

*Ващук И.Н., канд. пед. наук,  
доцент отделения ЭСТТиАТП филиала РГУ нефти газа  
(НИУ) им. И.М.Губкина,  
Заельская Н.А.,  
старший преподаватель кафедры КБМОИС,  
Надточий Н.С.,  
старший преподаватель кафедры КБМОИС,  
Оренбургский государственный университет,  
Россия, г. Оренбург*

**ИСПОЛЬЗОВАНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ  
ДЛЯ РЕШЕНИЯ ПРОБЛЕМЫ АВТОМАТИЗАЦИИ  
В ПРОДВИЖЕНИИ ИНТЕРНЕТ-РЕСУРСОВ**

*Аннотация: В статье рассматриваются актуальные вопросы, связанные с разработкой системы правил, которые позволили бы осуществить автоматизацию в процессе оптимизации предоставления услуг платного хостинга, что позволит разбивать тарифные планы на классы..*

*Ключевые слова: продвижение сайтов, интеллектуальные технологии, SEO, интеллектуальное поисковое продвижение.*

*Vashchuk I. N., Cand. PED. sciences', associate Professor of the Department of Esttiatp branch of the Russian state University of oil and gas (NIU). I. M. Gubkin, Zael'skaya N. A., senior lecturer, Department of CSMSIS, Nadtochiy N. S., senior lecturer, Department of CSMSIS, Orenburg state University, Russia, Orenburg*

*Annotation: The article deals with topical issues related to the development of a system of rules that would allow automation in the process of optimizing the provision of paid hosting services, which will allow splitting tariff plans into classes.*

*Keywords: website promotion, intelligent technologies, SEO, intelligent search engine promotion.*

Автоматизация процессов в продвижении сайтов – тема, которая становится все более актуальной с каждым годом, поскольку с развитием поисковых алгоритмов SEO-специалистам приходится работать с растущим количеством факторов, влияющих на позиции сайта в выдаче. Если десять лет назад конкуренция между сайтами в интернете была ниже и структура сайтов – проще, то сегодня оптимизаторам зачастую приходится продвигать сайты с определенной «историей болезни», с огромным числом накопленных за годы страниц и т.д.

Автоматизация процессов в SEO-компании решает как минимум три ключевые задачи: экономия времени специалистов, которым без роботов приходится вручную проводить много рутинной работы; повышение качества услуг (робот, в отличие от человека, не может пропустить что-то важное по невнимательности, также мощности, объемы обрабатываемых роботом данных дают более точную аналитическую картину, нежели проведенный вручную анализ небольшой выборки сайтов); масштабируемость бизнеса (автоматизированные процессы позволяют, во-первых, увеличивать в разы количество проектов, находящихся в одновременной работе, во-вторых, решают проблемы, которые возникают обычно при уходе специалиста или при расширении штата).

В основу мы предлагаем разработку системы правил, которые позволили бы осуществить автоматизацию в процессе оптимизации предоставления услуг платного хостинга, что позволит разбивать тарифные планы на классы. Цель – классификация тарифных планов на услуги хостинга.

Исходные данные определяются такими переменными, как Цена в месяц = [0.99; 49], Цена в год = [10.15; 499], Объем = [5;2500], Трафик = [0.4;1000], Количество сервисов = [1;4], Виртуальных серверов = [1;1000], FTP входов = [1;1000].

Промежуточные цели определяются следующими переменными: Comp1 = большая, средняя, малая; Comp2 = большая, средняя, малая, Comp3 = большая, средняя, малая.

Для проведения анализа использовался метод главных компонент, что дало возможность по  $p$  исходным признакам выделить в общем случае  $p$  главных компонент

Введем исходные данные в программу STATISTICA (50x7). Исходная сводка анализа метода главных компонент (МГК) представлена в таблице 1.

Таблица 1 -Исходная сводка МГК

Component Number	Eigen Value	Percent of Variance	Cumulative Percentage
1	3.10802	44.400	44.400
2	1.45952	20.850	65.251
3	0.913763	13.054	78.304
4	0.797564	11.394	89.698
5	0.506804	7.240	96.938
6	0.20519	2.931	99.870
7	0.00913	0.130	100.000
Number of complete cases : 50			

Из полученной сводки заключаем, что анализу подвергаются переменные цена/месяц, цена/год, количество сервисов, количество ftp входов, объем, трафик, количество виртуальных серверов и что число объектов составляет 50.

Далее следует информация непосредственно МГК собственные значения главных компонент, упорядоченные по величине (Eigenvalue), процент дисперсии, приходящийся на каждую выделенную главную компоненту (Percent of Variance), накопленный процент дисперсии (Cumulative Percentage).

Приведенные цифры говорят о том, что уже первые три главные компоненты описывают 78,304 % дисперсии исходных данных.

Для более детального анализа проделали еще ряд операций. Получили веса признаков в главных компонентах, таблица2.

Таблица 2 -Веса признаков в главных компонентах

	Component1	Component2	Component3
Цена/мес	<b>0.5386</b>	0.1130	0.0844
Цена/год	<b>0.5258</b>	0.1691	0.0541
Объем	<b>0.51509</b>	-0.02169	0.103268
Трафик	0.1791	<b>0.5622</b>	-0.1295
FTP входов	0.1571	<b>-0.6791</b>	-0.1488
Виртуальных серверов	0.2588	-0.2142	<b>-0.7796</b>
Сервисов	0.21025	<b>-0.3672</b>	<b>0.5766</b>

Как следует из полученных цифр, в первой компоненте наблюдается наибольшая зависимость от цены за месяц, цены за год, и объема. Во второй главной компоненте наблюдается обратно пропорциональная зависимость от количества ftp входов и количества сервисов, а также прямо пропорциональная зависимость от трафика. В третьей главной компоненте наблюдается прямо пропорциональная зависимость от количества сервисов, а также обратно пропорциональная зависимость от количества виртуальных серверов.

Выбор значащих компонент и определение названия для них представлено ниже.

1) Выберем  $p=3$  главных компонент.

2) Определим названия для них по формуле:  $K_{ij} = \frac{\sum_{i=1}^p w_{ij}^2 [w_{kj}]}{\sum_{i=1}^p w_{ij}^2 [w_j]}$ ,

где  $[w_{kj}]$  – подмножество участвующих в названии весовых коэффициентов  $j$ -й компоненты;

$[w_j]$  – все весовые коэффициенты  $j$ -й компоненты.

$$w_{ij} = \begin{vmatrix} 0.5386 & 0.1130 & 0.0844 \\ 0.5258 & 0.1691 & 0.0541 \\ 0.51509 & -0.02169 & 0.103268 \\ 0.1791 & 0.5622 & -0.1295 \\ 0.1571 & -0.6791 & -0.1488 \\ 0.2588 & -0.2142 & -0.7796 \\ 0.21025 & -0.3672 & 0.5766 \end{vmatrix}$$

Для первой, второй и третьей компоненты имеем соответственно:

$$\sum_{i=1}^n w_{i1}^2 = 1$$

$$\sum_{i=1,2,3} w_{i1}^2 = 0,8318733$$

$$K_{H1} = 0,8318733$$

$$\sum_{i=1}^n w_{i2}^2 = 1$$

$$\sum_{i=4,5,7} w_{i2}^2 = 0.9120815$$

$$K_{H2} = 0.9120815$$

$$\sum_{i=1}^n w_{i3}^2 = 1$$

$$\sum_{i=6,7} w_{i3}^2 = 0.9402437$$

$$K_{H3} = 0.9402437$$

$K_{H1}$  принадлежит интервалу  $[0,75; 0,95]$ , значит, первая главная компонента определяется (более чем на 83,1%) следующими показателями: цена за месяц, цена за год, и объема выделяемой памяти.

$K_{и2}$  принадлежит интервалу  $[0,75; 0,95]$ , значит, вторая главная компонента определяется (более чем на 91,2%) следующими показателями: трафик, количество ftp входов, и количество сервисов.

$K_{и3}$  принадлежит интервалу  $[0,75; 0,95]$ , значит, третья главная компонента определяется (более чем на 94%) следующими показателями: количество виртуальных серверов и количество сервисов.

Учитывая зависимость компонент от признаков, охарактеризуем каждый из классов.

В первый класс входят тарифы с малой ценой, малым размером выделяемой памяти, различным лимитом на трафик, различным количеством ftp входов, различным количеством сервисов и различным количеством виртуальных серверов.

Во второй класс входят тарифы с средней ценой, средним размером выделяемой памяти, различным лимитом на трафик, различным количеством ftp входов, высоким количеством сервисов и средним количеством виртуальных серверов.

В третий класс входят тарифы с высокой ценой, высоким размером выделяемой памяти, различным лимитом на трафик, различным количеством ftp входов, высоким количеством сервисов и высоким количеством виртуальных серверов.

Таким образом мы можем говорить о возможности построения дерева решений для интеллектуального поискового продвижения.

#### **Использованные источники:**

1. Brett Lantz. Machine Learning with R. Packt Publishing, Birmingham - Mumbai, 2013
2. Quinlan J.R. C4.5: Programs for Machine learning // Morgan Kaufmann Publishers. 1993.
3. Quinlan, J. R. Induction of Decision Trees // Machine Learning. Kluwer Academic Publishers. 1986. № 1. P. 81–106.
4. Шахиди А. Деревья решений — общие принципы работы. URL: <http://www.basegroup.ru/library/analysis/tree/description>.