

УДК 004

DOI: 10.30987/article\_5bd17b41214911.88208762

П.В. Пестерев, А.Г. Янишевская

## МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И АЛГОРИТМ ОПРЕДЕЛЕНИЯ ПОКАЗАТЕЛЯ СТРАНИЦЫ ДЛЯ РАНЖИРОВАНИЯ В РЕЗУЛЬТАТАХ ВЫДАЧИ МУЛЬТИАГЕНТНОЙ ПОИСКОВОЙ СИСТЕМЫ

Описываются разработанные математическая модель и алгоритм определения показателя страницы на основе пользовательских действий для ранжирования в результатах поисковой выдачи разработанной мультиагентной поисковой системы предприятия. Производится сравнение предложенной модели с классическим вариантом ранжирова-

ния результатов на основе только количества посещений.

**Ключевые слова:** поисковая система предприятия, мультиагентная система, результаты поисковой выдачи, факторы ранжирования, показатель отказов, показатель времени, показатель возвращения к повторному поиску, показатель возвращения не из поисковой системы.

P.V. Pesterev, A.G. Yanishevskaya

## SIMULATOR AND ALGORITHM FOR PAGE INDEX DEFINITION FOR RANKING IN RESULTS OF MULTI-AGENT SEARCH SYSTEM ISSUE

In the paper there is described a simulator and an algorithm for the page index definition on the basis of user actions for ranking in results of search issue of the company developed multi-agent search system. The application of the simulator and the algorithm offered allows making the systemized ranking of the results of the search issue that, in its turn, increases search quality and its speed. In the simulator offered for the definition of a page index value there are used such indices as an index of failures, an index of time spent on a page, an index of return to a repeated search, an index of return to a page not from a search system. All values obtained are used for the definition of a page position in the further search iterations on the basis of the algorithm of processing the results of a search issue the structural diagram of which is presented in the paper.

The process of page index definition on the basis of user actions with the use of the search system is described and presented in the paper as a structural diagram of an algorithm. Besides, in the paper on the basis of experimental investigations it is shown that the use of the simulator offered in comparison with the classical option of result ranking on the basis of the visits number only results in more reliable results of the search issue. Furthermore, the application of this simulator makes page ranking more efficient already after first search iterations.

**Key words:** company search system, multi-agent system, results of search issue, ranking factors, failure index, time index, index of return to repeated search, index of return from non-search systems.

### Введение

САПР – автоматизированная система, обеспечивающая реализацию информационной технологии выполнения функций проектирования [1] и представляющая собой организационно-техническую систему, состоящую из персонала и комплекса технических, программных и иных средств автоматизации процессов проектирования [2]. Любым предприятиям или организациям, внедряющим программно-аппаратный комплекс САПР [3; 4], необходимо обеспечить наличие компьютеров для реализации серверных приложений, сервисов и пользовательских данных. Эффективным решением в этом случае явля-

ется внедрение поисковой системы [5; 6], представляющей собой компьютерную систему поиска информации, взаимодействующую с пользователем посредством интерфейса. Поскольку такая система является сложной и состоит из большого количества программ, то эффективным является использование теории мультиагентных систем [7-12]. Таким образом, была разработана мультиагентная поисковая система предприятия [13]. По умолчанию последовательность страниц в результатах поисковой выдачи носит произвольный характер, что затрудняет поиск и увеличивает его продолжительность. Для того что-

бы последовательность страниц в результатах выдачи данной системы носила систематизированный характер, предлагается

### Постановка задачи

Поскольку можно сделать вывод, что систематизированное ранжирование результатов поисковой выдачи является актуальной задачей, решение которой позволит повысить качество поиска и его скорость, то предлагается разработать математическую модель и алгоритм определения показателя страницы для ранжирова-

### Теоретические основы

В основе математической модели определения показателя страницы для ранжирования в результатах выдачи мультиагентной поисковой системы лежат показатель отказов, показатель времени, проведенного на странице, показатель возвращения к повторному поиску и показатель возвращения на страницу не из поисковой системы.

Показатель отказов (bounce rate) – термин веб-аналитики, описывающий процентное соотношение количества посетителей сайта, покинувших его после посещения первой страницы [14]. В случае с поисковой системой предприятия, где все страницы в результатах выдачи представляют собой одностраничный документ, такое решение не подходит. Поэтому на каждой странице была добавлена строчка с чекбоксом, приведение которого в активное состояние определяет, что поиск завершен, а страница содержит необходимые пользователю данные. Соответственно если чекбокс остался в неактивном состоянии, то предполагается, что результат не соответствует тому, что искал пользователь. Таким образом, показатель отказов в разработанной мультиагентной поисковой системе предприятия представляет собой отношение количества активаций чекбокса к общему количеству посещений данной страницы через поисковую систему. Формула для расчета значения показателя отказов имеет следующий вид:

$$I_{br} = \frac{n_{sc}}{n_{ms}}, \quad (1)$$

каждой странице присваивать показатель на основе пользовательских действий при ее посещении.

ния в результатах поисковой выдачи мультиагентной поисковой системы. Также предлагается провести экспериментальные исследования для возможности сравнения предложенной модели с классическим вариантом ранжирования результатов на основе только количества посещений.

где  $n_{sc}$  – количество нажатий на чекбокс при посещении данной страницы, определяющее завершение поиска;  $n_{ms}$  – общее количество посещений страницы через поисковую систему.

Показатель времени, проведенного на странице, определяется на основе общего времени, проведенного пользователями на ней при переходе через поисковую систему, которое рассчитывается по формуле

$$t_{tt} = \sum_{i=1}^{n_{ms}} t_{op_i}, \quad (2)$$

где  $t_{op}$  – время посещения отдельным пользователем страницы с данными.

Показатель времени, проведенного на странице, представляет собой отношение общего времени пребывания пользователей на странице при переходе через поисковую систему к произведению времени принятия решения, равного 90 с, и общего количества посещений страницы через поисковую систему. Значение 90 с взято на основе анализа пользовательских действий. Пользователь уходит со страницы в случае несоответствия информации на ней той, которую он ищет, как правило, в первые 8 с. Положительное же решение принимается примерно в течение первой минуты или после посещения и анализа других страниц. Если же время, проведенное пользователем на странице, превысило 90 с, то высока вероятность того, что он оставил ее открытой, по факту прекратив ее посещение, и дальнейшее время в расчет не берется. Таким образом, формула для

расчета значения показателя времени, проведенного на странице, имеет следующий вид:

$$I_{at} = \frac{t_{tt}}{n_{ms} t_{dm}}, \quad (3)$$

где  $t_{tt}$  – общее время пребывания пользователей на странице с данными при переходе через поисковую систему, рассчитываемое по формуле (2);  $t_{dm} = 90$  с – время принятия решения.

Если в формулу (3) подставить формулу (2), то получим следующую формулу для расчета показателя времени, проведенного на странице:

$$I_{at} = \frac{\sum_{i=1}^{n_{ms}} t_{op_i}}{n_{ms} t_{dm}}. \quad (4)$$

Возвращение к повторному поиску в предложенной модели означает возврат со страницы в поисковую систему с последующим переходом из нее на другую страницу. Разница с показателем отказов заключается в том, что возвращение к повторному поиску означает, что клиент закрыл страницу и продолжил поиск, не найдя на странице нужной информации. В то же время неактивированный чекбокс может означать, что пользователь либо прекратил поиск вообще, либо отложил решение на потом. Для расчета показателя учитывается общее количество посещений страницы через поисковую систему и количество посещений через поисковую систему, после которых был продолжен поиск с последующим переходом на другую страницу. Таким образом, формула для расчета значения показателя возвращения к повторному поиску имеет следующий вид:

$$I_{sa} = \frac{n_{ms} - n_{cs}}{n_{ms}} = I - \frac{n_{cs}}{n_{ms}}, \quad (5)$$

где  $n_{cs}$  – количество посещений страницы, после которых был продолжен поиск с последующим переходом на другую страницу.

Возвращение на страницу не из поисковой системы может означать, что пользователь знает о ее содержимом, а

описанные на ней компонент, деталь или любая другая информация соответствуют тому, что он ищет. Возможно, пользователь посещал ее ранее и сохранил в закладках. Для расчета данного показателя учитывается количество всех входов на страницу и количество входов на нее только из поисковой системы. Формула для расчета значения показателя имеет следующий вид:

$$I_{vns} = \frac{n_{mv} - n_{ms}}{n_{mv}} = I - \frac{n_{ms}}{n_{mv}}, \quad (6)$$

где  $n_{mv}$  – общее количество посещений страницы пользователями.

Перерасчет всех значений, необходимых для определения значения показателя страницы, а также самого этого значения производится по завершении посещения страницы. Во время текущего поиска берутся значения страниц, рассчитанные ранее. Изначально всем показателям присвоено значение 0 (до первого посещения). Для определения значения показателя страницы производится суммирование значений всех описанных показателей (показателя отказов, показателя времени, проведенного на странице, показателя возвращения к повторному поиску и показателя возвращения на страницу не из поисковой системы). Таким образом, формула для расчета значения показателя страницы имеет следующий вид:

$$I_{rv} = I_{br} + I_{at} + I_{sa} + I_{vns}, \quad (7)$$

где  $I_{br}$  – показатель отказов, рассчитываемый по формуле (1);  $I_{at}$  – показатель времени, проведенного на странице, рассчитываемый по формуле (4);  $I_{sa}$  – показатель возвращения к повторному поиску, рассчитываемый по формуле (5);  $I_{vns}$  – показатель возвращения на страницу не из поисковой системы, рассчитываемый по формуле (6).

При подстановке в формулу (7) формул (1), (4), (5) и (6) получим следующую формулу для расчета значения показателя страницы:

$$I_{rv} = \frac{n_{sc}}{n_{ms}} + \frac{\sum_{i=1}^{n_{ms}} t_{opi}}{n_{ms} t_{dm}} + 1 - \frac{n_{cs}}{n_{ms}} + 1 - \frac{n_{ms}}{n_{mv}} = 2 + \frac{n_{sc} - n_{cs}}{n_{ms}} - \frac{n_{ms}}{n_{mv}} + \frac{\sum_{i=1}^{n_{ms}} t_{opi}}{n_{ms} t_{dm}}. \quad (8)$$

Чем выше значение показателя страницы, тем выше ее позиция в результатах поисковой выдачи. Страницы с одинаковыми значениями показателей ранжируются в результатах поисковой выдачи в произвольном порядке. Расчет новых значений производится на основе пользователь-

ских действий в результатах поисковой выдачи, а их перезапись в базу данных - после завершения посещения страницы. На рис. 1 представлена структурная схема алгоритма обработки результатов поисковой выдачи.

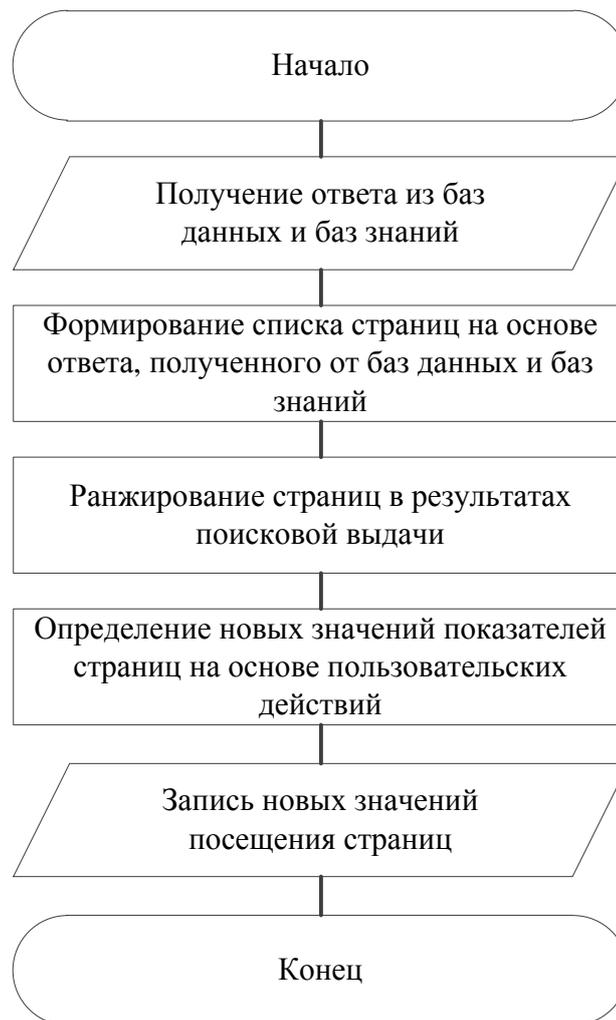


Рис. 1. Структурная схема алгоритма обработки результатов поисковой выдачи

Расчет нового значения показателя страницы начинается с вывода данных об объекте на открытую пользователем страницу. После этого производится определение текущего значения общего количества

посещений страницы с последующим прибавлением к этому значению 1. Затем определяется, каким образом пользователь попал на страницу: через поисковую систему или нет. Если переход осуществлен

не через поисковую систему, то производится перерасчет значения показателя страницы по формуле (8) и запись полученных значений в базу данных с последующим завершением алгоритма. Если же переход был осуществлен из поисковой системы, то определяются текущие значения общего времени пребывания пользователей на странице при переходе через поисковую систему, количества активаций чекбокса при посещении страницы через поисковую систему, количества посещений страницы из поисковой системы, после которых был продолжен поиск с последующим переходом на другую страницу, и общего количества посещений страницы через поисковую систему. Далее запускается отсчет времени пребывания пользователя на странице с 0 секунд до тех пор, пока пользователь находится на ней. Если он покинул ее, то отсчет времени останавливается, а полученное время фиксируется, после чего определяется, покинул ли пользователь поисковую систему. Если уход с поисковой системы не был осуществлен, то определяется, ввел ли пользователь новый запрос в поисковую систему. Если запрос не был введен, то проверяется переход на другую страницу из поиска. Если переход не был осуществлен, то проверяется повторный переход на ту же страницу. Если повторный вход не зафиксирован, то алгоритм возвращается к моменту проверки, покинул ли пользователь поисковую систему. Если же повторный вход на страницу был зафиксирован, то запускается отсчет времени пребывания пользователя на странице. Начальной точ-

кой отсчета в этом случае является время, зафиксированное ранее на моменте, когда пользователь перешел со страницы назад в поисковую систему. При этом алгоритм возвращается на этап ожидания, пока пользователь не покинет страницу. При выходе из поисковой системы или вводе нового запроса в ней производится проверка, провел ли пользователь на странице больше 90 с или нет. Если же после посещения страницы был осуществлен переход на другую страницу через поисковую систему, то сначала к текущему значению количества посещений страницы через поисковую систему, после которых был продолжен поиск с последующим переходом на другую страницу, прибавляется 1. Если пользователь пробыл на странице больше 90 с, то фиксируется значение времени пребывания 90 с, в ином случае фиксируется рассчитанное время. Затем определяется, был ли активирован чекбокс. При активном чекбоксе к текущему значению количества активаций чекбокса при посещении пользователями страницы через поисковую систему прибавляется 1. В ином случае прибавление не производится. Далее прибавляется 1 к общему количеству посещений страницы через поисковую систему. На основе полученных значений по формуле (8) рассчитывается новое значение показателя страницы. Все полученные значения записываются в базу данных вместо старых значений, после чего алгоритм завершается. На рис. 2 представлена структурная схема алгоритма определения значения показателя страницы.

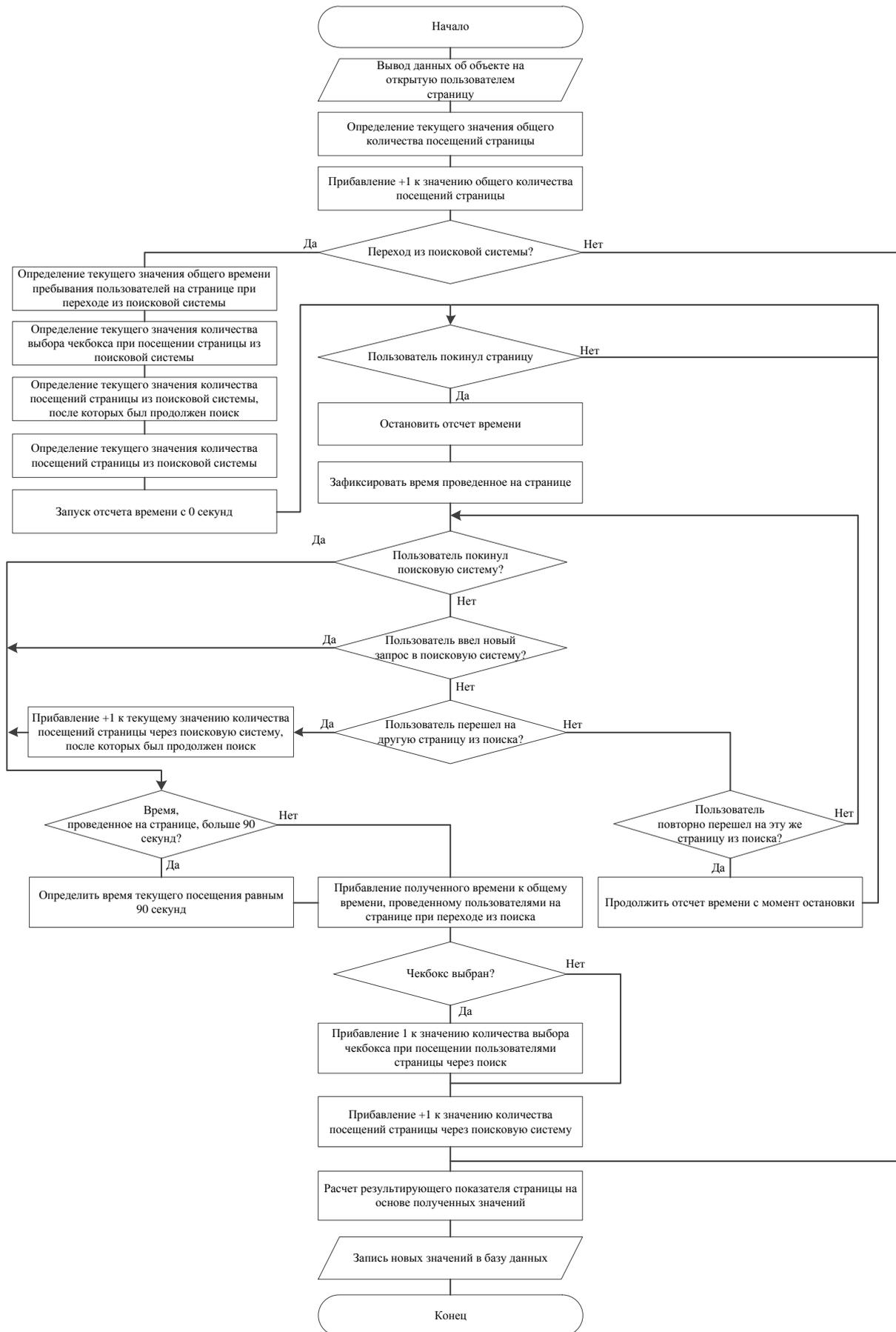


Рис. 2. Структурная схема алгоритма определения значения показателя страницы

### Результаты экспериментов

В таблице представлены значения показателя страницы при изменении значений общего количества посещений страницы, количества посещений страницы через поисковую систему (указано в процентах от общего количества посещений), среднего времени посещения страницы, количества активаций чекбокса при посе-

щении страницы через поисковую систему (указано в процентах от количества посещений через поисковую систему) и количества посещений, после которых был продолжен поиск с последующим переходом на другую страницу (указано в процентах от количества посещений через поисковую систему).

Таблица

Зависимость показателя страницы от значений показателей, определяемых пользовательскими действиями

Общее количество посещений страницы	Количество посещений через поисковую систему, %	Среднее время посещения страницы, с	Количество активаций чекбокса при посещении страницы через поисковую систему, %	Количество посещений страницы через поисковую систему, после которых был продолжен поиск, %	Показатель страницы
10000	95	45	20	50	1,25
	95	45	20	40	1,35
	95	45	30	50	1,35
	95	60	20	50	1,42
	90	45	20	50	1,3
1000	95	45	20	50	1,25
	95	45	20	40	1,35
	95	45	30	50	1,35
	95	60	20	50	1,42
	90	45	20	50	1,3
100	95	45	20	50	1,25
	95	45	20	40	1,35
	95	45	30	50	1,35
	95	60	20	50	1,42
	90	45	20	50	1,3
10	95	45	20	50	1,25
	95	45	20	40	1,35
	95	45	30	50	1,35
	95	60	20	50	1,42
	90	45	20	50	1,3

### Обсуждение результатов

Данные, представленные в таблице, показывают, что положение страницы в результатах поисковой выдачи зависит не от общего количества ее посещений, а от процентного соотношения количества посещений страницы через поисковую систему и общего количества посещений,

среднего времени посещения страницы, процентного соотношения количества активаций чекбокса при посещении страницы и количества посещений страницы через поисковую систему, процентного соотношения количества посещений страницы, после которых был продолжен поиск, и

количества посещений страницы через поисковую систему. Результаты экспериментов показывают, что применение предложенной модели приводит к более достоверным результатам поисковой выдачи в сравнении с классической моделью определения положения страницы на основе количества посещений. Так, например, страница, которую посетили 10 раз, может оказаться выше страницы, которую посе-

тили 1000 раз. Благодаря этому страницы с данными, добавленными позже, не зате-ряются в результатах поисковой выдачи и будут находиться в случае соответствия информации на них пользовательским ожиданиям. Кроме того, использование данной модели делает ранжирование страниц более эффективным уже после первых итераций поиска.

### Заключение

Использование в поисковой системе предприятия предложенных математической модели и алгоритма определения показателя страницы приводит к тому, что последовательность страниц в результатах поисковой выдачи носит систематизированный характер, что повышает качество поиска и его скорость. Кроме того, применение предложенной модели делает ранжирование результатов поисковой выдачи более достоверным и точным в сравнении с моделью, в которой положение страницы

определяется только на основе количества посещений. Тем не менее стоит учитывать, что работа данной системы увеличивает нагрузку на каналы передачи данных и может замедлить работу поискового приложения в целом, поэтому ее внедрение будет обоснованным только в том случае, если на предприятии хранится большое количество информации, которая может быть найдена по одинаковым запросам, но при этом отличается содержанием.

### СПИСОК ЛИТЕРАТУРЫ

1. ГОСТ 34.003-90. Информационная технология. Комплекс стандартов на автоматизированные системы. Термины и определения. - М., 2009. - 16 с.
2. ГОСТ 23501.101-87. Системы автоматизированного проектирования. Основные положения. - М., 1988. - 11 с.
3. Голев, В. Три составные части САПР предприятия: аппаратное обеспечение / В. Голев // САПР и графика. - 2000. - № 11. - URL: <https://sapr.ru/article/8100> (дата обращения: 02.06.2018).
4. Zubkova, T.M. Creation of system of computer-aided design for technological objects / T.M. Zubkova, M.A. Tokareva, N.Z. Sultanov // International Conference Information Technologies in Business and Industry 2018. IOP Conf. Series: Journal of Physics: Conf. Series 1015 (2018) 052031.
5. Segev, El. Google and the Digital Divide: The Biases of Online Knowledge / El. Segev. - Oxford: Chandos Publishing, 2010. - 171 p.
6. Birialtsev, E. Intelligent search in Big Data / E. Birialtsev, N. Bukharaev, A. Gusenkov // IOP Conf. Series: Journal of Physics: Conf. Series 913 (2017) 012010.
7. Гуревич, Л.А. Мультиагентные системы / Л.А. Гуревич, А.Н. Вахитов // Введение в Computer Science. - 2005. - С. 116-139.
8. Gwiazda, A. Modeling of a production system using the multi-agent approach / A. Gwiazda, A. Se-
9. Srinivasan, S. Multi-agent based decision Support System using Data Mining and Case Based Reasoning / S. Srinivasan, Jagjit Singh, Vivek Kumar // IJCSI International Journal of Computer Science Issues. - 2011. - Vol. 8. - Issue 4. - № 2.
10. Gyurjyan, V. AF ECS. Multi-Agent Framework for Experiment Control Systems / V. Gyurjyan, D. Abbott, G. Heyes, E. Jastrzembki, C. Timmer, E. Wolin // Journal of Physics: Conference Series 119 (2008) 022025. International Conference on Computing in High Energy and Nuclear Physics (CHEP'07).
11. Kovalev, I.V. Formation and control system by the specialized data in information networks / I.V. Kovalev, P.V. Zelenkov, M.V. Karaseva, V.V. Brezitskaya, D.I. Kovalev // IOP Conf. Series: Materials Science and Engineering 122 (2016) 012018. XIX International Scientific Conference Reshetnev Readings 2015.
12. Zhang, Xin-Jian. Successive lag synchronization on dynamical networks with communication delay / Xin-Jian Zhang, Ai-Ju Wei, Ke-Zan Li. // Chinese Physics B. - 2016. - Vol. 25. - № 3.
13. Пестерев, П.В. Структура и алгебраическая модель информационной поисковой системы предприятия на основе мультиагентной системы / П.В. Пестерев // Информационные технологии в науке и производстве: материалы всерос. мо-

- лод. науч.-техн. конф. - Омск: Изд-во ОмГТУ, 2018. - С.193-198.
14. Показатель отказов. Справка - Google Analytics. - URL: <https://support.google.com/analytics/answer/1009409?hl=ru> (дата обращения: 02.06.2018).
1. RSS 34.003-90. *Information Technology. Complex of Standards for Automated Systems. Terms and Definitions.* – М., 2009. – pp. 16.
  2. RSS 23501.101-87. *Systems of Computer Aided Design. Basic Regulations.* – М., 1988. – pp. 11.
  3. Golev, V. Three component parts of company CAD systems: hardware / V.Golev // *CAD Systems and Graphics.* – 2000. – No.11. - URL: <https://sapr.ru/article/8100> (address date: 02.06.2018.).
  4. Zubkova, T.M. Creation of system of computer-aided design for technological objects / T.M. Zubkova, M.A. Tokareva, N.Z. Sultanov // International Conference Information Technologies in Business and Industry 2018. IOP Conf. Series: Journal of Physics: Conf. Series 1015 (2018) 052031.
  5. Segev, El. Google and the Digital Divide: The Biases of Online Knowledge / El. Segev. - Oxford: Chandos Publishing, 2010. - 171 p.
  6. Birialtsev, E. Intelligent search in Big Data / E. Birialtsev, N. Bukharaev, A. Gusenkov // IOP Conf. Series: Journal of Physics: Conf. Series 913 (2017) 012010.
  7. Gurevich, L.A. Multi-agent systems / L.A. Gurevich, A.N. Vakhitov // *Introduction in Computer Science.* – 2005. – pp. 116-139.
  8. Gwiazda, A. Modeling of a production system using the multi-agent approach / A. Gwiazda, A. Sekala, W. Banas // IOP Conf. Series: Materials Science and Engineering 227 (2017) 012052.
  9. Srinivasan, S. Multi-agent based decision Support System using Data Mining and Case Based Reasoning / S. Srinivasan, Jagjit Singh, Vivek Kumar // IJCSI International Journal of Computer Science Issues. - 2011. - Vol. 8. - Issue 4. - № 2.
  10. Gyurjyan, V. AF ECS. Multi-Agent Framework for Experiment Control Systems / V. Gyurjyan, D. Abbott, G. Heyes, E. Jastrzembki, C. Timmer, E. Wolin // Journal of Physics: Conference Series 119 (2008) 022025. International Conference on Computing in High Energy and Nuclear Physics (CHEP'07).
  11. Kovalev, I.V. Formation and control system by the specialized data in information networks / I.V. Kovalev, P.V. Zelenkov, M.V. Karaseva, V.V. Brezitskaya, D.I. Kovalev // IOP Conf. Series: Materials Science and Engineering 122 (2016) 012018. XIX International Scientific Conference Reshetnev Readings 2015.
  12. Zhang, Xin-Jian. Successive lag synchronization on dynamical networks with communication delay / Xin-Jian Zhang, Ai-Ju Wei, Ke-Zan Li. // Chinese Physics B. - 2016. - Vol. 25. - № 3.
  13. Pesterev, P.V. Structure and algebraic model of company information search system based on multi-agent system / P.V. Pesterev // *Information Technologies in Science and Production: Proceedings of the All-Russian Youth Scientific tech. Conf.* – Омск: Омск STU Publishing House, 2018. – pp. 193-198.
  14. *Failure Index. Reference* - Google Analytics. - URL: <https://support.google.com/analytics/answer/1009409?hl=ru> (address date: 02.06.2018).

Статья поступила в редакцию 18.06.18.

Рецензент: д.т.н., доцент Сибирского государственного университета телекоммуникаций и информатики  
Розулина Л.Г.

Статья принята к публикации 14.08.18.

#### СВЕДЕНИЯ ОБ АВТОРАХ:

**Пестерев Павел Викторович**, аспирант Омского государственного технического университета, e-mail: [scoolptor@mail.ru](mailto:scoolptor@mail.ru).

**Pesterev Paul Victorovich**, Post graduate student, Omsk State Technical University, e-mail: [scoolptor@mail.ru](mailto:scoolptor@mail.ru).

**Янишевская Анна Генриховна**, д.т.н., доцент, профессор кафедры «Инженерная геометрия и САПР» Омского государственного технического университета, e-mail: [anna-yanish@mail.ru](mailto:anna-yanish@mail.ru).

**Yanisheskaya Anna Genrihovna**, Dr. Sc. Tech, Assistant Prof., Prof of the Dep. "Engineering Geometry and CAD Systems", Omsk State Technical University, e-mail: [anna-yanish@mail.ru](mailto:anna-yanish@mail.ru).