

МАТЕМАТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ В ЭКОНОМИКЕ

УДК 338

JEL C69

ИССЛЕДОВАНИЕ И РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ АНАЛИЗА И ЗАГРУЗКИ ДАННЫХ ПРИ ПОМОЩИ НЕЧЕТКОГО СРАВНЕНИЯ СТРОК

Яхонтова Ирина Михайловна, канд. экон. наук, доц.
Нилова Надежда Михайловна, ст. преп.

Кубанский государственный аграрный университет, ул. Калинина, 13, Краснодар,
Россия, 350044; e-mail: i.yahontova@yandex.ru; nillova_nm@mail.ru

Предмет: нечеткое сравнение строк, или просто неточное сравнение – это процесс поиска похожих, но необязательно в точности совпадающих строк. Сущность использования методов нечеткого сравнения строк на предприятии заключается в предоставлении возможности эффективной загрузки данных первичных документов в учетную систему предприятия. *Цель:* современные требования к развитию бизнеса достаточно высоки с точки зрения организации электронного документооборота. Статья посвящена исследованию математических методов сопоставления текстовых строк и разработке методики анализа и загрузки данных в информационную систему предприятия. *Дизайн исследования:* поступающие первичные документы, вне зависимости от стандартности их оформления, должны быть загружены в учетную систему предприятия на первоначальном этапе. Такой подход позволит соблюдать принцип фиксирования информации у источника, а также повысить эффективность документооборота в целом за счет ускорения получения данных из документов всеми службами предприятия. Существует большое количество методов анализа текстовых фрагментов, однако наиболее эффективными из них считаются методы нечеткого сравнения строк. Таким образом, применение научного подхода при выполнении операций загрузки данных позволит обеспечить безошибочность и скорость данного процесса. *Результаты:* в результате проведенного исследования было выявлено, что повышение экономической эффективности учетных процессов предприятия напрямую зависит от скорости и безошибочности загрузки данных из первичной документации в информационную систему предприятия. Была разработана методика

анализа и загрузки данных при помощи методик нечеткого сравнения строк.

Ключевые слова: учетные процессы, первичная документация, данные, метод, управление, информационная система, нечеткое сравнение строк, анализ, загрузка.

DOI: 10.17308/meps/2078-9017/2023/11/8-19

Введение

В условиях цифровой трансформации экономики повышение эффективности работы предприятий напрямую зависит от цифровизации бизнес-процессов. Причем речь идет не только о стандартизации исполняемых процессов, но и внедрении подхода к организации работы с максимально возможным количеством автоматизированно выполняемых операций. Так, например, информационный процесс ввода информации трансформируется в процесс загрузки данных. Это позволит обеспечить безошибочность вводимых данных, а также улучшить документопотоки в целом уже на первоначальном этапе жизненного цикла документа.

Совершенствование документооборота всегда выступало одним из приоритетных направлений повышения эффективности работы предприятия. Результаты исполнения процессов предприятия, будь то основные, обеспечивающие или управленческие, отражаются в различного рода документах. Сопровождение документа на всех этапах жизненного цикла предприятия – от создания и регистрации и до прекращения деятельности фирмы, с применением современных научных методик и цифровых решений позволит сделать документопотоки оптимальными за счет сокращения временных и трудовых затрат, прежде всего [2]. С целью же оптимизации процесса работы с документацией следует начинать с первоначального ввода содержимого документа в информационную систему предприятия. На этот счет существует достаточно большое количество научных исследований и практических разработок. Однако, следует предположить, что применение математических методов, таких как нечеткое сравнение строк, позволит резко увеличить эффективность данного процесса.

Современные тенденции в организации работы корпоративных информационных систем подразумевают выполнение автоматизированной загрузки данных из первичной документации с целью ускорения попадания данных из входящих документов в учетную систему. В процессе выполнения такой загрузки возможно дублирование данных, что в большинстве случаев недопустимо и приводит к серьезным, недопустимым ошибкам. Часто проблема поиска дублирующихся записей стоит в системах хранения и обработки учетных данных, где возможно частичное совпадение сведений о клиентах. Решить данную проблему средствами СУБД не представляется возможным. Все известные методы идентификации объектов в базах данных оперируют точным равенством сравниваемых полей и бессильны при нали-

чии ошибок и пропусков данных. Поэтому все чаще специалисты склоняются к применению методов нечеткого сравнения строк при загрузке данных [5].

Методология исследования: математические методы анализа и нечеткого сравнения строк, разработки и исследования отечественных и зарубежных ученых в области управления процессом анализа и загрузки данных.

Моделирование и разработка информационной системы

Нечеткое сравнение строк, или просто неточное сравнение – это процесс поиска похожих, но необязательно в точности совпадающих строк. Сущность использования методов нечеткого сравнения строк заключается в предоставлении пользователю возможности определения идентичности (схожести) двух наборов данных.

Применение методов нечеткого сравнения строк в информационной системе предприятия реализуется с помощью популярных инструментальных сред разработки цифровых решений, например, «1С» или «Python».

Таким образом, применение научного подхода при выполнении операций загрузки данных позволит обеспечить безошибочность и скорость данного процесса.

В современном предприятии, осуществляющем торгово-закупочную деятельность, предполагается постоянный поиск поставщиков товара с целью расширения ассортимента, оптимизации закупочной логистики, поиска предложений по более низким ценам. В результате один и тот же товар (или товары) могут поставлять разные поставщики. В настоящее время все торговые предприятия оснащены товароучетной информационной системой, в которой осуществляется учет торговых операций, а оборот отдельных видов и категорий товаров обязательно учитывается в государственных информационных системах, таких как ЕГАИС, Честный знак и т.д. При этом учетному подразделению, осуществляющему ввод первичной (закупочной) документации, ставится задача по вводу первичных закупочных документов в информационную систему предприятия. Для автоматизации этих процессов широко начинают применяться системы электронного документооборота (ЭДО) и системы автоматизированной загрузки первичной информации. Именование товаров в первичных документах у поставщиков может очень сильно отличаться, а также слова в названии могут быть переставлены местами, иметь специальные сокращения, также часто применяют в названии размер или тип групповой упаковки – например, «20 шт/уп» и т.д.

Проблемой этих загрузок является сопоставление (поиск и связь) загружаемой номенклатуры с имеющейся в базе данных предприятия для принятия решения о создании новой записи в справочнике «Номенклатура», особенно при отсутствии у товара глобального уникального реквизита – например, штрихового кода стандарта «EAN-13» или кода «GTIN» – глобальный номер товарной продукции в единой международной базе товаров «GS1» – международная организация, ведающая вопросами стандартизации учёта и штрихового кодирования логистических единиц. Так как примене-

ние идентификаторов типа «GTIN» и(или) «КиЗ» обязательно только для маркируемой продукции средствами идентификации и подлежащих учету в системах «Честный знак», «МДЛП», «ЕГАИС» и т.д., остальные товары, которые также представляют значительный объем (например, промышленные товары метизной группы), особенно производимые малыми партиями и (или) субъектами малого предпринимательства, не имеют уникальных отличительных признаков, что, при бесконтрольной автоматизированной загрузке, приводит к появлению дублей номенклатурных единиц в номенклатурном справочнике.

Решаются эти проблемы путем первичного ручного сопоставления наименований, применение различных ключевых составных признаков – например, структуры («Код поставщика», «Поставщик», «Артикул»), что приводит к чуть избыточному размеру хранения информации, но работает только при постоянном поставщике. Вторым способом решения являются автоматизированные методы поиска и рекомендуемого сопоставления.

Системы электронного обмена документами еще применяются не часто – в основном для оборота маркированного товара, а довольно часто используется выгрузка первичных документов в формате «XLS». Загрузка такого формата чаще всего имеет 2 проблемы:

- отличия расположения колонок данных у разных поставщиков;
- разбиение документа на печатные страницы, с выводом постраничного итога.

Следовательно, актуальна разработка собственного цифрового решения автоматизированного анализа и загрузки первичной документации.

Рассмотрим предметную область исследования. Оптимизация документопотоков, а соответственно, и базы учетных данных предприятия начинается, прежде всего, с поиска и устранения дублирующихся записей. Поиск по полному совпадению наименований номенклатур, контрагентов и т.п. часто не дают желаемых результатов. Поэтому нечеткое сопоставление записей в настоящее время выступает одним из очевидно актуальных способов работы с базами данных любого предприятия. Согласно исследованиям ученых в области организации оптимальных баз данных, следует выделять требования к алгоритму нечеткого сопоставления строк как части алгоритма нечеткого сопоставления записей. Так поиск всех возможных дублирующихся наименований с учетом пробелов и дефисов, с учетом возможных сокращений и пропусков в части имени и максимально возможная простота алгоритма поиска, анализа и сравнения можно обозначить как приоритетные направления повышения эффективности базы документов предприятия [3].

Что касается научного обоснования применения алгоритмов нечеткого сравнения строк, фонетических или орфографических, то выбор наиболее подходящего возможен только с учетом детального изучения объекта исследования и предметной области.

Объектом исследования данной работы определена учетная систе-

ма предприятия [1]. Документационное оформление хозяйственных операций выступает процессом с меняющейся спецификой в зависимости от сферы деятельности предприятия и сложившихся стандартов оформления документов. Вне зависимости от вышеперечисленных условий, часто специалисту приходится сталкиваться с проблемами организации грамотного оформления документов в информационной системе предприятия. Так самой распространенной проблемой выступают точность и оперативность внесения данных в учетную систему. При обработке больших объемов первичной документации: товарных накладных, счетов-фактур, актов приема-передачи и т.д. сложно в кратчайшие сроки отразить в учете верные данные в соответствии с нормативными требованиями и спецификой конкретной корпоративной системы. И главной проблемой выступает, безусловно, всецелая зависимость от человеческого фактора. При ручном вводе данных нередки неточности, опечатки, ошибки, дублирование данных. Все это отражается потом в отчетности.

В корпоративных информационных системах, построенных на основе «1С: Предприятие 8.3», возможна загрузка данных из первичных документов. Однако существенным недостатком процесса загрузки выступает анализ текстовых фрагментов только по штрихкоду или полному соответствию. То есть, если специалист предприятия допустил ошибку и завел номенклатурную позицию одного и того же товара дважды или даже трижды, сокращая название или набирая его с маленькой или большой буквы, то поиск соответствий текста не даст результата. То есть, специалист таким образом может завести очень много карточек одного и того же товара, порождая засорение базы. Это влечет за собой некорректное отображение остатков, дублирование записей, проблемы с ведением бухгалтерского учета и проведением инвентаризаций. Таким образом, необходимо решать проблему с применением других нестандартных методик поиска текста в названиях номенклатурных позиций. И одним из таких методов выступает нечеткий поиск.

Нечеткий поиск обычно входит в состав любой поисковой системы, несмотря на то, что реализация его алгоритма значительно сложнее, чем просто поиск по точному совпадению.

В языках программирования строки сравниваются очень просто если строка отличается хотя бы на один символ, то возвращает «false». Но чаще всего нужно получать не просто дискретное значение, а дифференцированное – например, в процентах, потому как строки «тест» и «тестирование» гораздо ближе к друг другу, чем «тест» и «абвг». Для решения данной задачи существует множество различных методов и алгоритмов нечеткого сравнения строк.

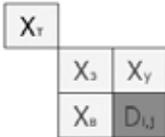
Задача нечеткого поиска может ставиться, как по заданному слову найти в тексте размера «N» все слова, совпадающие с этим словом (или начинающиеся с этого слова) с учетом «k» возможных различий». Например, при запросе «Машина» с учетом двух возможных ошибок най-

ти слова «Машинка», «Машина», «Малина», «Калина» и так далее.

Подобные алгоритмы обычно имеют особую численную оценку – так называемую «метрику» – т.е. функцию, отображающую «расстояния» между двумя словами, по которой можно судить о степени их сходства. Математическое определение метрики – функция расстояния между элементами набора. Если расстояние равно нулю, то элементы эквивалентны, иначе они отличаются друг от друга. Существует несколько метрик расстояния, но в алгоритмах машинного обучения и анализа чаще всего используются «расстояние Минковского», «расстояние Хемминга», «расстояние Левенштейна» и «расстояние Дамерау–Левенштейна» [4].

Обзор существующих математических методов нечеткого сравнения строк показал, что самым оптимальным методом поиска является метод – расстояние Дамерау–Левенштейна». Расстояние Левенштейна активно используется для исправления ошибок в словах, поиска дубликатов текстов, сравнения геномов и прочих полезных операций с символьными последовательностями. Данный метод выдает самые лучшие результаты, относительно несложные алгоритмы вычисления, а также нормированный результат, который можно переводить в проценты.

«Расстояние Дамерау–Левенштейна» (названо в честь ученых Фредерика Дамерау и Владимира Левенштейна) — это мера разницы двух строк символов, определяемая как минимальное количество операций вставки, удаления, замены и транспозиции (перестановки двух соседних символов), необходимых для перевода одной строки в другую. Является модификацией «расстояния Левенштейна»: к операциям вставки, удаления и замены символов, определенных в «расстоянии Левенштейна» добавлена операция транспозиции (перестановки) символов.



$$D_{ij} = \min(X_i + 1, X_j + 1, X_i + C_{\text{замена}}, X_i + C_{\text{удаление}})$$

$$C_{\text{замена}} = \begin{cases} 1, & \text{если } S_1[i] \neq S_2[j] \\ 0, & \text{иначе} \end{cases}$$

$$C_{\text{транспозиция}} = \begin{cases} 1, & \text{если } S_1[i] = S_2[j-1] \text{ и } S_1[i-1] = S_2[j] \\ \infty, & \text{иначе} \end{cases}$$

Рис. 1. Математическое представление метода Дамерау–Левенштейна

Расстояние по Левенштейну между двумя словами определяется как минимальное количество односимвольных операций, необходимых для преобразования одного слова в другое. «Расстояние Дамерау–Левенштейна»

работает точно так же, как и «расстояние Левенштейна», но здесь добавлена четвертая односимвольная операция, которая называется «транспозиция» – замена местами двух символов (например актер – катер). Это частично решает проблему больших расстояний при перестановке слов, но усложняет алгоритм нахождения минимального числа операций.

Чтобы вычислять такое расстояние, достаточно немного модифицировать алгоритм нахождения обычного «расстояния Левенштейна» следующим образом: хранить не две, а три последних строки матрицы.

Анализ функционального состава стандартных программных средств для ведения бухгалтерского учета на предприятиях и проведенное исследование методов нечеткого сравнения строк позволяет предложить следующую методику анализа и загрузки данных в учетную систему. Методика управления данным процессом предполагает специализированный подход к построению алгоритма загрузки. Данный подход включает в себя следующие шаги:

- нормализация источника загрузки – а именно нормализация структуры загружаемой записи (сопоставление колонок), описание спецификации областей загрузки;
- чтение данных во временной таблице и пометка всех строк как новых элементов;
- при наличии уникального ключевого значения – сопоставление загружаемых данных с имеющимися в базе данных, пометка оставшихся, как новых элементов к созданию;
- поиск в отмеченных новых элементах по полю «наименование» с применением метода нечеткого сравнения строк и вывода коэффициента совпадения;
- принятие решения по границе признака совпадения о необходимости создания элементов;
- при множественном совпадении – вывод найденных для выбора.

Методика управления процессом анализа и загрузки данных в информационную систему предприятия на основе нечеткого сравнения строк представлена на рисунке 2.

Представленная методика позволяет быстро и без потерь данных загрузить информацию из первичной документации в учетную систему предприятия. Помимо штатных средств поиска на соответствие текстовых фрагментов, а именно по полному совпадению и штрихкоду, в разрабатываемой информационной системе планируется реализовать поиск по нечеткому совпадению, что значительно повысит эффективность процесса загрузки.

Таким образом, такая методика позволит рационализировать базу первичной документации и базу данных за счет исключения дублирующихся записей. Сам же процесс внесения данных из первичных документов станет более оптимальным в связи с применением научного подхода к нему. Так

автоматизированная загрузка данных об операциях из первичных документов сократит трудовые и временные затраты, а также уменьшит количество ошибок и несоответствий в учетных записях предприятия.

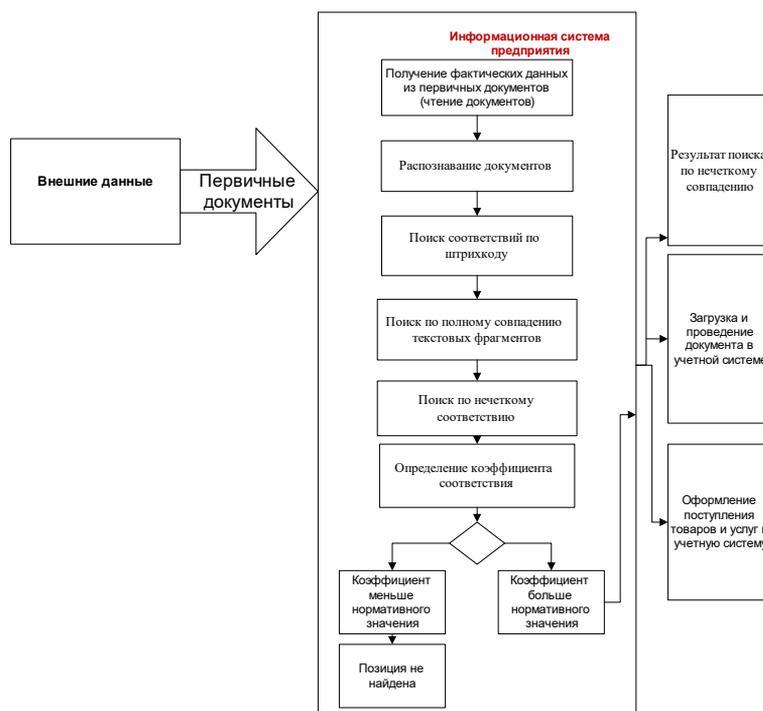


Рис. 2. Методика управления процессом анализа и загрузки данных в информационную систему предприятия на основе нечеткого сравнения строк

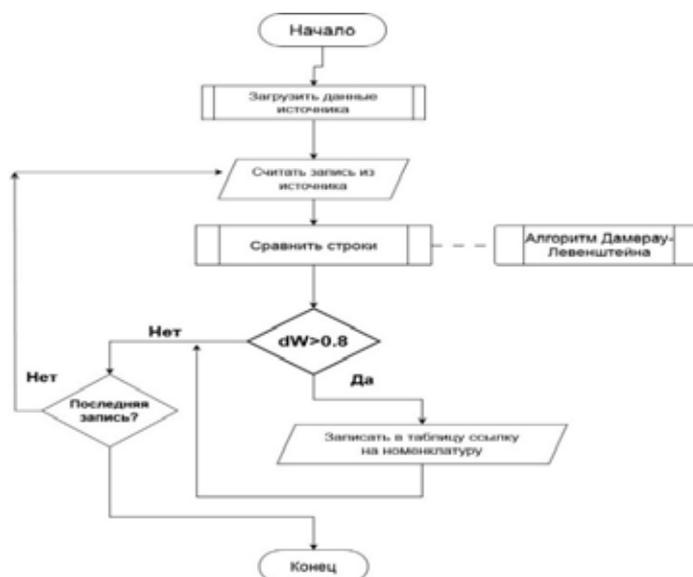


Рис. 3. Алгоритм сопоставления наименований товаров с данными из загружаемого источника

Алгоритм загрузки подразумевает построчное чтение данных из источника и вычисление значений расстояний для каждой записи номенклатуры. Значение расстояние определяется методом «расстояние Дамерау–Левенштейна», который преобразуется в коэффициент похожести. При совпадении от 0.8 считается, что данные похожи, и ссылка записывается в таблицу.

Исходя из приведенной методики, можно предложить разработать информационную систему для загрузки данных в документ «Поступление товаров» из внешних источников, представляющих собой файлы формата xls или dbf. Система должна реализовать интерактивную настройку связи колонок источника с колонками назначения, а также интерактивное задание областей загрузки. Поиск номенклатуры для сопоставления загружаемых данных будет осуществляться при помощи методов нечеткого поиска строк, основанных на методике «расстояние Дамерау–Левенштейна». В качестве инструментального средства разработки предлагается среда «1С: Предприятие 8.3», что объясняется программной совместимостью с большинством корпоративных информационных систем страны.

Заключение

В результате проведенного исследования разработана методика управления процессом анализа и загрузки данных в информационную систему предприятия на основе нечеткого сравнения строк. По итогам работы следует сделать вывод о том, что управление процессом анализа и загрузки данных с применением разработанной методики позволит повысить эффективность учетных процессов предприятия в целом.

Список источников

1. Айрапетов С.С. Современные методы оптимизации бизнес-процессов / Айрапетов С.С., Нилова Н.М. // *Информационное общество: современное состояние и перспективы развития: сборник материалов XIII международного форума*, Краснодар, 13–18 июля 2020 года. Краснодар, КубГАУ, 2020, с. 215–217.
2. Великанова Л.О. Применение мобильных технологий для автоматизации бизнес-процессов на торговом предприятии / Л.О. Великанова, А.Р. Гайвук // *Экономическое прогнозирование: модели и методы: материалы XII международной научно-практической конференции*, Воронеж, 17–19 ноября 2016 года. Воронеж, Воронежский Центральный научно-технический институт – филиал ФГБУ «РЭА» Минэнерго России, 2016, с. 314–317.
3. Исследование проблемы оценки эффективности инвестиционных проектов в условиях цифровой трансформации бизнеса / Л.О. Великанова, И.М. Яхонтова, А.В. Коваленко, А.С. Маликов // *Современная экономика: проблемы и решения*, 2022, no. 1(145), с. 38–48.
4. Исследование и разработка информационной системы ранжирования альтернатив / Е.В. Попова, П.А. Кочкарова, Н.С. Курносова [и др.] // *Современная экономика: проблемы и решения*, 2023, no. 7 (163), с. 17–25.
5. Крамаренко Т.А. *Разработка бизнес-приложений: учебник* / Т.А. Крамаренко, Е.А. Иванова. Краснодар, КубГАУ, 2021. 234 с.
6. Кузнецова О.Д. Система документооборота предприятия в эффективной бизнес-системе / Кузнецова О.Д., Емельянова С.Е., Нилова Н.М. // *Информационное общество: современное состояние и перспективы развития: сборник материалов XIII международного форума*, Краснодар, 13–18 июля 2020 года. Краснодар, КубГАУ, 2020, с. 242–244.
7. Мортон С.Т. Обработка неструктури-

рированных текстов. Поиск, организация и манипулирование / С.Т. Мортон, Г.С. Игрерсолл, Э.Л. Фэррис // ДМК Пресс, 2015. 414 с.

8. Попова Е.В. Цифровизация в экономике: проблемы и перспективы / Е.В. Попова, В.А. Тютюнникова, А.А. Коробкин // *Вектор современной науки: сборник тезисов по материалам Международной научно-практической конференции студентов и молодых ученых*, Краснодар, 15 ноября 2022 года. Краснодар, Кубанский государственный аграрный университет имени И.Т. Трубилина, 2022, с. 643-645.

9. Прикладная и компьютерная лингвистика / ред. Митренина О.В., Ландо Т.М. // *Ленанд*, 2016. 320 с.

10. Татьяна А.А. Программное обеспечение для автоматизации бухгалтерского учета / А.А. Татьяна, Д.Н. Савинская // *Информационное общество: современное состояние и перспективы развития: сборник материалов VI международного форума*, Краснодар, 28–30 декабря 2015 года / Редакцион-

ная коллегия: Попова Е.В., Замотайлова Д.А., Курносое С.А., Рахметова Р.У., Рогачев А.Ф., Тинякова В.И., Темирбулатов П.И., Тамбиева Д.А., Топсахалова Ф.Н-Г., Улезько А.В. Краснодар, Кубанский государственный аграрный университет имени И.Т. Трубилина, 2016, с. 7-9.

11. Федоров Н.В., Яхонтова И.М. Создание бланков стандартных документов на VBA // *Информационное общество: современное состояние и перспективы развития: сборник материалов VII Международного форума*. Краснодар, КубГАУ, 2016, с. 195-198.

12. Якунина М.С. Применение информационно-компьютерных технологий в экономике / М.С. Якунина, Л.О. Великанова // *Информационное общество: современное состояние и перспективы развития: сборник материалов IX студенческого международного форума*, Краснодар, 26–30 июня 2017 года. Краснодар, Кубанский государственный аграрный университет имени И.Т. Трубилина, 2017, с. 153-155.

RESEARCH AND DEVELOPMENT OF AN INFORMATION SYSTEM FOR ANALYZING AND LOADING DATA USING FUZZY STRING COMPARISON

Yakhontova Irina Mikhailovna, Cand. Sci. (Econ.), Assoc. Prof.

Nilova Nadezhda Mikhailovna, Assist. Prof.

Kuban State Agrarian University, Kalinina st., 13, Krasnodar, Russia, 350044; e-mail: i.yahontova@yandex.ru; nilova_nm@mail.ru

Importance: fuzzy string comparison, or simply inaccurate comparison, is the process of searching for similar, but not necessarily exactly matching strings. The essence of using fuzzy string comparison methods in an enterprise is to provide an opportunity for efficient loading of primary document data into the enterprise accounting system. *Purpose:* modern requirements for business development are quite high from the point of view of the organization of electronic document management. The article is devoted to the study of mathematical methods for comparing text strings and the development of methods for analyzing and uploading data into the enterprise information system. *Research design:* incoming primary documents, regardless of the standard of their design, must be uploaded to the company's accounting system at the initial stage. This approach will allow observing the principle of recording information at the source, as well as improving the efficiency of document management in general by speeding up the receipt of data from documents by all services of the enterprise. There are a large number of methods for analyzing text fragments, but the most effective of them are methods of fuzzy string comparison. Thus, the application of a scientific approach when performing data loading operations will ensure the accuracy and speed of this process. *Results:* as a result of the conducted research, it was revealed that the increase in the economic efficiency of the accounting processes of the enterprise directly depends on the speed and error-free loading of data from the primary documentation into the information system of the enterprise. The methodology of data analysis and loading was developed.

Keywords: accounting processes, primary documentation, data, method, management, information system, fuzzy string comparison, analysis, loading.

References

1. Airapetov S.S. Modern methods of optimization of business processes / Airapetov S.S., Nilova N.M. *Information society: current state and prospects of development: Collection of materials of the XIII International Forum*, Krasnodar, July 13-18, 2020. Krasnodar, KubSAU, 2020, pp. 215-217. (In Russ.)

2. Velikanova L.O. Application of mobile technologies for automation of business processes at a commercial enterprise / L.O. Velikanova, A.R. Gaivuk. *Economic forecasting: Models and methods : Materials of the XII International Scientific and Practical Conference, Voronezh, November 17-19, 2016*. Voronezh, Voronezh Central Scientific and Technical Institute – branch of the Federal State Budgetary Institution «REA» of the Ministry of Energy of Russia, 2016, pp. 314-317. (In Russ.)
3. Investigation of the problem of evaluating the effectiveness of investment projects in the conditions of digital business transformation / L.O. Velikanova, I.M. Yakhontova, A.V. Kovalenko, A.S. Malikov. *Modern Economics: problems and solutions*, 2022, no. 1(145), pp. 38-48. (In Russ.)
4. Research and development of an information system for ranking alternatives / E.V. Popova, P.A. Kochkarova, N.S. Kurnosova [et al.]. *Modern economics: problems and solutions*, 2023, no. 7 (163), pp. 17-25. (In Russ.)
5. Kramarenko T.A. *Development of business applications: textbook* / T.A. Kramarenko, E.A. Ivanova. Krasnodar, KubSAU, 2021. 234 p. (In Russ.)
6. Kuznetsova O.D. Document management system of the enterprise in an effective business system / Kuznetsova O.D., Emelyanova S.E., Nilova N.M. *Information society: current state and prospects of development: Collection of materials of the XIII International Forum*, Krasnodar, July 13-18, 2020. Krasnodar, KubSAU, 2020, pp. 242-244. (In Russ.)
7. Morton S.T. *Processing of unstructured texts. Search, organization and manipulation* / S.T. Morton, G.S. Igrersoll, E.L. Farris. DMK Press, 2015. 414 p. (In Russ.)
8. Popova E.V. Digitalization in the economy: problems and prospects / E.V. Popova, V.A. Tyutyunnikova, A.A. Korobkin. *Vector of modern science: A collection of abstracts based on the materials of the International Scientific and Practical Conference of Students and Young scientists*, Krasnodar, November 15, 2022. Krasnodar, Kuban State Agrarian University named after I.T. Trubilin, 2022, pp. 643-645. (In Russ.)
9. Applied and Computational Linguistics. / ed . Mitrenina O.V., Lando T.M. // Lenand, 2016. 320 p. (In Russ.)
10. Tatyana A.A. Software for automation of accounting / A.A. Tatyana, D.N. Savinskaya. *Information Society: current state and prospects of development: Collection of materials of the VI International Forum*, Krasnodar, December 28-30, 2015 / Editorial Board: Popova E.V., Zamotailova D.A., Kurnosov S.A., Rakhmetova R.U., Rogachev A.F., Tinyakova V.I., Temirbulatov P.I., Tambieva D.A., Topsakhalova F.N.-G., Ulezko A.V. Krasnodar, Kuban State Agrarian University named after I.T. Trubilin, 2016, pp. 7-9 (In Russ.)
11. Fedorov N.V., Yakhontova I.M. Creation of forms of standard documents for VBA / O.Y. Fedorov, I.M. Yakhontova. *Information Society: current state and prospects of development: collection of materials of the VII International Forum*. Krasnodar, KubSAU, 2016, pp. 195-198. (In Russ.)
12. Yakunina M.S. Application of information and computer technologies in economics / M.S. Yakunina, L.O. Velikanova. *Information society: current state and prospects of development : Collection of materials of the IX Student International Forum*, Krasnodar, June 26-30, 2017. Krasnodar, Kuban State Agrarian University named after I.T. Trubilin, 2017, pp. 153-155. (In Russ.)