

УДК 004.942

Аналитическая марковская модель информационного поиска в облачных хранилищах с учетом нечеткой формулировки критериев качества и предпочтений пользователей

Паращук И. Б., Саяркин Л. А.

Постановка задачи: реализация адекватных и эффективных аналитических моделей процесса смены состояний показателей качества реализации поисковых запросов пользователей в больших облачных хранилищах с учетом нечеткой формулировки критериев качества поиска и предпочтений этих пользователей. **Целью работы** является решение задачи устранения несоответствия между требованиями к современным аналитическим моделям реализации поисковых запросов и ограниченными возможностями существующих формальных моделей для математического описания поисковых систем, учитывающего нечеткость критериев качества поиска и неопределенность пользовательских предпочтений в системах данного класса. **Используемые методы:** подход ориентирован на использование управляемых цепей Маркова в форме разностных стохастических уравнений в сочетании с математически корректными алгоритмами идентификации нечетко заданных вероятностей переходов из состояния в состояние, обусловленных нечетко сформулированными критериями качества поиска и предпочтениями пользователей. **Результаты и их новизна:** элементами новизны представленной модели является использование универсального механизма аналитического моделирования, основанного на вероятностно-временной модели изменения состояний процесса реализации поисковых запросов и связанных с ним показателей качества. Этот механизм учитывает динамическую природу, вероятностный характер и нестационарность процесса поиска, учитывает нечетко сформулированные критерии качества и пользовательские предпочтения и нивелирует этот вид неопределенности, используя построение множества недоминирующих альтернатив на базе нечетких отношений предпочтения. **Практическая значимость:** представленный в работе подход к моделированию может быть применен для реальной разработки и практического построения аналитических моделей, описывающих процесс изменения состояния показателей качества выполнения пользовательских поисковых запросов в крупных облачных хранилищах с учетом нечеткой формулировки критериев качества поиска и предпочтений пользователей.

Ключевые слова: информационный поиск, нечеткое множество, облачное хранилище, марковская модель, критерий качества, пользователь.

Введение

Во всем мире строятся глобальные, доступные, гибкие и относительно безопасные распределенные системы облачного хранения данных с «зоной охвата», если не в континентальном масштабе, то в рамках гигантских территорий или в широких границах национальных пространств.

Такие структуры часто охватывают и обслуживают крупные ведомства или компании, призваны обеспечить предобработку и, главное, доступ и дли-

Библиографическая ссылка на статью:

Паращук И. Б., Саяркин Л. А. Аналитическая марковская модель информационного поиска в облачных хранилищах с учетом нечеткой формулировки критериев качества и предпочтений пользователей // Системы управления, связи и безопасности. 2025. № 2. С. 1-17. DOI: 10.24412/2410-9916-2025-2-001-017

Reference for citation:

Parashchuk I. B., Sayarkin L. A. Analytical Markov model of information retrieval in cloud storages taking into account fuzzy formulation of quality criteria and user preferences. *Systems of Control, Communication and Security*, 2025, no. 2, pp. 1-17 (in Russian). DOI: 10.24412/2410-9916-2025-2-001-017

тельное хранение очень больших объемов информации с использованием автоматизированной системы самообслуживания для облачных хранилищ, привлекаемая для этих целей последние достижения в области технологий сохранения данных, технологий обеспечения информационной и промышленной безопасности, ускоренной разработки и развертывания новых приложений и передовых компьютерных технологий [1-3].

Системы облачного хранения данных, крупные облачные хранилища (ОХ), используют специальные программные средства (включая операционные системы) и являются важнейшими элементами современной информационно-телекоммуникационной архитектуры как государственного, ведомственного, так и мирового единого информационного пространства [4, 5].

Крупные распределенные ОХ позволяют не только каждому конкретному пользователю (ведомству, департаменту, фирме) экономить ресурсы, но и делают доступными для них ряд новых уникальных услуг, которые могут предоставляться временно или на постоянной основе [3, 4].

Это, в свою очередь, позволяет использовать ОХ для реализации целевых национальных программ, например, с помощью защищенных «облаков» планируется архитектура «электронного правительства», «электронного образования» и т.п. Как ожидается, государственная единая облачная платформа («Гос-Облако») будет запущена в эксплуатацию с января 2025 года.

При этом, к крупным ОХ, к «облакам», способным хранить и обрабатывать на самом деле громадные объемы информации, действительно Большие Данные, на наш взгляд, следует отнести «Яндекс Диск» (до 3 Тбайт), «Облако Mail.ru» (до 4 Тбайт), Vlomp (до 10 Тбайт), Koofr (до 20 Тбайт) и iCloud (до 12 Тбайт) [6, 7].

Особую значимость при работе с крупными ОХ приобретают вопросы организации защиты информации и разграничения доступа к ней, но эти задачи успешно решаются отечественными и зарубежными учеными и практиками, как на уровне организационном, так и в программно-технической сфере реализации хранения данных и удовлетворения поисковых запросов пользователей [1, 8-10].

Вместе с тем, наряду с вопросами безопасности, остаются проблемными вопросы обеспечения качества (оперативности и релевантности) информационного поиска на ресурсах крупных ОХ, особенно, с учетом нечеткой формулировки (идентификации) критериев качества реализации поисковых запросов и нечеткости (лингвистической неопределенности) предпочтений пользователей.

Поисковый информационный запрос представляет собой формализованную, но не всегда конкретную, однозначную и четко сформулированную заявку (просьбу) пользователя на поиск данных, удовлетворяющих его насущную информационную потребность. Это вид запроса, с помощью которого пользователи ищут информацию по интересующей их теме, цель такого запроса – получение знаний или решения проблемы.

Еще одна проблема состоит в том, что запрашиваемая информация, в которой нуждаются пользователи, находится на территориально распределенных серверах ОХ, среди множества иных, рассредоточенных в пространстве, зача-

стую Больших Данных, и оперативный поиск релевантных ответов на запросы становится нетривиальной задачей.

Предметная особенность задачи, обуславливающая ее актуальность, заключается в том, что в современном цифровом мире ОХ стали неотъемлемой частью жизни общества и государства. Отдельные люди и сообщества, организации, департаменты, министерства, вузы, ритейлеры и промышленные предприятия хранят в облаке все: от фотографий и документов до музыки и видео. Именно поэтому найти нужную информацию или файл в этом безграничном пространстве Больших Данных может быть непросто, тем более, при нечеткой формулировке критериев качества поиска и предпочтений пользователей.

Именно поэтому продолжает оставаться востребованной задача оперативного и релевантного поиска на больших информационных ресурсах распределенных ОХ. Подобные хранилища, несмотря на то, что имеют специальные теги и метки, каталоги и словари для Больших Данных, несмотря на то, что имеют свои дополнительные эффективные средства (сервисы) навигации, фильтрации и сортировки, позволяющие сузить область поиска, в современном цифровом мире, с его лавинообразным ростом объемов информации, представляют собой сложный «черный ящик».

К этому можно добавить тот факт, что в условиях ограниченного времени на реализацию поисковых запросов в ОХ, часто имеет место некоторая нечеткость критериев поиска, нечеткость границ охвата поиском и даже нечеткость конкретных целей (пожеланий и предпочтений) пользователей при формулировке поисковых запросов.

Объективно существующая нечеткость разнородных критериев и показателей качества информационного поиска на ресурсах крупных ОХ, необходимость их компромиссного согласования в рамках конкретного поискового запроса, потребность в реализации поиска с учетом важности «весов» этих критериев (показателей) в рамках иерархии предпочтений пользователей, а также потребность в разграничении по очередности и значимости возможных (слабо формализуемых, нечетких) альтернатив в процедурах и механизмах информационного поиска, делает данную задачу актуальной, хотя и, безусловно, сложно решаемой [11-14].

Известны различные отечественные и зарубежные подходы к решению задач информационного поиска, как при однозначно сформулированных запросах, так и в условиях нечеткости критериев и предпочтений поиска. Большинство этих подходов опирается на классические, традиционные методологические механизмы обработки запросов и на математические инструменты обработки нечетких множеств в интересах оперативного и релевантного поиска, включая интеллектуальные инструменты, хорошо описанные в [11, 15-19].

Однако практическая реализация подобных традиционных подходов к обработке нечетких данных в интересах информационного поиска затруднена, поскольку реальные задачи реализации поисковых запросов на Больших Данных с учетом нечеткой формулировки критериев качества и предпочтений пользователей требуют предварительного сбора и предобработки большое объ-

ема статистики для выработки правдоподобных значений функций принадлежности нечетких множеств.

Интересен ряд работ, непосредственно, напрямую посвященных проблемам поиска информации методами нечетких множеств [20-22]. В этих работах осуществляется последовательное, поэтапное приведение нечетких данных (критериев и условий поиска) к четкому, однозначному, количественному виду. Простота и практическая реализуемость подобных подходов обусловили их активную разработку и применение в области информационного поиска, но подобные методы не ориентированы на необходимость учета многокритериального, а иногда, и противоречивого, характера запросов и предпочтений пользователей, особенно на больших массивах данных (на Больших Данных), что присуще современным облачным хранилищам.

В этой связи перспективным, по нашему мнению, направлением совершенствования методов и средств анализа и повышения качества информационного поиска в ОХ, является разработка адекватных и эффективных аналитических моделей процесса смены состояний показателей качества реализации поисковых запросов пользователей с учетом нечеткой формулировки критериев качества и предпочтений этих пользователей.

Аналитические модели, наряду со статистическими и имитационными, относятся к математическим моделям, причем аналитические модели представляют собой отчетливые выражения конкретных параметров (например, параметров качества поиска), как функции начальных и переменных состояния (например, в нашем случае – переменных состояния показателей качества поиска) [23]. Обычно они имеют вид дифференциально-алгебраических либо разностных уравнений, описывающих явные зависимости с помощью конкретных формул и неравенств [23, 24]. Марковские модели процесса смены состояний показателей качества, по нашему мнению, принадлежат именно к классу аналитических моделей.

Данные модели, на наш взгляд, позволят аналитически корректно расчлени процесс информационного поиска на элементарные подпроцессы, выделить все действующие на процедуру смены состояний показателей качества поиска, внутренние и внешние факторы с учетом нечеткости задач (предпочтений пользователей) и критериев поиска.

Это, в свою очередь, позволит рассмотреть и детально проанализировать возможные режимы информационного поиска в крупных ОХ, выбрать оптимальные управляющие воздействия, составить объективный прогноз по повышению качества реализации поисковых запросов.

Модель информационного поиска с учетом нечеткой формулировки критериев качества и предпочтений пользователей

Другими словами, представляется целесообразным, с точки зрения теории и практики повышения качества информационного поиска в крупных распределенных ОХ, устранить противоречие между требованиями к современным

аналитическим моделям реализации поисковых запросов и узким спектром возможностей современных формальных моделей для математического описания поисковых систем с учетом нечеткой формулировки критериев качества поиска и нечеткой дефиниции поисковых предпочтений пользователей систем такого класса.

Практика показывает, что учет ряда объективных ограничений на уровни и параметры потенциальной модели, приводит к необходимости поиска достаточно строгого математического описания процесса информационного поиска в динамике. Наиболее полно данные ограничения, по нашему мнению, могут быть приняты во внимание при формировании аналитических марковских моделей [25-29].

Известно, что марковские модели процессов обладают большой универсальностью, при этом сочетание теории марковских процессов с теорией переменных состояния и с теорией нечетких множеств открывает широкие возможности для исследования и детального, адекватного формального описания процесса смены состояний процесса реализации поисковых запросов с учетом нечеткой формулировки критериев качества и предпочтений пользователей.

Кроме того известно, что существует возможность построения математических, вероятностно-временных моделей смены состояний процесса информационного поиска, как аналитических моделей смены состояний показателей качества реализации поисковых запросов на основе марковских цепей (последовательностей). Это позволит добиться более высокой, по сравнению с существующей, степени адекватности модели данного процесса при сокращении размерности его математического описания и с учетом нечеткой формулировки критериев качества поиска и нечеткого представления поисковых предпочтений пользователей [30, 31].

Речь идет об аналитическом описании свойств систем и процессов на языке теории множеств и функциональных пространств, о векторе переменных состояния $\tilde{x}(k+1)$ показателей качества поиска, наблюдаемых через вектор выходных переменных, через вектор наблюдения за процессом (процессом смены состояний конкретного показателя качества поиска x). При этом под пространством состояний «показателя качества» понимается метрическое пространство, каждый элемент которого представляет собой конечномерный вектор переменных состояния этого показателя. Иными словами, процесс поиска можно описать, описывая процесс изменения состояний (значений) показателей его качества во времени. По сути, под состоянием показателя понимается факт нахождения его значений в данный момент в том или ином диапазоне значений (например, «отлично», «хорошо», «плохо», т.е. «значение соответствует требованиям», «на грани», «не соответствует»).

Таким образом можно утверждать, что для построения аналитической, вероятностно-временной модели смены состояний процесса информационного поиска (смены состояний показателей качества поиска), учитывающей динамический и вероятностный характер, нестационарность этого процесса и управления им, учитывающей нечеткие формулировки критериев качества и предпо-

чтений пользователей, применимы управляемые цепи Маркова с нечеткими переходами из состояния в состояние, описываемые в форме разностных стохастических уравнений [31].

Для разрывных марковских процессов такие модели опираются на формальное описание процедур реализации поисковых запросов в виде разностных стохастических уравнений, которые для дискретных моделей смены состояний показателей качества информационного поиска в форме управляемых марковских цепей с нечеткими переходами, имеют вид [31]:

$$\tilde{x}(k+1) = C^T(k+1) \bar{\Theta}(k+1); \quad (1)$$

$$\bar{\Theta}(k+1) = \tilde{\phi}^T(k+1, k, u) \bar{\Theta}(k) + \Delta \bar{\Theta}(k+1); \quad (2)$$

$$\bar{z}(k+1) = H(k, x(k)) \bar{\Theta}(k+1) + \bar{\omega}(k+1), \quad (3)$$

где равенство (1) – уравнение состояния конкретного показателя качества информационного поиска на $(k+1)$ -м шаге реализации поискового запроса, в котором: $\tilde{x}(k+1)$ – вектор-столбец значений конкретного показателя качества информационного поиска x на $(k+1)$ -м шаге, в котором все элементы кроме одного (реального значения показателя качества информационного поиска x на данном шаге) равны 0; $C^T(k+1)$ – транспонированная диагональная квадратная матрица (порядка m) возможных значений показателя качества информационного поиска x на $(k+1)$ -м шаге реализации поискового запроса, причем число m (строк и столбцов) зависит от выбранного числа состояний (глубины моделирования); $\bar{\Theta}(k+1)$ – вспомогательный вектор-столбец индикаторов состояния показателя качества информационного поиска x на $(k+1)$ -м шаге, вводимый для удобства записи динамики перехода показателя из состояния в состояние. Равенство (2) – уравнение состояния вспомогательного вектора индикаторов, в котором: $\bar{\Theta}(k)$ – вектор-столбец значений индикаторов состояния показателя качества информационного поиска x на предыдущем шаге; $\Delta \bar{\Theta}(k+1)$ – вектор-столбец компенсационных добавок (приращений) индикаторов состояния, элементы которого предназначены для компенсации нецелочисленной части равенства (2) и получены в результате коррекции исходного шума возбуждения с математическим ожиданием и дисперсией, соответствующими начальному состоянию моделируемого процесса реализации поисковых запросов [31].

При этом необходимо детализировать важный отличительный аспект данной модели – ее способность учитывать нечеткую формулировку критериев качества и предпочтений пользователей. Например, при формулировке должностными лицами в рамках специальной военной операции нечеткого запроса типа: «найти цифровую карту Мариуполя», где не до конца понятно какую карту – туристическую, с номерами домов, с высотами, крупномасштабную (как для спортивного ориентирования) или иную. В этом и иных подобных случаях, чтобы система поиска не выдала сотню ненужных пользователю ссылок, необходимо уточнить его предпочтения.

С точки зрения математики, необходимо четко и однозначно идентифицировать (верифицировать) значения вероятностей переходов $\tilde{\phi}^T(k+1, k, u)$ из

выражения (2), позволяя, тем самым, корректно формализовать элементы модели, нивелировать нечеткость разнородных критериев и показателей качества информационного поиска на ресурсах крупных ОХ, обеспечить их компромиссное согласование в рамках конкретного поискового запроса, осуществить описание процесса поиска с учетом важности «весов» этих критериев (показателей) в рамках иерархии предпочтений пользователей, а также разграничить по очередности и значимости возможные (слабо формализуемые, нечеткие) альтернативы в процедурах и механизмах информационного поиска.

Таким образом, научную новизну и практическую значимость данной аналитической марковской модели информационного поиска с нечеткими переходами из состояния в состояние придает именно элемент $\tilde{\phi}^T(k+1, k, u)$ в равенстве (2) – транспонированная квадратная матрица (порядка m) вероятностей перехода процесса реализации поисковых запросов, обуславливающего смену состояний конкретного показателя качества информационного поиска x на $(k+1)$ -м шаге:

$$\tilde{\phi}^T(k+1, k, u) = \begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} & \cdots & \tilde{p}_{1m} \\ \tilde{p}_{21} & \tilde{p}_{22} & \cdots & \tilde{p}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_{m1} & \tilde{p}_{m2} & \cdots & \tilde{p}_{mm} \end{pmatrix}. \quad (4)$$

Эта матрица нечетко заданных переходных вероятностей, например, для трех состояний ($m=3$) имеет вид:

$$\tilde{\phi}^T(k+1, k, u) = \begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} & \tilde{p}_{13} \\ \tilde{p}_{21} & \tilde{p}_{22} & \tilde{p}_{23} \\ \tilde{p}_{31} & \tilde{p}_{32} & \tilde{p}_{33} \end{pmatrix}, \quad (5)$$

где каждый из элементов – нечетко заданное значение вероятности перехода конкретного показателя качества информационного поиска x из состояния в состояние, обусловленное нечеткими знаниями экспертов об этих значениях, что, в свою очередь, обусловлено нечеткой формулировкой критериев качества и предпочтений пользователей.

Каждое нечетко заданное значение вероятности перехода конкретного показателя качества информационного поиска x из состояния в состояние в рамках данной модели, может и должно быть четко идентифицировано (верифицировано) с использованием известных алгоритмов принятия решений, предполагающих построение множества недоминирующих альтернатив на базе нечетких отношений предпочтения. Эти алгоритмы известны, исходные данные для их работы и последовательность решения подобных задач, ориентированных на поиск компромиссов при принятии решений, связанных с несоизмеримыми и противоречащими друг другу критериями (показателями) качества и целями оценки качества информационного поиска, детально описаны в работах [12, 13].

С точки зрения физического смысла этой математической операции, данный подход к моделированию позволяет четко и однозначно идентифицировать

(верифицировать) значения вероятностей переходов $\tilde{\phi}^T(k+1, k, u)$, позволяя, тем самым, корректно формализовать элементы модели, нивелировать нечеткость разнородных критериев и показателей качества информационного поиска на ресурсах крупных ОХ, обеспечить их компромиссное согласование в рамках конкретного поискового запроса, осуществить описание процесса поиска с учетом важности «весов» этих критериев (показателей) в рамках иерархии предпочтений пользователей, а также разграничить по очередности и значимости возможные (слабо формализуемые, нечеткие) альтернативы в процедурах и механизмах информационного поиска.

С учетом этого, получение текущих значений индикаторов состояния показателя качества информационного поиска x на $(k+1)$ -м шаге реализации поискового запроса осуществляется путем линейной процедуры [31, 32]:

$$\|\Delta\Theta(k+1)\| = A - \|\Theta^{\text{мн}}(k+1)\|; \quad (6)$$

$$\Delta\vec{\Theta}(k+1) = \|\Delta\Theta(k+1)\| \vec{\Theta}^{\text{пр}}(k+1); \quad (7)$$

$$\vec{\Theta}(k+1) = \vec{\Theta}^{\text{мн}}(k+1) + \Delta\vec{\Theta}(k+1), \quad (8)$$

где равенство (6) описывает механизм получения элементов m -мерной матрицы компенсационных добавок $\|\Delta\Theta(k+1)\|$ и содержит элементы: A – m -мерная единичная диагональная матрица; $\|\Theta^{\text{мн}}(k+1)\|$ – m -мерная матрица, столбцами которой являются вектора нецелочисленных значений индикаторов $\vec{\Theta}^{\text{мн}}(k+1)$, повторенные m раз. Равенство (7) описывает порядок выбора вектора компенсационных добавок $\Delta\vec{\Theta}(k+1)$ из матрицы компенсационных добавок $\|\Delta\Theta(k+1)\|$ с помощью вектора $\vec{\Theta}^{\text{пр}}(k+1)$ предварительных значений индикаторов состояний моделируемого процесса информационного поиска. Равенство (8) завершает алгоритм вычислений вектора индикаторов состояний $\vec{\Theta}(k+1)$ и является формальной унифицированной записью уравнения состояния моделируемого процесса информационного поиска, причем, элементы вектора нецелочисленных значений индикаторов состояния, находятся в соответствии с модифицированным, с учетом рассмотренных ранее нечетко заданных значений вероятности перехода, равенством:

$$\vec{\Theta}^{\text{мн}}(k+1) = \tilde{\phi}^T(k+1, k, u) \vec{\Theta}(k). \quad (9)$$

Данное равенство является модификацией известного уравнения Колмогорова-Чепмена с учетом нечеткой формулировки критериев качества и предпочтений пользователей, выражающихся в нечетко заданных вероятностях переходов $\tilde{\phi}^T(k+1, k, u)$ и получено в результате преобразования выражения (2) [31, 32].

При этом в равенстве (9): $\vec{\Theta}^{\text{мн}}(k+1)$ – вектор нецелочисленных значений индикаторов состояния моделируемого процесса (процесса смены состояний показателей качества реализации поисковых запросов), имеющий математический смысл пошаговых и, в конечном итоге, финальных нечетко заданных ве-

роятностей нахождения случайного процесса в том или ином состоянии; $\vec{\Theta}(k)$ – вектор целочисленных значений индикаторов состояния моделируемого процесса на предыдущем шаге.

Равенство (9) для отдельного i -го (из m возможных) нецелочисленного индикатора состояния $\Theta_i^{\text{мн}}(k+1)$ в общем случае имеет вид:

$$\begin{aligned} \Theta_i^{\text{мн}}(k+1) = & \Theta_1(k)(\tilde{p}_{i1} - \tilde{p}_{im}) + \Theta_2(k)(\tilde{p}_{i2} - \tilde{p}_{im}) + \dots \\ & \dots + \Theta_{m-1}(k)(\tilde{p}_{im-1} - \tilde{p}_{im}) + \tilde{p}_{im}, \end{aligned} \quad (10)$$

где $i = 1, 2, 3, \dots, m$, а $\tilde{p}_{i1} \dots \tilde{p}_{im}$ – нечетко заданные и нуждающиеся в идентификации (верификации) элементы матрицы $\vec{\phi}^T(k+1, k, u)$ вероятностей перехода процесса из одного состояния в другое.

Равенство (3) в системе уравнений (1)-(3), описывающих смену состояний процесса информационного поиска на $(k+1)$ -м шаге реализации поискового запроса, является уравнением наблюдения за процессом (процессом смены состояний конкретного показателя качества поиска x), где $H(k, \vec{x}(k))$ – диагональная квадратная матрица (порядка m) наблюдаемых значений процесса, а $\vec{\omega}(k+1)$ – вектор белых шумов наблюдения с нулевым средним и матрицей дисперсии $\delta_{\omega}(k+1)$.

Таким образом, выражениями (1)-(10) представлена аналитическая модель (в терминах управляемых марковских цепей с нечеткими переходами в форме разностных стохастических уравнений), математически корректно описывающая динамику изменения состояний показателей качества информационного поиска $\vec{x}(k+1)$ на $(k+1)$ -м шаге реализации поисковых запросов и учитывающая нечеткую формулировку критериев качества и предпочтений пользователей.

При этом в роли конкретных показателей качества x информационного поиска на больших ОХ, в роли элементов вектора $\vec{x}(k+1)$ могут выступать: x_1 – оперативность (время реализации поисковых запросов, временной интервал поиска); оперативность (время реализации поисковых запросов, временной интервал поиска); x_2 – уровень соответствия данного запроса требованиям пользователя – релевантность; x_3 – физический размер (объем) данных, которые необходимо найти; x_4 – уровень достоверности (адекватность, точность) результатов поиска [13].

Если обозначить x_1 – оперативность (время реализации поисковых запросов, временной интервал поиска) через ее эквивалент в виде численного показателя качества, например, $(\Delta \bar{t}_{\text{рпз}}(k))$ – среднее время реализации поисковых запросов, можно интерпретировать форму записи модели смены состояний данного показателя качества в виде:

$$\vec{t}_{\text{рпз}}^{\sim}(k+1) = C_{\vec{t}_{\text{рпз}}}^T(k+1) \vec{\Theta}_{\vec{t}_{\text{рпз}}}(k+1); \quad (11)$$

$$\vec{\Theta}_{\vec{t}_{\text{рпз}}}(k+1) = \vec{\phi}_{\vec{t}_{\text{рпз}}}^T(k+1, k, u) \vec{\Theta}_{\vec{t}_{\text{рпз}}}(k) + \Delta \vec{\Theta}_{\vec{t}_{\text{рпз}}}(k+1); \quad (12)$$

$$\vec{z}_{\text{рпз}}^-(k+1) = \mathbf{H}_{\text{рпз}}^-(k, \vec{x}(k)) \vec{\Theta}_{\text{рпз}}^-(k+1) + \vec{\omega}_{\text{рпз}}^-(k+1), \quad (13)$$

где $\vec{\phi}_{\text{рпз}}^T(k+1, k, u)$ в равенстве (12) – матрица нечетко заданных значений вероятностей перехода этого конкретного показателя качества информационного поиска ($\Delta \vec{t}_{\text{рпз}}^-(k)$) – среднего времени реализации поисковых запросов, из состояния в состояние. Данный вид неопределенности в предложенной модели обусловлен нечеткими знаниями экспертов об этих значениях, что связано с нечеткой формулировкой критериев качества и предпочтений пользователей. Как и показано при описании равенств (4) и (5), эти нечетко заданные значения вероятности перехода в рамках данной модели, могут быть заранее четко идентифицированы (верифицированы) с использованием известных алгоритмов принятия решений, предполагающих построение множества недоминирующих альтернатив на базе нечетких отношений предпочтения.

Все остальные элементы равенств (11)-(13) имеют ранее описанный в работе физический смысл.

Ориентируясь на структуру системы показателей качества, характеризующих процессы реализации поисковых запросов в больших ОХ, с учетом описанного аналитического марковского аппарата моделирования (1)-(10), а также того факта, что вероятностно-временная модель смены состояний показателей качества информационного поиска изначально базируется на детерминированных переменных, следующим шагом может стать формулировка идеи формирования структуры (архитектуры) и построение обобщенного программно-аппаратного прототипа для практического моделирования процесса информационного поиска в облачных хранилищах с учетом нечеткой формулировки критериев качества и предпочтений пользователей.

Выводы

Рассмотренный подход к моделированию может быть использован для решения задач построения адекватных и эффективных аналитических моделей процесса смены состояний показателей качества реализации поисковых запросов пользователей в больших ОХ с учетом нечеткой формулировки критериев качества поиска и предпочтений этих пользователей. Решение подобных задач часто сталкивается с проблемой необходимости учета нечеткости критериев качества поиска на больших объемах данных, нечеткости границ охвата поиском и даже нечеткости конкретных целей пользователей при формулировке поисковых запросов.

Предложен, по сути, универсальный механизм аналитического моделирования, основанный на использовании вероятностно-временной модели смены состояний процесса информационного поиска (смены состояний показателей качества поиска), учитывающей динамический и вероятностный характер, нестационарность этого процесса и управления им, учитывающей нечеткие формулировки критериев качества и предпочтений пользователей.

Предложенный подход к моделированию основан на применении управляемых цепей Маркова с нечеткими переходами из состояния в состояние, описываемых в форме разностных стохастических уравнений. Данный подход позволяет не только корректно формализовать элементы модели, нивелировать нечеткость разнородных критериев и показателей качества информационного поиска на ресурсах крупных ОХ, но и обеспечить их компромиссное согласование в рамках конкретного поискового запроса, осуществить моделирование процесса поиска с учетом важности «весов» этих критериев (показателей) в рамках иерархии предпочтений пользователей, а также разграничить по очередности и значимости возможные (зачастую, слабо формализуемые, нечеткие) альтернативы в процедурах и механизмах информационного поиска.

Литература

1. Bokefod J. D., Bhiz A. S., Satarkar P. A., Modani D. G. Development of a secure cloud storage system for IoT data storage using role-based encryption // *Procedia Computer Science*. 2016. Vol. 89. P. 43-50.
2. Аньель Х., Монтес Д., Родейро Иглесиа Х. Переход в облако. Практическое руководство по организации облачных вычислений для ученых и ИТ-специалистов. М.: Альпина PRO, 2022. 138 с.
3. Орбан С. Путешествие к облаку. Советы по использованию облачных технологий от лидеров ИТ-рынка. М.: Альпина PRO, 2022. 293 с.
4. Васяткин М. А., Белоус К. В. Облачное хранилище данных // *StudNet*. 2020. № 10. С. 1-12.
5. Маркелов А. А. OpenStac. Практическое знакомство с облачной операционной системой. М.: ДМК-Пресс, 2018. 306 с.
6. Игнатьева А. В. 20 облачных хранилищ: платные и бесплатные сервисы для хранения файлов и обмена ими // *Skillbox Media* [Электронный ресурс]. 16.04.2024. – URL: <https://skillbox.ru/media/management/20-oblachnykh-khranilishch-platnye-i-besplatnye-servisy-dlya-khraneniya-faylov-i-obmena-imi/?ysclid=m56ehsrqpw8552> (дата обращения: 27.12.2024).
7. Кодолов П. А. Облачное хранилище данных // *Наука, техника и образование*. 2016. № 4 (22). С. 51-53.
8. Паращук И. Б., Саенко И. Б., Пантюхин О. И. Доверенные системы для разграничения доступа к информации в облачных инфраструктурах // *Научные технологии в космических исследованиях Земли*. 2018. Т. 10. № 6. С. 68-75.
9. Дроздова И. И., Жилин В. В. Безопасность облачных хранилищ // *Технические науки в России и за рубежом: материалы VII Международной научной конференции* (г. Москва, ноябрь 2017 года). М.: Буки-Веди, 2017. С. 16-18.
10. Зияев П. В., Паращук И. Б., Ткаченко В. В. Анализ тенденций применения и вопросы обеспечения информационной безопасности систем хранения данных для облачных технологий // *Региональная информатика и*

информационная безопасность. Сборник трудов. Выпуск 4. СПб.: СПОИСУ, 2017. С. 93-95.

11. Целых А. Н., Котов Э. М., Целых А. А. Метод информационного поиска на основе нечеткого сходства ситуаций // Известия Южного федерального университета. Технические науки. 2014. № 6 (155). С. 74-78.

12. Винокуров А. С., Белов И. В., Баженов Р. И. Использование метода нечеткого отношения предпочтения для принятия оптимального решения по выбору цифрового фотоаппарата // Современная техника и технологии. 2014. № 11. – URL: <https://technology.snauka.ru/2014/11/4868> (дата обращения: 12.03.2025).

13. Паращук И. Б., Михайличенко А. В., Саяркин Л. А. Реализация поисковых информационных запросов в крупных дата-центрах с использованием математических методов теории нечетких множеств // Системы синхронизации, формирования и обработки сигналов. 2024. Т. 15. № 3. С. 20-27.

14. Пронина В. А., Панкова Л. А. Семантический текстовый поиск, основанный на теории нечетких множеств // Искусственный интеллект и принятие решений. 2013. № 3. С. 19-23.

15. Урвачева В. А. Обзор методов информационного поиска // Вестник Таганрогского государственного педагогического института имени А. П. Чехова. 2016. № 2. С. 457-463.

16. Солонин Е. Б. Интеллектуальные технологии поиска и анализа данных. Екатеринбург: Уральский федеральный университет, 2015. 43 с.

17. Manning C. D., Raghavan P., Schütze H. An Introduction to Information Retrieval. Cambridge: Cambridge University Press, 2009. 581 p.

18. Buttcher S., Clarke C. L. A., Cormack G. V. Information Retrieval: Implementing and Evaluating Search Engines. Cambridge: MIT Press, 2016. 632 p.

19. Afebende G. B. Information search strategies as correlate of information resource utilization in academic libraries // International Journal of Library and Information Science Studies. 2019. Vol. 5. No. 5. P. 34-42.

20. Рыжов А. П. Модели поиска информации методами нечетких множеств. М.: МГУ им. М. В. Ломоносова, 2004. 95 с.

21. Kraft D., Colvin E. Fuzzy Information Retrieval // Synthesis Lectures on Information Concepts Retrieval and Services. 2017. № 9 (1). P. 46-63.

22. Cross V. Fuzzy information retrieval // Journal of Intelligent Information Systems. 1994. Vol. 3. P. 29-56.

23. Боев В. Д. Моделирование в AnyLogic. Учебное пособие для вузов. М.: Юрайт. 2018. 298 с.

24. Сушков Ю. А. Аналитические модели систем. Учебное пособие. СПб.: Санкт-Петербургский государственный университет, 2016. 67 с.

25. Зейфман А. И. Марковские цепи и модели с непрерывным временем. М.: Элекс-КМ. 2008. 168 с.

26. Артеменков С. Л., Алхимов В. И., Баранов С. Н., Беляева О. Б., Думин П. Н., Корниенко П. А., Куравский Л. С., Малых С. Б., Марголис А. А., Мармалюк П. А., Панфилова А. С., Попков С. И., Юрьев Г. А., Юрьева Н. Е.

Марковские модели в задачах диагностики и прогнозирования. М.: Московский государственный психолого-педагогический университет, 2017. 197 с.

27. Галажинская О. Н., Моисеева С. П. Теория случайных процессов: учебное пособие. Часть 2: Марковские процессы. Томск: Издательский Дом Томского государственного университета, 2016. 126 с.

28. Паращук И. Б., Ковальченко Д. А. К вопросу математического описания динамики изменения показателей качества телекоммуникационной системы марковскими последовательностями // Проблемы автоматизации. Региональное управление. Связь и автоматика («ПАРУСА-2014»). Сборник трудов III Всероссийской научной конференции молодых ученых, аспирантов и студентов. (20-21 ноября 2014 г. Геленджик). Таганрог: ЮФУ, 2014. Т. 1. С. 164-167.

29. Саенко И. Б., Паращук И. Б., Кондрашов Ю. В., Лаута О. С., Домбровский Я. А., Попов А. И., Крюкова Е. С. Моделирование и проектирование систем: учебник. СПб.: ВАС, 2023. 472 с.

30. Матвеев М. Г., Алейникова Н. А., Громковский А. А. Дискретная однородная цепь Маркова для нечетких состояний // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2022. № 4. С. 119-131.

31. Паращук И. Б., Бобрик И. П. Нечеткие множества в задачах анализа сетей связи. СПб.: ВУС, 2001. 80 с.

32. Паращук И. Б. Моделирование процесса отклонений показателей качества сети многоканальной радиосвязи от требуемых значений // Радиотехника. 1998. № 11. С. 6-9.

References

1. Bokefod J. D., Bhiz A. S., Satarkar P. A., Modani D. G. Development of a secure cloud storage system for IoT data storage using role-based encryption. *Procedia Computer Science*, 2016, vol. 89, pp. 43-50.

2. Agnel H., Montes D., Rodeiro Iglesia X. *Perekhod v oblako. Prakticheskoe rukovodstvo po organizatsii oblachnikh vi-chislenii dlya uchenikh i IT-spetsialistov* [Switching to the cloud. A practical guide to organizing cloud computing for scientists and IT specialists]. Moscow, Alpina PRO, 2022. 138 p. (in Russian).

3. Orban S. *Puteshestvie k oblaku. Soveti po ispolzovaniyu oblachnikh tekhnologii ot liderov IT-rinka* [Journey to the cloud. Tips on using cloud technologies from the leaders of the IT market]. Moscow, Alpina PRO, 2022. 293 p. (in Russian).

4. Vasyatkin M. A., Belous K. V. Oblachnoe khranilishche dannikh [Cloud data storage]. *StudNet*, 2020, no. 10, pp. 1-12 (in Russian).

5. Markelov A. A. *OpenStac. Prakticheskoe znakomstvo s oblachnoi operatsionnoi sistemoi* [OpenStac. Practical introduction to the cloud operating system]. Moscow, DMK-Press, 2018. 306 p. (in Russian).

6. Ignatieva A. V. *20 oblachnikh khranilishch: platnie i besplatnie servisi dlya khraneniya failov i obmena imi* [20 cloud storages: paid and free services for storing

and sharing files]. *Skillbox Media*, 16 April 2024. Available at: <https://skillbox.ru/media/management/20-oblachnykh-khranilishch-platnye-i-besplatnye-servisy-dlya-khraneniya-faylov-i-obmena-imi/?ysclid=m56ehsrqpw8552> (accessed 27 December 2024) (in Russian).

7. Kodolov P. A. Oblachnoe khranilishche dannikh [Cloud data storage]. *Science, Technology and Education*, 2016, vol. 22, no. 4, pp. 51-53 (in Russian).

8. Parashchuk I. B., Saenko I. B., Pantyukhin O. I. Trusted systems for access control to information in cloud infrastructures. *H&ES Research*, 2018, vol. 10, no. 6, pp. 68-75 (in Russian).

9. Drozdova I. I., Zhilin V. V. Bezopasnost oblacloudnykh khranilishch [Security of cloud storage]. *Tekhnicheskie nauki v Rossii i za rubezhom. Materiali 7 Mezhdunarodnoi nauchnoi konferentsii* [Technical sciences in Russia and abroad. Proceedings of the 7th International Scientific Conference]. Moscow, Buki-Vedi, 2017, pp. 16-18 (in Russian).

10. Ziyaev P. V., Parashchuk I. B., Tkachenko V. V. Analiz tendentsii primeneniya i voprosi obespecheniya informatsionnoi bezopasnosti sistem khraneniya dannikh dlya oblacloudnykh tekhnologii [Analysis of application trends and issues of ensuring information security of data storage systems for cloud technologies]. *Regional'naya informatika i informatsionnaya bezopasnost. Sbornik trudov. Vipusk 4.* [Regional informatics and information security. Collection of works. Issue 4th]. Saint-Petersburg, Saint-Petersburg Society of Computer Science, Computer Technology, Communication Systems and Management, 2017, pp. 93-95 (in Russian).

11. Tselykh A. N., Kotov E. M., Tselykh A. A. Metod informatsionnogo poiska na osnove nechetkogo skhodstva situatsii [Information retrieval method based on fuzzy similarity of situations]. *Izvestiya SFedU. Engineering Sciences*, 2014, no. 6 (155), pp. 74-78 (in Russian).

12. Vinokurov A. S., Belov I. V., Bazhenov R. I. The use of fuzzy preference relations to make better decisions on choosing a digital camera. *Modern technics and technologies*. 2014, no. 11. Available at: <https://technology.snauka.ru/2014/11/4868> (accessed 12 March 2025) (in Russian).

13. Parashchuk I. B., Mikhailichenko A. V., Sayarkin L. A. Realizatsiya poiskovykh informatsionnykh zaprosov v krupnykh data-tsentrakh s ispolzovaniem matematicheskikh metodov teorii nechetkikh mnozhestv [Implementation of search information queries in large data centers using mathematical methods of fuzzy set theory]. *Sistemy sinhronizatsii, formirovaniya i obrabotki signalov*, 2024, vol. 15, no. 3, pp. 20-27 (in Russian).

14. Pronina V. A., Pankova L. A. Semantic text retrieval based on fuzzy set theory. *Artificial Intelligence and Decision Making*, 2013, no. 3. pp. 19-23 (in Russian).

15. Urvacheva V. A. Obzor metodov informatsionnogo poiska [Review of information retrieval methods]. *Bulletin of the Taganrog State Pedagogical Institute named after A. P. Chekhov*, Taganrog, 2016, no. 2, pp. 457-463 (in Russian).

16. Solonin E. B. *Intellektualnie tekhnologii poiska i analiza dannikh* [Intelligent technologies for data search and analysis]. Ekaterinburg, Ural Federal University, 2015. 43 p. (in Russian).
17. Manning C. D., Raghavan P., Schutze H. *An Introduction to Information Retrieval*. Cambridge, Cambridge University Press, 2009. 581 p.
18. Buttcher S., Clarke C. L. A., Cormack G. V. *Information Retrieval: Implementing and Evaluating Search Engines*. Cambridge, MIT Press, 2016. 632 p.
19. Afebende G. B. Information search strategies as correlate of information resource utilization in academic libraries. *International Journal of Library and Information Science Studies*, 2019, vol. 5, no. 5, pp. 34-42.
20. Ryzhov A. P. *Modeli poiska informatsii metodami nechetkikh mnozhestv* [Information retrieval models using fuzzy set methods]. Moscow, Lomonosov Moscow State University, 2004, 95 p. (in Russian).
21. Kraft D., Colvin E. Fuzzy Information Retrieval. *Synthesis Lectures on Information Concepts Retrieval and Services*, 2017, no. 9 (1), pp. 46-63.
22. Cross V. Fuzzy information retrieval. *Journal of Intelligent Information Systems*, 1994, vol. 3, pp. 29-56.
23. Boev V. D. *Modelirovanie v AnyLogic* [Modeling in AnyLogic]. Moscow, Yurait, 2018. 298 p. (in Russian).
24. Sushkov Yu. A. *Analiticheskie modeli sistem* [Analytical models of systems]. Saint-Petersburg, Saint-Petersburg University, 2016. 67 p. (in Russian).
25. Zeifman A. I. *Markovskie tsepi i modeli s neprerivnim vremenem* [Markov chains and models with continuous time]. Moscow, Eleks-KM, 2008. 168 p. (in Russian).
26. Artemenkov S. L., Alkhimov V. I., Baranov S. N., Belyaeva O. B., Dumin P. N., Kornienko P. A., Kuravsky L. S., Malykh S. B., Margolis A. A., Marmalyuk P. A., Panfilova A. S., Popkov S. I., Yuryev G. A., Yuryeva N. E. *Markovskie modeli v zadachakh diagnostiki i prognozirovaniya* [Markov models in diagnostic and forecasting problems]. Moscow, Moscow State University of Psychology and Education, 2017. 197 p. (in Russian).
27. Galazhinskaya O. N., Moiseeva S. P. *Teoriya sluchainikh protsessov. Chast 2: Markovskie protsessi* [Theory of random processes. Part 2: Markov processes]. Tomsk, Publishing House of Tomsk State University, 2016. 126 p. (in Russian).
28. Parashchuk I. B., Kovalchenko D. A. K voprosu matematicheskogo opisaniya dinamiki izmeneniya pokazatelei kachestva telekommunikatsionnoi sistemi Markovskimi posledovatel'nostyami [On the issue of mathematical description of the dynamics of changes in the quality indicators of a telecommunication system by Markov sequences]. *Problemi avtomatizatsii. Regionalnoe upravlenie. Svyaz i avtomatika ("PARUSA-2014")*. *Sbornik trudov 3 Vserossiiskoi nauchnoi konferentsii molodikh uchenikh, aspirantov i studentov* [Problems of automation. Regional management. Communications and automation (SAILS-2014). Proceedings of the 3rd All-Russian Scientific Conference of Young scientists, postgraduates and students]. Taganrog, Southern Federal University, 2014, vol. 1. pp. 164-167 (in Russian).

29. Saenko I. B., Parashchuk I. B., Kondrashov Yu. V., Lauta O. S., Dombrovsky Ya. A., Popov A. I., Kryukova E. S. *Modelirovanie i proektirovanie sistem* [Modeling and system design]. Saint-Petersburg, Military Academy of the Signal Corps, 2023. 472 p. (in Russian).

30. Matveev M. G., Aleynikova N. A., Gromkovsky A. A. Discrete homogeneous Markov chain for fuzzy states. *Proceedings of Voronezh State University. Series: System analysis and information Technologies*, 2022, no. 4, pp. 119-131 (in Russian).

31. Parashchuk I. B., Bobrik I. P. *Nechetkie mnozhestva v zadachakh analiza setei svyazi* [Fuzzy sets in problems of communication network analysis]. Saint-Petersburg, Military Academy of the Signal Corps, 2001. 80 p. (in Russian).

32. Parashchuk I. B. Modelirovanie protsessa otklonenii pokazatelei kachestva seti mnogokanalnoi radiosvyazi ot trebuemikh znachenii [Modeling the process of deviations of quality indicators of a multichannel radio network from the required values]. *Radiotekhnika*, 1998, no. 11, pp. 6-9 (in Russian).

Статья поступила 17 января 2025 г.

Информация об авторах

Паращук Игорь Борисович – доктор технических наук, профессор. Профессор кафедры автоматизированных систем специального назначения. Военная академия связи. Область научных интересов: информационный поиск; сбор и обработка информации. E-mail: shchuk@rambler.ru

Саяркин Леонид Андреевич – соискатель ученой степени кандидата технических наук. Адъюнкт кафедры автоматизированных систем специального назначения. Военная академия связи. Область научных интересов: информационный поиск; сбор и обработка информации. E-mail: leonid.sayarkin@ya.ru

Адрес: 194064, Россия, г. Санкт-Петербург, Тихорецкий пр., д. 3.

Analytical Markov model of information retrieval in cloud storages taking into account fuzzy formulation of quality criteria and user preferences

I. B. Parashchuk, L. A. Sayarkin

Purpose: implementation of adequate and effective analytical models of the process of changing the states of the quality indicators of the implementation of user search queries in large cloud storages, taking into account the fuzzy formulation of search quality criteria and preferences of these users. The aim of the work is to solve the problem of eliminating the discrepancy between the requirements for modern analytical models of search query implementation and the limited capabilities of existing formal models for the mathematical description of search systems, taking into account the fuzzy search quality criteria and the uncertainty of user preferences in systems of this class. **Methods:** the approach is focused on the use of controlled Markov chains in the form of difference stochastic equations in combination with mathematically correct algorithms for identifying fuzzy probabilities of transitions from state to state, due to fuzzy search quality

criteria and user preferences. **Results and novelty:** the elements of novelty of the presented model are the use of a universal mechanism of analytical modeling based on a probabilistic-temporal model of changing the states of the process of implementing search queries and the associated quality indicators. This mechanism takes into account the dynamic nature, probabilistic character and non-stationarity of the search process, takes into account fuzzy quality criteria and user preferences and neutralizes this type of uncertainty using the construction of a set of non-dominant alternatives based on fuzzy preference relations. **Practical relevance:** the modeling approach presented in the work can be applied to the actual development and practical construction of analytical models describing the process of changing the state of quality indicators for the execution of user search queries in large cloud storages, taking into account the fuzzy formulation of search quality criteria and user preferences.

Key words: information retrieval, fuzzy set, cloud storage, Markov model, quality criteria, user.

Information about Authors

Igor Borisovich Parashchuk – Dr. habil. of Engineering Sciences, Professor. Professor at the Department of Automated Special Purpose System. Military Academy of the Signal Corps. Field of research: information search; data acquisition. E-mail: shchuk@rambler.ru

Leonid Andreevich Sayarkin – Doctoral Student. The postgraduate student of the Department of Automated Special Purpose System. Military Academy of the Signal Corps. Field of research: information search; data acquisition. E-mail: leonid.sayarkin@ya.ru

Address: Russia, 194064, Saint-Petersburg, Tihoreckiy prospekt, 3.