

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное
учреждение высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий

(наименование института полностью)

Кафедра «Прикладная математика и информатика»

(наименование кафедры)

02.03.03 Математическое обеспечение и администрирование
информационных систем

(код и наименование направления подготовки, специальности)

Технология программирования

(направленность (профиль)/специализация)

БАКАЛАВРСКАЯ РАБОТА

на тему Разработка программного модуля для реализации формирования
эффективных ключевых фраз для контекстной рекламы для ООО «Рудник»

Студент(ка)

А.В. Симонов

(И.О. Фамилия)

(личная подпись)

Руководитель

А.Б. Кузьмичев

(И.О. Фамилия)

(личная подпись)

Консультант

О.А. Парфенова

(И.О. Фамилия)

(личная подпись)

Допустить к защите

Заведующий кафедрой

к.т.н, доцент, А.В. Очеповский

(степень, звание, И.О. Фамилия)

(личная подпись)

«_____» _____ 20__ г.

Тольятти 2018

АННОТАЦИЯ

Тема выпускной квалификационной работы: Разработка программного модуля для реализации формирования эффективных ключевых фраз для контекстной рекламы для ООО «Рудник».

Объектом исследования является прогнозирование и оценка возможной эффективности отдельно взятых, тематически подобранных, ключевых фраз в рамках системы «Яндекс.Директ».

Предметом исследования является гипотеза возможности использования математических методов для прогнозирования и оценки эффективности ключевых фраз.

Цель бакалаврской работы заключается в проверке данной гипотезы и разработке программного модуля формирования эффективных ключевых фраз.

Выдвинута следующая задача: произвести анализ явления эффективности ключевой фразы, разработать алгоритм оценки и прогнозирования эффективности ключевой фразы и реализовать на его основе программный модуль.

Работа состоит из трех разделов.

Первый раздел направлен на изучение предметной области и постановки задачи на реализацию модуля формирования ключевых фраз.

Второй содержит описание и разработку алгоритма, используемого модулем, для оценки и прогнозирования эффективности ключевой фразы.

Третий раздел описывает проектирование и реализацию программного модуля через диаграммы классов, компонентов и последовательностей.

Выпускная квалификационная работа выполнена на 58 страницах, состоит из введения, трех глав, заключения и списка литературы. Список использованной литературы содержит 28 источников (4 на английском языке) Также работа содержит 12 рисунков, 14 таблиц и одно приложение.

ABSTRACT

The title of the graduation work is «The Development of Software for the Implementation of Generate Productive Keywords for Contextual Advertising for LLC «Rudnik»».

The object of the graduation work is to predict and evaluate the possible effectiveness of individual, thematically selected keyword in the «Yandex.Direct» system.

The subject of the work is the hypothesis of the possibility of using mathematical methods to predict and evaluate the effectiveness of keywords.

The purpose of the work is to test this hypothesis and develop a software for generate productive keywords.

The following task are proposed: to analyze the phenomenon of the key phrase efficiency, to develop an algorithm for assessing and predicting the key phrase efficiency and to implement a software module on its basis. The work consists of three sections.

The first section is aimed at analyzing of the subject area and the formulation of the problem for the implementation of the module of generate keywords.

The second contains a description and development of the algorithm used by the module to evaluate and predict the effectiveness of the keywords.

The third section describes the design and implementation of the program module through diagrams of classes, diagrams of components and diagrams of sequences.

The graduation work is done on 58 pages, consists of an introduction, three chapters, conclusion and references. The list of references contains 28 sources (4 in English) also the work contains 12 figures, 14 tables and one notice.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	6
1 АНАЛИЗ ОСНОВ КОНТЕКСТНОГО РЕКЛАМИРОВАНИЯ И ЭФФЕКТИВНОСТИ КЛЮЧЕВОЙ ФРАЗЫ В РАМКАХ СИСТЕМЫ «ЯНДЕКС.ДИРЕКТ».....	8
1.1 Анализ реализации контекстной рекламы на примере системы «Яндекс.Директ»	8
1.2 Анализ сервиса контроля и управления рекламным аккаунтом «Яндекс.Директ» «Rudvert».....	10
1.3 Определение понятия ключевой фразы в рамках системы «Яндекс.Директ»	11
1.4 Анализ существующих решений в формировании ключевых фраз.....	14
1.5 Определение показателей эффективности ключевой фразы.....	15
1.6 Определение признаков эффективности ключевой фразы	17
1.7 Выработка требований к новой технологии формирования ключевых фраз.....	19
1.8 Постановка задачи на разработку модуля формирования ключевых фраз.....	21
2 РАЗРАБОТКА АЛГОРИТМА ВЫЧИСЛЕНИЯ И ПРОГНОЗИРОВАНИЯ КОЭФФИЦИЕНТА ЭФФЕКТИВНОСТИ КЛЮЧЕВОЙ ФРАЗЫ.....	23
2.1 Выбор математического аппарата для вычисления коэффициента эффективности ключевой фразы	23
2.2 Выбор математического аппарата для прогнозирования коэффициента эффективности ключевой фразы	24
2.3 Построение модели вычисления коэффициента эффективности ключевой фразы математическим методом главных компонент.....	26
2.4 Построение модели прогнозирования коэффициента эффективности по параметрам ключевой фразы нелинейным регрессионным анализом.....	32
3 ПРОЕКТИРОВАНИЕ ТЕХНОЛОГИИ ФОРМИРОВАНИЯ ЭФФЕКТИВНЫХ КЛЮЧЕВЫХ ФРАЗ	40
3.1 Проектирование диаграммы последовательностей модуля формирования эффективных ключевых фраз.....	40
3.2 Проектирование архитектуры модуля формирования эффективных ключевых фраз.....	41
3.3 Проектирование внутренней структуры модуля формирования эффективных ключевых фраз.....	43
3.4 Выбор средств реализации модуля формирования эффективных ключевых фраз.....	46

3.5 Реализация классов Model и Parameter.....	48
3.6 Реализация классов пакета MGen	51
3.7 Реализация управления модулем формирования эффективных ключевых фраз.....	52
ЗАКЛЮЧЕНИЕ	54
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ.....	56
ПРИЛОЖЕНИЕ А. Носитель с исходным кодом программы.....	59

ВВЕДЕНИЕ

Формирование ключевых фраз является одной из основных задач SEO оптимизации и продвижения веб-приложения в сети Интернет. Ключевые фразы должны четко и ёмко описывать тематику веб-ресурса и сферу товаров, услуг или контента, предоставляемых ресурсом.

Решения вопроса формирования эффективных ключевых фраз существуют у сервисов, направленных на поддержку и контроль рекламного аккаунта в системах контекстной рекламы. Разнообразные методы предлагают специалисты этой области: SEO оптимизаторы или директологи.

Однако, большинство решений либо требуют существенного понимания принципа работы контекстной рекламы от пользователя, либо полностью основаны на экспериментальном опыте и постепенном наращивании эффективных ключевых фраз.

Актуальность данной работы заключается в необходимости автоматизации формирования тематических и, при этом, наиболее эффективных, ключевых фраз для легкого вхождения неопытной аудитории сервиса «Rudvert» в контекстное рекламирование и построении или расширении рекламных кампаний в «Яндекс.Директ».

Новизна заключается в выполнении задачи оценки и прогнозирования эффективности ключевых фраз путем обработки и изучения их статистических и фактических параметров и построении модели оценки и прогнозирования на основе опытной выборки известных ключевых фраз.

Объектом исследования в данной работе является прогнозирование и оценка возможной эффективности отдельно взятых, тематически подобранных, ключевых фраз, используемых в рамках системы контекстной рекламы «Яндекс.Директ».

Предметом исследования является гипотеза возможности использования математических методов для прогнозирования и оценки эффективности ключевых фраз. Её основой является предположение о

существовании зависимостей между статистическими данными (олицетворяющими эффективность) и данными по ключевой фразе, доступными до её введения в работающую рекламную кампанию.

Целью работы является проверка данной гипотезы и разработка алгоритма прогнозирования эффективности фраз с последующей реализацией программного модуля на базе алгоритма, работающего в рамках сервиса контроля и управления рекламным аккаунтом «Яндекс.Директ» «Rudvert».

Выдвинуты следующие задачи:

1. Произвести анализ явления эффективности ключевой фразы, в рамках системы контекстной рекламы «Яндекс.Директ».
2. Разработать алгоритм прогнозирования эффективности ключевой фразы и проверить эффективность прогнозирования.
3. Реализовать на основе алгоритма модуль, обслуживающий сервис «Rudvert» с позиции формирования эффективных ключевых фраз.

Данная работа состоит из трех глав.

В первой главе рассматриваются теоретические основы контекстной рекламы, предметная область в виде определения понятия ключевой фразы как объекта системы «Яндекс.Директ» и сеть сервиса «Rudvert». Также производится анализ явления эффективности ключевой фразы и её параметров, предположительно способных влиять на эффективность. В конце формируются требования к новой реализуемой технологии формирования ключевых фраз.

Во второй главе рассматривается выбор математического аппарата и разработка алгоритма, на его основе, для построения моделей вычисления и прогнозирования коэффициента эффективности ключевой фразы.

В третьей главе рассматривается реализация программного модуля. Его внутренняя архитектура и способ взаимодействия с сервисом «Rudvert».

1 АНАЛИЗ ОСНОВ КОНТЕКСТНОГО РЕКЛАМИРОВАНИЯ И ЭФФЕКТИВНОСТИ КЛЮЧЕВОЙ ФРАЗЫ В РАМКАХ СИСТЕМЫ «ЯНДЕКС.ДИРЕКТ»

1.1 Анализ реализации контекстной рекламы на примере системы «Яндекс.Директ»

Контекстная реклама является типом интернет-рекламы, где рекламный материал демонстрируется пользователю в соответствии с содержанием или контекстом рекламируемого ресурса [1].

Для определения соответствия рекламного материала площадке или запросу обычно используется принцип ключевых фраз. Ключевые фразы являются ориентиром и для поисковых систем [1].

Все известные поисковые машины используют системы контекстной рекламы («Яндекс.Директ» [19], «Google AdWords» [18]). Они позволяют размещать рекламу на страницах с результатами поиска по определённым ключевым фразам, а также на площадках, установивших блоки контекстной рекламы на своих страницах или в мобильных приложениях.

Одной из основных проблем использования контекстной рекламы с позиции рекламодателя является достаточная сложность её структуры (многоуровневые объекты: кампании, группы, объявления и т.д.) и принципов работы (рекламный аукцион, анализ эффективности по конверсии и CTR и т.д.), относительно других распространителей рекламного материала.

Так, на примере системы «Яндекс.Директ», выглядит базовая структура рекламного аккаунта (рисунок 1.1).

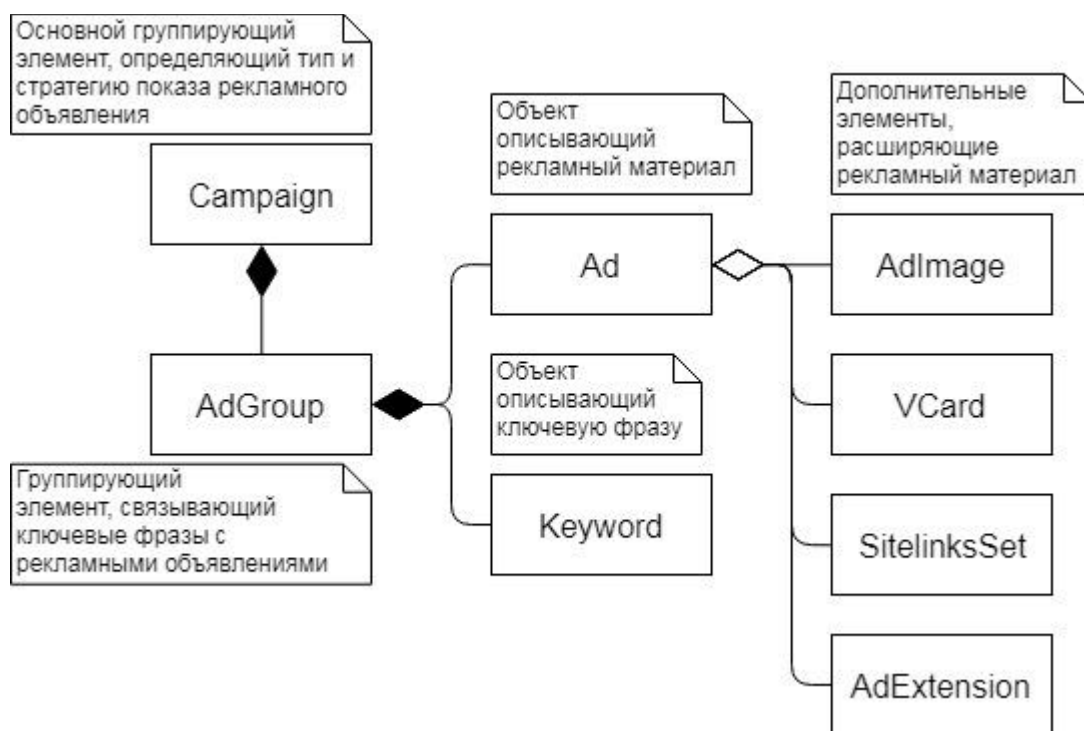


Рисунок 1.1 – Основные объекты в системе контекстной рекламы
«Яндекс.Директ»

При этом основой рекламного аккаунта, без которой рекламирования при помощи сервиса «Яндекс.Директ» просто невозможно, являются следующие объекты:

1. Campaigns – рекламные кампании, используемые для основной группировки рекламирования, обычно используемые для разделения по регионам показа или по брендам, или типам предлагаемых товаров или услуг.
2. AdGroups – группы объявлений, связывающие объявления и ключевые фраз, по которым данные объявления будут показываться.
3. Keywords – ключевые фразы, а также статистика и прочие данные на них в рамках данной группы объявлений.
4. Ads – объявление, являющиеся рекламным материалом, состоящим, как минимум, из заголовка, текста описания рекламируемого товара или услуги и ссылки на ресурс рекламодателя.

Совокупность данных объектов в связке называют (структурированным) семантическом ядром рекламного аккаунта. Данная

основа рекламного аккаунта распространена и встречается не только в системе «Яндекс.Директ» [5].

Разработка семантического ядра подразумевает формирование корректной и эффективной базы ключевых фраз и их распределение в состав групп и кампании.

На данный момент, существует проблема простого вхождения новых пользователей в сферу контекстного рекламирования. Необходимость рекламирования продукта или услуги является необходимостью для предпринимателя [26]. При этом бюджет и сроки, определенные на рекламу, ограничены.

1.2 Анализ сервиса контроля и управления рекламным аккаунтом «Яндекс.Директ» «Rudvert»

«Rudvert» [24] – это веб-ресурс (сайт) разрабатываемый компанией ООО «Рудник», направленный на контроль и управление рекламным аккаунтом «Яндекс.Директ» пользователя. Предоставляет следующие услуги:

- быстрое вхождение и адаптация нового неопытного пользователя в контекстное рекламирование на площадках Яндекса;
- сбор статистики по рекламному аккаунту и её предоставление пользователю в удобном для него формате;
- предоставление более удобного и интерактивного клиентского интерфейса в сравнении со стандартным интерфейсом Яндекс;
- контроль расходования рекламного бюджета пользователя по индивидуальным настройкам;
- контроль ставок по ключевым фразам пользователя в рамках аукциона «Яндекс.Директ» по индивидуальным настройкам;
- расширение рекламного аккаунта при помощи специальных инструментов.

«Rudvert» разработан на базе CMS «Joomla!». В качестве СУБД используется MySQL и нереляционная высокопроизводительная СУБД

«Redis» (для хранения статистики по ключевым фразам). «Rudvert» написан на языке программирования PHP 7 и базируется на веб-сервере «Apache».

Для доступа к данным клиента «Яндекс.Директ», сервис «Rudvert» использует «Яндекс.Директ» API. Так, для использования функционала сервиса, пользователю необходимо дать доступ к аккаунту «Яндекс.Директ» в виде OAuth-токена, выдаваемого Яндексом со согласия зарегистрированного пользователя.

В рамках данной работы необходимо разработать алгоритм, позволяющий оценить эффективность ключевой фразы, посредством анализа уже имеющихся в базе сервиса ключевых фраз (со статистикой, демонстрирующей эффективность) и реализовать модуль для «Rudvert» на основе данного алгоритма.

1.3 Определение понятия ключевой фразы в рамках системы «Яндекс.Директ»

Прежде чем вывести определение ключевой фразы в рамках системы «Яндекс.Директ», выведем общее определение.

Ключевая фраза – это семантическая конструкция, направленная на краткое описание содержимого веб-ресурса. Ключевые фразы зачастую используются поисковыми машинами для определения тематики ресурса [1].

Ключевая фраза в контекстной рекламе определяет семантику поискового запроса, ожидаемого от пользователя или его истории (запросов), при котором будет демонстрироваться определённый рекламный материал [9].

Ключевая фраза в рамках системы «Яндекс.Директ» представляет из себя объект в составе групп объявлений. Объекты ключевых фраз поддерживаются сервисом Keywords «Яндекс.Директ» API 5 версии [9].

Кроме того, данные по ключевой фразе можно получить от других объектов в системе «Яндекс.Директ» API. Рисунок 1.2 представляет

диаграмму объектов, реализующих ключевые фразы, как элементы системы «Яндекс.Директ» [9].

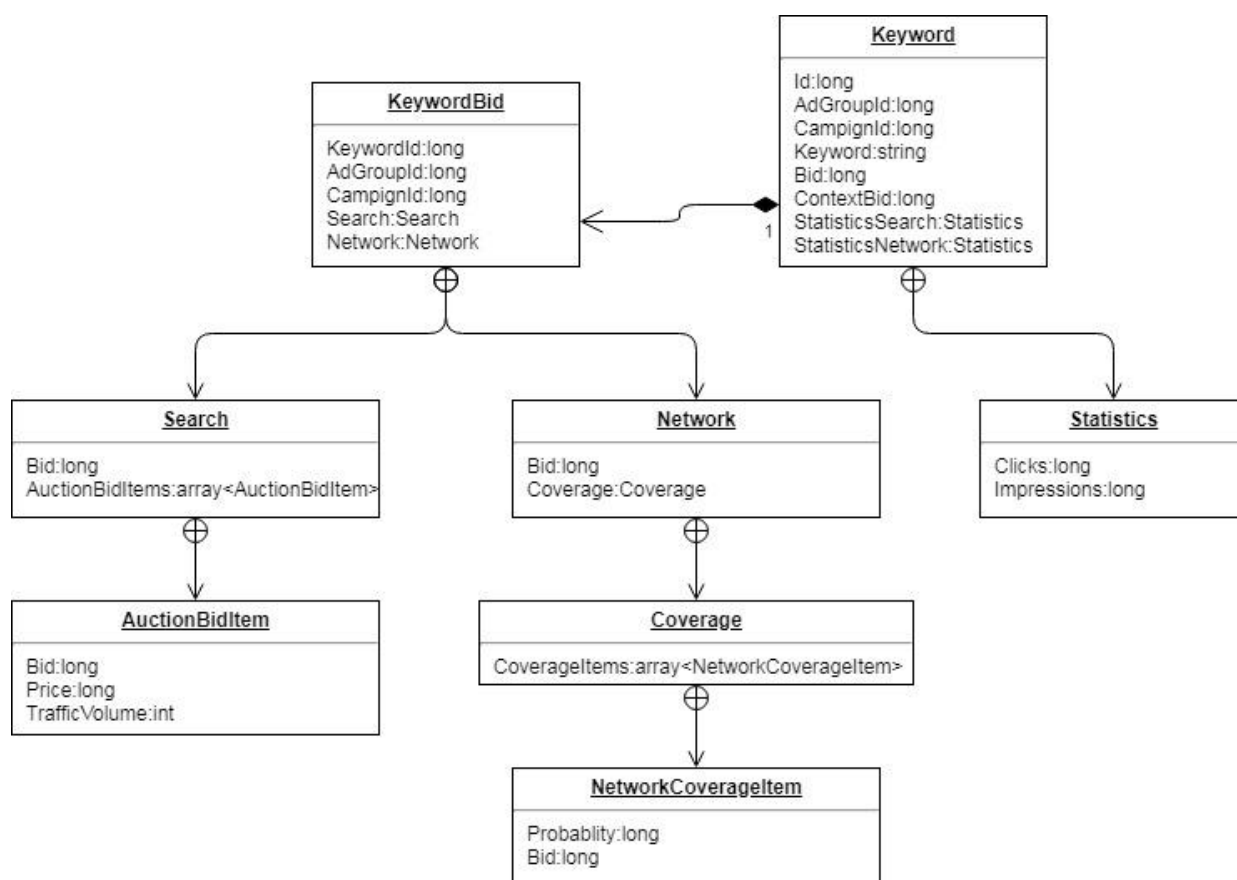


Рисунок 1.2 – Диаграмма объектов, определяющих ключевую фразу в системе «Яндекс.Директ»

В таблице 1.1 приведена спецификация диаграммы (рисунок 1.2).

Таблица 1.1 – Атрибуты объекта Keyword в системе «Яндекс.Директ»

Объект	Атрибут	Значение
Keyword	long Id, AdGroupId, CampaignId	Идентификаторы ключевой фразы, её группы объявлений и рекламной кампании
	string Keyword	Ключевая фраза (семантическая конструкция).
	long Bid, ContextBid	Установленная ставки для показа по фразе на Поиске (ленте поисковых ответов Яндекс) и РСЯ (рекламных сетях Яндекса).
	Statistics StatisticsSearch, StatisticsNetwork	Статистика показов и кликов всех объявлений группы по данной ключевой фразе на Поиске и РСЯ.

Объект	Атрибут	Значение
Statistics	long Impressions	Количество показов по данной ключевой фразе.
	long Clicks	Количество кликов по данной ключевой фразе.
KeywordBid	long KeywordId, AdGroupId, CampaignId	Дублирующие идентификаторы объекта Keyword.
	Search Search	Результаты аукционных торгов по ключевой фразе с учетом скидки Яндекса по Поиску.
	Network Network	Результаты аукционных торгов по ключевой фразе с учетом скидки Яндекса по РСЯ.
Search	long Bid	Ставка объекта Keyword.
	array<AuctionBidItem> AuctionBidItems	Ставки по аукциону Поиска «Яндекс.Директ»
Network	long Bid	Ставка объекта Keyword.
	array<CoverageNetworkItem> Coverage	Ставки по охвату РСЯ «Яндекс.Директ»
AuctionBidItem	long Bid	Рекомендуемая ставка для данной позиции.
	long Price	Минимальная ставка для данной позиции.
	int TrafficVolume	Индекс позиции на Поиске (1, 2, 3...8).
CoverageNetworkItem	Long Bid	Ставка для данного охвата РСЯ.
	long Probability	Процентный охват РСЯ (20, 50 или 100 %).

Ключевая фраза, существующая в рамках «Яндекс.Директ», представляет из себя разветвлённый объект, имеющий большое количество разнообразных статистических и фактических параметров.

В рамках данной работы ключевая фраза будет рассматриваться не как семантическая конструкция, а именно как элемент системы «Яндекс.Директ»

1.4 Анализ существующих решений в формировании ключевых фраз

Формирование тематических ключевых фраз под конкретную рекламную кампанию веб-ресурса может быть выполнен, как и вручную (оперируя опытом и знаниями в работе систем контекстной рекламы), так и посредством специального или схожего по задачам инструментария.

Как правило, оба варианта решения задачи формирования ключевых фраз используют средства, предоставляемые системами контекстной рекламы: Вордстат [23] (Яндекс) и Keyword Planner [21] (Google). Эти средства предназначены для получения вложенных поисковых запросов и статистики по конкретной ключевой фразе.

Полноценные решения в формировании ключевых фраз имеются у всех систем контроля и управления рекламным аккаунтом («Топвизор» [20], «Простор» [17] и т.д.), но, очевидно, данные решения являются коммерческой тайной и не доступны для использования независимо от самого сервиса.

В качестве независимых от систем контроля и управления рекламным аккаунтом решений задачи формирования именно эффективных фраз рассмотрим следующие сервисы:

«Keyword Collector» [16] – программа-помощник для составления семантического ядра во всех популярных на русскоязычном рынке системах контекстной рекламы. Разработан и обеспечивается компанией «LegatoSoft».

«Keyword Tool» [22] – иностранный наиболее крупный сервис, предлагающий услуги формирования ключевых тематических фраз. Преподносит себя как альтернатива стандартным средствам. Разработан и обеспечивается компанией «Key Tools Limited».

Рассмотрим возможность использования данных решений (как альтернатива собственной разработки) для реализации модуля формирования эффективных ключевых фраз.

Критериями оценки рассматриваемых решений будут следующие параметры:

1. Возможность интеграции функционала в веб-приложение (3 балла), притом 0 баллов – это полное отсутствие какого-либо API, в то время как 3 балла – это наличие полноценного API, позволяющего использовать весь функционал.

2. Полнота данных, позволяющих оценить эффективность фраз (10 баллов), притом 0 баллов – это полное отсутствие данных, которые могут указывать на эффективность, в то время как 10 баллов – это явный числовой показатель эффективности.

3. Дешевизна реализации или приобретения технологии (5 баллов), а именно, чем дешевле реализация, тем выше балл оценки.

4. Сравнительный анализ представлен в таблице 1.2.

Таблица 1.2 – Сравнительный анализ выбора решения задачи формирования ключевых фраз

Критерий оценки	Keyword Collector	Keyword Tool	Собственная разработка
Возможность интеграции функционала (0 - 3)	0	2	3
Полнота данных, позволяющих оценить эффективность фраз (0 – 10)	9	4	10
Дешевизна реализации или приобретения технологии (0 - 5)	4	0	3
Итого	13	6	16

На основе таблицы 1.2 делаем вывод, что известные готовые решения не удовлетворяют минимальным требованиям, что говорит о необходимости реализации собственного решения.

1.5 Определение показателей эффективности ключевой фразы

Эффективность рекламной кампании в контекстной рекламе определяется её окупаемостью и прибылью, получаемой от посетителей рекламируемого ресурса, попавших на него посредством рекламных

объявлений, и совершивших указанное рекламодателем целевое действие (покупка товара, прочтение статьи, оформление подписки и т.д.).

Ценой цели считается отношение затраченного на клики по рекламным объявлениям бюджета к количеству выполненных целевых действий посетителями [9].

Среднее от вырученных с привлечённых пользователей выгоды к цене цели и есть главный показатель эффективности всей рекламной кампании.

Анализ эффективности рекламной кампании целиком требует наличие ресурса рекламодача с установленным на нем счетчиком метрики системы контекстной рекламы, отслеживающем поведение пользователя и фиксирующего выполнение целевого действия [9].

С позиции семантического ядра эффективность может быть определена демонстрацией рекламного объявления наибольшему количеству потенциально заинтересованной аудитории, что повышает шанс выполнения целевого действия. Таблица 1.3. содержит статистические данные об рекламном аккаунте, демонстрирующие эффективность его семантического ядра «Яндекс.Директ».

Таблица 1.3. – Данные, демонстрирующие эффективность семантического ядра рекламного аккаунта «Яндекс.Директ»

Параметр	Обозначение	Определение
Количество показов	Shows (Sh_{kw})	Количество демонстраций пользователю рекламных объявлений в ленте результатов поиска или в сетях распространителя рекламы, вызванных определенной ключевой фразой.
Количество кликов	Clicks (Cl_{kw})	Количество переходов на рекламируемый ресурс, совершенных пользователями через рекламные объявления, показанных по определенной ключевой фразой.
Кликабельность	CTR (CTR_{kw})	Является процентным соотношением кликов к показам по определенной ключевой фразе. Рассчитывается по формуле: $CTR_{kw} = \frac{Cl_{kw}}{Sh_{kw}} \cdot 100\%$
Средняя	MidBid	Средняя цена торгов за «усредненную»

Параметр	Обозначение	Определение
стоимость клика	(MB_{kw})	позицию на Поиске и РСЯ для одного клика по рекламному объявлению. Является денежной единицей в валюте рекламодателя.

Эти данные доступны для базовых элементов рекламного аккаунта рекурсивным способом. Так количество показов рекламной кампании имеет значение суммы показов групп объявлений в составе данной кампании, которые получают своё значение от показов вложенных в них ключевых фраз [9]. Так, истинным источником и владельцем данных параметров является ключевая фраза.

Параметры, приведенные таблицей 1.3 достаточно точно демонстрируют эффективность фразы только в связке друг с другом. Для примера этого достаточно представить ключевую фразу с одним показом объявления, по которому пользователь тут же перешел на ресурс рекламодателя.

Параметр CTR в такой ситуации будет равен 100% (один клик к одному показу), в то время, как у действительно эффективных фраз он не достигает и 40%. Подобную аналогию можно привести с любым параметром.

Учитывая это, для корректной оценки эффективности ключевой фразы необходимо построение нового параметра. В рамках данной работы данный параметр будет именоваться коэффициентом эффективности ключевой фразы (Productive или pc_{kw}) и будет вычисляться с учетом зависимостей параметров, влияющих на эффективность ключевой фразы. Результирующий численный параметр позволит более явно и просто производить оценку эффективности ключевой фразы в системе «Яндекс.Директ».

1.6 Определение признаков эффективности ключевой фразы

Поскольку точно не известно, какие именно свойства ключевой фразы влияют на её эффективность в контекстном рекламировании необходимо

изучить и выявить корреляцию к эффективности наибольшего количества параметров.

Параметры необходимо брать из числа общедоступных, чтобы выполнение прогнозирования не требовало поиска дополнительной информации по ней.

Таковыми можно считать те параметры, которые можно выделить из самой ключевой фразы, как семантической конструкции, а также параметры, получаемые через «Яндекс.Директ».

Из последнего, через доступ к сервису Вордстат по «Яндекс.Директ» API, возможно получить данные о количестве поисковых запросов за последний месяц по любой фразе.

Рассмотрим и подберем наиболее очевидные параметры, теоретически влияющие на эффективность ключевой фразы в рамках контекстной рекламы.

Таблица 1.4 содержит список тех параметров, которые будут проверяться на коррелируемость с эффективностью.

Таблица 1.4. – Возможные признаки эффективности ключевых фраз

Параметр	Обозначение	Предположительная зависимость
Посимвольная длина фразы	Length (L_{kw})	Конкретизация или упрощение ключевой фразы влияет одновременно на популярность и охват заинтересованной аудитории
Количество поисковых запросов за последний месяц	Wordstat (W_{kw})	Очевидное влияние на популярность

Отметим, что все параметры, представленные в таблице 1.4, являются очевидно, численными. В случае, если параметры, описанные в таблице 1.4 не влияют на эффективность ключевой фразы или влияют недостаточно сильно, необходимо продолжить поиск новых и, обязательно, численных параметров и источников параметров для любой произвольной фразы.

1.7 Выработка требований к новой технологии формирования ключевых фраз

Для разработки требований к новой технологии выбрана система классификации требований FURPS+. Данная система предполагает функциональные (то, что система должна делать) и нефункциональные (ограничение, накладываемые на систему) требования [6]. Включает в себя:

1. Требование к удобству использования (Usability).
2. Требование к надежности (Reliability).
3. Требование к производительности (Performance).
4. Требование к поддержке (Supportability).
5. Требование к проектированию (Design requirements).
6. Требование к реализации (Implementation requirements).
7. Требование к интерфейсу (Interface requirements).
8. Физические требования (Physical requirements).

Из представленного списка не будут рассматриваться: требования к удобству использования (клиентом системы является другая система), требования к реализации и физические требования (отсутствие жестких требований).

Описание требований представлено на таблице 1.5.

Таблица 1.5 – Требования к новой технологии по классификации FURPS+

Требование	Полезность	Обоснование
Функциональные требования		
Выполнение прогнозирования и формирование отчета в виде оценок предполагаемой эффективности по, полученным в запросе от сервиса, ключевым фразам.	Критичное	Основная задача модуля – обслуживание сервиса «Rudvert», в виде прогнозирования эффективности ключевых фраз.
Выполнение по запросу сервиса самообслуживания, в виде пересчета математических моделей, используемых для оценки	Важное	Необходим пересчет используемых математических моделей с учетом обновленных статистических данных для

Требование	Полезность	Обоснование
прогнозирования эффективности ключевых фраз по новым данным.		повышения точности прогнозирования.
Требования к надежности		
Введение полных отчетов не только о выполнении программы, но и хода вычислений при пересчете математических моделей.	Важное	Способствует быстрому устранению ошибок.
Требование к производительности		
Быстрое выполнение прогнозирования и формирование отчета.	Критичное	Прогнозирование эффективности новых ключевых фраз происходит в момент запроса пользователя. Это значит, что пользователь в реальном времени ожидает выполнения запроса.
Требование к поддержке		
Добавление и удаление используемых признаков эффективности (программно)	Полезное	Хоть модуль не предполагает графического интерфейса, необходима возможность с минимумом усилий добавить или удалить используемый в прогнозировании параметр ключевой фразы.
Требование к проектированию		
Независимость частей модуля самообслуживания и обработки запросов сервиса.	Критично	Разделение двух составляющие модуля позволит тестировать или обрабатывать ошибки части самообслуживания, не прекращая обработки запросов сервиса.
Требование к интерфейсу		
Работа по протоколу HTTP	Критично	Необходимо для возможности расширения использования модуля для сторонних, помимо сервиса, клиентов.
Блокировка входной точки при помощи токена	Критичное	Ограничение доступа к функционалу модуля.

На основе анализа требований необходимо реализовать 2 функциональных требования и 6 нефункциональных ограничений, представленных в таблице 1.3.

1.8 Постановка задачи на разработку модуля формирования ключевых фраз

Целью данной работы является реализация модуля для сервиса контроля и управления рекламным аккаунтом «Яндекс.Директ» «Rudvert». Данный модуль отвечает за оценку предположительной эффективности, полученных в запросе сервиса, ключевых фраз. Модуль должен отвечать всем требованиям, выдвинутым в разделе 1.7 по системе требований FURPS+.

Предлагается следующая последовательность выполнения задачи разработки программного модуля формирования ключевых фраз:

1. Выбрать математический аппарат, способный реализовать вычисление единого численного параметра эффективности ключевой фразы по её параметрам, в определенной мере демонстрирующим эффективность.
2. Выбрать математический аппарат, способный к прогнозированию одного параметра (коэффициент эффективности) ключевой фразы по другим, влияющим на вычисляемый, параметрам. При этом выбранный метод прогнозирования должен корректно работать с нелинейной зависимостью влияющих параметров.
3. Используя выбранный математический аппарат, разработать алгоритмы вычисления коэффициента эффективности для известных фраз и последующего его прогнозирования для новых ключевых фраз.
4. На основе полученных вычислений необходимо произвести проверку надежности разработанных алгоритмов. В случае больших ошибок пересмотреть выбранный математический аппарат и выбранные используемые данные.

5. Реализовать модуль на базе разработанных алгоритмов, начиная с проектирования архитектуры реализуемого модуля.
6. Выбрать технологии, используемые для реализации модуля.
7. Спроектировать систему взаимодействия модуля с сервисом «Rudvert».
8. Конкретизировать архитектуру проработкой классов, их методов и свойств.

При построении модуля необходимо учитывать требования проектирования, а именно разделить модуль на две независимые составляющие: реализация математического аппарата и реализация управления модулем.

На основе результатов необходимо выдвинуть итог в виде рассмотрения недостатков и преимуществ реализованного модуля и подытожить рассмотрения выдвинутой гипотезы возможности использования математических методов для прогнозирования эффективности фразы в контекстной рекламе.

2 РАЗРАБОТКА АЛГОРИТМА ВЫЧИСЛЕНИЯ И ПРОГНОЗИРОВАНИЯ КОЭФФИЦИЕНТА ЭФФЕКТИВНОСТИ КЛЮЧЕВОЙ ФРАЗЫ

2.1 Выбор математического аппарата для вычисления коэффициента эффективности ключевой фразы

Реализация вычисления и последующего прогнозирования коэффициента эффективности ключевой фразы необходим математический аппарат, отвечающий всем условиям и не конфликтующий с математической составляющей параметров, используемых при выполнении вычислений.

В первую очередь необходимо обозначить задачи, которые должен выполнять выбираемый математический аппарат.

Для вычисления коэффициента эффективности необходимо объединение рассмотренных ранее параметров ключевой фразы (Shows, Clicks, CTR, MidBid) с учетом их влияния друг на друга. Данная задача может быть рассмотрена как уменьшение размерности данных с потерей наименьшего количества информации.

Для уменьшения размерности данных наиболее популярными считаются следующие методы:

- метод главных компонент (principal component analysis, PCA) — один из основных способов уменьшить размерность через вычисление сингулярного разложения матрицы данных или вычисление собственных векторов значений ковариационной матрицы исходных данных [3];
- многомерное шкалирование — метод анализа и визуализации данных с помощью расположения точек, соответствующих изучаемым (шкалируемым) объектам, в пространстве меньшей размерности чем пространство признаков объектов [3].

Проведем сравнительный анализ данных методов с позиции реализуемой задачи по следующим критериям:

1. Простота освоения (10 баллов) включает в себя оценку объема доступной информации и затратность освоения всех аспектов использования, при этом чем выше балл, тем легче освоение.

2. Требовательность к вычислительной машине (10 баллов) является оценкой реализованного программного метода с позиции затратности на вычисление, при этом чем выше балл, тем метод менее затратен.

Результат сравнительного анализа представлен в таблице 2.1.

Таблица 2.1 – Результат сравнительного анализа методов уменьшения размерности данных

Критерий оценки	Метод главных компонент	Многомерное шкалирование
Простота освоения	8	5
Требовательность к вычислительной машине	5	2
Итого	13	7

Как видно по таблице 2.1 наилучшим решением будет выбор метода главных компонент. Стоит отметить, что нелинейные методы (метод упругих карт, поиск наилучшей проекции и т.д.) не рассматривались, так как изучаемые параметры относятся к эффективности линейно, а это означает очевидную избыточность данных методов.

2.2 Выбор математического аппарата для прогнозирования коэффициента эффективности ключевой фразы

Прогнозирование коэффициента эффективности по, рассмотренным ранее признакам, очевидно, является задачей математического прогнозирования.

В первую очередь, необходимо определить метод и модель прогнозирования.

Метод прогнозирования представляет собой последовательность действий, которые нужно совершить для получения модели прогнозирования [4].

Модель прогнозирования есть функциональное представление, адекватно описывающее исследуемый процесс и являющееся основой для получения его будущих значений [4].

В рамках данной задачи выбран форматизированный метод, по причине очевидной невозможности автоматизации интуитивного.

Модели предметной области используют законы предметной области, в то время как модель временных рядов стремятся найти зависимость будущего от прошлого внутри самого процесса [4]. В рамках выполняемой задачи необходим метод временных рядов, поскольку влияние выбранных признаков на изучаемый параметр точно неизвестно.

В статистических моделях зависимость задается уравнением, в то время, как в структурных некоторой структурой и правилами перехода по ней [2]. В рамках данной задачи удобнее использовать статистическую модель в связи с простой реализации прогнозирования вычислением коэффициента подстановкой признаков в заранее выработанное уравнение.

Помимо этого, выбираемая модель прогнозирования должна корректно работать с нелинейными параметрами, так как используемые признаки нелинейны относительно эффективности. Также модель должна допускать возможность работы с множеством признаков (быть множественной).

Удовлетворяют всем указанным выше требованиям модели регрессионного анализа и экспоненциального сглаживания.

Критерии выбора модели прогнозирования аналогичны выдвинутым ранее (в разделе 2.1):

1. Простота освоения (10 баллов).
2. Требовательность к вычислительной машине (10 баллов).

Результат сравнительного анализа представлен в таблице 2.2.

Таблица 2.2 – Результат сравнительного анализа моделей прогнозирования

Критерий оценки	Регрессия	Экспоненциальное сглаживание
Простота освоения	5	5
Требовательность к вычислительной машине	7	5
Итого	13	10

Как видно по таблице 2.2 наилучшим решением будет выбор модели (нелинейного множественного) регрессионного анализа для прогнозирования коэффициента эффективности ключевой фразы.

2.3 Построение модели вычисления коэффициента эффективности ключевой фразы математическим методом главных компонент

Метод главных компонент — один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Изобретён Карлом Пирсоном в 1901 году. Применяется во многих областях, в том числе, в эконометрике, биоинформатике, обработке изображений и сжатии данных [28].

Суть метода заключается в вычислении главных компонент меньшей размерности, чем исходные переменные и выражается он формулой (1).

$$pc = \sum_{i=1}^n c_i y_i \quad (1)$$

где pc — главный компонент, рассчитанный для определенного объекта выборки и проецирующий значения признаков на меньшую размерность;

c_i — заранее рассчитанный вес (или коэффициент) признака (или объясняющей переменной) i , определяющий влияние на главную компоненту pc ;

y_i — значение признака i для определённого объекта.

Вычисление коэффициента эффективности ключевой фразы (как главной компоненты) будет заключаться в расчете степени влияния

статистических параметров (как весов признаков) и формирования формулы, подобной (1).

Вычисление главных компонент может быть сведено к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных.

Выведем ход действий, реализующие вычисление коэффициента эффективности методом главных компонент. На рисунке 2.1 представлена блок-схема построения модели методом главных компонент.

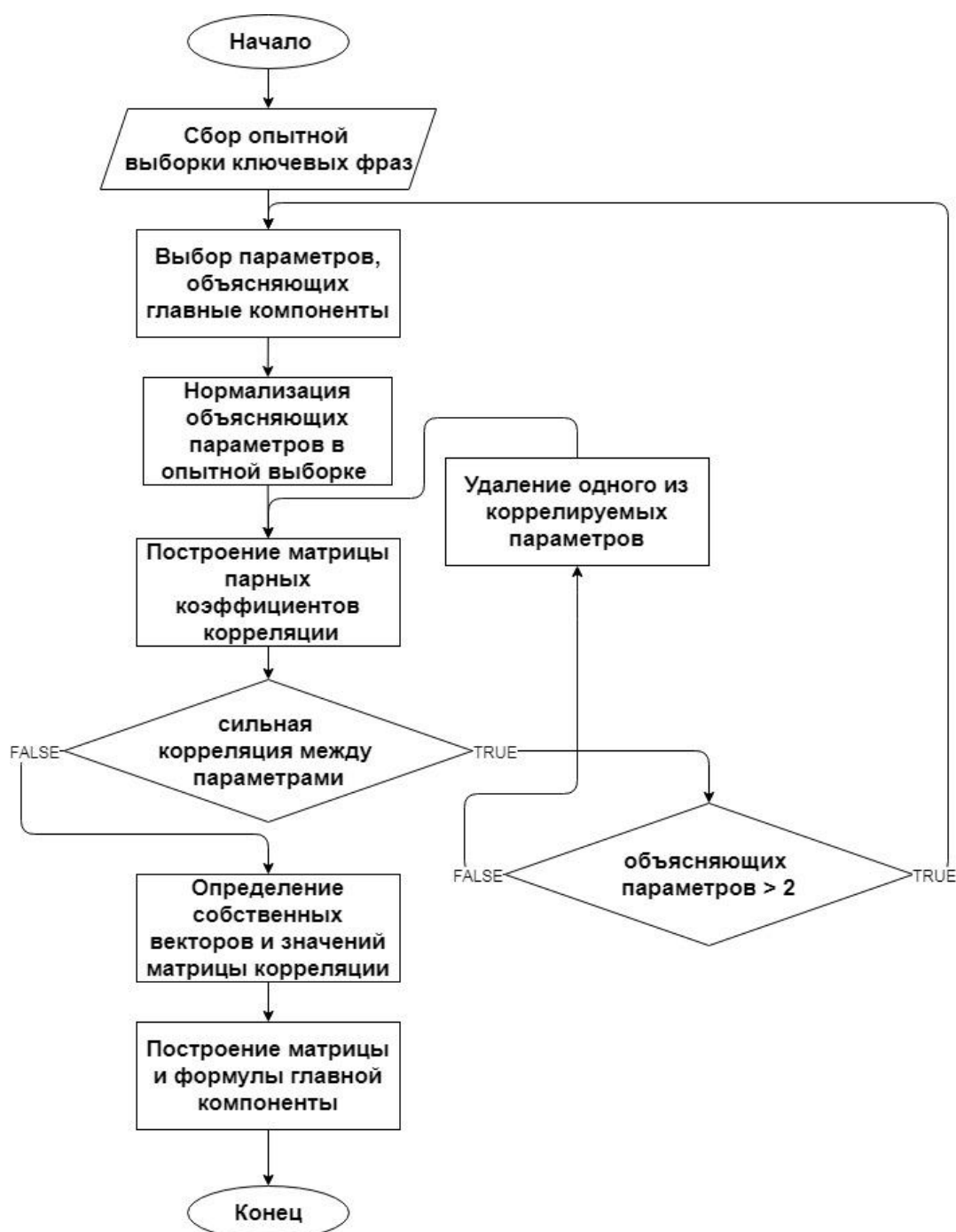


Рисунок 2.1 – Блок-схема построения модели вычисления коэффициента эффективности ключевой фразы

В качестве опытной выборки взята постоянно пополняемая база ключевых фраз сервиса «Rudvert». В выборку входят, не нулевые (имеющие хоть сколько-нибудь показов и кликов) в количестве двухсот, ключевые фразы разных клиентов и разного качества аккаунтов. Параметрами, определяющими коэффициент эффективности, выбраны Shows, Clicks, MidBid (CTR априори коррелируется с Shows и Clicks).

Далее необходима подготовка данных: центрирование и стандартизация (нормализация) данных, производимая для всех параметров каждому объекту выборки по формуле (2), что означает переход к новой величине, для которой математическое отклонение равно нулю, а дисперсия – единице [8].

$$Y = \frac{y - \bar{y}}{\sigma} \quad (2)$$

где Y – новое нормализованное значение параметра для определённого объекта;

y – исходное значение параметра для определённого объекта;

$\bar{y}$ – математическое ожидание распределения параметра по всей опытной выборке объектов;

σ – стандартное отклонение распределения параметра по всей опытной выборке объектов.

Следующим шагом является определение матрицы парных коэффициентов корреляции между изучаемыми параметрами. При этом необходима именно редуцированная матрица.

Редуцированная матрица – это матрица парных коэффициентов корреляции, на главной диагонали которой расположены не единицы полной корреляции или оценки полной дисперсии, а их редуцированные, несколько уменьшенные величины – значения оценок общностей [2].

Для построения редуцированной матрицы необходимо построить обычную корреляционную матрицу, а потом заполнить диагональ матрицы по формуле (3).

$$r_j = \frac{1}{n} \sum_{i=1}^n r_{ij} \quad (3)$$

где r_j – диагональное значение матрицы (j диагональ);

n – размер матрицы (матрица обязана быть квадратной);

r_{ij} – значение ячейки ij матрицы.

Таблица 2.3 содержит вычисленную для выборки ключевых фраз редуцированную матрицу парных корреляционных коэффициентов изучаемых параметров.

Таблица 2.3 – Редуцированная матрица парных коэффициентов корреляции между изучаемыми параметрами

	Shows	Clicks	MidBid
Shows	0,22	0,42	0,02
Clicks	0,42	0,24	0,07
MidBid	0,02	0,07	0,04

На данном этапе необходимо проверить тесноту линейной стохастической связи между признаками исследуемых объектов. Считается, что корреляция от 80% и выше указывает на необходимость замены, как минимум, одного признака из пары. В данном примере по таблицу 2.3 видно, что наиболее сильная корреляция (между Clicks и MidBid) составляет не более 70%, а это значит, что данные параметры можно использовать.

Далее требуется определить собственные значения и вектора для матрицы парных коэффициентов корреляции (таблица 2.3).

Собственным значением (или характеристическим числом) квадратной матрицы R порядка m называют такое число λ , что бы можно было подобрать такой m -мерный ненулевой собственный вектор u , что $R_u = \lambda_u$ [9].

Собственный вектор, в свою очередь – это ненулевой вектор, который при умножении на некоторую квадратную матрицу превращается в самого же себя с числовым коэффициентом (собственным значением) [9].

Составляем систему уравнений для определения координат собственных векторов (4).

$$\begin{aligned} 0,22 - \lambda x_1 + 0,42x_2 + 0,02x_3 &= 0 \\ 0,42x_1 + 0,24 - \lambda x_2 + 0,07x_3 &= 0 \\ 0,02x_1 + 0,07x_2 + 0,04 - \lambda x_3 &= 0 \end{aligned} \quad (4)$$

где λ – собственное значение определяемого собственного вектора для данной матрицы;

x_1, x_2, x_3 – переменные, определяющие собственный вектор по λ .

Помимо этого, составляем характеристическое уравнение матрицы (5).

$$\begin{vmatrix} 0,22 - \lambda & 0,42 & 0,02 \\ 0,42 & 0,24 - \lambda & 0,07 \\ 0,02 & 0,07 & 0,04 - \lambda \end{vmatrix} = 0 \quad (5)$$

Находим определитель матрицы (5), приравниваем полученное выражение к нулю (6).

$$\begin{aligned} &0,22 - \lambda \cdot 0,24 - \lambda \cdot 0,04 - \lambda - 0,07 \cdot 0,07 - \\ &- 0,42 \cdot 0,42 \cdot 0,04 - \lambda - 0,07 \cdot 0,02 + \\ &+ 0,02 \cdot (0,42 \cdot 0,07 - (0,24 - \lambda) \cdot 0,02) = 0 \end{aligned} \quad (6)$$

И преобразуем его в нормальный вид (7)

$$-\lambda^3 + 0,5 \cdot \lambda^2 + 0,1105 \cdot \lambda - 0,0049 = 0 \quad (7)$$

Корнями уравнения (4) являются значения λ , равные 0.61, -0.23 и 0.03. Зная их можно получить все три собственных вектора данной матрицы (таблица 2.3), но в этом нет необходимости, поскольку наилучший собственный вектор, с позиции описания наиболее приближенной к исходным данным главной компоненте, считается вектор с наибольшим собственным значением [28] (вектор, построенный на λ , равным 0.61).

Финальным шагом является формирования формулы главного компонента. Для построения матрицы коэффициентов (весов) параметров по главной компоненте используют формулу (8) [7].

$$\bar{\alpha} = \bar{\lambda} \cdot \bar{u} \quad (8)$$

где $\bar{\alpha}$ – коэффициент влияния определённого параметра на главный компонент;

$\bar{u}$ – значения определённого параметра в выбранном собственном векторе редуцированной матрицы;

λ – выбранное собственное значение от собственного вектора.

В результате построения по формуле (8) получаем матрицу главной компоненты (9).

$$A = \begin{matrix} 5,06 \\ 4,99 \\ 0,78 \end{matrix} \quad (9)$$

Итогом, через стандартную формулировку метода главных компонент – формулу (1) с подстановкой матрицы (9) получаем формулу вычисления единого численного коэффициента эффективности ключевой фразы (10).

$$pc_{kw} = 5,06 \cdot Sh_{kw} + 4,99 \cdot Cl_{kw} + 0,78 \cdot (MAX MB_{kw} - MB_{kw}) \quad (10)$$

где pc_{kw} – коэффициент эффективности ключевой фразы;

Sh_{kw} – количество показов по ключевой фразе;

MB_{kw} – средняя стоимость клика по ключевой фразе;

$MAX(MB_{kw})$ – наибольшее значение MB_{kw} в выборке ключевых фраз.

В качестве примера представлена таблица 2.4, содержащая (ненормализованные) данные и рассчитанный, при помощи полученной математической модели, коэффициент эффективности для шести случайных фраз, взятых из выборки.

Таблица 2.4 – Пример использования метода главных компонент

	Ключевая фраза	Clicks	Shows	MidBid	Productive
--	----------------	--------	-------	--------	------------

	Ключевая фраза	Clicks	Shows	MidBid	Productive
1	номера купить красивые +на автомобиль	31	37632	23,5	9,5
2	купить камаз сельхозник +с прицепом	11	10493	266,3	-4,7
3	электромонтаж официальный сайт адреса магазинов	1	2635	19,4	-6,7
4	российские алюминиевые лодки купить	127	19416	67,2	12,3
5	купить эвакуатор	11	5596	22,4	-4,6
6	купить печь +для бани	3	5071	25	-5,6

По примеру (таблица 2.4) видно, что полученный коэффициент демонстрирует достаточную оценку эффективности ключевой фразы. Так, самой эффективной (из предложенных в таблице) будет фраза 4, что видно по большому количеству кликов к относительно малой ставке (в отличии от фразы 3 с большой ставкой и малым количеством кликов).

Вычислив коэффициент эффективности для всей выборки и отсортировав её по убыванию можно пронаблюдать действенность коэффициента. Первыми остались ключевые фразы с хорошим количеством кликов и показов, в то время как последними оказались фразы, набравшие не больше десятка показов и единицы кликов.

2.4 Построение модели прогнозирования коэффициента эффективности по параметрам ключевой фразы нелинейным регрессионным анализом

Регрессионный анализ — статистический метод исследования влияния одной или нескольких независимых переменных на зависимую. Независимые переменные иначе называют регрессорами или объясняющими переменными, а зависимые переменные — откликом или объясняемая переменная [25].

Терминология зависимых и независимых переменных отражает лишь математическую зависимость переменных, а не причинно-следственные отношения [25].

Для прогнозирования, рассчитанного методом главных компонент, коэффициента эффективности по выбранным доступным параметрам построим нелинейную множественную регрессионную модель.

Нелинейная множественная регрессионная модель описывается формулой (11) [25].

$$y = a + \sum_{i=1}^n F_i(x_i) \quad (11)$$

где y – прогнозируемый параметр (отклик) объекта выборки;

$F_i x_i$ – функция от значения i -признака (регрессор);

a – свободный (от регрессоров) член регрессии;

n – количество регрессоров.

Построим ход действий для построения нелинейной множественной регрессионной модели (рисунок 2.2).



Рисунок 2.2 – Блок-схема алгоритма построения модели прогнозирования коэффициента эффективности ключевой фразы

Произведем анализ линейности функций регрессоров по опытной выборке по графикам (облаку) распределения регрессоров и отклика.

Рисунок 2.3 демонстрирует график распределения параметра Wordstat к вычисленному коэффициенту эффективности.

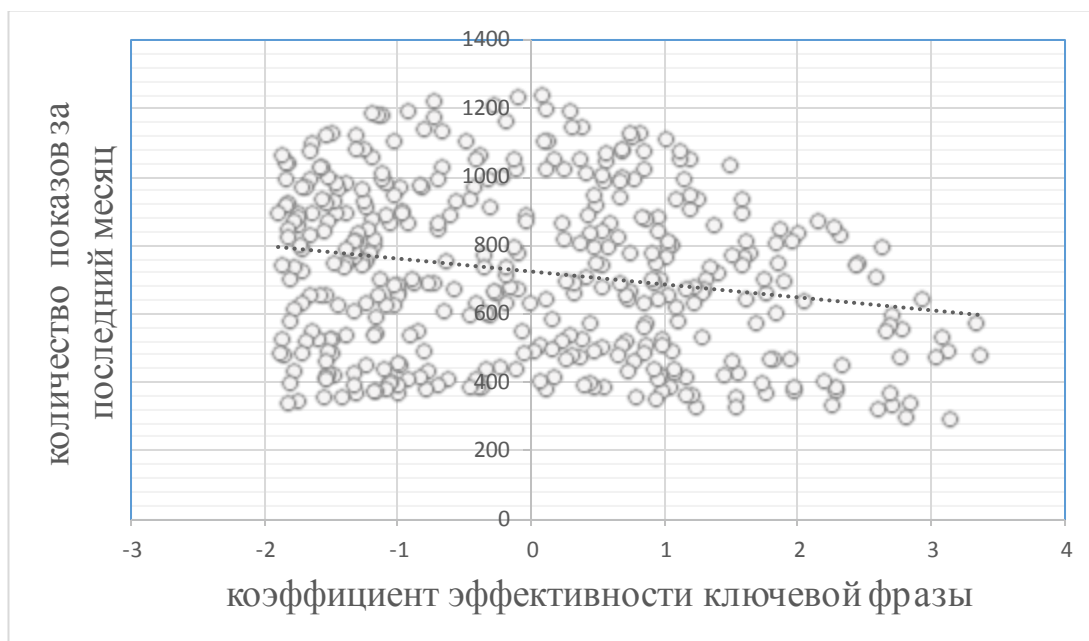


Рисунок 2.3 – График распределения параметра Wordstat к коэффициенту эффективности для ключевых фраз в выборке

График (рисунок 2.3) демонстрирует нелинейность параметра с одним экстремумом. И это не удивительно так как, низкой эффективностью обладают ключевые фраз либо совершенно непопулярные, либо слишком обобщающие и захватывающие слишком большую аудиторию.

Аналогичная ситуация и со вторым параметром Length (рисунок 2.4).

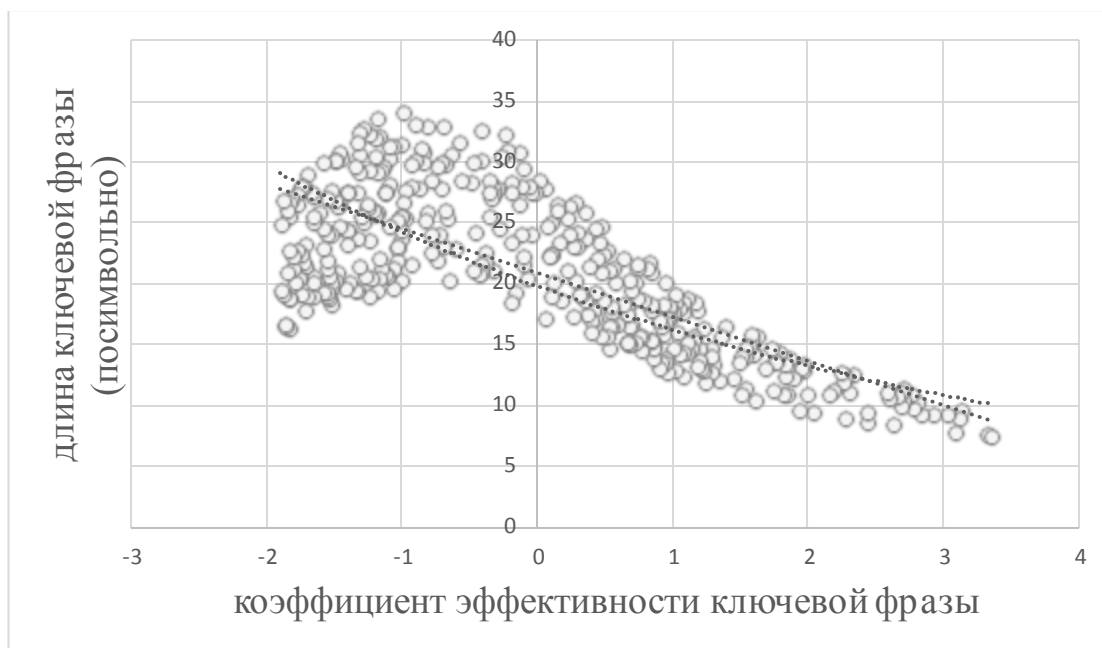


Рисунок 2.4 – График распределения параметра Length к коэффициенту эффективности для ключевых фраз в выборке

Визуальная оценка графиков используемых регрессоров позволяет раскрыть функции от них через параболическую функцию (12) [25].

$$F(x) = a + bx^2 + cx \quad (12)$$

Опишем зависимость признаков к отклику формулой (13) [25].

$$y = a + \sum_{i=1}^n b_i x_i^2 + c_i x_i \quad (13)$$

где y – прогнозируемый параметр (отклик) для определённого объекта выборки;

b_i, c_i – веса i - регрессора в разных степенях;

x_i – значение параметра i – регрессора для определённого объекта;

a – свободный (от регрессоров) член регрессии;

n – количество регрессоров ($n = 2$).

Далее приводим функции объясняющих переменных к линейному виду путем замены переменных. Теперь отклик линейно зависим от четырех регрессоров ($x_3 = x_1^2, x_4 = x_2^2$) [25].

Используем линейный метод наименьших квадратов. Согласно методу наименьших квадратов, параметры уравнения регрессии находятся по формуле (14) [27].

$$A = X^T \cdot X^{-1} \cdot X^T Z \quad (14)$$

где A – матрица коэффициентов регрессоров;

X – матрица регрессоров опытной выборки;

Z – матрица откликов опытной выборки.

Итоговую матрицу, полученную формулой (14), необходимо подставить в уравнение зависимости отклика (13). Полученное уравнение (15) позволяет рассчитать предсказанное значение коэффициента эффективности.\

$$pc_{kw} = 4,566 - 0,857 \cdot W_{kw} - 0,231 \cdot L_{kw} + \\ + 0,0042 \cdot W_{kw}^2 + 0,0024 \cdot L_{kw}^2 \quad (15)$$

где pc_{kw} – спрогнозированный рассчитанный коэффициент эффективности ключевой фразы;

W_{kw} – количество поисковых запросов за последний месяц по ключевой фразе;

L_{kw} – длина ключевой фразы в символах.

Таким образом, по формуле (15) можно рассчитать приблизительный коэффициент эффективности ключевой фразы, зная её символьную длину и месячное количество поисковых запросов.

Следует проверить результат регрессионного анализа. В первую очередь, для построения регрессионной статистики по модели, вычислим коэффициент детерминации формуле (16), представляющую из себя отношение объясненной суммы квадратов отклонений к необъясненной [2].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n y_i^2} \quad (16)$$

где R^2 – коэффициент детерминации полученной регрессионной модели;
 y_i – истинное значение прогнозируемого параметра i объекта опытной выборки;

$\hat{y}_i$ – спрогнозированное вычисленное значение прогнозируемого параметра i объекта опытной выборки;

$\bar{y}$ – математическое ожидание по прогнозируемому параметру в рамках опытной выборки;

Коэффициент детерминации демонстрирует процент данных, объясненных регрессионным уравнением. В рамках этой выборки $R^2 = 0,64$.

Далее рассчитаем множественный коэффициент корреляции, характеризующий тесноту связи между зависимой переменной и предиктором. Берется как корень из коэффициента детерминации. В данном случае множественный коэффициент корреляции $R = 0,80$, что показывает достаточную связь между опытными данными и расчетными [2].

Вычислим стандартную ошибку, являющуюся квадратным корнем из выборочной остаточной дисперсии (17).

$$S_e = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - p - 1}} \quad (17)$$

где S_e – стандартная ошибка полученной регрессионной модели;

n – количество объектов в опытной выборке;

p – количество регрессоров в регрессионной модели;

Для данного уравнения регрессии стандартная ошибка $S_e = 6,52$.

Оценка статистической значимости регрессии, в виде дисперсионного анализа, производится следующим образом [9]:

1. Выдвигается нулевая гипотеза о том, что уравнение в целом статистически незначимо.
2. Определяется фактическое значение F-критерия, по формуле (18).
3. Определяется табличное значение (по таблицам распределения Фишера) по заданному уровню значимости ($\alpha = 0,05$).
4. Если фактическое значение F-критерия меньше табличного, то нет основания отклонять нулевую гипотезу.
5. Принимается решение об отклонении нулевой гипотезы (и доверии статистической значимости регрессии).

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - p - 1}{p} \quad (18)$$

где F – F-критерий;

n – количество объектов в опытной выборке;

p – количество регрессоров в регрессионной модели;

R^2 – коэффициент детерминации полученной регрессионной модели.

В рамках данной выборки, после вычисления F-критерия формулой (18), определены $F = 2,74$, а $F_{table} = 0,21$. Так как табличное значение меньше фактического можно говорить о адекватности модели к опытным данным.

Таким образом, прогнозирование коэффициента эффективности ключевой фразы будет производиться уравнением регрессии (15).

Пример представлена таблица 2.5, содержащая уже рассчитаны ранее (таблица 2.4) коэффициент эффективности и его прогноз по полученной математической модели.

Таблица 2.5 – Пример прогнозирования коэффициента эффективности

	Ключевая фраза	Length	Wordstat	Productive	Прогноз Productive
1	номера купить красивые +на автомобиль	37	800	9,5	9,6
2	купить камаз сельхозник +с прицепом	35	50	-4,7	-0,08

	Ключевая фраза	Length	Wordstat	Productive	Прогноз Productive
3	электромонтаж официальный сайт адреса магазинов	35	52	-6,7	-0,06
4	российские алюминиевые лодки купить	47	28	12,3	5,4
5	купить эвакуатор	16	120	-4,6	-8,37
6	купить печь +для бани	21	87	-5,6	6,36

По примеру (таблица 2.5) видно довольно слабую численную точность прогнозирования, но при этом, если применить сортировку по убыванию, видно, что полученный коэффициент достаточен для определения эффективности в рамках определенной выборки.

Так все прогнозируемо слабые фразы окажутся в конце списка, а наиболее эффективные будут первыми, что и является задачей построенной модели.

Построена математическая прогнозируемая модель (15) методом множественной нелинейной регрессии (через метод наименьших квадратов). Оценка полученной модели произведена при помощи коэффициента детерминации ($R^2 = 0,64$) и множественного коэффициента корреляции ($R = 0,80$), адекватность которых была подтверждена при помощи критерия Стьюдента. При помощи данной модели произведено прогнозирование уже известных коэффициентов ключевых фраз (таблица 2.5). Результатом является вывод о низкой численной точности, но сохранении значения относительной эффективности в рамках определённой выборки.

3 ПРОЕКТИРОВАНИЕ ТЕХНОЛОГИИ ФОРМИРОВАНИЯ ЭФФЕКТИВНЫХ КЛЮЧЕВЫХ ФРАЗ

3.1 Проектирование диаграммы последовательностей модуля формирования эффективных ключевых фраз

Первостепенно необходимо определить архитектурную структуру реализуемого модуля. Для этого определим роль модуля в системе. Рисунок 3.1 демонстрирует диаграмму последовательности, определяющую роль модуля в системе сервиса «Rudvert», а именно часть оценки эффективности полученных ключевых фраз.

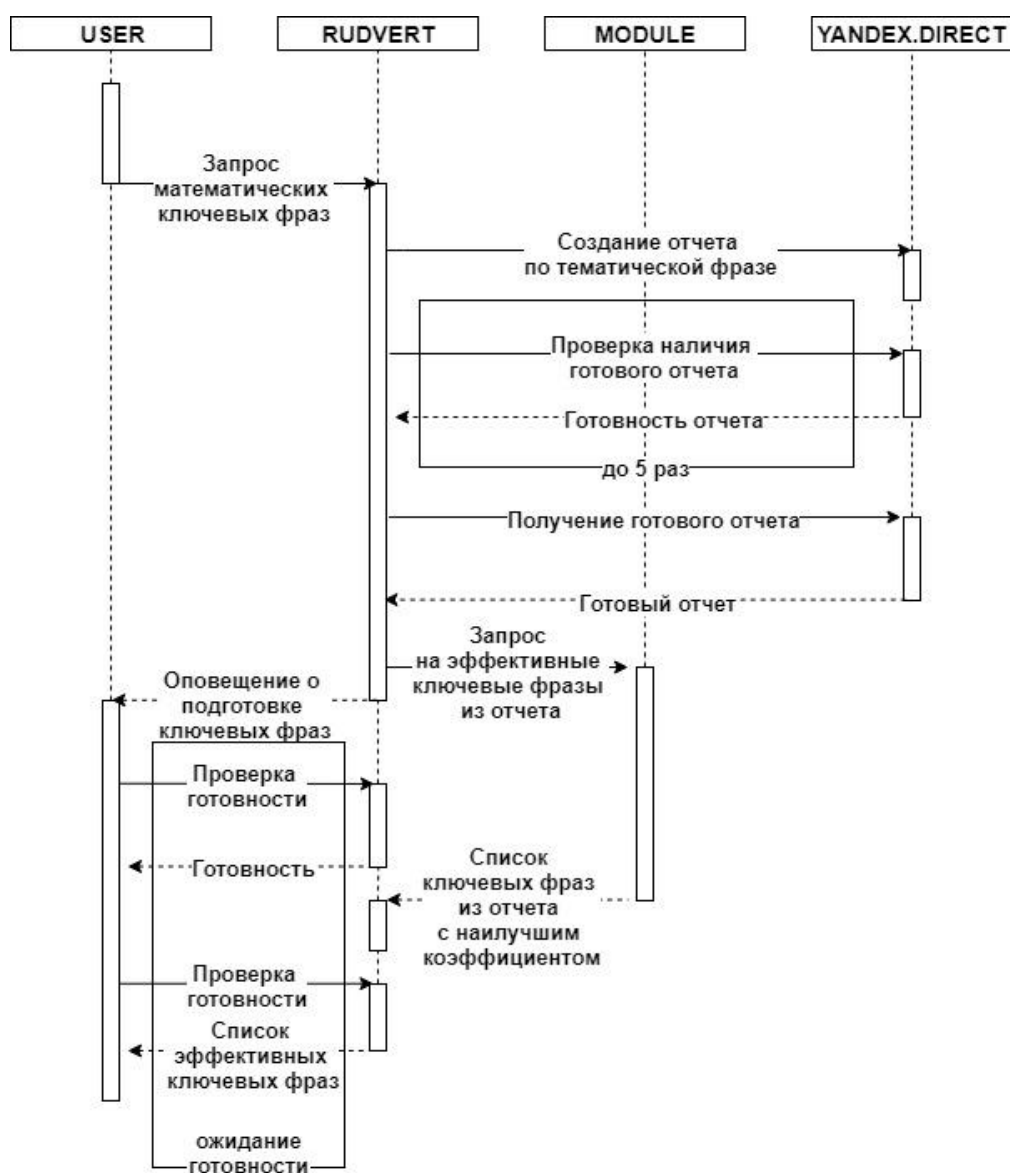


Рисунок 3.1 – Диаграмма последовательности для системы генерации
ключевых фраз пользователю

Диаграмма последовательности (рисунок 3.1) показывает, как пользователь делает запрос через интерфейс «Rudvert», передавая список базисных фраз, на основе которых будет формироваться новое семантическое ядро. Кроме этого пользователь выбирает желаемый регион показов, который «Rudvert» интерпретирует в идентификатор региона, предусмотренный «Яндекс.Директ».

Далее Rudvert обращается к «Яндекс.Директ» через API, а именно к методу CreateNewWordstatReport [9] (API 4 Live версии), вкладывая в запрос базисные фразы и идентификаторы регионов.

После, с периодичностью в 2 секунды, тот же скрипт запрашивает список готовых отчетов методом GetWordstatReportList [9], ожидая в списке нужный идентификатор отчета (полученный в ответ на CreateNewWordstatReport). Как только последний обнаружен, методом getWordstatReport [9], «Rudvert» запрашивает отчет.

В отчет включены массивы SearchedWith (вложенные фразы) и SearchedAlso (подобные фразы). За исключением самой фразы передается статистика поисковых запросов в Яндекс за последний месяц.

Вложенные фразы со статистикой передаются модулю формирования эффективных ключевых фраз. Сформированный список фраз модуль возвращает тому же скрипту.

Пока последний производит прогнозирования эффективности ключевых фраз, клиентская часть интерфейса Rudvert с определенной периодичностью запрашивает готовый список.

После получения списка ключевых фраз, в следующий же запрос клиентской части он помещается.

3.2 Проектирование архитектуры модуля формирования эффективных ключевых фраз

Помимо вычисления эффективности вложенных фраз, модуль поддерживает актуальность используемой модели прогнозирования и

опытную выборку. MasterController запрашивает прогноз эффективности, прикладывая фразы, полученные по тематике пользователя у сервиса Вордстат (через «Яндекс.Директ» API).

3.3 Проектирование внутренней структуры модуля формирования эффективных ключевых фраз

Помимо понимания архитектуры реализуемого модуля требуется обозначить его внутреннее устройство. Рассмотрим рисунок 3.4, содержащий диаграмму классов вычислительной части модуля (пакет calculate).

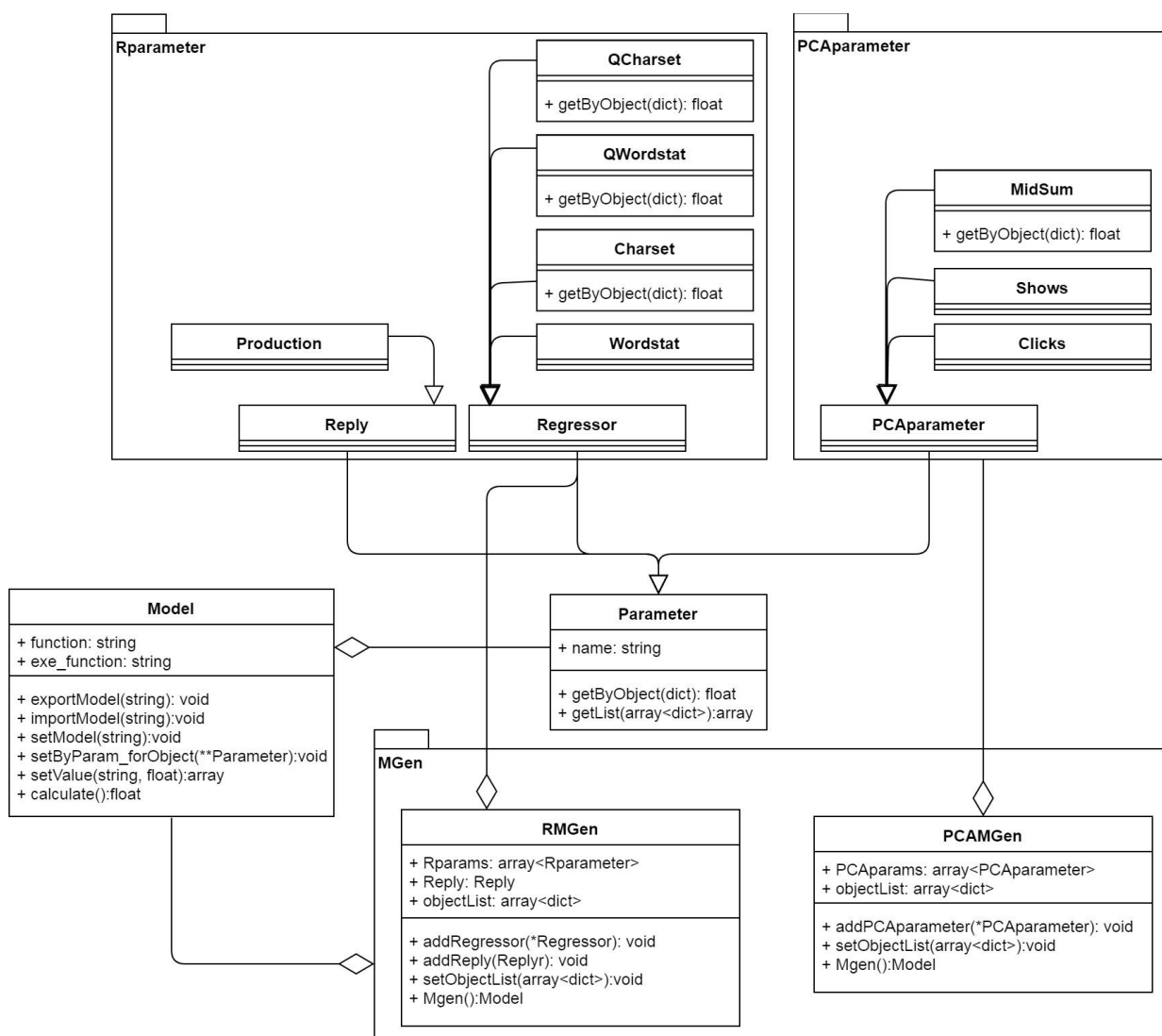


Рисунок 3.3 – Диаграмма классов пакет calculate

Таблица 3.1 содержит часть спецификации диаграммы классов (рисунок 3.3), реализующих вычислительную часть модуля формирования

эффективных фраз (пакета calculate). Элементы, не рассмотренные в спецификации (таблица 3.1), являются потомками, практически полностью или полностью копирующими логику родителя. Их необходимость определяется придерживанию объектно-ориентированному стилю написания программного кода, реализацией возможности расширения и создания наглядности архитектуры реализуемого программного модуля.

Таблица 3.1 – Основная часть спецификации пакета calculate

Класс	Аргументы	Методы
Parameter	+ string name – название описываемого параметра.	+ float getByObject (dict) – получение параметра из объекта выборки
Model	+ string function – функция модели. + string exe_function – копия функции для подстановки параметров	+ exportModel(string) – экспорт модули в файл. + importModel(string) – импорт модели из файла + setModel(string) – установка атрибута function + setByParam_forObject(**Parameter) – подстановка значений параметров объекта выборки. + array setValue(string, float) – установка значения параметру + float calculate() – вычисление функции модели со значениями параметров
RMGen	+ array<Rparameter> Rparams – лист регрессоров. + Reply Reply – отклик + array<dict> objectList – лист объектов выборки	+ addRegressor(*Regressor) – добавить регрессор + addReply(Reply) – добавить отклик + setObjectList(array<dict>) – установить лист объектов выборки + Model Mgen() – сформировать модель
PCAGen	+ array<PCApamet er> PCAparams – лист объясняющих параметров. + array<dict> objectList – лист объектов выборки.	+ addPCApamet er(*PCApamet er) – добавить объясняющую переменную. + setObjectList(array<dict>) – установить лист объектов выборки + Model Mgen() – сформировать модель

Пакет Mget содержит непосредственно логику расчета и прогнозирования коэффициента эффективности ключевой фразы. Результатом работы обоих классов пакета является экземпляр класса Model, содержащий формулу, позволяющую рассчитать искомую переменную по атрибутам объекта.

Пакеты PCParameter и Rparameter содержат классы, описывающие логику получения определённых параметров из объекта (словаря, содержащего в качестве атрибутов параметры ключевой фразы). Все классы пакетов являются потомками Parameter.

Рассмотрим взаимодействие пакета calculate с системой управления модуля (пакетом controller). Рисунок 3.4 содержит диаграмму классов, описывающих данный аспект.

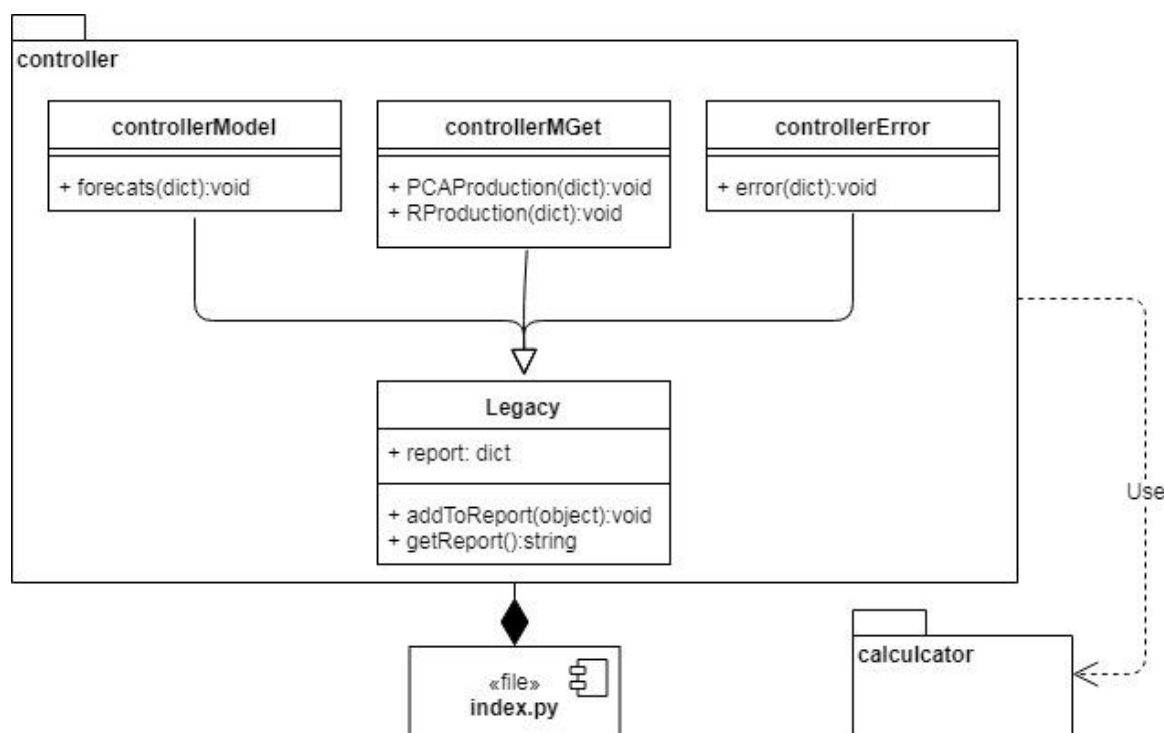


Рисунок 3.4 – Диаграмма классов пакета controller

Спецификация диаграммы классов (рисунок 3.4) системы управления модулем формирования эффективных ключевых фраз в таблице 3.2.

Таблица 3.2 – Спецификация классов пакета controller

Класс	Методы
Legacy	+ addToReport(object) – добавить объект в отчет + string getReport() – отдать отчет.
controllerModel	+ forecats(dict) – спрогнозировать эффективность полученных ключевых фраз.
controllerMGet	+ PCAProduction(dict) – пересчитать модель оценки ключевых фраз. + RProduction(dict) – пересчитать модель прогнозирования эффективности ключевой фразы.
controllerError	+ error(dict) – сформировать отчет об ошибках.

Как видно по диаграмме (рисунок 3.4), входная точка (файл index) использует пакет controller, который в свою очередь использует пакет calculate.

Пакет controller содержит класс-контроллеры, наследуемые классом Legacy, чьи методы используются для реализации представления (вывода), реализованного как JSON строка отчета, во входной точке.

3.4 Выбор средств реализации модуля формирования эффективных ключевых фраз

Для того, чтобы приступить к непосредственной реализации модуля необходимо определить язык программирования для реализации модуля. Рассмотрим три языка применяемых в разработке web-приложений: PHP [15], Java [13], Python [14].

Введем критерии оценки:

1. Наличие опыта (10 баллов) – наличие знаний и навыков работы на данном языке программирования.
2. Направленность на web-приложения (5 баллов) – наличие инструментов для разработки web-приложения.
3. Наличие готовых решений для выбранного математического аппарата (10 баллов) – наличие библиотек, направленных на матричную

математику и математическую статистику в данном языке программирования.

4. Скорость выполнения операций (10 баллов) – производительность с позиции выполнения многоитерационных циклов и математических вычислений.

Сравнительный анализ представлен в таблице 3.3.

Таблица 3.3 – Сравнительный анализ языков программирования

Критерий	PHP	Python	Java
Наличие опыта	8	5	3
Направленность на web-приложения	5	4	2
Наличие готовых решений для выбранного математического аппарата	2	8	8
Скорость выполнения операций	2	6	8
Итого	17	23	21

На основе проведенного анализа был выбран язык программирования Python. Далее необходимо выбрать библиотеку для выполнения математических операций для Python.

Рассмотрим три библиотеки, реализованные на Python, направленные на математические операции Scikit-learn [11], SymPy [12] и NumPy [10].

Введем критерии оценки:

1. Поддержка разработчиками (5 баллов) – актуальность использования библиотеки и наличие постоянного обновления и её улучшения.

2. Оптимизация (5 баллов) – оптимизация и затратность библиотеки.

3. Функциональность (10 баллов) – возможности библиотеки.

4. Наличие информации по библиотеке (10 баллов) – наличие информации и актуальной подробной документации по библиотеке.

Сравнительный анализ представлен в таблице 3.4.

Таблица 3.4 – Сравнительный анализ математических библиотек Python

Критерий	Scikit-learn	SymPy	NumPy
Поддержка разработчиками	5	5	2
Оптимизация	4	3	3
Функциональность	8	6	4
Наличие информации по библиотеке	7	8	6
Итого	24	22	15

На основе проведенного анализа была выбрана Python библиотека для математических операций Scikit-learn.

Таким образом программный модуль будет реализован на языке программирования Python с применением сторонней библиотеки для выполнения математических операций Scikit-learn.

3.5 Реализация классов Model и Parameter

Класс Model в рамках пакета calculator предназначен для хранения, выполнения и восстановления математической модели, сформированной MGen частью модуля.

Как видно по листингу данного класса (программный код в приложении) методы importModel и exportModel записывают или считывают (из файла) атрибут function, описывающий функцию нахождения искомого значения. Хранение в файле было выбрано в связи с отсутствием сильной необходимости содержать в рамках модуля собственную базу данных.

Использование атрибута function производится в методах setByParam_forObject, setValue и calculate.

Общий принцип заключается в подмене маркированной части строки function на конкретное значение свойства объекта (setByParam_forObject, setValue) и последующие выполнение данной строки как программного кода (calculate).

Метод setByParam_forObject получает неограниченное количество именованных атрибутов и из них ожидает следующие переменные:

- object – исследуемый объект;

- `parameters` – кортеж (не редактируемый список) объектов классов параметров (классов, наследуемых класс `Parameter`).

Далее, при помощи метода `getByObject`, описывающего получение значения по каждому параметру для данного объекта (описанного словарём (неупорядоченная коллекция) `object`) значения параметров помещаются в математическую модель (копия `function`, атрибут `exe_function`, необходимый для возможности восстановить модель в рамках экземпляра `Model`) методом `setValue`. Последний выполняет подстановку значения в строку `exe_function` вместо обернутого символами «#» названия параметра.

Итогом, метод `calculate` выполняет строку `exe_function`, как программный код, предварительно проверяя наличие всех параметров, и возвращает значение искомого параметра.

Как видно из листингов класса `Model` (программный код в приложении), он напрямую связан с классом `Parameter`, а точнее с его потомками.

Класс `Parameter` предназначен для определения логики получения, определенного параметра объекта по нему же. Напрямую в рамках модуля не используется и предназначен для последующего наследования, хоть, при этом, и не является абстрактным и может порождать собственные экземпляры. Содержит методы `getByObject` и `getList`.

Метод `getByObject`, как и описывалось ранее, содержит инструкцию по нахождению значения параметра в рамках объекта. В классе `Parameter` данный метод описывает самую тривиальную ситуацию – когда значение параметра явно определено в словаре `object` с ключом, одноименным атрибуту `name` класса `Parameter`. Нетривиальные примеры можно увидеть в переопределенных методах потомков `MidBid` и `Length`.

Обе вариации метода `getByObject` пытаются получить параметр, отталкиваясь от других свойств объекта и, в случае если это невозможно, пробуют родительский метод.

Рисунок 3.5 содержит диаграмму классов, исходящих из класса Parameter для параметров, определённых в рамках данной работы.

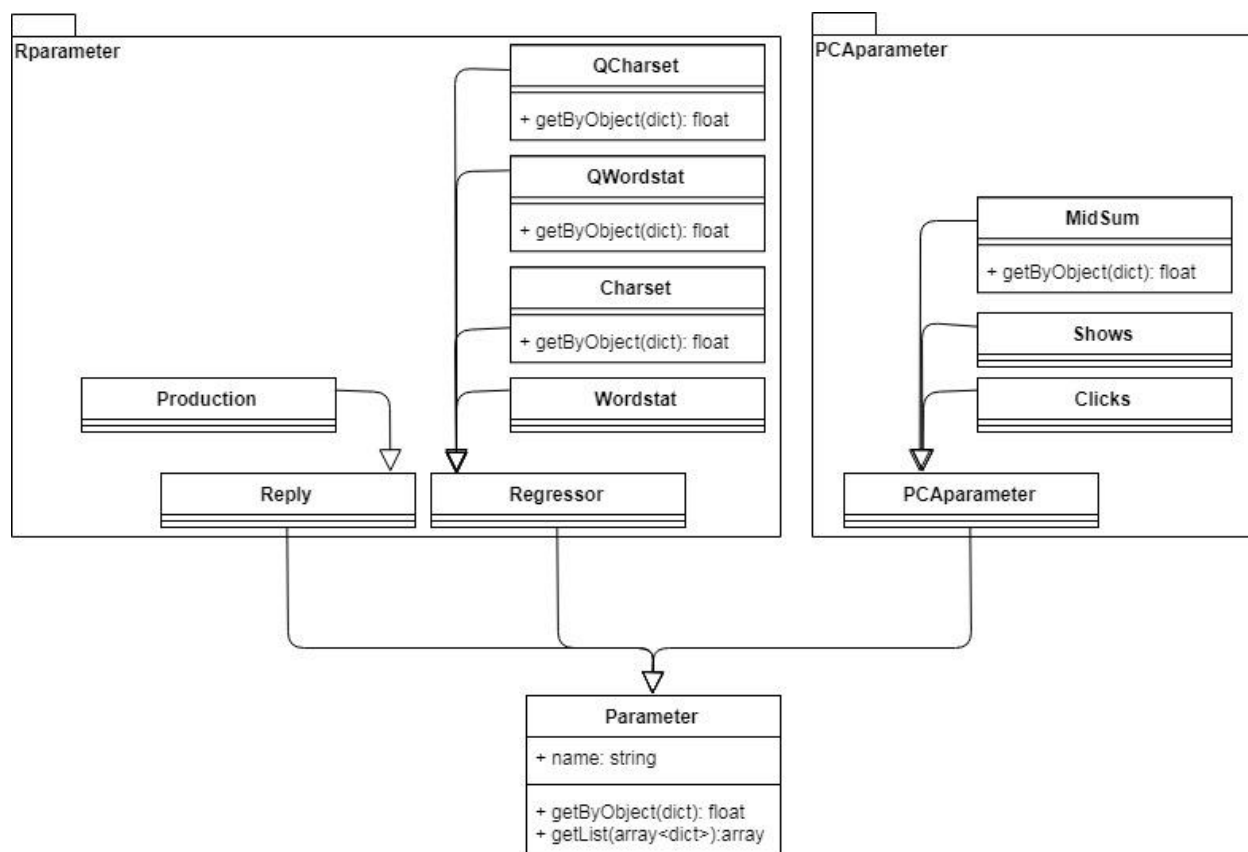


Рисунок 3.5 – Диаграмма классов-потомков Parameter

Так, из Parameter исходят Reply (отклик) и Regressor (регрессор) в пакете Rparameter (пакет, описывающий параметры, участвующие в регрессионном анализе) и PCparameter (описывающая переменная) для метода главных компонент в пакете PCparameter. Из последних исходят классы, описывающие конкретные параметры (Clicks, Wordstat, Production и т.д.).

Данная многоуровневая структура необходима для возможности внести изменения в логику обработки и хранения параметров на любом её уровне. Так, при необходимости, возможно внести дополнительную обработку, например, всем регрессорам прогнозирующей модели, или полностью преобразовать логику получения конкретных параметров.

3.6 Реализация классов пакета MGen

Пакет MGen состоит из двух классов: RMGen и PCAMGen, задача которых заключается в формировании модели (в виде экземпляра класса Model) регрессионным анализом и методом главных компонент, соответственно.

Оба класса работают по общему принципу, а именно предварительный ввод данных и расчет модели методом Mgen.

Единый для обоих классов метод setObjectList принимает массив, состоящий из словарей, описывающих объекты опытной выборки. Именно над этими данными будут производиться вычисления.

Методы addRegressor, addReply для PMGetn addPCAPparameter для PCAMGen ожидают соответствующие экземпляры классов-параметров (потомков Parameter) в качестве аргументов.

Метод Mgen класса PCAMGen использует библиотеку sklearn, а точнее класс PCA пакета decomposition. Помимо sklearn, используется библиотеку numpy, которая устанавливается в комплекте с sklearn.

Метод сначала собирает все необходимые данные в многомерный массив через цикл, проходящий по объектам-параметрам, каждый из которых вкладывает в этот массив лист параметров по каждому объекту выборки.

Далее данные помещаются в экземпляр класса sklearn.decomposition.PCA и производятся вычисления. Итог вычисления в виде массива коэффициентов весов теперь находится в атрибуте components_ экземпляра PCA.

Финальным действием является формирования function и помещение его в экземпляр Model, который и возвращается, завершая метод.

Теперь рассмотрим метод Mgen класса RMGen.

Важно отметить то, что данный метод выполняет линейную регрессию. Так, для нелинейной регрессии необходима линеинизация параметров (если это возможно). В рамках предложенных параметров задачу линеинизации выполняют классы-параметры QWordstat и QLength, рассмотренные выше.

Так же, как и у PCAMGen, Mgen метод использует библиотеки sklearn и numpy. Класс LinearRegression (пакет sklearn.linear_model) реализует вычисления линейной модели регрессии методом наименьших квадратов.

Аналогично Mgen для PCAMGen, данные собираются при помощи экземпляров классов-параметров в единый массив, который помещается в экземпляр LinearRegression с последующим вычислением коэффициентов весов. Разница заключается в разделении отклика (reply) от регрессоров (regressions). Формирование модели аналогично Mgen для PCAMGen.

Фактически, Mgen методы, и данные классы в целом, выполняют задачу прослойки между функционалом библиотеки sklearn и представлением данных, используемых в рамках программного модуля. Так, каким бы не был параметр и как бы сложно он не вычислялся он легко впишется в вычисления модели (если описан потомком Parameter), а результатом вычислений предсказуемо будет экземпляр Model.

3.7 Реализация управления модулем формирования эффективных ключевых фраз

Управление программным модулем реализовано через систему контроллеров. Рисунок 3.6 демонстрирует диаграмму классов-контроллеров в рамках модуля.

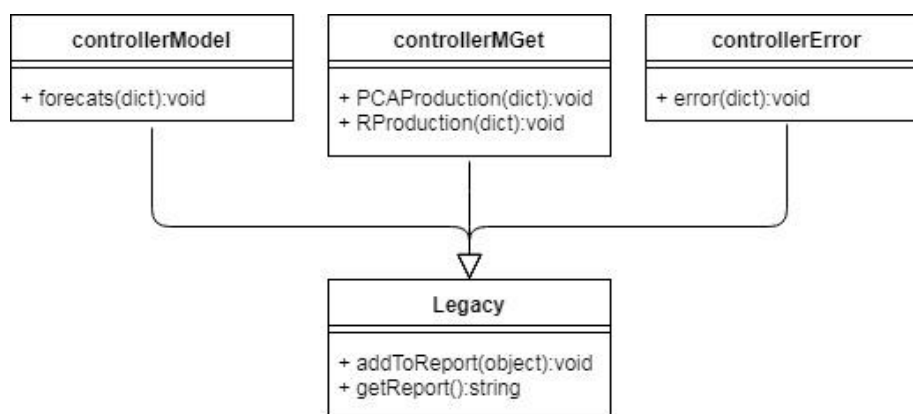


Рисунок 3.6 – Диаграмма классов-контроллеров

Запуск контроллера происходит в входном файле `index.py`. Данный файл содержит реализацию сокета при помощи библиотеки `tornado`. Входная точка ожидает три параметра:

- `controller` – имя используемого контроллера;
- `task` – метод выбранного контроллера;
- `data` – словарь (переданный в JSON), содержащий данные, которые будут вложены в метод контроллера как аргумент.

В случаях ошибочного выполнения или недостатка входных данных выбирается контроллер `Error` (`controllerError`) и его метод `error`, ожидающий во входном словаре сообщение об ошибке.

Методы `PCAProduction` и `RProduction` контроллера `MGet` выполняют пересчет математических моделей, используемых модулем. На выход возвращают «ОК» (или сообщение об ошибке).

Метод `forecast` контроллера `Model` выполняет прогнозирование полученной выборки. Возвращает ту же выборку, добавив к каждому объекту атрибут `Forecast`, содержащий спрогнозированный коэффициент эффективности ключевой фразы (`Production`).

Данная система управления является временной в рамках всей реализации, направленной на выполнение минимальных задач. С последующим тестированием и развитием модуля система управления будет заменена на более безопасную и отказоустойчивую.

ЗАКЛЮЧЕНИЕ

В рамках данной работы рассмотрена гипотеза о возможности использования математических методов для оценки прогнозирования эффективности ключевой фразы в рамках системы контекстной рекламы Яндекс.Директ. Разработан алгоритм оценки и прогноза эффективности ключевой фразы по статистическим и фактическим параметрам в рамках системы «Яндекс.Директ» и спроектирован и реализован модуль для сервиса «Rudvert», использующий данный алгоритм.

Через анализ сферы контекстного рекламирования, на примере «Яндекс.Директ», выдвинута проблема простого вхождения пользователей-рекламодателей в сферу контекстного рекламирования.

При рассмотрении сервиса «Rudvert», разрабатываемого в рамках ООО «Рудник» определена необходимость в алгоритме оценки и прогнозирования эффективности ключевых фраз.

Частичным анализом архитектуры системы «Яндекс.Директ» API через диаграмму объектов Keyword (и связанных с ним объектов) определено понимание ключевой фразы, как объекта системы.

Анализ готовых решений продемонстрировал необходимость реализации собственного решения.

Так же определены параметры, указывающие на эффективность ключевой фразы, при этом являясь доступными на момент необходимости прогнозирования для любой произвольной фразы: количество поисковых запросов за последний месяц (Wordstat) и посимвольная длина фразы (Length).

На основе выбранного, по наиболее важным критериям, математического аппарата (метод главных компонент для оценки ключевой фразы и регрессионный нелинейный метод для прогнозирования), разработан и проверен алгоритм, сначала оценки эффективности, и следом

прогнозирования этой оценки (представленной численным коэффициентом) для известных ключевых фраз.

Результатом разработки алгоритма являются модели оценки (через метод главных компонент) и прогнозирования (через метод нелинейной множественной регрессии).

Приведен пример оценки и прогнозирования эффективности шести случайных (из опытной выборки) ключевых фраз, по которым произведена оценка построенных математических моделей.

В рамках проектирования и реализации программного модуля (для пересчета и использования математических моделей оценки и прогнозирования ключевой фразы) представлены диаграммы последовательности, компонентов и классов.

В ходе выбора средств реализации определенны язык программирования и математическая библиотека в рамках выбранного языка (программный язык Python и библиотека Scikit-learn).

Итогом выполнен программный модуль (исходный код представлен в приложении).

В ходе данной работы, рассматриваемая гипотеза возможности прогнозирования эффективности ключевой фразы, через её фактические и статистические параметры, получила определенное подтверждение в виде некоторого успеха выполнения прогнозирования известных фраз.

Следующим шагом является улучшение реализованного модуля через постоянное добавление в модель прогнозирования новых параметров, потенциально влияющих на эффективность и при этом общедоступных для любой произвольно составленной, но при этом семантически значимой, фразы.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

Научная и методическая литература

1. Бабаев А. Контекстная реклама: учебник / А. Бабаев, Н. Евдокимов, А. Иванов – СПб.: Питер, 2015 – 304 с.
2. Магнус Я.Р. Эконометрика - Начальный курс: учебник / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий – М.: Дело, 2017 – 504 с.
3. Орлов А.И. Прикладная статистика: учебник для вузов / А.И. Орлов – М.: Экзамен, 2014 – 482 с.
4. Садовникова Н. А. Анализ временных рядов и прогнозирование: учебно-методический комплекс / Н. А. Садовникова, Р. А. Шмойлова – М.: Евразийский открытый институт, 2013 – 259 с.
5. Смирнов В.В. Прибыльная контекстная реклама. Быстрый способ привлечения клиентов с помощью «Яндекс.Директ» а / В.В. Смирнов – М.: Манн, Иванов и Фербер, 2013 – 192 с.

Электронные ресурсы

6. Capturing Architectural Requirements [Электронный ресурс]. — Режим доступа: <https://www.ibm.com/developerworks/rational/library/4706.html#N100A7> (дата обращения: 01.04.2018)
7. Eigenvectors and Eigenvalues [Электронный ресурс]. — Режим доступа: <http://setosa.io/ev/eigenvectors-and-eigenvalues/> (дата обращения: 01.05.2018)
8. STANDARDIZE function [Электронный ресурс]. — Режим доступа: <https://support.office.com/en-us/article/standardize-function-81d66554-2d54-40ec-ba83-6437108ee775> (дата обращения: 01.04.2018)
9. Документация API Директа — Технологии Яндекса [Электронный ресурс]. — Режим доступа: <https://tech.yandex.ru/direct/doc/dg/objects/keyword-docpage/> (дата обращения: 01.04.2018)

10. Документация NumPy [Электронный ресурс]. — Режим доступа: <http://www.numpy.org> (дата обращения: 12.05.2018)
11. Документация Scikit-learn [Электронный ресурс]. — Режим доступа: <http://scikit-learn.org/stable/index.html> (дата обращения: 12.05.2018)
12. Документация SymPy [Электронный ресурс]. — Режим доступа: <http://www.sympy.org/en/index.html> (дата обращения: 12.05.2018)
13. Сайт Oracle Java [Электронный ресурс]. — Режим доступа: <https://www.java.com/ru/> (дата обращения: 10.05.2018)
14. Сайт Python Software Foundation [Электронный ресурс]. — Режим доступа: <https://www.python.org> (дата обращения: 10.05.2018)
15. Сайт The PHP Group [Электронный ресурс]. — Режим доступа: <http://php.net/> (дата обращения: 10.05.2018)
16. Сайт программы Key Collector [Электронный ресурс]. — Режим доступа: <http://www.key-collector.ru/> (дата обращения: 09.04.2018)
17. Сайт рекламного интернет-агентства ПРОСТОР [Электронный ресурс]. — Режим доступа: <https://prosto-r.ru/> (дата обращения: 09.04.2018)
18. Сервис контекстной рекламы Google Adwords [Электронный ресурс]. — Режим доступа: <https://adwords.google.com/> (дата обращения: 01.04.2018)
19. Сервис контекстной рекламы «Яндекс.Директ» [Электронный ресурс]. — Режим доступа: <https://direct.yandex.ru> (дата обращения: 01.04.2018 г.).
20. Сервис поисковой аналитики Топвизор [Электронный ресурс]. — Режим доступа: <https://topvisor.ru/> (дата обращения: 05.04.2018)
21. Сервис статистики по ключевым фразам Keyword Planner – Google AdWords [Электронный ресурс]. — Режим доступа: <https://adwords.google.com/aw/> (дата обращения: 05.04.2018)
22. Сервис статистики по ключевым фразам Keyword Tool [Электронный ресурс]. — Режим доступа: <https://keywordtool.io/> (дата обращения: 09.04.2018)

23. Сервис статистики по ключевым фразам Яндекс.Вордстат [Электронный ресурс]. — Режим доступа: <https://wordstat.yandex.ru/> (дата обращения: 05.04.2018)

24. Сервис управления рекламным аккаунтом Rudvert [Электронный ресурс]. — Режим доступа: <https://rudvert.ru/> (дата обращения: 01.04.2018)

Литература на иностранном языке

25. Darlington, R. Regression Analysis and Linear Models Concepts, Applications, and Implementation / R. Darlington, A. Hayes – New York: The Guilford Press, 2016 – 661 с.

26. Davis, H. Google Advertising Tools: Cashing in with AdSense, AdWords, and the Google APIs / H. Davis – California: O'Reilly, 2016 – 368 с.

27. Devore, J.L. Probability and Statistics for Engineering and the Sciences / J.L. Devore – Boston, MA: Cengage Learning, 2012 – 504 с.

28. Jolliffe, I.T. Principal Component Analysis / I.T. Jolliffe – Berlin: Springer, 2012 – 488 с.

ПРИЛОЖЕНИЕ А.

Носитель с исходным кодом программы

К данной работе прикреплён компакт-диск формата CD-R, на котором находится следующее содержимое:

1. Исходный код реализованного программного модуля формирования эффективных ключевых фраз в рамках сервиса «Rudvert».
2. Текст пояснительной записки ВКР.