

МРНТИ 78.15.00; 78.01.21

<https://doi.org/10.58805/kazutb.v.1.18-85>**ПОИСК СЕМАНТИЧЕСКОЙ СВЯЗИ МЕЖДУ ОБЪЕКТАМИ****А.Т. Мазакова², Ш.А. Джомартова², Т.Ж. Мазаков^{1,2}, Д.К. Мухаев²,
Е.К. Мергенгали²**¹Институт информационных и вычислительных технологий КН МНВО РК, Алматы, Казахстан,²Казахский Национальный университет им. аль-Фараби, г. Алматы, Казахстан,
jomartova@mail.ru

Проблема аналитической обработки информации из множества информационной систем является одним из важнейших аспектов развития современного общества. В настоящее время решение этой проблемы в области разработки и эксплуатации информационных систем различного назначения (военных, технических, экономических, медицинских, социальных и др.) связано с разработкой различных аналитических продуктов, позволяющих выделять значимую информация для пользователей из множества всевозможных данных. В статье исследована возможность выделения связанной между собой группы лиц из некоторого однородного массива данных на основе применения матричного исчисления.

Разработано программное обеспечение на языке программирования Python. Применены два подхода: 1) основанный на методах матричного исчисления и 2) метод Дейкстра. Эффективность предложенных алгоритмов продемонстрировано на модельных задачах.

Ключевые слова: матрица, поиск информации, семантика, сортировка, Python.

**ОБЪЕКТІЛЕР АРАСЫНДАҒЫ СЕМАНТИКАЛЫҚ
БАЙЛАНЫСТЫ ІЗДЕУ****Ә.Т. Мазақова², Ш.А. Джомартова², Т.Ж. Мазаков^{1,2}, Д.К. Мухаев²,
Е.К. Мергенгали²**¹Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан,²әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,
jomartova@mail.ru

Көптеген ақпараттық жүйелерден ақпаратты аналитикалық өңдеу мәселесі қазіргі қоғам дамуының маңызды аспектілерінің бірі болып табылады. Қазіргі уақытта әр түрлі мақсаттағы ақпараттық жүйелерді (әскери, техникалық, экономикалық, медициналық, әлеуметтік және т.б.) әзірлеу және пайдалану саласындағы осы мәселені шешу пайдаланушыларға әр түрлі мәліметтерден маңызды ақпарат бөлуге мүмкіндік беретін әртүрлі аналитикалық өнімдерді әзірлеумен тікелей байланысты. Мақалада матрицалық есептеуді қолдану негізінде біртекті мәліметтер массивінен өзара байланысты адамдар тобын оқшаулау мүмкіндігі зерттелген.

Python бағдарламалау тілінде программалық жасақтама жасалды. Мұнда екі тәсіл қолданылады: 1) матрицалық есептеу әдістеріне негізделген тәсіл және 2) Дайкстра әдісі. Ұсынылған алгоритмдердің тиімділігі модельдік есептерде көрсетілген.

Түйін сөздер: матрица, ақпаратты іздеу, семантика, сұрыптау, Python.

FINDING THE SEMANTIC CONNECTION BETWEEN OBJECTS

**A.T. Mazakova², Sh.A. Jomartova², T.Zh. Mazakov^{1,2}, D.K. Mukhayev²,
Ye.K. Mergengali²**

¹RSE Institute of Information and Computational Technologies MSHE RK CS,
Almaty, Kazakhstan,

²Al-Farabi Kazakh National University, Almaty, Kazakhstan,
jomartova@mail.ru

The problem of analytical processing of information from multiple information systems is one of the most important aspects of the development of modern society. Currently, the solution to this problem in the development and operation of information systems for various purposes (military, technical, economic, medical, social, etc.) is associated with the development of various analytical products that allow to allocate meaningful information for users from the set of all kinds of data. This article explores the possibility of selecting a related group of individuals from a homogeneous data set by applying matrix calculus.

Developed software in the Python programming language. Two approaches were used: 1) based on the methods of matrix calculus and 2) the Dijkstra method. The effectiveness of the proposed algorithms is demonstrated on model problems.

Keywords: matrix, information retrieval, semantics, sorting, Python.

Введение. Интенсивное развитие сетевых технологий в современных процессах информатизации привело к стремительному росту информационных сетей и расширению их функциональных возможностей. Наглядным примером может служить распространение глобальной сети INTERNET, успешно преодолевающей межгосударственные и иные границы. Абонентами сети стали специалисты самых различных профессий. С целью удовлетворения своих информационных потребностей они могут обращаться к различным базам данных, независимо от их местонахождения. Если еще недавно пользователи имели дело с базами данных, в состав которых входили десятки, сотни тысяч записей, то в условиях глобаль-

ных сетей это число возросло на несколько порядков. Это обстоятельство ставит перед проектировщиками современных автоматизированных систем документального поиска (АСДП) ряд новых проблем, решение которых возможно лишь путем разработки принципиально новых методов и средств информационного поиска работ [1-2].

Материалы и методы. В работах [3-5] рассматривается широкий спектр вопросов, связанных с методологией, организацией и технологиями информационно-аналитической работы. Показаны возможности использования аналитического инструментария для исследования социально-политических и экономических процессов, прогнозирования и организации эффектив-

ного функционирования и развития систем управления предприятиями и учреждениями, совершенствования процессов принятия управленческих решений. На уровне «живого знания» в широком культурно-историческом контексте раскрывается сущность интеллектуальных технологий, приемов прикладной аналитической работы. Представлена характеристика зарубежных и отечественных аналитических центров.

В работах [6-8] представлены описания различных аналитических инструментов, которые позволяют решать задачи извлечения, очистки, анализа, моделирования и визуализации данных, прогнозирования событий и создания систем отчетности.

Вопросам добывания, сбора, учета, систематизации, анализа и обобщения, а также описанию методов аналитических исследований посвящены следующие работы [9-12].

Несмотря на то, что использование различных аналитических программных продуктов в настоящее время приобретает все большую популярность, необходимы новые алгоритмы, методы и средства получения дополнительной информации.

Целью представленной работы является

разработка алгоритма выделения группы лиц, объединенных определенными признаками из больших баз данных и его программная реализация.

Результаты и обсуждение. В работе [3] показано соотношение между информационной и аналитической работой. Поскольку грань между этими понятиями размыта, информационно-аналитическая деятельность воспринимается как единое целое, хотя на самом деле есть существенная разница (Рис. 1). Как правило, затруднений в этом информационном обеспечении значительно меньше, чем на этапе собственно аналитической обработки информации. Каждый сотрудник аналитического подразделения умеет собирать (добывать) информацию, вести её первичную обработку – систематизацию, классификацию, организовать хранение.

Сложности начинаются тогда, когда нужно из этой информации получить новые знания, принять эффективное управленческое решение, написать аналитический документ. Для этого нужны соответствующие навыки аналитической обработки информации, включая новые программно-аналитические инструментари.



Рис. 1- Соотношение между информационной и аналитической работой [3].

Для разработки программно-аналитических средств применяются различные математические модели. Научных исследования в данной работе основаны на применении теории графов и методов вычислений [13-15].

Входные и выходные данные организованы в виде отдельных файлов и содержат информацию одного из перечисленных ниже типов:

- параметрические данные;
- исходные данные, представленные в файле в заданном формате;
- результативный файл, содержащий результаты численных расчетов;

Для представления исходных объектов используется следующий формат исходных данных. В исходный файл построчно записывается следующая информация: 1) имя первого объекта, 2) имя второго объекта и 3) уточняющая информация. Все поля разделяются символами табуляции.

Из исходной информация об объектах исследования формируется список объектов

V представленный в виде N -мерного вектора. Длина вектора N зависит от исходных данных и автоматически формируется в процессе считывания данных из исходного файла.

Далее формируется матрица связей A размерности $N * N$. Вначале матрица A обнуляется. Далее построчно просматривается исходная матрица. Элементу $A[i,j]$ присваивается значение равное 1, если первый объект строки совпадает с $V[i]$ и второй объект строки совпадает с $V[j]$.

Для выявления связей между объектами применены два подхода: 1) на основе матричного исчисления и 2) на основе метода Дейкстры.

При вызове программы появляется заставка, в которой представлена организационно-разработчик и название программы.

Далее после выбора режима «Расчет» появляется следующая форма для ввода параметрической информации (Рис. 2).

The image shows a screenshot of a software application window. The title bar reads "Казакский Национальный Университет имени аль-Фараби". The main area of the window contains two input fields. The first field is labeled "Введите название файла" and contains the text "avia". The second field is labeled "Введите количество итераций" and contains the number "4". Below these fields are two buttons: "Рассчитать в файл" and "Выход".

Рис. 2- Форма ввода параметрических данных

В форме ввода необходимо заполнить имя исходного файла (без указания расширения txt). Также надо указать максимальное количество итераций-уровней вложенностей при поиске зависимостей между объектами. По умолчанию равно 3.

Далее появляется всплывающее окошко (Рис. 3), который позволяет вывести в результирующий файл содержимое исходного файла.

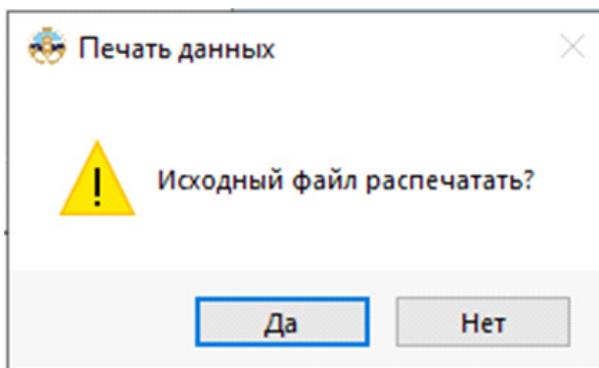


Рис. 3- Окошко выбора режима печати исходной информации

Окошко, представленное на рисунке 4, свидетельствует о нормальном завершении текущего расчета.

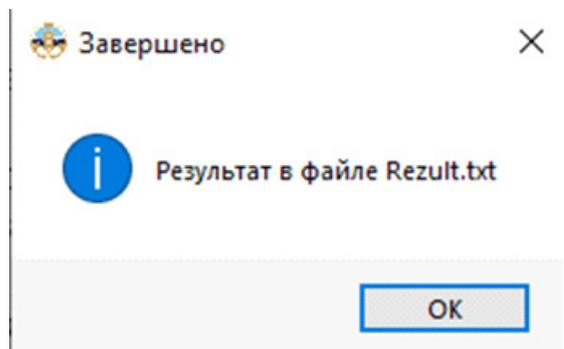


Рис. 4- Окошко сигнализации о завершении расчета

Далее появляется окно, представленное на рисунке 2, для выбора продолжения: вы-

полнение нового расчета или выхода из программы.

В результирующий файл записываются численные расчеты о взаимосвязанности объектов.

Ниже представлены результаты нахождения связей между объектами различной природы двумя программами.

1) определение группы лиц, связанных между собой телефонными переговорами, матричным методом

Программа нахождения семантических связей между объектами матричным методом:

имя программы – Semantic.exe
язык разработки – Python

Python Дата 21/2/2023 Время 12:5

Список объектов

- А
- Б
- В
- Г
- Д
- Е

Исходные данные из файла a2.txt

Количество данных = 4

- А – Б – Прим1
- Б – В – Прим2
- В – Г – Прим3
- Д – Е – Прим4

Количество итераций = 3

Матрица связей

Строка 1	0	1	0	0	0	0
Строка 2	1	0	1	0	0	0
Строка 3	0	1	0	1	0	0
Строка 4	0	0	1	0	0	0
Строка 5	0	0	0	0	0	1
Строка 6	0	0	0	0	1	0

Результативная матрица связей

Строка 1	0	2	1	1	0	0
Строка 2	1	0	2	1	0	0
Строка 3	1	2	0	1	0	0

Строка 4 1 1 2 0 0 0

Строка 5 0 0 0 0 0 1

Строка 6 0 0 0 0 1 0

Результаты определения связей между объектами

Объект А связан:

Б

В

Г

Объект Б связан:

А

В

Г

Объект В связан:

А

Б

Г

Объект Г связан:

А

Б

В

Объект Д связан:

Е

Объект Е связан:

Д

2) определение возможности вылета из одного аэропорта в другие.

Программа нахождения семантических связей между объектами методом Дейкстры:

имя программы – SemDeik.exe

язык разработки – Python

Python Дата 26/2/2023 Время 12:5

Список объектов

Алматы

Англия

Астана

Бишкек

Москва

Ташкент

Томск

Матрица связей

Строка 1 0 0 1 0 0 0 0

Строка 2 0 0 1 0 0 0 0

Строка 3 1 1 0 0 0 0 1

Строка 4 0 0 0 0 1 0 0

Строка 5 0 0 0 1 0 1 0

Строка 6 0 0 0 0 1 0 0

Строка 7 0 0 1 0 0 0 0

Результаты определения связей между объектами

Объект Алматы связан с:

Англия. Вес = 2

Астана. Вес = 1

Томск. Вес = 2

Объект Англия связан с:

Алматы. Вес = 2

Астана. Вес = 1

Томск. Вес = 2

Объект Астана связан с:

Алматы. Вес = 1

Англия. Вес = 1

Томск. Вес = 1

Объект Бишкек связан с:

Москва. Вес = 1

Ташкент. Вес = 2

Объект Москва связан с:

Бишкек. Вес = 1

Ташкент. Вес = 1

Объект Ташкент связан с:

Бишкек. Вес = 2

Москва. Вес = 1

Объект Томск связан с:

Алматы. Вес = 2

Англия. Вес = 2

Астана. Вес = 1

Выводы. На основе системы программирования Python разработаны два аналитических приложения, позволяющие выявлять семантические связи между объектами.

На модельной задаче «*определения группы лиц, связанных между собой телефонными переговорами, матричным методом*» продемонстрирована работа про-

граммы выделения связей между лицами. На примере видно, что объект «А» связан с объектом «Б» напрямую, с объектом «В» через объект «Б» и с объектом «Г» через объекты «б» и «В». Со следующими объектами «Д» и «Е» объект «А» не связан.

На модельной задаче *«определение возможности вылета из одного аэропорта в другие»* продемонстрирована работа программы, основанная на методе Дейкстры, выделения связей между аэропортами. В отличие от предыдущего примера, здесь дополнительно выводится вес связи. Так например, из «Алматы» можно напрямую вылететь из «Астаны». Из «Алматы» в «Ан-

глию» и «Томск» через одну пересадку.

Эффективность предложенных программ продемонстрирована на двух примерах.

В дальнейшем полученные программы могут быть использованы для изучения связей между объектами других социально-экономических систем.

Работа выполнена за счет средств программно-целевого финансирования научных исследований на 2023-2025 годы по проекту ИРН АР19676966 «Разработка программно-аппаратного комплекса психофизиологического отбора и реабилитации снайперов».

Литература

1. Зупарова Л.Б., Зайцева Т.А. Аналитико-синтетическая переработка информации. – М.: Изд-во «ФАИР», 2008. – 400 с.
2. Демидов А.А., Захаров Ю.Н. Информационно-аналитические системы поддержки принятия решений в органах государственной власти и местного самоуправления. Основы проектирования и внедрения. – СПб.: НИУ ИТМО, 2012. – 100 с.
3. Курносов Ю.В. Алгебра аналитики. Секреты мастерства в аналитической работе. – М.: Изд-во «Русская аналитическая школа», 2015. – 387 с.
4. Курносов Ю.В. Азбука аналитики. – М.: Изд-во «Русская аналитическая школа», 2013. – 280 с.
5. Курносов Ю.В. Аналитика как интеллектуальное оружие. – М.: Изд-во «Русская аналитическая школа», 2012. – 1200 с.
6. Белов В.С. Информационно-аналитические системы. – М.: Московский гос. университет экономики, статистики и информатики, 2004. -116 с.
7. Калачев Г.А., Стасюк О.Н. Информационно-аналитические системы. – Омск: Изд-во СибаДИ, 2010. – 101 с.
8. Курносов Ю.В., Конотопов П.Ю. Аналитика: методология, технология и организация информационно-аналитической работы. – М.: РУСАКИ, 2004. – 512 с.
9. Левкин И.М., Микадзе С.Ю. Добывание и обработка информации в деловой разведки. – СПб: Университет ИТМО, 2015. – 460 с.
10. Филиппов В.А. Интеллектуальный анализ данных. Методы и средства. – М.: Эдиториал УРСС, 2001. – 52 с.
11. Кузнецов И.Н. Учебник по информационно-аналитической работе. Информация: сбор, защита, анализ. – М.: Яуза, 2001. – 320 с.
12. Пилко И.С. (науч. редактор). Методы информационно-аналитической. – Кемерово: КемГУКИ, 2010. – 228 с.

13. Кремер Н.Ш., Фридман М.Н., Тришин И.М. Линейная алгебра. – М.: Юрайт, 2022. – 423 с.
14. Алексеев В.Е., Таланов В.А. Графы и алгоритмы, Структуры данных, Модели вычислений. – М.: Бином, 2012. – 320 с.
15. Касьянов В.Н., Евстигнеев В.А. Графы в программировании. Обработка, визуализация и применение. – СПб.: БХВ-Петербург, 2003. – 1104 с.

References

1. Zuparova L.B., Zaitseva T.A. Analytic and synthetic processing of information. – M.: FAIR Publishing House, 2008. – 400 p.
2. Demidov A.A., Zakharov Y.N. Information and analytical systems to support decision-making in public authorities and local self-government. Fundamentals of design and implementation. – SPb.: NRU ITMO, 2012. – 100 p.
3. Kurnosov Y.V. Algebra of Analytics. Secrets of mastery in analytical work. – M.: Russian Analytical School Publishing House, 2015. – 387 p.
4. Kurnosov Y.V. The ABC of Analytics. – M.: Russian Analytical School Publishing House, 2013. – 280 p.
5. Kurnosov Yu.V. Analytics as an Intellectual Weapon. – M.: Russian Analytical School Publishing House, 2012. – 1200 p.
6. Belov V.S. Information and analytical systems. – M.: Moscow State University of Economics, Statistics and Informatics, 2004. – 116 p.
7. Kalachev G.A., Stasiuk O.N. Information and analytical systems. – Omsk: SibADI Publishing House, 2010. – 101 p.
8. Kurnosov Y.V., Konotopov P.Y. Analitika: methodology, technology and organization of information-analytical work. – M.: RUSAKY, 2004. – 512 p.
9. Levkin I.M., Mikadze S.Y. Information extraction and processing in business intelligence. – SPb.: ITMO University, 2015. – 460 p.
10. Filippov V.A. Intelligent Data Analysis. Methods and Means. – M.: Editorial URSS, 2001. – 52 p.
11. Kuznetsov I.N. Textbook on information-analytical work. Information: collection, protection, analysis. – M.: Yauza, 2001. – 320 p.
12. Pilko I.S. (scientific editor). Methods of information-analytical. – Kemerovo: KemGUKI, 2010. – 228 p.
13. Kremer N.Sh., Friedman M.N., Trishin I.M. Linear algebra. – M.: Yurite, 2022. – 423 p.
14. Alexeev V.E., Talanov V.A. Graphs and algorithms, Data structures, Computation models. – M.: Binom, 2012. – 320 p.
15. Kasyanov V.N., Evstigneev V.A. Graphs in Programming. Processing, visualization and application. – SPb.: BXV-Peterburg, 2003. – 1104 p.

Сведения об авторах

Мазакова Айгерим Талгатовна – докторант НАО Казахского национального университета имени аль-Фараби, aigerym97@mail.ru;

Джомартова Шолпан Абдрақовна – доктор технических наук, доцент НАО Казахского национального университета имени аль-Фараби, jomartova@mail.ru;

Мазаков Талгат Жақупович – доктор физико-математических наук, главный научный сотрудник Института Информационных и вычислительных технологий КН МНВО РК, профессор НАО Казахского национального университета имени аль-Фараби, tmazakov@mail.ru;

Мухаев Дарын Какенулы – докторант НАО КазНУ имени аль-Фараби; daryn.mukhayev@gmail.com

Мергенгали Естай Куанышулы – докторант НАО КазНУ имени аль-Фараби; estony.9999@gmail.com

Information about the authors

Aigerim Mazakova – doctoral student at Al-Farabi Kazakh National University, Almaty, Kazakhstan; aigerym97@mail.ru;

Sholpan Jomartova – Al-Farabi Kazakh National University, doctor of technical sciences, ass.professor, Almaty, Kazakhstan; jomartova@mail.ru;

Talgat Mazakov – Al-Farabi Kazakh National University, doctor of physical and mathematical sciences, professor, Almaty, Kazakhstan, Chief Researcher at the RSE Institute of Information and Computational Technologies of the National Academy of Sciences of the Republic of Kazakhstan; tmazakov@mail.ru;

Daryn Mukhayev – doctoral student at Al-Farabi Kazakh National University, Almaty, Kazakhstan; daryn.mukhayev@gmail.com;

Yestay Mergengali – doctoral student at Al-Farabi Kazakh National University, Almaty, Kazakhstan; estony.9999@gmail.com