



TABIIY TILGA ISHLOV BERISH TIZIMLARI

Sultonov Shohnazar

sultonovshonazar@gmail.com

ToshDO‘TAU Kompyuter lingvistikasi
mutaxassisligi magistranti

Annotatsiya: Mazkur maqola tabiiy tilni ishlov berish tizimlariga bag'ishlangan bo'lib, bu tizimlarning ilmiy va amaliy ahamiyatini yoritadi. Tabiiy tilni ishlov berish (TTIB) so'zi kompyuterlar va boshqa sun'iy intellekt tizimlari yordamida inson tilini tushunish, qayta ishlash va tarjima qilish jarayonlarini o'z ichiga oladi. Maqolada TTIB tizimlarining asosiy konseptlari, texnologiyalari va usullari tahlil qilingan. Shuningdek, TTIB tizimlarining turli sohalarida, jumladan, tilni avtomatik tarjima qilish, matnni tahlil qilish, xabarlarini tushunish va sentiment tahlili kabi amaliy qo'llanilishini ko'rib chiqishadi. Tizimlarning samaradorligini oshirish uchun ishlatiladigan zamonaviy usullar, xususan, mashina o'rganish va chuqur o'rganish texnologiyalari haqida ham so'z yuritiladi. Maqolada tilni ishlov berish tizimlarining rivojlanishi va ularning kelajakdagi imkoniyatlari ham o'rganiladi. TTIB sohasidagi eng yangi yutuqlar, ularning ijtimoiy, madaniy va texnologik sohalarida qo'llanilishi hamda yuzaga kelayotgan muammolar ham maqolada batafsil ko'rib chiqiladi.

Abstract: This article is devoted to natural language processing systems and highlights the scientific and practical significance of these systems. The term natural language processing (NLP) includes the processes of understanding, processing and translating human language using computers and other artificial intelligence systems. The article analyzes the main concepts, technologies and methods of NLP systems. It also considers the practical application of NLP systems in various fields, including automatic language translation, text analysis, message understanding and sentiment analysis. Modern methods used to improve the efficiency of systems, in particular, machine learning and deep learning technologies, are also discussed. The article also studies the development of language processing systems and their future potential. The latest achievements in the field of NLP, their applications in social, cultural and technological spheres, as well as emerging problems are also considered in detail in the article.

Kalit so'zlar: tizimli analiz, tilni qayta ishlashning amaliy dasturlari, ko'p tildagi tizimlar, tabiiy tilni yaratish, tilni qayta ishlash

Keywords: systems analysis, language processing applications, multilingual systems, natural language generation, language processing

Kirish. Tabiiy tilga ishlov berish (TTIB) tizimlarining muvaffaqiyati, katta qismda, matnlarning ma'nosini to'g'ri tushunish qobiliyatiga bog'liq. Semantik analiz – TTIB ning muhim tarkibiy qismi bo'lib, matnning semantik tuzilishini, so'zlar va iboralarning ma'nosini, ular orasidagi munosabatlarni aniqlashga xizmat qiladi. Ushbu bobda ko'rib chiqilganidek, semantik analizning turli usullari – leksik-semantik resurslardan foydalanish, kontekstual tahlil, mashinaviy o'rganish algoritmlari – matnni tushunishda muhim rol o'ynaydi. Ammo, O'zbek tili kabi kam resursli tillar uchun semantik analizni amalga oshirishda hali ham muammolar mavjud. Lexik-semantik resurslarning yetishmasligi, murakkab grammatik tuzilmalar va ko'p ma'noli so'zlarning ko'pligi semantik analizning aniqligi va samaradorligiga ta'sir qiladi. Kelajakdagi tadqiqotlar O'zbek tili uchun yuqori aniqlikdagi semantik analizatorlarni yaratishga, mavjud resurslarni boyitishga va yangi, zamonaviy usullarni qo'llashga qaratilishi lozim.

Asosiy qism. Tabiiy tilga ishlov berish (Natural Language Processing - NLP) tizimlari, kompyuterlarning inson tilini tushunishi, tahlil qilishi va ishlab chiqishi uchun ishlatiladi. Bu sohada quyidagi asosiy yo'nalishlar mavjud:

1. *Tilni tushunish (Natural Language Understanding - NLU):* Bu inson tilidagi ma'lumotlarni analiz qilish va ularni kompyuter tushunadigan formatga aylantirish jarayonidir. Masalan, savolni aniqlash yoki kontekstni tushunish.

2. *Tilni ishlab chiqarish (Natural Language Generation - NLG):* Bu jarayon kompyuterga inson tilida ma'lumotlar yaratish imkonini beradi. Masalan, avtomatik matn yozish yoki xulosalar tayyorlash.

3. *Matnni tahlil qilish (Text Analysis):* Bu jarayonda matn ichidagi ma'lumotlarni, tuzilmani va kontekstni tahlil qilish uchun turli metodlar va algoritmlar qo'llaniladi.

4. *Muloqot tizimlari (Conversational Agents):* Chatbotlar va virtual yordamchilar bu toifaga kiradi. Ular foydalanuvchilar bilan interaktiv muloqot qilish imkoniyatini beradi.

5. *Mashina tarjimasi (Machine Translation):* Bu texnologiya bir tildagi matnni boshqa tilga avtomatik ravishda tarjima qilish uchun ishlatiladi, masalan, Google Translate.

6. *Ovozli tanish va generatsiya:* Ovozli buyruqlarni tanish va ovozli javoblarni ishlab chiqish.

Quyida tabiiy tilga ishlov berish sohasining ushbu yo'nalishlariga alohida to'xtalib o'tamiz.

Tilni tushunish (Natural Language Understanding - NLU) — bu tabiiy tilga ishlov berishning bir bo'limi bo'lib, kompyuterlarning inson tilidagi matn yoki nutqni tushunishi va tahlil qilishini ta'minlaydi. NLU jarayonlari quyidagi asosiy komponentlardan iborat:

1. **Maqsadni aniqlash:** Foydalanuvchining so'rovi yoki savolining maqsadini aniqlash. Masalan, foydalanuvchi "Bugun havo qanday?" deb so'raganda, bu savol havo ma'lumotlariga qiziqish bildiradi.
2. **Obyektlarni tanib olish:** Matndagi muhim elementlarni, masalan, odamlar, joylar yoki vaqtlarni aniqlash. Bu jarayon "Entity Recognition" deb ataladi.
3. **Sentiment tahlili:** Matnning ruhiyatini yoki hissiyotini aniqlash. Misol uchun, ijobiy, salbiy yoki neytral munosabatlarni tushunish.
4. **Sintaktik tahlil:** Matnning grammatik tuzilishini tahlil qilish. Bu jarayonda jumladagi so'zlar va ularning o'zaro aloqalari tahlil qilinadi.
5. **Semantik tahlil:** Matnning ma'nosini chuqurroq tushunish. Bu bosqichda so'zlar orasidagi bog'lanishlar va kontekst o'rganiladi.
6. **Kontekstni tushunish:** Ushbu jarayon, matnning oldingi va keyingi qismlarini hisobga olgan holda, foydalanuvchining maqsadini aniqlashni o'z ichiga oladi.

Tabiiy tilni ishlab chiqarish (Natural Language Generation - NLG) — bu tabiiy tilga ishlov berishning bir qismi bo'lib, kompyuterlarning ma'lumotlarni inson tilida ifodalashini ta'minlaydi. NLG tizimlari avtomatik ravishda matn yaratish uchun turli jarayonlarni o'z ichiga oladi. Asosiy komponentlar quyidagilardan iborat:

1. Ma'lumotni tahlil qilish: Avval, NLG tizimi uchun kerakli ma'lumotlar yig'iladi va tahlil qilinadi. Bu bosqichda qaysi ma'lumotlar kerakligini aniqlash muhimdir.

2. Rejalashtirish: Bu bosqichda tizim ma'lumotlardan qanday xulosa chiqarish kerakligini va qanday jummalarni yaratishi kerakligini rejalashtiradi. U qanday tuzilishga ega bo'lishi, qanday ma'lumotlarni ta'kidlashi kerakligi haqida qaror qabul qiladi.

3. Boshqaruv (Scripting): Rejaga muvofiq, tizim turli matn elementlarini va jummalarni birlashtiradi. Bu bosqichda grammatika va stil jihatlari inobatga olinadi.

4. Tilni ishlab chiqarish: Ushbu jarayon, tanlangan ma'lumotlarni, rejalashtirilgan tuzilmani va boshqaruvni birlashtirib, inson tilida to'g'ri va tushunarli matn yaratishni o'z ichiga oladi.

5. Post-ishi: Yaratilgan matnni tahrir qilish va yaxshilash. Bu bosqichda ko'proq ravonlik, mantiqiylik va grammatik to'g'rilik nazorat qilinadi.

NLG tizimlari ko'plab sohalarda qo'llaniladi, masalan:

- Avtomatik xulosa tayyorlash: Ma'lumotlar to'plamidan xulosalar chiqarish, masalan, hisobotlar va statistikalar.

- Chatbotlar va virtual yordamchilar: Foydalanuvchilarga javob berish uchun matn yaratish.

- Ovozli yordamchilar: Ovozli javoblar ishlab chiqarish.

- Ma'lumotlar vizualizatsiyasi: Murakkab ma'lumotlarni oddiy va tushunarli tarzda taqdim etish.

Matnni tahlil qilish (Text Analysis) — bu matndagi ma'lumotlarni, tuzilishni va ma'noni tahlil qilish jarayoni. Ushbu jarayon tabiiy tilga ishlov berish (NLP) ning

muhim qismi bo'lib, ko'plab sohalarda qo'llaniladi, jumladan, axborot qidirish, sentiment tahlili, va muloqot tizimlarida. Matnni tahlil qilish jarayoni bir necha asosiy bosqichlardan iborat:

1. *Preprocessing (dastlabki ishlov berish)*: Matnni tahlil qilishdan oldin, uni tayyorlash zarur. Bu bosqichda quyidagi jarayonlar amalga oshiriladi:

- Tozalash: Keraksiz belgilar va so'zlarni olib tashlash (masalan, nuqtalar, raqamlar, turli belgilar, internet linklari, html teglar).

- Tokenizatsiya: Matnni so'z yoki iboralar kabi kichik birlashtiruvchi qismlarga bo'lish.

- Stemming yoki lemmatization: So'zlarni ularning ildiz shakliga qaytarish yoki grammatika jihatidan to'g'ri shaklga keltirish.

2. *Tahlil qilish*: Bu bosqichda turli metodlar va algoritmlar yordamida matnning ma'lum jihatlari tahlil qilinadi. Masalan:

- *Nomlangan obyektlarni aniqlash (Named Entity Recognition)*: Matndagi nomlangan obyektlarni ya'ni odamlar, joylar, ko'chalar, binolar yoki tashkilotlar kabi obyektlarni ajratib olish.

- *Sentiment tahlili*: Matnning ijobiy, salbiy yoki neytral ruhiyatini aniqlash.

- *Sintaktik tahlil*: Matnning grammatik tuzilishini o'rganish.

3. *Ma'lumotlarni vizualizatsiya qilish*: Tahlil qilingan ma'lumotlar vizual shaklda taqdim etilishi mumkin, masalan, grafiklar yoki diagrammalar yordamida.

4. *Natijalarni chiqarish*: Tahlil natijalari asosida xulosalar chiqarish va ulardan foydalanish. Bu bosqichda foydalanuvchilarga yoki tizimlarga ma'lumotlar bilan bog'liq qarorlar qabul qilishda yordam beriladi.

Matnni tahlil qilish jarayonlari ko'plab sohalarda qo'llaniladi, masalan:

- *Marketologiya*: Mijozlar fikrlarini va sharhlarini tahlil qilish.

- *Ijtimoiy media monitoringi*: Foydalanuvchilarning postlarini va fikrlarini tahlil qilish.

- *O'qituvchilar va ilmiy tadqiqotchilar*: Matnlarni tahlil qilish orqali ilmiy tadqiqotlar va natijalarni chiqarish.

Matnni tahlil qilish jarayonlari uchun ko'plab vositalar va kutubxonalar mavjud, masalan, Python dasturlash tilida NLTK, spaCy, va TextBlob.

Muloqot tizimlari yoki suhbat agentlari — sun'iy intellekt asosida ishlaydigan dasturiy ta'minot bo'lib, ular odamlar bilan tabiiy tilda muloqot qilishga mo'ljallangan. Bu tizimlar ko'pincha chatbotlar, virtual yordamchilar yoki muloqot botlari sifatida tanilgan va turli vazifalarni bajarishga yordam beradi. Quyida muloqot tizimlarining turlarini ko'rishimiz mumkin:

- *Qoidaga asoslangan agentlar*: Ushbu agentlar oldindan belgilangan qoidalar va sxemalar asosida ishlaydi. Ular aniq skriptlar va muloqotlar ketma-ketligini bajaradi, lekin kutilmagan yoki yangi savollarga javob bera olmaydi.

- *Sun'iy intellekt asosidagi agentlar*: Ushbu agentlar tabiiy tilga ishlov berish (NLP) texnologiyalari yordamida ishlaydi. Ular foydalanuvchi maqsadini aniqlash,

mashinali o'qitish va chuqur o'qitish algoritmlaridan foydalanib, murakkab va noaniq savollarga javob bera oladi.

Muqolqot tizimlarini turli sohalarida qo'llash mumkin:

Xizmat ko'rsatish: Kompaniyalar muloqot agentlarini mijozlarga xizmat ko'rsatish uchun ishlatib, tez va qulay javoblar taqdim etishadi.

Savdo va marketing: Chatbotlar mahsulotlar haqida ma'lumot berish, mijozlarni jalb qilish va savdo jarayonini avtomatlashtirish uchun qo'llaniladi.

Ta'lim: Virtual yordamchilar talabalarga mavzuni tushuntirish, mashqlar berish va savollariga javob berishda yordam beradi.

Sog'liqni saqlash: Tibbiyotda bemorlarga maslahat berish, dori-darmonlar bo'yicha eslatmalar berish va diagnostik ma'lumotlar yig'ish uchun ishlatiladi.

Muloqot tizimlari yordamida insonlar bilan doimiy va samarali muloqot qilish imkoniyatlari kengayib, ularni turli sohalarida qo'llash imkoniyati yanada ortmoqda.

Mashina tarjimasi — bu kompyuter dasturlari yordamida matn yoki nutqni bir tildan boshqa tilga avtomatik ravishda tarjima qilish jarayonidir. Bu texnologiya sun'iy intellekt va tabiiy tilga ishlov berish (NLP) sohalaridagi rivojlanishlarga asoslanadi. Mashina tarimasini uch xil yondashuv asosida amalga oshirish mumkin:

- *Qoidalarga asoslangan mashina tarjimasi (Rules based on Machine Translation RBMT):* Ushbu yondashuv oldindan belgilangan grammatik qoidalar va lug'at asosida ishlaydi. U manba tilining grammatik tuzilmalarini maqsad tiliga o'zgartiradi. Bu usul aniq va qattiq qoidalarga asoslangan, ammo yangi yoki murakkab iboralar bilan ishlashda qiyinchiliklarga duch kelish mumkin.

- *Statistik mashina tarjimasi (Statistical Machine Translation SMT):* Bu yondashuv katta hajmdagi ikki tilli korpuslardan statistik ma'lumotlarni yig'ib, tarjima jarayonini amalga oshiradi. SMT biror jumlaning qaysi ehtimol bilan qanday tarjima qilinishini aniqlash uchun statistik modellarni qo'llaydi.

- *Neural mashina tarjimasi (Neural Machine Translation NMT):* Bu yondashuv chuqur o'qitish texnologiyalaridan foydalangan holda ishlaydi. NMT tizimlari neyron tarmoqlar yordamida tilni tushunish va tarjima qilishni o'rgatadi, bu esa ularning kontekstni tushunish va yanada tabiiyroq tarjimalar qilish imkoniyatini beradi.

Mashina tarjimasi tizimini ishlab chiqishning afzalliklari quyidagilardan iborat:

- *Tezlik va samaradorlik:* Mashina tarjimasi qo'l bilan tarjima qilishga qaraganda ancha tezroq bo'ladi, bu katta hajmdagi matnlarni qisqa vaqt ichida tarjima qilish imkonini beradi.

- *Qulaylik:* Bu texnologiya turli tabiiy tillardagi foydalanuvchilar o'rtasida aloqa o'rnatishni osonlashtiradi va global muloqot uchun qulay sharoit yaratadi.

- *Doimiy o'rganish va yaxshilanish:* Nevron tarmoq modellari asosida ishlovchi tizimlar o'z tajribalaridan o'rganib, vaqt o'tishi bilan o'z tarjima sifatini yaxshilay oladi.

Mashina tarjimasini amalga oshirish jarayonida ham qator qiyinchiliklar va cheklovlar mavjud:

- *Til va madaniyat farqlari:* Ba'zan matnni to'g'ri tarjima qilish uchun tilning nozikliklari va madaniy kontekstni tushunish zarur bo'ladi, bu esa mashina tarjimasini uchun qiyinchilik tug'dirishi mumkin.
- *Qiyin iboralar va murakkab grammatik tuzilmalar:* Mashina tarjimasini tizimlari murakkab sintaksis va ko'p ma'noli so'zlar bilan ishlashda xatolarga yo'l qo'yishi mumkin.
- *Korpus va ma'lumot yetishmovchiligi:* Tarjima tizimining sifatli ishlashi uchun katta hajmdagi yuqori sifatli o'quv ma'lumotlari talab etiladi, lekin har doim ham bunday ma'lumotlar mavjud bo'lmasligi mumkin.

Mashina tarjimasining turli sohalarda qo'llash mumkin:

Internet tarjima xizmatlari: Google Translate, DeepL, Microsoft Translator kabi mashhur onlayn tarjima xizmatlari dunyo bo'ylab millionlab foydalanuvchilar uchun tezkor va qulay tarjima imkoniyatlarini taqdim etadi.

Ovozli tarjima qurilmalari: Portativ qurilmalar yoki smartfon dasturlari yordamida biror nutqni real vaqt rejimida boshqa tilda tinglash va tushunish imkoniyati mavjud.

Professional tarjima yordamchilari: Ba'zi dasturlar professional tarjimonlarga qo'shimcha vosita sifatida ishlatiladi, ularning ish jarayonini tezlashtiradi va ularga tahliliy ma'lumotlar taqdim etadi.

Mashina tarjimasini kelajakda yanada rivojlanib, yanada aniqroq, tabiiyroq va samaraliroq tarjimalar qilish imkoniyatiga ega bo'lishi kutilmoqda. Jahonda ko'plab tabiiy tillarda bunday tizimlar ishlab chiqilgan va bugungi kunda foydalanilmoqda. NLP tizimlari odatda mashinali o'qitish (machine learning) va sun'iy intellekt (AI) metodlari yordamida rivojlanadi.

Xulosa: Ta'qdim etilgan ma'lumotda matnni tahlil qilish va muloqot tizimlari, ularning asosiy tamoyillari, qo'llanilishi va sohalari o'rganildi. Matnni tahlil qilish, tabiiy tilga ishlov berish (NLP)ning asosiy qismi bo'lib, axborot qidirish, sentiment tahlili va muloqot tizimlarida muhim o'rin tutadi. U bir necha bosqichlardan iborat: dastlabki ishlov berish, tahlil qilish, ma'lumotlarni vizuallashtirish va natijalarni chiqarish.

Matnni tahlil qilish jarayonida turli texnikalar qo'llaniladi, jumladan, matnni tozalash, tokenizatsiya, stemming yoki lemmatizatsiya, nomlangan obyektlarni aniqlash, sentiment tahlili va sintaktik tahlil. Bu texnikalar ma'lumotlarni yig'ish, tahlil qilish va xulosalar chiqarishda yordam beradi.

Muloqot tizimlari esa, sun'iy intellektga asoslangan dasturiy ta'minot bo'lib, insonlar bilan tabiiy tilda muloqot qilish imkonini beradi. Ular qoidalarga asoslangan va sun'iy intellektga asoslangan agentlarga bo'linadi. Sun'iy intellektga asoslangan agentlar NLP texnologiyalari yordamida murakkab va noaniq savollarga



javob berishga qodir. Muloqot tizimlari xizmat ko'rsatish, savdo, marketing, ta'lim va sog'liqni saqlash kabi turli sohalarda keng qo'llanilmoqda.

Foydalanilgan adabiyotlar

1. Abdurahmonova N. Automatic morphological analyze for English-Uzbek system // Известия Кыргызский государственный технический университет им. И.Раззакова. Теоретический и прикладной научно-технический журнал. № 2 (38) – Бишкек, 2016. – С. 12-18.;
2. Mengliyev B., Hamroyeva Sh. Turkiy tillardagi morfologik analizatorlar xususiyada// O'zbek tilini dunyo miqyosida keng targ'ib qilish bo'yicha hamkorlik istiqbollari // Xalqaro ilmiy-amaliy anjuman materiallari. – Toshkent, 2020, oktabr. – B. 360-364;
3. Hamroyeva Sh. O'zbek tili morfologik analizatorining lingvistik. – GlobeEdit, 2020;
4. Elov B. Hamroyeva SH. MORFOLOGIK ANALIZATORNI YARATISH USULLARI//Til va madaniyat 2022. 4-son.