



## КРАТКИЕ СООБЩЕНИЯ

УДК 004.89

# Использование самообучающегося алгоритма с элементами искусственного интеллекта на основе цепей Маркова для корректировки семантического ядра сайтов научного и коммерческого направления

**Колотовкин И.С.\***

Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация

ORCID: <https://orcid.org/0000-0002-6126-4849>

mail: [is@kolotovkin.pro](mailto:is@kolotovkin.pro)

В статье рассматривается эффективное использование программного средства с элементами искусственного интеллекта с опорой на цепи Маркова. На основе богатого личного опыта автора проводится описание типовых проблем, в данной сфере возникающих. В данной статье описаны только рекомендации по применению алгоритма.

**Ключевые слова:** семантическое ядро, цепи Маркова, ключевые фразы, искусственный интеллект.

### Для цитаты:

*Колотовкин И.С.* Использование самообучающегося алгоритма с элементами искусственного интеллекта на основе цепей Маркова для корректировки семантического ядра сайтов научного и коммерческого направления. 2022. Том 12. № 2. С. 103–109. DOI: <https://doi.org/10.17759/mda.2022120205>

## 1. ВВЕДЕНИЕ

Создание сайтов высокого уровня как правило опирается на расчет и оптимизацию семантического ядра. Семантическое ядро – это перечень основных ключевых фраз, используемых посетителями интернета при поиске необходимой им информации.

\**Колотовкин Игорь Сергеевич*, аспирант, преподаватель, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0002-6126-4849>, e-mail: [is@kolotovkin.pro](mailto:is@kolotovkin.pro)



Оптимальный выбор таких поисковых запросов, которые позволят обеспечить сайту максимум посещаемости при минимуме затрат, дело сложное и часто в достаточной мере субъективное. Тем не менее использование в процессе формирования семантического ядра искусственного интеллекта возможно и, при умелом, подходе весьма эффективно.

О достоинствах и недостатках применения средств программирования при создании семантического ядра были представлены ранее в исследованиях информационного контента пользователями в Сети Интернет [1], семантического ядра информационного средства сети Интернет [2], а также в докладе о задачах кластеризации запросов при составлении семантического ядра [3] на XVII Всероссийской научной конференции «Нейрокомпьютеры и их применение».

Но, когда семантическое ядро уже создано и сайт успешно функционирует, можно ли считать работу с семантическим ядром завершённой? Нет, если сайт предполагает развитие и расширение своей сферы деятельности. Что это означает? Владельцы коммерческого продающего сайта могут расширить ассортимент предлагаемой покупателям продукции. А это означает необходимость сбора ключевых фраз по этим новым товарам.

Еще важнее и сложнее оказывается задача, если сайт содержит научную информацию, так как наука каждый день расширяет горизонты познаваемого, появляются новые экспериментальные данные, гипотезы и теории, которые требуют освещения в научных кругах. А значит появляются новые в запросе фразы, на которые должен правильно реагировать сайт.

Именно здесь в полной мере было бы разумно использование самообучающегося алгоритма с элементами искусственного интеллекта на основе цепей Маркова.

## 2. ОТ НЕДОСТАТКОВ К ДОСТОИНСТВАМ

Если при создании семантического ядра мы указывали значимые недостатки использования искусственного интеллекта, то почему же при работе с уже имеющимся ядром, можно говорить о том, что достоинства перевешивают эти недостатки? Потому что:

- 1) длительность процесса;
- 2) длительность процесса;
- 3) длительность процесса.

Может показаться, что приведенном выше списке есть какая-то ошибка. Нет, в данном списке нет ошибки. Рассмотрим эти три пункта подробнее.

### Пункт первый

**Длительность** процесса, как временной показатель существования необходимости работы с семантическим ядром.

Успешный сайт может существовать годами. И если мы хотим, чтобы человек вел работу по постоянной регулировке семантического ядра, необходимо выделить отдельную рабочую единицу – специалиста, который в безостановочном режиме будет



фактически проводить постоянное создание нового семантического ядра. В ускоренном формате, но процессы при этом точно такие же. Это, безусловно, неэкономично и в конечном счете неэффективно.

Компьютер же может выполнять все те же действия параллельно с другими функциями, практически не используя ценных ресурсов, так как современные компьютеры достаточно мощные. Таким образом экономятся финансы и трудозатраты при не слишком значительной потере в качестве. Этой потерей качества в данном случае можно пренебречь.

То есть недостатки искусственного интеллекта перед человеческим разумом в творческом процессе при ограниченном времени работы становятся его преимуществом при использовании в длительном рутинном функционировании.

#### Пункт второй

**Длительность** процесса, как временные трудозатраты.

В этом случае мы имеем в виду отрицательную длительность. А именно значительно большие затраты времени на корректировку ядра в том случае, если работает человек. Даже учитывая возможность использования оператором неких компьютерных программ, сбор, обработка, проверка и формирование кластера запросов при работе человека значительно превышают по времени те же самые действия, производимые компьютером.

А поскольку (особенно в отношении научных сайтов) объемы новой информации весьма велики, человек просто физически может не успевать обрабатывать все происходящие изменения.

К тому же появление новых знаний влечет за собой вхождение в научный обиход новых терминов, понятий, фразеологических оборотов. Если специалист, работающий с сайтом, с ними не знаком, он может отбраковывать их или неправильно классифицировать. А изучение всех новаций потребует от него опять-таки дополнительных затрат времени.

Таким образом быстрота работы искусственного интеллекта является его несомненным достоинством.

#### Пункт третий

**Длительность** процесса, как фактор, позволяющий эффективно внедрять самообучающийся искусственный интеллект.

Поскольку при создании семантического ядра мы имеем дело с конкретным объемом информации и ограниченным временем на его формирование, а после переходим к другому сайту, возможно весьма далекому по тематике, то полноценное использование самообучающегося искусственного интеллекта весьма затруднено – времени на обучение может просто не хватить для формирования достаточно высокого качества работы. Что можно считать значимым и трудно преодолимым недостатком.

При ведении же определенного сайта, наличествует достаточный временной ресурс, позволяющий искусственному интеллекту постоянно совершенствоваться в поиске и обработке запросов конкретного направления. Это дает возможность не только более эффективно использовать его, но и активно повышать качество производимых



действий. Чем нивелируем и относительный минус (недостаток), указанный в первом пункте. А это уже неоспоримое достоинство. В табл. 1 приведены сравнительные показатели для человека (специалиста) и компьютера с ИИ.

Таблица 1

### Примерные\* затраты времени, труда и качество результата

Показатели в зависимости от сложности сайта	Исполнитель	
	Человек	Компьютер (самообучающийся ИИ**)
Временные затраты	8–24 часа	1–4 часа
Работа специалиста	Требуется постоянно	Не требуется (за исключением установки процесса)
Скорость работы	Возможно замедление, зависящее от человеческого фактора	Ускорение процесса при использовании самообучающегося ИИ
Масштабируемость (увеличение необходимого объема работы)	Требуется найти и обучить дополнительного специалиста	Требуется увеличить мощность используемого компьютера (купить более мощный/дополнительный)
Возможность ошибки	Возможны (постоянно действующий человеческий фактор)	Возможны (недочеты программного обеспечения) – с течением времени сводятся к нулю
Качество работы	Высокое/достаточное	Приемлемое/повышающееся со временем до достаточного

\* – эмпирические данные по результатам работы с разработчиками семантического ядра и программным обеспечением разного уровня.

\*\* – искусственный интеллект.

## 3. ИСПОЛЬЗОВАНИЕ САМООБУЧАЮЩЕГОСЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ОСНОВЕ ЦЕПЕЙ МАРКОВА

При систематизации интернет запросов для корректировки семантического ядра самообучающийся искусственный интеллект может эффективно использовать в работе цепи Маркова.

С одной стороны Марковские цепи [4] удобны в поиске новых запросов, сходных с уже существующими при общей известной сфере их направленности. При чем отсутствие «памяти» помогает не зависеть от ранее использованного контента, что позволяет работать с большими массивами информации абсолютно свободно.

Это важно особенно для научных сайтов, где поток новой информации не должен зависеть от уже существующих в сети материалов по определенной теме.

С другой стороны Марковские цепи помогают ранжировать выбранные запросы, исходя из их теоретической посещаемости, основанной на прогнозировании вероят-

ностных переходов между страницами сайта. Это качество просто необходимо для коммерческих сайтов, где на основе активности посещений страницы распределяется информационный контент.

На рис. 1 представлены поведенческие факторы с учетом Марковской цепи.

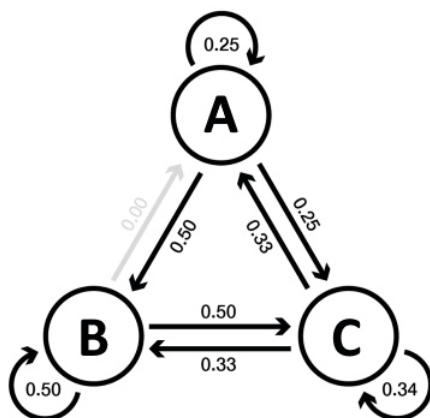


Рис. 1. Поведенческие факторы ранжирования ключевых фраз

На рис.1 приняты следующие обозначения:

- A – Посетитель не приходит на сайт по поисковой выдаче;
- B – Посетитель заходит на сайт по поисковой выдаче;
- C – Посетитель переходит на другую страницу сайта по ссылке на первой посещенной странице.

При начальной вероятности зайти и не зайти на сайт в первый раз равной 0,5.

В целом как одна, так и другая возможность, предоставляемая цепями Маркова, полезна и актуальна для всех сайтов вне зависимости от их направленности.

С другой стороны, самообучающийся искусственный интеллект позволяет сам, по своей сущности, а именно вне «задействования» Марковских цепей, использовать отсутствующую в этом методе прогнозирования «память», т.е. формировать релевантный прогноз не только на основании текущего события, но и уже накопленных ранее данных.

То есть, если при использовании исключительно Марковской цепи, появляется возможность свободно собирать ключевые фразы не зависимо от их веса посещаемости, то самообучающийся искусственный интеллект вне данного метода способен осуществлять ранжирование ключевых фраз по их эффективности относительно друг друга.

Таким образом именно взаимодействие двух начал – самообучающегося алгоритма и искусственного интеллекта на основе цепей Маркова, позволяет успешно использовать компьютерные программы для сбора, обработки, классификации ключевых фраз на постоянно действующей основе. Что в значительной мере повышает качество прогнозирования, а значит и качество семантического ядра. А так же



предоставляет возможность экономить денежные средства, расходуемые на обслуживание, рекламу и дополнительную раскрутку длительно функционирующего сайта.

#### 4. ВЫВОД

Длительно существующие сайты небезосновательно требуют для поддержания своей эффективности постоянной корректировки семантического ядра.

Использование труда человека при долговременной работе с семантическим ядром такого сайта на сегодняшний день можно считать более затратным по финансам и времени, чем использование информационных технологий, а значит, в конечном итоге, и не достаточно эффективным.

Использование программного средства с элементами искусственного интеллекта при корректировке семантического ядра действующих сайтов научного и коммерческого направления является экономически выгодным и весьма эффективным решением, так как исключает человеческий фактор и как риск ошибок при оптимизации ключевых фраз, и как более затратный вариант процесса при равнозначном, а в дальнейшем и меньшем качестве результата.

При этом основанное на цепях Маркова программное обеспечение, собирающее первоначальную базу данных поисковых запросов, используемое в этой работе повышает её результативность и «действенность» относительно существующих на данный момент программ.

Таким образом можно полагать, что необходимо разрабатывать программные средства с элементами искусственного интеллекта с опорой на цепи Маркова для оптимизации работы по корректировке действующих сайтов научного и коммерческого направлений.

#### *Литература*

1. Рыков А.С. Методы системного анализа: Многокритериальная и нечеткая оптимизация, моделирование и экспертные оценки. – М.: Экономика, 1999.
2. Search Engine Optimization (SEO) Secrets / Danny Dover, Erik Dafforn. – Indianapolis: John Wiley & Sons Limited, 2011.
3. Колотовкин И.С., Кулик С.Д., Шевченко А.А. Нейронные сети в задачах кластеризации запросов при составлении семантического ядра // Тезисы докладов. XVII Всероссийская научная конференция «Нейрокомпьютеры и их применение» НКП-2019. Москва, 19 марта 2019 г. – М.: ФГБОУ ВО МГППУ, 2019. – С.262–263.
4. Кельберт М.Я., Сухов Ю.М. Вероятность и статистика в примерах и задачах. Т. II: Марковские цепи как отправная точка теории случайных процессов и их приложения. – М.: МЦНМО, 2009.



# Using a Self-Learning Algorithm with Elements of Artificial Intelligence Based on Markov Chains to Correct the Semantic Core of Scientific and Commercial Sites

**Igor S. Kolotovkin\***

Moscow State University of Psychology & Education (MSUPE)

Moscow, Russian Federation

ORCID: <https://orcid.org/0000-0002-6126-4849>

mail: [is@kolotovkin.pro](mailto:is@kolotovkin.pro)

The article discusses the effective use of a software tool with elements of artificial intelligence based on Markov chains. Based on the rich personal experience of the author, a description of typical problems arising in this area is given. This article describes only recommendations for using the algorithm.

Keywords: semantic core, Markov chains, key phrases, artificial intelligence.

## **For citation:**

Kolotovkin I.S. Using a Self-Learning Algorithm with Elements of Artificial Intelligence Based on Markov Chains to Correct the Semantic Core of Scientific and Commercial Sites. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2022. Vol. 12, no. 2, pp. 103–109. DOI: <https://doi.org/10.17759/mda.2022120205> (In Russ., abstr. in Engl.).

## **References**

1. Rykov A.S. Methods of system analysis: Multicriteria and fuzzy optimization, modeling and expert evaluations. – M.: Economics, 1999.
2. Search Engine Optimization (SEO) Secrets / Danny Dover, Erik Dafforn. – Indianapolis: John Wiley & Sons Limited, 2011.
3. Kolotovkin I.S., Kulik S.D., Shevchenko A.A., Neural networks in the problems of query clustering when compiling a semantic core // Abstracts. XVII All-Russian Scientific Conference “Neurocomputers and their applications” NKP-2019. Moscow, March 19, 2019 – M.: FGBOU VO MGPPU, 2019. – P. 262–263.
4. Kelbert M. Ya., Sukhov Yu. M. Probability and statistics in examples and problems. Vol. II: Markov chains as a starting point for the theory of random processes and their applications. – M.: MTSNMO, 2009.

\***Igor S. Kolotovkin**, Post-Graduate Student, Lecturer, Moscow State University of Psychology & Education (MSUPE), Moscow, Russian Federation, ORCID: <https://orcid.org/0000-0002-6126-4849>, e-mail: [is@kolotovkin.pro](mailto:is@kolotovkin.pro)