

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



NLP ALGORITMLARI
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

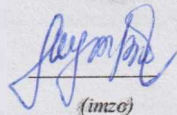
Sharq xalqlari tillari va adabiyoti
instituti direktori:



(imzo)

X.V. Mirzaxmedova

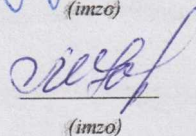
Tarjimonshunoslik, tilshunoslik va
xalqaro jurnalistika oliy maktabi boshlig'i:



(imzo)

S.T. Mustafayeva

Arm boshlig'i:



(imzo)

M. Yuldasheva

Fan/modul kodi NA1406		O'quv yili 2026/2027	Semestr 4	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)
	NLP algoritmlari	72	36	36	108	180
2.	<p align="center">I. Fanning mazmuni:</p> <p>Fanning maqsadi: NLP (Natural Language Processing) algoritmlari fanining asosiy maqsadi — talabalarga tabiiy tilni qayta ishlash usullari va algoritmlarini o'rgatishdir. Bu maqsadga erishish orqali talabalar inson va kompyuter o'rtasida samarali til aloqasini tashkil etuvchi algoritmlarni ishlab chiqish va amaliyotda qo'llash ko'nikmalariga ega bo'ladilar. Shuningdek, ularni tabiiy til matnlaridan foydali ma'lumotlarni olish, tahlil qilish va qo'llash uchun zamonaviy texnologiyalar bilan tanishtirish.</p> <p>Fan vazifasi - Talabalarga tabiiy tilni qayta ishlashning asosiy tushunchalari, tamoyillari va metodlarini tushuntirish. Til modellari, matn tahlili va leksik tahlil kabi jarayonlarni o'rgatish, jumladan, morfologik, sintaktik va semantik tahlil usullarini ko'rsatish. Python va boshqa dasturlash tillarida NLP algoritmlarini ishlab chiqish va ularni amaliyotda qo'llash bo'yicha ko'nikmalarni rivojlantirish. Talabalarni NLP loyihalari uchun matn ma'lumotlarini yig'ish, tayyorlash va tozalash bo'yicha jarayonlar bilan tanishtirish. Tabiiy til matnlarini tahlil qilish jarayonida so'zlarni asosiy shaklga keltirish usullarini o'rgatish. Matnlarni tasniflash, sentiment tahlili, mavzularni aniqlash va klasterlash uchun algoritmlarni qo'llash bo'yicha bilimlarni berish. Talabalarni NLP sohasida amaliy loyihalarni ishlab chiqish va taqdim etish orqali ularning bilim va ko'nikmalarini mustahkamlash. Transformer modellar (BERT, GPT va boshqalar), so'nggi algoritmlar va yondashuvlar bilan tanishtirish.</p> <p>Mazkur fan dasturi xalqaro tan olingan reytinglarda birinchi top 300 talik ro'yxatga kiruvchi Heidelberg University Heidelberg Informatika instituti (Heidelberg University 51 ARWU) xorijiy tajribasini inobatga olgan holda takomillashtirildi. https://backend.uni-heidelberg.de/de/dokumente/modulhandbuch-computerlinguistik-ba-2022-07-27/download</p>					

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

II.1. Fan tarkibiga quyidagi mavzular kiradi:

1-mavzu. Kirish va NLP asoslari

Ushbu mavzu talabalarni tabiiy tilni qayta ishlash (NLP) dunyosiga kiritadi. Darsda NLPning nima ekani, uning asosiy tushunchalari va amaliy qo'llanilishi ko'rib chiqiladi. Talabalar NLP algoritmlarining qanday ishlashi, turli tillarni qayta ishlash va tahlil qilish uchun qanday yondashuvlar mavjudligi haqida umumiy ma'lumot oladilar. Shuningdek, NLPning asosiy qo'llanilish sohalari, jumladan, matn tasnifi, sentiment tahlili, chat-botlar va avtomatik tarjima kabi texnologiyalar bilan tanishadilar.

2- mavzu. Tabiiy tilga oid asosiy tushunchalar

Tabiiy tilga oid asosiy tushunchalar mavzusi tabiiy tilni qayta ishlashning (NLP) asosiy elementlarini o'rgatishga qaratilgan. Bu darsda tilning tarkibiy qismlari (morfologiya, sintaksis, semantika), ularning vazifalari va matnlarni tahlil qilishdagi roli tushuntiriladi. Shuningdek, grammatik tuzilish, lug'at boyligi va til modellari bilan ishlashning asosiy tushunchalari ko'rib chiqiladi. Bu bilimlar NLP algoritmlarini yaratishda zarur bo'lgan bazaviy asosni shakllantiradi.

3- mavzu. Python va NLP kutubxonalari

Ushbu mavzu Python dasturlash tilida tabiiy tilni qayta ishlash (NLP) bo'yicha asosiy kutubxonalar bilan tanishtiradi. Dars davomida talabalar NLTK, spaCy, TextBlob, va Transformers kabi mashhur NLP vositalarini o'rganadilar. Ushbu kutubxonalar matnlarni oldindan qayta ishlash, til modellarini yaratish, matn tahlili va boshqa NLP vazifalarini bajarishda qanday qo'llanilishi haqida amaliy misollar keltiriladi. Bu kutubxonalar bilan tanishish talabalarni NLP bo'yicha dasturlarni tez va samarali yaratishga tayyorlaydi.

4- mavzu. Matnlarni oldindan qayta ishlash

Matnlarni oldindan qayta ishlash - bu tabiiy tilni qayta ishlash (NLP) jarayonining muhim bosqichi bo'lib, xom ma'lumotlarni tushunarli va ishlov berishga yaroqli holatga keltirishga yordam beradi. Ushbu jarayon turli usullarni o'z ichiga oladi, jumladan, tokenlash (matnni so'zlarga yoki gaplarga ajratish), stemming va lemmatizatsiya (so'zlarni asosiy shakliga keltirish), stop-so'zlarni olib tashlash (kam ahamiyatga ega bo'lgan so'zlar) va ma'lumotlarni tozalash (noto'g'ri yoki keraksiz belgilarni olib tashlash). Bu jarayonlar algoritmlarning aniq va samarali ishlashi uchun matnni tartibga solish va optimallashtirishni ta'minlaydi.

5- mavzu. Til modellari: n-gram va ehtimollik usullari

Ushbu mavzuda til modellarining asoslari va ularni yaratish usullari ko'rib chiqiladi. N-gram modellar so'zlarning ketma-ketligini aniqlash orqali matn ichidagi ehtimollik munosabatlarini tahlil qiladi. Bu texnika matnlarni taxmin qilish, yangi matn yaratish va so'zlarni bir-biri bilan bog'lashda ishlatiladi. Dars davomida n-gram modellarining ishlash prinsiplari, ularning amaliy qo'llanilishi va cheklovlari o'rganiladi. Shuningdek, ehtimollik usullaridan foydalanib, til modellari qanday qurilishi haqida tushuncha beriladi.

6- mavzu. Ma'lumotlarni to'plash va tayyorlash

Ushbu mavzu ma'lumotlarni qayta ishlash jarayonining dastlabki bosqichlari haqida tushuncha beradi. Talabalar ma'lumotlarni yig'ishning turli usullari, ularni tahlil qilish uchun tayyorlash jarayonlari bilan tanishadilar. Ushbu bosqichda ma'lumotlar tozalash, formatlash, yo'qolgan qiymatlarni to'ldirish va boshqa oldindan ishlov berish usullari ko'rib chiqiladi. Yaxshi tayyorlangan ma'lumot to'plamlari keyingi tahlil va algoritmlar uchun sifatli natijalarga erishishga yordam beradi.

7- mavzu. Vektorlashtirish usullari

Ushbu mavzuda matnlarni raqamli ko'rinishga o'tkazish usullari o'rganiladi. Vektorlashtirish usullari so'zlarni va matnlarni kompyuter tushinishi uchun raqamli vektorlar sifatida ifodalaydi. Talabalar "Bag of Words" (BoW), "Term Frequency-Inverse Document Frequency" (TF-IDF) va "Word Embeddings" (masalan, Word2Vec, GloVe) kabi usullar bilan tanishadilar. Bu usullar matnlarni tahlil qilish, taqqoslash va sinflarga ajratish uchun zarur bo'lgan asosiy tushunchalarni shakllantirishga yordam beradi.

8- mavzu. Word2Vec, GloVe va boshqa word embeddings

Ushbu darsda talabalar Word2Vec va GloVe kabi so'z embedding usullari bilan tanishadilar. Word embeddings — bu so'zlarni raqamli formatda ifodalash usullari bo'lib, ularning ma'nolari va kontekstlariga mos ravishda joylashuvini hisobga oladi. Dars davomida talabalar shuningdek, boshqa word embedding usullari, masalan, FastText va ELMo bilan tanishadilar. Ushbu usullar so'zlarning kontekstiga qarab turli xil embeddinglar yaratish imkoniyatini taqdim etadi.

9- mavzu. Sintaktik va semantik tahlil

Talabalar dars davomida sintaktik va semantik tahlil usullarini o'rganadilar, ularning nazariy asoslari, texnikalari va amaliy qo'llanilishini ko'rib chiqadilar. O'qitish jarayonida sintaktik va semantik tahlilning NLP dasturlaridagi qo'llanilishi, muammolari va yechimlari haqida muhokama olib boriladi. Bu ko'nikmalar talabalarni tabiiy tilni samarali va aniq tahlil qilishga tayyorlaydi.

10- mavzu. Sentiment tahlili va matn tasnifi

Sentiment tahlili va matn tasnifi, tabiiy tilni qayta ishlash (NLP) sohasining muhim bo'laklaridan biridir. Ushbu jarayonlar matnlarning mohiyatini tushunish va ularni ma'lum kategoriyalarga ajratish imkonini beradi. **Sentiment tahlili**- Bu jarayon matn yoki xabarning ijobiy, salbiy yoki neytral kayfiyatini aniqlashga qaratilgan. **Matn tasnifi**- Bu jarayon matnni oldindan belgilangan sinflarga (masalan, sport, iqtisod, siyosat) ajratish uchun amalga oshiriladi. Sentiment tahlili va matn tasnifi reklama kompaniyalarini baholash, mijozlar fikrlarini o'rganish va ijtimoiy media tahlili kabi ko'plab sohalarda qo'llaniladi. Ushbu jarayonlar orqali kompaniyalar va tashkilotlar o'z mahsulot va xizmatlarini yaxshilash va mijozlar ta'ablarini yanada samarali qondirish imkonini topadilar.

11- mavzu. Klasterlash, mavzularni aniqlash va matnlarni vektorlashtirish

Klasterlash va mavzularni aniqlash — tabiiy tilni qayta ishlash va ma'lumotlar tahlilining muhim jarayonlari bo'lib, katta hajmdagi matnlarni samarali ravishda tahlil qilish imkonini beradi.

Klasterlash — bu o'xshash xususiyatlarga ega bo'lgan ob'ektlarni (masalan, matnlarni) bir guruhga birlashtirish jarayonidir. Ushbu jarayonda K-means yoki Hierarchical Clustering kabi algoritmlar yordamida ma'lumotlar o'zaro yaqinliklariga ko'ra guruhlanadi. Klasterlash yordamida foydalanuvchi fikrlarini yoki sharhlarni o'xshashlik darajasiga ko'ra tahlil qilish va guruhlash mumkin.

Mavzularni aniqlash — matnlar ichidagi asosiy mavzularni aniqlash va tasniflash jarayonidir. Bu jarayon Latent Dirichlet Allocation (LDA) kabi statistik usullar yordamida amalga oshiriladi. Mavzularni aniqlash orqali matnning asosiy ma'no'lari va ularning o'zaro bog'lanishlari tushuniladi.

12- mavzu. NLP'da Deep Learning, Neyron Tarmoqlar va Transformer arxitekturalari

Ushbu mavzu tabiiy tilni qayta ishlash (NLP) sohasida chuqur o'rganish (Deep Learning) va neyron tarmoqlarning asosiy tushunchalarini o'z ichiga oladi. Deep Learning metodlari murakkab va katta hajmdagi ma'lumotlardan mustaqil ravishda o'rganishga qodir, bu esa NLP vazifalarini, masalan, matnni tasniflash, tarjima qilish, va suhbat tizimlarini yaratishda qo'llaniladi. Neyron tarmoqlar, xususan, konvolyutsion (CNN) va takroriy (RNN) tarmoqlar, bu jarayonda o'ziga xos ahamiyatga ega, chunki ular ma'lumotlardagi murakkab aloqalarni aniqlashda samaradorlikni oshiradi. Ushbu mavzu, shuningdek, transfer o'qitish (transfer learning) va pre-trained modellar, masalan, BERT va GPT, yordamida NLPdagi eng so'nggi yutuqlarni ham ko'rib chiqadi. Talabalar NLPda Deep Learning va neyron tarmoqlar yordamida qanday qilib til modellari va ilovalarini yaratishni o'rganadilar.

13- mavzu. Transformer modellar: BERT, GPT va boshqalar, ularni oldindan o'qitish va nozik sozlash

Ushbu mavzu tabiiy tilni qayta ishlash (NLP) sohasida transformator modellarining ahamiyati va ularning turli xillari, jumladan BERT (Bidirectional Encoder Representations from Transformers) va GPT (Generative Pre-trained Transformer) kabi ilg'or modellarni o'rganishga qaratilgan. Transformer modellar ko'plab NLP vazifalarini yechish uchun zamonaviy yondashuv sifatida tanilgan va quyidagi asosiy jihatlar bilan tavsiflanadi: Transformer arxitekturasi, BERT modeli, GPT modeli. Ushbu mavzu talabalarni zamonaviy transformator modellarini o'rganishga va ularning afzalliklari, imkoniyatlari va foydalanish sohaslarini chuqur tushunishga tayyorlaydi. Bu bilimlar talabalarni NLP sohasida eng ilg'or texnologiyalar bilan ishlashga yordam beradi.

14- mavzu. Tili modellari va tilni generatsiya qilish

Ushbu mavzu tabiiy tilni qayta ishlash (NLP) sohasida til modellari va matn generatsiyasi jarayonlariga bag'ishlangan. Tili modellari — bu kompyuterlar uchun tilni tushunishga va yaratishga yordam beruvchi statistik va matematik modellar. Ushbu mavzu quyidagi asosiy jihatlarini o'z ichiga oladi: Tili modellari tushunchasi. Matn generatsiyasi. Transformer modellar. Ushbu mavzu talabalarni til modellari va matn generatsiyasi jarayonlarining

asosiy tushunchalari bilan tanishtiradi va zamonaviy texnologiyalarni qo'llash orqali tilni qanday tushunish va yaratish mumkinligini o'rganishga yordam beradi.

15- mavzu. Chatbotlar va suhbat botlari yaratish

Ushbu mavzu chatbotlar va suhbat botlarini yaratish jarayoniga bag'ishlangan. Chatbotlar — foydalanuvchilar bilan muloqotda bo'lish uchun dasturlashtirilgan dasturlar bo'lib, ular avtomatik ravishda savollarga javob berish, ma'lumot taqdim etish va turli xizmatlar ko'rsatish imkoniyatini beradi. Mavzu talabalarni chatbotlar va suhbat botlarini yaratish jarayonini chuqur tushunishga yordam beradi va ularning dasturlash, dizayn va foydalanuvchi tajribasini oshirish ko'nikmalarini rivojlantiradi.

16- mavzu. Tabiiy til tahlilida ma'lumotlar bilan ishlash

Ushbu mavzu tabiiy tilni qayta ishlash (NLP) jarayonida ma'lumotlar bilan ishlashning muhim jihatlari o'rganishga qaratilgan. Tabiiy til tahlili ko'plab bosqichlarni o'z ichiga oladi, va har bir bosqichda ma'lumotlarni to'plash, tozalash, va tayyorlash jarayonlari katta ahamiyatga ega. Mavzu talabalarni tabiiy til tahlilida ma'lumotlar bilan ishlash jarayonlari haqida chuqur tushuncha berish va amaliy ko'nikmalarni shakllantirishga yordam beradi. Talabalar o'z loyihalarida ma'lumotlarni samarali tayyorlash va tahlil qilish imkoniyatini egallaydilar.

17- mavzu. Amaliy NLP loyihalari: Sentiment tablili loyihasi

Ushbu mavzu sentiment tahlili loyihasini amalga oshirishga bag'ishlangan bo'lib, u tabiiy tilni qayta ishlash (NLP) yordamida matnning his-tuyg'ularini aniqlash va tahlil qilish jarayonini o'rganishni o'z ichiga oladi. Loyihaning asosiy maqsadi turli manbalardan olingan matnlarda ijobiy, salbiy yoki neytral his-tuyg'ularni aniqlashdir. Mavzu talabalarni sentiment tahlilining nazariy va amaliy jihatlari bilan tanishtirib, ularning muayyan loyihalarni amalga oshirish qobiliyatlarini rivojlantiradi. Talabalar loyiha orqali real muammolarni hal qilish va ularni amaliyotda qo'llash imkoniyatiga ega bo'ladilar.

18- mavzu. Amaliy NLP loyihalari: Chatbot yaratish

Ushbu mavzu amaliy tabiiy tilni qayta ishlash (NLP) loyihasi sifatida chatbot yaratishga bag'ishlangan. Chatbotlar — foydalanuvchilar bilan muloqot qilish uchun mo'ljallangan dasturlar bo'lib, ular avtomatik ravishda savollarga javob berish, ma'lumot taqdim etish yoki vazifalarni bajarish imkonini beradi. Mavzu talabalarni chatbotlarni yaratish jarayoni bilan tanishtiradi, ularga amaliy ko'nikmalarni rivojlantirish imkoniyatini beradi va zamonaviy kommunikatsiya texnologiyalari haqida bilimlarni chuqurlashtirishga yordam beradi.

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

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

1-mavzu. Kirish va NLP asoslari

Ushbu mavzu talabalarni tabiiy tilni qayta ishlash (NLP) dunyosiga kiritadi. Darsda NLPning nima ekani, uning asosiy tushunchalari va amaliy qo'llanilishi ko'rib chiqiladi.

Talabalar NLP algoritmlarining qanday ishlashi, turli tillarni qayta ishlash va tahlil qilish uchun qanday yondashuvlar mavjudligi haqida umumiy ma'lumot oladilar. Shuningdek, NLPning asosiy qo'llanilish sohalari, jumladan, matn tasnifi, sentiment tahlili, chat-botlar va avtomatik tarjima kabi texnologiyalar bilan tanishadilar.

2- mavzu. Tabiiy tilga oid asosiy tushunchalar

Tabiiy tilga oid asosiy tushunchalar mavzusi tabiiy tilni qayta ishlashning (NLP) asosiy elementlarini o'rgatishga qaratilgan. Bu darsda tilning tarkibiy qismlari (morfologiya, sintaksis, semantika), ularning vazifalari va matnlarni tahlil qilishdagi roli tushuntiriladi. Shuningdek, grammatik tuzilish, lug'at boyligi va til modellari bilan ishlashning asosiy tushunchalari ko'rib chiqiladi. Bu bilimlar NLP algoritmlarini yaratishda zarur bo'lgan bazaviy asosni shakllantiradi.

3- mavzu. Python va NLP kutubxonalari

Ushbu mavzu Python dasturlash tilida tabiiy tilni qayta ishlash (NLP) bo'yicha asosiy kutubxonalar bilan tanishtiradi. Dars davomida talabalar NLTK, spaCy, TextBlob, va Transformers kabi mashhur NLP vositalarini o'rganadilar. Ushbu kutubxonalar matnlarni oldindan qayta ishlash, til modellarini yaratish, matn tahlili va boshqa NLP vazifalarini bajarishda qanday qo'llanilishi haqida amaliy misollar keltiriladi. Bu kutubxonalar bilan tanishish talabalarni NLP bo'yicha dasturlarni tez va samarali yaratishga tayyorlaydi.

4- mavzu. Matnlarni oldindan qayta ishlash

Matnlarni oldindan qayta ishlash - bu tabiiy tilni qayta ishlash (NLP) jarayonining muhim bosqichi bo'lib, xom ma'lumotlarni tushunarli va ishlov berishga yaroqli holatga keltirishga yordam beradi. Ushbu jarayon turli usullarni o'z ichiga oladi, jumladan, tokenlash (matnni so'zlarga yoki gaplarga ajratish), stemming va lemmatizatsiya (so'zlarni asosiy shakliga keltirish), stop-so'zlarni olib tashlash (kam ahamiyatga ega bo'lgan so'zlar) va ma'lumotlarni tozalash (noto'g'ri yoki keraksiz belgilarni olib tashlash). Bu jarayonlar algoritmlarning aniq va samarali ishlashi uchun matnni tartibga solish va optimallashtirishni ta'minlaydi.

5- mavzu. Til modellari: n-gram va ehtimollik usullari

Ushbu mavzuda til modellarining asoslari va ularni yaratish usullari ko'rib chiqiladi. N-gram modellar so'zlarning ketma-ketligini aniqlash orqali matn ichidagi ehtimollik munosabatlarini tahlil qiladi. Bu texnika matnlarni taxmin qilish, yangi matn yaratish va so'zlarni bir-biri bilan bog'lashda ishlatiladi. Dars davomida n-gram modellarining ishlash prinsiplari, ularning amaliy qo'llanilishi va cheklovlari o'rganiladi. Shuningdek, ehtimollik usullaridan foydalanib, til modellari qanday qurilishi haqida tushuncha beriladi.

6- mavzu. Ma'lumotlarni to'plash va tayyorlash

Ushbu mavzu ma'lumotlarni qayta ishlash jarayonining dastlabki bosqichlari haqida tushuncha beradi. Talabalar ma'lumotlarni yig'ishning turli usullari, ularni tahlil qilish uchun tayyorlash jarayonlari bilan tanishadilar. Ushbu bosqichda ma'lumotlar tozalash, formatlash, yo'qolgan qiymatlarni to'ldirish va boshqa oldindan ishlov berish usullari

ko'rib chiqiladi. Yaxshi tayyorlangan ma'lumot to'plamlari keyingi tahlil va algoritmlar uchun sifatli natijalarga erishishga yordam beradi.

7- mavzu. Vektorlashtirish usullari

Ushbu mavzuda matnlarni raqamli ko'rinishga o'tkazish usullari o'rganiladi. Vektorlashtirish usullari so'zlarni va matnlarni kompyuter tushunishi uchun raqamli vektorlar sifatida ifodalaydi. Talabalar "Bag of Words" (BoW), "Term Frequency-Inverse Document Frequency" (TF-IDF) va "Word Embeddings" (masalan, Word2Vec, GloVe) kabi usullar bilan tanishadilar. Bu usullar matnlarni tahlil qilish, taqqoslash va sinflarga ajratish uchun zarur bo'lgan asosiy tushunchalarni shakllantirishga yordam beradi.

8- mavzu. Word2Vec, GloVe va boshqa word embeddings

Ushbu darsda talabalar Word2Vec va GloVe kabi so'z embedding usullari bilan tanishadilar. Word embeddings — bu so'zlarni raqamli formatda ifodalash usullari bo'lib, ularning ma'nolari va kontekstlariga mos ravishda joylashuvini hisobga oladi. Dars davomida talabalar shuningdek, boshqa word embedding usullari, masalan, FastText va ELMo bilan tanishadilar. Ushbu usullar so'zlarning kontekstiga qarab turli xil embeddinglar yaratish imkoniyatini taqdim etadi.

9- mavzu. Sintaktik va semantik tahlil

Talabalar dars davomida sintaktik va semantik tahlil usullarini o'rganadilar, ularning nazariy asoslari, texnikalari va amaliy qo'llanilishini ko'rib chiqadilar. O'qitish jarayonida sintaktik va semantik tahlilning NLP dasturlaridagi qo'llanilishi, muammolari va yechimlari haqida muhokama olib boriladi. Bu ko'nikmalar talabalarni tabiiy tilni samarali va aniq tahlil qilishga tayyorlaydi.

10- mavzu. Sentiment tahlili va matn tasnifi

Sentiment tahlili va matn tasnifi, tabiiy tilni qayta ishlash (NLP) sohasining muhim bo'laklaridan biridir. Ushbu jarayonlar matnlarning mohiyatini tushunish va ularni ma'lum kategoriyalarga ajratish imkonini beradi. **Sentiment tahlili**- Bu jarayon matn yoki xabarning ijobiy, salbiy yoki neytral kayfiyatini aniqlashga qaratilgan. **Matn tasnifi**- Bu jarayon matnni oldindan belgilangan sinflarga (masalan, sport, iqtisod, siyosat) ajratish uchun amalga oshiriladi. Sentiment tahlili va matn tasnifi reklama kampaniyalarini baholash, mijozlar fikrlarini o'rganish va ijtimoiy media tahlili kabi ko'plab sohalarda qo'llaniladi. Ushbu jarayonlar orqali kompaniyalar va tashkilotlar o'z mahsulot va xizmatlarini yaxshilash va mijozlar talablarini yanada samarali qondirish imkonini topadilar.

11- mavzu. Klasterlash, mavzularni aniqlash va matnlarni vektorlashtirish

Klasterlash va mavzularni aniqlash — tabiiy tilni qayta ishlash va ma'lumotlar tahlilining muhim jarayonlari bo'lib, katta hajmdagi matnlarni samarali ravishda tahlil qilish imkonini beradi.

Klasterlash — bu o'xshash xususiyatlarga ega bo'lgan ob'ektlarni (masalan, matnlarni) bir guruhga birlashtirish jarayonidir. Ushbu jarayonda K-means yoki Hierarchical Clustering kabi algoritmlar yordamida ma'lumotlar o'zaro yaqinliklariga ko'ra guruhlanadi. Klasterlash

yordamida foydalanuvchi fikrlarini yoki sharhlarni o'xshashlik darajasiga ko'ra tahlil qilish va guruhlash mumkin.

Mavzularni aniqlash — matnlar ichidagi asosiy mavzularni aniqlash va tasniflash jarayonidir. Bu jarayon Latent Dirichlet Allocation (LDA) kabi statistik usullar yordamida amalga oshiriladi. Mavzularni aniqlash orqali matnning asosiy ma'nolari va ularning o'zaro bog'lanishlari tushuniladi.

12- mavzu. NLP'da Deep Learning, Neyron Tarmoqlar va Transformer arxitekturalari

Ushbu mavzu tabiiy tilni qayta ishlash (NLP) sohasida chuqur o'rganish (Deep Learning) va neyron tarmoqlarning asosiy tushunchalarini o'z ichiga oladi. Deep Learning metodlari murakkab va katta hajmdagi ma'lumotlardan mustaqil ravishda o'rganishga qodir, bu esa NLP vazifalarini, masalan, matni tasniflash, tarjima qilish, va suhbat tizimlarini yaratishda qo'llaniladi. Neyron tarmoqlar, xususan, konvolyutsion (CNN) va takroriy (RNN) tarmoqlar, bu jarayonda o'ziga xos ahamiyatga ega, chunki ular ma'lumotlardagi murakkab aloqalarni aniqlashda samaradorlikni oshiradi. Ushbu mavzu, shuningdek, transfer o'qitish (transfer learning) va pre-trained modellar, masalan, BERT va GPT, yordamida NLPdagi eng so'nggi yutuqlarni ham ko'rib chiqadi. Talabalar NLPda Deep Learning va neyron tarmoqlar yordamida qanday qilib til modellari va ilovalarini yaratishni o'rganadilar.

13- mavzu. Transformer modellar: BERT, GPT va boshqalar, ularni oldindan o'qitish va nozik sozlash

Ushbu mavzu tabiiy tilni qayta ishlash (NLP) sohasida transformator modellarining ahamiyati va ularning turli xillari, jumladan BERT (Bidirectional Encoder Representations from Transformers) va GPT (Generative Pre-trained Transformer) kabi ilg'or modellarni o'rganishga qaratilgan. Transformer modellar ko'plab NLP vazifalarini yechish uchun zamonaviy yondashuv sifatida tanilgan va quyidagi asosiy jihatlar bilan tavsiflanadi: Transformer arxitekturasi, BERT modeli, GPT modeli. Ushbu mavzu talabalarni zamonaviy transformator modellarini o'rganishga va ularning afzalliklari, imkoniyatlari va foydalanish sohasini chuqur tushunishga tayyorlaydi. Bu bilimlar talabalarni NLP sohasida eng ilg'or texnologiyalar bilan ishlashga yordam beradi.

14- mavzu. Tili modellari va tilni generatsiya qilish

Ushbu mavzu tabiiy tilni qayta ishlash (NLP) sohasida til modellari va matn generatsiyasi jarayonlariga bag'ishlangan. Tili modellari — bu kompyuterlar uchun tilni tushunishga va yaratishga yordam beruvchi statistik va matematik modellar. Ushbu mavzu quyidagi asosiy jihatlarni o'z ichiga oladi: Tili modellari tushunchasi. Matn generatsiyasi. Transformer modellar. Ushbu mavzu talabalarni til modellari va matn generatsiyasi jarayonlarining asosiy tushunchalari bilan tanishtiradi va zamonaviy texnologiyalarni qo'llash orqali tilni qanday tushunish va yaratish mumkinligini o'rganishga yordam beradi.

15- mavzu. Chatbotlar va suhbat botlari yaratish

Ushbu mavzu chatbotlar va suhbat botlarini yaratish jarayoniga bag'ishlangan. Chatbotlar — foydalanuvchilar bilan muloqotda bo'lish uchun dasturlashtirilgan dasturlar bo'lib, ular avtomatik ravishda savollarga javob berish, ma'lumot taqdim etish va turli xizmatlar ko'rsatish imkoniyatini beradi. Mavzu talabalarni chatbotlar va suhbat botlarini yaratish

jarayonini chuqur tushunishga yordam beradi va ularning dasturlash, dizayn va foydalanuvchi tajribasini oshirish ko'nikmalarini rivojlantiradi.

16- mavzu. Tabiiy til tahlilida ma'lumotlar bilan ishlash

Ushbu mavzu tabiiy tilni qayta ishlash (NLP) jarayonida ma'lumotlar bilan ishlashning muhim jihatlarini o'rganishga qaratilgan. Tabiiy til tahlili ko'plab bosqichlarni o'z ichiga oladi, va har bir bosqichda ma'lumotlarni to'plash, tozalash, va tayyorlash jarayonlari katta ahamiyatga ega. Mavzu talabalarni tabiiy til tahlilida ma'lumotlar bilan ishlash jarayonlari haqida chuqur tushuncha berish va amaliy ko'nikmalarni shakllantirishga yordam beradi. Talabalar o'z loyihalarida ma'lumotlarni samarali tayyorlash va tahlil qilish imkoniyatini egallaydilar.

17- mavzu. Amaliy NLP loyihalari: Sentiment tahlili loyihasi

Ushbu mavzu sentiment tahlili loyihasini amalga oshirishga bag'ishlangan bo'lib, u tabiiy tilni qayta ishlash (NLP) yordamida matnning his-tuyg'ularini aniqlash va tahlil qilish jarayonini o'rganishni o'z ichiga oladi. Loyihaning asosiy maqsadi turli manbalardan olingan matnlarda ijobiy, salbiy yoki neytral his-tuyg'ularni aniqlashdir. Mavzu talabalarni sentiment tahlilining nazariy va amaliy jihatlari bilan tanishtirib, ularning muayyan loyihalarni amalga oshirish qobiliyatlarini rivojlantiradi. Talabalar loyiha orqali real muammolarni hal qilish va ularni amaliyotda qo'llash imkoniyatiga ega bo'ladilar.

18- mavzu. Amaliy NLP loyihalari: Chatbot yaratish

Ushbu mavzu amaliy tabiiy tilni qayta ishlash (NLP) loyihasi sifatida chatbot yaratishga bag'ishlangan. Chatbotlar — foydalanuvchilar bilan muloqot qilish uchun mo'ljallangan dasturlar bo'lib, ular avtomatik ravishda savollarga javob berish, ma'lumot taqdim etish yoki vazifalarni bajarish imkonini beradi. Mavzu talabalarni chatbotlarni yaratish jarayoni bilan tanishtiradi, ularga amaliy ko'nikmalarni rivojlantirish imkoniyatini beradi va zamonaviy kommunikatsiya texnologiyalari haqida bilimlarni chuqurlashtirishga yordam beradi.

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:

Tabiiy tilga oid asosiy tushunchalar. Python va NLP kutubxonalari. Matnlarni oldindan qayta ishlash. Til modellari: n-gram va ehtimollik usullari. Ma'lumotlarni to'plash va tayyorlash. Vektorlashtirish usullari. Word2Vec, GloVe va boshqa word embeddings. Sintaktik va semantik tahlil. Sentiment tahlili va matn tasnifi. Klasterlash va mavzularni aniqlash. NLPda Deep Learning va Neyron Tarmoqlar. Transformer modellar: BERT, GPT va boshqalar. Tili modellari va tilni generatsiya qilish. Chatbotlar va suhbat botlari yaratish. Tabiiy til tahlilida ma'lumotlar bilan ishlash. Amaliy NLP loyihalari: Sentiment tahlili loyihasi. Amaliy NLP loyihalari: Chatbot yaratish.

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)

NLP algoritmlari fani o'qitilishi davomida talabalar quyidagi kompetensiyalarni shakllantiradilar:

1. Nazariy bilim:

- o Tabiiy tilni qayta ishlashning asosiy tushunchalari, jarayonlari va algoritmlari haqida chuqur bilimga ega bo'lish.
- o NLPning muhim tushunchalari, jumladan, tokenlash, lemmatizatsiya, stemming va sintaktik tahlil haqida bilish.

2. Amaliy ko'nikmalar:

- o Python dasturlash tilida NLP algoritmlarini amalga oshirish va foydalanish ko'nikmalarini rivojlantirish.
- o NLP kutubxonalari (NLTK, spaCy, TensorFlow va boshqalar) yordamida amaliy loyihalarni bajarish qobiliyati.

3. Ma'lumotlar bilan ishlash:

- o Turli formatdagi matnli ma'lumotlarni yig'ish, tozalash va tayyorlash jarayonlarini amalga oshirish.
- o Ma'lumotlarni to'plash va tayyorlash usullarini qo'llash orqali sifatli NLP modellarini yaratish.

4. Modelni o'qitish va baholash:

- o Chuqur o'rganish va neyron tarmoqlar yordamida NLP modellarini o'qitish jarayonlarini tushunish.

- o Modelning samaradorligini baholash va yaxshilash uchun metrikalar (aniqlik, recall, F1-score)dan foydalanish.

5. Muammolarni hal qilish:

- o NLP sohasidagi turli muammolarni aniqlash va ularni hal qilish uchun samarali algoritmlarni tanlash va qo'llash qobiliyati.
- o Chatbotlar, sentiment tahlili va matn tasnifi kabi amaliy muammolarni yechish.

6. Innovatsion fikrlash:

- o Yangi algoritmlar va usullarni ishlab chiqish va mavjud NLP yechimlarini takomillashtirish qobiliyati.
- o Tabiiy tilni qayta ishlash sohasida innovatsion loyihalarni amalga oshirishga tayyorlik.

7. Tajribalar va hamkorlik:

- o Guruhda ishlash va loyiha asosida tajriba almashish ko'nikmalarini rivojlantirish.
- o Muloqot va taqdimot ko'nikmalarini yaxshilash orqali o'z fikrlarini samarali ifoda etish.

5. VI. TA'LIM TEXNOLOGIYALARI VA METODLARI:

NLP (Tabiiy Tilni Qayta Ishlash) algoritmlari bo'yicha ta'lim berish jarayonida zamonaviy ta'lim texnologiyalari va metodlari quyidagilardan iborat:

1. Interaktiv o'qitish:

- o O'quvchilarga faol ishtirok etish imkoniyatini berish, muhokama va guruhviy ishlarni amalga oshirish orqali ko'nikmalarini rivojlantirish.
- o Talabalarni qiziqtiruvchi savollar va masalalar orqali o'rganish jarayonini jonlantirish.

2. Amaliy mashg'ulotlar:

- o NLP algoritmlarini dasturlash, tahlil qilish va ishlab chiqish uchun amaliy mashg'ulotlar.
- o Real dunyo muammolarini hal qilishga qaratilgan loyihalarni va tajribalar.

3. Online platformalar:

- o Moodle, Coursera yoki edX kabi onlayn ta'lim platformalaridan foydalanish orqali o'quv jarayonini yanada qulay va samarali qilish.
- o Onlayn kurslar va seminarlar orqali zamonaviy NLP texnologiyalarini o'rganish.

4. Qiyin vazifalar asosida o'qitish (Project-Based Learning):

- o Talabalarga real hayotdagi NLP loyihalarini yechishga qaratilgan qiyin vazifalar berish.
- o Talabalar o'z fikrlarini ifoda etish va muammolarni hal qilishda mustaqil fikrlashni rivojlantirish.

5. Virtual va kengaytirilgan haqiqat (VR/AR):

- o VR va AR texnologiyalari yordamida o'quvchilarga NLP jarayonlarini vizuallashtirish va amaliy tajribalar o'tkazish.
- o Foydalanuvchi interfeysi va tabiiy til bilan ishlashda interaktiv simulyatsiyalar yaratish.

6. Tizimli yondashuv (Blended Learning):

- o An'anaviy darslar va onlayn resurslar, masalan, video darslar va interaktiv materiallar yordamida o'qitish.
- o Har xil ta'lim usullarini birlashtirish orqali o'quvchilarning ko'nikmalarini kengaytirish.

7. Mentorlik va maslahat:

- o Tajribali mutaxassislar va o'qituvchilar tomonidan talabalar uchun mentorlik qilish.

	<ul style="list-style-type: none"> o O'quvchilarga qo'llab-quvvatlash va fikr bildirish orqali ularning o'rganish jarayonini yaxshilash. <p>8. O'z-o'zini o'rganish (Self-Directed Learning):</p> <ul style="list-style-type: none"> o Talabalarni mustaqil o'rganish va yangi texnologiyalarni o'zlashtirishga undash. o Internet manbalaridan foydalanish, ilmiy maqolalar va tadqiqotlarni o'qish orqali o'z bilimlarini kengaytirish.
<p>6.</p>	<p style="text-align: center;">VII. Kreditlarni olish uchun talablar:</p> <p>NLP algoritmlari bo'yicha topshiriqlarni muvaffaqiyatli bajarish uchun talabalar quyidagi asosiy ko'nikmalarga ega bo'lishlari zarur.</p> <ol style="list-style-type: none"> 1. Dasturlash ko'nikmalari: <ul style="list-style-type: none"> o Python dasturlash tilini chuqur bilish, shuningdek, NLP kutubxonalari (masalan, NLTK, spaCy, TensorFlow) bilan ishlay olish. o Dasturiy ta'minotni yozish, tuzatish va muammolarni hal qilish ko'nikmalari. 2. Matnni qayta ishlash ko'nikmalari: <ul style="list-style-type: none"> o Tabiiy tilni qayta ishlash jarayonlari (tokenlash, lemmatizatsiya, stemming, sintaktik tahlil va boshqalar) bilan tanishish va ularni amaliyotda qo'llay olish. o Matnli ma'lumotlarni yig'ish, tozalash va tayyorlash usullarini bilish. 3. Statistika va ma'lumotlar tahlili: <ul style="list-style-type: none"> o Statistika va ma'lumotlar tahlilining asosiy tushunchalari bilan tanishish. o Ma'lumotlarni tahlil qilish va natijalarni interpretatsiya qilish ko'nikmalarini rivojlantirish. 4. Mashina o'rganishi: <ul style="list-style-type: none"> o Mashina o'rganish algoritmlari va ularning NLPdagi qo'llanilishi haqida bilish. o Modellarini o'qitish, baholash va optimallashtirish jarayonlarini tushunish. 5. Muammolarni hal qilish: <ul style="list-style-type: none"> o NLP sohasidagi turli muammolarni aniqlash va samarali yechimlar ishlab chiqish ko'nikmalari. o Talabalar mustaqil va jamoaviy ravishda muammolarni hal qilishga tayyor bo'lishlari kerak. 6. Ijodiy fikrlash: <ul style="list-style-type: none"> o Yangi g'oyalarni ishlab chiqish va mavjud yechimlarni takomillashtirishda ijodiy yondashuvni qo'llash. o Foydalanuvchilarning ehtiyojlarini qondirish va innovatsion loyihalar ishlab chiqish. 7. Tajribalarni o'tkazish: <ul style="list-style-type: none"> o Tajriba o'tkazish va natijalarni tahlil qilish ko'nikmalarini rivojlantirish. o Tajribalar orqali olingan natijalarni asoslab berish va ularni amaliyotga qo'llash. 8. Muloqot va taqdimot: <ul style="list-style-type: none"> o O'z fikrlarini aniq va ravon ifoda etish, jamoaviy muhokamalarda ishtirok etish ko'nikmalari. o Loyihalar va natijalarni samarali taqdim etish, vizualizatsiya qilish ko'nikmalari.
<p>7.</p>	<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. Hands-On Large Language Models by Jay Alamar and Maarten Grootendorst Copyright © 2024 Jay Alamar and Maarten Pieter Grootendorst. All rights reserved. Printed in the United States of America. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. O'Reilly books may be purchased for educational, business, or sales promotional use.

2. Хобсон Лейн, Ханнес Халке, Коул Ховард X68 Обработка естественного языка в действии. — СПб.: Питер, 2020. — 576 с.: ил. — (Серия «Для профессионалов»). ISBN 978-5-4461-1371-2
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

Tavsiya qilinadigan 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. Daniel Jurafsky, James H. Martin. *Speech and Language Processing Copyright ©2024. All rights reserved. Draft of August 20, 2024*
7. Nitin Indurkha, Fred J. Damerau. *HANDBOOK OF NATURAL LANGUAGE PROCESSING SECOND EDITION* Chapman & Hall/CRC Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742 © 2010 by Taylor and Francis Group, LLC Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business No claim to original U.S. Government works Printed in the United States of America on acid-free paper 10 9 8 7 6 5 4 3 2 1 International Standard Book Number-13: 978-1-4200-8593-8 (Ebook-PDF)
8. Гольдберг И. Нейросетевые методы в обработке естественного языка / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2019. – 282 с.: ил. ISBN 978-5-97060-754-1
9. Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие /. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с. ISBN 978-5-9909752-1-7
10. Steven Bird, Ewan Klein, and Edward Loper - *Natural Language Processing with Python* by Steven Bird, Ewan Klein, and Edward Loper Copyright © 2009. All rights reserved. Printed in the United States of America. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472
11. Akshay Kulkarni, Adarsha Shivananda. *Natural Language Processing Recipes* ISBN-13 (pbk): 978-1-4842-4266-7 ISBN-13 (electronic): 978-1-4842-4267-4 <https://doi.org/10.1007/978-1-4842-4267-4> Library of Congress Control Number: 2019931849 Copyright © 2019 by Akshay Kulkarni and Adarsha Shivananda

Scopusdan maqolalar

- 12.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>
- 13.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>

Axborot manbalar:

<https://www.nltk.org/>
<https://new.mohirdev.uz/ru/rlp>
<https://web.stanford.edu/~jura/sky/slp3/>
<https://uzbekdevs.uz/>
<https://pythonworld.ru/>

8. Fanning o'quv dasturi Tarjimashunoslik, tilshunoslik va xalqaro jurnalistika oliy maktabida ishlab chiqilgan va 2025 yil - 11-06 dagi 24 sonli bayonnomasi bilan ma'qullangan.
O'quv dasturi Toshkent davlat sharqshunoslik universiteti Kengashining 2022 yil 28.06.11 sonli bayoni bilan tasdiqlangan.
9. Fau/modul uchun ma'sullar:
1. Aliyazov 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 kafedrasi professori, f.f.d