yozilgan ijobiy yoki salbiy fikrlarni aniqlash. Darsda sentiment tahlilini amalga oshirish uchun turli texnikalar, masalan, so'zlar asosida tahlil qilish, mashina o'rganish algoritmlari (k-nn, qaror daraxtlari) va neyron tarmoqlari yordamida tahlil qilish o'rganiladi. Sentiment tahlili reklama, marketing, ijtimoiy tarmoqlar va mijozlar fikrini o'rganishda keng qo'llaniladi. Dars davomida talabalar matn tahlil qilish, hissiy holatlarni aniqlash va amaliyotda qo'llash ko'nikmalarini egallaydi.

#### **10- mavzu. Statistik lingvistika**

Statistik lingvistika — bu tilshunoslik va statistik analizni birlashtirgan fan bo'lib, tilning strukturasi va xususiyatlarini statistik metodlar yordamida o'rganishga qaratilgan. Ushbu fan talabalarni til ma'lumotlarini statistik tahlil qilish, til modellari yaratish va tilga oid empirik ma'lumotlardan foydalanishni o'rgatadi. Statistik lingvistika tildagi so'zlar, frazalar, grammatik tuzilmalar va boshqa til elementlari orasidagi statistik aloqalarni o'rganadi. Masalan, so'zlarning tez-tez takrorlanishi, so'zlarning bir-biri bilan bog'lanish ehtimolliklari va tilning strukturalarini tahlil qilish. Darsda turli statistik usullar, masalan, ehtimollar nazariyasi, regressiya tahlili, klasterlash va matn tasniflash kabi metodlar o'rganiladi. Bu usullar yordamida tilni modellashtirish va tahlil qilish mumkin. Statistik lingvistika tabiiy tilni qayta ishlash, mashina o'rganish, tilni o'rganish va ko'plab amaliyotlarda (avtomatik tarjima, sentiment tahlili, so'z tanish) qo'llaniladi. Dars davomida talabalar tilni statistik tahlil qilish, model yaratish va amaliyotda foydalanish ko'nikmalarini egallaydi.

#### **11- mavzu. N-gramma modellar**

N-gramma modellar — bu tabiiy tilni qayta ishlashda so'zlar yoki belgilar ketma-ketligini o'rganish va tahlil qilish uchun ishlatiladigan statistik modellaridir. Ushbu fan talabalarni N-gramma modellarini yaratish va ularni tilni modellashtirish, tahlil qilish va bashorat qilishda qo'llashga o'rgatadi. N-gramma modeli — bu tilning so'zlar yoki harflar ketma-ketliklarini statistika asosida tahlil qilishni ta'minlaydi. "N" soni ketma-ketlikdagi so'zlar yoki belgilar sonini bildiradi (masalan, unigramma — bitta so'z, bigramma — ikki so'z, trigramma — uch so'z va hokazo). Darsda N-gramma modellarini tilni bashorat qilish, matn tasniflash, so'zlarni aniqlash, avtomatik tarjima va boshqa tabiiy tilni qayta ishlash vazifalarida qanday ishlatish o'rganiladi. N-gramma modellar tilni qayta ishlash va mashina o'rganish tizimlarida

keng qo'llaniladi, chunki ular tilning statistik xususiyatlarini o'rganish va so'zlar orasidagi aloqalarni aniqlashda samarali vositadir. Dars davomida talabalar N-gramma modellarini yaratish va amaliyotda qo'llash ko'nikmalarini egallaydi.

#### **12- mavzu. Markov zanjirlari va ularning qo'llanilishi**

Markov zanjirlari va ularning qo'llanilishi — bu matematik model va algoritmlar yordamida tizimlar va jarayonlarning ehtimoliy xususiyatlarini o'rganadigan fan. Markov zanjirlari, asosan, tizimlarning kelajak holati faqat hozirgi holatga bog'liq bo'lganda ishlatiladi, ya'ni ularning xotirasi yo'q (Markov xususiyati). Markov zanjiri — bu hodisalar yoki holatlar ketma-ketligidan iborat bo'lib, har bir keyingi holat faqat avvalgi holatga bog'liq bo'ladi. Ushbu model ehtimollar yordamida tizimning kelajakdagi holatlarini bashorat qilishni ta'minlaydi. Markov zanjirlari tabiiy tilni qayta ishlashda, masalan, so'zlar yoki frazalarning ketma-ketligini bashorat qilishda, shuningdek, mashina o'rganish, genomika, iqtisodiyot, o'yin nazariyasi va boshqalarda keng qo'llaniladi. Markov zanjirlari tizimlar va jarayonlarning harakatini va rivojlanishini tushunishga yordam beradi. Dars davomida talabalar Markov zanjirlarini yaratish, ularni tahlil qilish va amaliyotda qo'llash ko'nikmalarini egallaydi.

#### **13- mavzu. Fuzzy logic va tilshunoslikda qo'llanilish**

Fuzzy logic va tilshunoslikda qo'llanilishi — bu noaniqlik va aniqlikni hisobga olgan holda qarorlar qabul qilish va tizimlarni modellashtirish jarayonlarini o'rgatadigan fan. Fuzzy logic (noaniq mantiq) — bu an'anaviy mantiqdan farqli o'laroq, qiymatlarni faqat "ha" yoki "yo'q" bilan ifodalash o'rniga, ularni o'rtacha qiymatlar bilan ifodalashni ta'minlaydi. Fuzzy logic — bu tizimlar va jarayonlarni matematik jihatdan modellashtirishda noaniqliklarni hisobga olish uchun ishlatiladigan mantiqiy tizimdir. U klassik mantiqdagi aniq "ha" yoki "yo'q" qarorlar o'rniga, so'zlar va tushunchalar o'rtasidagi noaniqlikni (masalan, "issiq", "salqin", "yaxshi", "yomon") aniqlashga imkon beradi. Tilshunoslikda qo'llanilishi: Fuzzy logic tilshunoslikda semantik tahlil, so'zlarning noaniq ma'nolarini aniqlash, va tabiiy tilni qayta ishlashda noaniqliklarni hisobga olishda qo'llaniladi. Masalan, so'zlarni ma'no jihatidan tasniflashda yoki kontekstga asoslangan ma'no topishda fuzzy logic tizimlari yordam beradi. Fuzzy logic tilshunoslikda noaniqlikni hal qilish va tabiiy tilni qayta ishlashni yanada samarali qilish uchun muhim vosita bo'lib xizmat qiladi. Dars davomida talabalar fuzzy logic tizimlarini yaratish va ularni tilshunoslikda amaliyotda qo'llash ko'nikmalarini egallaydi.

#### **14- mavzu. Ovozli yordamchilar va dialog tizimlari**

Ovozli yordamchilar va dialog tizimlari — bu tabiiy tilni qayta ishlash va sun'iy intellekt sohasiga oid fan bo'lib, kompyuterlar bilan insonlar o'rtasida ovozli muloqotni tashkil etish va avtomatik ravishda suhbatlashadigan tizimlarni yaratishni o'rgatadi. Ovozli yordamchilar — bu foydalanuvchining ovozli buyruqlariga javob beradigan va turli vazifalarni bajaradigan tizimlardir. Misollar sifatida Siri, Alexa, Google Assistant kabi tizimlar keltirilishi mumkin. Bu tizimlar foydalanuvchining talablarini tushunib, ma'lumot taqdim etish, xabarlar yuborish, musiqalar o'ynatish

kabi vazifalarni bajaradi. Dialog tizimlari — bu foydalanuvchilar bilan ikki tomonlama muloqotni amalga oshiradigan tizimlardir. Ular ovozli yoki matnli suhbatni amalga oshirish, so'rovlarni tushunish va tegishli javoblarni berish orqali ishlaydi. Dialog tizimlari sun'iy intellekt va tabiiy tilni qayta ishlash usullaridan foydalanadi. Ovozli yordamchilar va dialog tizimlari kundalik hayotda, biznes, mijozlarga xizmat ko'rsatish, ta'lim va boshqa sohalarda keng qo'llaniladi. Dars davomida talabalar ovozli yordamchilarni yaratish, dialog tizimlarining ishlash prinsiplarini tushunish va ularni amaliyotda qo'llash ko'nikmalarini egallaydi.

#### **15- mavzu. Mashina o'qitishning asoslari**

Mashina o'qitishning asoslari — bu sun'iy intellekt va kompyuter fanlari sohasida, kompyuterlarga ma'lumotlardan o'rganish va qarorlar qabul qilishni o'rgatadigan fan. Ushbu mavzu talabalarni mashina o'qitishning asosiy tushunchalari, algoritmlari va amaliyotlarini o'rganishga yo'naltiradi.

Mashina o'qitish — bu kompyuterlarga ma'lumotlar asosida o'zini-o'zi takomillashtirish va mustaqil qarorlar qabul qilish imkonini beradigan texnologiyadir. Bu jarayonda tizimlar turli algoritmlar yordamida o'rgatilib, yangi ma'lumotlarga asoslangan bashoratlar qilishga qodir bo'ladi. Dars davomida talabalar nazariy va amaliy jihatdan mashina o'qitishning turli metodlarini o'rganadilar. Mashina o'qitish texnologiyalari turli sohalarda, masalan, tibbiyot, moliya, sanoat, tabiiy tilni qayta ishlash va ko'plab boshqa sohalarda qo'llaniladi. Dars davomida talabalar mashina o'qitishning asosiy texnikalari va ularni turli vazifalarda qo'llash ko'nikmalarini egallaydi.

#### **16- mavzu. Tabiiy tilni qayta ishlashda algoritmik yondashuvlar**

Tabiiy tilni qayta ishlashda algoritmik yondashuvlar — bu tabiiy tilni qayta ishlash (NLP) sohasidagi algoritmlar va ularning tilshunoslik muammolarini hal qilishdagi qo'llanilishini o'rganadigan fan. Ushbu mavzu talabalarni tabiiy tilni avtomatik tushunish va qayta ishlashda ishlatiladigan asosiy algoritmik usullar bilan tanishtiradi. Algoritmik yondashuvlar — bu tabiiy tilni qayta ishlashda tilni tahlil qilish, ma'lumotlarni chiqarish, matnни tasniflash, tarjima qilish, so'zlar va frazalarning ma'nolarini tushunish kabi vazifalarni hal qilish uchun ishlatiladigan algoritmlar yig'indisidir. Ushbu algoritmlar kompyuterga tilni tushunish va insonlar bilan samarali muloqot qilish imkonini beradi.

Darsda quyidagi asosiy algoritmik yondashuvlar o'rganiladi:

Matnни tahlil qilish va segmentatsiya (so'zlarni ajratish va tuzish).

Sintaktik tahlil (grammatik strukturalarni aniqlash).

Semantik tahlil (so'zlar va jumlarlar ma'nosini tushunish).

Mashina tarjimasi va sentiment tahlili kabi ilovalar uchun algoritmlar.

Tabiiy tilni qayta ishlashda algoritmik yondashuvlar avtomatik tarjima, matnни tasniflash, sentiment tahlili va boshqa NLP vazifalarida keng qo'llaniladi. Dars davomida talabalar NLP muammolarini hal qilishda samarali algoritmlar va metodlarni qo'llash ko'nikmalarini egallaydi.

#### **17- mavzu. Kengaytirilgan til modellari**

Kengaytirilgan til modellari — bu tabiiy tilni qayta ishlash (NLP) sohasidagi ilg'or yondashuvlarni o'rgatadigan fan bo'lib, tilning murakkab tuzilmalarini tahlil qilish va ishlash uchun zamonaviy modellarni qo'llashni o'rganadi. Ushbu mavzu talabalarni tilni avtomatik qayta ishlashda so'nggi yutuqlar va metodlar bilan tanishtiradi. Kengaytirilgan til modellari — bu tilni tushunish va ishlab chiqishdagi ilg'or yondashuvlar bo'lib, ularni sun'iy neyron tarmoqlari, transformatik modellar va boshqa chuqur o'rganish usullari orqali amalga oshirish mumkin. Bunday modellar tilning murakkab tuzilmalari, konteksti, semantikasini va sintaksisini aniqroq tushunishga imkon beradi.

Asosiy modellar va usullar:

Transformatorlar (masalan, BERT, GPT) — bu modellar tilning kontekstual tuzilishini o'rganishga imkon beradi va ko'plab NLP vazifalarida yuqori natijalar beradi.

Pre-trening va finetuning — modelni avvaldan o'rganish va maxsus vazifalarga moslashtirish jarayonlari.

Chuqur o'rganish va neyron tarmoqlari — tilni murakkab holatlarda tahlil qilish uchun ishlatiladigan ilg'or usullar.

Kengaytirilgan til modellari tabiiy tilni qayta ishlashda yanada aniq va samarali natijalarga erishish imkonini beradi. Dars davomida talabalar bu modellarni yaratish, o'rganish va amaliyotda qo'llash ko'nikmalarini egallaydi.

### **18- mavzu. Leksikal ma'lumotlarni qayta ishlash**

Leksikal ma'lumotlarni qayta ishlash — bu tabiiy tilni qayta ishlash (NLP) sohasidagi fan bo'lib, so'zlar va ularning ma'nolarini avtomatik tahlil qilish, strukturaviy va semantik jihatlarni aniqlash jarayonlarini o'rgatadi. Ushbu mavzu talabalarni leksikal ma'lumotlar bilan ishlash va ularni turli tilshunoslik vazifalarida samarali qo'llashga o'rgatadi. Semantik tahlil — so'zlarning ma'nolarini va ular orasidagi munosabatlarni aniqlash. So'zlarning vektorizatsiyasi — so'zlar yoki iboralarni matematik ko'rinishda ifodalash (masalan, Word2Vec, GloVe, FastText). Leksikal ma'lumotlarni qayta ishlash so'zlarni aniq va samarali tushunish va ishlashni ta'minlaydi. Bu texnologiyalari avtomatik tarjima, sentiment tahlili, so'z tanish, ma'lumotlarni chiqarish kabi turli NLP ilovalarida qo'llaniladi. Dars davomida talabalar leksikal ma'lumotlarni qayta ishlash usullarini o'rganib, ularni amaliyotda qo'llash ko'nikmalarini egallaydi.

### **19- mavzu. Chastota va Zipf qonuni**

Chastota va Zipf qonuni — bu tabiiy tilni qayta ishlash (NLP) va lingvistika sohasida so'zlar va belgilarning chastotasini o'rganish hamda Zipf qonunining amaliy qo'llanilishini tushuntiruvchi fan. Ushbu mavzu so'zlar va ularning tahlilini statistik asosda o'rganishga yo'naltirilgan bo'lib, tilni qayta ishlashda statistik metodlar va qonuniyatlarni qo'llashni o'rgatadi. Chastota — bu so'z yoki belgi qanday tez-tez uchrashini o'lchashni anglatadi. Tabiiy til matnlarida ba'zi so'zlar juda tez-tez, boshqalari esa kamroq uchraydi. So'zlarning chastotasini o'rganish, ularning tildagi roli va ahamiyatini tushunishda muhimdir. Zipf qonuni — bu statistik qonuniyat bo'lib, so'zlarning chastotasi va ularning rangida ko'rsatilgan o'rni o'rtasida

muayyan tenglama mavjudligini ta'kidlaydi. Zipf qonuniga ko'ra, bir tilning eng ko'p ishlatiladigan so'zi boshqa so'zlarga nisbatan eng yuqori chastotaga ega bo'ladi va bu tartib sekin-asta kamayadi. Zipf qonuni va so'z chastotasi tilni tahlil qilishda, masalan, so'zlarni tasniflash, ma'lumotlarni chiqarish, klasterlash va til modellari yaratishda keng qo'llaniladi. Dars davomida talabalar so'z chastotasini o'rganish, Zipf qonunini tushunish va uni amaliyotda qo'llash ko'nikmalarini egallaydi.

## **20- mavzu. Matnlarni klasterlash va guruhlash**

Matnlarni klasterlash va guruhlash — bu tabiiy tilni qayta ishlash (NLP) sohasida matnlarni o'rganish va tasniflashga asoslangan fan bo'lib, ma'lumotlar orasidagi o'xshashliklar asosida ularni guruhlarga ajratish metodlarini o'rgatadi. Ushbu mavzu matnlarni avtomatik tahlil qilish va ma'lumotlarni samarali qayta ishlash uchun turli algoritmlarni qo'llashni o'rganadi. Matnlarni klasterlash — bu ma'lumotlar to'plamini (masalan, hujjatlar yoki xabarlar) o'xshashliklariga qarab guruhlash jarayonidir. Klasterlashda matnlar orasidagi o'xshashliklar va farqlar hisobga olinadi, va natijada bir-biriga o'xshash matnlar bir guruhga kiritiladi.

Guruhlash usullari: Matnlarni klasterlashda turli algoritmlar qo'llaniladi:

K-means klasterlash — eng keng tarqalgan klasterlash algoritmlaridan biri bo'lib, matnlarni K guruhga ajratadi.

Hierarchical clustering — matnlarni ierarxik tarzda guruhlash va turli darajalarda klasterlar yaratish.

DBSCAN — zichlikka asoslangan klasterlash algoritmi.

Ahamiyati: Matnlarni klasterlash va guruhlash texnologiyalari ma'lumotlarni tahlil qilish, hujjatlarni tasniflash, sentiment tahlili, tavsiya tizimlari yaratish va ko'plab boshqa NLP ilovalarida qo'llaniladi. Dars davomida talabalar matnlarni klasterlash va guruhlash algoritmlarini o'rganib, amaliyotda qo'llash ko'nikmalarini egallaydi.

## **21- mavzu. Avtomatik til identifikatsiyasi Semantik tarmoqlar va ontologiyalar**

"Avtomatik til identifikatsiyasi" (Avtomatik nutqni tanib olish) fanining maqsadi — nutqni avtomatik ravishda tanish va nutq egasini aniqlash texnologiyalari va metodlari bilan tanishtirishdir. Talabalar nutq signalini qayta ishlash va tahlil qilish asoslari, shuningdek, nutqni tanish va identifikatsiya qilish uchun ishlatiladigan algoritmlar va tizimlar bilan tanishadilar.

Mavzu davomida nutqni qayta ishlashning nazariy asoslari, nutqni tanish va identifikatsiya qilish uchun mashinani o'rganish metodlari, shuningdek, bu texnologiyalarning xavfsizlik tizimlari, sog'liqni saqlash, mijozlarga xizmat ko'rsatish va boshqa sohalarda qo'llanilishi kabi amaliy jihatlari ko'rib chiqiladi. Mavzu davomida talabalar semantik tarmoqlar va ontologiyalarni yaratish va ulardan foydalanishning nazariy va amaliy jihatlari o'rganadilar.

Semantik tarmoqlar — bu bilimlarni tasvirlash va ularga bog'liq munosabatlarni ifodalashda ishlatiladigan graf asosidagi modeldir. Ontologiyalar esa, ma'lumotlarni tasvirlash uchun bilimlar tizimi, ob'ektlar, ularning xususiyatlari va ular orasidagi munosabatlarni aniq belgilaydigan tuzilmalardir.

Mavzu da ontologiyalarni yaratish, ular bilan ishlash va semantik tarmoqlar yordamida ma'lumotlarni tahlil qilish va izlash metodlari o'rgatiladi. Bu

texnologiyalar sun'iy intellekt, ma'lumotlar bazalari, ma'lumotlarni boshqarish va boshqa sohalarda keng qo'llaniladi.

## **22- mavzu. Suhbat robotlari va dialog tizimlari Ko'p tilli tizimlar**

"Suhbat robotlari va dialog tizimlari" mavzusi suhbat robotlari (chatbotlar) va interaktiv dialog tizimlarini yaratish, ularning ishlash tamoyillari hamda amaliy qo'llanilishini o'rganishga qaratilgan. Mavzu davomida talabalar tabiiy tilni qayta ishlash (NLP), mashinani o'rganish, sun'iy intellekt va boshqa texnologiyalar yordamida suhbat robotlarini yaratish usullarini o'rganadilar.

Mavzuda dialog tizimlarining tuzilishi, foydalanuvchilar bilan muloqot qilish uchun algoritmlar, suhbat robotlarini trening qilish, ulardan ma'lumot olish va ularga javob berish usullari kabi mavzular ko'rib chiqiladi. Suhbat robotlari va dialog tizimlari mijozlarga xizmat ko'rsatish, elektron tijorat, ta'lim va boshqa ko'plab sohalarda keng qo'llaniladi. Mavzu davomida talabalar ko'p tilli tizimlar, ularning qurilishi, tilni qayta ishlash (NLP), tarjima, va semantik tahlil kabi sohalarda qo'llaniladigan metodlar bilan tanishadilar. Mavzuda ko'p tilli tizimlar uchun ma'lumotlarni qayta ishlash, tilni avtomatik tarjima qilish, til modeli yaratish va tillararo muloqotni samarali tashkil etish usullari ko'rib chiqiladi. Ushbu texnologiyalar xalqaro biznes, ta'lim, madaniyat, va boshqa sohalarda muhim ahamiyatga ega.

## **23- mavzu. Ovozli tilni qayta ishlash Tiling tahlilida sun'iy intellekt**

"Ovozli tilni qayta ishlash" mavzu nutq va ovozli ma'lumotlarni qayta ishlash texnologiyalarini o'rganishga qaratilgan. Mavzu davomida talabalar ovozli ma'lumotlarni tahlil qilish, ovozni aniqlash, nutqni matnga aylantirish (speech-to-text), matnni ovozga aylantirish (text-to-speech), nutqni tanish va boshqa ovozli tizimlar yaratish bo'yicha nazariy va amaliy bilimlarga ega bo'ladilar.

Mavzuda nutqni qayta ishlashning asosiy usullari, ovozli interfeyslar, sun'iy intellekt yordamida ovozli tizimlarni yaratish, shuningdek, ovozli tizimlarning amaliy qo'llanilishi, masalan, mijozlarga xizmat ko'rsatish, xavfsizlik va sog'liqni saqlash sohalaridagi yondashuvlar ko'rib chiqiladi. Mavzu davomida talabalar tilni tahlil qilishda ishlatiladigan sun'iy intellekt metodlari, shu jumladan mashinani o'rganish, chuqur o'rganish (deep learning) va tabiiy tilni qayta ishlashning ilg'or usullari bilan tanishadilar.

Mavzuda tilni semantik, sintaktik va pragmatik tahlil qilish, til modellari yaratish, ma'lumotlarni avtomatik tahlil qilish, tarjima qilish va tilni avtomatik ravishda tushunish kabi mavzular ko'rib chiqiladi. Sun'iy intellekt yordamida tilni tahlil qilish texnologiyalari ilmiy izlanishlar, biznes, ta'lim, sog'liqni saqlash va boshqa sohalarda keng qo'llaniladi.

## **24- mavzu. Korpusni tahlil qilish va til modellari yaratish Matematik modellash va tilshunoslik**

"Korpusni tahlil qilish va til modellari yaratish" mavzusi tilshunoslik va sun'iy intellektning integratsiyasini o'rganishga qaratilgan. Mavzu davomida talabalar katta hajmdagi matnlar (korpuslar)ni tahlil qilish, ularni o'rganish va til modellari yaratish metodlarini o'rganadilar.

Mavzuda tilni statistik va ma'lumotlarga asoslangan tahlil qilish, korpusdan olingan ma'lumotlar yordamida til modellari (masalan, so'z birikmalari, grammatika, sintaksis va semantika)ni yaratish, hamda tabiiy tilni qayta ishlashda (NLP) qo'llaniladigan metodlar, algoritmlar va texnologiyalar ko'rib chiqiladi. Korpusni tahlil qilish va til modellari yaratish mashinani o'rganish, tarjima tizimlari, avtomatik matn tahlili, va boshqa sohalarda keng qo'llaniladi. "Matematik modellash va tilshunoslik" mavzusi tilshunoslik va matematik modellashning o'zaro aloqasini o'rganishga qaratilgan. Mavzu davomida talabalar tilni matematik jihatdan tasvirlash, tilshunoslikdagi masalalarni matematik modellar yordamida yechish va tilni tahlil qilishda matematik yondashuvlarni qo'llashni o'rganadilar.

Mavzuda tilning sintaktik, semantik va pragmatik aspektlarini matematik modellar yordamida ifodalash, til modellari yaratish, algoritmlar va optimizatsiya usullarini qo'llash, shuningdek, tabiiy tilni qayta ishlash (NLP)da matematik metodlardan foydalanish kabi mavzular ko'rib chiqiladi. Bu mavzu sun'iy intellekt, kompyuter lingvistikasi, ma'lumotlar tahlili va boshqa sohalarda keng qo'llaniladi.

### III. Mustaqil ta'lim\*

#### III.1. Mustaqil ta'lim va mustaqil ishlar

Talabalarni mustaqil ta'lim shaklini tashkil etishga qo'yilgan talablar O'zbekiston Respublikasi oliy ta'lim, fan va innovatsiyalar vazirligining 2024-yil 29-apreldagi 136-sonli "Oliy ta'lim muassasalari talabalari mustaqil ta'limini tashkil etish bo'yicha namunaviy tartibni tasdiqlash to'g'risida"gi buyrug'i asosida ishlab chiqilgan.

Mustaqil ta'limni baholash semestr davomida berilgan topshiriq asosida bajarilgan ishlarni HEMISda ilova qilish, shuningdek, oraliq va yakuniy test va savollarga javob berish asosida oshiriladi.

*"Kompyuter lingvistikasi"* fanidan mustaqil ta'lim uchun tavsiya etiladigan mavzular:

Tabiiy tilni avtomatik tarjima qilish. Sentiment tahlili. Statistik lingvistikasi  
N-gramma modellar Markov zanjirlari va ularning qo'llanilishi Fuzzy logic va tilshunoslikda qo'llanilish Ovozli yordamchilar va dialog tizimlari Mashina o'qitishning asoslari Tabiiy tilni qayta ishlashda algoritmik yondashuvlar Kengaytirilgan til modellari Leksikal ma'lumotlarni qayta ishlash Chastota va Zipf qonuni Matnlarni klasterlash va guruhlash Avtomatik til identifikatsiyasi Semantik tarmoqlar va ontologiyalar Suhbat robotlari va dialog tizimlari Ko'p tilli tizimlar Ovozli tilni qayta ishlash Tiling tahlilida sun'iy intellect Korpusni tahlil qilish va til modellari yaratish Matematik modellash va tilshunoslik

Talabalarni mustaqil ta'lim shaklini tashkil etishga qo'yilgan talablar O'zbekiston Respublikasi oliy ta'lim, fan va innovatsiyalar vazirligining 2024-yil 29-apreldagi 136-sonli "Oliy ta'lim muassasalari talabalari mustaqil ta'limini tashkil etish bo'yicha namunaviy tartibni tasdiqlash to'g'risida"gi buyrug'i asosida ishlab chiqilgan.

Mustaqil ta'limni baholash semestr davomida berilgan topshiriq asosida bajarilgan ishlarni HEMISda ilova qilish, shuningdek, oraliq va yakuniy test va savollarga javob berish asosida oshiriladi.

*\* Izoh: Mustaqil ta'lim ishi mavzulari fan doirasida o'zgarishi mumkin.*

**4. V. Fan o'qitilishining natijalari (shakllanadigan kompetensiyalar)**

**Tilni matematik jihatdan modellashtirish:** Talabalar tilshunoslik masalalarini matematik usullar yordamida tasvirlash va modellashtirishni o'rganadilar, tilning sintaktik va semantik strukturalarini matematik modellar asosida tahlil qilish imkoniyatiga ega bo'ladilar.

**Algoritmik fikrlash:** Talabalar matematik modellar yaratish va ularni algoritmik yondashuvlar yordamida yechish ko'nikmalarini rivojlantiradilar, shuningdek, tilni qayta ishlashda algoritmikni qo'llashni o'rganadilar.

**Kompyuter lingvistika va tabiiy tilni qayta ishlash:** Talabalar matematik modellardan foydalangan holda tabiiy tilni qayta ishlash, tilni avtomatik tahlil qilish, semantik va sintaktik strukturalarni aniqlash, til modellari yaratish kabi ko'nikmalarni egallaydilar.

**Matematik usullarni tilshunoslikka qo'llash:** Talabalar tilshunoslikda matematik yondashuvlarni qo'llash orqali tilni tushunish, tahlil qilish va o'rganishning yangi usullarini o'zlashtiradilar.

**Ma'lumotlar tahlili va optimizatsiya:** Talabalar matematik modellar va statistik metodlarni qo'llash orqali tilshunoslik ma'lumotlarini tahlil qilish va optimizatsiya qilish ko'nikmalariga ega bo'ladilar.

**Mashinani o'rganish va tilni qayta ishlashdagi amaliy qo'llanilishi:** Talabalar matematik modellarini mashinani o'rganish (machine learning) va tilni qayta ishlash (NLP) texnologiyalarida qo'llashni o'rganadilar.

**5. VI. TA'LIM TEXNOLOGIYALARI VA METODLARI:**

**Interaktiv ta'lim metodlari:**

Talabalar o'z bilimlarini amaliy mashqlar orqali rivojlantiradilar. Bu metodlar orqali talabalar mavzularni o'rganishda faollik ko'rsatadilar, muammolarni mustaqil ravishda hal qilishga undaydilar. Masalan, guruhli ishlash, munozaralar va interaktiv faoliyatlar orqali ta'lim olib boriladi.

**Masalalar va misollar yechish:**

O'qitish jarayonida onlayn resurslar, ta'lim platformalari (masalan, Moodle, Google

|    |  |
|----|--|
|    | <p>Classroom), ma'lumotlarni vizualizatsiya qilish vositalari (Matplotlib, Seaborn, Plotly) va tahlil vositalaridan foydalaniladi. Bu texnologiyalar talabalarni zamonaviy ilmiy tadqiqotlar va amaliy masalalar bilan tanishtiradi. O'qitish jarayonida talabalar mashina o'qitish va sun'iy intellekt metodlarini o'rganadilar, masalan, lingvistik ma'lumotlarni tasniflash va klasterlash, semantik tahlil va lingvistik model yaratishda. Talabalarni matematik modellar va sun'iy intellekt asosidagi tizimlar bilan tanishtiradi. O'quv jarayonida nazariy bilimlar amaliy masalalar orqali mustahkamlanadi. Talabalar matematik modellarni haqiqiy lingvistik ma'lumotlar ustida qo'llab, amaliy loyihalarni amalga oshiradilar. Bu usul talabalar uchun kengaytirilgan o'rganish imkoniyatini yaratadi.</p>   |
| 6. | <p style="text-align: center;"><b>VII. Kreditlarni olish uchun talablar:</b></p> <p>Topshiriqni muvaffaqiyatli bajarish uchun asosiy ko'nikmalar:</p> <p><b>Nazariy bilimlar:</b> Tabiiy tilni kompyuterlar yordamida qayta ishlashning asosiy tushunchalari va usullarini bilish. Matnni tahlil qilish, semantik va sintaktik strukturalarni aniqlash, leksik analiz va morfologik tahlil kabi asosiy NLP metodlarini tushunish.</p> <p><b>Amaliy ko'nikmalar:</b> Python, Java yoki boshqa dasturlash tillarida tabiiy tilni qayta ishlashga oid dasturlarni yozish. NLP algoritmlarini ishlab chiqish, masalan, so'zlarni ajratish (tokenization), so'zning ma'nosini aniqlash (lemmatization), sintaktik tahlil (parsing) va ma'lumotlarni qayta ishlash. Mashinani o'rganish algoritmlarini tabiiy tilni qayta ishlash uchun qo'llash (masalan, tasniflash, klasterlash). Chuqur o'rganish (deep learning) usullarini qo'llab, til modellari yaratish va ularni optimallashtirish. Matnni semantik tahlil qilish, ma'lumotlarni tushunish va tilga oid murakkab masalalarni hal qilish. Ma'lumotlar bazalarida tilni semantik jihatdan izlash va ularni analiz qilish.</p> <p><b>Muloqot ko'nikmalari:</b> O'z qarorlarini aniq va ravshan tushuntirish qobiliyati.</p> |
| 7. | <p style="text-align: center;"><b>VIII. ASOSIY VA QO'SHIMCHA O'QUV ADABIYOTLAR HAMDA AXBOROT MANBALARI</b></p> <p style="text-align: center;"><b>Asosiy adabiyotlar</b></p> <p>1. Pushpak Bhattacharyya Aditya Joshi Published by Wiley India Pvt. Ltd., 4436/7, Ansari Road, Daryaganj, New Delhi-110002. Printed at: Yash Printographics Edition: 2023 ISBN: 978-93-5746-238-9 ISBN: 978-93-5746-239-6 (ebk) Copyright © 2023 by Wiley India Pvt. Ltd.</p>   |

3 ✓ Zoran Gacovski Natural Language Processing Arcler Press 224 Shoreacres Road Burlington, ON L7L 2H2 Canada www.arclerpress.com Email: orders@arclereducation.com e-book Edition 2021 ISBN: 978-1-77407-979-9 (e-book)

3 ✓ Daniel Jurafsky. James H. Martin Stanford University University of Colorado at Boulder Copyright ©2023. All rights reserved. Draft of January 7, 2023

#### **Qo'shimcha adabiyotlar**

4. Mirziyoyev Sh.M. Hozirgi zamon va Yangi O'zbekiston. - Toshkent: O'zbekiston, 2024.
5. Mirziyoyev Sh.M. Yangi O'zbekistonda taraqqiyot strategiyasi asosida demokratik islohotlar yo'lini qat'iy davom ettiramiz. 6-jild. - Toshkent: O'zbekiston, 2023
6. Saidov D. Python dasturlash tili. O'quv - uslubiy qo'llanma, 2019
7. Barbara H.Partee Mathematical methods in linguistics Department of Linguistics University of Washington 1987 ISBN 90-277-2244-7
8. Григорьев Ю.Д., Мартыненко Г.Я. Последовательности типа Фибоначчи. Теория и прикладные аспекты Год издания:2024 ISBN 978-5-507-50228-8 516 с.
9. Николаев И. С., Митренина О. В., Ландо Т. М. (Ред.). Прикладная и компьютерная лингвистика. М.: URSS. 2016. 320 с. ISBN 978-5-9710-3472-8
10. Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов : учеб. пособие / Т. В. Батура ; Новосиб. гос. ун-т. – Новосибирск : РИЦ НГУ, 2016. – 166 с.
11. Nitin Indurkha, Fred J. Damerau Handbook of natural language processing 2010 by Taylor and Francis Group, LLC Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business International Standard Book Number-13: 978-1-4200-8593-8 (Ebook-PDF)

#### **Scopusdan maqolalar**

12. What Do Self-Supervised Speech Models Know About Words? Open Access Ankita Pasad, Chung-Ming Chien, Shane Settle, Karen Livescu Transactions of the Association for Computational Linguistics (2024) 12: 372–391. <https://doi.org/10.1162/tacl.a.00656>

[https://direct.mit.edu/tacl/article/doi/10.1162/tacl\\_a\\_00656/120586/What-Do-Self-Supervised-Speech-Models-Know-About](https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00656/120586/What-Do-Self-Supervised-Speech-Models-Know-About)

13. Scope Ambiguities in Large Language Models Gaurav Kamath, Sebastian Schuster, Sowmya Vajjala, Siva Reddy Transactions of the Association for Computational Linguistics (2024) 12: 738–754. [https://doi.org/10.1162/tacl\\_a\\_00670](https://doi.org/10.1162/tacl_a_00670)  
[https://direct.mit.edu/tacl/article/doi/10.1162/tacl\\_a\\_00670/121540/Scope-Ambiguities-in-Large-Language-Models](https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00670/121540/Scope-Ambiguities-in-Large-Language-Models)

**Axborot manbalar:**

<https://www.w3schools.com/>

<https://www.tutorialsteacher.com/>

<https://uzbekdevs.uz/>

<https://www.studytonight.com/>

8. Fanning o'quv dasturi Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabida ishlab chiqilgan va 2025 yil "18" 06 dagi 24-sonli bayonnomasi bilan ma'qullangan.

O'quv dasturi Toshkent davlat sharqshunoslik universiteti Kengashining 2025 yil "28" 06-4 sonli bayoni bilan tasdiqlangan.

9. **Fan/modul uchun ma'sullar:**

1. Allanyazov R.B.– TDSHU Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi o'qituvchisi

10. **Taqrizchilar:**

1. Usmanova Shoira Rustamovna- Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi professori, f.f.d.

2. N.Abdurahmonova - O'zMU, Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasida professori, f.f.d

Self