

ИССЛЕДОВАНИЕ КЛЮЧЕВЫХ СЛОВ КАК ИНСТРУМЕНТ ОПТИМИЗАЦИИ УПРАВЛЕНИЯ ЭЛЕКТРОННЫМИ ДОКУМЕНТАМИ

Баканова Н.Б.¹, Усманова И.В.¹

¹ФГБОУ ВПО «Пензенский государственный университет», Пенза, Россия (440026, г. Пенза, ул. Красная, 40), e-mail: inoup@pnzgu.ru

В статье предлагается методика анализа текстов документов, являющаяся основой автоматизированной информационной системы аудита нормативных документов организации. Цель анализа — проверка соответствия положений, закрепленных в действующих нормативных документах, реальной ситуации, сложившейся в организации. Процедура анализа включает два последовательных этапа: исследовательский и аналитический. Целью исследовательского этапа является выделение в рассматриваемом тексте основных лингвистических конструкций и определение в них действующих субъектов, выполняемых ими действий, объектов, на которые направлены действия, а также свойств субъектов, объектов и действий. Полученные результаты являются исходными данными для второго (аналитического) этапа методики анализа текстов документов, целью которого является формулирование рекомендаций по оптимизации бизнес-процессов и электронных документопотоков. Предложенная методика анализа текстов документов позволяет привести внутренние нормативные документы в соответствие с реальным распределением функциональных обязанностей между сотрудниками, а также повысить эффективность функционирования организации в целом.

Ключевые слова: система электронного документооборота, ключевые слова, аудит нормативных документов, анализ текста, семантическая сеть, тезаурус, состав документов, интеллектуальные системы.

KEYWORD RESEARCH IS OPTIMIZATION TOOLS ELECTRONIC DOCUMENT MANAGEMENT

Bakanova N.B.¹, Usmanova I.V.¹

¹Penza State University, Penza, Russia (440026, Penza, Street Krasnaya, 40), e-mail: inoup@pnzgu.ru

In this paper we propose a text analysis methods, which is the basis of an automated information system audit regulations of the organization. The purpose of the analysis - to check compliance with provisions laid down in the existing normative documents, the real situation in the organization. The test procedure involves two successive stages: research and analytical. The aim of the research stage is the selection in the text under consideration the basic linguistic structures and the definition of actors in them, they perform actions, objects, on which the action, as well as the properties of subjects, objects and actions. The results obtained are the initial data for the second (analytical) phase of text analysis methods, which aims to formulate recommendations to optimize business processes and electronic route of the documents. The proposed method of analysis of texts can cause internal regulations in line with the actual distribution of functional responsibilities between the staff, and to increase the efficiency of the organization as a whole.

Keywords: electronic document management system, keywords, auditing normative documents, text analysis, semantic network, thesaurus, documents consist, intelligent systems.

Электронный документооборот — это реалия современного делового мира. Эффективность управления электронными документами во многом определяет успешность бизнеса. В свою очередь, для повышения эффективности управления электронными документами в современных условиях динамично изменяющегося рынка необходимо проводить мониторинг бизнес-процессов и своевременно оптимизировать документооборот предприятия.

Процедуры проектирования и оптимизации электронного документооборота организации должны основываться на моделировании маршрутов движения документов. Последнюю задачу предлагается решить с помощью исследования ключевых слов [1]. Оно

заключается в сопоставлении наборов ключевых слов, содержащихся в документах и/или описывающих фактически исполняемые должностные обязанности конкретного сотрудника, представленные в положениях о подразделениях, должностных регламентах или инструкциях сотрудников и др. [2]. В случае отсутствия нормативных документов необходимо детальное обследование организации с целью выявления существующих бизнес-процессов и порождаемых ими документопотоков.

«Рутинные» операции лингвистического анализа текстов, выполняемые вручную, являются источником значительного количества ошибок и неточностей и делают сам процесс анализа документов очень сложным и длительным. Поэтому необходимым условием успешного аудита локальной нормативной базы деятельности организации является наличие программных средств, позволяющих автоматизировать процесс анализа текстов документов [3].

В данной статье предлагается методика анализа текстов документов, являющаяся основой автоматизированной информационной системы аудита нормативных документов организации. Цель анализа — проверка соответствия положений, закрепленных в действующих нормативных документах, реальной ситуации, сложившейся в организации, например в связи с внедрением системы электронного документооборота. Исходными данными служат тексты нормативных документов и/или результаты обследования организации, а также описание предметной области деятельности организации и сферы делопроизводства в виде семантической сети понятий [1; 5].

Для анализа текстов документов с помощью предлагаемой методики необходимы следующие словари.

1. Словарь неинформативных лексических единиц. Сюда входят союзы, предлоги, артикли и т.д. и общеупотребительные термины. Например, предлоги «по», «в», «из» и пр.
2. Словарь существительных, обозначающих виды деятельности. Например, «работа», «подготовка», «управление», «обеспечение», «организация» и пр.
3. Словарь устойчивых словосочетаний. Под этим термином понимается устойчивая конструкция из нескольких последовательно расположенных слов, имеющая определённый смысл только в рамках этой конструкции. Многие названия документов будут включены в этот словарь. Например, «должностная инструкция», «административный регламент», «пояснительная записка» и др. В словаре для каждого термина указывается часть речи основного слова.
4. Словарь замены синонимов на дескрипторы. Здесь дескриптор — это единый термин, объединяющий различные описания одного и того же объекта. Например, синонимы «структурное подразделение организации», «структурное подразделение» или просто

«подразделение» всегда заменяются единым дескриптором «структурное подразделение»; а «документационное обеспечение управления» и «ДОУ» — дескриптором «ДОУ».

Процедура анализа включает два последовательных этапа: исследовательский и аналитический.

Целью исследовательского этапа является выделение в рассматриваемом тексте основных лингвистических конструкций и определение в них действующих субъектов, выполняемых ими действий, объектов, на которые направлены действия, а также свойств субъектов, объектов и действий. Исследовательский этап методики включает следующие шаги.

1. Подготовка текста: из текста необходимо исключить маркировку списков, проверить орфографию и пунктуацию и др.
2. Построение последовательного списка слов текста документа, с сохранением пунктуации предложений.
3. Определение грамматических и морфологических признаков слов (часть речи, число, время, род и др.).
4. Приведение слов к словарной форме.
5. Выявление устойчивых словосочетаний. Для этого предлагается проводить последовательное сравнение множества слов, выделенных в результате анализа текста с записями словаря устойчивых словосочетаний.
6. Выявление имён существительных, обозначающих виды деятельности, путём сравнения полученных результатов с данными словаря видов деятельности.
7. Составление неделимых конструкций из последовательностей слов и словосочетаний с сохранением порядка их следования в тексте. Неделимыми считаются устойчивые обороты речи, которые имеют смысл в контексте данного документа и характеризуют объекты и субъекты предметной области, а также виды деятельности, связывающие их. Цель данного шага - выявление элементов структуры текста в их взаимосвязи с информационным тезаурусом исследуемой области для дальнейшего семантического анализа текста.

Предлагается выделять следующие конструкции текста: S, S+S, A+S; V; Sp, Sp+Sp, A+Sp+Sp; PR+S, PR+Sp; где S — имя существительное, Sp — существительное из словаря видов деятельности, A — прилагательное, V — глагол, PR — предлог.

8. Замена синонимов на дескрипторы.
9. Отнесение элементов преобразованного текста к одной из следующих категорий, представленных в таблице 1.

Таблица 1 — Соответствие категорий ключевых слов конструкциям текста

Категория	Конструкция	Категория	Конструкция
-----------	-------------	-----------	-------------

	текста		текста
Субъект	S; S+S; A+S	Объект	S; S+S; A+S
Роль	V	Свойство	PR+S; PR+Sp
Вид деятельности	Sp; Sp+Sp; A+Sp		

Полученные элементы текста далее назовём ключевыми словами документа.

Рассмотрим пример реализации исследовательского этапа анализа фрагмента текста должностной инструкции главного бухгалтера: «Главный бухгалтер обеспечивает контроль за соблюдением порядка оформления первичных учётных документов» (таблица 2).

Таблица 2 — Реализация исследовательского этапа анализа фрагмента текста должностной инструкции главного бухгалтера

Последовательный список слов, приведённых к нормальной форме	Список слов с учётом устойчивых словосочетаний и синонимов	Морфологические признаки с учётом существительных, содержащих виды деятельности	Категоризация
Главный	Главный бухгалтер	S	Субъект
Бухгалтер			
Обеспечивать	Обеспечивать	V	Роль
Контроль	Контроль	Sp	Вид деятельности
За	За	PR	Свойство вида деятельности
Соблюдение	Соблюдение	Sp	
Порядок	Порядок оформления первичных учётных документов	S	Объект
Оформление		S	
Первичный			
Учётный			
Документ			

Результаты проведённого исследования текста приведены в таблице 3.

Таблица 3 — Результат исследования фрагмента текста должностной инструкции главного бухгалтера

Документ	Субъект	Роль	Вид деятельности	Свойство вида деятельности	Объект
Должностная инструкция	Главный бухгалтер	Обеспечивать	Контроль	За соблюдением	Порядок оформления первичных учётных документов

Полученные результаты являются исходными данными для второго (аналитического) этапа методики анализа текстов документов, целью которой является формулирование рекомендаций по оптимизации бизнес-процессов и электронных документопотоков.

Для анализа применяются операции над множествами (объединение, пересечение, вычитание). В качестве элементов множеств рассматриваются ключевые слова различных

документов и/или ключевые слова, полученные по результатам обследования предметной области и отнесённые к категориям «субъект», «объект», «роль», «вид деятельности».

Аналізу могут подвергаться: документы системы менеджмента качества организации; инструкции (в том числе должностные); положения о структурных подразделениях; документы, образовавшиеся в оперативной деятельности организации.

Сравнение ключевых слов документов позволяет найти:

- общие вопросы, рассматриваемые несколькими субъектами (должностными лицами);
- вопросы, рассматриваемые одним субъектом, но не рассматриваемые другим;
- вопросы, не рассматриваемые ни одним из субъектов.

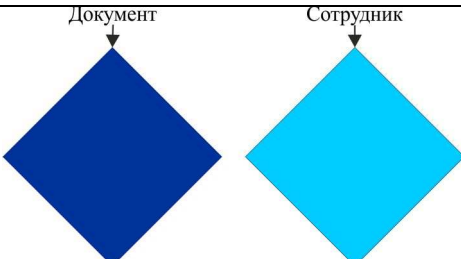
Отдельную проблему составляет интерпретация полученных результатов и выработка рекомендаций по улучшению сложившейся ситуации. Для снижения зависимости от уровня квалификации пользователя предлагается разработка комплекса стандартных автоматизированных процедур анализа полученных результатов, основными из которых являются:

- анализ полноты состава документов, сопровождающих выполнение бизнес-процесса;
- проверка наличия дублирования работ, выполняемых должностными лицами;
- выявление недокументируемых участков деятельности структурных подразделений;
- проверка актуальности содержания нормативных документов, регламентирующих основные бизнес-процессы и делопроизводство.

Перечисленные процедуры предполагают попарное сравнение интересующих пользователя документов и дополняются наборами рекомендаций по устранению выявляемых проблем.

На рисунке 1 представлены разновидности ситуаций, возникающих при сопоставлении ключевых слов распорядительного документа, например приказа, и должностных инструкций сотрудников, участвующих в согласовании документа.

Таблица 4 — Примеры ситуаций, возникающих при сопоставлении ключевых слов распорядительного документа и должностных инструкций сотрудников, участвующих в его согласовании

№ п/п	Описание ситуации	Графическая иллюстрация
1.	$T_d \cap T_c = \emptyset$	

№ п/п	Описание ситуации	Графическая иллюстрация
2.	$(T_{\delta} \cap T_{c1}) \cap (T_{\delta} \cap T_{c2}) = \emptyset$	
3.	$(T_{\delta} \cap T_{c1}) \cap (T_{\delta} \cap T_{c2}) \neq \emptyset$, причём $ (T_{\delta} \cap T_{c1}) \gg (T_{\delta} \cap T_{c2}) $	
4.	$(T_{\delta} \cap T_{c1}) \cap (T_{\delta} \cap T_{c2}) \neq \emptyset$, причём $ (T_{\delta} \cap T_{c1}) \approx (T_{\delta} \cap T_{c2}) $	
5.	$(T_{\delta} \cap T_{c1}) \cap (T_{\delta} \cap T_{c2}) \neq \emptyset$, причём $T_{c2} \in T_{c1}$	

В таблице использованы следующие обозначения:

- T_{δ}, T_c — множества ключевых слов распорядительного документа и должностной инструкции сотрудника соответственно;
- T_{c1}, T_{c2} — множества ключевых слов должностных инструкций первого и второго сотрудников соответственно.

Очевидно, что ситуации 1 и 2 предполагают однозначное решение: в случае 1 сотрудник не должен принимать участие в согласовании документа; в случае 2 каждый из участвующих сотрудников отвечает за свой участок работы, и процесс оформления документа является корректным. В ситуациях 3-5 в принятии решения о маршруте оформления документа должен участвовать эксперт.

Указанная задача имеет смысловой характер, не является структурированной, и поэтому для её решения могут быть использованы модели и методы представления знаний и логического вывода, применяемые при создании интеллектуальных систем [4].

Автоматизация формирования логических выводов позволяет уменьшить требования к квалификации специалистов, выполняющих аудит внутренних нормативных документов или анализ документооборота организации, а также снизить временные затраты на выполнение этой работы. Для ситуации 3 (таблица 4) фрагмент сети продукций приведён в таблице 5.

Таблица 5 — Фрагмент сети продукций для формулирования рекомендаций

Семантические утверждения	Статус утверждения	Графическая интерпретация
1. $ T_{\partial} \cap T_{c1} \gg T_{\partial} \cap T_{c2} $	Факт	
2. Оба сотрудника работают в одном подразделении	Факт	
3. Оба сотрудника занимают должности одного уровня	Факт	
4. Исключить второго сотрудника из списка согласующих	Рекомендация	
5. Внести изменения в должностную инструкцию второго сотрудника	Рекомендация	

Узлами сети продукций являются семантические утверждения, которые представляют собой предложения естественного языка, фиксирующие некоторые факты или действия [4].

Предложенная методика анализа текстов документов позволяет не только привести внутренние нормативные документы в соответствие с реальным распределением функциональных обязанностей между сотрудниками организации, но и оптимизировать документопотоки и бизнес-процессы, что способствует повышению эффективности функционирования организации в целом. Применение методики обеспечивает возможность:

- уменьшения общего времени аудита нормативных документов за счет автоматизации выполнения рутинных операций по сопоставлению текстов документов;
- снижения требований к уровню квалификации должностного лица, выполняющего анализ документооборота организации;
- поддержания актуальности нормативных документов путём их периодического аудита;
- повышения эффективности электронного документооборота и бизнес-процессов организации в целом за счёт оптимизации состава и содержания документов, а также маршрутов их прохождения.

Таким образом, исследование ключевых слов с помощью предложенной методики является эффективным инструментом управления электронными документами организации.

Список литературы

1. Усманова И.В., Баканова Н.Б. Метод ключевых слов как инструмент решения задачи реинжиниринга документопотоков организации // Известия ПГПУ им. В.Г. Белинского. - 2012. - № 30. - С. 302—309.
2. Усманова И.В., Баканова Н.Б. Методика анализа текстов документов организации // Современные технологии документооборота в бизнесе, производстве и управлении : сборник статей XIII Международной научно-практической конференции. – Пенза : Приволжский Дом знаний, 2013. - С. 84—86.
3. Усманова И.В., Баканова Н.Б. Оптимизация маршрутов прохождения документов на основе анализа ключевых слов их текстов // Materiály VIII mezinárodní vědecko - praktická conference «Dny vědy - 2012». Díl 86. Moderní informační technologie: Praha. Publishing House «Education and Science» s.r.o. 80 stran. Str. 47—52.
4. Усманова И.В., Коровина Л.В., Соколова О.Г. Настройка экспертной системы анализа документооборота на особенности организации // Современные проблемы науки и образования. – 2013. – № 6. - URL: www.science-education.ru/113-10995 (дата обращения: 19.02.2014).
5. Фионова Л.Р. Использование семантической модели для построения сети понятий в автоматизированной системе тестирования знаний при аттестации армейских кадров // Известия высших учебных заведений. Поволжский регион. Технические науки. Специальный выпуск. – 2008. - С. 243-253.

Рецензенты:

Фионова Л.Р., д.т.н., профессор, декан факультета вычислительной техники, заведующая кафедрой «Информационное обеспечение управления и производства» ФГБОУ ВПО «Пензенский государственный университет», г. Пенза.

Косников Ю.Н., д.т.н., профессор, заведующий кафедрой «Информационно-вычислительные системы» ФГБОУ ВПО «Пензенский государственный университет», г. Пенза.