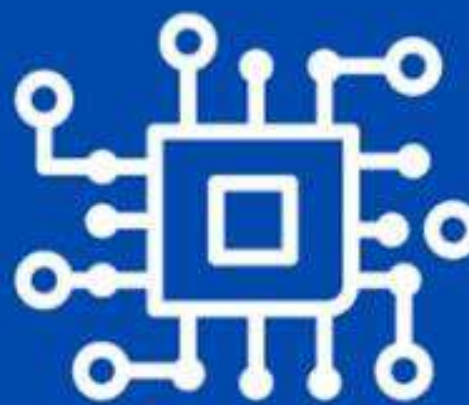




TECH SCIENCE

ISSN 3030-3702

**TEXNIKA FANLARINING
DOLZARB MASALALARI**

**TOPICAL ISSUES OF TECHNICAL
SCIENCES**



№ 9 (4) 2026

BOSH MUHARRIR:

ABBAZOV ILHOM ZAPIROVICH

TAHRIR HAY'ATI:

Usmankulov Alisher Kadirkulovich — Texnika fanlari doktori, professor, Jizzax, O'zbekiston;

Fayziyev Xomitxon – texnika fanlari doktori, professor, Toshkent, O'zbekiston;

Rashidov Yusuf Karimovich – texnika fanlari doktori, professor, Toshkent, O'zbekiston;

Adizov Bobirjon Zamirovich – Texnika fanlari doktori, professor, Toshkent, O'zbekiston;

Abdunazarov Jamshid Nurmuxamatovich — Texnika fanlari doktori, dotsent, Jizzax, O'zbekiston;

Eshonkulov Ravshan Abdurazakovich — Texnika fanlari doktori, dotsent, Qashqadaryo, O'zbekiston;

Umarov Shavkat Isomiddinovich – Texnika fanlari doktori, professor, Jizzax, O'zbekiston;

Bozorov G'ayrat Rashidovich – Texnika fanlari doktori, Buxoro, O'zbekiston;

Maxmudov Muxtor Jamolovich – Texnika fanlari doktori, Buxoro, O'zbekiston;

Asatov Nurmuxammad Abdunazarovich – Texnika fanlari nomzodi, professor, Jizzax, O'zbekiston;

Mamayev G'ulom Ibroximovich – Texnika fanlari bo'yicha falsafa doktori (PhD), Jizzax, O'zbekiston;

Ochilov Abduraxim Abdurasulovich – Texnika fanlari bo'yicha falsafa doktori (PhD), Buxoro, O'zbekiston.

OAK Ro'yxati

Mazkur jurnal O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi Oliy attestatsiya komissiyasi Rayosatining 2025-yil 8-maydagi 370-son qarori bilan texnika fanlari bo'yicha ilmiy darajalar yuzasidan dissertatsiyalar asosiy natijalarini chop etish tavsiya etilgan ilmiy nashrlar ro'yxatiga kiritilgan.

Muassislar: "SCIENCEPROBLEMS TEAM" mas'uliyati cheklangan jamiyati;
Jizzax politexnika instituti.

TECHSCIENCE.UZ- TEXNIKA**FANLARINING DOLZARB****MASALALARI** elektron jurnali

15.09.2023-yilda 130343-sonli

guvohnoma bilan davlat ro'yxatidan o'tkazilgan.

TAHRIRIYAT MANZILI:

Toshkent shahri, Yakkasaroy tumani, Kichik

Beshyog'och ko'chasi, 70/10-uy.

Elektron manzil:

scienceproblems.uz@gmail.com

Barcha huquqlar himoyalangan.

© Scienceproblems team, 2026-yil

© Mualliflar jamoasi, 2026-yil

MUNDARIJA

Ataboyeva Shaxrizoda

EMPIRIK TAHLILLAR ASOSIDA OLIY TA'LIM BOSHQARUVI VA INDIVIDUAL
TA'LIM JARAYONLARINI TAKOMILLASHTIRISHDA SUN'YI INTELLEKTGA
ASOSLANGAN OPTIMALLASHTIRISH ALGORITMLARI 4-11

Turgunbaev Rashid

METAEXTRACT - KO'P TILLI ILMIY JURNALLARDAN BIBLIOGRAFIK
METAMA'LUMOTLARNI QOIDALARGA ASOSLANGAN EKSTRAKSIYA QILISH TIZIMI 12-18

Muqimova Anaxon

ADVANCED REAL-TIME TRAFFIC SIGN SEGMENTATION AND CLASSIFICATION USING
HYBRID DEEP CONVOLUTIONAL ARCHITECTURES ON GTSDB AND BENCHMARK
DATASETS 19-24

Pulatov G'iyos, Kabildjanov Aleksandr, Pulatova Gulxayo

SUST SHAKLLANGAN JARAYONLARNI BASHORATLASHDA INTELLEKTUAL
ALGORITMLARNI QIYOSIY TAHLILI 25-33

Usarkulova Feruza

AN INTEGRATED MODEL FOR AI-BASED FORECASTING OF TECHNICAL SERVICE
DEMAND AND OPTIMAL RESOURCE ALLOCATION IN AGRO-INDUSTRIAL
COMPLEXES 34-43

Botirov Munisbek

NUMERICAL MODELING AND OPTIMIZATION OF GROUNDWATER FILTRATION
PROCESSES USING COMPUTATIONAL ALGORITHMS 44-53

Гулямов Шухрат, Зайниддинова Зебинисо, Караханова Алсу

РАЗРАБОТКА МЕТОДОВ ПОСТРОЕНИЯ ВИРТУАЛЬНЫХ ПРИБОРОВ
ЭКОЛОГИЧЕСКОГО КОНТРОЛЯ КАЧЕСТВА ПРОМЫШЛЕННОЙ ПРОДУКЦИИ 54-61

Каршиев Хумоюн, Хасанов Абдирашид

ПЕРСПЕКТИВНЫЕ СПОСОБЫ ПЕРЕРАБОТКИ КОБАЛЬТСОДЕРЖАЩИХ
ОТХОДОВ ЦИНКОВОГО ПРОИЗВОДСТВА 62-73

Xaytbayev Aybek

TAKOMILLASHTIRILGAN KUZATUVCHI QIDIRUV ALGORITMI ASOSIDA SIMSIZ SENSOR
TARMOQLARIDA ENERGIYA SAMARADOR KLASTER BOSH TUGUNINI TANLASH 74-88

Zokirov Islomjon, Nurillaev Ahadullo, Abdullaev Shavkat

INTEGRATED MATLAB/SIMULINK-SIMSCAPE MODELING AND PI-BASED CLOSED-LOOP
CONTROL OF A MICRO-HYDROPOWER PLANT 89-101

Qarshiboyev Asqar, Tursunov Ilhom

KOMPRESSOR USKUNALARINING OPTIMAL ISH REJIMINI ANIQLASH VA ELEKTR ENERGIYASI
ISTE'MOLI SAMARADORLIGINI BAHOLASH 102-112

METAEXTRACT - KO'P TILLI ILMIY JURNALLARDAN BIBLIOGRAFIK METAMA'LUMOTLARNI QOIDALARGA ASOSLANGAN EKSTRAKSIYA QILISH TIZIMI

Turgunbaev Rashid

Qo'qon davlat universiteti

ORCID: 0000-0003-1443-7900

Email: atmdsmi@gmail.com

Annotatsiya. Ushbu maqolada ilmiy jurnal maqolalaridan bibliografik metama'lumotlarni (sarlavha, mualliflar, annotatsiya, kalit so'zlar, DOI va boshqalar) avtomatik ekstraktsiya qilish uchun ishlab chiqilgan qoidalarga asoslangan dasturiy tizimning arxitekturasini va amalga oshirilishi tavsiflanadi. Tizim ikkita o'zaro bog'liq, ammo mustaqil ishga tushiriladigan Java Swing desktop dasturidan - qoida yaratish vositasi (PDFTextLabeler) va ekstraktsiya vositasi (MetaExtract) - iborat bo'lib, ular umumiy ExtractionConfig deb nomlangan, Jackson kutubxonasi yordamida JSON formatiga serializatsiya qilinadigan konfiguratsiya sxemasi orqali bog'lanadi. Foydalanuvchi jurnalning bitta namunaviy maqolasida metama'lumot maydonlarini geometrik (to'rtburchak) va chiziqli (kolontitul/adabiyotlar chegarasi) rejimlarda belgilaydi, so'ngra hosil bo'lgan qoidalar shu jurnalning istalgan sonidagi barcha maqolalariga qayta-qayta qo'llaniladi. Muallif ismlarini aniqlash ishonchligini oshirish maqsadida geometrik qoidalar Apache OpenNLP kutubxonasining lotin va kirill yozuvlari uchun o'qitilgan NameFinderME modeli bilan birlashtirilgan aralash rejimda ishlaydi. Tizim skanerlangan hujjatlar uchun Tess4J (Tesseract OCR) mexanizmini, matn ajratishning zaxira yo'li sifatida iText kutubxonasini hamda Open Journal Systems (OJS) platformasiga to'g'ridan-to'g'ri import qilinadigan XML eksportini qo'llab-quvvatlaydi. Maqolada tizimning to'rt qatlamli arxitekturasini, ExtractionConfig sxemasining ichki tuzilishi, asosiy dasturiy modullar va ular orasidagi ma'lumotlar oqimi, shuningdek amalga oshirish jarayonida yuzaga kelgan bir qator amaliy muammolar (masalan, tipografik tutuq belgilarining grafik variativligi) va ularning yechimlari batafsil yoritiladi.

Kalit so'zlar. metama'lumotlarni ekstraktsiya qilish; qoidalarga asoslangan tizim; Apache PDFBox; Apache OpenNLP; NameFinderME; Java Swing; Open Journal Systems; kam resursli tillar.

METAEXTRACT - A RULE-BASED EXTRACTION SYSTEM FOR BIBLIOGRAPHIC METADATA FROM MULTILINGUAL SCIENTIFIC JOURNALS

Turgunbaev Rashid

Kokand State University

Annotation. This paper describes the architecture and implementation of a rule-based software system designed to automatically extract bibliographic metadata (title, authors, annotation, keywords, DOI, etc.) from scientific journal articles. The system consists of two interrelated but independently executable Java Swing desktop applications - a rule generator (PDFTextLabeler) and an extraction tool (MetaExtract), which are connected through a common configuration schema called ExtractionConfig, serialized to JSON format using the Jackson library. The user defines metadata fields in geometric (rectangular) and linear (header/reference border) modes for a single sample journal article, and then the resulting rules are applied repeatedly to all articles in any issue of this journal. To increase the reliability of author name recognition, geometric rules work in a mixed mode, combined with the NameFinderME model trained for Latin and Cyrillic scripts from the Apache OpenNLP library. The system supports the Tess4J (Tesseract OCR) engine for scanned documents, the iText library as a fallback for text extraction, and XML export for direct import to the Open Journal Systems (OJS) platform. The article describes in detail the four-layer architecture of the system, the internal structure of the ExtractionConfig scheme, the main software modules and the data flow between them, as well as a number of practical problems that arose during the implementation (for example, graphic variability of typographical punctuation marks) and their solutions.

Keywords: metadata extraction; rule-based system; Apache PDFBox; Apache OpenNLP; NameFinderME; Java Swing; Open Journal Systems; low-resource languages.

DOI: <https://doi.org/10.47390/ts-v4i9y2026N02>

Kirish

Ilmiy jurnal maqolalaridan bibliografik metama'lumotlarni (sarlavha, mualliflar, annotatsiya, kalit so'zlar, adabiyotlar ro'yxati va h.k.) qo'lda kiritish tahririyat xodimlari va kutubxonachilar uchun sezilarli vaqt va resurs sarflaydigan jarayon hisoblanadi, ayniqsa jurnal arxivi o'nlab yoki yuzlab sonlarni o'z ichiga olgan hollarda. Bu muammoni hal qilish uchun ikkita asosiy yondashuv mavjud: chuqur neyron tarmoqlarga asoslangan universal ekstraktorlar (masalan, GROBID, CERMINE [1]) va qoidalarga asoslangan tizimlar [4]. Universal modellar odatda ingliz tilidagi yirik annotatsiyalangan korpuslar asosida o'qitilgan bo'lib [2], o'zbek lotin, o'zbek kirill va rus tilidagi ilmiy nashrlar uchun bunday hajmdagi tayyor o'quv ma'lumotlari mavjud emas. Shu sababli kam resursli tillar uchun qoidalarga asoslangan yondashuv nafaqat amaliy, balki hozircha real ishlaydigan yagona alternativ hisoblanadi [12].

Ushbu maqolada taqdim etilayotgan MetaExtract tizimi aynan shu muammoga javob sifatida ishlab chiqilgan. Tizimning asosiy g'oyasi shundan iboratki, har bir ilmiy jurnal o'z maqolalarida ma'lum bir standartlashtirilgan joylashtirishga ega bo'ladi, bu esa jurnalga xos qoidalar to'plamini bir marta yaratib, uni jurnalning qolgan barcha maqolalariga qayta-qayta qo'llash imkonini beradi. Maqolaning maqsadi tizimning dasturiy arxitekturasini, texnologik yechimlarini va asosiy modullarini muhandislik nuqtai nazaridan tahlil qilishdir.

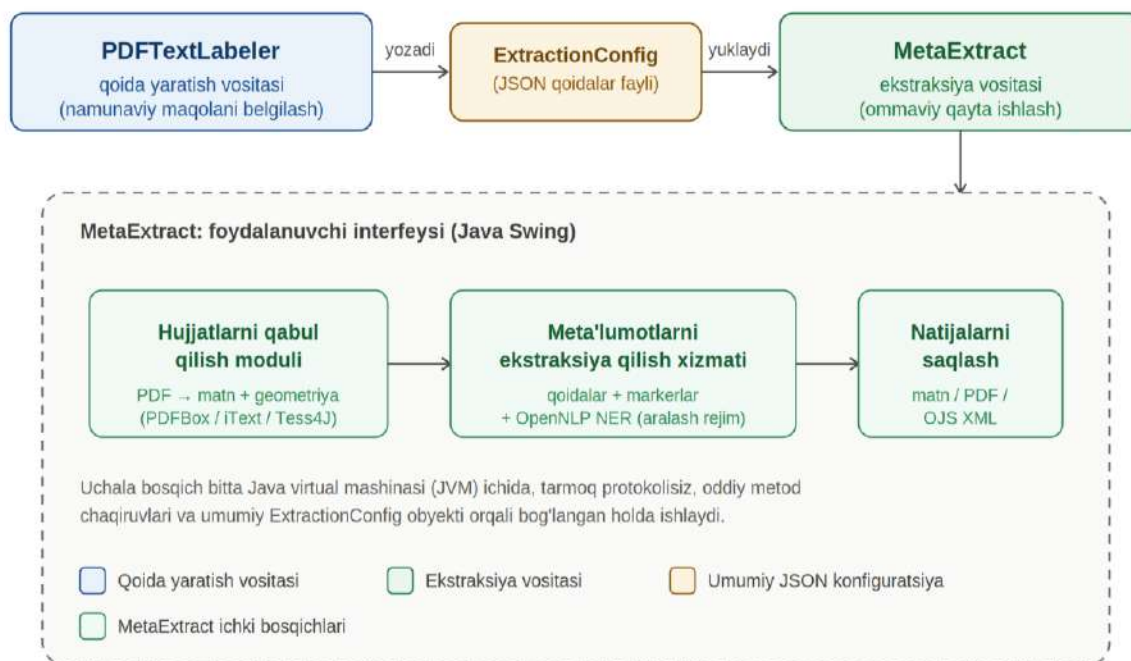
Tizim arxitekturasini

Tizim yagona Java Virtual Mashinasi doirasida ishlaydigan, ammo mustaqil ishga tushiriladigan ikkita hamkor dasturdan iborat - PDFTextLabeler (qoida yaratish vositasi) va MetaExtract (ekstraksiya vositasi) [6]. Ikkala dastur ham ExtractionConfig deb nomlangan Java klasslari to'plami orqali bog'langan bo'lib, bu klasslar Jackson kutubxonasining ObjectMapper vositasi yordamida JSON faylga serializatsiya qilinadi. Bunday fayl-asoslangan integratsiya tizimni tarmoqdan mustaqil, o'rnatilishi va tarqatilishi sodda qiladi hamda qoidalar faylini istalgan matn muharriri orqali qo'lda tahrirlash imkonini saqlab qoladi.

Tizimning matn va geometrik ma'lumotlarni ekstraksiya qilish qatlami Apache PDFBox kutubxonasiga asoslangan; PDFBox orqali matnni to'g'ri ajratib bo'lmagan holatlar uchun iText kutubxonasi zaxira mexanizmi sifatida ishlatiladi [7]. Skanerlangan (matn qatlamisiz) hujjatlar uchun Tess4J orqali Tesseract OCR mexanizmi ishga tushiriladi. Muallif ismlarini tekshirishda geometrik va marker so'zlarga asoslangan qoidalarga qo'shimcha sifatida Apache OpenNLP kutubxonasining lotin va kirill yozuvlari uchun o'qitilgan langdetect va NameFinderME modellari qo'llaniladi - bu "aralash rejim" tizimga faqat joylashuvga emas, balki qo'shimcha lingvistik signalga ham tayanish imkonini beradi [3]. Foydalanuvchi interfeysi Java Swing kutubxonasi asosida qurilgan, loyihaning barcha tashqi bog'liqliklari esa Apache Ant qurish tizimi orqali boshqarilib, yakuniy JAR fayl ichiga birlashtiriladi [9].

Shunday qilib, tizim arxitekturasini to'rtta funksional qatlamga ajratish mumkin - (1) hujjatlarni qabul qilish va dastlabki qayta ishlash (PDF → matn + geometriya), (2) metama'lumotlarni ekstraksiya qilish (qoidalar + marker so'zlar + NER), (3) qoidalar va natijalarni JSON/matn/PDF/XML fayllar ko'rinishida saqlash, va (4) Java Swing asosidagi foydalanuvchi interfeysi. Bu to'rt qatlam bitta Java ilovasi ichida metodlar va klasslar

ko‘rinishida amalga oshirilgan bo‘lib, ular orasidagi bog‘lanish oddiy metod chaqiruvlari va umumiy ExtractionConfig obyektini orqali ta’minlanadi (1-rasm).



1-rasm. MetaExtract tizimining umumiy arxitekturasini: PDFTextLabeler va MetaExtract dasturlarini bog‘lovchi ExtractionConfig (JSON) fayli hamda MetaExtract ichidagi uch bosqichli qayta ishlash ketma-ketligi

ExtractionConfig konfiguratsiya sxemasi

Tizimning markaziy saqlash birligi - har bir jurnal uchun alohida yaratiladigan ExtractionConfig JSON fayli (“qoidalar fayli”) hisoblanadi [5]. Ushbu sxema quyidagi asosiy tarkibiy qismlardan iborat: jurnal nomi (journalName), ekstraksiya ishonchlik ko‘rsatkichi (confidence) va besh kichik sxemaga bo‘lingan qoidalar bloki (rules): maqola strukturasi (articleStructure), metama’lumot markerlari (metadataMarkers), joylashuv qoidalari (layoutRules), OCR sozlamalari (ocrSettings) va chiqish formatlari sozlamalari (outputOptions).

Metama’lumot markerlari (metadataMarkers) annotatsiya, kalit so‘zlar va adabiyotlar bo‘limlarini matn ichida aniqlash uchun ishlatiladigan so‘z-belgilarni saqlaydi; har bir marker turi to‘rtta til varianti - o‘zbek lotini, o‘zbek kirill, rus va ingliz - uchun alohida saqlanadi, chunki ko‘p tilli jurnallar bitta maqolada annotatsiya va kalit so‘zlarni bir necha tilda ketma-ket keltiradi [10]. Joylashuv qoidalari (layoutRules) sarlavha, annotatsiya va kalit so‘zlar uchun uzunlik chegaralarini hamda yuqori/quyi kolontitullarning Y-koordinatalarini saqlaydi - geometrik chegara matn mazmunidan qat’i nazar, ma’lum joylashuv zonasidagi barcha qatorlarni izchil ravishda chiqarib tashlash imkonini beradi, bu matn-marker asosidagi yondashuvdan (masalan, “sahifa raqami” so‘zini qidirish) ko‘ra ishonchliroq, chunki quyi kolontitulda ko‘pincha faqat raqam yoki qisqartma keladi (1-jadval).

1-jadval

Qoidalar fayli (ExtractionConfig)ning asosiy maydonlari

Maydon	Turi	Tavsifi
journalName	String	Jurnal nomi

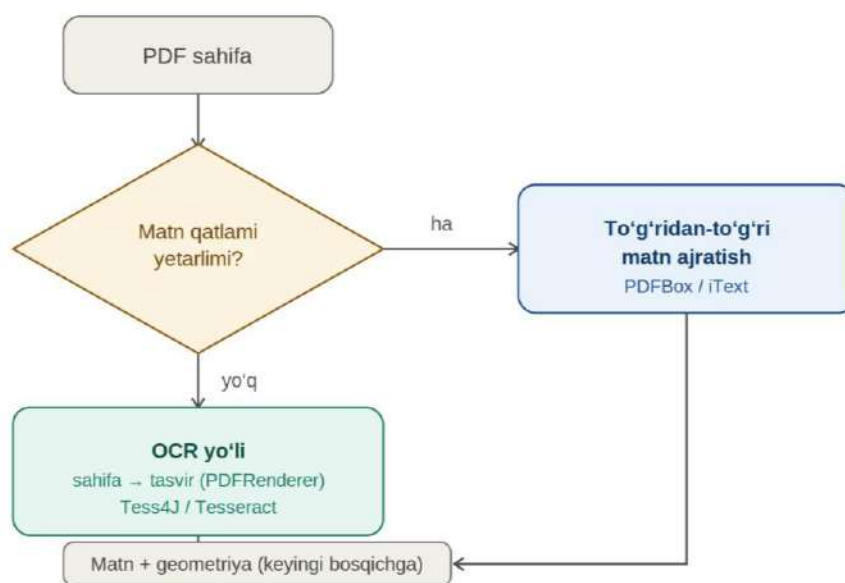
confidence	double	Ekstraksiya ishonchlilik ko'rsatkichi (status panelida ko'rsatiladi)
articleStructure.totalArticles	int	Jurnaldagi maqolalar soni
articleStructure.startPage / endPage	int	Ekstraksiya boshlanadigan va tugaydigan sahifalar (endPage = 0 — hujjat oxirigacha)
articleStructure.sectionMapping	List<SectionMapping>	OJS bo'lim nomi va unga tegishli maqolalar diapazoni
metadataMarkers.{abstract, keyword,reference}Markers	LanguageMarkers	Har biri 4 til varianti: o'zbek lotin, o'zbek kirill, rus, ingliz
metadataMarkers.doi.enabled	boolean	DOI qidiruvini yoqish/o'chirish (naqshning o'zi dasturda qattiq belgilangan)
layoutRules.{title,abstract, keyword}Extraction.*Length	int	Sarlavha, annotatsiya va kalit so'z uchun uzunlik chegaralari
layoutRules.headerY / footerY / dpi	int	Yuqori/quyi kolontitul Y-koordinatasi va aks ettirish aniqligi
ocrSettings.enabled / languages	boolean / String[]	OCR yoqilganligi va tanlangan tillar ro'yxati
outputOptions.textFiles.enabled	boolean	Oddiy matn fayllari chiqishi
outputOptions.researchBible.{enabled,website}	boolean / String	"Research Bible" formatidagi chiqish
outputOptions.ojs321 / ojs340	OJSSettings	OJS 3.2.1 / 3.4.0 uchun XML eksport sozlamalari (jild, son, yil va h.k.)

Jackson kutubxonasining standart xatti-harakati amaliy ahamiyatga ega: agar qoidalar fayli eski formatda bo'lib, unda yangi qo'shilgan maydon mavjud bo'lmasa, ObjectMapper bunday maydon uchun sinf ichida belgilangan standart qiymatni qo'llaydi va yuklashni xatosiz yakunlaydi. Bu, tizim rivojlanishi davomida sxemaga yangi sozlamalar qo'shilganda ham, avval yaratilgan eski qoidalar fayllarining ishlashda davom etishini ta'minlaydi - bu dasturiy ta'minotning orqaga qarab moslashuvchanligi (backward compatibility) nuqtai nazaridan muhim dizayn qarori hisoblanadi.

Asosiy dasturiy modullar

Hujjatlarni qabul qilish moduli tizimning kirish nuqtasi bo'lib, foydalanuvchi tanlangan PDF fayl PDFBox kutubxonasi (PDDocument) yordamida ochiladi. Har bir matn bo'lagi (TextPosition obyekti) o'zining x va y koordinatalari, kengligi, balandligi hamda shrift o'lchami bilan birga qaytariladi - aynan shu geometrik ma'lumotlar keyinchalik qoidalarini qo'llash bosqichida maydonlarning joylashuvini aniqlash uchun asos bo'lib xizmat qiladi (2-rasm). Agar sahifadan ajratib olingan matn hajmi kutilganidan sezilarli darajada kam bo'lsa, tizim OCR yo'lga o'tadi - PDFRenderer yordamida sahifa BufferedImage tasviriga aylantiriladi va Tess4J orqali Tesseract chaqiriladi. Tizim OCR yo'lini faqat zarur bo'lgan hollarda ishga tushiradi,

chunki OCR to'g'ridan-to'g'ri matn qatlamidan o'qishga qaraganda sezilarli darajada sekinroq va natija sifati jihatidan kamroq ishonchli.



2-rasm. Hujjatlarni qabul qilish modulida matn ajratish yo'lini tanlash sxemasi, matn qatlami yetarli bo'lsa to'g'ridan-to'g'ri ekstraksiya, aks holda OCR yo'li ishga tushadi

Metama'lumotlarni ekstraksiya qilish mantig'i ikki bosqichda ishlaydi: qoida yaratish (PDFTextLabeler ichida) va qoidani qo'llash (MetaExtract ichida). Sarlavha va mualliflarni aniqlashda geometrik mezonlar (sahifaning yuqori qismi, markazlashtirilgan joylashuv, shrift o'lchami) ustuvor rol o'ynaydi, muallif ismlarini tekshirishda esa qo'shimcha ravishda OpenNLP NameFinderME modeli ishlatiladi [8]. Annotatsiya va kalit so'zlarni ajratishda tizim marker so'zga mos keluvchi qatorni topib, undan keyingi matnni navbatdagi marker uchraguncha to'playdi. Amalga oshirish jarayonida aniqlangan real muammolardan biri - turli tipografik tutuq belgilarining (Unicode variantlarining) faqat ba'zilar tan olinishi natijasida ba'zi marker so'zlarning to'g'ri aniqlanmasligi edi; bu muammo mavjud normalizatsiya funksiyasini marker taqqoslashning barcha nuqtalarida izchil qo'llash orqali bartaraf etildi. DOI identifikatorini aniqlash uchun standart "10.xxxx/yyyy" ko'rinishidagi muntazam ifoda (regex) qo'llaniladi.

Ma'lumotlarni saqlash xizmati Jackson kutubxonasi asosidagi JSON serializatsiya/deserializatsiya mexanizmi hamda natijalarni fayl ko'rinishida yozib chiquvchi moduldan iborat. Natijalarni yozish qismi outputOptions sozlamalariga qarab bir nechta formatni bir vaqtning o'zida hosil qilishi mumkin: har bir maqola uchun alohida matn fayli, ajratib olingan alohida PDF fayl (copyPDF funksiyasi), "Research Bible" formatidagi chiqish hamda OJS ning 3.2.1 va 3.4.0 versiyalariga mos XML fayllar. OJS uchun hosil qilingan XML fayl "Native XML plugin" moduliga yuklanadi, so'ngra OJS ushbu faylni o'qib, maqolalarni tegishli bo'lim, jild va sonlarga avtomatik joylashtiradi.

Muhokama

Taqdim etilgan arxitekturaning bir qator muhandislik jihatdan qiziqarli xususiyatlari mavjud. Birinchidan, ikkita mustaqil dastur (PDFTextLabeler va MetaExtract) o'rtasidagi integratsiyani tarmoq protokoli emas, balki oddiy fayl formati (JSON) orqali amalga oshirish qarori kutubxona va jurnal tahririyatlari kabi IT-infratuzilmasi cheklangan tashkilotlar uchun amaliy jihatdan asoslangan, chunki bunday tashkilotlarda odatda alohida server yoki tarmoq

xavfsizligi bo'yicha mutaxassislar mavjud bo'lmaydi [4]. Ikkinchidan, geometrik qoidalar va statistik NER signalini birlashtirgan aralash rejim faqat geometriyaga yoki faqat marker so'zlarga tayanishning har birining alohida kamchiliklarini qoplaydi: faqat geometriya biroz burilib skanerlangan yoki formatlash jihatidan farq qiluvchi sahifalarda xato berishi mumkin, faqat marker so'zlar esa marker matnda takror uchragan hollarda chalkashlik keltirib chiqarishi mumkin [11].

Tizimning cheklovlari ham mavjud - qoidalar bir jurnalga xos bo'lgani sababli, yangi jurnal uchun har safar namunaviy maqolada dastlabki belgilash talab qilinadi, bu esa universal, o'qitishni talab qilmaydigan modellarga nisbatan qo'shimcha bosqich hisoblanadi [10]. Bundan tashqari, normalizatsiya funksiyasi hozirda faqat ba'zi matn-taqqoslash nuqtalarida qo'llanilgan bo'lib, uni muallif ismlari va manzillarni aniqlashning barcha nuqtalarida izchil qo'llash tizimning navbatdagi takomillashtirish yo'nalishlaridan biri sifatida belgilangan.

Xulosa

MetaExtract tizimi ko'p tilli (o'zbek lotin, o'zbek kirill, rus, ingliz) ilmiy jurnallardan bibliografik metama'lumotlarni ekstraksiya qilish uchun qoidalarga asoslangan, ammo statistik NER signali bilan boyitilgan gibrid arxitekturaga ega [8]. Ikkita hamkor Java Swing dasturi, umumiy JSON konfiguratsiya sxemasi va OJS bilan to'g'ridan-to'g'ri XML integratsiyasi tizimni o'rnatish va joriy etish jihatidan sodda, ammo amaliy jihatdan samarali yechimga aylantiradi. Tizimning kuchli tomoni texnologik murakkabligida emas, balki soddaligi va aynan maqsadli vazifaga - kam resursli tillardagi ilmiy jurnallardan metama'lumotlarni ishonchli tarzda ekstraksiya qilishga - aniq yo'naltirilganligida namoyon bo'ladi.

Adabiyotlar/Литература/References:

1. Tkaczyk D, Szostek P, Fedoryszak M, Dendek PJ, Bolikowski Ł. CERMINE: automatic extraction of structured metadata from scientific literature. *International Journal on Document Analysis and Recognition (IJDAR)*. 2015 Dec;18(4):317-35.
2. Kong H, Yoon H, Seol J, Hyun M, Lee H, Kim S, Choi W. Annotated open corpus construction and BERT-based approach for automatic metadata extraction from Korean academic papers. *IEEE Access*. 2022 Dec 29;11:825-38.
3. Ahmed MW, Afzal MT. FLAG-PDFe: Features oriented metadata extraction framework for scientific publications. *IEEE Access*. 2020 May 27;8:99458-69.
4. Skluzacek TJ, Chard K, Foster I. Automated metadata extraction: challenges and opportunities. In *2022 IEEE 18th International Conference on e-Science (e-Science) 2022 Oct 11 (pp. 495-500)*. IEEE.
5. Suryani MA, Hahne S, Beth C, Wallmann K, Renz M. DAF: Data Acquisition Framework to Support Information Extraction from Scientific Publications. In *KDIR 2023 (pp. 468-476)*.
6. Hashmi AM, Afzal MT, ur Rehman S. Rule based approach to extract metadata from scientific PDF documents. In *2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA) 2020 Nov 25 (pp. 1-4)*. IEEE.
7. Alamoudi A, Alomari A, Alwarthan S. A rule-based information extraction approach for extracting metadata from PDF books. *ICIC Express Letters, Part B: Applications*. 2021 Feb;12(2):121-32.
8. Raval P, Bhaidasna H. Metadata Extraction From Scholarly Document Using Deep Learning. In *2025 3rd International Conference on Inventive Computing and Informatics (ICICI) 2025 Jun 4 (pp. 1-5)*. IEEE.

9. Ingram WA, Wu J, Kahu SY, Manzoor JA, Banerjee B, Ahuja A, Choudhury MH, Salsabil L, Shields W, Fox EA. Building datasets to support information extraction and structure parsing from electronic theses and dissertations. *International Journal on Digital Libraries*. 2024 Jun;25(2):175-96.
10. Skluzacek TJ, Chen M, Hsu E, Chard K, Foster I. Models and metrics for mining meaningful metadata. In *International Conference on Computational Science 2022 Jun 15* (pp. 417-430). Cham: Springer International Publishing.
11. Boukhers Z, Bouabdallah A. Vision and natural language for metadata extraction from scientific PDF documents: A multimodal approach. In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries 2022 Jun 20* (pp. 1-5).
12. T. J. Skluzacek, "Automated metadata extraction can make data swamps more navigable," Ph.D. dissertation, The University of Chicago, 2022.

TECHSCIENCE.UZ

**TEXNIKA FANLARINING DOLZARB
MASALALARI**

№ 9 (4)-2026

TOPICAL ISSUES OF TECHNICAL SCIENCES

**TECHSCIENCE.UZ- TEXNIKA
FANLARINING DOLZARB MASALALARI**
elektron jurnali 15.09.2023-yilda 130346-
sonli guvohnoma bilan davlat ro'yxatidan
o'tkazilgan.

Muassislar: "SCIENCEPROBLEMS TEAM"
mas'uliyati cheklangan jamiyati;
Jizzax politexnika instituti.

TAHRIRIYAT MANZILI:

Toshkent shahri, Yakkasaroy tumani, Kichik
Beshyog'och ko'chasi, 70/10-uy.

Elektron manzil:

scienceproblems.uz@gmail.com