

*Майоров К.Н., Кочарян Т.С. Разработка алгоритма анализа естественности ссылочного профиля сайтов // Академия педагогических идей «Новация». Серия: Студенческий научный вестник. – 2017. – № 08 (август). – АРТ 366-эл. – 0,1 п.л. - URL: <http://akademnova.ru/page/875550>*

**РУБРИКА: ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ**

**УДК 004.8**

**Майоров Константин Николаевич**

Магистр 09.04.04, факультет ИВТ

*Научный руководитель:* Ложкин А.Г., д.т.н.проф. каф. ПО ИжГТУ  
ФГБОУ ВО «Ижевский государственный технический университет  
имени М.Т. Калашникова»

gibiskus@gmail.com

**Кочарян Тигран Саркисович**

Магистр 09.04.04, факультет ИВТ

*Научный руководитель:* Ложкин А.Г., д.т.н.проф. каф. ПО ИжГТУ  
ФГБОУ ВО «Ижевский государственный технический университет  
имени М.Т. Калашникова»

tigran0105@gmail.com

**РАЗРАБОТКА АЛГОРИТМА АНАЛИЗА ЕСТЕСТВЕННОСТИ  
ССЫЛОЧНОГО ПРОФИЛЯ САЙТОВ**

*Аннотация:* Рассматривается алгоритм анализа ссылочного профиля сайтов по естественности входящих ссылок. Алгоритм предназначен для исследования ссылочного профиля сайта и позволяет по большому количеству факторов определить являются ли входящие ссылки сайта естественными или оплаченными.

*Ключевые слова:* система анализа ссылочного профиля, ссылочный профиль, ссылочные факторы

**Maiorov Konstantin Nikolaevich**

Master of 09.04.04, Faculty of Computer Science

*Supervisor:* Lozhkin A.G., DSc., Prof. FGBOU VO ISTU

FGBOU VO «Kalashnikov Izhevsk State Technical University»

gibiskus@gmail.com

**Kocharyan Tigran Sarkisovich**

Master of 09.04.04, Faculty of Computer Science

*Supervisor:* Lozhkin A.G., DSc., Prof. FGBOU VO ISTU

FGBOU VO «Kalashnikov Izhevsk State Technical University»

tigran0105@gmail.com

**DEVELOPMENT OF AN ALGORITHM FOR ANALYZING THE  
NATURALNESS OF A REFERENCED SITE PROFILE**

*Annotation:* An algorithm for analyzing the link profile of sites by the naturalness of incoming links is considered. The algorithm is designed to investigate the site's link profile and allows, by a large number of factors, to determine whether the incoming site links are natural or paid for.

*Keywords:* Analysis system of site link profile, link profile, link factors.

Ссылки — важнейшая составляющая интернета. Они связывают между собой страницы всемирной паутины и являются тропинками, по которым ходят пользователи.

Поисковые системы ценят ссылки, которые ставятся для того, чтобы помочь посетителям сайта. Те же ссылки, которые появляются на сайтах за определенную плату (особенно когда они размещаются с использованием ссылочных бирж и предназначены для поисковых систем, а не посетителей сайта), мешают поисковым системам корректно определять степень доверия к сайтам и их релевантность запросам пользователей. Поисковые системы стараются не учитывать купленные ссылки, а при их большом количестве могут даже наложить на сайт определенные санкции и понизить его в поисковом рейтинге.

Поэтому сегодня для успешного продвижения сайта на первую страницу поисковой выдачи просто необходимо понимать, что из себя представляет ссылочный профиль сайта, насколько он отличается от профиля сайтов, находящихся на первой странице поисковой выдачи, какие ссылки являются естественными, а какие куплены.

В данной работе рассматривается алгоритм для анализа естественности ссылочного профиля исследуемого сайта. На основе различных ссылочных факторов алгоритм распознает являются ли входящие ссылки естественными, оценивает ссылочный профиль в целом и сопоставляет с ссылочными профилями конкурентов, находящихся на первой странице поисковой выдачи.

Задача по анализу ссылочного профиля не является новой. Существуют различные программные решения по анализу ссылок Ahrefs, Majestic SEO, Check Trust. Однако данные программы не позволяют оценить естественность ссылок, а также не обладают возможностью генерировать рекомендации по улучшению ссылочного профиля. Следовательно, реализация системы комплексного анализа ссылочного профиля является вполне оправданной.

Исходя из поставленной цели, в работе решались следующие задачи:

- построить модель ссылочного профиля;
- разработка алгоритма идентификации тематики сайта-донора и страницы со ссылкой;
- разработка алгоритма классификации ссылок по естественности;
- разработка ПО для генерации рекомендаций по улучшению ссылочного профиля.

Классификация ссылок на предмет естественности происходит из учета большого количества факторов (Рис.1). К таким факторам можно отнести: расположение ссылки в теле страницы, тип анкера, схожесть тематик страниц [1], схожесть тематик сайтов, наличие переходов пользователей по этой ссылке и других.

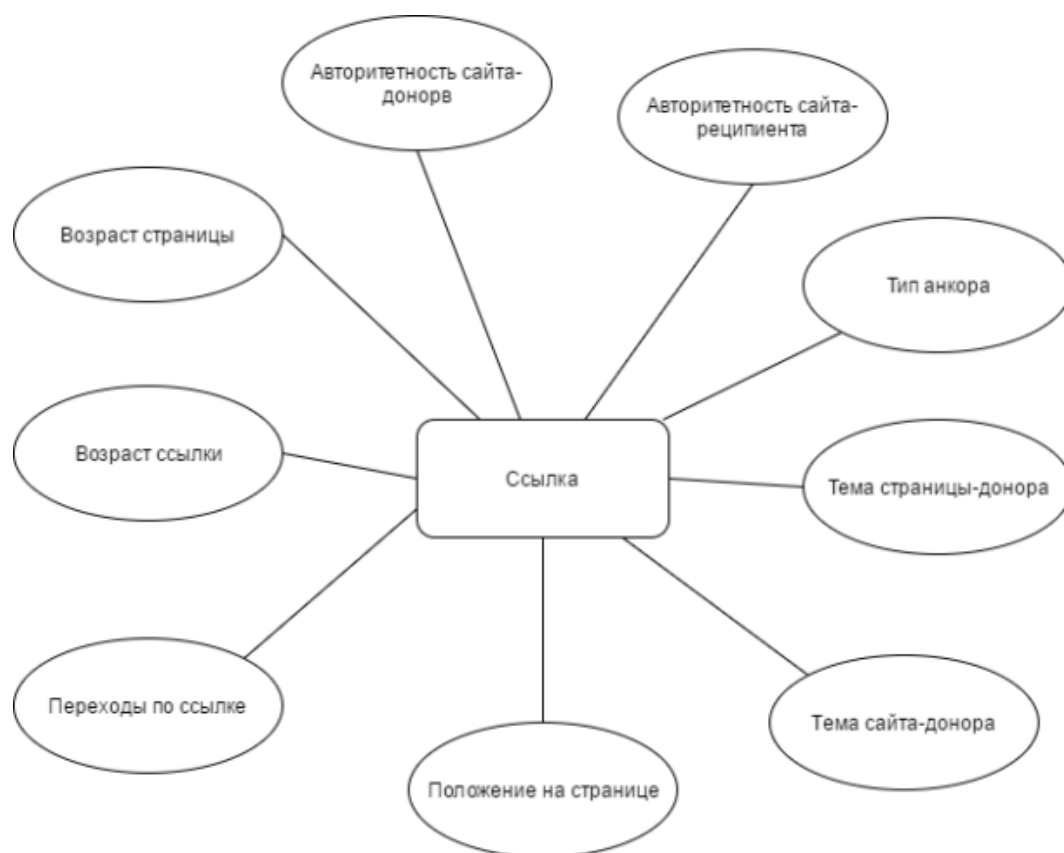


Рис. 1. Ссылочные факторы

Каждый фактор имеет свой вес, который формируется в процессе обучения системы на обучающих выборках. Входными сигналами нейронной сети являются различные ссылочные факторы, описанные выше. Каждый входной сигнал имеет свой вес  $w_{np}$ , который передается в скрытый слой  $S_1..S_n$ . Каждый скрытый слой также имеет свой вес  $W_{np}$ . На выходе мы получаем класс ссылки: естественная или купленная.

Алгоритм feedforward (прямопроходящий)[2] в цикле обходит все узлы скрытого слоя и для каждого из них вычисляет сумму величин выходных сигналов от узлов входного слоя, помноженных на вес соответствующей связи. Выходной сигнал каждого скрытого узла — это результат применения функции тангенса к взвешенной сумме входных сигналов [3]. Этот сигнал передается на выходной уровень. Выходной уровень делает то же самое — умножает полученные от предыдущего уровня сигналы на веса связей и применяет функцию тангенса для получения окончательного результата.

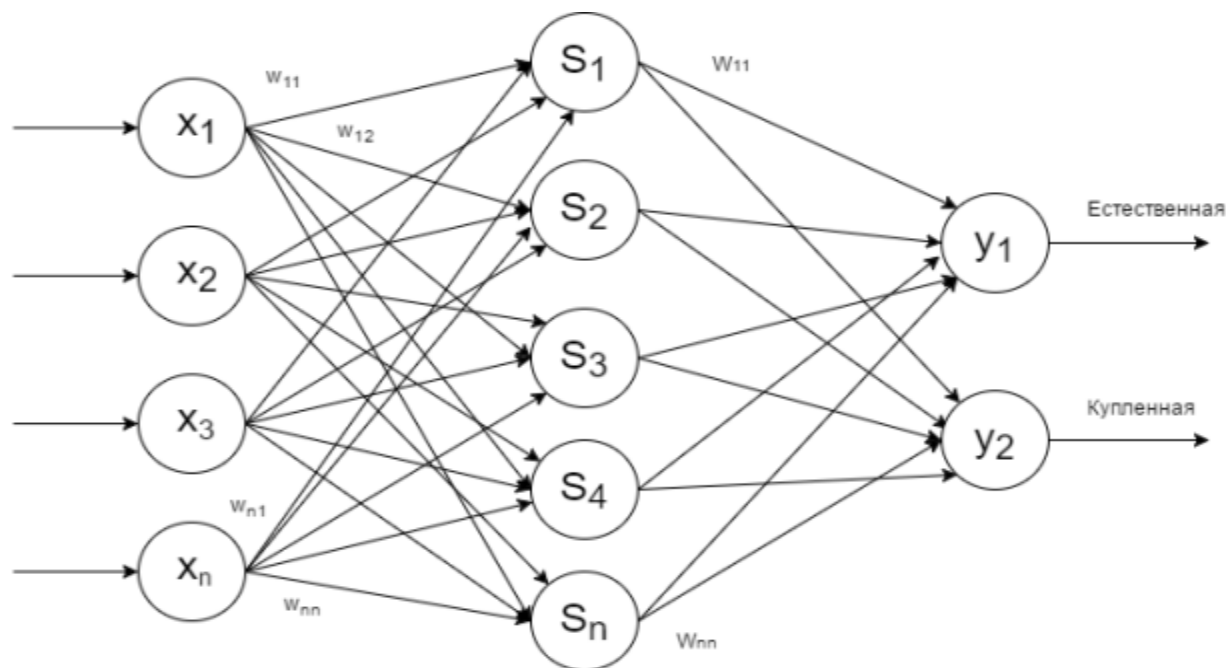


Рис. 2. Схема нейронной сети для классификации ссылки

Алгоритм реализован в виде веб-приложения.

Для разработки веб-приложения используется язык Python [4]. Этот язык, позволяет качественно и на достаточном уровне выполнить приложение необходимой сложности, а также обладает богатым набором библиотек для работы с веб-страницами. Для управления данными применяется фреймворк Django.

В качестве СУБД используется свободная система управления базами данных PostgreSQL [5]. PostgreSQL – это свободно распространяемая объектно-реляционная система управления базами данных (ORDBMS), наиболее развитая из открытых СУБД в мире и являющаяся реальной альтернативой коммерческим базам данных.

В качестве веб-сервера выбран Apache 2.2 HTTP — сервер, принадлежащий к категории свободного ПО [6]. Основными достоинствами Apache считаются надёжность и гибкость конфигурации. Он позволяет подключать внешние модули для предоставления данных, использовать СУБД для аутентификации пользователей, модифицировать сообщения об ошибках.

#### **Список использованной литературы:**

1. A. Lozhkin, P. Bozek, V. Lyalin, V. Tarasov, M. Tothova, R. Sultanov Reverse and Direct Methods for Solving the Characteristic Equation. AIP Conference Proceedings, v. 1738, 480140(2016) doi: 10.1063/1.49544935
2. Каллан Р. Основные концепции нейронных сетей = The Essence of Neural Networks First Edition. — М.: Вильямс, 2001. — 288 с. — ISBN 5–8459–0210-X.
3. Lozhkin A.G., Shubin V.A., Suslov Y.B., Bimakov E.V. In the issue of robots design/ Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering ElConRus 2017, 3 February 2017, Pp. 930-933.
4. Марк Саммерфилд. Программирование на Python 3. Подробное руководство. М.: Издательство «Символ-Плюс», 2009.

5. Малыгина М. П. Базы данных: основы, проектирование, использование. С.-П.: БХВ-Петербург, 2004.

6. Скот Хоккинс. Администрирование web-сервера Apache. М.: Издательский дом «Вильямс», 2001.

*Дата поступления в редакцию: 13.08.2017 г.*

*Опубликовано: 13.08.2017 г.*

*© Академия педагогических идей «Новация». Серия «Студенческий научный вестник», электронный журнал, 2017*

*© Майоров К.Н., Кочарян Т.С., 2017*