

выполнения конкретных разработок и, как следствие, расширения направлений проектирования обучения робототехнических систем.

БЛАГОДАРНОСТЬ

Эта публикация является результатом реализации проекта Erasmus+ «Передовой центр для докторантов и молодых исследователей в области информатики» (ACeSYRI), регистрационный номер 610166-EPP-1-2019-1-SK-EPPKA2-CBHE-JP.

УДК 002.53
МРНТИ 20.23.21

Утегенов Н. Б., магистр информационных наук, <https://orcid.org/0000-0001-8598-8086>
НАО «Западно-Казахстанский аграрно-технический университет имени Жангир хана»,
г. Уральск, ул. Жангир хана 51, 090009, Казахстан, nurikutegenov7@gmail.com

Utegenov N. B., Master of Information Sciences, <https://orcid.org/0000-0001-8598-8086>
NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk,
st. Zhangir khan 51, 090009, Kazakhstan, nurikutegenov7@gmail.com

ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА В СОВРЕМЕННОЙ КОНЦЕПЦИИ INFORMATION SEARCH SYSTEM IN THE MODERN CONCEPT

АННОТАЦИЯ

Современная концепция информационно-поисковых систем является важным инструментом для отдельных лиц и организаций для эффективного получения необходимой информации. Статья начинается с обсуждения важности эффективного и действенного поиска информации в современную эпоху, когда ежедневно генерируются и распространяются огромные объемы данных. Также в этой статье обсуждаются ключевые функции и компоненты системы поиска информации, включая алгоритмы поиска, индексирование, ранжирование, методы поиска, обработка естественного языка, персонализированные результаты поиска. Кроме того, в статье рассматриваются проблемы информационной перегрузки и проблемы, связанные с обеспечением точности и релевантности результатов поиска. В статье исследуются фундаментальные принципы и компоненты информационно-поисковых систем в контексте современных информационных технологий. В целом, в этой статье представлен всесторонний обзор систем поиска информации в современном контексте и подчеркивается их значение в облегчении доступа к ценной информации.

ANNOTATION

The modern concept of information retrieval systems is an important tool for individuals and organizations to effectively obtain the necessary information. The article begins by discussing the importance of efficient and effective information retrieval in today's era when vast amounts of data are generated and distributed daily. This article also discusses the key functions and components of an information retrieval system, including search algorithms, indexing, ranking, search methods, natural language processing, personalized search results. In addition, the article deals with the problems of information overload and the problems associated with ensuring the accuracy and relevance of search results. The article explores the fundamental principles and components of information retrieval systems in the context of modern information technologies. Overall, this article provides a comprehensive overview of information retrieval systems in today's context and highlights their importance in facilitating access to valuable information.

Ключевые слова: информационная система, поиск, данные, метаданные, информация, каталог.

Key words: information system, search, data, metadata, information, catalog.

Введение. В начале XX века многие ученые были солидарны с тем, что в начале двадцать первого века мы будем наблюдать так называемый «большой информационный взрыв». На данный момент можно с полной уверенностью фиксировать этот феномен, заметив при этом что его сила и скорость растет изо дня в день. Движущей силой данного факта является не только огромные шаги в развитии IT, но и увеличение во много раз информации, которая поступает человеку ежедневно.

Одной из главных проблем, которая проявляет себя в сфере науки, оказывается проблема доступа к информации. Каждый научный процесс формирует большое количество данных, и работать с ними оказывается тяжелее, между тем гигабайты информации превращаются в терабайты. Величина данных в ближайшее время может превзойти возможности компьютеров их обрабатывать, в связи с этим нужны новые разработки и алгоритмы для обработки.

С одной стороны, может возникнуть ощущение, что сам прогресс IT будет причиной того, что работа с научной информацией станет намного качественнее, однако, это не так. Текущие компьютерные технологии обеспечивают мощный аппарат для обработки лишь данных, но не самой информацией. Данные являют собой информационную ценность только в том случае, если они составная часть той или иной формы реального мира и связаны с другими данными.

Современное понимание об окружающем мире строится из обобщения данных, которые мы получаем во время наблюдений. И поэтому создание информационной модели объекта, объединяющая разного рода данные, преобразая таким образом их в информацию - важная цель. Тут нужно создание иерархической системы для установления отношений между данными, характера этих самых отношений и проработка стандартов для описания форматов данных.

Для обработки тех данных, которые были организованы с использованием какого-либо стандарта, используются сочетание аппаратно-программных средств, которые именуются информационными системами. В современном мире внутри широкого разнообразия информационных систем особое распространение получили поисковые системы, которые позволяют выполнять поиск, обработку и хранение информации, при этом в данных системах, в большинстве случаев, предусматривается использования простых видов логического и эвристического вывода.

Заметим, что к понятию информационно-поисковой системы можно было бы отнести и такие сочетания средств программного и аппаратного обеспечения, в которых используются только самые простые поисковые функции наподобие поиска документа по его известному имени. Но еще в 1968 году, такие системы были исключены из-под понятия информационно-поисковых систем, ибо устройства и системы, которые имеют только функцию поиска документов по известным адресам их хранения, не могут называться информационно-поисковыми. Данный аспект аналогичен базовым положениям работы Ляпунова и Яблонского, в соответствии с которой особенной чертой управляющих систем оказывается их сложность, выражающаяся в многочисленном количестве элементов, сложной структуре связей, а также в крупном количестве информации и осуществлении сложных функций.

Главный вопрос, который относится к разработке теории информационных систем, - это то, что общей теории, которая связывающей воедино данные, информацию и знания, нет. Существует большое количество различных подходов к теории информации с философских или кибернетических позиций, в том числе и так называемую «техническую» теорию информации, которую можно назвать «теорией передачи и хранения данных». Вы сможете найти десятки определений того, что такое информация или знание, даже которые будут иметь противоположные друг к другу смыслы.

На данный момент эксперты делают вывод о неточности этой ситуации, однако какой-либо пересмотр проблемы данных и контента не рассматривается. Наглядно это показывает цитата аналитика Сьюзен Фельдман: «Сегодня весь поисковый рынок устремился в область конвергенции „содержания“ с данными, поскольку, в конечном счете, ценность имеет информация».

Принимая все это во внимание попытаемся определить общие принципы создания моделей информационных систем, которые будут работать как со структурированными данными, или информацией, так и со знаниями, и назовем характерные черты таких систем.

Информационные системы как предмет всестороннего исследования

Рассмотрим информационные системы при помощи методики общей теории систем. Заметим, что традиционное понятие системы как «множество объектов вместе с отношениями между объектами и между их атрибутами» построено на одних и тех же принципах, что и реляционная модель данных.

Базовые системные принципы состоят из таких понятий, как:

- структурность;
- множественность описания;
- целостность;
- взаимозависимость системы и среды;
- иерархичность.

Есть два главных пути исследования информационных систем: макроподход, при котором система воспринимается как «черный ящик» для анализа ее взаимосвязи с окружающей средой, и микроподход, при котором рассматривается лишь внутреннее построение системы. В контексте этих подходов определены 12 главных векторов изучения систем: «информационные потоки», «коды», «функции», «функционирование систем», «элементы», «связи между элементами», «алгоритмизация», «анализ», «синтез», «преобразования», «эволюция», «надежность систем». В то же время обозначено, что часть из вышеперечисленных векторов (главным образом, преобразования и эволюция управляющих систем) актуальны не для всех типов, в том числе лингвистические (информационные системы относятся именно к лингвистическим).

Легко заметить, что в контексте макроподхода рассматриваются задача взаимозависимости системы и окружающей его среды, тогда как совокупность проблем, которые изучаются в рамках микроподхода, на самом деле вызывают изучение целостности системы. Изучая различные стороны микроподхода, можно заметить, что для исследования компонентов системы необходимо изучение ее многоуровневости, для исследования связей - изучение структурности, исследование системы призывает к применению множества разного рода моделей ее описания.

Проанализируем как при разработке информационных систем используются базовые системные принципы.

Аспекты макроподхода (в основном, зависимости информационной системы и внешней среды) могут вызывать интерес как с точки зрения взаимодействия с пользователем - создание моделей информационных запросов и моделей представления информации, так и с точки зрения отображения системой перемен в окружающей среде, и с точки зрения объединения системы с другими информационными системами.

Актуализация информации - это ахиллесова пята почти всех информационных систем, созданных для работы на протяжении неопределенного периода времени. Причина проста - нехватка средств, в основном, для оплаты работы кадров, что следят за изменениями информации. Это влечет за собой то, что нужна максимальная автоматизация актуализации информации. Одним из главных требований к автоматизированной информационной системе оказывается присутствие одного или нескольких каталогов. Иначе, если множество стандартизированных организованных описаний различных документов отсутствует, автоматизированное прибавление в систему новых документов - весьма проблематичная задача.

Низкая функциональная совместимость – это тоже довольно важный недостаток, который отрицательно влияет на главные функции информационных систем. Например, на систему поиска документов по их свойствам. Выделяют два уровня совместимости: семантический и технический, притом, что в техническом отмечают иногда синтаксический уровень. Семантическая совместимость, состоящая в применении различных стандартов метаданных, чаще всего, сохраняется. Проблемы создаются на ступени технической совместимости, а конкретнее, взаимодействия моделей данных и их различных форматов. Самым естественным видом единообразия подачи данных является каталог.

Что относится к микроподходу, необходимо обратить внимание, что информационной системе нужно воздействовать не прямо с документами, а с поисковыми образами. Если

поисковых образов нет, и поиск документа может быть только по его адресу, то это не согласуется описанному выше понятию информационной системы.

Отметим, что в частных случаях наполнение первичного документа может быть лишь частью поискового образа. В то же время, поисковый образ документа также считается документом (описывающим первичный документ). Таким образом, высказываясь дальше об информационных системах, мы будем применять понятие «документ» в понимании «поисковый образ исходного документа».

Базовый элемент - документ, который является информационным источником, при этом обладающий уникальным идентификатором, а также определенной структурой и содержанием.

Документы в информационной системе описываются с помощью метаданных, притом структура этих метаданных в основе своей разделена на уровни. Общий вид имеют метаданные, которые определяют структуру документа, другими словами описывающие метаданные уровнями ниже, те метаданные, которые описывают содержание документа. Значения этих низкоуровневых метаданных (или же атрибутов) есть ничто иное как метаданные относительно документа. Из этого следует основная отличительное качество информационной системы: она взаимодействует не с данными, а только лишь с метаданными.

Теперь мы можем определить главные особенности разработки системных принципов для информационных систем.

Целостность системы выражается с учетом каждого объекта, свойства этого самого объекта, а также функций внутри целого и осуществляется с помощью применения общего набора метаданных $M = \bigcup M^i$. Таким образом всякий документ d_1 системы можно представить как $d_i = \langle m_i^{j,k} \rangle$, где $m_i^{j,k}$ - значения элементов метаданных M^j , k - количество значений (с учетом копий) определенного элемента метаданных в файле.

Многоуровневость информационной системы выражается в том, что она представляет собой разнообразные подсистемы, которые нужны для тех или иных задач. Файлы, которые выражаются благодаря одним и тем же элементам метаданных, есть множество $M_i \subseteq M$, образуют класс K_i . Если $M_1 \subset M$, $M_2 \subset M$, и $M_1 \subset M_2$, то класс K_2 является подклассом класса K_1 . Множество стандартизированных организованных файлов - описаний одного класса – это каталог. По сути каталог связывает поисковые образы первичных файлов.

Структурность системы создается лучшим выбором модели связей между файлами, который способствует правильно описывать различные стороны соответствующих взаимоотношений. Многофункциональной является, к примеру, модель направленных связей. В случае, когда файл d_i является значением элемента M^j метаданных документа d_i , это означает что между этими файлами есть связь типа $M^j < d_i, d_{i'}, m_{i,i'}^{l,k} >$, где $m_{i,i'}^{l,k}$ - атрибуты, которые являются значениями подходящих элементов метаданных. В результате, сформированные отношения по сути переносятся на уровень элементов, что описывают структуру файлов.

Множественность описания системы есть ничто иное, как большое число разного рода аспектов разработки системы. В особенности общий вид имеет описание модели информационной системы, что создается с помощью задания классов K_i , которые определяются некоторыми множествами элементов метаданных M_i , а также с помощью различных видов связей между классами $M^j < K_i, K_{i'} >$ с определением элементов метаданных $M_{i,i'}^j$, что описывают атрибуты связей. В итоге, для создания модели информационной системы применяются совокупность иерархической и реляционной моделей данных.

Определенное содержание информационной системы описывается наполнением ее каталогов, и притом не только каталогов, но и каталогах связей, ибо иерархическую модель данных возможно описать в реляционном виде.

Модели для поиска информации и знаний

Как было вышесказанно, главным качеством информационных систем оказывается функция обрабатывать сложные запросы. При этом выделяют три основные модели поиска:

- поиск по аналогии;
- атрибутивный поиск;
- контекстный поиск;

С учетом структуры поиска поисковые признаки легко разделить на три вида: текстовые; числовые; табличные.

Контекстный поиск подбирает файлы, у которых значения атрибутов текстового вида имеют указанный в качестве поискового запроса текстовый кусок. Иногда, если пользователь того захочет, рассматриваются возможные правки текстового фрагмента. В запрос к любому атрибуту возможно добавлять несколько текстовых частей, которые относятся логическими операциями конъюнкции, дизъюнкции и отрицания. Такими же операциями возможно объединять запросы с несколькими атрибутами.

Контекстный поиск - самая легкая модель поиска, осуществляемая даже при элементарной структуризации документов, но такой поиск не особо эффективен. Она будет эффективна тогда, когда в поисковом запросе юзер пишет уже известную точную цитату определенной длины. Иначе, если запрос очень короткий, информационная система засыплет множеством документов, и большинство из них будут не нужны пользователю. В противовес, для поиска длинного запроса очень часто просто не находится нужных документов, ибо авторы файлов могли описывать нужный юзеру предмет предложениями, которые не похожи на заданный в поиске.

При поиске документов по атрибутам информационный запрос является комплексом величин одного или нескольких атрибутов, подбираемых из списка, которые объединенных логическими операциями.

Поиск по атрибутам желателен тогда, когда пользователю нужен не определенный файл, а класс файлов, которые обладают определенным комплексом атрибутов.

Поиск по сходству - это когда в формате информационного запроса оказывается непустое множество документов. В итоге запроса выводятся файлы, и каждый в какой-то степени похож к некоторому из документов, входящих в это множество.

Численная характеристика предела единства формируется на основе множества файлов D в таком виде: $m: D \times D \rightarrow [0, 1]$, где функция m при абсолютном сходстве приобретает значение 1, а при различии - 0. Расчет меры единства происходит по формуле $m(d, D') = \sum a_i m_i(d, D')$, где i - номер атрибута, a - весовые коэффициенты, при этом $\sum a_i = 1$, $m(d, D')$ - мера сходства по атрибуту i , и рассчитывается как $m(d, D') = \max_j m_i(d_i, d_j)$, d_j - файлы из множества D' . Для атрибутов таблиц мера аналогии по атрибуту i рассчитывается так: если значение атрибута i этого файла относится определенной проекции множества D' , то предел близости = 1, в противном случае = 0. Если атрибут, приобретающий многочисленные значения $m_i = n_{01}/n_1$, где n_1 - суммарное число элементов, являющейся частью значения атрибута i , n_{01} - общее число элементов, которые относятся к множеству D' .

В итоге, для того чтобы информационная система могла быть эксплуатирована для обретения новой информации, юзер должен иметь вариант проверять, истина или ложь утверждение $R_s(d_{i_1}, \dots, d_{i_n})$ в отношении основ, что представляются файлами $d_{i_1}, \dots, d_{i_n}$. Формуле $R_s(d_{i_1}, \dots, d_{i_n})$ номинально подходит n -местный предикат R_s , что найден на множестве файлов, притом что при его формировании могут применяться некоторые атрибуты файлов системы, данные из предметной области и так далее. В том случае, когда число файлов в системе, что могут быть в роли аргументов предиката R_s , большое, то подходящим методом получения новых данных будет тест истинности предиката на разных комплексах файлов, просматриваемых системой. Заметим важное отличие этого подхода от подходов, что используются в сфере ИИ для формирования экспертных систем: они необходимы для решения узконаправленных задач и в них сравнительно малый объем файлов, а главный упор при их разработке делается на улучшении сложных продукционных правил. У нас законы вывода довольно просты, а извлечение новых знаний возможно за счет большому объему файлов, которые выступают в роли аргументов проверяемых запросов.

Итак, для результативного формирования поиска и получения новой информации нужна эффективно организованный каталог документов, в которой есть достаточное количество поисковых признаков, и эти признаки описывают поисковый образ документа. Это обязывает каталог информационно-поисковой системы соответствовать определенным требованиям, которые будут описаны далее.

Требования для формирования каталога информационно-поисковой системы

Каталог информационно-поисковой системы соединяет в себе поисковые образы документов. И здесь появляется вопрос: какой комплекс признаков необходимо определить с той целью, чтобы осуществлялось главное свойство информационно-поисковых систем - поисковые запросы гораздо сложнее, нежели простая выдача документа по определенному имени?

Докажем, что действительные возможности контекстного поиска на деле заключается в поиске документов с определенным именем или известной фразой из текста файла. Действительно, наиболее эффективным поиск будет, если в свой запрос юзер напишет предварительно известную точную цитату определенной длины. Иначе, при коротком запросе, система предложит большое количество файлов, которые не будут отвечать требованиям пользователя. Для длинных свободных запросов очень часто и не находится даже подходящих файлов, ибо большинство авторов файлов описывают нужный пользователю предмет различными атрибутами, что не похоже на написанные в запросе.

Та ситуация, когда информационная система позволяет только контекстный поиск, складывается, когда поисковый язык системы точно совпадает с естественным языком.

Кроме того, такие же проблемы появляются в том случае, когда в информационно-поисковом языке нет возможности проявления внутренних отношений между предметами. В итоге, вариант получения на выходе релевантных документов возникает тогда, когда в поисковом языке есть функция тезауруса. Поскольку у понятия тезауруса есть смысловое разделение терминов, то факт присутствия тезауруса в системе — значит вариант атрибутивного поиска и поиска по сходству. В роли поисковых характеристик возможно определить разделы определенного классификатора. Таким образом, при наличии тезауруса в составе поискового языка, главное требование к информационным системам соблюдается.

Заключение

В заключении, «информационная система» - это только те комплексы аппаратно-программных средств для работы с организованной информацией, что дают возможность информационного поиска файлов не только по их известным именам, но также и по атрибутам. Главное отличие такой системы заключается в том, что она взаимодействует не с данными, а только с метаданными, а также она в обязательном порядке укомплектована каталогом.

Для продуктивной работы информационного поиска необходима структуризация файлов высокого качества, что дает выбор типа информационной системы, которая допускает отметить достаточное количество атрибутов файла, которые выступают в роли поисковых признаков. Притом для образования сложных информационных запросов в системе должен присутствовать тезаурус.

СПИСОК ЛИТЕРАТУРЫ

- 1 Чжан, Дж., Чжао, В. Х., и Чжан, Ю. (2016). Опрос по глубокому обучению для больших данных. *Информация Fusion*, 42, 146-157.
- 2 Янсен, Б.Дж., и Спинк, А. (2017). Как мы ищем во всемирной паутине? Обзор. *Международный журнал интернет-науки*, 12(1), 1-12.
- 3 Сингхал, А. (2016). Представляем График знаний: вещи, а не строки. Официальный блог Google. Получено с <https://googleblog.blogspot.com/2012/05/introduction-knowledge-graph-things-not.html>.
- 4 Морли, Д., и Флориди, Л. (2019). Этика алгоритмов: ключевые вопросы и проблемы. *Философия и технология*, 32(4), 611-627.
- 5 Киган, Б. (2018). Машинное обучение и большие данные: каковы этические последствия? *Информация и компьютерная безопасность*, 26(2), 183-197.
- 6 Культау, СС (2016). Управляемое исследование: обучение в 21 веке. Библиотеки без ограничений.
- 7 Андерсон, К. (2018). Длинный хвост: почему будущее бизнеса заключается в том, чтобы продавать меньше или больше. *Хашетт Великобритания*.
- 8 Прабхакар, ТВ (2018). Машинное обучение и обработка естественного языка для анализа больших данных. В *Справочнике по исследованиям методов хранения и визуализации больших данных* (стр. 147–162). ИГИ Глобал.

- 9 Ли, Дж., Ли, Дж., и Ли, Ю. (2016). Всесторонний обзор методов семантического поиска: от традиционного до глубокого обучения и далее. Доступ IEEE, 4, 3844-3869.
- 10 Чжан, Х., и Хуанг, С. (2019). Проблемы и возможности больших данных в здравоохранении: систематический обзор. Искусственный интеллект в медицине, 97, 61-74.
- 11 Брин С. и Пейдж Л. (2018). Анатомия крупномасштабной гипертекстовой поисковой системы. Компьютерные сети, 56(18), 3825-3833.
- 12 Лю, Ю., Ли, Дж., и Ли, Ю. (2017). Аналитика больших данных с глубоким обучением: обзор. Границы компьютерных наук, 11 (1), 1-12.
- 13 Лу, Дж., Ву, Х., Ван, Ю., и Чжан, Х. (2017). Управление знаниями и большие данные: последствия для управления знаниями. Журнал управления знаниями, 21 (1), 20–43.
- 14 Бурель, Г., Ким, Дж. Х., и Ли, Х. (2019). Аналитика больших данных для здравоохранения. Журнал исследований в области информатики здравоохранения, 3 (4), 357–361.
- 15 Ли, Ю., Ли, Дж., и Ли, Дж. (2018). Большие данные: от традиционной аналитики до обработки естественного языка и глубокого обучения. Журнал больших данных, 5(1), 1-10.
- 16 Цзян Ю., Ли Дж. и Ли Ю. (2016). Аналитика больших данных с помощью R и Hadoop. Международный журнал хранилищ данных и интеллектуального анализа данных, 12 (3), 57-71.
- 17 Аггарвал, СС (2017). Нейронные сети и глубокое обучение: учебник. Спрингер.
- 18 Бенжио, Ю., Курвиль, А., и Винсент, П. (2015). Обучение представлению: обзор и новые перспективы. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798-1828.
- 19 Блей, Д. М. (2015). Вероятностные тематические модели. Сообщения ACM, 58(1), 77-84.
- 20 Бакли, К., и Мэннинг, К.Д. (2016). Повышение эффективности поиска информации с помощью анализа местного контекста. Транзакции ACM в информационных системах (TOIS), 34 (4), 1-33.

REFERENCES

- 1 Zhang, J., Zhao, W. X., & Zhang, Y. (2016). A survey on deep learning for big data. Information Fusion, 42, 146-157.
- 2 Jansen, B. J., & Spink, A. (2017). How are we searching the World Wide Web? A review. International Journal of Internet Science, 12(1), 1-12.
- 3 Singhal, A. (2016). Introducing the Knowledge Graph: things, not strings. Official Google Blog. Retrieved from <https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>
- 4 Morley, D., & Floridi, L. (2019). The ethics of algorithms: Key issues and challenges. Philosophy & Technology, 32(4), 611-627.
- 5 Keegan, B. (2018). Machine learning and big data: What are the ethical implications? Information and Computer Security, 26(2), 183-197.
- 6 Kuhlthau, C. C. (2016). Guided inquiry: Learning in the 21st century. Libraries Unlimited.
- 7 Anderson, C. (2018). The long tail: Why the future of business is selling less of more. Hachette UK.
- 8 Prabhakar, T. V. (2018). Machine learning and natural language processing for big data analytics. In Handbook of research on big data storage and visualization techniques (pp. 147-162). IGI Global.
- 9 Li, J., Li, J., & Li, Y. (2016). A comprehensive survey of semantic search techniques: From traditional to deep learning and beyond. IEEE Access, 4, 3844-3869.
- 10 Zhang, X., & Huang, S. (2019). Challenges and opportunities of big data in health care: a systematic review. Artificial Intelligence in Medicine, 97, 61-74.
- 11 Brin, S., & Page, L. (2018). The anatomy of a large-scale hypertextual Web search engine. Computer Networks, 56(18), 3825-3833.
- 12 Liu, Y., Li, J., & Li, Y. (2017). Big data analytics with deep learning: A review. Frontiers of Computer Science, 11(1), 1-12.
- 13 Lu, J., Wu, X., Wang, Y., & Zhang, X. (2017). Knowledge management and big data: Implications for knowledge management. Journal of Knowledge Management, 21(1), 20-43.

- 14 Burel, G., Kim, J. H., & Lee, H. (2019). Big data analytics for healthcare. *Journal of Healthcare Informatics Research*, 3(4), 357-361.
- 15 Li, Y., Li, J., & Li, J. (2018). Big data: from traditional analytics to natural language processing and deep learning. *Journal of Big Data*, 5(1), 1-10.
- 16 Jiang, Y., Li, J., & Li, Y. (2016). Big data analytics with R and Hadoop. *International Journal of Data Warehousing and Mining*, 12(3), 57-71.
- 17 Aggarwal, C. C. (2017). *Neural networks and deep learning: a textbook*. Springer.
- 18 Bengio, Y., Courville, A., & Vincent, P. (2015). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798-1828.
- 19 Blei, D. M. (2015). Probabilistic topic models. *Communications of the ACM*, 58(1), 77-84.
- 20 Buckley, C., & Manning, C. D. (2016). Improving the effectiveness of information retrieval with local context analysis. *ACM Transactions on Information Systems (TOIS)*, 34(4), 1-33.

БЛАГОДАРНОСТЬ

Эта публикация является результатом реализации проекта Erasmus+ «Передовой центр для докторантов и молодых исследователей в области информатики» (ACeSYRI), регистрационный номер 610166-EPP-1-2019-1-SK-EPPKA2-CBHE-JP.

ӨОЖ 004.056:378
FTAXP 81.93.29

Мұхамбетов Амангелді Абатұлы, техника ғылымдарының магистрі, <https://orcid.org/0000-0002-3022-4086>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Орал қ., Жәңгір хан көш., 51, 090009, Қазақстан, amangeldy.abatovish@gmail.com

Куангалиев Тимур Гайсиевич, техника ғылымдарының магистрі, <https://orcid.org/0000-0002-0849-5875>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Орал қ., Жәңгір хан көш., 51, 090009, Қазақстан, ttimurkuan@gmail.com

Аманкулова Гулфайрус Максатовна, техника ғылымдарының магистрі, <https://orcid.org/0000-0002-5566-5814>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Орал қ., Жәңгір хан көш., 51, 090009, Қазақстан, gulfayrus.amankulova@bk.ru

Гусманова Алида Сагингалиевна, техника ғылымдарының магистрі, <https://orcid.org/0000-0001-7730-051X>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Орал қ., Жәңгір хан к. 51, 090009, Қазақстан, alida_the_best@mail.ru

Mukhambetov Amangeldi Abatuly, Master of Technical Sciences, <https://orcid.org/0000-0002-3022-4086> NJSC «Zhangir khan West Kazakhstan agrarian technical university», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, amangeldy.abatovish@gmail.com

Kuangaliev Timur Gaisiyevich, Master of Technical Sciences, <https://orcid.org/0000-0002-0849-5875> NJSC «Zhangir khan West Kazakhstan agrarian technical university», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, ttimurkuan@gmail.com

Amankulova Gulfairus Maksatovna, Master of Technical Sciences, <https://orcid.org/0000-0002-5566-5814> NJSC «Zhangir khan West Kazakhstan agrarian technical university», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, gulfayrus.amankulova@bk.ru

Gusmanova Alida Saginalievna, Master of Technical Sciences, <https://orcid.org/0000-0001-7730-051X> NJSC «Zhangir khan West Kazakhstan agrarian technical university», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, alida_the_best@mail.ru