

ISSN 2708-2032
e-ISSN 2708-2040



**INTERNATIONAL
UNIVERSITY**

**INTERNATIONAL
JOURNAL OF INFORMATION
& COMMUNICATION TECHNOLOGIES**

**Volume 2, Issue 1
March 2021**

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN



**INTERNATIONAL JOURNAL OF
INFORMATION AND COMMUNICATION
TECHNOLOGIES**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ
ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

Том 2, Выпуск 1
Март 2021

Главный редактор – Ректор АО МУИТ,
профессор, д.т.н.
Ускенбаева Р.К.

Заместитель главного редактора –
проректор по НиМД, PhD, ассоц.профессор
Дайнеко Е

Отв. Секретарь – PhD, ассоц.профессор, директор департамента по науке
Кальпеева Ж.Б.

ЧЛЕНЫ РЕДКОЛЛЕГИИ:

Отельбаев М. д.т.н., профессор, АО «МУИТ», Рысбайулы Б., д.т.н., профессор, АО «МУИТ», Куандыков А.А., д.т.н., профессор, АО «МУИТ», Синчев Б.К., д.т.н., профессор, АО «МУИТ», Дузбаев Н.Т., PhD, проректор по ЦИИ, АО «МУИТ», Ыдырыс А., PhD, заведующая кафедрой «МКМ», АО «МУИТ», Касымова А.Б., PhD, заведующая кафедрой «ИС», АО «МУИТ», Шильдибеков Е.Ж., PhD, заведующий кафедрой «ЭиБ», АО «МУИТ», Ипалакова М.Т., к.т.н., ассоц. профессор, заведующая кафедрой «КИИБ», АО «МУИТ», Айтмагамбетов А.З., к.т.н., профессор, АО «МУИТ», Амиргалиева С.Н., д.т.н., профессор, АО «МУИТ», Ниязгулова А.А., к.ф.н., заведующая кафедрой «МиИК», АО «МУИТ», Молдагулова А.Н., к.т.н., ассоциированный профессор, АО «МУИТ», Джоламанова Б.Д., ассоциированный профессор, АО «МУИТ», Prof. Young Im Cho, PhD, Gachon University, South Korea, Prof. Michele Pagano, PhD, University of Pisa, Italy, Tadeusz Wallas, Ph.D., D.Litt., Adam Mickiewicz University in Poznań, Тихвинский В.О., д.э.н., профессор, МТУСИ, Россия, Масалович А., к.ф.-м.н., Президент Консорциума Инфорус, Россия, Lucio Tommaso De Paolis is the Research Director of the Augmented and Virtual Laboratory (AVR Lab) of the Department of Engineering for Innovation, University of Salento and the Responsible of the research group on “Advanced Virtual Reality Application in Medicine” of the DREAM, a multidisciplinary research laboratory of the Hospital of Lecce (Italy), Liz Bacon, Professor, Deputy Principal and Deputy Vice-Chancellor, Abertay University (Great Britain).

Издание зарегистрировано Министерством информации и общественного развития Республики Казахстан. Свидетельство о постановке на учет № KZ82VPY00020475 от 20.02.2020 г.

Журнал зарегистрирован в Международном центре по регистрации сериальных изданий ISSN (ЮНЕСКО, г. Париж, Франция)

Выходит 4 раза в год.

УЧРЕДИТЕЛЬ:

АО «Международный университет информационных технологий»

ISSN 2708-2032 (print)
ISSN 2708-2040 (online)

СОДЕРЖАНИЕ

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНЖЕНЕРИЯ ЗНАНИЙ

<i>Серікбай Е.Е., Таир Р.Т., Куандыков А.А, Нальгожина Н.Ж.</i>	
Разработка бизнес-процесса для оцифровки автомобильных дорог	10
<i>Жуманбаева С.К., Пащенко Г.Н.</i>	
Проектирование и разработка информационной системы для обработки научных трудов	18
<i>Айтбекова М.Б., Пащенко Г.Н.</i>	
Разработка информационной системы для расчетов рейтингов успеваемости студентов	24
<i>Алимжанова Л.М., Омарова А.Ш., Таштамышева А.Э.</i>	
Исследование актуальных проблем при переходе обучения в онлайн-формат.....	34
<i>Найзабаева Л., Аринова М. С.</i>	
Интеллектуальный анализ и прогнозирование токсичных элементов в почве	39
<i>Имангалиева А.А., Пащенко Г.Н.</i>	
Проектирование и разработка информационной системы для управления научно - образовательной деятельностью Университета	46

ИНФОКОММУНИКАЦИОННЫЕ СЕТИ И КИБЕРБЕЗОПАСНОСТЬ

<i>Базарбеков И.М., Шарипов Б.Ж.</i>	
Система smart кампус в университете: требования, преимущества и недостатки	53
<i>Кенескызы К., Ескермес С.Б.</i>	
Метод машинного обучения для обратных задач теплопроводности	59
<i>Алимжанова Л.М., Панарина А.В.</i>	
Ценность IT-аутсорсинга для клиента	65
<i>Мамен Е.К., Айтим А.К., Аднабеков А.Х., Абиев А.Б., Мустафина А.К.</i>	
Разработка умного холодильника без продавца	71
<i>Ауезова А.С., Муратова К.Н., Синчев Б.</i>	
Методы информационного поиска неструктурированных данных	79
<i>Надіров Н.Қ., Дүйсебекова К.С.</i>	
Разработка системы формирования профиля клиента на основе bigdata с использованием семантического анализа	85
<i>Бердыкулова Г.М.</i>	
Что такое научная статья и как не надо ее писать?	96

ЦИФРОВЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

Мукан Б., Саркамбаева Ш.Г.

Развитие проектной деятельности в малом и среднем бизнесе в Казахстане 104

Мейрманова Л.Е.

Эффективность применения управления проектами в управлении человеческими ресурсам 113

Алдибекова Н.Б., Тяп А.В., Омаров И.Г., Мохамед А.Х., Алимжанова Л.М.

Использование математического моделирования и программного обеспечения в управлении проектными рисками 119

Омарова А.Ш., Махбаева Ә.Н.

Проблемы и влияние коммуникаций в управлении проектами 129

Момбекова А., Бекболова М.Б.

Влияние аналитики больших данных на эффективность процессов управления проектами ..137

Булантаев А.М., Мусахан Х.Б., Молдагулова А.Н., Сембина Г.К.

Прогноз ожидаемых убытков банка при предоставлении кредита 145

Омарова Е.Г.

Алгоритм автоматизации классификации финансовых активов при розничном кредитовании 150

CONTENTS

SOFTWARE DEVELOPMENT AND KNOWLEDGE ENGINEERING

<i>Alin G.T., Rakhymzhanova N.K.</i> Software development project management: project risk management	9
<i>Zhumanbaeva S.K., Pachshenko G.N.</i> Desining and development of information system for the processing scientific works	19
<i>Aitbekova M.B., Pachshenko G.N.</i> Development of the information system for calculating students' performance rating.....	25
<i>Alimzhanova L.M., Omarova A.S., Tashtamysheva A.E.</i> Investigation of topical problems during the transition of learning in online format.....	35
<i>Naizabayeva L., Arinova M.S.</i> Intellectual analysis and prediction of toxic elements in the soil	40
<i>Imangalieva A., Pachshenko G.N.</i> Development of an information system for managing research and educational activi-ties of the university.....	47
<i>Serikbay Y.Y., Tair R.T., Kuandykov A.A., Nalgozhina N.Zh.</i> Development of a business process for digitizing roads.....	54

INFOCOMMUNICATION NETWORKS AND CYBERSECURITY

<i>Bazarbekov I.M., Sharipov B.Zh.</i> Smart campus system in the university: requirements, advantages and disadvantages.....	62
<i>Keneskyzy K., Yeskermes S.B.</i> Machine learning method for inverse heat conduction problems	68
<i>Alimzhanova L.M., Panarina A.V.</i> The value of IT outsourcing for the client	74
<i>Mamen Y.K., Aitim A.K., Adnabekov A.H., Abiyev A.B., Mustafina A.K.</i> Development of a smart refrigerator without a seller	80
<i>Auyezova A.S., Muratova K.N., Sinchev B.</i> Methods of information search for unstructured data	88
<i>Nadirov N. To., Duisenbekova K. S.</i> Development of a client profile formation system based on Big data using semantic analysis	94
<i>Berdykulova G.M.</i> What is an academic article or how not to write it?.....	105

DIGITAL TECHNOLOGIES IN ECONOMICS AND MANAGEMENT

<i>Mukan B., Sarkambaeva S.G.</i> Development of project activities in small and medium-sized businesses in Kazakhstan	113
<i>Meirmanova L.Ye.</i> Efficiency application project management in human resource management.....	122

<i>Aldibekova N.B., Tyan A.V., Omarov I.G., Mohamed A. Hamada., Alimzhanova L.M.</i>	
Using mathematical modelling and software programming in project risk management	128
<i>Omarova A., Makhbayeva A.</i>	
Challenges and impact of communication in project management	138
<i>Mombekova A., Bekbolova M.B.</i>	
The impact of big data analytics on the effectiveness of project management processes	146
<i>Bulantayev A.M., Musakhan K.B., Moldagulova A.N., Sembina G.K.</i>	
Forecasting expected bank losses at granting a loan.....	154
<i>Omarova Ye.G.</i>	
Algorithm for automating the classification of financial assets in retail lending	159

МАЗМҰНЫ

БАҒДАРЛАМАЛЫҚ ҚАМТАМА ҚҰРУ ЖӘНЕ БІЛІМ ТЕХНИКАСЫ

<i>Алин Г.Т., Рахимжанова Н.К.</i>	
Бағдарламалық даму жобасын басқару: жобаның тәуекелділігін басқару	9
<i>Жуманбаева С.К., Пащенко Г.Н.</i>	
Ғылыми еңбектерді өңдеуге арналған ақпараттық жүйені жобалау және зерттеу	19
<i>Айтбекова М.Б., Пащенко Г.Н.</i>	
Оқушылардың үлгерімін бағалауға арналған ақпараттық жүйе құру	25
<i>Алимжанова Л.М., Омарова А.Ш., Таштамышева А.Э.</i>	
Білім берудің онлайн форматқа көшуіндегі өзекті мәселелерді зерттеу	35
<i>Найзабаева Л., Аринова М. С.</i>	
Жер қыртысындағы улы элементтерді интеллектуалды талдау және болжау	40
<i>Иманғалиева А.А., Пащенко Г.Н.</i>	
Университеттің ғылым және білім беру қызметін басқару ақпараттық жүйесін жобалау және әзірлеу	47
<i>Серікбай Е.Е., Таур Р.Т., Қуандықов А.А., Нальгожина Н.Ж.</i>	
Жолдарды цифрландырудың бизнес процесін дамыту	54

ИНФОКОММУНИКАЦИЯЛЫҚ ЖЕЛІЛЕР ЖӘНЕ КИБЕРҚАУІПСІЗДІК

<i>Базарбеков И.М., Шарипов Б.Ж.</i>	
Университеттегі Smart кампус системасы: талаптар, артықшылықтары мен кемшіліктері ...	62
<i>Кенесқызы К., Ескермес С.Б.</i>	
Кері жылу өткізгіштік есептеріне арналған машиналық оқыту әдісі	68
<i>Алимжанова Л.М., Панарина А.В.</i>	
Клиент үшін ІТ аутсорсингтің мәні.....	74
<i>Мамен Е.К., Әйтім Ә.Қ., Аднабеков А.Х., Абиев А.Б., Мустафина А.Қ.</i>	
Сатушысыз ақылды тоңазытқышты құру	80
<i>Ауезова Ә.С., Муратова К.Н., Синчев Б.</i>	
Құрылымданбаған деректерді ақпараттық іздеу әдістері	88
<i>Нәдіров Н.Қ., Дүйсебекова К.С.</i>	
Семантикалық талдауды қолдану арқылы Bigdata негізінде клиент профилін құру жүйесін қалыптастыру	94
<i>Бердыкулова Г.М.</i>	
Ғылыми мақала дегеніміз не немесе оны қалай жазбауға болады?	105

ЭКОНОМИКА ЖӘНЕ МЕНЕДЖМЕНТТЕГІ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

Мұқан Б., Саркамбаева Ш.Г.

Қазақстандағы шағын және орта бизнесте жобалау қызметін дамыту 113

Мейрманова Л.Е.

Адам ресурстарын басқаруда жобаларды басқаруды қолданудың тиімділігі 122

Алдибекова Н.Б., Тяп А.В., Омаров И.Г., Мохамед А.Х., Алимжанова Л.М.

Жобалық тәуекелдерді басқаруда математикалық модельдеу мен бағдарламалық жасақтаманы қолдану 128

Омарова А.Ш., Махбаева Ә.Н.

Жобаларды басқарудағы коммуникацияның мәселелері мен әсері 138

Момбекова А., Бекболова М.Б.

Үлкен деректер аналитикасының жобаны басқару процестерінің тиімділігіне әсері 146

Булантаев А.М., Мұсахан Х.Б., Молдагулова А.Н., Сембина Г.К.

Несие беру үшін банктен күтілетін шығындардың болжауы 154

Омарова Е. Ғ.

Бөлшек несиелеудегі қаржы активтерін жіктеуді автоматтандыру алгоритмі 159

Ауезова А.С. *, Муратова К.Н., Синчев Б.

Международный университет информационных технологий, Алматы, Казахстан

МЕТОДЫ ИНФОРМАЦИОННОГО ПОИСКА НЕСТРУКТУРИРОВАННЫХ ДАННЫХ

Аннотация. В статье предложен новый метод, применяемый для решения задачи информационного поиска неструктурированных (текстовых) данных. Поиск документов осуществляется по ключевым словам, на естественном языке, применяемым в поисковых машинах. Таким образом, на основе полиномиальных алгоритмов создается универсальная машина выборки по нескольким ключам с улучшенными характеристиками по времени и пространству. Данная предлагаемая машина может быть применена для обработки больших данных в различных областях экономики. Для достижения цели в основе новых полиномиальных алгоритмов использована задача о сумме подмножеств, которая относится к классу NP-complete. Эти алгоритмы значительно эффективнее по времени и пространству существующих лучших полиномиальных и экспоненциальных алгоритмов.

Ключевые слова: поиск, метод, алгоритм, неструктурированная информация, поисковая машина, полиномиальные алгоритмы, NP-complete, большие данные

Введение

В настоящее время существуют важные практические и теоретические задачи: теория алгоритмов (theory of algorithms), задача поиска (search problem), задача выполнимости (satisfiability problem), задача принятия решений (decision problem), задача кодирования (encryption problem), задача шифрования (encipherment problem) и другие. Одним из ключевых мест является анализ времени работы алгоритмов и используемая ими память. Задача нахождения k -мерного подмножества из n -мерного множества, сумма элементов которого равна некоторому заданному числу S , является NP-полной (NP-complete). Вычислительная сложность этой задачи зависит от двух параметров – количества n элементов исходного множества и точности p (определяется как количество двоичных разрядов в числах, составляющих множество). Задача становится лёгкой только при очень малых значениях параметров n и p . Если n (количество входных данных) мало, то полный перебор вполне приемлем. Если параметр p (количество разрядов в числах множества) мал, можно использовать динамическое программирование при решении задачи о сумме подмножеств. В классических работах определены время работы алгоритма $T=O(2^{n/2})$ и потребляемая память $M=O(2^{n/4})$, которые не позволяют применять полученные результаты на практике при больших значениях n и на поиск решения необходимо затратить экспоненциальное время. Основным недостатком известных табличных методов (сумм) является построение каждой строки таблицы по свойству определяемым каждым ключевым словом.

К примеру, одна из самых крупных поисковых систем Google рассматривает более 200 различных показателей при оценке веб-сайтов, включая текст, внутренние ссылки, удобство использования веб-сайта и информационную архитектуру. Чтобы получить и сохранить долю рынка онлайн-поиска, поисковым системам необходимо убедиться, что они предоставляют результаты, соответствующие тому, что ищут их пользователи. Они делают это, поддерживая базы данных веб-страниц, которые они разрабатывают с помощью автоматизированных программ, известных как «пауки» или «роботы» для сбора информации. Поисковые системы используют сложные алгоритмы для оценки веб-сайтов и веб-страниц и присваивают им рейтинг по релевантным поисковым фразам. Эти алгоритмы тщательно охраняются и часто обновляются. Поскольку потребители и организации все больше полагаются на поисковые системы при выборе товаров, услуг и поставщиков, в которых они нуждаются, важность поисковых машин с меньшими затратами для современного бизнеса только возрастает.

Методы сбора первичной (исходной) информации

Среди существующих поисковых систем самыми популярными являются два принципа информационного поиска: поиск по ключевым терминам и поиск на основе кластерных методов и векторных моделей [1],[9]. В большинстве случаев информационно-поисковые системы используют метод поиска по ключевым словам, как основной, а кластерные методы в качестве дополнения [10], что существенно повышает эффективность и точность поиска информации. В системах, использующих поиск по ключевым словам, как Google, Яндекс, AstaVista, Yahoo, процесс поиска сводится к поиску во множестве документов входящих каждого из заданных ключевых слов, с последующей сортировкой этих документов по степени релевантности. В подобных системах процесс поиска сводится к поиску во множестве документов входящих каждого из заданных ключевых слов, с последующей сортировкой этих документов по степени релевантности.

Приведем методы информационного поиска: последовательный поиск, точный поиск по алгоритму Бойера-Мура, информационно-поисковые системы, строящие поисковый индекс, хеширующие методы, инвертированные файлы, сигнатурные файлы, суффиксные массивы, деревья, суффиксные деревья, кластерные методы, тернарные деревья поиска и векторные модели, триангуляционные деревья и др.

Эти подходы позволяют перейти к самым современным методам информационного поиска.

Совершенствование работы поисковых систем невозможно без сбора достоверной информации и последующего ее анализа. Поэтому в общей теории и практике информационного поиска повышенное внимание уделяется теоретическим исследованиям, включающим в себя методы сбора информации и ее анализа.

Теоретические исследования используют научную базу таких дисциплин как теория алгоритмов, теория поиска, теория шифрования, теория кодирования, теория принятия решений и др. К основным методам, применяемым для решения научных и практических задач следует отнести:

- технология обработки и анализа больших данных;
- методы исследования вычислительных операций;
- методы алгоритмической разрешимости.

В рамках совершенствования поисковых систем пристальное внимание уделяется дискретным оптимизационным задачам в комбинаторике, такие как задача о ранце.

Описание основных научных вопросов и гипотез

Вопрос о равноправии классов задач сложности P и NP (также общеизвестный как проблема перебора) — это одна из важных открытых проблем теории алгоритмов уже больше чем пяти десятилетий. Если задачи сложности P и NP получить утвердительный ответ, это будет означать, что теоретически возможно решать многие сложные задачи существенно быстрее, чем сейчас.

В разделе теории о вычислительной сложности алгоритмов рассматриваются отношения между классами P и NP . Она изучает наиболее общие ресурсы для решения некоторой задачи. Ресурсы — это время (количество шагов) и память (необходимый объем памяти для решения задачи). Проблема равенства классов P и NP входит в семь задач тысячелетия, за решение которой Математический институт Клэя назначил премию в миллион долларов США.

Теоретические и прикладные исследования проблемы базируются на методах системного анализа, Web Service, Big Data, Hadoop/Map Reduce.

Машина, созданная по данной методике, базируется на основные положения по разработке поисковых систем, а доступ к ним осуществляется на основе поискового запроса на естественном языке.

Реализация машины осуществляется на основе современной теории класса NP-complete.

Исследование функциональных качеств машины проводится на основе контрольных баз данных, а корректность работы проверяется и доказывается методами и полиномиальными алгоритмами. Результаты научного исследования по данному проекту позволят:

- развивать теорию алгоритмов и практику поиска произвольной информации;
- решить сложные задачи, связанные с равенством классов P и NP;
- подготовить специалистов и молодых ученых в области информационных технологий;
- создать технологии и инструментарии для разработки поисковых машин.

На сегодняшний день имеются компании, которые заинтересованы в результатах данного проекта. Это компании, которые занимаются большими данными.

Обоснование научной новизны

В n -элементном множестве целых чисел найти подмножество, сумма элементов которого близка сертификату S . Эта задача о сумме подмножеств связана с проблемой ранца, которая решена Р. Беллманом [1] в 1956 году. Им предложен псевдополиномиальный алгоритм с временем выполнения $T=O(nS)$ на основе метода динамического программирования. Поскольку этот алгоритм псевдополиномиального времени является фундаментальной частью информационных технологий, а Subset Sum является одной из основных NP-трудных проблем, важно решить, можно ли улучшить время выполнения $O(nS)$.

Спустя 43 года Д. Писинжер [2] улучшил решение этой задачи за время $T=O(nS/\log S)$, применяя битовое представление сертификата S .

Недавно Килильяс и Сюй представили в [3] псевдополиномиальный алгоритм «разделяй и властвуй», который вычисляет все реализуемые суммы до целого числа в оценке времени $O(\min(\sigma, S))$ выполнения алгоритма, где σ - сумма всех элементов исходного множества. Они используют хеширование для быстрого решения задачи о сумме подмножеств, если исходное множество содержится в небольшом интервале. Затем они сводят общий случай к случаю малого интервала путем расщепления исходного множества на меньшие подмножества. Они считают, что новый алгоритм является самым быстрым общим детерминистическим алгоритмом для этой задачи. Однако решение исходной задачи находится с некоторой точностью.

В 2017 году К. Биргман предложил псевдополиномиальный рандомизированный алгоритм [4], работающий за время $O(n+S)$. Точнее, учитывая экземпляр задачи о сумме подмножеств (S) , вычисляется множество всех сумм $s \in S$, генерируемых небольшими подмножествами $Y \subseteq S$ размерности $|Y| \leq k$ с постоянной вероятностью принадлежности сертификата S множеству сумм $s \in Y$. В работе не указана величина вероятности.

Важно отметить, что детальный обзор современных результатов, содержащихся в более 60 научных трудах, по известным лучшим алгоритмам решения задачи о сумме подмножеств приведен в работах [3, 4].

Наряду с псевдополиномиальными алгоритмами имеются точные алгоритмы с экспоненциальным временем работы, когда в n -элементном множестве существует хотя одно подмножество, сумма элементов которого равна S . В первую очередь, к ним относятся работы Горовиц, Санни [5] с временем выполнения алгоритма и требуемым пространством и Шрепеля, Шамира [6] с временем выполнения алгоритма и требуемым пространством. При этом они использовали инвариантные соотношения.

В работах [7,8,9] введены понятия полиномиального (практического) алгоритма и его сложности. В [10] показано, что сложность работы алгоритма для задачи о сумме подмножеств составляет не менее $2^{n/2}$, где n — число переменных. Более детальное исследование сложности решения задачи о сумме подмножеств проведено в работах [11,12]. В них полу-

чены верхние и нижние оценки максимальной экспоненциальной сложности решения этой задачи методом ветвей и границ.

Сложность алгоритма напрямую используется при определении времени работы известных алгоритмов поискового запроса неструктурированной информации по многим ключевым словам [13,14] и методов обработки и анализа больших данных [15].

Важно подчеркнуть, что определение сложности решения задачи о сумме подмножеств является фундаментальной проблемой, используемой в качестве стандартной задачи, которая может быть решена за слабо полиномиальное время во многих дисциплинах по образовательным программам. В качестве основной задачи в теоретической информатике и учебных материалах [16,17] рассматривается задача k -SUM, которая является подзадачей задачи о сумме подмножеств и связана с методом перебора.

В работе [18] для параметра мощности $k=2$ разработаны уникальный метод и универсальные алгоритмы с временем и требуемым пространством для выборки искомого подмножества и для параметра мощности $k=3$. Для параметра сложности $k=2$ показана разрешимость равенства классов P и NP. Другими словами, найдено решение этой задачи, в которой проверка правильности решения задачи будет равна получению решения.

Описание методов исследования

Основные методы исследования базируются на современных методах и алгоритмах решения базовой задачи-задача о сумме подмножеств, которая относится к классу NP-complete. Решение хотя бы одной задачи из этого класса позволяет решить все оставшиеся задачи этого класса.

Постановка задачи. Задача о сумме подмножеств формулируется в виде:

$$\sum_{i=1}^n \alpha_i x_i = S, \alpha_i \in \{0,1\}, x_i \in X^n, i \in N, \quad (1)$$

где X^n – множество целых положительных чисел, размерность $n = |X^n|$, $x_i < +\infty$, N - множество натуральных чисел с размерностью $n = |N|$, $n < +\infty$, Подзадачей задачи (1) будем называть следующую задачу:

$$\sum_{i=1}^k \alpha_i x_i = S, \alpha_i \in \{0,1\}, x_i \in X^n, \alpha_i = 1, i \in K, \alpha_i = 0, i \in N \setminus K, \quad (2)$$

где $X^k \subseteq X^n$, $k = |X^k|$, $k \leq n$, $K \subseteq N$, $k=|K|$, K – подмножество индексов всех выбранных переменных $x_i \in X^n$, $N \setminus K$ - подмножество индексов всех невыбранных переменных $x_i \in X^n$ подзадачи (2). Отметим, что задача (1) является частным случаем задачи (2), когда множество $K = \emptyset$.

Предложены полиномиальные алгоритмы решения задачи о сумме подмножеств (2) на основе работ [18, 19] с разделением исходного множества X^n на два подмножества. Затем применяется к этим подмножествам в отдельности алгоритм генерации сочетания C_n^m для получения необходимых подмножеств и потом используется метод слияния для полученных подмножеств, из которых находим искомое подмножество X^k . Здесь $k=2m$.

По результатам проведенного анализа существующих поисковых систем были сделаны выводы о том, что алгоритмы поиска частично удовлетворяют современным требованиям сбора, хранения, обработки и анализа больших объемов данных.

Все поисковые процессы должны протекать с соблюдением трех основных характеристик больших данных:

1. Volume
2. Velocity
3. Variety

Все три характеристики должны быть применены к базовой задаче, как правило, с несколькими переменными и нетривиальным условием. Работы с большими данными технологии: поддержка Hadoop, интеграция с программно-аппаратными комплексами.

Заклучение

На основе полученных результатов можно сделать следующие выводы:

- приведенный обзор работ показывает, что важная задача о сумме подмножеств в теории сложности алгоритмов относится к основной из трудных проблем класса NP-complete и может быть использована по разработке поисковых систем;
- предложены полиномиальные алгоритмы решения задачи о сумме подмножеств (2) на основе работ [18, 19] с разделением исходного множества X^n на два подмножества;
- использованы алгоритм генерации сочетания C_n^m и метод слияния для получения искомого подмножества X^k .

ЛИТЕРАТУРА

1. Richard Bellman. Notes on the theory of dynamic programming iv - maximization over discrete sets. Naval Research Logistics Quarterly, 3(1-2):P.67-70, 1956.
2. David Pisinger. Linear time algorithms for knapsack problems with bounded weights. Journal of Algorithms, 33(1):1 – 14, 1999.
3. Konstantinos Koiliaris, Chao Xu. A Faster pseudopolynomial time algorithm for subset sum. To appear in SODA '17, 2017. // arXiv:1610.04712v2[cs.Ds] 8 Jan 2017.-18p.
4. Karl Bringmann. A near-linear pseudopolynomial time algorithm for subset sum. To appear in SODA '17, 2017. // arXiv:1610.04712v2[cs.Ds] 8 Jan 2017.-18p.
5. E. Horowitz, S. Sanni. Computing Partitions with Application to the Knapsack Problem //Journal of the ACM(JACM), 1974, T21, pp.277-292
6. R. Schroepfel, A. Shamir. A $T=O(2^{n/2})$, $S=O(2^{n/4})$ Algorithm for Certain NP-Complete Problem // SIAM Journal on Computing, 1981, Vol.10, № 3, pp.456-464
7. Cobham A. The intrinsic computational difficulty of functions. //In Proceedings of the Con-gress for logic, methodology and philosophy of science.-NorthHoiLand, 1964.P.24-30.
8. Egmonds J. Parths, treers and flowers. // Canadian Journal of mathematics. -1965. Vol.17. – P.449-467.
9. Николаев А.В. Геометрический подход к задаче о разрезе. -Ярославль, ЯрГУ, 2014, -68с.
10. Robert G. Jeroslow. Trivial integer programs unsolvable by branch-and-bound // Mathematical Programming. — 1974. — V. 6. — P. 105-109
11. Krishnamoorth B. Bounds on the size of branch-and-bound proofs for integer knapsacks // OR Letters. — 2008. — V. 36, № 1. — P. 19-25.
12. Колпаков Р.М., Посыпкин МА. Верхняя оценка числа ветвлений для задачи о сумме под-множеств. //Матем. Вопросы кибернетики. Вып.18.-М.: Физматлит, 2013.-с.213-226
13. C. J. van Rijsbergen. "Information Retrieval." Dept. of Computer Science. University of Glas-gow, 1979
14. A. Adamansky. Overview of full-text search methods and algorithms. -Novosibirsk, Novosibirsk State University, 2018, -26p.
15. Min Chen, Shiwen Mao, Yin Zhang, Victor C.M. Leung. Big Data. Related Technologies, Chal-lenges, and Future Prospects. — Springer, 2014. — 100 p.
16. *Лифшиц Юрий*. Точные алгоритмы и открытые проблемы // Современные задачи *теоретической информатики*. [Электронный ресурс] URL: <http://download.yandex.ru/class/> (дата обращения: 15.01.2021)
17. Куликов Александр. Алгоритмы NP-трудных задач. Лекции. //Computer Science клуб Санкт-Петербургского отделения при Математическом Институте имени В.А. Стеклова, 2009. -с.13-16
18. B. Sinchev, A.B. Sinchev, J. Akzhanova, A.M. Mukhanova. New methods of information search.
19. // News of the National Academy of Sciences of Kazakhstan, Series of Geology and Technical Sciences, Volume 3, Number 435 (2019), pp. 240-246
B. Sinchev, A.B. Sinchev, J. Akzhanova, Y. Issekeshv, A.M. Mukhanova. Polynomial time algorithms for solving NP-complete problems. // News of the National Academy of Sciences of Kazakhstan, Series of Geology and Technical Sciences, Volume 3, Number 441 (2020), pp.97-101
20. S.G. Akl. Adaptive and optimal algorithms for enumerating perinutations and combinations. //The Computer Journal, 30 (1987), pp. 433-436
21. B. Sinchev, A. B. Sinchev, Z.A. Akzhanova. Computing network architecture for reducing a computing operation time and memory usage associated with determining, from a set of data el-ements, a subset of at least two data elements, associated with a target computing operation re-sult. // Патент в USPTO, 2019.-38p.

Ауезова Ә.С.*, Муратова К.Н., Синчев Б.

Құрылымданбаған деректерді ақпараттық іздеу әдістері

Ауезова Әнел Саттарқызы, PhD-докторанты, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының лекторы.

Муратова Камила Нурланқызы, магистр, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының лекторы.

Синчев Бахтгерей, т.ғ.д., Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының профессоры.

Андатпа. Мақалада құрылымданбаған (мәтіндік) деректерді ақпараттық іздеу мәселесін шешу үшін қолданылатын жаңа әдістер қарастырылған. Құжаттарды іздеу негізгі сөздер бойынша іздеу машиналарында қолданылатын табиғи тілде жүзеге асырылады. Осылайша көпмүшелік алгоритмдер негізінде уақыт пен кеңістік бойынша жақсартылған сипаттамалары бар бірнеше кілттер бойынша әмбебап іріктеу машинасы жасалады. Бұл ұсынылған машина экономиканың әртүрлі салаларында үлкен деректерді өңдеу үшін қолданылуы мүмкін. Мақсатқа жету үшін жаңа полиномиялық алгоритмдер негізінде NP-complete класына жататын ішкі жиындардың қосындысы туралы есеп пайдаланылды. Полиномиялық және экспоненциалды алгоритмдердің уақыты мен кеңістігінде, осы алгоритмдер жақсырақ және едәуір тиімді.

Түйін сөздер: іздеу, әдіс, алгоритм, құрылымданбаған ақпарат, іздеу машинасы, көпмүшелік алгоритмдер, NP-толық, үлкен деректер (Big data)

Auyezova A.S. *, Muratova K.N., Sinchev B.

Methods of information search for unstructured data

Anel S. Auyezova, PhD-doctoral student, lecturer, «Information Systems» Department, Inter-national Information Technology University.

Kamila N. Muratova, Master of Tech. Sci., lecturer, «Information Systems» Department, International Information Technology University.

Bakhtgerey Sinchev, Doct. of Tech. Sci., Professor, «Information Systems» Department, In-ternational Information Technology University.

Abstract. The article discusses new methods used to solve the problem of information retrieval of unstructured (text) data. Search for documents is carried out by keywords in the natural language used in search engines. Thus, on the basis of polynomial algorithms, a universal multi-key sampling machine is created with improved time and space characteristics. This proposed machine can be used for processing big data in various areas of the economy. To achieve this goal, the new poly-nomial algorithms are based on the problem of the sum of subsets, which belongs to the NP-complete class. These algorithms are significantly more time and space efficient than the existing best polynomial and exponential algorithms.

Keywords: search, method, algorithm, unstructured information, search engine, polynomial algorithms, NP-complete, big data

Сведения об авторах:

Ауезова Анель Саттаровна, PhD-докторант, лектор кафедры «Информационных систем» Международного университета информационных технологий.

Муратова Камила Нурлановна, магистр, лектор кафедры «Информационных систем» Международного университета информационных технологий.

Синчев Бахтгерей, д.т.н., профессор кафедры «Информационных систем» Международного университета информационных технологий.

INTERNATIONAL JOURNAL OF INFORMATION AND
COMMUNICATION TECHNOLOGIES

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

Ответственный за выпуск	Есбергенов Досым Бектенович
Редакторы	Далабаева Айсара Касымбековна Садганова Эльмира Абуовна
Компьютерная верстка	Туратауова Айжаркын Ахметовна
Компьютерный дизайн	Туратауова Айжаркын Ахметовна

Редакция журнала не несет ответственности за
недостоверные сведения в статье и
неточную информацию по цитируемой литературе

Подписано в печать 26.03.2021 г.
Тираж 500 экз. Формат 60x84 1/16. Бумага тип.
Уч.-изд.л. 10.1. Заказ №158

Издание международный университет информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59