

УДК 004.9

РЕШЕНИЕ ПРОБЛЕМЫ ДУБЛЕЙ В СПРАВОЧНИКАХ НА ПРИМЕРЕ БАЗЫ ДАННЫХ «1С:БУХГАЛТЕРИЯ» КОМПАНИИ ООО «РЕВЕРС ПЛЮС»

Фальтина Е.В., студентка гр. ПИм-221, II курс

Научный руководитель: Буйная Е.В., к.э.н, доцент

Кузбасский государственный технический университет
имени Т.Ф. Горбачева
г. Кемерово

Перед началом работы с программой любой организации необходимо заполнить первичную информацию. Часто в компаниях нет жесткого регулирования пополнения таких данных. В связи с этим возникает проблема дублей, которая ведет к производным сложностям учета: пересортице, отрицательным остаткам, избыточной инвентаризации.

Поэтому очень важно систематизировать данные и исправить все ошибки, чтобы не потерять прибыль из-за проведения операций не с теми данными.

Для решения данной проблемы хорошо подойдет метод косинусной дистанции [1]. Косинусная дистанция (расстояние) — это когда из единицы вычитаем косинусное подобие. Косинусное подобие — метрика, используемая для определения сходства между двумя векторами. Она вычисляется как косинус угла между двумя векторами и может принимать значения от -1 до 1 . Значение 1 означает полное сходство, а значение -1 — полное несходство.

Данный метод заключается в предварительной обработке данных, представлении их в виде векторов и расчета косинусной дистанции между ними. Далее анализируются полученные числа и выбираются самые наименьшие значения — соответствующие им вектора-данные и будут ближайшими по смыслу словами к вводимым строкам. Пользователь, видя на экране похожие фразы, сможет выбрать какое-то из них и таким образом избежать дублирования.

Основная программа будет написана на Python, а пользователь будет работать в программе «1С: Бухгалтерия». Для взаимодействия между двух этих программ будет использовано расширение Бром API (Brom) [2]. Это интеграционная компонента, которая предоставляет программный интерфейс взаимодействия (API) с информационными системами на базе 1С:Предприятие. Компонента представляет собой расширение (файл формата CFE), которое устанавливается поверх произвольной конфигурации 1С. На стороне Python используется библиотека, которая обеспечивает взаимодействие с расширением «Бром» для 1С:Предприятие. Библиотека позволяет:

- получать данные различных ссылочных коллекций из 1С (справочников, документов, ПВХ и тд...) с учетом сложных условий отбора и сортировки;
- модифицировать данные коллекций на стороне 1С (добавлять, удалять и редактировать элементы справочников, документов и пр. объектов);
- получать и изменять значения констант на стороне 1С;
- получать данные из 1С с помощью произвольных запросов на языке 1С:Предприятие;
- вызывать процедуры и функции серверных модулей 1С (общих модулей и модулей менеджеров) с передачей произвольного количества параметров вызова и с передачей полученного значения обратно в качестве результата вызова;
- исполнять произвольный код 1С на стороне сервера с возможностью передачи параметров и получения возвращаемого значения;
- взаимодействовать с объектами ссылочного типа на клиентской стороне.

Во время разработки программы были спроектированы бизнес-процессы. Основной бизнес-процесс содержит операции внесения данных, их обработку, расчет результатов и представление их в программе 1С. На Рисунок 1 представлена схема данного бизнес-процесса. Далее каждый процесс будет детализирован с помощью собственных схем.



Рисунок 1 - Бизнес-процесс «Основной»

Предобработка данных. На рисунке Рисунок 2 представлен процесс предобработки данных. Он содержит основные процедуры по изменению слов, для дальнейшей корректной работы с ними. Процесс транслитерации переводит все иностранные слова на русский язык. На следующем этапе удаляются все предлоги, частицы и т.п. После чего из данных удаляются все пунктуационные знаки – запятые, точки и т.д. Также необходимо все слова перевести в

начальную форму и разбить их на последовательности по две буквы (биплеты [3]). На этом процесс обработки будет завершен. Для работы с данными будут использоваться соответствующие библиотеки. За все процедуры отвечает технический пользователь, то есть их выполняет программа без участия пользователя.

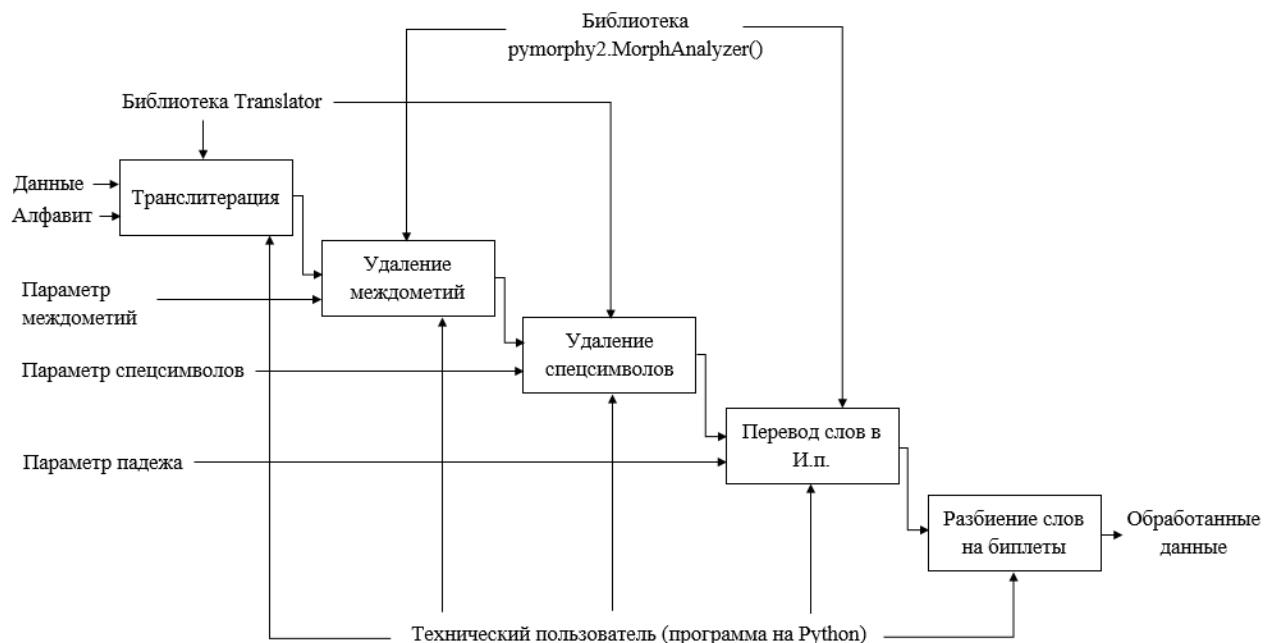


Рисунок 2- Бизнес-процесс «Предобработка данных»

Расчет результатов. В бизнес-процессе расчета результатов сначала данные переводятся в гистограммы-вектора [4], как показано на Рисунок 3. Далее производятся основные расчеты с обработанными данными по формулам – рассчитывается косинусная дистанция между векторами и сравнивается со всеми остальными. На вход подаются обработанные на предыдущем этапе данные и словарь – количество различных биплетов. На выходе выводятся похожие строки с наименьшим значением косинусной дистанции.



Рисунок 3 - Бизнес-процесс «Расчет результатов»

Список литературы:

1. Кристиан С. Пероне, «Машинное обучение:: косинусное сходство векторных пространственных моделей (часть III)» в Terra Incognita.
2. ITworks Group Бром API v.1.0-beta / ITworks Group [Электронный ресурс] // Клиент для Python : [сайт]. — URL: <https://brom.itworks.group/documentation/python/> (дата обращения: 19.03.2024).
3. 1C-RarusTechDay 2023 — 6-я открытая техническая конференция для специалистов 1C // 1C-Папус URL: (<https://rarus.ru/events/20230720-1c-rarustechday-2023-587081/>, б.д.) (дата обращения: 19.03.2024).
4. Г.И.Фалин, А.И.Фалин. О гистограмме и её свойствах. Математика в профильной школе. ФРАКТАЛ, 2013, №1, стр.44-62
5. 1С: Бухгалтерия [Электронный ресурс]. — Режим доступа <https://its.1c.ru/db/lector/browse/2/-1/1> (дата обращения: 19.03.2024)