

УДК 004.451.53

ОНТОЛОГИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ВНУТРЕННЕГО ДОКУМЕНТА ПО ОСНОВНОЙ ДЕЯТЕЛЬНОСТИ ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ

В. В. Чуйкова¹ [0009-0000-6784-8135], М. О. Таныгин² [0000-0002-4099-1414]

^{1, 2}Юго-западный государственный университет, Курская обл., Курск, 305040,
Россия

¹chuikova_v_v@bk.ru, ²tanygin@yandex.ru

Аннотация

Проанализирована структура служебных записок кафедры высшего учебного заведения с целью упрощения ретроспективного поиска среди массива документации и выделения групп однотипных документов. Установленные правила классификации служебных записок на основе формальных признаков позволят значительно упростить процесс поиска и работы с ними, а также создания новых типовых документов.

Ключевые слова: онтология документа, структура документа, онтологическая модель, анализ структуры документов.

ВВЕДЕНИЕ

Нормативные документы имеют определённую логическую структуру. Например, как правило, приказы делятся на логические разделы, научные статьи состоят из аннотации, введения, обзора существующих работ и других секций, а служебные записки обычно содержат краткую информацию о каком-либо процессе или событии. Информация о логической структуре полезна для автоматического анализа документа.

В условиях увеличения объёма внутренней служебной документации необходимо быстро находить подобные документы для создания на их основе новых, а также легко ориентироваться в организационно-распорядительной документации (ОРД).

ПРОБЛЕМАТИКА

В [1] авторы предложили извлекать логическую структуру из сканированных документов. Этот процесс имеет ряд трудностей: сканированный документ может не содержать текстового слоя, часть документа может быть частично отформатирована и, самое главное, документы имеют разную структуру. В [2] представлено решение задачи оперативного формирования документов, основанное на формальном моделировании информационной структуры и процессов генерации документов. В [3] авторы попытались выделить физическую и логическую структуры PDF-документа для его дальнейшего распознавания и обработки. В [4] рассмотрены структура данных и модель документа в Microsoft Word и обоснована концептуальную модель программного обеспечения для нормоконтроля печатных работ. В [5] предложено выделять из сценариев видеоигр необходимые сущности и передавать их дальнейшим шагам алгоритма, который по текстовым описаниям будет генерировать игровые ресурсы. Такая декомпозиция позволяет оптимизировать процесс создания сценария игры.

Актуальность настоящего исследования заключается в необходимости проведения ретроспективного поиска среди организационно-распорядительной документации. Поисковой запрос обычно формируется простыми разговорными фразами и на основе субъективных понятий, тогда как в документе текст носит официально-деловой характер, с соответствующей терминологией. Соответственно, необходимы правила классификации ОРД и описания их смысла на основе формальных признаков.

СТРУКТУРА СЛУЖЕБНЫХ ЗАПИСОК В ОРГАНИЗАЦИИ: ЧЕМ ОНИ ОТЛИЧАЮТСЯ

Традиционно служебная записка представляет собой документ, созданный в каком-либо текстовом редакторе. Отличает его от обычных документов то, что в его состав входят только структурированный текст и таблицы, а графика, математические выражения, диаграммы в документах такого рода не допускаются [6, ГОСТ Р. 7.0.97-2016]. В настоящей статье мы рассматриваем сферу образования, а конкретно, документальное обеспечение подготовки специалистов в высших учебных заведениях.

Служебные записки (СЗ) имеют следующую общую структуру:

1. Автор (от кого пишется СЗ).
2. Адресат (кому направляется документ).
3. Заголовок «Служебная записка».
4. Основной текст.
5. Подпись автора и/или визы задействованных сотрудников.
6. Резолюция адресата (опционально), накладываемая после анализа содержания СЗ.

На рис. 1 представлена типовая СЗ кафедры в рамках учебного процесса.

Кафедра	Ректору ФИО
---------	----------------

СЛУЖЕБНАЯ ЗАПИСКА
07.02.2025 г. № 07.07.23 / 53

Об отчислении аспиранта

Прошу отчислить ФИО из состава аспирантов университета в связи с невыполнением обучающимся по образовательной программе высшего образования – программе подготовки научно-педагогических кадров в аспирантуре (ОП ВО) обязанностей по добросовестному освоению ОП ВО и выполнению учебного плана с 14.02.2025 года.

Зав. кафедрой	ФИО

Рис. 1. Типовая СЗ кафедры в рамках учебного процесса

Пункты 1–6 являются, по сути, атрибутами. Если проводить систематизацию документов только на их основании, то с практической точки зрения в этом не так много смысла. Действительно, так ли важно, что в конкретную дату на имя определённого сотрудника из разных структурных подразделений организации поступило определённое количество СЗ. Обычно взаимосвязи «Автор – Адресат» определяются организационно-штатной структурой конкретного учебного заведения,

а не характером сведений, содержащихся в служебной записке, тогда как основной текст содержит информацию, которая будет содержаться в поисковых запросах, о затрагиваемых событиях, процессах, действиях, которые и определяют назначение каждого конкретного документа. В то же время даже такой достаточно сложно формализуемый элемент документа может содержать типовые лингвистические конструкции, которые как определяют отношение «Автор – Адресат», так и позволяют провести первичную классификацию документа.

Мы предположили, что для формализации типа документа надо рассмотреть виды связок «глагол (действие) – дополнение (тип действия)».

Было проанализировано более 700 служебных записок организации за несколько лет. Выявлены следующие связки: «прошу – зачислить», «прошу – отчислить», «прошу – рассмотреть», «прошу – заверить», «прошу – назначить», «прошу – утвердить». Их можно обобщить в связку вида «Прошу – дополнение».

Кроме того, установлены более редкие связки: «предлагаю – следующее», «предлагаю – состав», «прошу – разрешения», «предлагается», «довожу – до вашего сведения», «прошу – установить (надбавку)», «прошу – возместить».

По результатам проведённого анализа выделены обобщённые связки трёх типов, которые наиболее характерны для служебных записок:

1. «Прошу – дополнение»;
2. «Предлагаю – дополнение»;
3. «Довожу до сведения (информирование)».

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ

Для построения онтологической модели [7] проанализируем эти сочетания и выделим основные сущности служебной записки.

Если рассматривать более глубоко содержимое основной части СЗ, то для него, с учётом описанных выше атрибутов (таких как автор, адресат, заголовок «служебная записка», основной текст, подпись автора и/или визы задействованных сотрудников, резолюция адресата) можно выделить следующие сущности.

1. Объект воздействия (на что хотим повлиять) представлен на рис. 2.



Рис. 2. Объект воздействия

2. Действия, которые инициирует служебная записка, показаны на рис. 3.

Действия	Адресат
Прошу	Требуется адресат просьбы (ранг выше/один уровень)
Предлагаю	Требуется адресат для предлагаемой информации (ранг выше/один уровень)
Довожу до сведения	Требуется адресат (ранг любой)

Рис. 3. Действия

3. «Третьи лица» (кому переадресовано)

- ✓ структурные подразделения;
- ✓ работники.

4. Результат рассмотрения адресатом рис. 4

Результат рассмотрения адресатом

- ⇒ Инициация действия
- ⇒ Отказ
- ⇒ Переадресация другому лицу
- ⇒ Инициация анализа ситуации

Рис. 4. Результат рассмотрения адресатом

Определим объекты, классы и отношения между ними.

1. Объекты представлены на рис. 5.

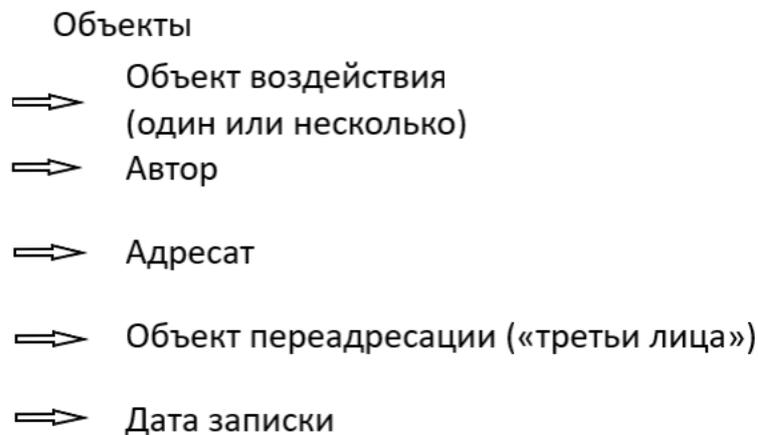


Рис. 5. Объекты

2. Классы показаны на рис. 6.

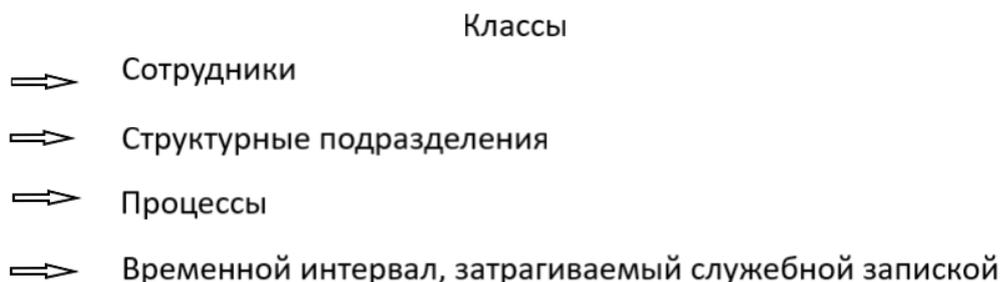


Рис. 6. Классы

Все объекты входят в те или иные классы. Так, «Объект воздействия» (один или несколько) может входить во все три класса: он может быть и сотрудником, и структурным подразделением, и процессом. Объект «Автор» может быть только сотрудником. Объекты «Автор» и «Адресат» также могут быть только сотрудниками. «Объект переадресации» («третьи лица») может быть сотрудником или структурным подразделением. Объект «Дата записки», относящийся к классу «Временной интервал, затрагиваемый служебной запиской», косвенно относится к процессам, характеризуя период времени влияния на «Объект воздействия» и в целом актуальность служебной записки.

3. Отношения между объектами:

- ✓ «Автор – Адресат» (рис. 7).



Рис. 7. Отношение «Автор – Адресат»

- ✓ «Адресат – Объект переадресации»
- ✓ «Адресат – Время»
- ✓ «Автор – Объект переадресации» представлено на рис. 8.



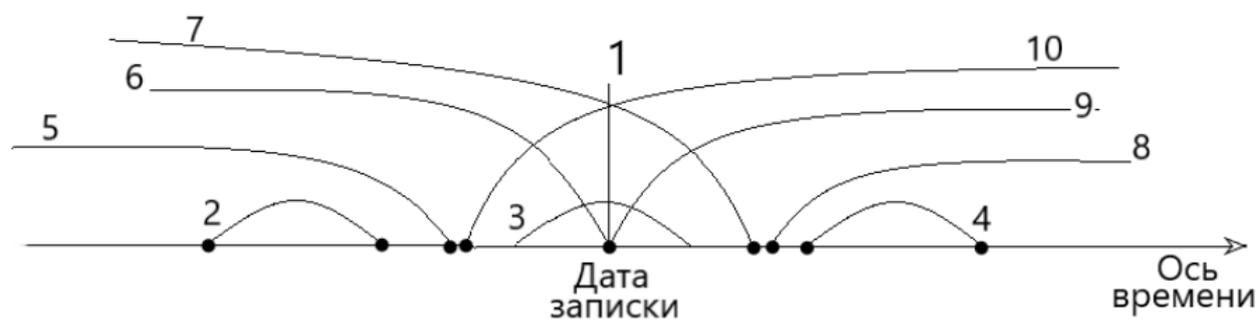
Рис. 8. Отношение «Автор – Объект переадресации»

Для объекта «Дата записки» можно выделить три свойства, определяющие её, этот объект характеризуется следующим.

1. Время «актуальности» записки

- ✓ Текущее;
- ✓ Интервал в прошлом;
- ✓ Интервал в будущем.

2. Для отношения «Дата записки – Объект воздействия» можно выделить 10 характеристик, зависящих от даты инициализации служебной записки и её расположения на оси времени относительно времени, в течение которого документ актуален (рис. 9).



- ✓ 1 - открытый интервал в текущий момент времени
- ✓ 2, 3, 4 - три закрытых интервала
- ✓ 5, 6, 7 - три открытых интервала в прошлом
- ✓ 8, 9, 10 - три открытых интервала в будущем

Рис. 9. Характеристики для отношения «Дата записки – Объект воздействия»

Такая подробная декомпозиция документа, выделение объектов и определение отношений между объектами нужны, чтобы получить возможность легко идентифицировать служебные записки и строить логические выводы о том, какая именно служебная записка перед нами, а также быстро провести поиск и найти нужную. Кроме того, был проведён поиск среди имеющегося массива служебных записок и найдены значения количественных показателей СЗ обозначенных видов (табл. 1).

Таблица 1. Количество СЗ разных видов

Смысловые связи в СЗ	Количество
1. «Прошу – зачислить»	4%
2. «Прошу – отчислить»	4%
3. «Прошу – рассмотреть»	5%
4. «Прошу – заверить»	6%
5. «Прошу – назначить»	10%
6. «Прошу – утвердить»	7%
7. «Прошу – назначить»	4%
8. «Предлагаю – следующее»	7%

9. «Предлагаю – состав»	6%
10. «Прошу – разрешения»	5%
11. «Предлагается»	7%
12. «Довожу – до вашего сведения»	21%
13. «Прошу – установить (надбавку)»	9%
14. «Прошу – возместить»	5%

Можно сделать вывод, о том что наибольшее количество документов типа «служебная записка» встречается в исследуемой сфере деятельности (образование) со связками «Прошу – дополнение» (табл. 2).

Таблица 2 – Количество СЗ обобщённых типов

Обобщённые связки в СЗ	Количество
1. «Прошу – дополнение»	59%
2. «Довожу – до сведения» (информирование)	21%
3. «Предлагаю – дополнение»	20%

ЗАКЛЮЧЕНИЕ

Используя представленную онтологическую модель документа, основанную на формальной структуре служебных записок, можно легко организовывать хранение, автоматическую обработку, ретроспективный поиск в архивах и найти документы, имеющие одинаковые связки типа «Прошу – дополнение», «Предлагаю – дополнение», «Довожу до сведения (информирование)» в зависимости от того, что именно требуется найти. Очевидно, что СЗ с какими-либо предложениями мы не будем искать среди СЗ со связками «Прошу – дополнение» и «Довожу до сведения (информирование)». Кроме того, традиционная структура привязки служебной записки к дате её создания не всегда позволяет корректно установить тот временной интервал, который в ней рассмотрен или затрагивается. Предложенное нами описание отношения «Объект воздействия – Дата записки» позволяет добавлять объектам, выделенным в составе служебной записки, свойство «темпоральности», то есть изменчивости свойств объекта во времени. С учётом

того, что под объектами мы понимаем не только работников, но и целые структурные подразделения и процессы, это даст возможность использовать предложенную модель поиска не только в электронных архивах, но и современных CRM-системах.

СПИСОК ЛИТЕРАТУРЫ

1. Богатенкова А.О., Козлов И.С., Беляева О.В., Перминов А.И. Извлечение логической структуры из сканированных документов // Труды ИСП РАН. 2020. №4. URL: <https://cyberleninka.ru/article/n/izvlechenie-logicheskoy-struktury-iz-skanirovannyh-dokumentov> (дата обращения: 24.02.2025).
 2. Пенькова Т.Г. Модели и методы оперативного формирования документов // Вычислительные технологии. 2009. Т. 14. № 2. С. 98–109.
 3. Беляева О.В., Перминов А.И., Козлов И.С. Использование синтетических данных для тонкой настройки моделей сегментации документов // Труды ИСП РАН. 2020. Т. 32, вып. 4. С. 189–202.
[https://doi.org/10.15514/ISPRAS-2020-32\(4\)-14](https://doi.org/10.15514/ISPRAS-2020-32(4)-14) (дата обращения: 24.02.2025).
 4. Виноградская М.Ю., Крючок А.Ю. Описание структуры документов Microsoft Word для разработки программ нормоконтроля // Вестник Калужского университета. 2019. №3 (44). С. 113–116.
 5. Нурлыгаянов Н.Р., Кугуракова В.В. Подход к созданию корпуса текстов видеоигр на основе универсальной структуры // Электронные библиотеки. 2024. Т. 27. №4. С. 578–597.
 6. ГОСТ Р. 7.0.97-2016. Национальный стандарт Российской Федерации. Система стандартов по информации, библиотечному и издательскому делу. Организационно-распорядительная документация. Требования к оформлению документов. URL: <https://docs.cntd.ru/document/1200142871>.
URL: <http://publication.pravo.gov.ru/Document/View/0001202002070036>.
 7. Горшков С. Введение в онтологическое моделирование. Ревизия 2.4. ООО «ТриниДата», 2014–2018, 150 с.
-

INTERNAL DOCUMENT ONTOLOGICAL REPRESENTATION IN THE MAIN ACTIVITY OF THE EDUCATIONAL ORGANIZATION

V. V. Chuikova¹ [0009-0000-6784-8135], M. O. Tanygin² [0000-0002-4099-1414]

^{1,2}South-West State University

¹chuikova_v_v@bk.ru, ²tanygin@yandex.ru

Abstract

The paper analyzes the structure of the department's memos to simplify a retrospective search among an array of documentation and identify groups of similar documents, which will greatly simplify the process of searching and working with them, as well as the creation of new standard documents.

Keywords: *document ontology, document structure, ontological models, document structure analysis.*

REFERENCES

1. Bogatenkova A.O., Kozlov I.S., Belyaeva O.V., Perminov A.I. Izvlechenie logicheskoy struktury iz skanirovannykh dokumentov // Trudy ISP RAN. 2020. №4. URL: <https://cyberleninka.ru/article/n/izvlechenie-logicheskoy-struktury-iz-skanirovannykh-dokumentov> (data obrashcheniya: 24.02.2025).
2. Pen'kova T.G. Modeli i metody operativnogo formirovaniya dokumentov // Vychislitel'nye tekhnologii. 2009. T. 14. № 2. S. 98–109.
3. Belyaeva O.V., Perminov A.I., Kozlov I.S. Ispol'zovanie sinteticheskikh dannykh dlya tonkoj nastrojki modelej segmentacii dokumentov // Trudy ISP RAN. 2020. T. 32, vyp. 4. S. 189–202. [https://doi.org/10.15514/ISPRAS-2020-32\(4\)-14](https://doi.org/10.15514/ISPRAS-2020-32(4)-14) (data obrashcheniya: 26.02.2025).
4. Vinogradskaya M.Yu., Kryuchok A.Yu. Opisanie struktury dokumentov Microsoft Word dlya razrabotki programm normokontrolya // Vestnik Kaluzhskogo universiteta. 2019. №3 (44). S. 113–116.
5. Nurlygayanov N.R., Kugurakova V.V. Podhod k sozdaniyu korpusa tekstov videoigr na osnove universal'noj struktury // Elektronnye biblioteki. 2024. T. 27. №4. S. 578–597.

6. GOST R. 7.0.97-2016. Nacional'nyj standart Rossijskoj Federacii. Sistema standartov po informacii, bibliotechnomu i izdatel'skomu delu. Organizacionno-rasporjaditel'naya dokumentaciya. Trebovaniya k oformleniyu dokumentov.

URL: <https://docs.cntd.ru/document/1200142871>.

URL: <http://publication.pravo.gov.ru/Document/View/0001202002070036>.

7. *Gorshkov S. Vvedenie v ontologicheskoe modelirovanie. Reviziya 2.4.* ООО «TriniData», 2014-2018, 150 s.

СВЕДЕНИЯ ОБ АВТОРАХ



ЧУЙКОВА Виктория Владимировна – старший преподаватель и аспирант кафедры «Информационная безопасность», ЮЗГУ. Область научных интересов: онтологические модели, семантический поиск, автоматизированные системы документооборота. Число научных публикаций – 15.

Viktoriia Vladimirovna CHUIKOVA – Senior Lecturer and aspirant of the Information security department, South-West State University. Research interests: ontological models, semantic search, automated document management systems. The number of publications – 15.

email: chuikova_v_v@bk.ru

ORCID: 0009-0000-6784-8135



ТАНЫГИН Максим Олегович – д. т. н., профессор кафедры «Информационная безопасность», ЮЗГУ. Область научных интересов: анонимные одноранговые сети, аутентификация в многоагентных системах, автоматизированные системы документооборота. Число научных публикаций – 119.

Maxim Olegovich TANYGIN – Doctor of technical sciences, professor. Professor of the Information security department, South-West State University. Research interests: anonymous peer-to-peer networks, authentication in multi-agent systems, automated document management systems. The number of publications – 119.

email: tanygin@yandex.ru

ORCID: 0000-0002-4099-1414

Материал поступил в редакцию 25 марта 2025 года
