

Автоматическая классификация обращений на основе анализа больших данных

Ерисов В.Д., магистрант,

ФГБОУ ВО «Московский политехнический университет», Москва, Россия

Суворов В.С., к.э.н., профессор,

ФГБОУ ВО «Московский политехнический университет», Москва, Россия

Царькова Н.И., к.п.н., доцент,

ФГБОУ ВО «Московский политехнический университет», Москва, Россия

Аннотация. В данной статье рассматривается применение технологий анализа Big Data к имеющимся данным по внутренним заявкам и запросам сотрудников компании. Целью исследования является автоматизация и оптимизация процесса обработки и распределения поступающих от сотрудников обращений для повышения скорости, и качества исполнения внутриорганизационных обязанностей структурными подразделениями компании. В качестве основных инструментов и платформ для анализа данных и проектировки программного обеспечения используются: Python, IDLE, Jupyter Notebook. В результате исследования был произведен анализ большого массива данных, их очистка и дальнейшее выделение ключевых слов и проблемных тем обращений. Помимо этого, используя высокоуровневый язык программирования Python и с применением операторов ветвления, было спроектировано программное обеспечение, которое производит автоматическую классификацию обращений сотрудников внутри компании.

Ключевые слова: Python, большие данные; программное обеспечение; анализ данных; автоматическое распределение обращений; структурные подразделения; классификация

Automatic classification of appeals Based on big data analysis

Erisov V.D., undergraduate student,

Moscow Polytechnic University, Moscow, Russia

Suvorov S.V., Candidate of Economic Sciences,

Moscow Polytechnic University, Moscow, Russia

Tsarkova N.I., Ph.D., Associate Professor,

Moscow Polytechnic University, Moscow, Russia

Annotation. This article discusses the application of Big Data analysis technologies to the available data on internal requests and requests of company employees. The purpose of the research is to automate and optimize the process of processing and distributing requests received from employees in order to improve the speed, accuracy and quality of performance of internal organizational duties by the company's structural divisions. The main tools and platforms for data analysis and software design are: Python, IDLE, and Jupyter Notebook. As a result of the research, a large array of data was analyzed, cleaned up, and further highlighted keywords and problematic topics of requests. In addition, using the high-level Python programming language and using branching operators, software was designed that automatically classifies employee requests within the company.

Key words: Python, big data; software; data analysis; automatic distribution of requests; structural divisions; classification

Введение

В связи с ростом объема получаемой разнородной информации возникает необходимость замены традиционных инструментов сбора, обработки, анализа и хранения данных на технологию Big Data [1]. Это позволяет более рационально организовать работу предприятия, повысить эффективность операций. В итоге есть все основания рассчитывать на более высокий уровень удовлетворения потребностей клиентов и, как следствие, на повышение прибыли. В ходе

подготовки отчета «Большие данные в крупных компаниях» директор по исследованиям ПА Том Давенпорт опросил топ-менеджеров более 50 компаний, чтобы понять, как они используют большие данные. [2] Он обнаружил, что, по их мнению, использование больших данных дает положительный эффект по следующим параметрам:

- Снижение цены.

Технологии больших данных, такие как Hadoop и облачная аналитика, позволяют существенно сократить затраты, а также позволяют определять более эффективные способы ведения бизнеса. [3]

- Более быстрое и правильное принятие решений.

Благодаря скорости программ и аналитике в сочетании с возможностью анализа новых источников данных предприятия могут более оперативно анализировать информацию и принимать решения.

- Новые товары и услуги.

С помощью аналитики больших данных все больше компаний создают новые продукты для удовлетворения потребностей клиентов и своих сотрудников.

Экономические условия информационного общества предъявляют к хозяйствующим субъектам настоятельное требование заменить рутинный нефизический труд человека работой информационно-технологических систем. Оно обусловлено, главным образом, двумя обстоятельствами. Во-первых, программно-аппаратному комплексу не нужно выплачивать заработную плату. И, во-вторых, он лишен недостатков, свойственных живым организмам. Качество его внимания в процессе решения стоящих перед ним задач не зависит от состояния здоровья, физиологических особенностей, настроения и различных психических процессов.

Сегодня замена нетворческого умственного труда человека работой информационно-технологических систем позволит предприятиям, организациям и учреждениям обеспечить свою конкурентоспособность в сфере организации

труда и является одной из ключевых задач в работе по модернизации системы менеджмента, что определяет актуальность темы настоящего исследования.

Объектом настоящего исследования является классификация заявок пользователей в структурные подразделения, к компетенции которых они относятся. Предметом его является автоматизация этого процесса.

Целью настоящего исследования является создание информационно-технологической системы, позволяющей полностью автоматизировать процесс классификации заявок. Достижение намеченной цели предполагает решение следующего комплекса задач:

- подготовка и очитка большого массива данных;
- определение ключевых слов и словосочетаний;
- категоризация обращений;
- разработка интерактивного интерфейса.

Результат исследования

По результатам предварительного анализа текущего процесса обработки обращений сотрудников внутри компании были выявлены следующие функции, представленные ниже (рис. 1):

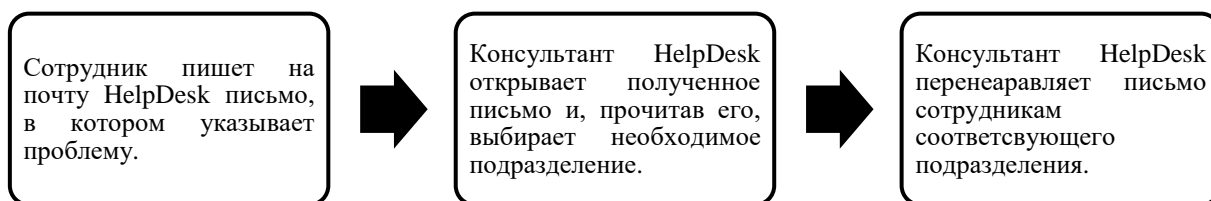


Рис. 1 – Текущий процесс обработки обращений

Текущий процесс имеет ряд недостатков. Очевидно, что между сотрудником, который оставляет заявку, и структурным подразделением, которое её должно исполнить, присутствует ещё одно звено – консультант службы поддержки, который непосредственно занимается распределением поступающих заявок в отделы, к компетенции которых они относятся.

Распределение заявок происходит путем выявления тематики обращения. Определение тематики обращения опирается на два фактора.

Первый фактор подразумевает четкую регламентацию соотношения логически правильно составленной заявки с использованием ключевых слов к словарю компетенций структурных подразделений. Это значит, что в компании существуют утверждённые нормы составления заявок, которые диктуют необходимый порядок употребления слов и словосочетаний. Вторым фактором отталкивается от интуитивного понимания корреляции между смысловой нагрузкой текста заявки и существующими структурными подразделениями с перечнями их компетенций. Это подразумевает, что, например, новые сотрудники, еще не успевшие полностью ознакомиться с внутренними порядками компании, могут писать заявки в свободной форме, не позволяющей консультанту сразу распознать тематику обращения.

Приняв во внимание оба фактора, можно утверждать, что в процессе распределения заявок возможны неточности и ошибки, обусловленные человеческим фактором. Консультант может забыть или перепутать компетенции структурных подразделений, неправильно соотнести ключевое слово или не найти связи в тексте сообщения.

Автоматизация процесса обработки обращений сотрудников не только сократит время перенаправления письма, что в свою очередь повлияет на скорость решения конкретной проблемы, но, также поможет избежать возникновения ошибки, обусловленной человеческим фактором.

В соответствии с поставленными задачами исследования был проведен первичный анализ имеющихся данных, необходимый для выявления специфики обращений сотрудников компании, проблемные темы, ключевые слова и словосочетания. Исходные данные включают следующие параметры: дата обращения; текст обращения; ответственное подразделение; идентификатор инициатора.

Все обращения распределяются между шестью структурными подразделениями со своими компетенциями (табл. 1).

Структурные подразделения и их компетенции

Структурное подразделение	Компетенции
Отдел системного обеспечения	Сломался/не работает/нет подключения/нужно выдать сотруднику/нужно подключить сотруднику: <ul style="list-style-type: none"> • Компьютер; • Принтер; • Интернет; • Телефон.
Отдел программного обеспечения	Не работает/нужно установить/нужно обновить/слетела/удалилась: <ul style="list-style-type: none"> • Программа; • Файл; • Приложение. Не работает СЭД (система электронного документооборота)
Отдел нормативно-справочной информации	Нужно добавить/удалить/изменить в базе данных 1С или СЭД: <ul style="list-style-type: none"> • Договор; • Контрагента; • Реквизиты.
Административно-хозяйственный отдел	Необходимо предоставить/заказать/поменять/починить: <ul style="list-style-type: none"> • Кресло; • Стол; • Тумбочку; • Шкаф. Необходимо сделать уборку в кабинете. Необходимо поменять воду в кулере. Необходимо предоставить корпоративное жильё. Не работает/нужно починить: <ul style="list-style-type: none"> • Свет; • Кондиционер; • Батарея.
Отдел делопроизводства	Необходимо предоставить/заказать: <ul style="list-style-type: none"> • Канцтовары. Нужно отправить почту.
Отдел корпоративного развития	Необходимо обучить: <ul style="list-style-type: none"> • Программы Microsoft; • 1С; • ГрандСмета.

Первым шагом исследования были подготовка и очистка данных. Были проверены целостность данных, наличие пустых показателей и некоренных записей. Дублирующие друг друга сообщения, отправленные в разные дни одним и тем же пользователем, не были удалены, так как в данном случае наличие двух идентичных сообщений с разницей во времени более чем в один

день свидетельствуют не об ошибке пользователя и случайном отправлении двух одинаковых писем, а о том, что ответ на обращение не получен или возникла идентичная проблема. Дубликаты сообщений с разницей во времени в несколько минут были удалены. Они составили менее 0,5% от общего количества записей.

Анализ имеющихся данных выявил (рис. 2), что структурное подразделение «Отдел программного обеспечения» получает наибольшее количество обращений, что составляет 32,83% от общего количества обращений за год. На втором месте по популярности находится отдел системного обеспечения, а на третьем – отдел нормативно-справочной информации.



Рис. 2 – Распределение количества обращений за год

Следующим шагом работы было выявление ключевых слов из текста обращений и дальнейшее формирование классов. Данные из столбца с текстом обращений были разделены по пробелам и знакам препинания. Затем они были очищены от ненужных для анализа слов. Например, были удалены приветствия, даты, подписи, персональные данные и т.д. В итоге были выделены ключевые слова и словосочетания, разбитые на группы, соответствующие ответственным за данную проблему подразделениям (рис. 3).

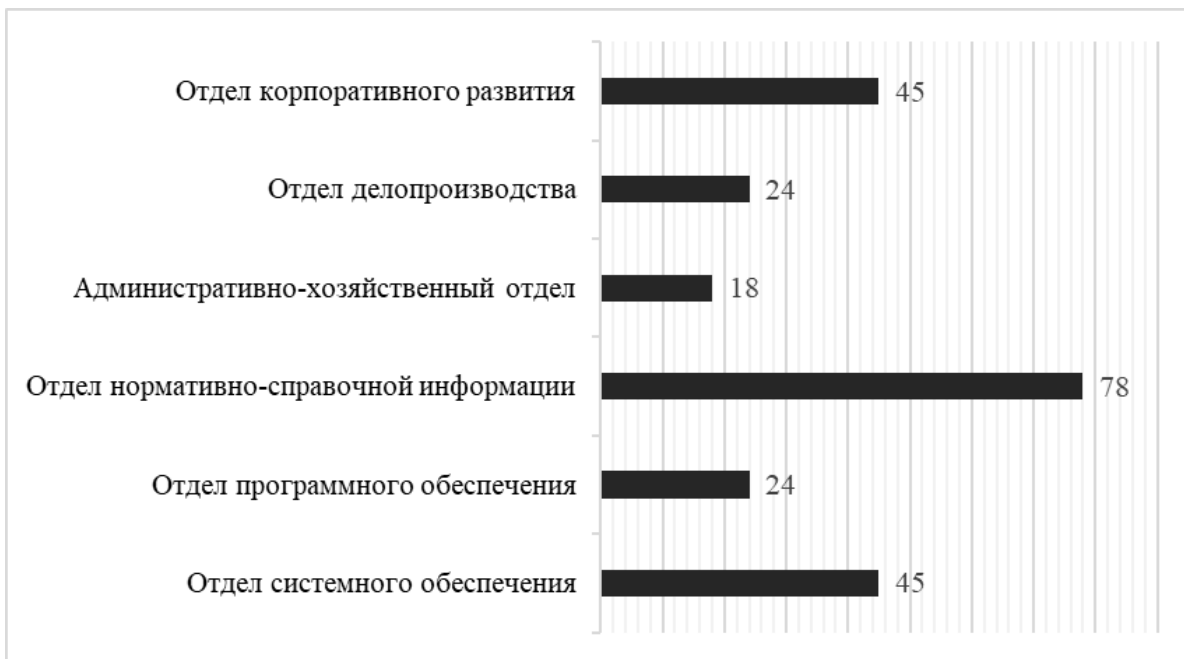


Рис. 3 – Ключевые слова

Помимо этого, было выявлено, что потенциальная проблема автоматизации обработки обращений может заключаться в большом количестве вариантов написания ключевого слова. Например, если у сотрудника проблема с программой 1С, то он может написать: 1С (где С – буква русского алфавита), 1С (где С – буква английского алфавита), 1 ЭС, 1S, Один эс и т.д. Для решения данной проблемы в группы ключевых слов следует внести все встречающиеся варианты написания. В дальнейшем будет необходимо поддерживать массив с ключевыми словами, добавляя новые проблемные темы и варианты написания ключевых слов.

Следующим шагом было создание программного обеспечения для автоматизации распределения обращений (рис. 4) Полученные в ходе анализа текстов обращений ключевые слова, распределённые по группам, были использованы для формирования списка условий для распределения. Программа была написана с использованием языка программирования Python 3 и операторов ветвления. Пользователю предлагается выбрать тематику своей проблемы, а затем написать ключевое слово, которое характеризует его обращение, учитывая выбранную им категорию. Затем происходит автоматический выбор нужного отдела.


```
Python 3.7.2 Shell
File Edit Shell Debug Options Window Help
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018, 22:20:52) [MSC v.1916 32 bit
(Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: C:/Users/Desktop/l11.py =====
Нажмите:
1 - если что-то сломалось/не работает/нет подключения/нужно выдать сотруднику/ну
жно подключить сотруднику
2 - если что-то не работает/нужно установить/нужно обновить/слетела/удалилась
3 - если нужно что-то добавить/удалить/изменить в базе данных 1С или СЭД
4 - если необходимо предоставить/заказать/поменять/починить
5 - если необходимо заказать курьера/канцтовары или отправить почту
6 - если необходимо обучить сотрудника

2
Что именно не работает/нужно установить/нужно обновить/слетело/удалилось? файл
Заявка перенаправлена специалистам отдела программного обеспечения
>>>
Нажмите:
1 - если что-то сломалось/не работает/нет подключения/нужно выдать сотруднику/ну
жно подключить сотруднику
2 - если что-то не работает/нужно установить/нужно обновить/слетела/удалилась
3 - если нужно что-то добавить/удалить/изменить в базе данных 1С или СЭД
4 - если необходимо предоставить/заказать/поменять/починить
5 - если необходимо заказать курьера/канцтовары или отправить почту
6 - если необходимо обучить сотрудника

6
Какой программе необходимо обучить сотрудника? 1 Эс
Заявка перенаправлена специалистам отдела корпоративного развития
```

Рис. 4 – Работа программы

В данной программе предусмотрены различные варианты написания проблемной темы, что позволяет повысить уровень ее точности и сократить количество возвратов обращений с просьбой переформулировать ключевое слово или выбрать другую тему для обращения.

Заключение

Таким образом, можно сделать вывод, что рассматриваемая проблема решена, автоматическое распределение обращений сотрудников происходит с помощью полученного в ходе big data анализа массива ключевых слов с различными вариантами их написания. Следует отметить, что по мере возникновения новых тем для обращений появится необходимость в анализе групп ключевых слов. Также представляется целесообразным ежегодно проводить профилактическое выявление новых вариантов написания существующих ключевых слов.

Библиографический список

1. Царькова Н.И., Ерисов В.Д., Пекова Е.А. Технология BigData как инструмент управления в межкультурной коммуникации // Управление экономическими системами: электронный научный журнал. – 2019. – № 7.
2. Суворов С.В., Царькова Н.И., Спиридонова А.К. Анализ больших данных компании Uber Technologies Inc с помощью технологии Data Mining // Управление экономическими системами: электронный научный журнал. – 2019. – № 7.
3. Hadoop — официальный сайт языка программирования Hadoop [Электронный ресурс]. Режим доступа: <https://hadoop.com>, свободный – (дата обращения: 04.11.2020).

References

1. Tsarkova N.I., Erisov V.D., Pekova E.A. Big Data technology as a management tool in inter-cultural communication // Management of economic systems: electronic scientific journal. – 2019. – № 7.
2. Suvorov S.V., Tsarkova N.I., Spiridonova A.K. Big data analysis of Uber Technologies Inc using Data Mining technology // Management of economic systems: electronic scientific journal, – 2019. – № 7.
3. Hadoop - the official site of the Hadoop programming language [Electronic resource]. Access mode: <https://hadoop.com>, free – (date of access: 04.11.2020).