

## МОДЕЛЬ АВТОМАТИЗАЦИИ АНАЛИЗА РАЗЛИЧНЫХ ГРУПП ДОКУМЕНТОВ ДЛЯ СИНТЕЗА СТРУКТУРЫ ДОКУМЕНТА В ЗАВИСИМОСТИ ОТ ВИДА ДЕЯТЕЛЬНОСТИ

И. А. СЕРГЕЕВ<sup>1</sup>, А. П. БЕЛТЮКОВ<sup>2</sup>

<sup>1</sup>sia104809@gmail.com, <sup>2</sup>belt.udsu@mail.ru

ФГБОУ ВО «Удмуртский государственный университет» (УдГУ)

*Поступила в редакцию 2 декабря 2021 г.*

**Аннотация.** Цель исследования – создание модели автоматизации анализа различных классов документов для синтеза модели отличий документов разных классов, обладающих различными атрибутами и признаками. Рассматриваются такие особенности документов, как форма представления и содержание различных классов документов в зависимости от сферы деятельности, в которой они применяются. Определение сферы применения документации различных видов деятельности и примеры. Основное внимание уделяется выявлению общих и специфических элементов информации различных классов документов, характерных для самых разных сфер деятельности. Устанавливаются отличия и выявляются отличительные классификационные признаки, характерные для документов различных сфер деятельности. В результате определены их отличительные особенности в зависимости от сфер деятельности и вида документа. Результаты работы планируется применить в автоматизации процесса анализа, трансформации и синтеза структуры документов.

**Ключевые слова:** документы; классификация; анализ; сферы деятельности; структура данных; признаки; обработка; модель; автоматизация; система.

Документ в трактовке Федерального закона «Об информации, информатизации и защите информации» от 27 июля 2006 г. и ГОСТа Р 7.0.8-2013 (см. «Национальный стандарт Российской Федерации. Система стандартов по информации, библиотечному и издательскому делу. Делопроизводство и архивное дело. Термины и определения» (утв. Приказом Росстандарта от 17.10.2013 N 1185-ст) – это зафиксированная на материальном носителе информация с реквизитами, позволяющая ее идентифицировать.

Законом «Об информации, информационных технологиях и о защите информации» определено понятие «документированная информация» – это зафиксированная

на материальном носителе путем документирования информация с реквизитами, позволяющими определить такую информацию или в установленных законодательством Российской Федерации случаях ее материальный носитель» (см. Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 N 149-ФЗ).

Документы являются крайне важным продуктом человеческой деятельности. Они фиксируют факты, события, результаты и многое другое. Документы редко являются целью человеческой деятельности, основная их задача – это отражение деятельности человека или общества. Документы исполь-

зуются во многих сферах человеческой жизни и являются важной частью информационных структур [1]. Эффективное использование цифровых технологий в современном мире является неоспоримым фактором роста конкурентоспособности компании. Практически все современные предприятия подвергаются всеобщей цифровизации. В настоящее время не существует предприятия, которое может осуществлять свою деятельность без использования широких возможностей Интернета и современных информационных технологий. Они значительно упрощают жизнь любого предприятия, и нередко бывает, что эффективность работы предприятия напрямую связана с уровнем технического оснащения, и в некоторых случаях невозможно вести бизнес, если техническое оснащение предприятия недостаточно [2].

Основная роль цифровых информационных технологий предприятия и их технического оснащения заключается в генерации, анализе и изменении документации и различной информации, непосредственно связанной с рабочим процессом [8].

Основную сложность для анализа документов представляет неоднородность документации. Каждый документ, независимо от его типа, представляет собой структуру данных. В большинстве случаев такие документы оформляются в довольно неформальном стиле, что в значительной степени усложняет их анализ. Сложность восприятия документов человеком (имеется в виду время и умственные усилия) зависит от сферы деятельности, в которой они используются. Кроме субъективной сложности, для документа есть и объективные – описательные сложности документов как структур данных.

В документах одной сферы деятельности можно проследить общие признаки и свой-

ства. Если же сравнивать аналогичные документы, но из разных сфер деятельности, то можно отметить некоторые отличия. Определение данных отличий позволит эффективнее анализировать и обрабатывать документы, относящиеся к разным типам и сферам деятельности.

Основными задачами данного исследования являются поиск и определение свойств и признаков документов различных сфер деятельности. Рассматриваются структуры документов и классификация признаков, которые являются отличительными. Также рассматриваются особенности различных групп документов как между собой, так и между документами из разных сфер деятельности [3].

В области делопроизводства каждый документ обладает определенным набором атрибутов, что позволяет ему подстраиваться в качестве элемента структуры документооборота организации. Правильный документ является необходимым условием для принятия эффективных управленческих решений на его основе. Деловая информация, в частности, документы, являются предметом труда и одновременно его результатом. В процессе документооборота участвует множество различных документов разных видов и форматов. Для эффективного оформления, хранения и обработки все документы делят по классам. Разделяют на виды и типы в зависимости от их признаков, функций и свойств. Первостепенная задача классификации – повышение эффективности работы с документами в масштабах организации, что благоприятно влияет на скорость их исполнения и контроль за производственными процессами организации [4].

Все документы, обращающиеся в организации, классифицируются по признакам, указанным в табл. 1.

Таблица 1

## Классификационные признаки [5]

## Classification signs [5]

Признак классификации	Определение признаков
Назначение	Входящие
	Исходящие
	Внутренние
Наименование	Акт, Приказ, Справка, Распоряжение и т.д.
Способ фиксации	В зависимости от типа материального носителя
Степени гласности	Открытые
	– Документы с ограниченным доступом;
	– Документы для служебного пользования;
	– Для личного использования;
	– Коммерческие;
	– Секретные;
Степени сложности (по различным мерам и границам сложности)	– Совершенно секретные;
	– Конфиденциальные
Происхождение	Простые
	Сложные
	Личные:
	– Именные;
	– Касающиеся конкретного лица
	Служебные:
– Государственные;	
– Общественные;	
– Частные	
Юридическая сила	Подлинник
	Дубликат
	Копия
	Заверенная копия
	Выписка
	Отпуск (Копия отправителя)
Сроки исполнения	Срочный
	Среднесрочный
	Не срочный
Степень унификации	Индивидуальные
	Трафаретные
	Типовые
Сроки хранения	Временный ≤ 10 лет
	Долговременный ≥ 10 лет
	Постоянного хранения
	До ликвидации организации
	До минования надобности
	До замены новыми
	До отмены
Степень обязательности	Обязательные
	Рекомендательные
	Информационные
Актуальность	Актуальный
	Неактуальный
	Без срока давности

Каждый классификационный признак также имеет свою определенную систему градации. В табл. 1 указаны классификационные признаки, кроме признака, связанно-

го со сферой применения, так как данная классификация по признаку применяется чаще, чем остальные. Для более полного представления она была вынесена в табл. 2.

## Классификация по сфере деятельности [5]

## Classification by field of activity [5]

Виды деятельности	Сфера применения	Определение	Примеры
Организационно-распорядительные	Нормативные	Документы, регламентирующие правила, касающиеся различных видов деятельности	Стандарты
	Организационные	Документы, утверждающие права, функции и задачи работников и руководителей	Приказы, постановления, распоряжения
	Организационно-правовые	Документы, обеспечивающие правовую основу деятельности организации	Договоры, штатное расписание, уставы
Финансово-бухгалтерские	Кадровые	Документы, регламентирующие отношения между работником и работодателем	Трудовые договоры и соглашения
	Коммерческие	Документы, применяемые для реализации торговой деятельности и совершения сделок	Контракты, договоры
	Бухгалтерские	Документы, отражающие совершаемые хозяйственные операции	Накладная, счет, авансовые отчеты
	Финансовые	Документы, отражающие финансовое состояние организации	Счет-фактура, счет-проформа, инвойс
	Торговые	Документы, регламентирующие процесс торгового обмена	Документы сферы платежей, счета и коммерческая корреспонденция
	Ценообразование	Документы, регламентирующие ценообразование и сметное нормирование	Договоры подряда
Производственно-технологические	По изобретениям и открытиям	Документы, регламентирующие проверку и регистрацию изобретений	Охранные и регистрационные документы
	Конструкторские	Документы, содержащие информацию об устройстве и определяющие его состав и производство	Сборочный чертеж, модель детали, спецификация, ведомость
	Технологические	Документы, регламентирующие процесс производства	Карты техпроцессов, маршрутная карта, тех. инструкции
	Проектно-сметные	Документы, содержащие основу проекта и обосновывающие те или иные требования и особенности	Архитектурные решения, декларации безопасности, спецификации
Информационные	Плановые	Документ, регламентирующий процесс установления целей и задач организации	Программы, планы, графики
	Информационно-справочные	Справочные документы, содержащие дополнительные сведения для принятия тех или иных решений	Докладная, справка, акт, письмо, предложение
Научно-исследовательские	Научно-технические	Документы, в которых отражена научно-исследовательская и техническая деятельность предприятия	Отчеты по темам НИР, патентная документация
	Учебно-методические	Документ, регламентирующий процесс проведения учебного занятия	Учебный план, расписание занятий, учебно-методическая документация

Цифровые документы, несмотря на некоторую однородность обработки документации, очень отличаются, если рассматривать их с позиции видов деятельности, в которых они используются. Организационные, финансово-бухгалтерские и производственно-технологические – все эти виды деятельности в основе имеют различные сферы применения. Каждой сфере характерны определенные форматы данных, такие как: DOCX, DOC, PDF, TIFF, XLSX, XLS, которые наиболее часто используются.

Каждый формат данных представляется в определенной форме представления. В зависимости от формы – текст, таблица или изображения – выбираются соответствующие методы для обработки. Разумеется, каждый из представленных форматов может содержать в основе своей две другие формы представления (текст, таблица или изображения).

В зависимости от вида деятельности каждый документ имеет определенные классификационные признаки, указанные в табл. 1, которые позволяют определить атрибуты документа и его место в системе документации организации. Помимо официальных классификационных признаков, в работе используется множество других признаков, атрибуты которых задают сами пользователи. Таким образом, количество признаков не ограничено.

Классификация документов по сфере применения является наиболее распространенной среди всех классификаций, но для того, чтобы отличить один документ от другого, необходимо изучить хотя бы первый абзац. Следует учесть, что любые документы можно классифицировать сразу по нескольким признакам. По наблюдению авторов, чаще всего классификация групп документов идет в определенном порядке – вначале документы классифицируют по признаку сферы назначения, затем по степени сложности и так далее по некоторому определенному порядку. Данный процесс вполне поддается автоматизации при грамотном анализе документа. Классификация по сложности зависит от принятых мер сложности и границ для значений этих мер [6].

В соответствии с международным стандартом ГОСТ 2.105-95, в документе могут содержаться формулы, таблицы, рисунки и текст. Данное утверждение регламентирует только содержимое официальных документов. Документы, оформленные неформально и внутри компании, могут содержать менее формализованное оформление и содержимое. На текущий момент бумажные документы имеют несколько больше прав, чем электронные, в результате чего весомая часть правил оформления документа ориентируется на информацию, которая способна быть отображена на бумаге, а именно: изображения, формулы, графики, таблицы и текст (см. ГОСТ 2.105-95. «Межгосударственный стандарт. Единая система конструкторской документации. Общие требования к текстовым документам» (введен в действие Постановлением Госстандарта РФ от 08.08.1995 N 426) (ред. от 22.06.2006)).

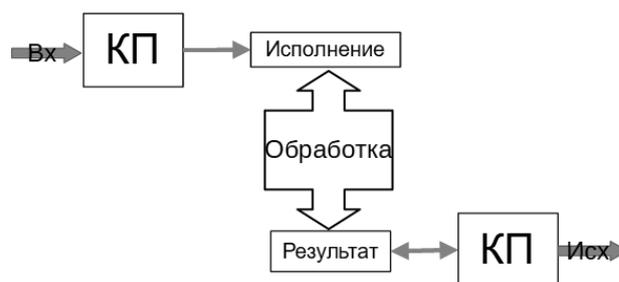


Рис. 1. Модель автоматизации

В зависимости от сферы деятельности выбираются формы представления информации – синтаксиса, на основе чего подбираются соответствующие классификационные признаки – содержание.

На рис. 1 изображена схема модели автоматизации – в основе поступающей информации лежит соответствующая форма представления, которая подается на вход. Там в зависимости от задачи, на основе классификационных признаков информация под воздействием методов распознавания и порождения распознается или же дополняется в зависимости от условия для исполнения. В дальнейшем она подается на обработку, где в зависимости от результата фильтрации классификационными признаками, она подается на выход или повторно возвращается на обработку, если условия классификационных признаков не выполнены.

Если условие выполняется лишь частично в соответствии с классификационными признаками информации, блоков для обработки может быть несколько, при этом в процессе информация будет дополняться, в результате блоков классификационных признаков может быть много.

Теоретически блока обработки может и не быть – достаточно лишь множества фильтров классификационных признаков.

Правила, регламентирующие оформление и содержимое документов, меняются менее быстро, так как данный аспект затрагивает достаточно большое количество стандартов. Электронные документы становятся более значимыми, благодаря удобству обработки передачи информации, из чего делается вывод, что в ближайшее десятилетие стандарты по оформлению документов будут сосредоточены на регламентировании правил оформления электронной документации, а не бумажной [7].

В результате полученной из настоящего исследования информации делаются выводы:

- множества документов разных сфер деятельности могут пересекаться между собой;

- на основе полученной информации демонстрируется как общность документов по всем признакам, так и индивидуальные особенности, на основании которых формируется семантическая модель документа, которая может быть использована для автоматического анализа документации.

#### СПИСОК ЛИТЕРАТУРЫ

1. **Корнеев И. К.** Делопроизводство. Образцы, документы. Организация и технология работы. М.: Проспект, 2021. 480 с. [ I. K. Korneev, *Office work. Samples, documents. Organization and technology of work*, (in Russian). Moscow: Prospect, 2021. ]

2. **Электронные** документы в делопроизводстве // Журнал «Справочник секретаря и офис-менеджера». [Электронный ресурс]. URL: <https://www.sekretariat.ru/article/210677-qqq-16-m10-elekt-ronnye-dokumenty-v-delo-proizvodstve> (дата обращения 2.11.2021). [ Electronic documents in business // “Secretary and Office Manager’s Handbook” magazine (2021, Nov. 2), [Online]. Available: <https://www.sekretariat.ru/article/210677-qqq-16-m10-elektronnye-dokumenty-v-delo-proizvodstve> ]

3. **Михеева Е. В.** Компьютерное делопроизводство в бухгалтерии: учебник. М.: Академия, 2019. 96 с. [ E. V. Mikhееva, *Computer business-production in accounting: textbook*, (in Russian). Moscow: Academia, 2019. ]

4. **Пожникова Н. М.** Практикум по предмету «Документы, корреспонденция и делопроизводство»: учебное пособие. М.: Академия, 2019. 208 с. [ N. M. Pozhnikova, *Workshop on the subject "Documents, Correspondence and Office Work": textbook*, (in Russian). Moscow: Academia, 2019. ]

5. **Виды** документов в делопроизводстве // Журнал «Справочник секретаря и офис-менеджера». [Электронный ресурс]. URL: <https://www.sekretariat.ru/article/210914-qqq-17-m5-vidy-dokumentov-v-delo-proizvodstve> (дата обращения 2.11.2021). [ Types of documents in office work // “Secretary and Office Manager’s Handbook” magazine (2021, Nov. 2), [Online]. Available: <https://www.sekretariat.ru/article/210914-qqq-17-m5-vidy-dokumentov-v-delo-proizvodstve> ]

6. **Чвилов В. В.** Судебное делопроизводство в цифровую эпоху. М.: ИД Городец, 2020. 352 с. [ V. V. Chvirov, *Litigation in the digital era*, (in Russian). Moscow: ID Gorodets, 2020. ]

7. **Колышкина Т. Б.** Деловые коммуникации, документооборот и делопроизводство. М.: Юрайт, 2020. 164 с. [ T. B. Kolyshkina, *Business communications, document flow and office production*, (in Russian). Moscow: Yurayt, 2020. ]

8. **Сергеев И. А.** Адаптивные информационные технологии автоматической классификации документов по их важности и критичности // Информационные технологии интеллектуальной поддержки принятия решений (ITIDS'2019): труды VII Всероссийской научной конференции (с приглашением зарубежных ученых). Уфа: УГАТУ, 2019. С. 27–32. [ I. A. Sergeev, “Adaptive information technologies for automatic classification of documents according to their importance and criticality”, (in Russian), in *Proc. 7<sup>th</sup> All-Russian Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS'2019)*, Ufa, 2019, pp. 27-32. ]

#### ОБ АВТОРАХ

**СЕРГЕЕВ Игорь Альбертович**, асп. каф. теоретических основ информатики. Дипл. инженер (ИжГТУ, 2015). Готовит дис. о автоматической подготовке оформления статей.

**БЕЛЬТЮКОВ Анатолий Петрович**, проф., зав. каф. теоретических основ информатики. Дипл. математик (Ленингр. гос. ун-т, 1975). Д-р физ.-мат. наук по тер. основам информатики (СПбГУ, 1993). Иссл. в обл. матем. логики и алгоритмической теории сложности.

#### METADATA

**Title:** A model for automating the analysis of various groups of documents for synthesizing the structure of a document depending on the type of activity.

**Authors:** I. A. Sergeev<sup>1</sup>, A. P. Belyukov<sup>2</sup>

**Affiliation:** Udmurt State University (UdsU), Russia.

**Email:** <sup>1</sup>sia104809@gmail.com, <sup>2</sup>belt.udsu@mail.ru.

**Language:** Russian.

**Source:** SIIT (scientific journal of Ufa State Aviation Technical University), vol. 4, no. 1 (8), pp. 5-11, 2022. ISSN 2686-7044 (Online), ISSN 2658-5014 (Print).

**Abstract:** The purpose of the study is to create a model for automating the analysis of various classes of documents for synthesizing a model of differences between documents of different classes that have different attributes and features from each other. The article examines the features of documents such as the form of presentation and the content of various classes of documents,

depending on the field of activity in which these documents are applied. De-termination of the scope of documentation of various activities and examples. The main attention is paid to identifying general and specific elements of information of various classes of documents, characteristic of various fields of activity. Differences are established and you are the distinctive classification features characteristic of documents in various fields of activity. As a result, the distinctive features of various classes of documentation were determined, depending on the field of activity and the type of document. The results of the work are planned to be applied in the automation of the process of analysis, transformation and synthesis of the structure of documents

**Key words:** documents; classification; analysis; areas of activity; data structure; signs; treatment; model; automation; system.

**About authors:**

**SERGEEV, Igor Albertovich**, Postgrad. Student, Dept. of Theoretical Foundations of Informatics. Dipl. Engineer (Izhevsk State Technical University named after M. T. Kalashnikov, 2015). Prepares dis. on the automatic preparation of article formatting.

**BELTYUKOV, Anatoly Petrovich**, Prof., chair Dept. of Theoretical Foundations of Informatics. Dipl. Mathematics (Leningr State Univ., 1975). Dr. of Phys.-Math. Sci. (SPbSU, 1993).