

## АДАПТИВНАЯ СИСТЕМА ДЛЯ ОПЕРАТИВНОЙ ОЦЕНКИ И КОРРЕКЦИИ ДОКУМЕНТОВ

**П. С. ВОЛЧЕЦКАЯ**

*(Белорусский государственный университет, г. Минск)*

**Аннотация.** В условиях постоянного роста количества документов и изменения их структуры возникает необходимость усовершенствования технологий обработки с целью сокращения расходов времени и средств. В докладе предлагается комплекс моделей и алгоритмов для построения адаптивной системы, в значительной мере решающей эту проблему на основе формализованных экспертных знаний.

**Ключевые слова:** интеллектуальная обработка текста, аналитические системы, системы поддержки принятия решений.

**Введение.** Принятие решений в информационном обществе основано на результатах анализа потоков документов, количество и объем которых постоянно увеличивается. Качество решений зависит от качества входных документов и эффективности средств их обработки. Критическим параметром является время синтеза решения [1].

Для крупных компаний разработаны достаточно эффективные дорогостоящие системы для анализа текста, например, SAS Visual Text Analytics. В государственных организациях, а также средних и малых компаниях проблема работы с документами не решена в должной мере по следующим причинам:

- постоянно увеличивающийся объем труда корректоров, редакторов, нормировщиков, преподавателей для проверки качества документов негативно влияет на качество результатов их труда;
- сложность проверки правильности терминологии, которая в XXI в. постоянно эволюционирует по мере технологических и других изменений (особенно в области IT);
- несоответствие защитных элементов или применение устаревшей терминологии в документе приводят к потере качества принимаемых на его основе решений [2];
- неполнота комплекта компонентов, составляющих современные документы, значительно увеличивает время синтеза решений [3, 4, 5].

Перечисленные факторы приводят к тому, что на вход специализированных систем типа SAS часто поступают априори нештатные по структуре и терминологии документы, снижающие эффективность решений, принимаемых на их основе.

В докладе предлагается адаптивная система для работы с документами, устраняющая негативные факторы и ориентированная на средние и малые компании.

**Постановка задачи.** Для постановки задачи обобщим типичные ситуации, характерные для современных организаций различного типа, включая научные и образовательные учреждения, издательства, рекрутинговые компании.

Пусть имеется организация, которая принимает решения на основе анализа входного потока цифровых документов  $D_1, D_2, \dots, D_k$ , формируется авторами  $A_1, A_2, \dots, A_k$  (людьми или ИИ типа GPT-4). Соответственно, сцена работы с документами включает авторов документа и корректора, который оценивает и, возможно, корректирует документ и отправляет его автору по каналу связи. Документы носят различный характер и структуру (научные отчеты, докладные, диссертации, дипломы и т.д.). Качество подготовки контента входных документов неизвестно. Документы принимаются корректором и результат обработки отправляется авторам.

Для ускорения понимания корректором правильности структуры, лексики, наличия средств защиты и передачи документа в систему автоматизированной обработки компании к документам предъявляются определенные требования корпоративного и государственного уровня.

Требуется разработать систему, обеспечивающую значительное сокращение времени и объема ручного труда на процессы преобразования входного документа неизвестного качества в документ, соответствующий государственным или корпоративным требованиям.

**Модели.** Как показано выше, поставленная задача носит междисциплинарный характер. Для такого рода задач обычно применяют системный подход, т.е. строят концептуальные модели, которые формируют цельный взгляд на проблему, и уточняют их в рамках методологий различных дисциплин до уровня программного кода.

Обобщая реальные ситуации, можно утверждать, что на верхнем уровне должна находиться модель сцены, включающей всех одушевленных и неодушевленных актеров:

$$\text{Scene} = (\text{Auth}, \text{Insp}, \text{Exp}, \text{sys}, \text{com}), \quad (1)$$

где Auth – автор документа;

Insp – лицо, оценивающее и корректирующее документ;

Exp – эксперт, знающий «горячие» требования к документу;

sys – искомая система;

com – коммуникации для общения участников.

Для практической реализации сцены необходимо уточнить состав и отношения участников.

Модель автора:

$$\text{Auth} = (\text{idAuth}, \text{doc}, \text{grdE} \langle V, U \rangle, \text{com}) \quad (2)$$

где  $\text{idAuth}$  – идентификатор автора;  
 $\text{doc}$  – созданный автором документ;  
 $\text{grdE} \langle V, U \rangle$  – оценка  $V$  документа экспертом или системой и соответствующее управляющее решение  $U$ .

Модель корректора:

$$\text{Insp} = (\text{idInsp}, \text{knInsp}, \text{com}), \quad (3)$$

где  $\text{idInsp}$  – идентификатор корректора;  
 $\text{knInsp}$  – знания корректора.

Модель эксперта:

$$\text{Exp} = (\text{idExp}, \text{knExp}, \text{com}), \quad (4)$$

где  $\text{idExp}$  – идентификатор эксперта;  
 $\text{knExp}$  – актуальные знания эксперта о конкретных типах документов.

Модель документа:

$$\text{doc} = (\text{idDoc}, \text{req}, K_1, K_2, \dots, K_n, V, U), \quad (5)$$

где  $\text{idDoc}$  – идентификатор документа определенного типа;  
 $\text{req}$  – требования к документу;  
 $K_1, K_2, \dots, K_n$  – компоненты документа определенного типа.

Модель целевой системы:

$$\text{sys} = (a_i, \text{lib}, \text{knMOD}, \text{dbIN}, \text{dbOUT}, \text{com}), \quad (6)$$

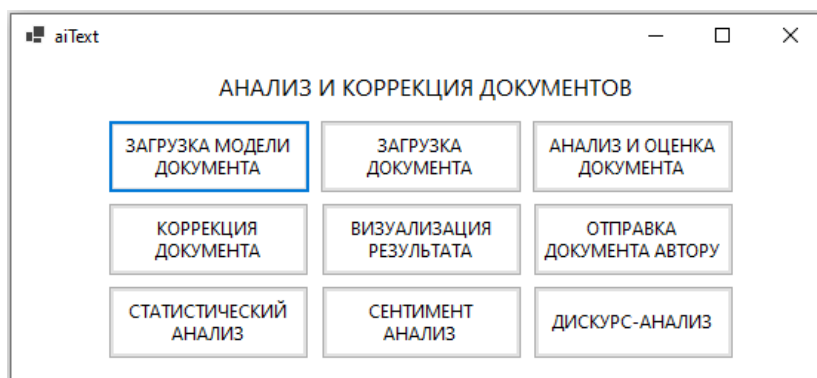
где  $a_i$  – управляющая система;  
 $\text{lib}$  – библиотека программных модулей, реализующих функции обработки текста;  
 $\text{knMOD}$  – база данных с моделями различных типов документов;  
 $\text{dbIN}$  – база данных входных документов;  
 $\text{dbOUT}$  – база данных выходных документов.

Для практического применения моделей необходимо разработать комплекс алгоритмов, обеспечивающих адаптацию целевой системы к характерной для настоящего времени динамике структуры документов.

**Алгоритмы.** Разработан следующий комплекс алгоритмов, ориентированный на реализацию в форме интероперабельных микросервисов:

- алгоритм управления сервисами системы;
- алгоритм формирования модели документа и построения соответствующей базы знаний на основе формализованных экспертных знаний;
- алгоритм чтения входного документа, определения его типа, оценки его качества и синтеза соответствующего управляющего решения на основе базы знаний;
- алгоритм коррекции документа в соответствии с управляющим решением;
- алгоритм визуализации результатов обработки документа для корректора;
- алгоритм записи откорректированного документа в выходную базу данных и отправка автору;
- служебные алгоритмы для автоматизации различных методов анализа документа для широкого спектра задач (статистический анализ, сентимент анализ и т.д.).

Алгоритмы реализованы на языке Python с применением библиотек Gensim, NLTK, NumPy, Pandas, Matplotlib в форме микросервисов и применены для построения как консольных, так и графических приложений. Например, на рисунке 1 представлен локальный графический вариант интерфейса, разработанный на базе WinForms.



**Рисунок 1. – Графический вариант главного меню**

Данный вариант меню обеспечивает доступ к программным и информационным ресурсам системы эксперту и корректору по отдельным паролям.

**Апробация системы.** Качество разработанного ПО определим на примере решения следующей практической задачи. Дано: 3 дипломные работы объемом 50 страниц в среднем каждая, поступивших корректору на предмет оценки их соответствия требованиям организации.

В результате измерений определено, что на проверку одного диплома корректор тратит в среднем 19 минут. Всего затраты времени составили 57 минут (3420 секунд).

Требуется провести сравнительный анализ качества работы и затрат времени на проверку дипломов контролером и системой. В систему экспертом предварительно загружена модель диплома DP2024. Результат решения задачи показан на рисунке 2. Система затратила на решение 41 секунду, что в 83 раза быстрее корректора.

Тип документа	Объем (стр)	Количество ошибок	Оценка документа	Управляющее решение
Диплом 1	48	34	Пять	Корректировать
Диплом 2	52	28	Шесть	Корректировать
Диплом 3	51	131	Три	Вернуть

Источник: dbIN  
 Тип документа: DP2024  
 Количество документов: 3  
 Время выполнения: 00:41:27

**Рисунок 2. – Результат обработки документов**

**Заключение.** На основании анализа проблемы и результата автоматизации можно утверждать, что в условиях изменчивости структуры и требований к документам адаптивные системы являются основным средством реагирования на турбулентность среды и сокращения затрат. Применение экспертных знаний для адаптации необходимо. В целом, такого рода системы можно рассматривать как дополнение адаптивными сервисами крупных аналитических систем, аналогичных SAS.

## ЛИТЕРАТУРА

1. Кузнецов, И.Н. Документационное обеспечение управления / И.Н. Кузнецов – М.: Юрайт, 2015. – 611 с.
2. Лейн, Х. Обработка естественного языка в действии / Х. Лейн, Х. Хапке, К. Ховард – СПб.: Питер, 2020. – 576 с.
3. О порядке оформления диссертации, диссертации в виде научного доклада, автореферата диссертации и публикаций по теме диссертации [Электронный ресурс]. – Режим доступа: Полное руководство по анализу текста [Электронный ресурс]. – Режим доступа: <https://asu-analitika.ru/analys-text>. – Дата доступа: 20.01.2024.
4. Полное руководство по анализу текста [Электронный ресурс]. – Режим доступа: <https://asu-analitika.ru/analys-text>. – Дата доступа: 22.02.2024.
5. Проект ГОСТ Р эталонная модель цифрового документооборота организации [Электронный ресурс]. – Режим доступа: <https://www.normacs.ru/Doclist/doc/235BA.html>. – Дата доступа: 05.02.2024.