

Манифест для разметки курсовых, ВКР и магистерских диссертаций (9.09.2022)

Необходимо обрабатывать магистерские и дипломные работы, задача извлечь текст в структурированном виде:

- Титульный лист (**title**)
- Содержание (**toc**)
- Главы / подглавы с их текстом (**named_item**, **named_item_continue**, **raw_text**, **footnote**, **page_id**)

Для этого необходимо решить задачу классификации строк документа. На вход будет подаваться изображение страницы документа с одной строкой, обведённой в рамку. Необходимо каждую строку пометить одним из следующих типов:

- Все строки титульного листа – **title**;
- Заголовок содержания и строки содержания – **toc**;
- Первые строки названий глав/подглав – **named_item**;
- Продолжения названий глав/подглав (в случае многострочных заголовков) – **named_item_continue**;
- Текст глав/подглав – **raw_text**;
- Номера страниц — **page_id**;
- Сноски – **footnote**;
- Остальные выделенные элементы, не являющиеся текстовыми, нужно отмечать как **other**.

Титульный лист – есть во всех работах. Содержит название университета, факультета, тему работы, имя автора, руководителя и год написания работы. Все текстовые строки титульного листа отмечаются как **title**. При этом рукописную подпись и другие нетекстовые элементы надо помечать как **other**

| |
|---|
| Федеральное государственное бюджетное образовательное учреждение высшего образования РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА и ГОСУДАРСТВЕННОЙ СЛУЖБЫ при ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ Институт государственной службы и управления |
| Направление подготовки: 38.04.01 Экономика Образовательная программа: Государственное регулирование экономики |
| Магистерская диссертация на тему: «Развитие мер поддержки несырьевого экспорта в Московской области» |
| Автор: обучающийся группы ЗМ18-35 заочной формы обучения  Двойченков Павел Александрович |
| Руководитель: Доктор экономических наук, профессор кафедры теории и практики государственного контроля ИГСУ РАНХиГС  Ишина Ирина Валериевна |
| Москва, 2021 г. |

Содержание – это план работы студента, только в развернутой форме и с демонстрацией места расположения данных (глав). Строка заголовка содержания и все строки содержания отмечаются как **toc**.

Пример содержания

| СОДЕРЖАНИЕ | |
|--|----|
| ВВЕДЕНИЕ | 4 |
| 1 ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ АНАЛИЗА ПОЛИТИЧЕСКОЙ СИСТЕМЫ ОБЩЕСТВА | 10 |
| 1.1 Политическая система общества: понятие, функции, модели | 10 |
| 1.2 Особенности политической системы РФ | 16 |
| 1.3. Правовые основы модернизации политической системы | 22 |
| 2 МОДЕРНИЗАЦИЯ ПСО РФ НА СОВРЕМЕННОМ ЭТАПЕ РАЗВИТИЯ | 30 |
| 2.1. Изменения в законодательной ветви власти РФ | 30 |
| 2.2. Обновление президентской республики | 35 |
| 2.3. Реконструкция ПСО в отдельных сферах государственного строительства | 42 |
| ЗАКЛЮЧЕНИЕ | 51 |
| БИБЛИОГРАФИЧЕСКИЙ СПИСОК | 54 |

Главы / подглавы: Главы разделяют текст документа на смысловые части. Например:

Пример начала главы «Введение»

Введение

Тема поддержки несырьевого экспорта Российской Федерации только набирает силу и приобретает свою актуальность переняв инициативу от курса страны, направленного на стратегию импортозамещения, которая активно продвигалась государством в последние несколько лет.

Заголовки глав нужно помечать как `named_item`. Обычно заголовки глав/подглав выделяются визуально (жирностью, размером шрифта, нумерацией), эти заголовки присутствуют в содержании документа.

Например, выделенную на изображении ниже строку нужно пометить **`named_item`**.

Введение

Тема поддержки несырьевого экспорта Российской Федерации только набирает силу и приобретает свою актуальность переняв инициативу от курса страны, направленного на стратегию импортозамещения, которая активно продвигалась государством в последние несколько лет.

Текст глав нужно помечать как **`raw_text`**. Пример **`raw_text`** дан ниже:

Введение

Тема поддержки несырьевого экспорта Российской Федерации только набирает силу и приобретает свою актуальность переняв инициативу от курса страны, направленного на стратегию импортозамещения, которая активно продвигалась государством в последние несколько лет.

Помимо примеров выше возможны и случаи, когда глава содержит многострочные заголовки и подглавы:

1 ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ АНАЛИЗА ПОЛИТИЧЕСКОЙ СИСТЕМЫ ОБЩЕСТВА

1.1 Политическая система общества: понятие, функции, модели

Прежде, чем обратиться к словарным определениям понятия «политическая система общества», разобраться с функциями и моделями политической системы общества (ПСО), необходимо озвучить несколько предварительных замечаний.

Выделенную на изображении ниже строку нужно пометить как **named_item** (это первая строка заголовка):

1 ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ АНАЛИЗА ПОЛИТИЧЕСКОЙ СИСТЕМЫ ОБЩЕСТВА

1.1 Политическая система общества: понятие, функции, модели

Прежде, чем обратиться к словарным определениям понятия

Следующую выделенную строку надо пометить как **named_item_continue**, так как это строка-продолжение заголовка (см. пример ниже).

1 ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ АНАЛИЗА ПОЛИТИЧЕСКОЙ СИСТЕМЫ ОБЩЕСТВА

1.1 Политическая система общества: понятие, функции, модели

Прежде, чем обратиться к словарным определениям понятия

Следующую выделенную строку нужно снова пометить как **named_item**, так как это первая строка заголовка подглавы (см. пример ниже).

1 ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ АНАЛИЗА ПОЛИТИЧЕСКОЙ СИСТЕМЫ ОБЩЕСТВА

1.1 Политическая система общества: понятие, функции, модели

Прежде, чем обратиться к словарным определениям понятия

Сноска - ссылка вне основного текста на источник информации, использованный при написании статьи, или комментарий. Сноски должны быть отмечены как **footnote**.

Например:

- аудит системы внутриорганизационного экспортного контроля и бизнес процессов в части экспортного контроля, включая анализ ВЭД, консультации и

¹ Официальный сайт Федеральной службы по техническому и экспортному контролю:
<https://fstec.ru/eksportnyj-kontrol/zakonodatelstvo/96-zakony/>

36

Например, на изображениях, показанных ниже, выделенные в рамку строки нужно пометить как **footnote**.

- аудит системы внутриорганизационного экспортного контроля и бизнес процессов в части экспортного контроля, включая анализ ВЭД, консультации и

¹ Официальный сайт Федеральной службы по техническому и экспортному контролю:
<https://fstec.ru/eksportnyj-kontrol/zakonodatelstvo/96-zakony/>

36

- аудит системы внутриорганизационного экспортного контроля и бизнес процессов в части экспортного контроля, включая анализ ВЭД, консультации и

¹ Официальный сайт Федеральной службы по техническому и экспортному контролю:
<https://fstec.ru/eksportnyj-kontrol/zakonodatelstvo/96-zakony/>

36

Номер страницы – текст с номером страницы, обычно расположенный внизу страницы. Номера страниц должны быть отмечены как **page_id**.

Пример:

- аудит системы внутриорганизационного экспортного контроля и бизнес процессов в части экспортного контроля, включая анализ ВЭД, консультации и

¹ Официальный сайт Федеральной службы по техническому и экспортному контролю:
<https://fstec.ru/eksportnyj-kontrol/zakonodatelstvo/96-zakony/>