

ПРОЕКТИРОВАНИЕ МОДЕЛЕЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ОТЗЫВОВ ВЕБ-САЙТА АВТОСАЛОНА

Н.В. Давыдик

УО «Гродненский государственный университет им. Я. Купалы», факультет математики и информатики, специальность «Управление информационными ресурсами», кафедра системного программирования и компьютерной безопасности.

Научный руководитель - Е.В. Косарева, кандидат физико-математических наук, доцент кафедры системного программирования и компьютерной безопасности.

Во введении указан объект исследования - система автоматизации процесса обработки отзывов веб-сайта автосалона. Осуществлен сравнительный анализ имеющихся в свободном доступе онлайн-сервисов для обработки текстов «Tone Analyzer» и «Анализ писем». Проанализирована предметная область и выявлены основные преимущества при обработке отзывов методами интеллектуального анализа. В основной части определены этапы работ по обработке отзывов и определению тематики методами интеллектуального анализа. Спроектированы модели: ER-диаграмм, предоставляющая возможность визуализировать проектирование базы данных; модель IDEF0, DFD модель - диаграмма потоковых данных системы и Use-case диаграмма (диаграмма вариантов использования) – представляет собой описание действий, выполняемых системой, являющиеся значимыми для конкретного актера. Целью исследования является изучение теоретических аспектов, касающиеся методов анализа естественного языка для обработки текста и проектирование автоматизированной системы обработки отзывов веб-сайта. Научная значимость исследования заключается в использовании интеллектуальных методов обработки естественного языка для анализа тональности отзывов и определения темы обращения пользователей веб-сайта автосалона, что позволяет сократить время на обработку отзыва сотрудником компании, получать аналитику об имидже компании и своевременно решать конфликтные ситуации.

Ключевые слова: процесс обработки отзывов, автоматизация процессов, проектирование систем.

Введение. Для эффективной деятельности любой компании одной из важных составляющих является конкурентоспособность и устойчивый имидж в глазах своих клиентов. А при выборе определенной услуги или продукции компании потребитель, не имея опыта с данной организацией, будет обращать первоочередное внимание на опыт других клиентов с данной организацией.

Данный опыт клиент может получить, прочитав отзывы на веб-сайте компании. И на основании этих отзывов формирует свое отношение к организации, предоставляющей товары или услуги.

При этом, следует учитывать, что текст отзыва может нести как положительную, так и отрицательную оценку продукции компании. И здесь большую роль играет то, как организация будет обрабатывать отрицательные отзывы.

Когда обработкой и определением проблемы отзыва занимается сотрудник компании, то в этом случае может быть затрачено много времени на данный процесс. А необработанные негативные отзывы, находящиеся на официальном веб-сайте компании, оказывают негативное влияние на имидж и конкурентоспособность.

При анализе онлайн сервисов «Анализ писем» и «Tone Analyzer», можно сделать вывод, что данные системы не предусматривают выявления тематики, что увеличивает время сотрудника компании на обработку отзыва и решение проблемы. При чем, обработка отзыва и определение его эмоциональной окраски осуществляется через тональный словарь, который содержит слова и их весовые коэффициенты [1-2]. Затем суммирует результат и получает вес всего текста, по которому и определяют является указанный текст положительным или отрицательным. Данный метод является одним из наиболее распространенных, однако мир меняется и вместе с ним меняется и лексический словарный запас людей. И для более корректной обработки отзыва необходимо постоянно обновлять словарь, добавляя в него актуальные слова.

В связи с этим, было принято решение о проектировании информационной системы автоматизированной обработки отзывов методом интеллектуального анализа. И анализироваться будет именно проблематика отзыва клиента с учетом тональности текста. Автоматизация сможет сэкономить время, а методы анализа естественного языка, за счет непрерывного накопления данных, позволят корректно определять тематику отзыва, что поспособствует нахождению рациональных решений для возникающих проблем.

Целью исследования является изучение теоретических аспектов, касающиеся методов анализа естественного языка для обработки текста и проектирование автоматизированной системы обработки отзывов

веб-сайта. Предметом исследования являются модели проектирования системы обработки отзывов: ER-диаграмма, модель IDEF0, модель потоковых данных и диаграмма вариантов использования.

Основная часть. Определены основные этапы обработки текста отзыва методами анализа естественного языка:

1. Определение тональности отзыва на основе рейтинга. Для тренировки компьютера собирается датасет отзывов с текстом и рейтингом. При этом, каждая оценка имеет свою интерпретацию. Отзыв с оценкой 5 будет получать от сотрудника компании благодарность за положительный отзыв. А отзывы с оценкой от 4 до 1 будут подвержены детальному рассмотрению и потребуют от сотрудника принятия решения по устранению проблемной ситуации продукта или услуги.

2. Первичная очистка текста отзыва, для упрощения дальнейшей обработки включает в себя: токенизацию (разбиения строк на маленькие осмысленные элементы – токены), лемматизацию (данный метода позволяют привести слово к его первоначальной форме [3]), удаление стоп-слов (часто используемые слова, сильно не влияющие на смысловую нагрузку текста слова, такие как предлоги, частицы, местоимения и др), векторизация (данный процесс позволяет представить текст в виде числового вектора для того, чтобы облегчить обработку отзыва).

3. Тематическое моделирование. Самым распространенным способом определения тематики является метод тематического моделирования, в основе которого лежит построение тематической модели. Данная модель на входе получает коллекцию документов, представленных в текстовом виде. А на выходе формируется вектор их чисел, который состоит из оценок, определяющих степень принадлежности документа к определенным темам. Размерность вектора формируется либо через количество тем, которых были заданы на входе в модель, либо сгенерировано автоматически самой моделью.

Проектирование информационной системы

ER-модель предоставляет возможность визуализировать проектирование базы данных для рассматриваемой предметной области и представлена в виде диаграммы «сущность-связь».

Для рассматриваемой предметной области было выделено 7 основных сущностей:

1. Сущность «Клиент» содержит информацию о клиенте (ФИО, почта) и имеет связь «один ко многим» с сущностью «Отзыв», так как один клиент может оставлять неограниченное количество отзывов.

2. Сущность «Отзыв» содержит неструктурированные данные текста и определённую тематику, также имеет связь с сущностью «Сотрудник» и «Клиент», так как за определенным отзыв закрепляется специалист, занимающийся определенной тематикой и клиент, который написал этот отзыв.

3. Сущность «Тематика» имеет наименование проблемы (темы обращения) клиента и синонимы, отнесенные к данной тематике, полученные в результате использования методов анализа естественного языка. И имеет связь «один ко многим» с сущностью «Решение», так как тема обращения может содержать несколько вариантов решения.

4. Есть промежуточная сущность «ТематикиОтзывов» она соединяет сущности «Отзыв» и «Тематика», так как отзыв может содержать несколько тем, так и одну и ту же тему могут поднимать различные пользователи в своих отзывах.

5. Сущность «Сотрудник» включает в себя информацию о сотруднике (ФИО, электронная почта, должность). И имеет связь «один ко многим» с сущностями «Отзыв» и «Тематика», так как один специалист может решать некоторые смежные проблемы клиентов и иметь большое количество отзывов для принятия решения.

6. Сущность «Решение» имеет наименование тематики/проблемы и описание возможного варианта решения и имеет связь с сущностями «Тематика» и «Ответ на отзыв».

7. Сущность «Ответ на отзыв» содержит решение специалиста по конкретной тематике и направляется к клиенту, поэтому данная сущность имеет связь с такими сущностями как: «Решение», «Сотруднику» и «Клиент».

Моделирование процесса автоматизации обработки отзывов *методологией IDEF0* включает в себя 3 уровня. Основным процессом системы является обработка отзыва. Данный процесс позволяет определить тональность текста (положительный или отрицательный), также определить тематику отзыва и отправить соответствующему специалисту для принятия решения по конкретной тематике отзыва.

В качестве «Входов» были представлены следующие элементы: текст отзыва клиента, данные о сотрудниках компании (имя, фамилия, должность, какой тематикой занимается) и словарь стоп-слов (содержит часто употребляющиеся слова, не несущие большую смысловую нагрузку для текста, так как междометия, предлоги и т.д.)

В качестве «Управления» определены: язык программирования, соответствующая для него документация (информация о выполняемых функциях, методах и т.д. языком программирования) и библиотеки языка программирования необходимые для реализации системы.

К «Механизмам» отнесены – информационная система и сотрудники. Но, так как речь идет об автоматизации системы, то практически все функции выполняет информационная система.

В качестве «Выходов» получаем: определенную тематику отзыва и соответствующий ответ специалиста по данной проблеме.

На втором уровне выполняется декомпозиция основного процесса и включает в себя 5 блоков (представлена на рисунке 1):

- определение тональности (на данном этапе был получен отзыв и его рейтингом, определенный клиентом, и в зависимости от оценки получаем тональность текста отзыва и отправляем сам текст на следующую обработку);
- первичная обработка и очистка текста отзыва (используется для облегчения последующего процесса обработка и более подробно будет рассмотрен на следующем уровне декомпозиции);
- определение тематики отзыва (определение на основе методов анализа естественного языка тематики и соответствующие ей синонимы);
- направление отзыва специалисту предусматривает определение сотрудника по теме, которой он занимается;
- анализ отзыва сотрудником (в ходе данного этапе определяется решение по тематике клиента).

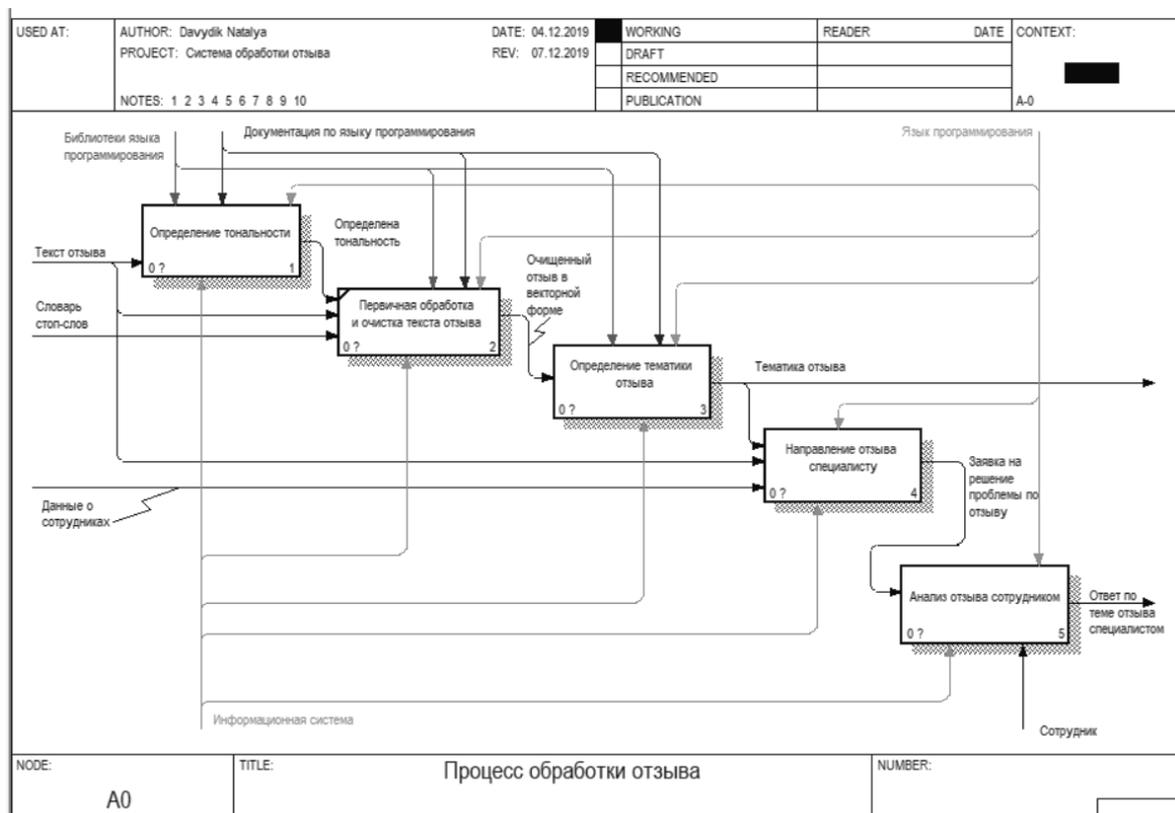


Рисунок 1 – Декомпозиция основного блока

На третьем уровне осуществляется декомпозиция блоков второго уровня с более подробным описанием действий.

Рассмотрим декомпозицию блока «Первичная обработка и очистка текста отзыва». На входе имеем текст отзыва и словарь стоп-слов, а на выходе – очищенный отзыв в векторной форме. Сам блок делится на 4 блока:

- удаление знаков пунктуации (токенизация);
- приведение слов к их первоначальной форме (лемматизация);
- удаление стоп-слов;
- векторизация отзыва.

Далее рассмотрим декомпозицию блока «Направление отзыва специалисту». В качестве входов отображены тематика отзыва, данные о сотрудниках и текст отзыва, а на выходе – заявка на решение проблемы по отзыву. Декомпозиция блока содержит 2 элемента:

- определение специалиста, отвечающего за определенную проблематику/тематику;
- формирование заявки специалисту на решение проблемы и ее отправка.

И последним рассмотрим декомпозицию блока «Анализ отзыва сотрудником». В качестве входов отображена заявка на решение проблемы по отзыву, полученную на выходе из предыдущего блока. Декомпозиция блока содержит 3 элемента:

- Анализ отзыва и полученных данных сотрудником.
- Принятие решений по данной тематике
- Составление ответа по проблеме и отправление ответа клиенту.

Далее рассмотрим моделирование с использованием *методологии DFD*. Диаграмма потоковых данных системы представлена на нулевом и первом уровне.

На нулевом уровне представлен основной процесс обработки отзыва. В качестве внешних сущностей определены:

- Клиент, который оставляет отзыв на сайте и получает ответ по представленной тематике.
- Сотрудник, который принимает решение по отзыву клиента и дает ответ на отзыв.

На уровне 1 представлена декомпозиция процесса обработки отзыва и отображена на рисунке 2. Основной блок разбивается на 5 процесса: определение тональности, первичная обработка и очистка текста отзыва, определение тематики отзыва, направление отзыва специалисту, формирование ответного письма клиенту.

Стрелками отображаются потоки данных в системе, которые указывают входящую и исходящую информацию.

Также в диаграмме использованы 2 хранилища данных:

- словарь стоп-слов (содержит часто употребляющиеся слова, не несущие большую смысловую нагрузку для текста, так как междометия, предлоги и т.д.);
- информация о сотрудниках (содержит информацию о сотрудниках, с помощью которой определяется сотрудник способный решить проблем клиента).

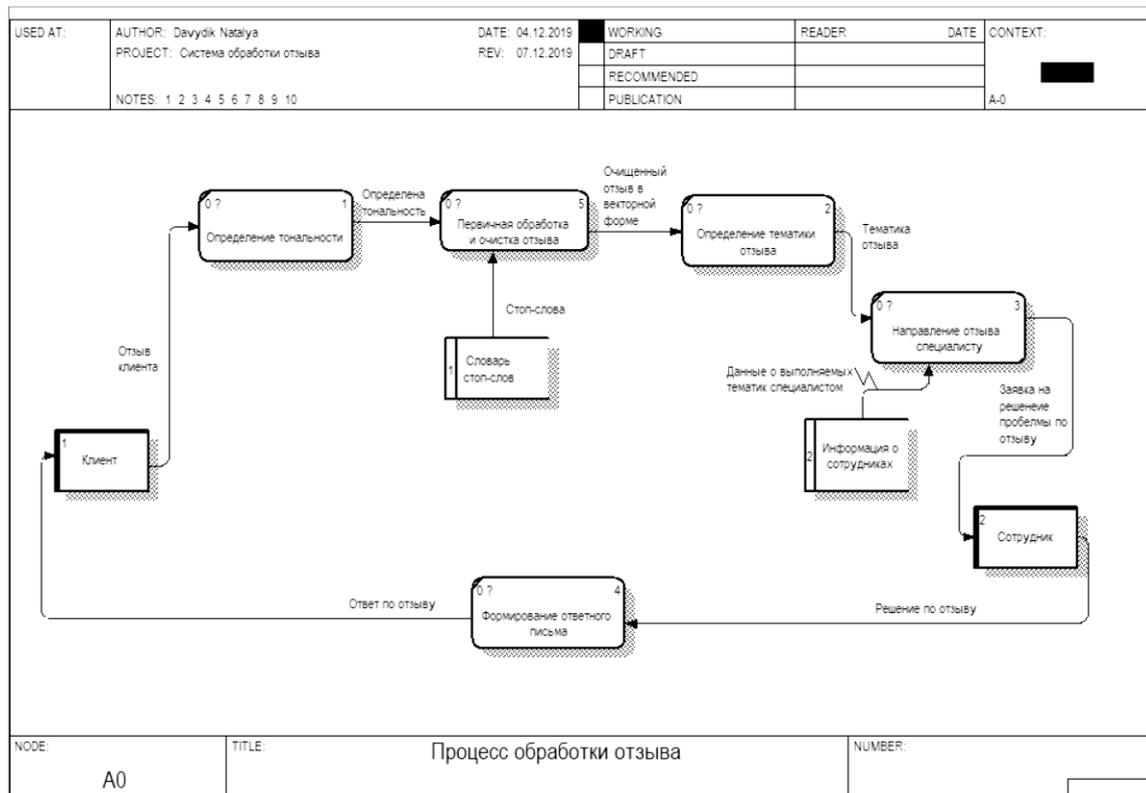


Рисунок 2 – Процесс обработки отзыва первый уровень

Use-case диаграмма – описание действий выполняемых системой, являющиеся значимыми для конкретного актера. А варианты использования (use case) отображены в виде эллипсов с именами [4].

Для проектируемой системы автоматизации процесса обработки отзывов веб-сайта были выделены 3 актера:

1. Клиент. Оставляет отзыв на сайте и получает ответ от сотрудника по тематике отзыва.
2. Информационная система. Принимает отзыв клиента, определяет тональность отзыва, осуществляет первичную очистку текста, формирует тему обращения клиента и отправляет заявку соответствующему специалисту для решения проблемной ситуации.
3. Сотрудник. Обрабатывает заявку и полученные данные по отзыву клиента, решает проблемную ситуацию и отправляет ответ по тематике клиенту.

Заключение. На сегодняшний день, практически у каждой организации есть официальный сайт, который предоставляет возможность не только разместить информацию о компании и предоставляемых товарах или услугах, а также получить реакцию клиента на осуществляемую деятельность. Одним из видов такой реакции является отзыв клиента.

Но не всегда клиенты оставляют положительные отзывы, что в дальнейшем может негативно повлиять на решение других людей. Таким образом, важно уметь работать с отзывами, обрабатывать их и давать своевременный ответ и решение по выделенной проблеме клиентом.

Проанализировав особенности и характеристики доступных онлайн сервисов «Анализ писем» и «Tone Analyzer», можно сделать вывод, что данные системы только анализируют тональности текста и не предусматривают выявление тематики, что увеличивает время сотрудника компании на обработку отзыва и решение проблемы.

В связи с этим, было принято решение о проектировании информационной системы автоматизированной обработки отзыва методом интеллектуального анализа.

В ходе исследования была спроектирована система автоматизации процесса обработки отзывов веб-сайта, с помощью моделей: ER-диаграмма, модель IDEF0, модель потоковых данных и диаграмма вариантов использования

Внедрение автоматизированной системы обработки отзыва методами интеллектуального анализа позволит сократить время на обработку отзыва сотрудником компании и своевременно реагировать на негативные отзывы и решать конфликтные ситуации.

Список литературы

1. Tone Analyzer [Электронный ресурс] – 2019 – Режим доступа: <https://tone-analyzer-demo.ng.bluemix.net/>. – Дата доступа: 25.11.2019.
2. Сервис «Анализ писем» [Электронный ресурс] – 2019 – Режим доступа: <http://www.analizpise.ru/>. – Дата доступа: 25.11.2019.
3. Toman, M. “Influence of word normalization on text classification”/ M. Toman, R. Tesar, and K. Jezek; Proceedings of InSciT, pp. 354–358, 2006.
4. Use Case Diagrams [Электронный ресурс] – 2018 – Режим доступа: <https://www.smartdraw.com/use-case-diagram/>. – Дата доступа: 04.12.2019.