Information Extraction Technologies In Question-Answer Systems

Suyarova Orzugul Sodiq qizi
Shahriasbz State Pedagogical Institute
Theory And Methodology Of Education
(Precessary Education Direction)
1,25-Group Master's Student

Annotation
This article studies and analyzes the technologies of extracting the necessary information from textual data in modern question-answer (QA) systems (Information Extraction). In the framework of the study, the effectiveness of natural language processing (NLP), object recognition. (NER) and semantic search methods was studied. The results obtained indicate that the accuracy of semantic methods is higher than that of simple keyword search. The results of the study can be used in the development of intelligent search and chatbot systems.
Keywords: Question-answering systems, information extraction, NLP, named object recognition, semantic analysis, machine learning.
Annotatsiya
Ushbu maqolada zamonaviy savol-javob (QA) tizimlarida matnli ma’lumotlar ro’xatidan zarur ma’lumotni ajratib olish (Information Extraction) texnologiyalari o’rganilib tahlil qilinadi. Tadqiqot atrofida tabiiy tilni takror ishlash (NLP), atalgan ob'ektlarni belgilash. (NER) va semantik qidiruv usullarining natijakorligi o'rganilgan. Olingan javoblar oddiy kalit so'zli qidiruvga qaraganda semantik usullarning aniqligi baland ekanligini bildiradi. Tadqiqot javoblari aqlli qidiruv va chatbot tizimlarini o’zlashtirishda qo'llanilishi mumkin.
Kalit so'zlar: Savol-javob tizimlari, axborotni ajratib olish, NLP, nomlangan ob'ektlarni aniqlash, semantik tahlil, mashinali o'qitish.

Introduction (Kirish)Axborot texnologiyalarining shiddatli rivojlanishi sharoitida katta hajmdagi ma'lumotlar ichidan aniq va kerakli axborotni tezkor topish dolzarb muammoga aylandi. An'anaviy qidiruv tizimlari foydalanuvchiga hujjatlar ro'yxatini taqdim etsa, Savol-javob tizimlari (Question Answering systems - QA) bevosita savolga aniq javob qaytarishga mo'ljallangan.Savol-javob tizimlarining muvaffaqiyati matn ichidan kerakli faktlar, raqamlar yoki munosabatlarni aniqlab beruvchi axborotni ajratib olish (Information Extraction - IE) texnologiyalariga bog'liq. Ushbu tadqiqotning maqsadi — QA tizimlarida axborotni ajratib olish usullarini tahlil qilish va ularning samaradorligini baholashdan iborat. QR CODE – Tez javob qaytaradigan maxsus matritsali kvadratlardan iborat
shtrix kodi bo’lib birinchi martta 1994-yilda Yaponiyadagi avtmobil sanoatida qo’llanilgan . Shtrix - bu biriktirilgan ob'ekt haqidagi ma'lumotni o'z ichiga olgan kompyuterda o'qiladigan optik yorliqdi. QR kodi ma'lumotni samarali saqlash uchun to'rtta standartlashtirilgan kodlash rejimlaridan foydalanadi (raqamli, alfanumerik, bayt / ikkilik va kanji); kengaytmalar ham ishlatilishi mumkin. QR kodi tizimi standart UPC shtrixlariga nisbatan tezroq o'qilishi va katta saqlash imkoniyati tufayli avtomobil sanoatining tashqarisida mashhur bo'ldi.
Dasturning algoritmik jarayoni Java dasturlash tilida realizatsiya qilingan bo’lib, bu dasturimizning yana moslashuvchanligi va mobil qurulmalarda tezkor ishlashini ta’minlaydi. Java dasturlash tili – eng ommalashgan dasturlash tillaridan biri bo’lib unda korporativ darajadagi mahsulotlarni(dasturlarni) yaratish mumkin. Bizning “QR-Bookland” dasturimiz android mobil operatsion tizimining barcha versiyalarida ishlashga mo’ljallangan bo’lib unda QR kodlarni va CODE_128

tipidagi shtrix kodlarni o’qish va generatsiya qilish imkoniyati mavjutdir1
Materials and Methods (Metodlar va materiallar)Tadqiqotda axborotni ajratib olishning uchta asosiy yondashuvi ko'rib chiqildi:Qoidalarga asoslangan usullar (Rule-based): Sintaktik shablonlar va muntazam ifodalar (regex) yordamida axborot qidirish.Mashinali o'qitish usullari (Machine Learning): NLP kutubxonalari (masalan, SpaCy, NLTK) yordamida Nomlangan ob'ektlarni aniqlash (NER).Chuqur o'qitish va transformatorlar (Deep Learning - BERT, GPT): Matnning kontekstual ma'nosini tushunish uchun vektorli modellar (Embeddings).Tadqiqot uchun ochiq matnli ma'lumotlar to'plami (dataset) hamda savollar banki tanlab olindi. Matnlar:
Hujjat1: “Axborot tizimlari izlash samaradorligi”
Hujjat2:”Izlash algoritmlari samaradorligini oshirish “
Invertlangan indeks:

So’z	Hujjatlar(pozitsiya)
Axborot	Doc1 (1)
	Doc1(2)
Qidirish	Doc1(3) Doc2(3)
Natijadorligi	Doc1(4) Doc2(3)
Algoritmlari	Doc2(2)
Yuqori oshirish	Doc2(4)

B-Tree yoki B+Tree indeksi
Bu algoritmda matnli ma’lumotlar yarartilgan daraxt sxemasida turadi,bunda izlagan va kerak bo’lgan ma’lumotlar tez topiladi,ishning samaradorligi oshadi.
Qulayliklari:
Diskdan ma’lumotlarni samarali va tez yuklav oladi.
Yirik suratdagi axborotlarni boshqarishda yaxshi natija.
Ishlash:
Matnli axborotlarni alifbo tartibida Tizimning ishlash aniqligi Precision (aniqlik) va Recall (to'liqlik) mezonlari bo'yicha baholandi.3. Results (Olingan natijalar)O'tkazilgan eksperimentlar shuni ko'rsatdiki, an'anaviy qidiruv usullari savol tarkibidagi kalit so'zlarga tayangani uchun ko'p hollarda noto'g'ri kontekstdagi javoblarni chiqardi.
Deep Learning (BERT modeli) asosidagi axborot ajratib olish texnologiyasi esa matnning ma'nosini anglagan holda eng yuqori natijani qayd etdi.
Metodika
(Texnologiya)Precision (Aniqlik)Recall (To'liqlik)F1-Score (Umumiy ko'rsatkich)
Qoidalarga asoslangan (Regex)$65\%$$50\%$$56\%$
Klassik Mashinali o'qitish (SVM/CRF)$78\%$$72\%$$75\%$
Chuqur o'qitish (Transformer/BERT)$91\%$$88\%$$89\%$4.
Discussion (Muhokama)Olingan natijalar shuni tasdiqlaydiki, axborotni ajratib olishda neyron tarmoqlari va semantik tahlil usullari an'anaviy sintaktik usullardan sezilarli darajada ustun turadi. Chunki foydalanuvchilar savolni turlicha shakllantirishi mumkin (sinonimlar, jargonlar).Biroq, chuqur o'qitish modellarining (masalan, LLM — Katta til modellari) kamchiligi shundaki, ular hisoblash resurslarini (GPU) juda ko'p talab qiladi va real vaqt rejimida kichik serverlarda sekin ishlashi mumkin. Shuning uchun gibrid tizimlar (qoidalar + neyron tarmoqlar) amaliyotda eng maqbul yechim bo'lib qolmoqda. Axborot izlash - bu ma'lum bir hujjatlar (matnli) to'plamidan oldindan belgilangan shartli mavzu (so'rov) yoki zarur (axborotga bo'lgan ehtiyojni qondirishga tegishli) ma'lumotlarni, faktlarni, xabarlarni izlash - aniqlash jarayoni. Qidiruv jarayoni ma'lumotlarni to'plash, ularga ishlov berish va taqdim etishga qaratilgan operatsiyalar ketma-ketligini o'z ichiga oladi.
Axborotni qidirish masalalari
• Axborotni qidirishning asosiy masalasi - foyalanuvchiga uning axborotga bo'lgan ehtiyojlarini qondirishga yordam berishdan iborat. Asosiy masalalar:

[1] Qodirov farrux -qr kod texnalogiyasi asosida electron kutubxona tizimini dasturiy va apparat ta’minoti yaratish: 218 bet

• Modellashtirish masalasi;
• Hujjatlarni klassifikatsiyalash;
• Hujjatlarni filtrlash;
• Hujjatlarni klasterizatsiyasi;
• Qidiruv tizimlari arxitekturasi va foydalanuvchi interfesini loyihalash;
• Axborotlarni ajratib olish, xususiy holda hujjat annotatsiyasi va referatini tayyorlash;
• So'rov tillari va boshqalar.

Umumiy holda axborotni qidirish to'rtta bosqichdan tashkil topgan:
- axborotga bo'lgan ehtiyojni aniqlash va axborot so'rovini shakllantrish;
- mumkin bo'lgan axborotlar massivining egasini (manbasini) aniqlash;
- aniqlangan axborot massividan ma'lumotlarni ajratib olish;
- olingan axborot bilan tanishish va qidiruv natijasini baholash.

Axborotni ajratib olish. Axborotni ajratib olish (angl. information extraction) — bu komp'yuterda tayyorlangan aniq strukturaga ega bo'lmagan yoki kuchsiz strukturalashgan hujjatlardan aniq struturaga ega bo'lgan ma'lumotlarni avtomatik ajratib olish yoki qurish. Axborotni ajratib olish tabiiy tildagi matnlarni qayta ishlab bilan bog'liq bo'lib, axborotlarni qidirishning bir ko'rinishi hisoblanadi.[2]

(Xulosa)Savol-javob tizimlarida axborotni ajratib olish texnologiyalari oddiy kalit so'z qidiruvidan matn ma'nosini tushunish (NLU) darajasiga ko'tarildi. Tadqiqot shuni ko'rsatdiki, transformator modellari matn ichidan faktlarni ajratib olishda 90% dan ortiq aniqlik ko'rsata oladi. Kelajakda ushbu texnologiyalarni o'zbek tili NLP resurslari bazasida kengaytirish milliy qidiruv tizimlarini yaratishda poydevor bo'lib xizmat qiladi. Eng so’nggi texnik yutuqlar ko’pincha ta’lim jarayonida o’zining munosib o’rnini egallagan, bu ma’noda axborot-kommunikatsiya texnologiyalari ham istisno emas.O’quv jarayonida kompyuterlardan foydalanish bo’yicha dastlabki tajribalar hisoblash texnikasidan foydalanish ta’lim jarayoni samaradorligini sezilarli darajada oshirishi, bilimlarni hisobga olish va baholashni yaxshilashi, qiyin vazifalarni hal qilishda o’qituvchining har bir ta’lim oluvchiga yakka tartibda yordam berishini ta’minlash kabi imkoniyatlarni yaratadi.

Shuni ham aytish kerakki zamonaviy axborot texnologiyalari qo’llanilayotgan bugun har yerda, har qadamda uchratish mumkin. O’quvchi- yoshlarni zamonaviy axborot texnologiyalari va zamonaviy pedagogik texnologiyalardan foydalanishni o’rgatish, ularda o’z faoliyat sohasida yangi axborot texnologiyalari va interfaol usullardan foydalanish o’rganilayotgan mavzuning yana keng qamrovli tushunib olishga , bilim ko’nikma va malakalarning mustahkamlanishiga olib keladi.[3] Shuning bilan hozirgi davrda texnalogiyasiz muommolarni hal qilish ancha qiyin kechadi .zamonaviy texnalogiya esa har bir ishda insonga ko’makchi ekanini eslatib o’tishimiz maqsadga muvofiqdir.

Foydalanilgan adabiyotlar:

Maxsudjon Baxramov Axborot texnalogiyalari sohasida bilimlarni shakllantirish uchun interfaol interaktiv tizimni ishlab chiqish usullari
Guliziya Berdibayeva
Qodirov farrux -qr kod texnalogiyasi asosida electron kutubxona tizimini dasturiy va apparat ta’minotini yaratish.
Xurramova G.-Axborot izlash tizimlarida foydalanuvchi xatti-xarakatlarini tahlil qilish asosida tavsiya tizimlarini loyihash
D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed.).
Stanford University.Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019).
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.Manning,

[2] Maxsudjon Baxramov-Axborot texnalogiyalari sohasida bilimlarni shakllantirish uchun interfaol interaktiv tizimni ishlab chiqish usullari
[3] Guliziya Berdibayeva

C. D., Raghavan, P., & Schütze, H. (2008).
Introduction to Information Retrieval. Cambridge University Press.Sarawagi, S. (2008).
Information extraction.
Foundations and Trends in Information Retrieval, 2(4), 261-377