TOSHKENT MENEJMENT VA IQTISODIYOT INSTITUTI

MANAGEMENT AND ECONOMICS SCIENTIFIC RESEARCH JOURNAL

2-TO'PLAM
1-MAXSUS SON/2025

MANAGEMENT AND ECONOMICS SCIENTIFIC RESEARCH JOURNAL

MENEJMENT VA IQTISODIYOT ILMIY-TADQIQOT JURNALI

НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЖУРНАЛ МЕНЕДЖМЕНТА И ЭКОНОМИКИ

journals@timeedu.uz

+998 95 651 90 00

MANAGEMENT AND ECONOMICS
SCIENTIFIC RESEARCH JOURNAL

+998 95 651 90 00
journals@timeedu.uz
Shota Rustaveli ko'chasi, 114
tmijournals.ndc-agency.uz
@TMII_ilmiy_bolim

MUHANDISLIK OYLIGI

“RAQAMLI TEXNOLOGIYALARNI RIVOJLANTIRISH VA INNOVATSION MUHANDISLIK YECHIMLARI: ZAMONAVIY TENDENSIYALAR VA ISTIQBOLLAR” MAVZUSIDA RESPUBLIKA ILMIY-AMALIY ANJUMANI TO’PLAMI MATERIALLARI

17-18 APREL, 2025
TOSHKENT

MUNDARIJA

1.	**Primova X.A., Xaniyev N.M., Nodirov M.T.** RFID TEXNOLOGIYASI ORQALI OLINGAN MA'LUMOT ASOSIDA QORAMOL KASALLIKLARINI ANIQLASHNING LOYIHALASH	6
2.	**Doshanova M.Y., Otaxanova B.I., Aliyev R.R.** USING ARTIFICIAL INTELLIGENCE TO AUTOMATE THE PROCESS OF COLLECTING AND ANALYZING DATA FROM ONLINE JOB POSTINGS	10
3.	**Eshonqulov E.Sh., Abdiyeva H.S., Abdurahmonov M.S.** SUN'IY YO'LDOSH TASVIRLARIDA PANSHARPENING YONDASHUVI NATIJALARINI BAHOLASH BAYONNOMALARI	23
4.	**Abduraxmanova Z.T.** O'ZBEKISTONDA QISHLOQ XO'JALIGINING RAQAMLI TRANSFORMATSIYASI	26
5.	**Abdiyeva K.S., Shamsiyeva K., Oblokulov S.** GOMPERTZ ALGORITHM FOR THE DETECTION OF BREAST CANCER USING MAMMOGRAPHIC IMAGES	31
6.	**Abdiyeva K.S, Shamsiyeva K., Olimjonova S.** TOPOGRAPHIC MAP BASED SEGMENTATION OF MEDICAL IMAGES	35
7.	**Radjabov B.A.** O'ZBEKISTONNING TASHQI SAVDODA LOGISTIKA XARAJATLARI: RAQAMLI YECHIMLAR VA ULARNING EKSPORT RAQOBATBARDOSHLIGIGA TA'SIRI	39
8.	**Primova X.A., Vaydullayeva M.F.** SUN'IY INTELLEKTNI TALQIN QILISH VA QAROR QABUL QILISH: NORAVSHAN FIKRLASH TIZIMLARINI HISOBLASH MODELLARINI CHUQUR O'RGANISH	49
9.	**Ulashov A.E., Obilov H.X.** MATEMATIK MASALALARNI ZAMONAVIY DASTURLASH TILLARI YORDAMIDA YECHISH	57
10.	**Mamatov N.S., Asqarov A.A.** HAVO SIFATI INDEKSINI BASHORATLASHNING REGRESSION MODELI	69
11.	**Mirpulatova L.M.** AXBOROT TEXNOLOGIYALARI VA RAQAMLI TRANSFORMATSIYANING BANK FAOLIYATIDAGI ROLI	75
12.	**Ниёзматова Н.А., Жалелов Қ.А., Абдуллаева Б.М.** НУТҚ СИГНАЛЛАРИДАН ШОВҚИНЛАРНИ БАРТАРАФ ЭТИШДА ФИЛЬТРЛАР ЖУФТЛИГИНИ ҚЎЛЛАШ ЁНДАШУВИ	84
13.	**Mamatov N.S., Jo'rayev I.A, Jalelova M.M.., Usarov J.A.** TIBBIY TASVIRLAR BAZALARI TAHLILI	88
14.	**Mamatov N.S., Jo'rayev I.A, Jalelova M.M., Jumayev B.J.** TIBBIY TASVIRLARNI TAHLIL QILISH USUL VA ALGORITMLARI	92
15.	**Mamatov N.S., Jalelova M.M., Jo'rayev I.A., Turakulova Sh.A.** TIBBIY 2D TASVIRLAR ASOSIDA 3D TASVIRLARNI SHAKLLANTIRISH USULLARI	96
16.	**Niyozmatova N.A., Madaminjonov A.D.** INSON HIS-TUYG'ULARINI ANIQLASH DASTURIY VOSITALARI VA QURILMALARI	99

17.	**Мухамедиева Д.Т., Маматов Н.С., Маматов А.А.** МАТНЛИ МАЪЛУМОТЛАРНИ ТАСНИФЛАШДА БЕЛГИЛАРНИ ШАКИЛЛАНТИРИШ УСУЛ ВА АЛГОРИТМЛАРИ	103
18.	**Raximov M.** KASB KOMPETENSIYASIDA RAQAMLI TA'LIM TRANSFORMATSIYASI	106
19.	**Turakulov O.X.** MATNLI MA'LUMOTLARNI QAYTA ISHLASH BOSQICHLARI	110
20.	**Turakulov O.X., Jalelov R.M.** MATNLI MA'LUMOTLARGA DASTLABKI ISHLOV BERISH ALGORITMLARI	116
21.	**Мухамедиева Д.К.** ПОСТРОЕНИЕ ДЕРЕВА МЕРКЛА	123
22.	**Mamatov N.S., Nuritdinov N.D., Almuradova N.A.** IJTIMOIY TIZIMLARDA RAQAMLI TRANSFORMATSIYA JARAYONLARINI MODELLASHTIRISH	127
23.	**Kabildjanov A.S., Pulatov G'.G.** OB-HAVO OMILLARINING YURAK-QON TOMIR KASALLIKLARIGA TA'SIRINING SUN'IY INTELLEKT ASOSIDA TAHLILI	130
24.	**Mamatov N.S., Kodirov E.S** KAFT TASVIRI ASOSIDA SHAXSNI TANIB OLISH YONDASHUVLARI TAHLILI	133
25.	**Raximov I.M.** MAHALLIY ISHLAB CHIQARUVCHILARNING XALQARO STANDARTLARGA MOSLASHUVI – IMKONIYATLAR, YUTUQLAR VA BARQAROR STRATEGIK YECHIMLAR	136
26.	**Raximov F.J.** TABIIY VA SUN'IY TOLALAR ASOSIDA QAYTA ISHLANGAN MAHSULOTLARNING TEXNOLOGIK SIFATI VA IQTISODIY SAMARADORLIGI	143
27.	**Mamatov N.S., Dusanov X.T.** NUTQ SIGNALINI XALAQITLARDAN TOZALASH USULLARI	150
28.	**Dauletov A.Y., Sabirova S.T., Oydinov S.SH.** ARXIVDA MATNLARNI QIDIRISH USULLARI TAHLILI	155
29.	**Toshpulatov M.** UNVEILING TOMORROW: EXPLORING THE FRONTIERS OF INNOVATION	162
30.	**Маматов Н.С., Ниёзматова Н.А., Тожибоева Ш.Х., Яхяев Б.Ю., Машанпин Т.В.** ИНСОН ҲИС-ТУЙҒУЛАРИНИ ЮЗ ТАСВИРИ ОРҚАЛИ АНИҚЛАШ ДАСТУРИЙ ВОСИТАЛАРИ ВА ҚУРИЛМАЛАРИ	165
31.	**Ниёзматова Н.А., Маматов Н.С., Турғунова Н.М.** МАТНЛИ МАЪЛУМОТЛАРНИ ТАСНИФЛАШ УСУЛ ВА АЛГОРИТМЛАРИ ТАҲЛИЛИ	172
32.	**Маматов Н.С., Валижонов И.Х., Шукруллоев Б.Р.** ТАСВИРЛАРНИ ТАСНИФЛАШДА ҚЎЛЛАНИЛАДИГАН УСУЛЛАР	185
33.	**Niyozmatova N.A., Xoitqulov A.A., Mamatov A.A.** MATNLI MA'LUMOTLARDAN FAKTLARNI AJRATISH TEXNOLOGIYALARI	178
34.	**Buribaeva Z.R.** DIGITAL TECHNOLOGIES AND ARTIFICAL INTELLIGIENCE IN EDUCATION: CHALLENGES AND OPPORTUNITIES	181

35.	**Anarova Sh.A., Ibrohimova Z.E., Boliyeva D.N.** R-FUNKSIYA USULINI (RFM) QO‘LLAGAN HOLDA KO‘P O‘LCHOVLI FRAKTAL TUZILISHDAGI TASVIRLARNI GEOMETRIK MODELLARINI ISHLAB CHIQISH	188
36.	**Ibrohimova Z.E., Qarshiboyev J.O‘.,Xaniyev N.M.** KOMPYUTER GRAFIKASINING GEOMETRIK ALMASHTIRISHLARI ASOSIDA MURAKKAB FRAKTAL TUZILISHLARNI MATEMATIK VA RAQAMLI MODELLASHTIRISHNI AMALGA OSHIRISH	192
37.	**Xalimov A.A.** SANOAT KORXONALARIDA QAYTA TIKLANUVCHI ENERGIYA MANBALARIDAN FOYDALANISH YO‘LLARI	198
38.	**Rustamova M.M.** INNOVATIVE TECHNOLOGIES AND INVESTMENT EFFICIENCY IN COTTON-TEXTILE CLUSTERS: OPTIMIZING COSTS THROUGH INDUSTRY 4.0 SOLUTIONS.	204
39.	**Suyarov A.M.** AXBOROTLARNI VIZUAL TAQDIM ETISHDA INFOGRAFIKA VOSITALARINI YARATISH USULLARI	214
40.	**Saminjonov N.A.** CHAKANA SAVDO FAOLIYATIDA RAQAMLI VA NEYROMARKETING STRATEGIYALARINING INTEGRATSIYASI	219
41.	**Alimov A.A.** INDUSTRY 4.0 TALABLARI ASOSIDA MUHANDISLIK TA’LIMINI TASHKIL ETISH MASALALARI	224
42.	**Sobirov A.A., Zaripbayeva A.A.** SANOAT KLASTERLARINI SHAKLLANISH MEXANIZMLARI VA ZAMONAVIY MODELLARI	228
43.	**Sharipov K.A., Karimov V.A.** AQLLI ENERGIYA TIZIMLARI VA QAYTA TIKLANADIGAN ENERGIYA MANBALARIDAN FOYDALANISH USULLARI VA SAMARADORLIGI	237
44.	**Маматов Н.С., Самижонов А.Н., Самижонов Б.Н.** ИНСОН ЭҲТИЁЖЛАРИНИ АНГЛОВЧИ ВА УЛАРГА ЖАВОБ БЕРИШ ҚОБИЛИЯТИГА ЭГА ЭМПАТИК РОБОТЛАР	246
45.	**Yusupov R.A., G‘ulomov A.X.** MAYSALARNI O‘RISH VA QOR TOZALASH MASHINASINI MASOFADAN BOSHQARISHNI AVTOMATLASHTIRISH	251
46.	**Yusupov R.A., Qurbonov Sh.B.** HELIX: FOYDALANUVCHIGA MOSLASHA OLADIGAN, KUNDALIK HAYOTDA YORDAM BERUVCHI SUN’IY INTELLEKT	254
47.	**Haydarov E.D., Dusmurodov Q.A.** OCHIQ WI-FI TARMOQLAR XAVFSIZLIGI	257
48.	**Haydarov E.D., Dusmurodov Q.A.** SIMSIZ TARMOQDA XAVFSIZLIK	265
49.	**Ro‘ziyeva F.K.** MAISHIY KIMYO TOVARLARI B2B TAQSIMOTIDA KO‘P KANALLI MARKETING STRATEGIYASINING AHAMIYATI	269
50.	**Qudratov U.I., Abdullayeva N.R.** MASHINANI O‘QITISH USULLARIDAN FOYDALANGAN HOLDA ZARARLI DASTURLARNI ANIQLASH	277

51.	**Muhamediyeva D.T., Raupova M.** ELLIPTIK EGRI CHIZIQ ASOSIDA XAVFSIZ KALIT ALMASHINUVI	279
52.	**Muhamediyeva D.T., Raupova M.** POST-KVANT KRIPTOGRAFIYA ASOSIDA POLINOMLAR BILAN SHIFRLASH	282
53.	**Muhamediyeva D.T., Raupova M.** KVANT KALIT ALMASHINUVI BILAN INTEGRATSIYALANGAN RSA SHIFRLASH	285
54.	**Mamayusupov A.A.** ZAMONAVIY RAQAMLI MUHITDA KIBERXAVFSIZLIK TAHDIDLARI VA ULARNING OLDINI OLISH YO‘LLARI	288
55.	**Тоиров Ш. А.** ОПТИМИЗАЦИЯ ФУНКЦИЙ С ПРИМЕНЕНИЕМ ЭВОЛЮЦИОННЫХ АЛГОРИТМОВ	291
56.	**Toirov Sh. A.** ARCHITECTURE OF OPERATORS IN QUANTUM ALGORITHMS	397
57.	**Olimov S.A.** TALABALARNING O‘ZLASHTIRISH NATIJALARINI BAHOLASHDA RAQAMLI TEXNOLOGIYALARNING O‘RNI	302
58.	**Olimov S.A.** OLIY TA'LIM MUASSASALARIDA O‘QUV JARAYONINI BOSHQARISHDA RAQAMLI TEXNOLOGIYALARNING AHAMIYATI	309
59.	**Uzakova G.Z.** ENGAGING ESP LEARNERS: STRATEGIES FOR PRODUCTIVE LESSONS	316
60.	**Shafkarov F.X.** SANATORIY-SOG‘LOMLASHTIRISH MUASSASALARINING ICHKI AUDITINI FUNDAMENTAL ASOSDA TASHKIL ETISH	321
61.	**Raximov Sh.F.** TIJORAT BANKLARIDA MIJOZLARNING SODIQLIK DASTURLARI ASOSIDA DEPOZITLARNI OSHIRISH YO’LLARI	329
62.	**Туракулова Ш.А,. Суюнова Н.** ОПИСАНИЕ КРАЙНЫХ ГИББСОВСКИХ МЕР ДЛЯ МОДЕЛИ ИЗИНГА С НЕНУЛЕВЫМ ВНЕШНИМ МАГНИТНЫМ ПОЛЕМ НА РЕШЕТКЕ БЕТЕ	337
63.	**Ulashov A.E.** ZAMONAVIY TA’LIM TIZIMINING RAQAMLI TRANSFORMATSIYASI: TENDENSIYALAR, MUAMMOLAR VA YECHIMLARI	344
64.	**Radjabov B.A.** O‘ZBEKISTONDA TA’LIMNING RAQAMLI TRANSFORMATSIYASI VA SUN’IY INTELLEKT	353
65.	**Mamatov N.S., Ibroximov S.R., Almuradova N.A.** O’QUV JARAYONIDA SUN’IY INTELLEKT VOSITALARIDAN FOYDALANISH	359
66.	**Kuvandikov J.T., Tojiyev A.H.** Axborot resurslaridan foydalanish ko‘rsatkichlarini normallashtirish asosida bilim oluvchilarning bilim darajasini obyektiv baholash usullari	364
67.	**Saidov S.F.** TA'LIMDA SUN'IY INTELLEKTDAN SAMARALI FOYDALANISH MASALALARI	374
68.	**Suyunova N.A.** TA’LIMDA MANTIQ TUSHUNCHASI VA O‘QUVCHILARDA MANTIQIY FIKRLASHNI RIVOJLANTIRISHNING INNOVATSION TEXNOLOGIYALARI	382

69.	**Kuvandikov J.T. Tojiyev A.H.** TA'LIM OLUVCHILAR BILIMINI BAHOLASHDA NORAVSHAN MANTIQDAN FOYDALANISH	389
70.	**Djabbarova M.D.** KASB–HUNAR TA'LIMIDA ELEKTRON TA'LIMNING PEDAGOGIK VA PSIXOLOGIK ASOSLARI	393
71.	**Anarkulova G.M., Abduraxmanova F.N.** OLIY TA'LIMDA MAXSUS FANLAR BO'YICHA MASHG'ULOTLARDA ISHCHAN VA ROLLI O'YINLAR METODINI QO'LLASH INNOVATSION YONDASHUV	396
72.	**Анаркулова Г.М., Абдиганиева З.С.** ПРОБЛЕМЫ ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ МЕТОДОВ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ СТУДЕНТОВ КАК ФАКТОР УСПЕВАЕМОСТИ	406
73.	**Obidov R.A.** TA'LIMDA KEYS TEXNOLOGIYALARINING AHAMIYATI	417
74.	**Narbayev Sh.K., Maxkamova N.X.** MAXSUS FANLARNI O'QITISHDA RAQAMLI VA INNOVATSION TEXNOLOGIYALARIDAN FOYDALANISHNING METODIK SHART-SHAROITLARI	422
75.	**Erhanova N.A.** MAKTABGACHA TA'LIM TASHKILOTLARIDA MULTIMEDIA RESURSLARIDAN FOYDALANISH SAMARADORLIGI	430
76.	**Xidoyatov M.B.** OLIY TA'LIM MUASSASALARIDA RAQAMLI MARKETING FAOLIYATINI BAHOLASH: AXBOROT TEXNOLOGIYALARIGA ASOSLANGAN YONDASHUVLAR	435
77.	**Темиров А.А.** АКЦИЯДОРЛИК ЖАМИЯТЛАРИ ФАОЛИЯТИГА РАҚАМЛИ БОШҚАРУВНИ ЖОРИЙ ЭТИШ ИСТИҚБОЛЛАРИ	444
78.	**В.А.Ахмадалиев** ЧОРВАЧИЛИК ХЎЖАЛИКЛАРИНИ ТАШКИЛ ЭТИШ ВА УЛАРГА ЕР АЖРАТИШ ТАРТИБИ	454
79.	**Носирова Н.Дж.** ЭКСПОРТНЫЙ ПОТЕНЦИАЛ СУБЪЕКТОВ МАЛОГО БИЗНЕСА УЗБЕКИСТАНА: СОВРЕМЕННОЕ СОСТОЯНИЕ И ПЕРСПЕКТИВЫ (2018–2025 ГГ)	459
80.	**Матлюбов Х.Ж.** ТРАНСФОРМАЦИЯ ПРОМЫШЛЕННОЙ ПОЛИТИКИ: ОСНОВНЫЕ ТЕНДЕНЦИИ И ВЫЗОВЫ	463
81.	**Sa'dulloyev H. I.** SANOAT KORXONALARINING EKO-MARKETING TIZIMINI BAHOLAH AMALIYOTI	469

USING ARTIFICIAL INTELLIGENCE TO AUTOMATE THE PROCESS OF COLLECTING AND ANALYZING DATA FROM ONLINE JOB POSTINGS

Doshanova M.Y.,
Associate Professor of the Department of SOIT, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,
Otaxanova B.I.,
Associate Professor of the Department of SOIT, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Aliyev R.R.
Student, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi

Annotatsiya. *Maqolada mavzu sohasining katta matnli korpusida o‘qituvchisiz o‘qitilgan neyron tarmoq tili modellaridan foydalangan holda jumla vektorlari va bilimlar bazasi obyektlarining semantik yaqinligini aniqlashga asoslangan onlayn ta’limdan foydalangan holda ma’lumot olish yondashuvi muhokama qilinadi. Matn korpusini belgilashning ko‘p mehnat talab qiladigan protseduralarisiz va qoidalarga asoslangan yondashuvlardan foydalanmasdan, joriy mehnat bozori talablarini tahlil qilish muammosini hal qilishda maqbul sifatga erishish imkonini beruvchi zamonaviy nazorat qilinadigan va nazoratsiz axborot olish usullari batafsil ko‘rib chiqilgan.*
Kalit so‘zlar: *tasniflash usuli, mashinani o‘qitish, neyron tarmoq til modellari, tabiiy tilni qayta ishlash, ma’lumot olish, obyektni tanib olish.*

Abstract. *The article discusses an approach to information extraction using online learning based on determining the semantic proximity of sentence vectors and knowledge base entities using neural network language models trained without a teacher on a large text corpus of the subject area. A detailed review of modern supervised and unsupervised information extraction methods is provided, which allow achieving acceptable quality in solving the problem of analyzing current labor market requirements without the labor-intensive procedure of text corpus tagging and without using rule-based approaches.*
Keywords: *classification method, machine learning, neural network language models, natural language processing, information extraction, entity recognition.*

Аннотация. *Предлагаемый в статье подход к извлечению информации основан на онлайн-обучении и использует меру семантической близости между векторами предложений и сущностями базы знаний. Для вычисления этой близости применяются нейросетевые языковые модели, обученные без учителя на обширном корпусе текстов, относящихся к предметной области. Приведен подробный обзор современных контролируемых и неконтролируемых методов извлечения информации, позволяющих добиться приемлемого качества при решении задачи анализа текущих требований рынка труда без трудоемкой процедуры разметки текстового корпуса и без использования подходов на основе правил.*
Ключевые слова: *метод классификации, машинное обучение, нейросетевые языковые модели, обработка естественного языка, извлечение информации, распознавание сущностей.*

Introduction

Today, the problem of extracting information from natural language texts is considered a relevant task. Traditional methods for this area of tasks are rule-based approaches, as well as supervised machine learning methods on text corpora tagged by experts. These approaches show the best quality in recognizing named entities and extracting facts. However, the possibilities of high-quality solutions to information extraction problems in individual subject areas using neural network models such as word2vec, fasttext, paragraph2vec, as well as other variants of distributed word representations, for example, GloVe, trained without a teacher on large text corpora, have not been fully explored. Although these approaches to text vectorization show better results in the problems of determining semantic proximity and resolving lexical ambiguity [26, 27].

The aim of the study is to experimentally evaluate the approach to determining the type of entity described by a fragment of text in natural language using neural network models trained without a teacher on large text corpora. The method is based on the hypothesis that the semantics of the entity types defined in the knowledge base corresponds to the semantics of the entities used in job descriptions by employers in online systems (for example, Ish-bor.uz, Ish.mehnat.uz), but it is necessary to take into account the difference between the vocabulary of professional standards and vacancies. The possibilities of using this approach are demonstrated using the example of the task of extracting entities from the texts of online recruitment system vacancies by comparing them by semantic proximity with the texts of professional standards to identify current labor market requirements in the IT industry. The novelty of the study lies in the application of approaches based on the use of neural network language models in the task of extracting entities from text, which does not require labor-intensive manual corpus tagging for training classifiers or writing a complex system of rules (without using rule-based approaches).

Overview of Information Extraction Methods

The term "information extraction" covers a wide range of natural language processing tasks. There are two main components of the information extraction process:

1) recognition of named entities;

2) extraction of relationships between entities.

Recognition of entities (usually named ones, such as persons, organizations, and locations) involves identifying individual phrases and sentences and defining them as mentions of an element of a particular type.

There are three approaches to entity recognition:

- gazetteers - primitive collections of various mentions of certain entities, acting as dictionaries;

- rule-based - rule-based systems that are broader in scope, but limited

by the number of instructions and templates given to them;

- machine learning algorithms - more complex models that can identify entities more flexibly, gradually memorizing new attributes of such elements during the learning process.

Today, the basic solution to the entity recognition problem is a combination of gazetteers, basic rules, and Conditional random field (CRF). CRF is one of the classic machine learning algorithms. Such a set of algorithms was used, for example, as a baseline in informal or slang text [1]. Most researchers used CRF, as well as classic feedforward neural networks (FFNN) and Markov algorithms. In addition to the texts themselves, many researchers also used word embedding values using the word2vec and GloVe algorithms. Unfortunately, compared to the baseline of 31%, researchers only managed to achieve a result of 57.6%, which indicates that there are certain difficulties in the entity extraction process today.

In [2], the authors Z. Zhang and J. Iria propose an algorithm for automatically constructing gazetteers based on Wikipedia and WordNet by identifying the entity type by moving up the hypernym hierarchy. The method shows rather weak results for such types of named entities as person and organization, but works better for geographic locations. In addition, it is limited by the data available in the mentioned systems. As far as we know, this method has not been developed. Rule-based systems are currently considered to be quite primitive, suitable only for automating the processes of extracting well-structured information. The main disadvantage of rule-based systems is their limitation - each new section of knowledge requires the development of its own set of rules that can take into account the specifics of the texts in this area, which requires the involvement of a large number of human resources. At the same time, the quality of more automated systems based on machine learning algorithms has increased enough to compete with the best rule-based systems. In the work [3], the authors S.A. Zahraa, C. Mark and H. Gholamreza declared the possibility of developing a rule-based system that can compare in quality with machine learning algorithms, if you first spend 8 man-weeks on developing rules for a specific subject area.

When talking about machine learning algorithms, it is worth distinguishing between supervised algorithms, which are trained on a sufficient volume of manually labeled examples, and unsupervised algorithms, which learn to recognize entities using only the information provided in the processed data and some previously known heuristics. Supervised algorithms have a drawback similar to rule-based systems: their training requires a fairly labor-intensive process of preparing training data. Among supervised machine learning algorithms, most classical methods reduce the task of entity recognition to the labeling of sequences and their subsequent element-by-element classification.

More specific examples include the above-mentioned CRF. CRF is one of the most popular models for searching for named entities, defining tags based on attributes, but taking into account both the current word and previous and subsequent words in the text. Thus, this algorithm underlies a number of popular sequence taggers. There are also sequence labeling algorithms based on maximum entropy [4], which predicts the label of a sequence element based on the probabilities of occurrence of certain attributes of a word and its predecessors, and Markov models [5], which perceive text labeling as a Markov process, where states are the desired classes, and the probabilities of the labels of the current element are determined by the previous state of the process [6]. More complex sequence classification algorithms can rely on complex neural network models, such as LSTM, which has become popular in working with text data due to its ability to take into account the history of sequences passed through it. Examples of the use of such models can be found in [7], where the use of bidirectional LSTM networks allows one to simultaneously take into account the attributes of both previous and subsequent words in a sentence when assigning an entity tag, and in [8], where the authors compare the performance of unidirectional and bidirectional LSTM using CRF at the network output to take into account the obtained tags of neighboring words and improve quality. Unlike supervised algorithms, unsupervised algorithms often perform entity detection in text based on searching for similar words in a document, in an attempt to identify named entities into common groups based on context. An example of this approach is [9], in which the authors use Word2vec to generate clusters of words with similar contexts. This approach shows better results compared to the classical CRF for languages with a low volume of labeled corpora (for example, Bengali). Another example is the use of Brown clusters - organizing words in a document into hierarchical clusters based on their distribution probabilities. However, supervised and unsupervised algorithms are often used together. For example, in [10], the author suggests using the above-mentioned word2vec model, whose word vectors are used as one of the attributes in the process of classifying named entities. The authors of [11] compare the quality of various options for using word2vec and Brown clusters as attributes for a CRF classifier for finding entities in medical texts.

Relationship Extraction

Extracting relationships between entities found in the text is the next logical step in extracting structured information from simple unstructured text. There are several approaches to solving this problem in the literature. One of these approaches is the approach based on the classification of possible candidate pairs of entities. Thus, in the work [7], the authors present their algorithm for classifying binary relations based on many different groups of attributes - statistical (frequency), linguistic, based

on existing knowledge - using SVM as a classifier of relationship types. Another example is the algorithm based on a convolutional neural network in the work [5]. Here, CNN is used to combine the attributes of individual words and obtain the attributes of the entire sentence, which are then used together to train a softmax classifier of relationship types. Another group of approaches is based on the use of kernels for the automated detection of patterns of certain types of relations. For example, the authors of [12] consider the use of a tree-kernel for constructing syntactic structures and determining the proximity between them using the presence of common subtrees in them. In another example [13], the authors propose a polynomial kernel-based method for automatically finding "interaction" words involved in relation patterns. A third group of approaches treats the relation extraction task as a generation task. More specifically, using the original unstructured text, such approaches generate a structured representation of the information contained in the text. Such generation usually falls under the term sequence2sequence or seq2seq - "sequence to sequence". A variation of seq2seq based on bidirectional recurrent networks (BiRNN) is used to extract relation triplets from unstructured text. This neural network uses confidence vectors to determine whether a particular triplet belongs to a particular relation type, as well as whether the entities are mentioned in the input sequence. This method is limited by the need to specify the exact number of relation types in advance, which reduces its flexibility when changing the domain. The authors of [14] consider the possibility of using seq2seq for automatic fact extraction from Wikipedia article texts in the format of its infobox elements. In this case, the authors use a separate generator for each element type. In [15], the authors propose using a CNN-LSTM encoder-decoder to transform an input sentence into a set of all relations present in it, in descending order of information content. The main drawback of this approach is the need for a sufficiently large training corpus, necessary for high-quality training of seq2seq. In addition, this corpus will be larger than the corpus for training a conventional classifier.

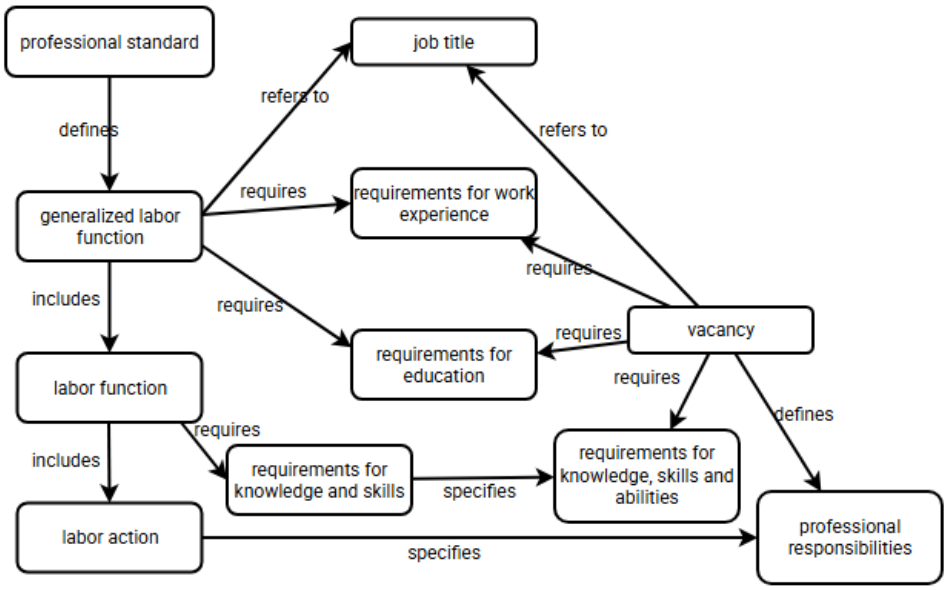
Method for Determining the Type of Entities

The study proposes to set the task of determining the type of entity for a fragment of the text of a vacancy by correlating it with elements of professional standards.

Conceptual Model

Figure 1 illustrates the conceptual model of the subject area describing the correspondence between elements of the description of vacancies and elements of professional standards.

Figure 1.
Conceptual model of entity relationships between professional standards and vacancies



This work proposes to define the following types of entities common in vacancy descriptions:

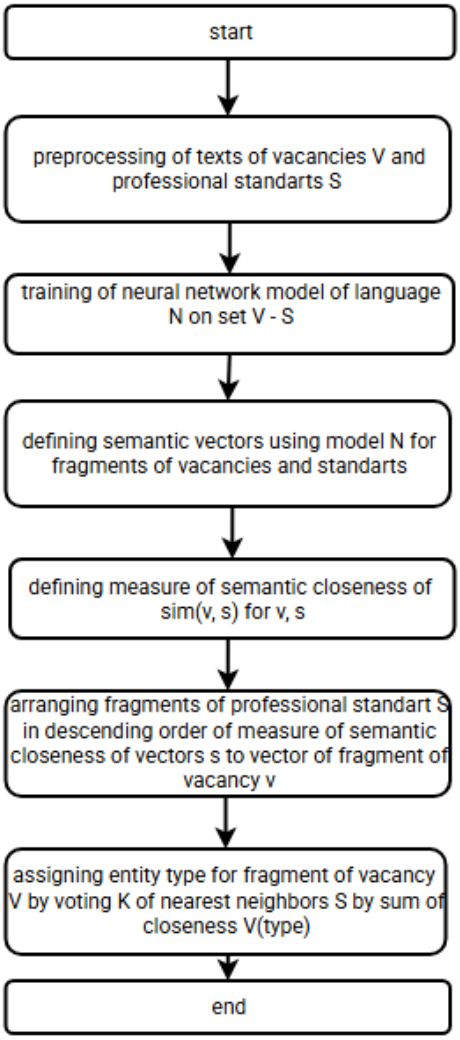
- labor actions (responsibilities);
- education requirements;
- knowledge/skills requirements;
- work experience requirements.

Algorithm for determining the type of entity

Figure 2 shows a general scheme of the algorithm for determining the type of entity based on determining the closest semantic elements of the texts of standards for each of the vacancy elements. In this case, the semantic proximity of the texts is determined based on the proximity of the vector representations of the texts generated by a neural network language model trained on a large corpus of vacancy texts and professional standards. Vector proximity is usually determined based on the cosine measure of the angle between the directions of two vectors: the closer the vectors are to each other, the smaller the angle between their directions, and therefore closer to 1 its cosine. For each element of the vacancy text, the approach is reduced to the following stages:

- determination of the closest elements of standards;
- voting of the classes of these “neighbors” to determine the class of the vacancy element.

Figure 2.
Algorithm for determining the entity type for a vacancy fragment based on voting of k-nearest neighbors from standard fragments



Thus, the main stage of the proposed approach is the vectorization of a text document. The experiment compares a number of different known vector representation algorithms: averaged word2vec, TF-IDF weighted averaged word2vec, SIF weighted averaged word2vec, paragraph2vec. These algorithms show high results in semantic proximity determination tasks, including for the Uzbek language, for example, in the framework of semantic proximity determination and lexical ambiguity resolution competitions.

Neural network language models

Word2Vec

The neural network approach to language modeling was proposed by a team of Google researchers led by T.Mikolov. This approach is presented in the form of two variations of a neural network architecture containing a single hidden layer. The final model, relying on the distribution hypothesis (language units with similar distributions have similar meanings), learns to match words and the contexts of their use. Training occurs without

the help of a teacher, using only unlabeled texts, producing a set of vectors of a given dimension for any word encountered during the training process. The resulting vectors reflect the proximity of these words: closer words have closer vectors and vice versa. The positive characteristics of this model are the low sparseness of the final vectors, the ability to specify their dimension, as well as the speed of operation (in comparison with more complex models that provide a similar level of quality). The main disadvantage is the inability to interpret the values of the coordinates of a vector. To obtain a vector representation of the entire text, it is necessary to combine the vector representations of individual words, which is usually done by taking the average value of the vectors.

Paragraph2Vec

Developing the idea of word2vec, T.Mikolov soon proposed a neural network model for vectorizing entire documents, called paragraph2vec or doc2vec. This model has an architecture similar to word2vec, with the only difference being that in addition to context words, the model also takes into account the context document, learning its vector representation during the training process. As a result, paragraph2vec is able to return vectors of entire texts that have a similar quality to the vectors of individual words in word2vec. At the same time, for previously unseen documents, a vector can be generated based on the words included in the document. Thus, using paragraph2vec, you can obtain vector representations of texts without any additional actions.

TF-IDF

Having become a classic algorithm in NLP, TF-IDF is an easy-to-understand and calculate scheme for weighting words in documents. TF-IDF is a combination of two simpler weights for a word: tf (term frequency) word frequency and idf (inverse document frequency). TF is the simplest frequency characteristic of a word in a corpus of documents, reflecting the frequency of its use in documents of a given set. The TF heuristic is based on the assumption that the more frequently a word is used, the more important it is. IDF is a slightly more complex frequency characteristic that shows how significant a word is for distinguishing between texts in the analyzed corpus. This weight measure attempts to correct the shortcoming of TF, due to which the weight of frequently used but unimportant function words increases. To this end, IDF is inversely proportional to the number of documents in which a word occurs, giving greater weight to words that occur only in individual documents, assuming that these words best describe such a document. Vector representations of words when using TF-IDF are one-hot vectors containing only one value different from 0, equal to the TF-IDF weight of this word. The dimension of such a vector is equal to the number of unique words in a particular corpus. In this case, to obtain vector representations of the entire document, a similar vector is created in which TF-IDF weights of all words

occurring in it are substituted instead of 0. The main disadvantage of such vectors is their extreme sparseness: a collection of documents may use all the words of a language (over 1 million), while each document may use only a small part of them (usually around 5–10 thousand). The advantages of TF-IDF are the simplicity of its calculation and absolute transparency in interpreting the values of the vectors. However, TF-IDF weights can be used as modifiers for other vector representations. In this case, the TF-IDF weight of a word is a lexical filter that determines the influence of this word on the final text vector, decreasing or increasing its contribution based on the "importance" of this word.

In our experiments, we use TF-IDF weighting to improve the quality of the averaged word2vec.

Smoothed Inverse Frequency (SIF)

Another form of text vectorization based on weighting word vector representations was proposed by scientists at Princeton University. SIF (smoothed inverse frequency) is a vectorization algorithm consisting of two stages.

The first stage is calculating the weights of words. Initially, the generative model determines the probabilities of generating a word at the current time, taking into account the current context of this word. Then the weights of each word are determined as:

$$weight = a/(a + p(w)), \quad (1)$$

where a is a parameter and $p(w)$is the probability of generating a word w.

The resulting weights are used to weight the original word vectors. The second stage is removing common components. This action is similar in motivation to IDF: frequently used words and word pairs receive large vectors, which cause anomalies in the document vectors obtained by averaging word vectors. Namely, the contribution of these word vectors causes an increase in the projection of the document vector onto directions that do not make any sense. To combat this influence, the authors propose removing the projection of the document vector onto the first principal component.

As a result, the model generates weighted average vectors for documents, which are quite simple to calculate, but more effective than many modern baselines.

Experiment

Next, we consider individual aspects of the experiment with an assessment of the quality of the implemented approach on a representative text corpus.

Characteristics of text corpora

A large corpus of texts of vacancies in the field of information technology from the platforms of online systems Ish-bor.uz and Ish.mehnat.uz for the last 3 years was prepared for training neural network models. Two corpora were prepared to assess the quality of the models:

1. Corpus of professional standards for the professions:

"programmer", "database administrator", "system administrator of information and communication systems", "specialist in testing in the field of information technology".

2. Corpus of fragments of 102 vacancies of the corresponding professions, in which the following types of entities were marked by 4 experts (each fragment describes only one entity):

- labor actions (responsibilities): 576 examples;
- education requirements: 40 examples;
- knowledge/skills requirements: 545 examples;
- work experience requirements: 53 examples.

Detailed characteristics of the text corpora are presented in Table 1.

Training parameters of neural network models

The implementation from the gensim library [16] was used to train neural network language models. Table 2 presents the training parameters of these models.

Table 1

Characteristics of text corpora used in the experiments

Corpus	**Number of documents**	**Number of fragments (sentences)**	**Number of tokens**	**Number of unique tokens (dictionary)**
Большой корпус вакансий	461 тыс.	618 тыс.	113 млн.	200 тыс.
Корпус проф. стандартов	4 стандарта	502	9 тыс.	1,1 тыс.
Тестовый корпус вакансий	102	648	7,2 тыс.	1,6 тыс.

Table 2.

Parameters for training neural network language models

Model	**Architecture**	**Dimension**	**Min. word frequency**	**Epochs**
Paragraph2vec	PV-DM	200	3	5
Paragraph2vec	PV-DBOW	200	3	5
Word2vec	skip-gram	300	3	5
Word2vec	CBOW	300	3	5

Text preprocessing

Before training the models, the source texts are processed according to the following principles:

1) multi-line texts are combined into one line;

2) texts are cleared of all characters that are not letters, numbers, spaces or some special characters;

3) each token is subjected to morphological analysis and brought to normal form (if possible);

4) for normalized tokens, a part of speech mark is added;

5) service parts of speech (conjunctions, prepositions and pronouns) are removed.

Experiment results

As follows from the results presented in Table 3, the application of various weighting modifications to the average word2vec vector did not lead to an increase in the quality of the solution to the entity type determination problem. The table also contains the values of k – the number of nearest neighbors, which yielded the best results for each of the models.

Table 3.

Results of comparing various neural network language models in determining the entity type by k nearest neighbors

Модель	**Precision**	**Recall**	**F1 (micro)**	**k (число ближайших соседей)**
Doc2vec (DBOW)	0.57	0.54	0.52	13
Doc2vec (DM)	0.48	0.50	0.48	13
Avr. Word2Vec (skip-gram)	0.70	0.68	0.69	12
Avr. Word2Vec (CBOW)	0.74	0.73	0.73	13
TFIDF+Word2Vec	0.65	0.63	0.63	13
SIF+Word2Vec	0.61	0.59	0.58	9

This is explained by the already mentioned difference in vocabulary between the two corpora: the professional standards corpus and the vacancy corpus. Classical weighting schemes (TF-IDF, SIF) were unable to adapt and qualitatively calculate inverse frequencies for terms from the dictionary when mapping vacancy elements to the space of professional standards elements. paragraph2vec showed itself to be significantly worse, which can be explained by the low quality of application of trained models of the PV-DBOW and PV-DM architectures to short texts of vacancy fragments (10–15 words on average), as well as the small number of examples of standard texts compared to the number of vacancy texts for high-quality training of document contexts.

Table 4.

Results of entity type recognition for each of the classes for the best Avr. Word2Vec (CBOW) model

Entity type	**Precision**	**Recall**	**F1 (micro)**
Labor action (duty)	0.79	0.71	0.75
Education requirements	0.98	0.89	0.93
Knowledge/skill requirements	0.69	0.72	0.71
Work experience requirements	0.51	0.88	0.64
Total (micro)	**0.74**	**0.73**	**0.73**
Total (macro)	**0.74**	**0.80**	**0.76**

However, it should be noted that even despite the complete absence of vacancy marking in the training data, mapping documents of one type to the space of documents of another type based on the definition of semantic similarity using neural network language models by online learning using the nearest neighbor method already gives the quality of entity type definition of 0.73 by the F-measure.

Conclusion

In this paper, we proposed and experimentally investigated a method for determining the entity type when extracting information from texts by calculating the semantic similarity of vectors obtained using neural network language models and determining the nearest neighbor entities from an automatically constructed knowledge base. The applicability of the method for determining four entity types from vacancy texts is demonstrated on a representative text corpus of vacancies and professional standards.

During the experiment, the best neural network model was determined - the averaged word2vec, trained using the CBOW algorithm, which shows the quality of micro-F1: 0.73 and macro-F1: 0.76 when determining four entity types. The main share of errors is related to the specifics of the wording of duties and requirements in vacancies, when employers mix the description of work actions with requirements for practical experience in applying work actions (skills).

The method has advantages in the low labor intensity of preparing a text corpus in comparison with traditional methods of learning with a teacher and methods based on rules. Also, the experiment showed the advantage of using word2vec model vectors without TF-IDF or SIF weighting schemes in conditions of limited vocabulary of texts from a knowledge base automatically generated from professional standards.

List of used literatures:

1. Al-Nabki, W., Eduardo, F., Enrique, A., & Laura F.-R. (2020). Improving named entity recognition in noisy user-generated text with local distance neighbor feature, Neurocomputing, Volume 382, 1-11.

2. Z., Zhang & J., Iria. (2019). A Novel Approach to Automatic Gazetteer Generation using Wikipedia, Proceedings of the 2009 Workshop on the People's Web Meets NLP, ACL-IJCNLP, 1-9.

3. Zahraa, S. A., Mark, C., & Gholamreza H. (2017). Multi-domain evaluation framework for named entity recognition tools, Computer Speech & Language, Volume 43, 34-55.

4. Rekia, K., Yu, Z., Weinan Zh., & Ting L. (2017). CCG supertagging via Bidirectional LSTM-CRF neural architecture, Neurocomputing, Volume 283, 31-37.

5. Wang, Y., Tong, H., Zhu, Z., & Li Y. (2022). Nested Named Entity Recognition: A Survey. ACM Transactions on Knowledge Discovery from Data, Online publication. 1-29.

6. Hu, X., & etw. (2023). Location Reference Recognition from Texts: A Survey and ComparisonACM Computing Surveys, Online publication. 1-37

7. Shao, Y., Hardmeier, C., & Nivre, J. (2016). Multilingual named entity recognition using hybrid neural networks. In The sixth Swedish language technology conference (SLTC).

8. Zhiheng, H., Wei, X., & Kai Yu. (2015). Bidirectional LSTM-CRF models for sequence tagging. CoRR, abs/1508.01991.

9. Luan, S., & Anderson F. (2021). An Entity Resolution Approach Based on Word Embeddings and Knowledge Bases for Microblog Texts. Association for Computing Machinery, Article 53, 1–8.

10. Kumarjeet, P., Pramit, M., & Vaishali, G. (2020). Named Entity Recognition using Word2vec, International Research Journal of Engineering and Technology (IRJET). Volume: 07, Issue: 09. 1818-1820.

11. Debora, N., Pikakshi, M., Elisabetta, F., Matteo, P., & Enza M. (2021). LearningToAdapt with word embeddings: Domain adaptation of Named Entity Recognition systems, Information Processing & Management, Volume 58, Issue 3.

12. Thien, H. N., Barbara, P., & Ralph, G. (2015). Semantic Representations for Domain Adaptation: A Case Study on the Tree Kernel-based Method for Relation Extraction, Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 635–644.

13. Wenhao, G., Xiao Y., Minhao Y., Kun H., Wenying P., & Zexuan Z. (2022). MarkerGenie: an NLP-enabled text-mining system for biomedical entity relation extraction, Bioinform Adv. 2(1):vbac035.

14. Wang, Zc., Wang, Zg., Li, Jz. et al. (2012). Knowledge extraction from Chinese wiki encyclopedias. J. Zhejiang Univ. - Sci. C 13, 268–280.

15. Nayak, Tapas & Ng, Hwee. (2019). Effective Modeling of Encoder-Decoder Architecture for Joint Entity and Relation Extraction. 10.48550/arXiv.1911.09886.

16. Library genism. https://pypi.org/project/gensim/ . Date of application [10.02.2025].

Ingliz tili muharriri: *Nosirova Nargiza Jamoliddin qizi*
Musahhih: *Vahobova Marfua Mirkamalovna*
Sahifalovchi va dizayner: *Zoirov Sardor Ziyodin o'g'li*

1-maxsus son

Mazkur jurnalda maqolalar chop etish uchun quyidagi havolalarga maqola, reklama, hikoya va boshqa ijodiy materiallar yuborishingiz mumkin.
Materiallar va reklamalar pullik asosda chop etiladi.
El. pochta: journals@timeedu.uz
Tel.: +998 95 651 90 00

“Management and Economics Scientific Research Journal” jurnali 19.09.2024-yildan O‘zbekiston Respublikasi Prezidenti Adminstratsiyasi huzuridagi Axborot va ommaviy kommunikatsiyalar agentligi tomonidan 404368-son reyestr raqami tartibi bo‘yicha ro‘yxatdan o‘tkazilgan. Litsenziya raqami: C-5669639
PNFL: 308906510

Manzilimiz: *Toshkent shahar, Yakkasaroy tumani*
Shota Rustaveli ko'chasi, 114