

✓

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




TABIIY TILNI QAYTA ISHLASH (NLP)
FANINING O‘QUV DASTURI

Bilim sohasi: 200000 – San’at va gumanitar soha

Ta’lim sohasi: 230000 – Tillar

Ta’lim mutaxassisligi: 70230401 – Kompyuter lingvistikasi


Sharq xalqlari tillari va adabiyoti
instituti direktori:



(imzo)

X.V.Mirzaxmedova

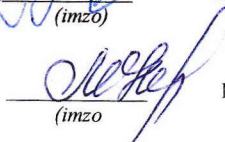
Tarjimashunoslik, tilshunoslik va
xalqaro jumalistika oliy maktabi boshlig'i:



(imzo)

S.T.Mustafayeva

ARM boshlig'i:



(imzo)

M.Yuldasheva

Fan/modul kodi TTQI 1104		O'quv yili 2025/2026	Semestr 1	YeCTS - Kreditlar 4		
Fan/modul turi Majburiy		Ta'lim tili O'zbek		Haftadagi dars soatlari 4		
1.	Fanning nomi	Auditoriya mashg'ulotlari (soat)	Ma'ruza	Amaliy	Mustaqil ta'lim (soat)	Jami yuklama (soat)
	Tabiiy tilni qayta ishlash (NLP)	48	24	24	72	120
2.	I. Fanning mazmuni:					
<p>Mazkur fanning maqsadi magistr'larga til texnologiyasida kompyuter texnologiyalaridan unumli foydalanish hamda sohaga doir masalalar (tillarni o'qitish, bilimlarni baholash, matnlarni tahrir qilish, tarjima dasturlarini yaratish, matnlarni qayta ishlash, matnlar korpusini yaratish, matnlarni ovozlashtirish hamda til va nutq hodisalarini modellashtirish kabilarni) ham lingvistik, ham kompyuter texnologiyalari yordamida hal qilish, shuningdek, mantiqiy to'g'ri, izchil fikrlashga o'rgatish, ularda nazariy bilimlarni shakllantirish va o'z yo'nalishlari bo'yicha amalda qo'llashga oid ko'nikmalarni hosil qilishni nazarda tutadi.</p> <p>Ushbu kurs magistr'larda tabiiy tilni qayta ishlash bo'yicha chuqur bilimlarni shakllantirish, kompyuter lingvistikasi tamoyillarini o'zlashtirish, matematik modellar tuzish bo'yicha bilimlarni oqilona qo'llash qobiliyatini rivojlantirish, matematik mantiq asosida fikrlashga o'rgatish, til va adabiyotga doir kompyuter dasturlarini yaratish, algoritim va kompyuter dasturlari asoslarini berish kabi vazifalarni qo'yadi. Ushbu maqsadga erishish uchun fan talabalarni nazariy bilimlar, amaliy ko'nikmalalar ilmiy jarayonlarga uslubiy yondashuv hamda ilmiy dunyoqarashini shakllantirish vazifalarini bajaradi.</p> <p>Fanning vazifasi – ushbu kurs magistr'larda NLP bo'yicha chuqur bilimlarni shakllantirish, kompyuter lingvistikasi tamoyillarini belgilash, til modellarini tuzish bo'yicha bilimlarni to'g'ri yo'naltirish, formallashtirish sohasidagi bilimlarni rejali qo'llash qobiliyatini rivojlantirish, matematik mantiq asosida fikrlashga o'rgatish, filologiyaga doir kompyuter dasturlarini yaratish, algoritim va kompyuter dasturlarining lingvistik ta'minotini yaratish asoslarini o'zlashtirish va ularning mazmun-mohiyatini ochib berish.</p> <p>Mazkur fan dasturi xalqaro tan olingan reytinglarda birinchi top 300 talik ro'yxatga kiruvchi Stanford Universiteti (AQShning Stanford Universitetida 2 ARWU)xorijiy tajribasini inobatga olgan holda takomillashtirildi.</p> <p>https://web.stanford.edu/class/cs224n/</p>						

3.

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

II.1. Fan tarkibiga quyidagi mavzular kiradi:

1-mavzu. "Tabiiy tilni qayta ishlash" faniga kirish

Fanning maqsadi va vazifasi. NLP fan sifatida rivojlanishida nazariy va amaliy tadqiqotlar. NLPning boshqa fanlar bilan aloqasi

2-mavzu. NLP ning asosiy yo'nalishlari

Kompyuter leksikografiyasi. Mashina tarjimasini. Korpus lingvistikasi. Kompyuter lingvodidaktikasi. NLP va lingvistik texnologiya

3-mavzu. NLP uchun fundamental algoritmlar tahlili. Lingvistik bilimlar bazasi

Regular ifodalar. n-gram til modeli. NaiveBayes algoritmi. Matn klassifikatsiyasi. Sentiment tahlil. Grammatik aksiomalar. Formal til nazariyasi. Formallashtirish tamoyillari va lingvistik bilimlar bazasidan foydalanish

4-mavzu. Lingvistik protsessorning grammatik komponentlari

Morfologik ma'lumotlar bazasi, semantik va sintaktik ma'lumotlar bazasi. Semantik vector

5-mavzu. Lingvistik bilimlar bazasini loyihalashning dasturiy bazasi

SQL ma'lumotlar bazasi, Access ma'lumotlar bazasi va so'rov yaratish

6-mavzu. Matnni lingvistik tahlil qilish algoritmlari. Sintaktik tahlil (Parsing)

O'zbek tilining morfologik qonuniyatlari. Morfologik ma'lumotlar bazasini yaratish. Matnning morfologik analizi. Sintaktik strukturalar. Grammatik nazariyalar. Matnning sintaktik modellari. Sintaktik aloqa munosabatlari. Diskurs va matn segmenti. Sintaktik tahlil algoritmlari

7-mavzu. Korpus lingvistikasi va tilni avtomatlashtirilgan tahlil qilish

Ushbu mavzu doirasida talabalar til korpuslari tushunchasi, ularning tuzilishi, turli yozma va og'zaki matnlar to'plamidan iboratligi bilan tanishadilar. Kursda korpuslar yordamida tilning morfologik, sintaktik va semantik tahlili, statistik metodlar bilan so'z va iboralar chastotasi, kontekstli ma'nolarni aniqlash, hamda avtomatlashtirilgan til tahlil vositalarini yaratish ko'nikmalari o'rgatiladi. Shu bilan birga Python va NLP kutubxonalari (nltk, spaCy, pandas) yordamida korpusni qayta ishlash, tokenizatsiya, lemmatizatsiya, so'zlarni aniqlash va ma'lumotlarni vizual tarzda tahlil qilish amaliy mashqlar bilan boyitiladi.

8-mavzu. Kompyuter leksikografiyasi. Mashina tarjimasini

Lug'at tipologiyasi. Tezaurus, ontologiya bilan ishlash. Terminologik lug'atlar. Maxsus dasturlar va lingvistik instrumentariyalar. Matnlarni qayta ishlash tizimida lug'atlardan foydalanish. Lingvistik (grammatik) lug'atlar Mashina tarjimasini tarixi: rivojlanish evolyutsiyasi. Mashina tarjimasini tizimlari va texnologiyasi. Qoidaga asoslangan, korpusga asoslangan mashina tarjimasini texnologiyasi. Neyromashina tarjimasini. Mashina tarjimasida lingvistik muammolar. Parallel korpuslar. Tarjima xotirasini yaratish masalalari

9-mavzu. Kompyuter lingvistikasida metodlar. Kompyuterda matnlarni qayta ishlash tizimlari.

Statistik metod. Modellashtirish metodi. Iyerarxik metod. Matn generatsiyasi. Qidiruv va ekspert tizimi haqida tushuncha. Dialogning kompyuterga oid modellari. Matnlarni ovozlashirish va ovozli ma'lumotlarni matn holatiga o'tkazish. Nutq sintezatori. Matnlarni avtomatik tahrirlash.

10 - mavzu Tayyorlangan til modellari va ularning NLPdagi roli: amaliy qo'llanilishi va integratsiyasi

Tayyorlangan til modellarining ishlash tamoyillari, ular yordamida matni avtomatik tushunish va javob berish tizimlari, yirik model va arxitekturalarning NLPdagi o'rni (masalan, GPT, BERT).

11 - mavzu Semantik tahlil va ma'lumotni chiqarish. Sentiment analizi

Semantik tahlilning metodlari, matndan ma'lumotni aniqlash va chiqarish usullari, til modellari va ma'lumotlar orasidagi bog'lanishni aniqlash. Matnlar asosida muallifning yoki fikrni ifodalovchi his-tuyg'ularni aniqlash va tasniflash. Sentiment analizi texnologiyalari, algoritmlar va ularning amaliy qo'llanilishi.

12 - mavzu Dialog tizimlari va tilga asoslangan botlar

Tabiiy tilni qayta ishlashda dialog tizimlarini yaratish, tilga asoslangan chatbotlar va ularning ishlab chiqilishi, foydalanuvchi bilan o'zaro aloqada NLP texnologiyalarining qo'llanilishi.

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

"Tabiiy tilni qayta ishlash (NLP)" fani bo'yicha amaliy mashg'ulotlar uchun quyidagi mavzular tavsiya etiladi:

1-mavzu. "Tabiiy tilni qayta ishlash" faniga kirish

Fanning maqsadi va vazifasi. NLP fan sifatida rivojlanishida nazariy va amaliy tadqiqotlar. NLPning boshqa fanlar bilan aloqasi

2-mavzu. NLP ning asosiy yo'nalishlari

Kompyuter leksikografiyasi. Mashina tarjimasini. Korpus lingvistikasi. Kompyuter lingvodidaktikasi. NLP va lingvistik texnologiya

3-mavzu. NLP uchun fundamental algoritmlar tahlili. Lingvistik bilimlar bazasi

Regular ifodalar. n-gram til modeli. NaiveBayes algoritmi. Matn klassifikatsiyasi. Sentiment tahlil. Grammatik aksiomalar. Formal til nazariyasi. Formallashtirish tamoyillari va lingvistik bilimlar bazasidan foydalanish

4-mavzu. Lingvistik protsessorning grammatik komponentlari

Morfologik ma'lumotlar bazasi, semantik va sintaktik ma'lumotlar bazasi. Semantik vector

5-mavzu. Lingvistik bilimlar bazasini loyihalashning dasturiy bazasi

SQL ma'lumotlar bazasi, Access ma'lumotlar bazasi va so'rov yaratish

6-mavzu. Matnni lingvistik tahlil qilish algoritmlari. Sintaktik tahlil (Parsing)

O'zbek tilining morfologik qonuniyatlari. Morfologik ma'lumotlar bazasini yaratish. Matnning morfologik analizi. Sintaktik strukturalar. Grammatik nazariyalar. Matnning sintaktik modellari. Sintaktik aloqa munosabatlari. Diskurs va matn segmenti. Sintaktik tahlil algoritmlari

7-mavzu. Korpus lingvistikasi

Korpus taksonomiyasi. Korpus lingvistikasining nazariy va amaliy masalalari. Korpus yaratish texnologiyasi. Metadata. Korpus ilmiy tadqiqot obyekti sifatida. Korpusning turli sohalarda qo'llanishi

8-mavzu. Kompyuter leksikografiyasi. Mashina tarjimasi

Lug'at tipologiyasi. Tezaurus, ontologiya bilan ishlash. Terminologik lug'atlar. Maxsus dasturlar va lingvistik instrumentariylar. Matnlarni qayta ishlash tizimida lug'atlardan foydalanish. Lingvistik (grammatik) lug'atlar Mashina tarjimasi tarixi: rivojlanish evolyutsiyasi. Mashina tarjimasi tizimlari va texnologiyasi. Qoidaga asoslangan, korpusga asoslangan mashina tarjimasi texnologiyasi. Neyromashina tarjimasi. Mashina tarjimasida lingvistik muammolar. Parallel korpuslar. Tarjima xotirasini yaratish masalalari

9-mavzu. Kompyuter lingvistikasida metodlar. Kompyuterda matnlarni qayta ishlash tizimlari.

Statistik metod. Modellashtirish metodi. Iyerarxik metod. Matn generatsiyasi. Qidiruv va ekspert tizimi haqida tushuncha. Dialogning kompyuterga oid modellari. Matnlarni ovozlashtirish va ovozli ma'lumotlarni matn holatiga o'tkazish. Nutq sintezatori. Matnlarni avtomatik tahrirlash.

10 - mavzu Tayyorlangan til modellari va ularning NLPdagi roli

Tayyorlangan til modellarining ishlash tamoyillari, ular yordamida matnni avtomatik tushunish va javob berish tizimlari, yirik model va arxitekturalarning NLPdagi o'rni (masalan, GPT, BERT).

11 - mavzu Semantik tahlil va ma'lumotni chiqarish. Sentiment analizi

Semantik tahlilning metodlari, matndan ma'lumotni aniqlash va chiqarish usullari, til modellari va ma'lumotlar orasidagi bog'lanishni aniqlash. Matnlar asosida muallifning yoki fikrni ifodalovchi his-tuyg'ularni aniqlash va tasniflash. Sentiment analizi texnologiyalari, algoritmlar va ularning amaliy qo'llanilishi.

12 - mavzu Dialog tizimlari va tilga asoslangan botlar

Tabiiy tilni qayta ishlashda dialog tizimlarini yaratish, tilga asoslangan chatbotlar va ularning ishlab chiqilishi, foydalanuvchi bilan o'zaro aloqada NLP texnologiyalarining qo'llanilishi.

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.

"NLP algoritmlari" fanidan mustaqil ta'lim uchun tavsiya etiladigan mavzular:

NLP ning asosiy yo'nalishlari NLP uchun fundamental algoritmlar tahlili Lingvistik bilimlar bazasi Lingvistik protsessorning grammatik komponentlari Lingvistik bilimlar bazasini loyihalashning dasturiy bazasi Matni lingvistik tahlil qilish algoritmlari. Sintaktik tahlil (Parsing) Korpus lingvistikasi

Kompyuter leksikografiyasi

Mashina tarjimai Kompyuter lingvistikasida metodlar. Kompyuterda matnlarni qayta ishlash tizimlari Tayyorlangan til modellari va ularning NLPdagi roli . Semantik tahlil va ma'lumotni chiqarish . Sentiment analizi. Dialog tizimlari va tilga asoslangan botlar

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)**
- turli maqsadlar uchun orborot va intellektual tizimlarining lingvistik va kognitiv komponentlarini yaratish va takomillashtirish (tezaurus, lingvistik ontologiyalar, ma'lumotlar bazalari, bilimlar bazalari, sintaktik tahlil tizimlari parserlar, semantik tahlil dasturlari, morfoanalizatorlar)ni bilishi; tilshunoslik resurslarini ishlab chiqish va takomillashtirish (matnti lug'atlar, fonetik- leksik, terminologik ma'lumotlar bazalari)ni bilishi; tizimlari yordamida amaliyotda og'zakj, youna va avtomatik tarjima qilish tizimini ishlab chiqish va joriy eta olishi, avtomatik tarjima vositalarini qo'llay olish, mashina tarjimai nazariyasini tahlil qilish va tadqiq etish, tarjimon dasturlari lingvistik ta'minoti va ma'lumotlar bazalari, dasturiy ta'minotini yaratish va takomillashtirish *haqida tasavvurga ega bo'lishi kerak;*

	<p>–tabiiy tilni avtomatik qayta ishlash tizimlarining lingvistik komponentlarini (ogʻzaki nutqni sintez qilish va tanib olish, matnни yaratish, avtomatik tarlima, avtomatik referatlash va annotatsiyalash), intellektual tizimlarni ishlab chiqishi; lingvistik bilimlarni formallashtirishda matematik asoslardan foydalana olish va lingvistik strukturalar hamda tilshunoslik bilimlarini tahlil va sintez qila olishi; kompyuter leksikografiyasi vazifalarini bajara olishi: elektron lugʻat bazasini yarata olishi, yangi avlod elektron lugʻatlarini, lingvistik lugʻatlarning mobil ilovalarini yarata olishni bilishi; til modellari va resuslarini tayyorlash, sunʼiy intellekt uchun tabiiy tilni modellashtira bilishi hamda amaliyotga qoʻllay olishi; ovoqli nutqni va tabiiy tildagi yozma matnlarni avtomatik qayta ishlashni nazarda tutuvchi electron tizimlarni <i>bilishi va ulardan foydalana olishi kerak</i>;</p> <p>– parallel korpuslardan foydalana olishi va ularning bilimlar va maʼlumotlar bazalarini yarta olishni bilishi; matnlarni xorijiy tillardan davlat va rasmiy trllarga hamda davlat va rasmiy tillardan chet tillariga malakali tarjima qila olishi; tabiiy tilga ishlov beruvchi zamonaviy dasturlar boʻyicha dolzarb masalalarni tahlil qilish va yoritish <i>kompetensiyasiga ega boʻlishi kerak</i></p>
5.	<p style="text-align: center;">VI. TAʼLIM TEXNOLOGIYALARI VA METODLARI:</p> <ul style="list-style-type: none"> • maʼruzalar; • interfaol keys-stadilar; • seminarlar (mantiqiy fiklash, tezkor savol-javoblar); • guruhlarda ishlash; • taqdimotlarni qilish; • individual loyihalar; <p>jamoa boʻlib ishlash va himoya qilish uchun loyihalar.</p>
6.	<p style="text-align: center;">VII. Kreditlarni olish uchun talablar:</p> <p>Fanga oid nazariy va uslubiy tushunchalarni toʻla oʻzlashtirish, tahlil natijalarini toʻgʻri aks ettira olish, oʻrganilayotgan jarayonlar haqida mustaqil mushohada yuritish va joriy, oraliq nazorat shakllarida berilgan vazifa va topshiriqlarni bajarish, yakuniy nazorat boʻyicha yozma ishni topshirish.</p>
7.	<p style="text-align: center;">VIII. ASOSIY VA QOʻSPHMCHA OʻQUV ADABIYOTLAR HAMDA AXBOROT MANBALARI</p> <p style="text-align: center;">Asosiy adabiyotlar</p> <ol style="list-style-type: none"> 1. Abduraxmonova N. Kompyuter lingvistikasi (darslik). – Toshkent: Nodirabegim, 2021. – 400 b. 2. Jurafsky D., J. H. Martin. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2023. – 620 p. 3. 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

Qo‘shimcha adabiyotlar

4. Mirziyoyev Sh.M. O‘zbekiston Respublikasi Prezidentining PF-6084 sonli farmoni. Mamlakatimizda o‘zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida. – Toshkent. 2020 yil, 20 oktyabr.
5. Mirziyoyev Sh. “Chet tillarini o‘qitish tizimini takomillashtirish chora-tadbirlari” videoselektor yig‘ilishi materiallari. – Toshkent. 2021 yil, 6 may.
6. Rahimov A. Kompyuter lingvistika asoslari. Toshkent: Akademnashr, 2011. 160 b.
7. Polatov A. Kompyuter lingvistikasi Toshkent: Akademnashr, 2014. 500 b.
8. Alexander Clark, Chris Fox, and Shalom Lappin. (Edited) The Handbook of Computational Linguistics and Natural Language Processing. 2010.
9. Srinivasa-Desikan B. Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras. – Packt Publishing Ltd, 2018.
10. Kurdi M. Z. Natural language processing and computational linguistics 2: semantics, discourse and applications. – John Wiley & Sons, 2017. – T. 2
11. Williams Nwagwu The Rise and Rise of Natural Language Processing Research, 1958-2021 Research Square November 29th, 2022
12. Safoora M , Hossein A , Najmeh S., Saeid Sh. Natural language processing (NLP) to facilitate abstract review in medical research: the application of BioBERT to exploring the 20-year use of NLP in medical research. Masoumi et al. Systematic Reviews (2024)

Scopusdan maqolalar

1. A Neighborhood Framework for Resource-Lean Content Flagging Open Access Sheikh Muhammad Sarwar, Dimitrina Zlatkova, Momchil Hardalov, Yoan Dinkov, Isabelle Augenstein, Preslav Nakov. Transactions of the Association for Computational Linguistics (2022) 10: 484–502.
https://doi.org/10.1162/tacl_a_00472
https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00472/110995/A-Neighborhood-Framework-for-Resource-Lean-Content
2. A Knowledge-Enhanced Pretraining Model for Commonsense Story Generation Open Access Jian Guan, Fei Huang, Zhihao Zhao, Xiaoyan Zhu, Minlie Huang Corresponding Author Transactions of the Association for Computational Linguistics (2020) 8: 93–108. https://doi.org/10.1162/tacl_a_00302
https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00302/43540/A-Knowledge-Enhanced-Pretraining-Model-for

Internet ma’lumotlari:

<http://modmorph.turklang.net/ru/>

http://www.promt.com/company/technology/hybrid_technology

<http://www.dialog-21.ru/Archive/Directions/JPTRANSF.htm>

<https://universaldependencies.org/>

<https://www.clarin.eu/resource-families/parallel-corpora>

	https://www.sketchengine.eu/quick-start-guide/parallel-corpus-lesson-7/ http://opus.nlpl.eu/ https://www.llas.ac.uk/resources/gpg/1444
8.	<p>Fanning o'quv dasturi Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabida ishlab chiqilgan va 202<u>5</u> yil "<u>18</u>" <u>06</u> dagi <u>24</u> sonli bayonnomasi bilan ma'qullangan.</p> <p>O'quv dasturi Toshkent davlat sharqshunoslik universiteti Kengashining 202<u>5</u> yil "<u>28</u>" <u>06</u> sonli bayoni bilan tasdiqlangan.</p>
9.	<p>Fan/modul uchun ma'sullar:</p> <p>Dauletov Adilbek Yusupbayevich Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi PhD, dotsent</p> <p>R.B.Allanyazov- Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi o'qituvchisi</p>
10.	<p>Taqrizchilar:</p> <p>S.T.Mustafayeva - Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabi f. f. d., dotsent</p> <p>N.Abdurahmonova - O'zMU, Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasida professori, f.f.d</p>

ms