

**O‘ZBEKISTON RESPUBLIKASI
OLIIY TA’LIM, FAN VA INNOVATSIYALAR VAZIRLIGI**

**ALISHER NAVOIY NOMIDAGI
TOSHKENT DAVLAT O‘ZBEK TILI VA ADABIYOTI
UNIVERSITETI**



«TASDIQLAYMAN»

**Alisher Navoiy nomidagi Toshkent
davlat o‘zbek tili va adabiyoti
universiteti rektori
Sh. Sirojiddinov**

25-avgust 2026-yil

**10.00.11 – Amaliy va kompyuter lingvistikasi ixtisosligi
bo‘yicha tayanch doktorantura (PhD)ga kirish imtihoni**

D A S T U R I

Toshkent – 2026

Ushbu dastur Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kengashida ko'rib chiqildi va tasdiqlandi (2026-yil 25-avgustdagi 1-sonli bayonnoma).

Tuzuvchilar:

f.f.f.d. (PhD) B. Elov

f.f.d., professor Z. Xolmanova

f.f.d., professor Sh. Hamroyeva

KIRISH

O‘zbek tilining davlat tili sifatidagi mavqeyini mustahkamlash, uning raqamli makondagi qo‘llanish doirasini kengaytirish, milliy til resurslari va sun‘iy intellektga asoslangan lingvistik texnologiyalarni yaratish yuqori malakali ilmiy kadrlar tayyorlashni talab etadi. O‘zbekiston Respublikasi Prezidentining 2019-yil 21-oktabrdagi PF-5850-son va 2020-yil 20-oktabrdagi PF-6084-son farmonlarida belgilangan vazifalar o‘zbek tilining zamonaviy axborot-kommunikatsiya muhitidagi funksional imkoniyatlarini kengaytirish, ilmiy tadqiqotlar natijalarini amaliyotga joriy etish hamda til texnologiyalarini rivojlantirish uchun metodologik asos bo‘lib xizmat qiladi.

10.00.11 – Amaliy va kompyuter lingvistikasi ixtisosligi til nazariyasi yutuqlarini amaliy sohalarga tatbiq etish, tabiiy tillarni formallashtirish, modellashtirish va avtomatik qayta ishlash usullarini ishlab chiqishga qaratilgan fanlararo ilmiy yo‘nalishdir. Ixtisoslik kompyuter lingvistikasi, sun‘iy intellekt, ma’lumotlar bazalari, axborot texnologiyalari, mantiq, semiotika, kognitiv fanlar, kibernetika, raqamli gumanitar fanlar va zamonaviy axborot-kommunikatsiya texnologiyalari bilan uzviy aloqada rivojlanadi.

Mazkur kirish imtihoni dasturi talabgorlarning amaliy va kompyuter lingvistikasi nazariyasi hamda metodologiyasini egallaganlik darajasini, til va nutq birliklarini formal ifodalash, lingvistik resurslar yaratish, tabiiy tilni qayta ishlash tizimlarini loyihalash, mashinali va chuqur o‘qitish modellarini tanlash, ilmiy tajribani tashkil etish hamda natijalarni ishonchli baholash kompetensiyalarini aniqlashga mo‘ljallangan.

Dasturda ixtisoslik pasportining barcha asosiy yo‘nalishlari: formal va hisoblash modellashtirish, statistik va neyron til modellari, korpuslar va raqamli til resurslari, mashina tarjimasini, katta til modellari va generativ sun‘iy intellekt, axborot qidirish va bilimlar grafi, nutq va multimodal texnologiyalar, kompyuter leksikografiyasi, intellektual tahrir-nashr, ekspert va tavsiya tizimlari, raqamli ta’lim hamda inson–kompyuter muloqoti masalalari o‘zaro bog‘liq tizim sifatida qamrab olingan.

Talabgor savollarga javob berishda tushunchalarni sanab o‘tish bilan cheklanmasdan, muammoni ilmiy jihatdan formallashtirishi, mavjud yondashuvlarni qiyoslashi, ularning afzallik va cheklovlarini ko‘rsatishi, o‘zbek tili materialida amaliy misol keltirishi, ma’lumotlar va baholash mezonlarini asoslab berishi hamda tadqiqot natijalarining takrorlanuvchanligi, etikligi va amaliy qiymatini baholay olishi lozim.

ASOSIY QISM

1. Amaliy va kompyuter lingvistikasining nazariy-metodologik asoslari

Amaliy lingvistika va kompyuter lingvistikasining obyekti, predmeti, maqsadi va vazifalari. Nazariy tilshunoslik natijalarini amaliy masalalarga tatbiq etish shakllari. Kompyuter lingvistikasida tilni ma'lumot, qoidalar, belgilar tizimi, ehtimollik taqsimoti va hisoblash jarayoni sifatida talqin qilish.

Qoidalariga asoslangan, statistik, korpusga asoslangan, mashinali o'qitish, neyron, katta til modellari va gibril paradigmlar. Paradigmlar almashinuvi, ularning epistemologik asoslari, tushuntiriluvchanlik va umumlashtirish masalalari.

Amaliy va kompyuter lingvistikasining informatika, sun'iy intellekt, ma'lumotlar ilmi, kognitiv fanlar, semiotika, mantiq, kibernetika, psixolingvistika, neyrolingvistika, sotsiolingvistika, raqamli gumanitar fanlar va inson-kompyuter muloqoti bilan aloqalari.

Til bilimini hisoblash tizimiga ko'chirish: lingvistik birlik, belgi, xususiyat, munosabat, kontekst va diskursning formal talqini. Lingvistik nazariya va muhandislik yechimlari o'rtasidagi muvozanat. Inson kognitiv va ijodiy nutqiy faoliyatini modellashtirishning imkoniyatlari va chegaralari.

O'zbek tili uchun amaliy va kompyuter lingvistikasi muammolari: agglyutinativ morfologiya, erkinroq so'z tartibi, yozuv va apostrof variantlari, lotin-kirill parallel qo'llanishi, dialektal farqlar, kod almashinuvi, terminologik barqarorlik, kam resurslilik va domen siljishi.

2. Til birliklarini formallashtirish, matematik va hisoblash modellashtirish

Formalizatsiya tushunchasi. Tabiiy til hodisasidan formal obyektga o'tish bosqichlari: kuzatuv, birliklarni ajratish, xususiyatlar tizimi, cheklovlar, qoidalar, model, algoritim va hisoblash tajribasi. Formal modelning adekvatligi, to'liqligi, izchilligi va hisoblanish imkoniyati.

Alfavit, satr, til, grammatika va avtomat tushunchalari. Muntazam ifodalar va chekli avtomatlar; kontekstsiz grammatikalar; xususiyatli grammatikalar; unifikatsiya; bog'liqlik grammatikasi; mantiqiy va semantik tasvirlar. Graf, daraxt, jadval va vektor ko'rinishidagi lingvistik modellar.

Algoritmik asoslar: ma'lumotlar tuzilmalari, qidirish, saralash, indekslash, dinamik dasturlash, graf algoritmlari, taxminiy qidirish va hisoblash murakkabligi. Vaqt va xotira murakkabligining real NLP tizimlariga ta'siri. Unicode, kodlash, yozuv tizimlari va satrlarni xavfsiz qayta ishlash.

Matematik asoslar: vektor va matritsalar, masofa va o'xshashlik, ehtimollar nazariyasi, shartli ehtimol, Bayes qoidasi, maksimal ehtimollik va maksimal aposterior baholash, entropiya, kross-entropiya, Kullback-Leibler tafovuti, optimallashtirish, gradient va regulashtirish.

Lingvistik va statistik-kombinator modellashtirish: chastota taqsimotlari, Zipf va Heaps qonunlari, kollokatsiya ko'rsatkichlari, n-gramlar, Markov taxmini, tekislash usullari, ehtimollik til modellari va perplexity. Kvantitativ lingvistika, statistik semantika va stilistika metodlari.

Eksperimental statistika: tanlanma va bosh to‘plam, dispersiya, korrelyatsiya, ishonch oralig‘i, parametrik va noparametrik testlar, bootstrap va randomizatsiya, ko‘p martalik taqqoslash muammosi. Miqdoriy natijani lingvistik talqin bilan bog‘lash.

3. Korpus lingvistikasi, raqamli til resurslari va ma’lumotlar bazalari

Korpus tushunchasi, uning representativligi, muvozanati, hajmi, janr-domen qamrovi va maqsadga mosligi. Umumiy, maxsus, milliy, mualliflik, parallel, qiyosiy, tarixiy, dinamik, o‘quv, nutqiy va multimodal korpuslar. Korpus dizayni va tanlanma strategiyasi.

Ma’lumotlarning hayotiy sikli: yig‘ish, huquqiy ruxsat, kelib chiqishni qayd etish, tozalash, deduplikatsiya, normalizatsiya, segmentatsiya, annotatsiya, verifikatsiya, versiyalash, nashr etish va qayta foydalanish. Metama’lumotlar, provenance, litsenziya, maxfiylik va saqlash siyosati.

Annotatsiya darajalari: grafik-ortografik, gap va token, lemma, so‘z turkumi va morfologik xususiyatlar, morfemik, sintaktik, semantik, pragmatik, diskurs, koreferensiya, hissiylik, termin va multimodal annotatsiya. BIO/BIOES, CoNLL-U, XML/TEI, JSON/JSONL va boshqa formatlar.

Annotatsiya metodologiyasi: tagset va yo‘riqnoma yaratish, pilot tajriba, ko‘r annotatsiya, ikki karra belgilash, annotatorlararo kelishuv, Cohen kappa, Fleiss kappa, Krippendorff alpha, adjudikatsiya, gold standart va audit. Avtomatik, yarim avtomatik va faol o‘qitishga asoslangan annotatsiya.

Lingvistik leksikografiya va terminografiya. Elektron lug‘at maqolasi, morfologik va semantik ma’lumotlar modeli, tezaurus, WordNet, terminologik baza, kollokatsiyalar bazasi, freym va ontologiya. Semantik tarmoqlar hamda bilimlar grafida tushuncha va munosabatlarni ifodalash.

Relyatsion, hujjatga yo‘naltirilgan, graf va vektorli ma’lumotlar bazalari. Indeksar, to‘liq matnli qidiruv, API, versiyalash, zaxiralash va kirish huquqlari. Avtomatlashtirilgan axborot tizimlari va ma’lumotlar bazalarining lingvistik ta’minoti.

O‘zbek tili milliy korpusi va xususiy korpuslar uchun arxitektura, qidiruv, konkordans, chastota, kollokatsiya, morfologik-sintaktik filtrlash va vizual tahlil imkoniyatlari. Raqamli gumanitar tadqiqotlarda korpus, stilometriya, mualliflikni aniqlash va tarixiy matnlarni qayta ishlash.

4. Matnni tayyorlash, tokenizatsiya va morfologik qayta ishlash

Matnni olish va tozalash: HTML/PDF/OCR manbalari, belgilarni kodlash, Unicode normalizatsiyasi, ko‘rinmas belgilar, bo‘shliq, satr uzilishi, shovqin va texnik segmentlarni aniqlash. Original va normallashtirilgan shaklni alohida saqlash prinsipi.

O‘zbek yozuvini normallashtirish: o‘ va g‘ harflari, apostrofning turli kod nuqtalari, lotin-kirill transliteratsiyasi, aralash yozuv, bosh-kichik harf, raqam, sana, qisqartma, URL, elektron manzil, hashtag va emoji. Normalizatsiyada ma’no va manba izini saqlash.

Gap chegarasini aniqlash va tokenizatsiya. Soʻz, soʻzshakl, token, grafema va ortografik harf tushunchalari. Klitiklar, qoʻshma birliklar, defisli shakllar, qisqartmalar va punktuatsiya. Oq boʻshliq, qoidaviy, statistik va neyron tokenizatorlar.

Subsoʻz segmentatsiyasi: byte-pair encoding, WordPiece, Unigram va SentencePiece. Lugʻat hajmi, kam uchraydigan birliklar, morfologik yaxlitlik, OOV va model samaradorligi oʻrtasidagi muvozanat. Oʻzbek agglyutinatív tuzilishida subsoʻz chegaralarini baholash.

Stemming, lemmatizatsiya, morfemik segmentatsiya, morfologik tahlil va sintez. Lugʻat va qoida, chekli holatli transduser, paradigmatic model, statistik va neyron analizatorlar. Soʻzshakl–lemma–morfologik xususiyatlar oʻrtasidagi bogʻlanish.

Soʻz turkumlarini va morfologik xususiyatlarni aniqlash. Universal va milliy tagsetlar, feature engineering, ketma-ketlikni belgilash, HMM, CRF, BiLSTM-CRF va Transformer modellar. Morfologik omonimiya, kontekstual disambiguatsiya va xatoliklar tahlili.

Baholash: token chegaralari uchun precision, recall va F1; lemmatizatsiya aniqligi; POS va FEATS uchun accuracy va macro-F1; toifalar kesimidagi natijalar, chalkashlik matritsasi, OOV va domenlararo barqarorlik.

5. Sintaktik, semantik, pragmatik va diskurs tahlili

Sintaktik tahlilning asosiy birliklari va modellari. Bevosita tashkil etuvchilar va bogʻliqlik sintaksisi. Universal Dependencies tamoyillari, sintaktik daraxt, bosh soʻz, bogʻlanish turi, proyektivlik va nonproyektivlik. Qoidaviy, oʻtishlarga asoslangan va grafaga asoslangan parserlar.

Sintaktik noaniqlik, ellipsis, koordinatsiya, qoʻshma gap, erkin soʻz tartibi va punktuatsiyaning tahlilga taʼsiri. Daraxtbank yaratish, annotatsiya kelishuvi, unlabeled va labeled attachment score, gap uzunligi va konstruksiya boʻyicha xatoliklar tahlili.

Leksik va kompozitsion semantika. Koʻp maʼnolilik, omonimiya, soʻz maʼnosini kontekstda aniqlash, semantik rollar, predikat-argument tuzilishi, valentlik, semantik freymlar, hodisa va munosabatlarni ajratib olish.

Nomlangan obyektlarni aniqlash, entitetni bilim bazasiga bogʻlash, munosabat va hodisalarni ajratib olish. Yassi, ichma-ich va kesishuvchi entitetlar. Lugʻat, qoida, klassik mashinali oʻqitish va neyron yondashuvlar. Entitet va munosabat darajasidagi baholash.

Koʻp soʻzli birliklar, terminlar, kollokatsiyalar va frazeologik birliklarni aniqlash. Statistik assotsiatsiya, sintaktik qolip, kontekstual tasvir va korpus dalillari. Leksik resurslar bilan integratsiya.

Anafora va koreferensiya. Eslatma birliklari, antecedent, zanjir, salience, sintaktik va semantik cheklovlar. Qoidaviy, klasterlash va end-to-end neyron modellar. Mention va cluster darajasidagi baholash.

Pragmatik tahlil: nutqiy akt, kommunikativ maqsad, deysis, presuppozitsiya, implikatura, relevantlik, konnotatsiya va kontekst. Pragmatik mazmunni avtomatik aniqlashdagi ma'lumot va baholash muammolari.

Diskurs va muloqot tahlili: koherensiya, kohesiya, ritorik munosabatlar, dialog aktlari, navbat almashinuvi, muloqot holati va foydalanuvchi niyati. Hissiy tahlil, aspektga asoslangan sentiment, emotsiya, kinoya, sarkazm, toksik va zararli mazmunni aniqlash.

6. Statistik mashinali o'qitish va neyron modellar

Matnni xususiyatlar orqali ifodalash: bag-of-words, n-gramlar, chastota, TF-IDF, belgilar va morfemalar, sintaktik va lug'aviy xususiyatlar. Xususiyat tanlash, o'lchamni kamaytirish va siyrak vektorlar.

Nazoratli o'qitish: Naive Bayes, logistik regressiya, tayanch vektorlar mashinasi, qaror daraxtlari va ansambl usullar. Nazoratsiz va yarim nazoratli o'qitish: klasterlash, latent mavzu modellari, self-training va co-training. Sinflar nomutanosibli va kalibrlash.

Ketma-ketlik modellar: n-gram til modeli, yashirin Markov modeli, maksimal entropiya va shartli tasodifiy maydon. Dekodlash, Viterbi algoritmi va strukturaviy bashorat.

So'zlarning taqsimot semantikasi va vektorli tasvirlari: Word2Vec, GloVe, FastText. O'xshashlik, analogiya, subso'z ma'lumoti, domen va tarafkashlik. Kontekstual tasvirlar: ELMo, BERT oilasi va ko'p tilli modellar.

Sun'iy neyron tarmoqlar: perceptron, MLP, embedding qatlami, aktivatsiya, yo'qotish funksiyasi, backpropagation, gradientga asoslangan optimallashtirish, learning rate, batch, epoch, regularlashtirish va erta to'xtatish.

CNN, RNN, LSTM, GRU, ikki yo'nalishli va gibrid modellar. Encoder-decoder, teacher forcing, attention va beam search. Ketma-ketlikni tasniflash, belgilash va generatsiya vazifalariga tatbiq.

Transformer arxitekturasi: self-attention, multi-head attention, pozitsion kodlash, maskalash, residual bog'lanish, layer normalization va feed-forward qatlam. Encoder, decoder va encoder-decoder modellarining farqi.

Oldindan o'qitilgan modellarni moslashtirish: feature-based foydalanish, to'liq fine-tuning, adapter, prompt/prefix tuning, LoRA va QLoRA. Transfer learning, tillararo ko'chirish, zero-shot va few-shot o'qitish. Kam resursli tillar uchun ma'lumotni ko'paytirish, distillatsiya, kvantlash va samarali inferensiya.

7. Katta til modellari, generativ sun'iy intellekt va intellektual muloqot agentlari

Katta til modelining asosiy tamoyillari: autoregressiv modellashtirish, maskalangan til modeli, parametr, kontekst oynasi, token byudjeti, pretraining ma'lumotlari, masshtablash va emergent xususiyatlar. Model imkoniyatini til kompetensiyasi bilan aynanlashtirishning metodologik xatarlari.

Instruksiyaga moslashtirish: supervised fine-tuning, inson fikr-mulohazasi asosida mustahkamlovchi o'qitish, preference optimization va DPO. Prompt

dizayni, zero-shot, one-shot, few-shot, in-context learning, reasoning prompting va javob formatini boshqarish.

Generatsiya strategiyalari: greedy, beam search, temperature, top-k, top-p, repetition penalty va stop condition. Anqlik, xilma-xillik, izchillik va boshqariluvchanlik o'rtasidagi muvozanat.

Tashqi bilim bilan boyitilgan generatsiya: hujjatlarni bo'laklash, embedding, indekslash, retriever, reranker, kontekst yig'ish, generator, manba ko'rsatish va javobni tekshirish. RAG tizimlarida qidiruv xatosi, kontekst shovqini va generatsiya xatosini ajratish.

Agent tizimlari: vazifani rejalashtirish, vosita chaqirish, xotira, tashqi API va ma'lumotlar bazasi bilan ishlash, ko'p bosqichli nazorat, xavfsiz bajarish va kuzatuvchanlik. Deterministik jarayon bilan generativ model o'rtasidagi chegaralarni belgilash.

Chatbot va dialogue system arxitekturasini: niyat va slotlarni aniqlash, dialog holatini kuzatish, siyosat, javob generatsiyasi, retrieval va generativ yondashuvlar, foydalanuvchi modeli, kontekstni saqlash va eskalatsiya.

Generativ tizimlarni baholash: faktologik to'g'rilik, manbaga muvofiqlik, mazmuniy moslik, to'liqlik, foydalilik, zararli mazmun, tarafkashlik va maxfiylik. Inson ekspertizasi, avtomatik metrikalar va katta til modelidan baholovchi sifatida foydalanishning imkoniyat hamda cheklovlari.

O'zbek tili uchun LLM: tokenizatsiya samaradorligi, korpus sifati, yozuvlararo ko'chirish, instruktsion dataset, terminologik va madaniy moslik, xavfsizlik testlari, benchmark va ochiq reproduktiv baholash.

8. Mashina tarjimasi va ko'p tilli intellektual platformalar

Mashina tarjimasi nazariyasi va rivojlanishi: qoidalarga asoslangan, misollarga asoslangan, statistik va neyron tarjima. Lingvistik bilim, parallel ma'lumot va umumiyashtirish o'rtasidagi nisbat.

Parallel korpuslar: hujjat, gap, ibora va token darajasida moslashtirish; 1:1, 1:N, N:1 va nomos segmentlar; shovqin, dublikat, noto'g'ri til, domen va tarjima sifati. Alignment algoritmlari va parallel gold qatlam.

Neyron mashina tarjimasi: encoder-decoder, attention, Transformer, subso'z segmentatsiyasi, teacher forcing, decoding va uzunlik normalizatsiyasi. Ko'p tilli, ko'p yo'nalishli, zero-shot va pivot tarjima.

Terminologik cheklov, glossariy, majburiy tarjima, transliteratsiya, nomlangan birliklar, raqam va formatni saqlash. Domain adaptation, back-translation, synthetic data va post-editing.

Tarjima sifatini baholash: inson bahosi, adequacy, fluency, error taxonomy, BLEU, chrF, TER va semantik metrikalar. Statistik ishonchlilik va segmentlar kesimidagi xatolar. O'zbek-ingliz va o'zbek-rus tarjimasida morfologiya, so'z tartibi va kontekst muammolari.

Tarjima platformasi arxitekturasini: terminologik baza, tarjima xotirasi, parallel korpus, model serveri, sifat nazorati, post-editing interfeysi, foydalanuvchi fikri va modelni uzluksiz yaxshilash.

9. Axborot qidirish, ma'lumotlarni ajratib olish, savol-javob va bilimlardan xulosa chiqarish

Axborot qidirish modeli: hujjat yig'ish, tozalash, indekslash, so'rovni tahlil qilish, matching, ranking va natijani taqdim etish. Invertlangan indeks, Boolean qidiruv, TF-IDF, BM25 va lingvistik normalizatsiya.

Semantik qidiruv: bi-encoder va cross-encoder, zich vektorli indeks, taxminiy eng yaqin qo'shni qidiruvi, hybrid retrieval, query expansion va reranking. Domen, til va so'rov uzunligining natijaga ta'siri.

Ma'lumotlarni avtomatik ajratib olish: nomlangan obyekt, munosabat, hodisa, atribut, jadval va hujjat strukturasi. Pipeline va joint model, distant supervision, weak supervision va faol o'qitish.

Savol-javob tizimlari: factoid, ro'yxatli, izohli va ko'p bosqichli savollar; extractive, retrieval-based va generative yondashuv. Savolni tushunish, dalil topish, javobni yig'ish va noaniqlikni ifodalash.

Semantik tarmoq, ontologiya va bilimlar grafi: sinf, obyekt, atribut, munosabat, aksioma va cheklovlar. Entitetlarni bog'lash, graph embedding, link prediction, qoidaviy va neyron xulosa chiqarish.

RAG va bilimga asoslangan generativ tizimlarda manba izini saqlash, dalilga tayanganlik, iqtibosning to'g'riligi va qarama-qarshi manbalarni boshqarish. Qidiruv, reranking va generatsiyani alohida hamda end-to-end baholash.

Baholash: precision@k, recall@k, MAP, MRR, nDCG, exact match, token F1, entity/relation F1, answer faithfulness va factuality. Qidiruv jurnalidan foydalanib foydalanuvchi xulqini tahlil qilish.

10. Nutq texnologiyalari va audio-lingvistik qayta ishlash

Nutq signalining lingvistik va akustik xususiyatlari. Diskretlash, spektr, mel-shkala, MFCC, prosodiya, urg'u, intonatsiya, pauza va fonetik birliklar. Audio korpus yaratish va vaqt bo'yicha annotatsiya.

Avtomatik nutqni tanish arxitekturası: akustik model, talaffuz lug'ati, til modeli va dekoder. HMM-DNN, CTC, RNN-Transducer, attention-based va end-to-end Transformer modellar. Streaming va offline tanish.

Nutq sintezi: matnni normallashtirish, grapheme-to-phoneme, prosodik model, akustik model va vokoder. Konkatenativ, parametrik va neyron TTS. Tabiiylik, tushunarliklik va ovoz identifikatsiyasi.

Spiker, dialekt, til, emotsiya va paralingvistik belgilarni aniqlash. Shovqin, mikrofon, kanal va nutq uslubiga chidamlilik. Audio augmentatsiya va domen moslashtirish.

Baholash: word error rate, character error rate, segment xatolari, real-time factor, MOS va tinglovchi ekspertizasi. O'zbek tilida fonema-grafema munosabati, sheva, o'zlashma so'zlar, son va qisqartmalarni o'qish muammolari.

Nutq texnologiyalarini call-markaz, tibbiyot, ta'lim, media, subtitr, ovozli qidiruv va yordamchi texnologiyalarda qo'llash. Maxfiylik, biometrik ma'lumot va rozilik masalalari.

11. Multimodal lingvistika

Multimodal ma'lumot va multimodal ma'no tushunchasi. Matn, nutq, tasvir, video, mimika, imo-ishora va kontekstni vaqt hamda obyekt darajasida moslashtirish. Modalitetlar o'rtasidagi komplementarlik, redundantlik va ziddiyat.

Multimodal arxitekturalar: early fusion, late fusion, joint representation, cross-attention va encoder-decoder. Matn-tasvir kontrastiv o'qitish, multimodal embedding va vision-language modellar.

Amaliy vazifalar: rasmga izoh yaratish, vizual savol-javob, video mazmunini tahlil qilish, OCR va hujjat tushunish, audiovizual nutqni tanish, multimodal hissiy tahlil, imo-ishora va subtitrl integratsiyasi.

Multimodal korpuslar: sinxronlashtirish, annotatsiya birliklari, vaqt oralig'i, bounding box, obyekt izi, nutq segmenti, dialog va emotsiya. Modalitetning yetishmasligi va shovqinli moslik.

Baholash, tarafkashlik, shaxsiy tasvir va ovozni himoya qilish, accessibility hamda zaif guruhlar uchun texnologiyalar. O'zbekcha multimodal ma'lumotlarning kamligi va transfer learning imkoniyatlari.

12. Amaliy intellektual tizimlar, raqamli ta'lim va inson-kompyuter muloqoti

Kompyuter leksikografiyasi va terminografiyasi: elektron lug'at, terminologik platforma, qidiruv, morfologik kengaytirish, misol tanlash, ma'no ajratish, kollokatsiya, audio va multimodal komponentlar. Lug'at maqolasini yarim avtomatik shakllantirish.

Intellektual tahrir-nashr tizimlari: imlo, punktuatsiya, grammatika, uslub, terminologik izchillik, plagiatga yaqinlik, transliteratsiya va format nazorati. Xatoni aniqlash va tuzatish, foydalanuvchi bilan iterativ ishlash va izoh berish.

Matnni avtomatik umumlashtirish: extractive va abstractive yondashuv, single-document va multi-document, query-focused va dialogue summarization. Faktologik to'g'rilik, qamrov, ortiqchalik va manbaga muvofiqlik.

Sun'iy intellekt asosidagi ekspert va tavsiya tizimlari: bilim bazasi, qoidalar, profil, kontekst, ranking, tushuntirish va feedback. Kontent, kollaborativ va gibrid tavsiya. Lingvistik belgilar va foydalanuvchi tilini modellashtirish.

Kompyuter lingvodidaktikasi va raqamli ta'lim: avtomatik mashq yaratish, javobni baholash, xato diagnostikasi, moslashtirilgan o'qitish, o'quvchi modeli, intelligent tutoring system va ta'lim analitikasi. Akademik halollik va generativ AI bilan ishlash.

Inson-kompyuter muloqoti: foydalanuvchi ehtiyoji, dialog dizayni, tushunarlilik, qisqalik, moslashuvchanlik, accessibility, ko'p tillilik va xatodan tiklanish. Raqamli kommunikatsiya uchun optimal til vositalari va foydalanuvchi tajribasini baholash.

Raqamli gumanitar fanlarda lingvistik texnologiyalar: tarixiy matnlar, arxivlar, stilometriya, mualliflik, mavzu dinamikasi, tarmoqlar va geolingvistik tahlil. Ilmiy talqinda hisoblash natijasining kontekstual cheklavlari.

Amaliy tizim arxitekturası: ma'lumotlar qatlami, lingvistik servislar, model serveri, API, foydalanuvchi interfeysi, loglash, monitoring, versiyalash, xavfsizlik, tezlik va masshtablash. Prototipdan ishlab chiqarish muhitiga o'tish.

13. Ilmiy tadqiqot metodologiyasi, baholash, takrorlanuvchanlik va mas'uliyatli sun'iy intellekt

Ilmiy muammo, obyekt, predmet, maqsad, vazifa, gipoteza, tadqiqot savoli, ilmiy yangilik, nazariy va amaliy ahamiyat. Adabiyotlarni tizimli izlash, tanqidiy sharhlash, mavjud bo'shliqni aniqlash va natijani ixtisoslik pasportiga bog'lash.

Ma'lumotlarga asoslangan tadqiqot dizayni: maqsadli populyatsiya, tanlanma, domen va janr, annotatsiya siyosati, train/dev/test ajratish, vaqt va manba bo'yicha split, kross-validatsiya, data leakage, dublikat va distribution shift.

Baseline va qiyosiy tajriba. Bir xil ma'lumot, preprocessing va baholash sharoitini ta'minlash. Giperparametr tanlash, seed, hisoblash resursi va eksperiment jurnalini yuritish. Ablatsiya, sensitivity va robustness sinovlari.

Vazifaga mos metrikani tanlash: accuracy, precision, recall, F1, macro/micro/weighted hisoblash, ROC-AUC, calibration; sequence labeling va parsing metrikalari; BLEU/chrF/COMET; ROUGE/BERTScore; WER/CER; IR va RAG metrikalari. Metrikaning xato narxi va foydalanuvchi maqsadiga mosligi.

Statistik ishonchlilik: ishonch oralig'i, bootstrap, randomizatsiya testi, effekt o'lchami va ko'p taqqoslash. Natija farqini amaliy ahamiyat bilan ajratish. Sifat tahlili, xatolar taksonomiyasi va ekspert muhokamasi.

Takrorlanuvchanlik va ochiq ilm: kod, konfiguratsiya, ma'lumot versiyasi, muhit, model checkpoint, seed, dataset card, model card, tajriba protokoli va litsenziya. Salbiy natija va cheklovlarni halol bayon qilish.

Tushuntiriluvchanlik va noaniqlik: feature importance, attentionni talqin qilish ehtiyotkorligi, LIME/SHAP, misolga asoslangan tushuntirish, confidence, calibration, selective prediction va human-in-the-loop.

Etik va huquqiy masalalar: mualliflik huquqi, shaxsiy ma'lumot, rozilik, maxfiylik, diskriminatsiya va tarafkashlik, toksik va zararli mazmun, til hamjamiyatiga zarar, ekologik va hisoblash xarajati. Xavfsizlik, prompt injection, ma'lumot sizishi va modeldan noto'g'ri foydalanish xavflari.

PhD tadqiqotini rejalashtirish: uch yillik ish paketi, ilmiy natijalar, resurs va dasturiy mahsulot, validatsiya, joriy etish, nashr strategiyasi, risklar va kutilayotgan ta'sir.

BAHOLASH MEZONLARI

Kirish imtihoni amaldagi dasturda belgilangan tartibga muvofiq og‘zaki shaklda o‘tkaziladi va 100 ballik mezon asosida baholanadi. Imtihon savollari mutaxassislikning mohiyatidan kelib chiqib qabul komissiyasi tomonidan tuziladi hamda universitet Kengashi tomonidan tasdiqlanadi.

Savollar nazariy-metodologik, formal-algoritmik va amaliy-tadqiqot bloklarining mutanosib qamrovini ta‘minlashi kerak. Talabgorning javobida tushunchaning ilmiy ta‘rifi, model yoki algoritmnining ishlash tamoyili, zarur ma‘lumot va resurslar, baholash usuli, cheklovlar hamda o‘zbek tili materialiga tatbiq misoli yoritilishi kutiladi.

Talabgorning yakuniy javobi quyidagi mezonlar asosida baholanadi:

Berilgan javobning mazmuni va sifati	Ball
Savolning nazariy va metodologik mohiyati to‘liq ochilgan; terminlar aniq qo‘llangan; yondashuvlar tanqidiy qiyoslangan; formal model, algoritmi yoki tizim arxitekturasini asoslangan; ma‘lumotlar, baseline, metrikalar va xatoliklar tahlili ko‘rsatilgan; o‘zbek tili materialida asosli misol keltirilgan; mustaqil ilmiy xulosa, cheklov va istiqbol bayon etilgan.	86–100
Savolning asosiy mazmuni to‘g‘ri yoritilgan; tushunchalar va yondashuvlar izchil tushuntirilgan; model yoki amaliy misol keltirilgan; baholash va cheklovlar qisman asoslangan, biroq ayrim metodologik yoki texnik jihatlar yetarlicha chuqur ochilmagan.	71–85
Savol mohiyati umumiy darajada anglangan; ayrim asosiy tushunchalar to‘g‘ri bayon qilingan, ammo yondashuvlar o‘rtasidagi bog‘lanish, formal yoki algoritmik asos, ma‘lumot va baholash mezonlari to‘liq ochilmagan; misol va xulosa yetarli emas.	55–70
Savolning faqat ayrim qirralari yoritilgan yoki asosiy tushunchalar chalkashtirilgan; javobda ilmiy mantiq, model, dalil, misol va xulosa mavjud emas; savolga mazmunan javob berilmagan.	0–54

TAVSIYA ETILADIGAN ADABIYOTLAR

I. Normativ-huquqiy va metodologik manbalar

1. O‘zbekiston Respublikasi Prezidentining 2019-yil 21-oktabrdagi “O‘zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to‘g‘risida”gi PF-5850-son Farmoni.
2. O‘zbekiston Respublikasi Prezidentining 2020-yil 20-oktabrdagi “Mamlakatimizda o‘zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida”gi PF-6084-son Farmoni.
3. O‘zbekiston Respublikasining “Ta’lim to‘g‘risida”gi Qonuni. O‘RQ-637, 2020-yil 23-sentabr.
4. O‘zbekiston Respublikasi Vazirlar Mahkamasining 2017-yil 22-maydagi “Oliy ta’limdan keyingi ta’lim tizimini yanada takomillashtirish chora-tadbirlari to‘g‘risida”gi 304-son qarori.
5. 10.00.11 – Amaliy va kompyuter lingvistikasi ixtisosligi pasporti.
6. Nurmonov A. Lingvistik tadqiq metodologiyasi va metodlari. – Toshkent, 2011.
7. Bushuy A.M., Safarov Sh. Til qurilishi: tahlil metodlari va metodologiyasi. – Toshkent: Fan, 2007.

II. O‘zbek tilidagi asosiy adabiyotlar

8. Po‘latov A.K., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2009.
9. Po‘latov A. Kompyuter lingvistikasi. – Toshkent: Akademnashr, 2011.
10. Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent: Akademnashr, 2011.
11. Mengliyev B., Ergasheva G., Hamroyeva Sh. Tilshunoslik asoslari. – Toshkent, 2025.
12. Hojiyev A. Tilshunoslik terminlarining izohli lug‘ati. – Toshkent, 2002.
13. Safarov Sh. Pragmalingvistika. – Toshkent, 2013.
14. Xolmanova Z. Tilshunoslik nazariyasi. – Toshkent, 2020.

III. Kompyuter lingvistikasi, NLP va sun’iy intellekt bo‘yicha adabiyotlar

15. Jurafsky D., Martin J.H. Speech and Language Processing. 2nd ed. – Upper Saddle River: Prentice Hall, 2009.
16. Eisenstein J. Introduction to Natural Language Processing. – Cambridge, MA: MIT Press, 2019.
17. Manning C.D., Schütze H. Foundations of Statistical Natural Language Processing. – Cambridge, MA: MIT Press, 1999.
18. Bird S., Klein E., Loper E. Natural Language Processing with Python. – Sebastopol: O’Reilly Media, 2009.
19. Goldberg Y. Neural Network Methods for Natural Language Processing. – Morgan & Claypool, 2017.
20. Goodfellow I., Bengio Y., Courville A. Deep Learning. – Cambridge, MA: MIT Press, 2016.
21. Murphy K.P. Probabilistic Machine Learning: An Introduction. – Cambridge, MA: MIT Press, 2022.
22. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. – Pearson, 2021.
23. Clark A., Fox C., Lappin S. (eds.). The Handbook of Computational Linguistics and Natural Language Processing. – Wiley-Blackwell, 2010.
24. Bender E.M., Lascarides A. Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics. – Morgan & Claypool, 2019.

25. Tunstall L., von Werra L., Wolf T. Natural Language Processing with Transformers. – Sebastopol: O'Reilly Media, 2022.
26. Aggarwal C.C. Machine Learning for Text. – Cham: Springer, 2018.

IV. Axborot qidirish, mashina tarjimasi, korpus va leksikografiya

27. Manning C.D., Raghavan P., Schütze H. Introduction to Information Retrieval. – Cambridge: Cambridge University Press, 2008.
28. Koehn P. Statistical Machine Translation. – Cambridge: Cambridge University Press, 2010.
29. Koehn P. Neural Machine Translation. – Cambridge: Cambridge University Press, 2020.
30. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. – Cambridge: Cambridge University Press, 2012.
31. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. – Cambridge: Cambridge University Press, 1998.
32. Ide N., Pustejovsky J. (eds.). Handbook of Linguistic Annotation. – Dordrecht: Springer, 2017.
33. Atkins B.T.S., Rundell M. The Oxford Guide to Practical Lexicography. – Oxford: Oxford University Press, 2008.
34. Gries S.Th. Statistics for Linguistics with R: A Practical Introduction. 3rd ed. – Berlin/Boston: De Gruyter Mouton, 2021.
35. Fellbaum C. (ed.). WordNet: An Electronic Lexical Database. – Cambridge, MA: MIT Press, 1998.
36. Schreibman S., Siemens R., Unsworth J. (eds.). A Companion to Digital Humanities. – Oxford: Blackwell, 2004.

V. Nutq, multimodal tizimlar va inson–kompyuter muloqoti

37. Huang X., Acero A., Hon H.-W. Spoken Language Processing. – Upper Saddle River: Prentice Hall, 2001.
38. Rabiner L., Schafer R. Theory and Applications of Digital Speech Processing. – Pearson, 2011.
39. Taylor P. Text-to-Speech Synthesis. – Cambridge: Cambridge University Press, 2009.
40. Dix A., Finlay J., Abowd G.D., Beale R. Human–Computer Interaction. 3rd ed. – Pearson, 2004.
41. Baltrušaitis T., Ahuja C., Morency L.-P. Multimodal Machine Learning: A Survey and Taxonomy // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2019. – Vol. 41, No. 2. – P. 423–443.

VI. Asosiy ilmiy maqolalar va zamonaviy metodlar

42. Vaswani A. et al. Attention Is All You Need // Advances in Neural Information Processing Systems. – 2017.
43. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // NAACL-HLT. – 2019.
44. Liu Y. et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. – 2019.
45. Conneau A. et al. Unsupervised Cross-lingual Representation Learning at Scale // ACL. – 2020.
46. Sennrich R., Haddow B., Birch A. Neural Machine Translation of Rare Words with Subword Units // ACL. – 2016.

47. Kudo T., Richardson J. SentencePiece: A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text Processing // EMNLP System Demonstrations. – 2018.
48. Mikolov T. et al. Distributed Representations of Words and Phrases and their Compositionality // NeurIPS. – 2013.
49. Pennington J., Socher R., Manning C.D. GloVe: Global Vectors for Word Representation // EMNLP. – 2014.
50. Bojanowski P. et al. Enriching Word Vectors with Subword Information // Transactions of the ACL. – 2017.
51. Lewis P. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // NeurIPS. – 2020.
52. Ouyang L. et al. Training Language Models to Follow Instructions with Human Feedback // NeurIPS. – 2022.
53. Hu E.J. et al. LoRA: Low-Rank Adaptation of Large Language Models // ICLR. – 2022.
54. Dettmers T. et al. QLoRA: Efficient Finetuning of Quantized LLMs // NeurIPS. – 2023.
55. Papineni K. et al. BLEU: A Method for Automatic Evaluation of Machine Translation // ACL. – 2002.
56. Lin C.-Y. ROUGE: A Package for Automatic Evaluation of Summaries // Text Summarization Branches Out. – 2004.
57. Zhang T. et al. BERTScore: Evaluating Text Generation with BERT // ICLR. – 2020.
58. Ribeiro M.T., Singh S., Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier // KDD. – 2016.
59. Lundberg S.M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // NeurIPS. – 2017.
60. Gebru T. et al. Datasheets for Datasets // Communications of the ACM. – 2021. – Vol. 64, No. 12. – P. 86–92.
61. Mitchell M. et al. Model Cards for Model Reporting // FAT*. – 2019.
62. Bender E.M. et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // FAccT. – 2021.
63. Baevski A. et al. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // NeurIPS. – 2020.
64. Graves A. et al. Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks // ICML. – 2006.

VII. Elektron standartlar va ilmiy resurslar

- Universal Dependencies: annotatsiya tamoyillari, tilga xos ko‘rsatmalar va CoNLL-U formati.
- Text Encoding Initiative (TEI P5) Guidelines.
- The Unicode Standard va Unicode Character Database.
- ISO 24611 Language Resource Management — Morpho-syntactic Annotation Framework (MAF).
- ISO 24612 Language Resource Management — Linguistic Annotation Framework (LAF).
- ISO 24615-1 Language Resource Management — Syntactic Annotation Framework (SynAF).
- ACL Anthology — kompyuter lingvistikasi bo‘yicha ilmiy maqolalar bazasi.

- CLARIN, ELRA/ELDA va Linguistic Data Consortium — til resurslari kataloglari.
- Hugging Face Hub — model, dataset va demo resurslari.
- NLTK, spaCy, Stanza, scikit-learn, PyTorch va TensorFlow rasmiy hujjatlari.

NAMUNAVIY SAVOLLAR

Quyidagi savollar tavsiya xarakteriga ega bo'lib, imtihon biletlarini shakllantirishda ixtisoslik pasporti va dastur bo'limlarining mutanosib qamrovini ta'minlash uchun xizmat qiladi.

1. Nazariy-metodologik asoslar

1. Amaliy lingvistika va kompyuter lingvistikasining obyekti, predmeti hamda vazifalarini qiyoslab tushuntiring.
2. 10.00.11 ixtisosligining fanlararo tabiatini informatika, sun'iy intellekt, mantiq, semiotika va kognitiv fanlar misolida tahlil qiling.
3. Amaliy va kompyuter lingvistikasining sinxron hamda diaxron rivojlanish bosqichlarini izohlang.
4. Qoidalariga asoslangan, statistik, neyron va gibrid paradigmalarni ilmiy bilish modeli sifatida qiyoslang.
5. Lingvistik nazariya bilan dasturiy muhandislik talablari o'rtasidagi muvozanat qanday ta'minlanadi?
6. Tilni belgilar tizimi, ehtimollik taqsimoti va hisoblash jarayoni sifatida talqin qilishning farqlarini ko'rsating.
7. Insonning kognitiv va ijodiy nutqiy faoliyatini hisoblash modeli yordamida ifodalashning imkoniyat va chegaralarini tahlil qiling.
8. Kompyuter lingvistikasida tushuntiriluvchanlik va bashorat aniqligi o'rtasidagi munosabatni izohlang.
9. Inson–kompyuter muloqotida optimal til vositasini yaratish uchun qaysi lingvistik mezonlar hisobga olinadi?
10. O'zbek tilining kam resursli va agglyutinativ til sifatidagi xususiyatlari tadqiqot metodologiyasiga qanday ta'sir qiladi?

2. Formal, matematik va statistik-kombinator modellash

1. Tabiiy til hodisasini formal modelga aylantirish bosqichlarini aniq misol asosida tushuntiring.
2. Alfavit, satr, formal til, grammatika va avtomat tushunchalarining NLPdagi o'zaro bog'lanishini izohlang.
3. Muntazam ifoda va chekli avtomatning o'zbek matnini qayta ishlashdagi qo'llanishiga misol keltiring.
4. Kontekstsiz grammatika va bog'liqlik grammatikasining ifodalash imkoniyatlarini qiyoslang.
5. Xususiyatli grammatika va unifikatsiya mexanizmi morfologik-sintaktik moslashuvni qanday ifodalaydi?
6. Lingvistik tuzilmalarni daraxt, graf, jadval va vektor ko'rinishida ifodalashning afzalliklarini tahlil qiling.
7. Shartli ehtimol, Bayes qoidasi, MLE va MAP baholashni NLP misolida tushuntiring.
8. Entropiya, kross-entropiya va perplexity o'rtasidagi bog'lanishni izohlang.
9. Kosinus o'xshashligi va masofa metrikalarini matnlarning mazmuniy yaqinligini baholashda qiyoslang.

10. Algoritmning vaqt va xotira murakkabligi real til tizimining arxitekturasiga qanday ta'sir qiladi?
11. Zipf va Heaps qonunlarining korpus va lug'at hajmini tahlil qilishdagi ahamiyatini tushuntiring.
12. Bootstrap, ishonch oralig'i va statistik ahamiyatni model natijalarini qiyoslashda qanday qo'llash mumkin?

3. Korpuslar, annotatsiya va raqamli til resurslari

1. Korpusning reprezentativligi va muvozanati qanday aniqlanadi?
2. Milliy, maxsus, parallel, tarixiy, dinamik, nutqiy va multimodal korpuslarni qiyoslang.
3. Korpus ma'lumotlarining kelib chiqishi va huquqiy maqomini hujjatlashtirish nega muhim?
4. Matnni tozalash, normalizatsiya va deduplikatsiya jarayonlarini bir-biridan farqlang.
5. Annotatsiya tagseti va yo'riqnomasini yaratish bosqichlarini bayon qiling.
6. Annotatorlararo kelishuv nima va Cohen kappa, Fleiss kappa hamda Krippendorff alpha qachon qo'llanadi?
7. Gold standart, ko'r annotatsiya va adjudikatsiya tushunchalarini izohlang.
8. BIO/BIOES, CoNLL-U, XML/TEI va JSONL formatlarining vazifalarini qiyoslang.
9. Elektron lug'at, tezaurus, WordNet, ontologiya va bilimlar grafi o'rtasidagi farqlarni tushuntiring.
10. Relyatsion, hujjat, graf va vektorli ma'lumotlar bazalarini lingvistik resurslar uchun qiyoslang.
11. O'zbek tili milliy korpusining tavsiya etiladigan funksional arxitekturasini ishlab chiqing.
12. Dataset card, versiyalash, litsenziya va ma'lumotlar maxfiyligi resursning ilmiy qiymatiga qanday ta'sir qiladi?

4. Matnni tayyorlash va morfologik qayta ishlash

1. Unicode normalizatsiyasi va apostrof variantlari o'zbek NLP tizimlarida qanday xatolarga sabab bo'ladi?
2. Lotin–kirill transliteratsiyasi bilan ortografik normalizatsiyani qiyoslang.
3. Gap chegarasini avtomatik aniqlashda qisqartma, son va punktuatsiya muammolarini tahlil qiling.
4. Token, so'zshakl, grafema va ortografik harf tushunchalarini farqlang.
5. O'zbek tilida defisli, apostrofli, qo'shma va klitik birliklarni tokenizatsiya qilish siyosatini taklif eting.
6. BPE, WordPiece va Unigram subso'z modellarini agglyutinativ til uchun qiyoslang.
7. Stemming, lemmatizatsiya va morfemik segmentatsiya o'rtasidagi farqlarni tushuntiring.
8. Morfologik analizator va morfologik sintezatorning kirish-chiqishlarini formal tavsiflang.
9. Chekli holatli transduser va neyron morfologik modelni qiyoslang.

10. POS va morfologik xususiyatlarni belgilashda universal hamda milliy tagsetlar qanday uyg'unlashtiriladi?
11. Morfologik omonimiyani bartaraf etishda kontekstning rolini tahlil qiling.
12. Tokenizatsiya, lemmatizatsiya va POS teglash tizimlari uchun baholash va xatoliklar tahlili rejasini tuzing.

5. Sintaksis, semantika, pragmatika va diskurs

1. Bevosita tashkil etuvchilar va bog'liqlik sintaksisini qiyoslang.
2. Universal Dependencies tizimining asosiy tamoyillari va o'zbek tili uchun ahamiyatini tushuntiring.
3. O'tishlarga asoslangan va grafga asoslangan dependency parserlarni qiyoslang.
4. Erkin so'z tartibi, ellipsis va koordinatsiya sintaktik tahlilda qanday muammo tug'diradi?
5. So'z ma'nosini kontekstda aniqlash vazifasini formallashtiring va baholash usulini taklif eting.
6. Semantik rollar va valentlik nazariyasining avtomatik semantik tahlildagi o'rini izohlang.
7. Yassi va ichma-ich NER uchun ma'lumot formati hamda model tanlovini qiyoslang.
8. Entitetni bilim bazasiga bog'lash jarayonining bosqichlarini tushuntiring.
9. Munosabat va hodisalarni ajratib olishda pipeline va joint yondashuvlarni qiyoslang.
10. Ko'p so'zli birlik va terminlarni aniqlashda statistik hamda lingvistik mezonlarni birlashtirish yo'lini taklif eting.
11. Anafora va koreferensiyani hal qilishda sintaktik, semantik va diskurs omillarining rolini tahlil qiling.
12. Nutqiy akt, deysis, presuppozitsiya va implikaturani avtomatik aniqlash muammolarini bayon qiling.
13. Diskurs koherensiyasi va dialog aktlarini modellashtirish usullarini tushuntiring.
14. Sentiment, aspektga asoslangan sentiment, emotsiya, sarkazm va toksik mazmun vazifalarini qiyoslang.

6. Mashinali o'qitish, neyron tarmoqlar va Transformer

1. Bag-of-words, n-gram va TF-IDF tasvirlarini qiyoslang.
2. Naive Bayes, logistik regressiya va SVM algoritmlarini matn tasnifi uchun qiyoslang.
3. Sinflar nomutanosibli model va metrikaga qanday ta'sir qiladi?
4. HMM va CRF ketma-ketlikni belgilash vazifasida qanday farqlanadi?
5. Klasterlash va mavzu modellashtirishning maqsadi hamda natijalarini baholash usullarini izohlang.
6. Word2Vec, GloVe va FastText modellarining farqli jihatlarini tahlil qiling.
7. Kontekstual embedding statik embeddingdan nimasi bilan farq qiladi?
8. MLP, CNN, RNN, LSTM va GRU arxitekturalarining til vazifalaridagi rolini qiyoslang.
9. Encoder-decoder, attention va teacher forcing mexanizmlarini tushuntiring.

10. Transformer arxitekturasining self-attention, multi-head attention va pozitsion kodlash komponentlarini izohlang.
11. BERT, RoBERTa va XLM-R modellarining o'qitish maqsadi hamda qo'llanishini qiyoslang.
12. Transfer learning va tillararo ko'chirish kam resursli o'zbek tili uchun qanday imkoniyat yaratadi?
13. To'liq fine-tuning, adapter, LoRA va QLoRA usullarini resurs talabi bo'yicha qiyoslang.
14. Ma'lumotni ko'paytirish, distillatsiya va kvantlash usullarining afzallik hamda xavflarini tahlil qiling.

7. Katta til modellari va generativ sun'iy intellekt

1. Autoregressiv til modeli va maskalangan til modelining o'qitish maqsadini qiyoslang.
2. Parametr soni, pretraining ma'lumoti va kontekst oynasi model imkoniyatiga qanday ta'sir qiladi?
3. Supervised fine-tuning, RLHF va DPO yondashuvlarini qiyoslang.
4. Zero-shot, one-shot, few-shot va in-context learning tushunchalarini izohlang.
5. Prompt dizaynida topshiriq, kontekst, cheklov va chiqish formatining rolini tushuntiring.
6. Greedy, beam, temperature, top-k va top-p dekodlash strategiyalarini qiyoslang.
7. RAG tizimining asosiy modullari va xato manbalarini tahlil qiling.
8. Vosita chaqiruvchi agentda rejalashtirish, xotira, xavfsizlik va kuzatuvchanlik qanday ta'minlanadi?
9. Chatbotda niyat, slot, dialog holati, siyosat va javob generatsiyasi qanday bog'lanadi?
10. Asossiz ma'lumot yaratish, faktologik xato va manbaga nomuvofiqlikni farqlang.
11. Inson ekspertizasi va LLM-as-a-judge yondashuvlarining imkoniyat hamda cheklovlarini tahlil qiling.
12. O'zbek tili uchun LLM benchmarkini yaratish rejasini taklif eting.

8. Mashina tarjimasini

1. Qoidalarga asoslangan, statistik va neyron mashina tarjimasini qiyoslang.
2. Parallel korpus yaratish va gaplarni moslashtirish bosqichlarini tushuntiring.
3. 1:1, 1:N, N:1 va nomos segmentlar parallel korpusda qanday belgilanadi?
4. Encoder-decoder va attention mashina tarjimasida qanday ishlaydi?
5. Transformerning neyron mashina tarjimasidagi ustunliklarini tahlil qiling.
6. Ko'p tilli, zero-shot va pivot tarjima yondashuvlarini qiyoslang.
7. Subsoz segmentatsiyasi va OOV muammosi tarjima sifatiga qanday ta'sir qiladi?
8. Terminologik cheklov, transliteratsiya va nomlangan birliklarni saqlash usullarini izohlang.
9. BLEU, chrF, TER va inson bahosini qiyoslang.
10. O'zbek-ingliz mashina tarjimasini uchun xatolar taksonomiyasini taklif eting.

9. Axborot qidirish, ma'lumot ajratish, savol-javob va bilimlar grafi

1. Invertlangan indeks va BM25 algoritmining ishlash tamoyilini tushuntiring.

2. Siyrak, zich va gibrid qidiruvni qiyoslang.
3. Bi-encoder, cross-encoder va rerankerning vazifalarini tushuntiring.
4. Vektorli indeks va taxminiy eng yaqin qo'shni qidiruvining afzalliklarini tahlil qiling.
5. Ma'lumotlarni ajratib olish pipeline'ini NER, relation extraction va event extraction misolida tavsiflang.
6. Extractive va generative savol-javob tizimlarini qiyoslang.
7. Ontologiya, semantik tarmoq va bilimlar grafi o'rtasidagi farqlarni izohlang.
8. Entitetlarni bilimlar grafigiga bog'lash va link prediction jarayonini tushuntiring.
9. Bilimlardan qoidaviy va neyron xulosa chiqarish usullarini qiyoslang.
10. RAG tizimida chunking, retrieval, reranking va generation bosqichlarini optimallashtirish rejasini tuzing.
11. Precision@k, recall@k, MAP, MRR va nDCG metrikalarini qiyoslang.
12. O'zbek tilidagi sohaviy semantik qidiruv tizimining arxitekturasini taklif eting.

10. Nutq texnologiyalari

1. Nutq signalining asosiy akustik va prosodik xususiyatlarini izohlang.
2. MFCC va mel-spektrogrammaning nutqni tanishdagi o'rnini tushuntiring.
3. An'anaviy ASR tizimidagi akustik model, talaffuz lug'ati, til modeli va dekodeer vazifalarini bayon qiling.
4. HMM-DNN, CTC, RNN-Transducer va attention-based ASR modellarini qiyoslang.
5. Streaming va offline nutqni tanish tizimlarining farqlarini tahlil qiling.
6. Matndan nutq sintezining normalizatsiya, fonemashtirish, akustik model va vokoder bosqichlarini tushuntiring.
7. Nutqiy korpus yaratish va vaqt bo'yicha annotatsiya qilish metodologiyasini bayon qiling.
8. WER, CER, real-time factor va MOS metrikalarini qiyoslang.
9. O'zbek ASR/TTS tizimlarida sheva, urg'u, son, qisqartma va o'zlashma so'zlar muammosini tahlil qiling.
10. Ovozli ma'lumot bilan ishlashda rozilik, maxfiylik va biometrik xavfsizlik talablari qanday ta'minlanadi?

11. Multimodal lingvistik

1. Multimodal ma'no va modalitetlararo moslashtirish tushunchalarini izohlang.
2. Early fusion, late fusion va cross-attention yondashuvlarini qiyoslang.
3. Matn-tasvir kontrastiv o'qitish va multimodal embedding qanday ishlaydi?
4. Rasmga izoh yaratish va vizual savol-javob vazifalarini qiyoslang.
5. OCR va NLP integratsiyasida hujjat tuzilishi qanday hisobga olinadi?
6. Multimodal hissiy tahlilda nutq, matn va yuz ifodasi signallarini birlashtirish usulini taklif eting.
7. Multimodal korpusda vaqt, obyekt va dialog annotatsiyasi qanday tashkil etiladi?
8. Multimodal tizimlarda maxfiylik, tarafkashlik va accessibility muammolarini tahlil qiling.

12. Amaliy tizimlar, raqamli ta'lim va inson-kompyuter muloqoti

1. Yangi avlod elektron lug'atining lingvistik va dasturiy arxitekturasini taklif eting.

2. Terminologik baza va terminlarni avtomatik ajratib olish tizimini qanday integratsiya qilish mumkin?
3. Imlo, grammatika va uslub xatolarini aniqlash hamda tuzatish bosqichlarini qiyoslang.
4. Extractive va abstractive umumlashtirish usullarini faktologik xavf nuqtai nazaridan tahlil qiling.
5. Intellektual tahrir-nashr tizimida foydalanuvchiga tushuntirish berish nega muhim?
6. Ekspert tizimi va tavsiya tizimining bilim hamda qaror mexanizmlarini qiyoslang.
7. Intellektual o'qitish tizimida o'quvchi modeli, xato diagnostikasi va moslashtirish qanday ishlaydi?
8. Generativ sun'iy intellektdan ta'limda foydalanishda akademik halollik qanday ta'minlanadi?
9. Inson–kompyuter muloqoti uchun tushunarlilik, xatodan tiklanish va accessibility mezonlarini izohlang.
10. Raqamli gumanitar tadqiqotda stilometriya va mualliflikni aniqlash metodlarini tahlil qiling.
11. Avtomatlashtirilgan axborot tizimi va ma'lumotlar bazasining lingvistik ta'minoti nimalardan iborat?
12. O'zbek tilida end-to-end amaliy NLP mahsulotini prototipdan ishlab chiqarishgacha olib borish arxitekturasini bayon qiling.

13. Ilmiy tadqiqot dizayni, baholash va etika

1. PhD tadqiqotida muammo, obyekt, predmet, maqsad, vazifa va gipotezani o'zaro mos shakllantirish usulini tushuntiring.
2. Adabiyotlar sharhi orqali tadqiqot bo'shlig'i va ilmiy yangilik qanday aniqlanadi?
3. Dataset uchun tanlanma, domen, annotatsiya va train/dev/test splitini asoslash rejasini tuzing.
4. Baseline nima va u zamonaviy model bilan adolatli qiyoslashda qanday rol o'ynaydi?
5. Accuracy, precision, recall, F1, macro va micro hisoblashni xato narxi nuqtai nazaridan qiyoslang.
6. Kross-validatsiya, vaqt bo'yicha split va manba bo'yicha split qachon qo'llanadi?
7. Data leakage va dublikatlar tajriba natijasini qanday buzadi?
8. Ablatsiya va sensitivity tahlili modelning qaysi xususiyatlarini ochib beradi?
9. Xatolar taksonomiyasi va sifat tahlilini tuzish metodologiyasini bayon qiling.
10. Ishonch oralig'i, statistik ahamiyat va effekt o'lchamini birgalikda talqin qiling.
11. Kod, ma'lumot, konfiguratsiya, seed, model card va dataset card takrorlanuvchanlikka qanday xizmat qiladi?
12. Tushuntiriluvchanlik, confidence va calibration tushunchalarini qiyoslang.
13. Mualliflik huquqi, shaxsiy ma'lumot, tarafkashlik va zararli mazmun bo'yicha xavf tahlilini tuzing.
14. 10.00.11 ixtisosligi doirasida o'z PhD tadqiqot loyihangizning muammosi, ma'lumotlari, modeli, baholash rejasi, yangiligi va amaliy natijasini asoslab bering.

Mazkur dastur 10.00.11 – Amaliy va kompyuter lingvistikasi ixtisosligining yangilangan ilmiy mazmunini aks ettiradi va kirish imtihoni savollarini shakllantirish uchun asos bo‘lib xizmat qiladi.

Ilmiy va texnologik taraqqiyotning yuqori sur‘atini hisobga olib, dasturdagi zamonaviy model, standart, metrika va dasturiy vositalar ro‘yxati universitet Kengashi qarori bilan davriy ravishda yangilanishi mumkin. Bunda ixtisoslik pasportining asosiy tadqiqot yo‘nalishlari, lingvistik va hisoblash komponentlari muvozanati hamda o‘zbek tili uchun ilmiy-amaliy ahamiyat saqlanishi shart.

Kirish imtihonida talabgorning tayyor javobni takrorlashi emas, balki ilmiy muammoni anglash, dalillash, formallashtirish, metod tanlash, tajriba va baholashni loyihalash hamda mas‘uliyatli xulosa chiqarish qobiliyati ustuvor baholanadi.