

Xudayberganova Munira

Kirish: Kompyuter lingvistikasi - tabiiy tillarni avtomatik qayta ishlash va tahlil qilish bilan shug'ullanadigan fan sohasidir. U sun'iy intellekt, mashina o'rganish va algoritmlar yordamida tilshunoslik qonuniyatlarini o'rganib, inson tilini tushunish, yaratish va talqin qilishga yordam beradi. Ushbu sohaning muhim vazifalaridan biri - kompyuterlar yordamida tezaurus lug'atlarini yaratishdir. Tezaurus lug'atlari so'zlarning semantik munosabatlarini o'z ichiga olgan ma'lumotlar to'plami bo'lib, ular tabiiy tilda ishlatiladigan so'zlarni to'g'ri talqin qilishda, sinonimlar va antonimlar, so'z ma'nolari orasidagi bog'lanishlarni aniqlashda qo'llaniladi.

Asosiy qism:

1. **Kompyuter lingvistikasi tushunchasi:** Kompyuter lingvistikasi tabiiy tillarni matematik va algoritmik asosda o'rganadi. Bu soha tilga oid ma'lumotlarni qayta ishlashda katta yutuqlarga erishgan bo'lib, u matnlarni avtomatik tahlil qilish, tarjima qilish, ovoqli buyruqlarni qabul qilish kabi texnologiyalarni rivojlantirishga xizmat qiladi. Kompyuter lingvistikasining asosiy vazifalaridan biri tilshunoslikda to'plangan bilimlarni algoritmlarga tatbiq etishdir.

2. **Tezaurus lug'atlari va ularning ahamiyati:** Tezaurus lug'atlari o'zaro sinonim, antonim yoki bog'langan ma'nolarga ega so'zlarni tizimlashtirish imkonini beradi. Ushbu lug'atlar tabiiy tilni qayta ishlash (NLP) jarayonida muhim rol o'ynaydi. So'zlar orasidagi ma'noviy bog'lanishlar ko'p hollarda til modellari va sun'iy intellekt algoritmlariga tushunarli bo'lishi uchun ishlatiladi. Masalan, tezaurus yordamida kompyuter bir so'zning o'rniga ma'nosi yaqin boshqa so'zni aniqlab bera oladi, bu esa avtomatik tarjima va so'z tanlash jarayonlarini yanada takomillashtiradi.

3. **Tezaurus lug'atlarini yaratishda kompyuter lingvistikasining roli:** Tezaurus lug'atlarini yaratishda kompyuter lingvistikasi dasturlar va algoritmlar orqali so'zlar orasidagi semantik munosabatlarni aniqlaydi. Dastlab so'zlar va ularning lug'atdagi shakllari o'rganiladi, so'ngra sinonimlar, antonimlar va boshqa bog'lanishlar asosida toifalarga bo'linadi. Bu jarayon ko'p hollarda sun'iy intellekt va mashina o'rganish algoritmlariga asoslanadi. Ushbu jarayon quyidagi bosqichlarni o'z ichiga oladi:

- **So'zlarni semantik tahlil qilish:** Kompyuter lingvistikasining semantik tahlil algoritmlari yordamida so'zlarning ma'nolari, ularning boshqa so'zlar bilan bog'liqligi o'rganiladi.

- **Sinonimlar va antonimlar tarmog'ini yaratish:** Mashina o'rganish yordamida lug'atlardagi sinonim va antonimlarni aniqlash va ularni o'zaro bog'lash orqali tezaurus yaratiladi.

- **Tezaurus lug'atlarining avtomatik yaratilishi:** Bu jarayonda algoritmlar va modellardan foydalanib katta hajmdagi matnli ma'lumotlardan so'zlar va ularning munosabatlari avtomatik tarzda olingan holda tezauruslar yaratib boriladi.

4. Amaliy misollar:

- **WordNet:** Bu ingliz tilidagi tezaurus lug'ati bo'lib, unda so'zlar orasidagi sinonimlar va antonimlar tarmoqlangan ko'rinishda keltirilgan. WordNet ko'pincha tabiiy tilni qayta ishlash sohasida tadqiqotlar uchun asosiy manba sifatida ishlatiladi.

- **Avtomatik tarjima tizimlari:** Tezaurus lug'atlari avtomatik tarjima tizimlarida so'zlarni to'g'ri tanlashda va sinonimlarni aniqlashda yordam beradi.

ILM FAN YANGILIKLARI KONFERENSIYASI

30-SENTABR

ANDIJON, 2024

Xulosa: Kompyuter lingvistikasi va tezaurus lugʻatlari yaratish tabiiy tillarni qayta ishlashning asosiy yoʻnalishlaridan biri boʻlib, sunʻiy intellektning rivojlanishiga katta hissa qoʻshadi. Tezauruslar til oʻrganishda, matnlarni tahlil qilishda va avtomatik tarjimalar yaratishda katta yordam beradi. Shu sababli, kelajakda kompyuter lingvistikasi sohasidagi izlanishlar va innovatsiyalar tezaurus lugʻatlarini yaratish jarayonini yanada samarali va rivojlangan texnologiyalar bilan taʼminlashi kutilmoqda.

Foydalanilgan adabiyotlar:

1. Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed.). Pearson.
2. Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
3. Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
4. Wilks, Y., & Slator, B. M. (1996). *Electric Words: Dictionaries, Computers, and Meanings*. MIT Press.
5. Kilgarriff, A., & Grefenstette, G. (2003). *Introduction to the Special Issue on the Web as Corpus*. *Computational Linguistics*, 29(3), 333-347.