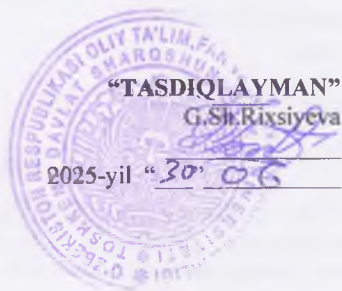


✓

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



STATISTIKA
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)

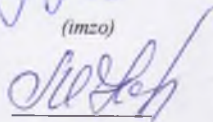
Tarjimashunoslik, tilshunoslik va
xalqaro jurnalistika oliy maktabi boshlig'i:



S.T. Mustafayeva

(imzo)

Arm boshlig'i:



M. Yuldasheva

(imzo)

Fan/modul kodi STST1206	O'quv yili 2025/2026	Semestr 2	YeCTS - Kreditlar 6			
Fan/modul turi Majburiy	Ta'lim tili O'zbek		Haftadagi dars soatlari 6			
1.	Fanning nomi	Auditoriya mashg'ulotlari (soat)	Ma'ruza	Amaliy	Mustaqil ta'lim (soat)	Jami yuklama (soat)
	Statistika	72	36	36	108	180
2.	<p>I. Fanning mazmuni:</p> <p>Fanning maqsadi: talabalarga til ma'lumotlarini tahlil qilish uchun statistik usullardan foydalanish bo'yicha chuqur bilim va ko'nikmalarni berish, shuningdek, tilshunoslikda statistik tadqiqotlar natijalarini izohlay olish qobiliyatini rivojlantirishdan iborat. Fan talabalarni tilshunoslik sohasida miqdoriy tadqiqotlar olib borishga va statistik modellardan foydalangan holda katta hajmdagi matnli ma'lumotlarni qayta ishlash usullarini o'zlashtirishga tayyorlashga qaratilgan.</p> <p>Fanning vazifasi: Talabalarga ehtimollar nazariyasi va statistik tahlil asoslarini o'rgatish. So'zlar va grammatika birliklarining chastotaviy taqsimotini hisoblash usullarini o'rgatish. Matnlarni tahlil qilishda statistik yondashuvlarni, masalan, POS-taggerlar, sintaktik tahlil va morfologik tahlil usullarini qo'llashni o'rgatish. So'zlar chastotasi, gap uzunligi, sintaktik strukturalarning chastotasi kabi lingvistik ko'rsatkichlarni tahlil qilish va talqin qilish. Matnlarni qayta ishlashda HMM (yashirin Markov modellari), n-gramma modellaridan foydalanishni o'rgatish. Statistik tahlilni avtomatlashtirish uchun dasturlash ko'nikmalarini rivojlantirish, ayniqsa Python va R kabi dasturlash tillarida. Lingvistik tahlil natijalarini vizualizatsiya qilish va ularni to'g'ri talqin qilishni o'rgatish. Lingvistik tadqiqotlarda statistik ma'lumotlarning ahamiyatini tushunish va ulardan foydalanishni shakllantirish.</p> <p>Mazkur fan dasturi xalqaro tan olingan reytinglarda birinchi top 300 talik ro'yxatga kiruvchi Boxumdagi Rur universiteti (Ruhr University Bochum 249 ARWU xorijiy tajribasini inobatga olgan holda takomillashtirildi.</p> <p>https://vvz.ruhr-uni-bochum.de/campus/all/module.asp?gguid=0x7A15936B26FB496AA558B6B9111F301B</p>					

3.

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

III.1. Fan tarkibiga quyidagi mavzular kiradi:

1-mavzu. Til statistikasiga kirish, Lingvostatistikaning asosiy tushunchalari

Lingvistik statistikaning asosiy tushunchalari va vazifalari.

Tilshunoslikda statistikasi qo'llash: matnlarni tahlil qilishdan tortib mashina tarjimasigacha.

Korpus lingvistikasi va uning statistika bilan bog'liqligi.

Mavzu bo'yicha ilmiy maqolalarni ko'rib chiqish.

2- mavzu. Tilshunoslikda ma'lumotlar turlari

Matnlar korpusi: izohli va izohsiz ma'lumotlar.

Lingvistik birliklarning turlari: so'zlar, morfemalar, sintaktik qurilmalar.

Metrik va kategorial ma'lumotlar.

Korpus va uning annotatsiyasi bilan tanishish.

3- mavzu. Til ma'lumotlarini chastotali tahlil qilish

So'z chastotasi va leksik variativlik tushunchasi.

Chastotali lug'atlar tuzish.

Sipf qonuni va uning tilshunoslikda qo'llanilishi.

Kichik korpus asosida chastotali lug'at tuzish.

4- mavzu. N-grammalar va ularning qo'llanilishi

So'zlar ketma-ketligini tahlil qilish uchun N-gramm modellari.

Matndagi bigramm va trigrammlarga misollar.

N-gramm modellar yordamida keyingi so'zni bashorat qilish.

Korpusni tahlil qilish uchun bigramm modelini amalga oshirish

5- mavzu. Kollokatsiyalar va ularning tahlili

Kollokatsiya tushunchasi.

O'zaro ma'lumot (MI) va turg'un so'z birikmalarini tahlil qilish uchun boshqa chora-tadbirlar.

O'zaro ma'lumotni hisoblashga misollar.

MI yordamida matn korpusidagi kollokatsiyalarni topish.

6- mavzu. Matnlarni qiyosiy tahlili.

Chastota va statistik tavsiflar bo'yicha matnlarni taqqoslash usullari.

Matnlarni qiyosiy tahlil qilish usullarini egallash.

Turli turdagi matnlar (badiiy, ilmiy, publitsistik) bilan ishlash ko'nikmalarini rivojlantirish.

7- mavzu. Matndan namunalar hajmini aniqlash va kuzatish xatolari

Matn tahlilida kuzatuv xatolarini tasniflash.

Tahlil natijalariga turli xil xatolarning ta'sirini baholash.

Matnli ma'lumotlar sifatini baholash vositalarini o'zlashtirish

8- mavzu. Tilshunoslikda ehtimollik modellari

Ehtimollik modellari: unigrammalar, bigrammalar va trigrammalar.

Markov zanjirlari va ularning tilshunoslikda qo'llanilishi.

Ehtimollik modellari asosida bashorat qilishga misollar.

Bigramm model yordamida matnni bashorat qilish.

9-mavzu. Yashirin markov modellari Hidden Markov Model

- Yashirin Markov modellari asoslari (YMM).

- HMMning tilshunoslikda qo'llanilishi: POS-tegging, sintaktik tahlil.

HMM da ehtimolliklarni hisoblashga misollar.

Matnni avtomatik teglash uchun HMM ni amalga oshirish.

10-mavzu.

Bayes modellari

Bayes teoremasining asoslari.

matn tahlili uchun bayes klassifikatori.

Matnli tasniflash uchun Bayes klassifikatorini amalga oshirish.

11- mavzu. Latent-semantik tahlil (LSA)

LSA metodining asoslari.

So'zlar va hujjatlar o'rtasidagi ma'noviy aloqalarni tahlil qilishda qo'llash.

Matnlarni klasterlashda LSA dan foydalanishga misol.

Matnlar to'plami uchun LSA tuzish va mavzularni tahlil qilish.

12- mavzu. Tematik modellashtirish (LDA)

Dirixle latent taqsimoti (LDA).

Matnlardagi mavzularni aniqlash uchun LDA dan foydalanish.

Modelni katta matn korpuslariga qo'llash.

Yangilik maqolalari korpusidagi mavzularni tahlil qilish uchun LDAni qo'llash.

13- mavzu. So'zlik modellari va vektorli tasavvurlar

Word2Vec, GloVe va FastText BERT va Transformer ga kirish.

So'zlarning vektor ko'rinishi va ularning modellarda qo'llanilishi.

Vektorli modellardan foydalangan holda so'zlarning semantik o'xshashligiga misollar.

Word2Vec modelini yasash va so'z vektorlarini tahlil qilish.

14- mavzu. Katta ma'lumotlarga asoslangan statistik tahlil

Katta korpus ma'lumotlarini tahlil qilish uchun statistik usullarni qo'llash.

Masshtablashtirish muammolari va ularning yechimlari.

Katta matn massivlari tahliliga misollar.

15- mavzu. Lingvistik statistikada mashinali o'qitish

Til ma'lumotlarini qayta ishlash va tahlil qilish uchun mashinali o'qitish usullaridan foydalanish.

16- mavzu. Korpus lingvistikasi va statistik tahlil, tabiiy tilni qayta ishlash (NLP) usullari

Korpus tuzish, tozalash va oldindan qayta ishlash. Annotatsiyalangan korpuslar asosida statistik tahlil

17- mavzu. Til modellari va neyron tarmoqlar (RNN, Transformer)

Klassik va zamonaviy neyron tarmoqlar asosidagi til modellari Tilni avtomatik tushunish va yaratishda statistik yondashuvlar

18- mavzu. Tilshunoslikda regressiya, Klasterlash usullari, Dimensiyani qisqartirish (PCA, t-SNE) usullari

Statistik regressiya modellari (masalan, logistik regressiya) So'zlar va matnlarni klasterlash orqali tematik tahlil qilish

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

"Statistika" fani bo'yicha amaliy mashg'ulotlar uchun quyidagi mavzular tavsiya etiladi:

1-mavzu. Til statistikasiga kirish, Lingvostatistikaning asosiy tushunchalari

Lingvistik statistikaning asosiy tushunchalari va vazifalari.

Tilshunoslikda statistikaning qo'llash: matnlarni tahlil qilishdan tortib mashina tarjimasigacha.

Korpus lingvistikasi va uning statistika bilan bog'liqligi.

Mavzu bo'yicha ilmiy maqolalarni ko'rib chiqish.

2- mavzu. Tilshunoslikda ma'lumotlar turlari

Matnlar korpusi: izohli va izohsiz ma'lumotlar.

Lingvistik birliklarning turlari: so'zlar, morfemalar, sintaktik qurilmalar.

Metrik va kategorial ma'lumotlar.

Korpus va uning annotatsiyasi bilan tanishish.

3- mavzu. Til ma'lumotlarini chastotali tahlil qilish

So'z chastotasi va leksik variativlik tushunchasi.

Chastotali lug'atlar tuzish.

Sipf qonuni va uning tilshunoslikda qo'llanilishi.

Kichik korpus asosida chastotali lug'at tuzish.

4- mavzu. N-grammalar va ularning qo'llanilishi

So'zlar ketma-ketligini tahlil qilish uchun N-gramm modellari.

Matndagi bigramm va trigrammlarga misollar.

N-gramm modellar yordamida keyingi soʻzni bashorat qilish.
Korpusni tahlil qilish uchun bigramm modelini amalga oshirish

5- mavzu. Kollokatsiyalar va ularning tahlili

Kollokatsiya tushunchasi.

Oʻzaro maʼlumot (MI) va turgʻun soʻz birikmalarini tahlil qilish uchun boshqa chora-tadbirlar.

Oʻzaro maʼlumotni hisoblashga misollar.

MI yordamida matn korpusidagi kollokatsiyalarni topish.

6- mavzu. Matnlarni qiyosiy tahlili.

Chastota va statistik tavsiflar boʻyicha matnlarni taqqoslash usullari.

Matnlarni qiyosiy tahlil qilish usullarini egallash.

Turli turdagi matnlar (badiiy, ilmiy, publitsistik) bilan ishlash koʻnikmalarini rivojlantirish.

7- mavzu. Matndan namunalarni hajmini aniqlash va kuzatish xatolari

Matn tahlilida kuzatuv xatolarini tasniflash.

Tahlil natijalariga turli xil xatolarning taʼsirini baholash.

Matnli maʼlumotlar sifatini baholash vositalarini oʻzlashtirish

8- mavzu. Tilshunoslikda ehtimollik modellari

Ehtimollik modellari: unigrammalar, bigrammalar va trigrammalar.

Markov zanjirlari va ularning tilshunoslikda qoʻllanilishi.

Ehtimollik modellari asosida bashorat qilishga misollar.

Bigramm model yordamida matnni bashorat qilish.

9- mavzu. Yashirin markov modellari Hidden Markov Model

• Yashirin Markov modellari asoslari (YMM).

• HMMning tilshunoslikda qoʻllanilishi: POS-tegging, sintaktik tahlil.

HMM da ehtimolliklarni hisoblashga misollar.

Matnni avtomatik teglash uchun HMM ni amalga oshirish.

10- mavzu.

Bayes modellari

Bayes teoremasining asoslari.

matn tahlili uchun bayes klassifikatori.

Matnli tasniflash uchun Bayes klassifikatorini amalga oshirish.

11- mavzu. Latent-semantik tahlil (LSA)

LSA metodining asoslari.

Soʻzlar va hujjatlar oʻrtasidagi maʼnoviy aloqalarni tahlil qilishda qoʻllash.

Matnlarni klasterlashda LSA dan foydalanishga misol.

Matnlar to'plami uchun LSA tuzish va mavzularni tahlil qilish.

12- mavzu. Tematik modellashtirish (LDA)

Dirixle latent taqsimoti (LDA).

Matnlardagi mavzularni aniqlash uchun LDA dan foydalanish.

Modelni katta matn korpuslariga qo'llash.

Yangilik maqolalari korpusidagi mavzularni tahlil qilish uchun LDAni qo'llash.

13- mavzu. So'zlik modellari va vektorli tasavvurlar

Word2Vec, GloVe va FastText BERT va Transformer ga kirish.

So'zlarning vektor ko'rinishi va ularning modellarda qo'llanilishi.

Vektorli modellardan foydalangan holda so'zlarning semantik o'xshashligiga misollar.

Word2Vec modelini yasash va so'z vektorlarini tahlil qilish.

14- mavzu. Katta ma'lumotlarga asoslangan statistik tahlil

Katta korpus ma'lumotlarini tahlil qilish uchun statistik usullarni qo'llash.

Masshtablashtirish muammolari va ularning yechimlari.

Katta matn massivlari tahliliga misollar.

15- mavzu. Lingvistik statistikada mashinali o'qitish

Til ma'lumotlarini qayta ishlash va tahlil qilish uchun mashinali o'qitish usullaridan foydalanish.

16- mavzu. Korpus lingvistikasi va statistik tahlil, tabiiy tilni qayta ishlash (NLP) usullari

Korpus tuzish, tozalash va oldindan qayta ishlash. Annotatsiyalangan korpuslar asosida statistik tahlil

17- mavzu. Til modellari va neyron tarmoqlar (RNN, Transformer)

Klassik va zamonaviy neyron tarmoqlar asosidagi til modellari Tilni avtomatik tushunish va yaratishda statistik yondashuvlar

18- mavzu. Tilshunoslikda regressiya, Klasterlash usullari, Dimensiyani qisqartirish (PCA, t-SNE) usullari

Statistik regressiya modellari (masalan, logistik regressiya) So'zlar va matnlarni klasterlash orqali tematik tahlil qilish

IV. Mustaqil ta'lim*

1.1. 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.

	<p>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.</p> <p>“Statistika” fanidan mustaqil ta'lim uchun tavsiya etiladigan mavzular:</p> <p>Tilshunoslikda ma'lumotlar turlari, Til ma'lumotlarini chastotali tahlil qilish, N-grammalar va ularning qo'llanilishi, Kollokatsiyalar va ularning tahlili, Matnlarni qiyosiy tahlili, Matndan namunalar hajmini aniqlash va kuzatish xatolari, Tilshunoslikda ehtimollik modellari, Yashirin markov modellari Hidden Markov Model, Bayes modellari, Latent-semantik tahlil (LSA), Tematik modellashtirish (LDA), So'zlik modellari va vektorli tasavvurlar, Katta ma'lumotlarga asoslangan statistik tahlil, Lingvistik statistikada mashinali o'qitish</p> <p><i>* Izoh: Mustaqil ta'lim ishi mavzulari fan doirasida o'zgarishi mumkin.</i></p>
4.	<p>V. Fan o'qitilishining natijalari (shakllanadigan kompetensiyalar)</p> <p>Lingvistik Ma'lumotlarni Tahlil Qilish Talabalar tilshunoslik va lingvistik tadqiqotlarda statistik usullarni qo'llashni biladilar. Til o'zgarishlari va tilshunoslik masalalarini statistik tahlil qilish qobiliyatiga ega bo'ladilar.</p> <p>Ma'lumotlarni To'plash va Tashkil etish Talabalar lingvistik tadqiqotlar uchun ma'lumotlarni yig'ish, tasniflash va tuzish ko'nikmalariga ega bo'ladilar. Anketalar, so'rovlar va boshqa usullar yordamida ma'lumotlarni samarali yig'ishni o'rganadilar.</p> <p>Statistik Ko'rsatkichlarni Hisoblash Talabalar deskriptiv statistika ko'rsatkichlarini (o'rtacha qiymat, median, mod, dispersiya va h.k.) hisoblashni biladilar va ularni lingvistik ma'lumotlarga qo'llay oladilar.</p> <p>Regressiya va Korrelatsiya Tahlillari Talabalar lingvistik tadqiqotlarda regressiya va korrelatsiya tahlillarini amalga oshirish qobiliyatiga ega bo'ladilar. Til o'zgarishlari va boshqa faktorlar orasidagi bog'liqliklarni tahlil qilishni o'rganadilar.</p>
5.	<p>VI. TA'LIM TEXNOLOGIYALARI VA METODLARI:</p> <p>Interaktiv ta'lim metodlari: Talabalarga lingvistik statistikaning nazariy asoslarini va metodologiyasini tushuntirish. Ma'ruzalar statistik tahlilning til tadqiqotlarida qanday qo'llanilishini misollar bilan ko'rsatadi.</p> <p>Masalalar va misollar yechish: Talabalar lingvistik ma'lumotlarni statistik usullar yordamida tahlil qilish, asosiy statistik ko'rsatkichlarni hisoblash va natijalarni tahlil qilishni amaliy mashg'ulotlarda</p>

	o'rganadilar. Grafik va diagrammalar yaratish orqali ma'lumotlarni vizualizatsiya qilish, ular orqali ma'lumotlarni tahlil qilish va sharhlash.
6.	<p style="text-align: center;">VII. Kreditlarni olish uchun talablar:</p> <p>Topshiriqni muvaffaqiyatli bajarish uchun asosiy ko'nikmalar:</p> <p>Statistik tahlil ko'nikmalari: Lingvistik ma'lumotlarga asoslangan statistik tahlilni mustaqil o'tkazish va natijalarni to'g'ri talqin qilish ko'nikmasi.</p> <p>Ehtimollik modellarini qo'llash: Markov zanjirlari, Zipf qonuni va boshqa ehtimollik modellari yordamida til hodisalarini modellashtirish va ularni lingvistik ma'lumotlarga tatbiq qilish qobiliyati.</p> <p>Ma'lumotlarni to'plash va tartibga solish: Lingvistik korpuslardan ma'lumotlar yig'ish, ularni to'g'ri formatlash va tahlil qilish uchun tayyorlash ko'nikmasi.</p> <p>Korpus lingvistikasidan foydalanish: Korpus lingvistikasining asosiy vositalarini qo'llash, matnlarni qayta ishlash va statistik tahlilga tayyorlash ko'nikmalari.</p> <p>Dasturiy vositalardan foydalanish: Python, R yoki boshqa statistik dasturlash tillarida tilshunoslik ma'lumotlarini tahlil qilish uchun skriptlar yozish va dasturlarni amaliyotda qo'llash ko'nikmasi.</p>
7.	<p style="text-align: center;">VIII. ASOSIY VA QO'SHIMCHA O'QUV ADABIYOTLAR HAMDA AXBOROT MANBALARI</p> <p style="text-align: center;">Asosiy adabiyotlar</p> <ol style="list-style-type: none"> 1. Statistics for Linguists: An Introduction Using R Bodo Winter First published 2020 by Routledge 52 Vanderbilt Avenue, New York, NY 10017 2. Васильев Юлий Обработка естественного языка. Python и spaCy на практике. — СПб.: Питер, 2021. — 256 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-4461-1506-8 <p style="text-align: center;">Tavsiya qilinadigan qo'shimcha adabiyotlar</p> <ol style="list-style-type: none"> 3. Mirziyoyev Sh.M. Hozirgi zamon va Yangi O'zbekiston. - Toshkent: O'zbekiston, 2024. 4. Mirziyoyev Sh.M. Yangi O'zbekistonda taraqqiyot strategiyasi asosida demokratik islohotlar yo'lini qat'iy davom ettiramiz. 6-jild. - Toshkent: O'zbekiston, 2023 5. Микулин Н.А. Теория вероятностей и математическая статистика учеб пособие 2002 г. стр -192. 6. Захаров В.П., Богданова С.Ю. Корпусная лингвистика 3-е изд. Издательство С-Петербург 2020.-234 с.

7. Алефиренко Н.Ф. Современные проблемы науки о языке. Издательство "ФЛИНТА" 2004 г.
8. Головин Б.Н. Язык и статистика М., «Просвещение» 1970 г
9. Бослаф С. Статистика для всех. август 2015 года Объем, стр: 588 ISBN: 978-5-97060-537-0
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

Scopusdan maqolalar

1. A Statistical Analysis of Summarization Evaluation Metrics Using Resampling Methods Open Access Daniel Deutsch, Rotem Dror, Dan Roth Transactions of the Association for Computational Linguistics (2021) 9: 1132–1146.
<https://doi.org/10.1162/tacl.a.00417>
<https://direct.mit.edu/tacl/article/doi/10.1162/tacl.a.00417/107833/A-Statistical-Analysis-of-Summarization-Evaluation>
2. Data-driven Model Generalizability in Crosslinguistic Low-resource Morphological Segmentation Open Access Zoey Liu, Emily Prud'hommeaux Transactions of the Association for Computational Linguistics (2022) 10: 393–413.
<https://doi.org/10.1162/tacl.a.00467>
<https://direct.mit.edu/tacl/article/doi/10.1162/tacl.a.00467/110437/Data-driven-Model-Generalizability-in>

Axborot manbalar:

1. <https://www.r-project.org/>
2. <https://www.w3schools.com/statistics/index.php>

8. Fanning o'quv dasturi Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabida ishlab chiqilgan va 2025 yil "18" 06 dagi 29 -sonli bayonnomasi bilan ma'qullangan.

O'quv dasturi Toshkent davlat sharqshunoslik universiteti Kengashining 2025 yil "28" 06 dagi 11 -- 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. Xamidov Xayrillo Xudoyorovich- Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi professori, f.f.d.

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

100