Динамика структуры указателей к стенографическим отчетам Государственной Думы начала XX в.: анализ корпуса текстов на основе разметки XML

**Человек:** Предметом исследования являются личные алфавитные указатели к стенографическим отчетам заседаний Государственной Думы начала XX в. Значение личных алфавитных указателей к стенографическим отчетам состоит в том, что их содержание в максимально структурированном виде отражает деятельность каждого депутата в Думе, что позволяет оценить деятельность парламентариев в количественном и качественном измерении. В статье прослежено изменение подходов к публикации информации, посвященной личным характеристикам и деятельности парламентариев, рассмотрена динамика структуры и содержания источника. В основе исследовательской методологии – анализ корпуса исторических текстов на основе технологии XML-разметки. Для решения задач исследования была специально разработана и применена схема XML-разметки текстов указателей, в структуре которой были предусмотрены теги для описания основных параметров источника – метаданных источника, личных характеристик депутатов и видов деятельности депутатов в ходе сессий I-IV созывов российского парламента. Анализ тегов разметки позволил представить степень подробности описания видов деятельности депутатов, а также изменения в структуре представленных данных, выявить постоянно отображаемую в источнике информацию о депутатах и сведения, которые варьировались в указателях разных сессий. Результаты данного исследования позволяют оценить информационный потенциал источника в его динамике, в перспективе восполнить неполноту данных сведениями из стенограмм, а также проанализировать вложенность тегов для классификации депутатов по характеристикам деятельности.

**Key words:** Указатель, стенографический отчет, Государственная Дума, XML-разметка, исторический корпус, структура источника, деятельность депутатов, XSD схема, контент-анализ, парламент

=================================

**FastText\_KMeans\_Clean:** В ходе проведения традиционного источниковедческого анализа был проанализирован порядок создания и публикации стенографических отчетов, отмечен личный вклад начальника думской канцелярии Я.В. Глинки и секретаря Думы Д.И Шаховского в формировании комплекса думской отчетности [3] и затронуты некоторые другие аспекты, влиявшие на содержание источника. <deputy id="226"> <deputyname> Леонов, Дмитрий Алексеевич </deputyname>. Так, наиболее детальные сведения изложены в указателях III созыва, тогда как наименее подробным является документ за 3 сессию IV созыва (в нем указаны данные лишь по 10 из 29 видов деятельности). Как видно из представленной таблицы (табл. Таким образом, реализованная XML-разметка текстов указателей показывает свою эффективность в изучении динамики структуры источника, а также предоставляет существенные возможности для преодоления неполноты данных указателей и дальнейшей работы по изучению деятельности парламентариев.

**Key words part:** 0.68

=================================

**FastText\_KMeans\_Raw/:** Цель статьи – изучение личных алфавитных указателей и динамики изменений их структуры с использованием методов "глубокой" разметки текстов на основе языка разметки XML, а также в определении информационного потенциала источника для дальнейшего использования в изучении истории российского парламента начала XX в. В ходе проведения традиционного источниковедческого анализа был проанализирован порядок создания и публикации стенографических отчетов, отмечен личный вклад начальника думской канцелярии Я.В. Глинки и секретаря Думы Д.И Шаховского в формировании комплекса думской отчетности [3] и затронуты некоторые другие аспекты, влиявшие на содержание источника. <deputy id="226"> <deputyname> Леонов, Дмитрий Алексеевич </deputyname>. То есть личные алфавитные указатели для полноценного использования при изучении описанных выше параметров деятельности депутата должны быть дополнены сведениями из стенографических отчетов, в которых публиковалась информация о поименном голосовании по различным вопросам повестки дня.

**Key words part:** 0.72

=================================

**FastText\_PageRank\_Clean/:** Историография парламентской истории России начала XX в. обширна. Целью проекта являлось проведение просопографического исследования депутатов Государственной Думы первого созыва. 1. Состав корпуса текстов по объему слов. 1. метаданные источника (созыв, заглавие, сессия, заседания, даты заседаний, издательство, год издания указателя);. 3. показатели, связанные с парламентской активностью депутата. <deputy id="226"> <deputyname> Леонов, Дмитрий Алексеевич </deputyname>. Скриншот фрагмента указателя, размеченного на основе XML. Динамика структуры источника.

**Key words part:** 0.64

=================================

**FastText\_PageRank\_Raw/:** Историография парламентской истории России начала XX в. обширна. 1. Состав корпуса текстов по объему слов. 1. метаданные источника (созыв, заглавие, сессия, заседания, даты заседаний, издательство, год издания указателя);. 3. показатели, связанные с парламентской активностью депутата. <deputy id="226"> <deputyname> Леонов, Дмитрий Алексеевич </deputyname>. Скриншот фрагмента указателя, размеченного на основе XML. Динамика структуры источника. Как видно из представленной таблицы (табл.

**Key words part:** 0.56

=================================

**Mixed\_ML\_TR/:** Цель статьи – изучение личных алфавитных указателей и динамики изменений их структуры с использованием методов "глубокой" разметки текстов на основе языка разметки XML, а также в определении информационного потенциала источника для дальнейшего использования в изучении истории российского парламента начала XX в. Методика данного исследования предполагала формирование исторического корпуса текстов на основе личных алфавитных указателей к стенографическим отчетам заседаний Государственной Думы I-IV созывов за 1906-1917 гг. Также более подробно в первых указателях изложены сведения о предложениях депутатов и внесенных поправках, тогда как в остальных указателях эти данные либо не публиковались совсем, либо указывались лишь в отдельных случаях. То есть личные алфавитные указатели для полноценного использования при изучении описанных выше параметров деятельности депутата должны быть дополнены сведениями из стенографических отчетов, в которых публиковалась информация о поименном голосовании по различным вопросам повестки дня. Отметим также изменения в структуре данных в указателях одного созыва, в частности, в указателях за 1-2 сессию III созыва не опубликованы сведения о депутатах, пропускавших заседания, тогда как в документах за 3-5 сессию эти сведения присутствуют. Социальные характеристики депутатов также не публиковались в одинаковом виде и имели как разную структуру данных в зависимости от созыва, так и разную полноту (либо эта информация отсутствовала в Указателях и публиковалась в иных приложениях к стенографическим отчетам). Таким образом, реализованная XML-разметка текстов указателей показывает свою эффективность в изучении динамики структуры источника, а также предоставляет существенные возможности для преодоления неполноты данных указателей и дальнейшей работы по изучению деятельности парламентариев.

**Key words part:** 0.88

=================================

**MultiLingual\_KMeans/:** Методика данного исследования предполагала формирование исторического корпуса текстов на основе личных алфавитных указателей к стенографическим отчетам заседаний Государственной Думы I-IV созывов за 1906-1917 гг. Также более подробно в первых указателях изложены сведения о предложениях депутатов и внесенных поправках, тогда как в остальных указателях эти данные либо не публиковались совсем, либо указывались лишь в отдельных случаях. То есть личные алфавитные указатели для полноценного использования при изучении описанных выше параметров деятельности депутата должны быть дополнены сведениями из стенографических отчетов, в которых публиковалась информация о поименном голосовании по различным вопросам повестки дня. Таким образом, реализованная XML-разметка текстов указателей показывает свою эффективность в изучении динамики структуры источника, а также предоставляет существенные возможности для преодоления неполноты данных указателей и дальнейшей работы по изучению деятельности парламентариев.

**Key words part:** 0.84

=================================

**Multilingual\_PageRank/:** Так, за III-IV созывы опубликовано 5 и 4 Указателей соответственно. Также отсутствует в корпусе указатель за экстренную сессию IV созыва, поскольку стенограммы дополнительной "секретной" сессии заседаний не публиковались. <deputy id="226"> <deputyname> Леонов, Дмитрий Алексеевич </deputyname>. Это наиболее очевидные виды деятельности по повестке дня. В Указателях за III-IV созывы эта ценная информация опускалась. 2), информация о голосовании совсем не встречается в ряде указателей (в том числе, нет данных о голосованиях в личных указателях I и II созывов, 2 сессии III созыва и нескольких сессий IV созыва). Вместе с тем, поименное голосование не практиковалось часто, и неполноту сведений о голосованиях можно объяснить самим характером голосований, когда результаты не фиксировались поименно, а было достаточно простого большинства для принятия решений, да и фамилии голосовавших коллективно в стенограммах не фиксировались. К таким относятся: "голосует", "возглас с места", "воздержался" (от голосования, например), "отпуск" (парламентарий мог пропускать заседания по этой причине) и "внеочередное заявление".

**Key words part:** 0.48

=================================

**RuBERT\_KMeans\_Without\_ST/:** Значение личных алфавитных указателей к стенографическим отчетам состоит в том, что их содержание в максимально структурированном виде отражает деятельность каждого депутата в Думе, что позволяет оценить деятельность парламентариев в количественном и качественном измерении. В ходе проведения традиционного источниковедческого анализа был проанализирован порядок создания и публикации стенографических отчетов, отмечен личный вклад начальника думской канцелярии Я.В. Глинки и секретаря Думы Д.И Шаховского в формировании комплекса думской отчетности [3] и затронуты некоторые другие аспекты, влиявшие на содержание источника. 2), информация о голосовании совсем не встречается в ряде указателей (в том числе, нет данных о голосованиях в личных указателях I и II созывов, 2 сессии III созыва и нескольких сессий IV созыва). Дальнейшее использование размеченных документов позволит изучать вложенность тегов, то есть выявлять наборы данных, свойственные отдельным депутатам и далее выделять типичные группы парламентариев согласно общности характеристик.

**Key words part:** 0.64

=================================

**RuBERT\_KMeans\_With\_ST/:** Структура разметки XML разрабатывалась с учетом необходимости решения конкретных источниковедческих задач, в частности, обеспечения возможностей изучения депутатов, принадлежавших к субэлите [10], то есть имеющим повторяющийся парламентский статус. Так, наиболее детальные сведения изложены в указателях III созыва, тогда как наименее подробным является документ за 3 сессию IV созыва (в нем указаны данные лишь по 10 из 29 видов деятельности). Так, исходя из сведений указателей для элемента <signed> ("подписал") были выявлены варианты документов для подписи, которые являются атрибутами и имеют вид в структуре разметки, как представлено в Таблице 3:. Особенностью указателя второго созыва является отсутствие информации о подписанных запросах, и заявления упоминаются только в общем виде без определения какого-то конкретного типа.

**Key words part:** 0.52

=================================

**RUBERT\_page\_rank\_Without\_ST/:** Историография парламентской истории России начала XX в. обширна. В дальнейшем их стали издавать посессионно. Динамика структуры источника. Это наиболее очевидные виды деятельности по повестке дня. Интересны факторы, влиявшие на полноту представленной в источнике информации.

**Key words part:** 0.48

=================================

**RUBERT\_page\_rank\_With\_ST/:** Историография парламентской истории России начала XX в. обширна. Схема XSD отражает основные тематические разделы указателей и включает в себя следующие данные:. 3. показатели, связанные с парламентской активностью депутата. Динамика структуры источника. К таким типам активности относятся "покидает заседание" и "меняется порядком выступления", встречающиеся в структуре указателей 2 и 1 раз соответственно.

**Key words part:** 0.6

=================================

**RUSBERT\_KMeans\_Without\_ST/:** В ходе проведения традиционного источниковедческого анализа был проанализирован порядок создания и публикации стенографических отчетов, отмечен личный вклад начальника думской канцелярии Я.В. Глинки и секретаря Думы Д.И Шаховского в формировании комплекса думской отчетности [3] и затронуты некоторые другие аспекты, влиявшие на содержание источника. Наиболее полно личные сведения раскрывают указатели III и IV созывов, в которых перечисленные данные дополняются целой совокупностью характеристик (принадлежность к партии, сословие, звание (чин), национальность, дата рождения, семейное положение, образование (в т.ч. учебное заведение), членство в комиссиях и комитетах (внедумская деятельность), владение собственностью (с указанием типа собственности и размеров имущества). Степень полноты данных о деятельности думцев, представленных в указателях, различается. Социальные характеристики депутатов также не публиковались в одинаковом виде и имели как разную структуру данных в зависимости от созыва, так и разную полноту (либо эта информация отсутствовала в Указателях и публиковалась в иных приложениях к стенографическим отчетам).

**Key words part:** 0.68

=================================

**RUSBERT\_KMeans\_With\_ST/:** Структура разметки XML разрабатывалась с учетом необходимости решения конкретных источниковедческих задач, в частности, обеспечения возможностей изучения депутатов, принадлежавших к субэлите [10], то есть имеющим повторяющийся парламентский статус. А поскольку в действительности эти случаи происходили с большей регулярностью, чем это отражено в указателях, то в данном случае очевидна неполнота данных источника, которая может быть восполнена на основании сведений из стенографических отчетов. Вместе с тем, поименное голосование не практиковалось часто, и неполноту сведений о голосованиях можно объяснить самим характером голосований, когда результаты не фиксировались поименно, а было достаточно простого большинства для принятия решений, да и фамилии голосовавших коллективно в стенограммах не фиксировались. Различия в структуре указателей имеются как между созывами, так и в структуре документов одного созыва.

**Key words part:** 0.6

=================================

**RUSBERT\_page\_rank\_Without\_ST/:** Динамика структуры источника. Таблица типов деятельности, соответствий тегов и их количества представлена ниже. Интересны факторы, влиявшие на полноту представленной в источнике информации. 2), информация о голосовании совсем не встречается в ряде указателей (в том числе, нет данных о голосованиях в личных указателях I и II созывов, 2 сессии III созыва и нескольких сессий IV созыва). В схеме XSD разметки указателей предусмотрены атрибуты, то есть уточняющие характеристики для некоторых элементов.

**Key words part:** 0.6

=================================

**RUSBERT\_page\_rank\_With\_ST/:** Динамика структуры источника. Это наиболее очевидные виды деятельности по повестке дня. Интересны факторы, влиявшие на полноту представленной в источнике информации. Как видно из представленной таблицы (табл. В схеме XSD разметки указателей предусмотрены атрибуты, то есть уточняющие характеристики для некоторых элементов.

**Key words part:** 0.6

=================================

**Simple\_PageRank/:** Значение личных алфавитных указателей к стенографическим отчетам состоит в том, что их содержание в максимально структурированном виде отражает деятельность каждого депутата в Думе, что позволяет оценить деятельность парламентариев в количественном и качественном измерении. Одно издание Указателей за 2 сессию IV созыва оказалось временно недоступным для включения в корпус ввиду того, что стенографические отчеты и указатели к ним стали библиографической редкостью и его добавление в корпус планируется сделать позднее. Наиболее полно личные сведения раскрывают указатели III и IV созывов, в которых перечисленные данные дополняются целой совокупностью характеристик (принадлежность к партии, сословие, звание (чин), национальность, дата рождения, семейное положение, образование (в т.ч. учебное заведение), членство в комиссиях и комитетах (внедумская деятельность), владение собственностью (с указанием типа собственности и размеров имущества). Так, сотрудники думской канцелярии выделяли некоторые описательные единицы, которые встречаются единично в текстах личных алфавитных указателей, хотя довольно регулярно отражаются в стенограммах заседаний. Полагаем, что указанные выше виды деятельности отражают значимые характеристики депутатов и могли бы способствовать более полной оценке роли каждого парламентария в собрании, его значимости, степени оказываемого ему доверия, если бы продолжали выделяться в последующих указателях. Также создана возможность для сравнения деятельности представителей субэлиты в Думе на протяжении нескольких созывов, поскольку тексты источников связаны на основе выделения персонального идентификатора для каждого депутата в структуре разметки.

**Key words part:** 0.72

=================================

**TextRank/:** Цель статьи – изучение личных алфавитных указателей и динамики изменений их структуры с использованием методов "глубокой" разметки текстов на основе языка разметки XML, а также в определении информационного потенциала источника для дальнейшего использования в изучении истории российского парламента начала XX в. Личные алфавитные указатели к стенографическим отчетам заседаний Государственной Думы относятся к высокоструктурированным источникам, основной принцип формирования которых сохранялся в течение всех четырех думских созывов начала XX в. Методика данного исследования предполагала формирование исторического корпуса текстов на основе личных алфавитных указателей к стенографическим отчетам заседаний Государственной Думы I-IV созывов за 1906-1917 гг. Отметим также изменения в структуре данных в указателях одного созыва, в частности, в указателях за 1-2 сессию III созыва не опубликованы сведения о депутатах, пропускавших заседания, тогда как в документах за 3-5 сессию эти сведения присутствуют. Социальные характеристики депутатов также не публиковались в одинаковом виде и имели как разную структуру данных в зависимости от созыва, так и разную полноту (либо эта информация отсутствовала в Указателях и публиковалась в иных приложениях к стенографическим отчетам). Вместе с тем, отсутствие в указателях информации о деятельности депутатов может быть восполнено на основе включения в указатели данных из текстов стенографических отчетов для получения более полной картины деятельности депутатов в Государственной думе.

**Key words part:** 0.84

=================================

**TF-IDF\_KMeans/:** Проектирование, создание и наполнение информационных систем "Российские парламентарии начала XX в" и "Стенографические отчеты Государственной Думы, 1906–1917" способствовали расширению возможностей применения компьютерных технологий для исследования парламентской истории. Отметим также изменения в структуре данных в указателях одного созыва, в частности, в указателях за 1-2 сессию III созыва не опубликованы сведения о депутатах, пропускавших заседания, тогда как в документах за 3-5 сессию эти сведения присутствуют. Вместе с тем, принятие резолюций, как особого типа документа, характерно и для деятельности депутатов последующих созывов [11]. Таким образом, реализованная XML-разметка текстов указателей показывает свою эффективность в изучении динамики структуры источника, а также предоставляет существенные возможности для преодоления неполноты данных указателей и дальнейшей работы по изучению деятельности парламентариев.

**Key words part:** 0.76

=================================

**Текст:** Стенографические отчеты заседаний Государственной Думы являются основным источником для изучения парламентской истории Российской Империи. Это подробное многотомное официальное издание стенограмм служило основой для формирования личных и предметных указателей к самим отчетам. Значение личных алфавитных указателей к стенографическим отчетам состоит в том, что их содержание в максимально структурированном виде отражает деятельность каждого депутата в Думе, что позволяет оценить деятельность парламентариев в количественном и качественном измерении. Указатели имеют четкую структуру в оформлении, что располагает к применению компьютерных методов в их изучении, создает высокий потенциал для извлечения информации нового уровня, получение которой традиционными способами весьма затруднительно. Анализ структуры указателей и ее изменений позволяет выявить возможности для дальнейшего изучения источника, оценить его информационный потенциал, представить степень подробности описания деятельностных характеристик депутатов, а также изменения в структуре представленных данных.. Цель статьи – изучение личных алфавитных указателей и динамики изменений их структуры с использованием методов «глубокой» разметки текстов на основе языка разметки XML, а также в определении информационного потенциала источника для дальнейшего использования в изучении истории российского парламента начала XX в.. . Историография. Историография парламентской истории России начала XX в. обширна. И в исторической науке к настоящему времени сформировалось особое направление, связанное с изучением российского парламентаризма на основе методов, подходов и инструментария, основанных на применении информационных технологий. Одним из первых таких исследований стал проект “Duma I”, выполненный московскими историками - Л.И. Бородкиным, Ю.Г. Григорьевой, Н.Б. Селунской и др. [1]. Целью проекта являлось проведение просопографического исследования депутатов Государственной Думы первого созыва. На основе проектирования, заполнения и анализа данных информационной системы авторами был изучен коллективный портрет парламентариев начала XX в.. Импульс источниковедческому анализу важного источника по думской истории начала XX в. – стенографических отчетов заседаний Государственной Думы – был придан реализацией целого ряда научно-исследовательских проектов учеными Пермского университета. В 2003–2018 гг. коллективом исследователей под руководством И.К. Кирьянова и С.И. Корниенко были рассмотрены различные аспекты истории российского парламента начала XX в. [2]. Помимо других ресурсов, источниковой основой осуществления данных проектов стали именно стенографические отчеты заседаний. В ходе проведения традиционного источниковедческого анализа был проанализирован порядок создания и публикации стенографических отчетов, отмечен личный вклад начальника думской канцелярии Я.В. Глинки и секретаря Думы Д.И Шаховского в формировании комплекса думской отчетности [3] и затронуты некоторые другие аспекты, влиявшие на содержание источника.. Важной составной частью работы над этими проектами было развитие информационной среды, включающей, в том числе, оцифрованные и структурированные источники, а также разработанный инструментарий для работы с данными. Проектирование, создание и наполнение информационных систем «Российские парламентарии начала XX в.» и «Стенографические отчеты Государственной Думы, 1906–1917» способствовали расширению возможностей применения компьютерных технологий для исследования парламентской истории. Указанные системы были интегрированы в единую платформу, а итоговым информационным продуктом стал продолжающий развиваться и наполняться научно-образовательный интернет-портал «Парламентская история дореволюционной России» [4-7].. Оцифровка источников и их организация на основе информационных технологий позволили применить количественные методы исследования и компьютеризированный контент-анализ к текстам источников.. Одним из первых исследователей, использовавших количественные методы для анализа думской документации стал А.Л. Кобринский, который изучал деятельность современного российского парламента [6]. Методология исследования была основана на технологии контент-анализа стенограмм заседаний Государственной Думы 1993-1995 гг., выделении в выступлениях ряда депутатов значимых словоформ и категорий и их подсчете, а также анализе распределения указанных единиц в текстах.. Таким образом, в изучении российской парламентской истории использовались различные компьютеризированные методы анализа источников, и накопленный опыт количественного анализа и интерпретации данных был использован в данном исследовании.. . Методология. Методология данного исследования основана на анализе исторических корпусов и использовании XML-разметки текстов источников, что впервые предпринимается по отношению к думской отчетности начала XX в.. Личные алфавитные указатели к стенографическим отчетам заседаний Государственной Думы относятся к высокоструктурированным источникам, основной принцип формирования которых сохранялся в течение всех четырех думских созывов начала XX в. Указатель является перечнем данных о депутатах и их деятельности в Думе, организованным в алфавитном порядке [9]. Методика данного исследования предполагала формирование исторического корпуса текстов на основе личных алфавитных указателей к стенографическим отчетам заседаний Государственной Думы I-IV созывов за 1906-1917 гг. Основные этапы этого процесса заключались в переводе источников в машиночитаемый формат, их редактировании, а также разметке текстов на основе языка разметки XML согласно разработанной схеме тегов, отвечающей целям и задачам исследования.. Указатели первоначально публиковались по одному тому за созыв (так было в I-II созывах). В дальнейшем их стали издавать посессионно. Так, за III-IV созывы опубликовано 5 и 4 Указателей соответственно. Одно издание Указателей за 2 сессию IV созыва оказалось временно недоступным для включения в корпус ввиду того, что стенографические отчеты и указатели к ним стали библиографической редкостью и его добавление в корпус планируется сделать позднее. Также отсутствует в корпусе указатель за экстренную сессию IV созыва, поскольку стенограммы дополнительной «секретной» сессии заседаний не публиковались. Работа над проектом продолжается, и на данный момент отредактировано и размечено 10 томов указателей, созданный корпус текстов имеет совокупный объем в 749 793 слова (табл.1).. . Табл. 1. Состав корпуса текстов по объему слов. Структура системы тегов (отражена в файле-схеме XSD) XML-разметки разрабатывалась параллельно с осуществлением разметки, поскольку за ее основу выбрана структура источника. Схема XSD отражает основные тематические разделы указателей и включает в себя следующие данные:. 1. метаданные источника (созыв, заглавие, сессия, заседания, даты заседаний, издательство, год издания указателя);. 2. личные характеристики депутата (ФИО, губерния избрания, признание правильности выборов, партия, возраст, социальный статус, сословие, курия избрания и некоторые другие сведения);. 3. показатели, связанные с парламентской активностью депутата.. Структура разметки XML разрабатывалась с учетом необходимости решения конкретных источниковедческих задач, в частности, обеспечения возможностей изучения депутатов, принадлежавших к субэлите [10], то есть имеющим повторяющийся парламентский статус. Изучение деятельности переизбиравшихся депутатов стало возможным благодаря введению параметра идентификации для каждого депутата. В структуре разметки идентификатор имеет номерное обозначение, уникальное для каждого парламентария, и выглядит так:. <deputy id="226"> <deputyname> Леонов, Дмитрий Алексеевич </deputyname>. Создание единого исторического корпуса личных алфавитных указателей позволило трансформировать разрозненные источники в единый массив связанных текстов, а использование XML-разметки, в том числе персональных идентификаторов, стало основой для организации текстов в табличном представлении, удобном для анализа (рис.1).. . . Рис. 1. Скриншот фрагмента указателя, размеченного на основе XML. . Анализ разметки. Динамика структуры источника. Личные характеристики депутатов публиковались в Указателях на протяжении всех созывов Государственной Думы, однако полнота данных в рамках сессий была различной. Так, в разделе с личными характеристиками депутатов указатели I-II созывов содержат лишь краткие данные (ФИО, губерния избрания, признание правильности выборов и указание принадлежности к отделу). Наиболее полно личные сведения раскрывают указатели III и IV созывов, в которых перечисленные данные дополняются целой совокупностью характеристик (принадлежность к партии, сословие, звание (чин), национальность, дата рождения, семейное положение, образование (в т.ч. учебное заведение), членство в комиссиях и комитетах (внедумская деятельность), владение собственностью (с указанием типа собственности и размеров имущества). В указателях III созыва эта информация публикуется только в указателе за 1 сессию, тогда как в остальных она упускается. В указателях IV созыва эти данные фиксируются на протяжении всех сессий.. В процессе работы над разметкой текстов указателей и при формировании схемы данных XSD всего корпуса текстов было определено 29 видов деятельности депутатов в Думе. Нами выделено 4 типа тегов, обозначающих разные виды деятельности депутатов, в зависимости от наличия соответствующих тегов в личных алфавитных указателях по созывам (I тип – теги представлены в указателях всех созывов, II тип – в указателях трех созывов, III – в указателях двух созывов, IV – в указателях одного созыва). Таблица типов деятельности, соответствий тегов и их количества представлена ниже.. Табл. 2 A-D. Типы деятельности депутатов Государственной Думы начала XX в., написание и значение тегов, количество тегов в текстах указателей по сессиям I-IV созывов.. Табл. 2 A. I тип тегов. . Табл. 2 B. II тип тегов. . Табл. 2 C. III тип тегов. . Табл. 2 D. IV тип тегов. . . Из таблицы видны типы представленной информации о деятельности депутатов, и 16 из 29, то есть большинство из них, выделены в личных алфавитных указателях трех или всех четырех созывов.. Степень полноты данных о деятельности думцев, представленных в указателях, различается. Так, наиболее детальные сведения изложены в указателях III созыва, тогда как наименее подробным является документ за 3 сессию IV созыва (в нем указаны данные лишь по 10 из 29 видов деятельности). Очевидно, что краткость данных в упомянутом указателе основывается на особенностях формирования документа, поскольку было проведено лишь 3 заседания в сессии.. Видами деятельности депутатов, которые указывались постоянно являются следующие: подписал документ, говорит по теме, вносит поправки, объясняется (по личному вопросу), говорит по порядку дня, отказывается от слова, замечания Председателя депутату, докладывает, находится под судебным преследованием, занимает должность, сообщение о самом депутате, объявляет от имени партии, отсутствует в зале заседания, отстранён от заседаний, заявляет, отказывается от должности. Это наиболее очевидные виды деятельности по повестке дня. Однако обращает на себя внимание тот факт, что в ряде сессий указатели совсем не отражают часть других значимых сведений.. Интересны факторы, влиявшие на полноту представленной в источнике информации. Одним из них является субъективный подход секретариата при выделении значимых пунктов депутатской деятельности и структуры личного алфавитного указателя. Первоначально канцелярия предприняла попытку выделения максимально возможного количества видов деятельности, и эта избыточность в дальнейшем корректировалась. Так, сотрудники думской канцелярии выделяли некоторые описательные единицы, которые встречаются единично в текстах личных алфавитных указателей, хотя довольно регулярно отражаются в стенограммах заседаний. К таким типам активности относятся «покидает заседание» и «меняется порядком выступления», встречающиеся в структуре указателей 2 и 1 раз соответственно. А поскольку в действительности эти случаи происходили с большей регулярностью, чем это отражено в указателях, то в данном случае очевидна неполнота данных источника, которая может быть восполнена на основании сведений из стенографических отчетов.. В указателях за первые два созыва были выделены и такие виды деятельности как «предлагается на должность», «информирует» (сообщает новость или некие дополнительные сведения), «объявляет как представитель региона», «лишается права голоса». Так, в документах подробно определены виды деятельности депутатов, которые содержат также сведения о том, на какие позиции были предложены депутаты, как проходила баллотировка, сколько шаров получил парламентарий при избрании на должность. В Указателях за III-IV созывы эта ценная информация опускалась. Также более подробно в первых указателях изложены сведения о предложениях депутатов и внесенных поправках, тогда как в остальных указателях эти данные либо не публиковались совсем, либо указывались лишь в отдельных случаях. Полагаем, что указанные выше виды деятельности отражают значимые характеристики депутатов и могли бы способствовать более полной оценке роли каждого парламентария в собрании, его значимости, степени оказываемого ему доверия, если бы продолжали выделяться в последующих указателях.. Другим, более существенным примером неполноты данных, является информация о депутатском голосовании. Как видно из представленной таблицы (табл. 2), информация о голосовании совсем не встречается в ряде указателей (в том числе, нет данных о голосованиях в личных указателях I и II созывов, 2 сессии III созыва и нескольких сессий IV созыва). А этот вид деятельности является существенным, поскольку непосредственно влиял на принимаемые в Думе решения. То есть личные алфавитные указатели для полноценного использования при изучении описанных выше параметров деятельности депутата должны быть дополнены сведениями из стенографических отчетов, в которых публиковалась информация о поименном голосовании по различным вопросам повестки дня. Вместе с тем, поименное голосование не практиковалось часто, и неполноту сведений о голосованиях можно объяснить самим характером голосований, когда результаты не фиксировались поименно, а было достаточно простого большинства для принятия решений, да и фамилии голосовавших коллективно в стенограммах не фиксировались.. Структура указателя со временем видоизменялась, и вводились новые сведения, данных о которых нет в указателях первых двух созывов Думы. К таким относятся: «голосует», «возглас с места», «воздержался» (от голосования, например), «отпуск» (парламентарий мог пропускать заседания по этой причине) и «внеочередное заявление». Причем о значимости этих видов деятельности говорят количественные характеристики тегов (см. Таблицу 2).. Отметим также изменения в структуре данных в указателях одного созыва, в частности, в указателях за 1-2 сессию III созыва не опубликованы сведения о депутатах, пропускавших заседания, тогда как в документах за 3-5 сессию эти сведения присутствуют.. В схеме XSD разметки указателей предусмотрены атрибуты, то есть уточняющие характеристики для некоторых элементов. Так, исходя из сведений указателей для элемента <signed> («подписал») были выявлены варианты документов для подписи, которые являются атрибутами и имеют вид в структуре разметки, как представлено в Таблице 3:. Табл. 3. Атрибуты элемента <signed> («подписал») и количественные данные по ним в личных алфавитных указателях. . . Подробность описания типов подписываемых документов депутатами в указателях также различается (см. Таблицу 3). Так, тип документа «резолюция» выделяется только в указателе первого созыва Думы, а «заявление с протестом» и «особое мнение» свойственны только ряду указателей III созыва нижней палаты. Вместе с тем, принятие резолюций, как особого типа документа, характерно и для деятельности депутатов последующих созывов [11]. Особенностью указателя второго созыва является отсутствие информации о подписанных запросах, и заявления упоминаются только в общем виде без определения какого-то конкретного типа.. . Заключение. Разметка текстов указателей на основе XML позволила проанализировать структуру указателей в динамике. Так, структура личных алфавитных указателей к стенографическим отчетам претерпела существенные изменения за время их публикации в 1906-1917 гг. Различия в структуре указателей имеются как между созывами, так и в структуре документов одного созыва. Структура первых двух изданных указателей (сессии I-II созывов) свидетельствует об опытном характере публикаций, когда указываемые первоначально данные в дальнейшем показались избыточными и упускались. Первоначально указатели содержали дополнительные сведения о процессах избрания, данные о которых опускались в дальнейшем. Социальные характеристики депутатов также не публиковались в одинаковом виде и имели как разную структуру данных в зависимости от созыва, так и разную полноту (либо эта информация отсутствовала в Указателях и публиковалась в иных приложениях к стенографическим отчетам). Вместе с тем, отсутствие в указателях информации о деятельности депутатов может быть восполнено на основе включения в указатели данных из текстов стенографических отчетов для получения более полной картины деятельности депутатов в Государственной думе.. Дальнейшее использование размеченных документов позволит изучать вложенность тегов, то есть выявлять наборы данных, свойственные отдельным депутатам и далее выделять типичные группы парламентариев согласно общности характеристик. Разметка источников на основе XML и формирование корпуса текстов обладают значительно большим потенциалом по сравнению с традиционным контент-анализом в исследовании структуры документов, фрагменты которых следует рассматривать отдельно друг от друга. Также создана возможность для сравнения деятельности представителей субэлиты в Думе на протяжении нескольких созывов, поскольку тексты источников связаны на основе выделения персонального идентификатора для каждого депутата в структуре разметки.. Таким образом, реализованная XML-разметка текстов указателей показывает свою эффективность в изучении динамики структуры источника, а также предоставляет существенные возможности для преодоления неполноты данных указателей и дальнейшей работы по изучению деятельности парламентариев.