**Описание процесса проведения исследований**

**Перевод НПА и НТД в цифровую форму**

Существует некоторая путаница в отношении использования таких терминов, как: оцифровка, цифровизация и цифровое преобразование.

Для любых документов можно выделить следующие “уровни” цифровизации:

1. Документ представлен в бумажном виде;
2. Документ представлен в виде файла, содержащего графическое изображение документа;
3. Документ представлен в виде файла, содержащего только текст (набор символов) документа;
4. Документ представлен в виде файла, помимо текста содержащего информацию о структуре документа (разделение на разделы, пункты и т.п.);
5. Документ представлен в виде файла, содержащего текст, структуру и разметку (выделение в тексте отдельных терминов);
6. Документ представлен в виде файла (набора файлов), содержащего машиночитаемые правила.

В настоящее время документы в цифровой форме в основном, представляют собой документы уровня 4. Процедура, позволяющая перевести документы с уровня 1 на уровень 4 называется “оптическим распознаванием символов” (optical character recognition, OCR). Процедура OCR в настоящее время реализована во многих программных продуктах, таких как Abbyy FineReader, CuneiForm, Adobe Acrobat DC и т.п. И не является предметом НИР.

Процедура, переводящая документы с уровня 4 на уровень 5 называется “разметкой документа” (annotation). Разметка документа это процесс добавления информации в текст документа, выполняемый вручную экспертами. Процесс добавления информации в текст обычно требует принятием экспертов решения о разметке того или иного участка текста, базирующегося как на самом тексте, так и на знаниях эксперта. Обычно процесс разметки документа состоит в выделении фрагмента текста, подлежащих аннотации и добавлению к выбранному фрагменту дополнительной информации из фиксированного классификатора.

Размеченные документы обычно используются в задачах искусственного интеллекта и машинного обучения связанных с задачами обработки текста на естественном языке.

Необходимо отметить, что содержимое классификатора, используемого для разметки документа зависит от задачи, в рамках решения которой проводится разметка документа.

В дальнейшем, используя документы уровня 5, можно приступать к формированию машиночитаемых правил.

Документ уровня 2-4 обычно хранятся в виде одного файла, размещаемого на файловой системе компьютера. Документ уровня 5 или 6 может быть представлен несколькими файлами и, как правило, для хранения и работы с этими документами используются специализированное программное обеспечение или информационные системы.

Процесс перевода НПА и НТД в цифровую форму можно разделить на следующие этапы:

1. Формирование машиночитаемого корпуса документов (OCR) - формирование корпуса документов, удовлетворяющих требованиям уровня 4;
2. Разметка корпуса документов - классификация содержимого документов в соответствии с поставленной задачей и перевод документов на уровень 5;
3. Формирование машиночитаемых правил на основании размеченных документов.

На основе опыта Сингапура в области автоматизированных проверок была выработана система классификации документов, используемая для разметки корпуса документов в рамках задачи оцифровки нормативов для автоматизированной проверки информационной модели строительного объекта на основе требований НТД. Система классификации документов приведена в таблице \_\_\_\_.

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Тип** | **Цветовое обозначение** | **Комментарий** |
| 1 | Понятие |  | Задает термин, обозначающей понятие строительной отрасли, к которой задается требование |
| 2 | Уровень обязательности  (глагол) |  | Задаёт уровень обязательности требования |
| 3 | Требование |  | Задаёт непосредственно требование к понятию |
| 4 | Условие применимости |  | Задает условие, при которых необходимо проверять данное требование |
| 5 | Характеристика понятия |  | Задает характеристику понятия, к которой задаётся требование |

Например:

В жилых зданиях с одной лестничной клеткой устраиваются аварийные выходы из квартир с учетом уровня их расположения.

Внутренний диаметр ствола мусоропровода должен быть не менее 400 мм, при этом величина размера диагонали ковша загрузочного клапана не должна превышать величину 0,9 от размера внутреннего диаметра труб ствола для исключения образования засоров внутри ствола.

Далее, после того как выполнена разметка документа, выявленные требования формализуются и оформляются в виде таблицы. Единицей разметки является предложение в документе. Структура таблицы приведена в таблице\_\_\_.

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование поля** | **Назначение поля** |
| 1 | Понятие |  |
| 2 | Характеристика |  |
| 3 | Наименование объекта и характеристики в ЭМ СО | Класс и свойство объекта из электронной модели строительного объекта, которому соответствует выделенное понятие |
| 4 | Условие применимости |  |
| 5 | Правило проверки для условия применимости в терминах ЭМ СО | Указывается либо имя ранее определенного правила либо формула для проведения проверки |
| 6 | Уровень обязательности |  |
| 7 | Требование |  |
| 8 | Правило проверки требования |  |
| 9 | Предельное значение |  |
| 10 | Единица измерения предельного значения |  |

Например:

В жилых зданиях с одной лестничной клеткой устраиваются аварийные выходы из квартир с учетом уровня их расположения

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Понятие | Характеристика | Наименование объекта и характеристики в ЭМ СО | Условие применимости | Правило проверки для условия применимости в терминах ЭМ СО | Уровень обязательности | Требование | Правило проверки требования | Предельное значение | Единица измерения предельного значения |
| Жилое зданий |  | residential\_building | с одной лестничной клеткой | residential\_building  .stairwell = 1 | Обязательно | устраиваются аварийные выходы из квартир с учетом уровня их расположения | Содержит | emergency\_exit | шт. |

Внутренний диаметр ствола мусоропровода должен быть не менее 400 мм, при этом величина размера диагонали ковша загрузочного клапана не должна превышать величину 0,9 от размера внутреннего диаметра труб ствола для исключения образования засоров внутри ствола.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Понятие | Характеристика | Наименование объекта и характеристики в ЭМ СО | Условие применимости | Правило проверки для условия применимости в терминах ЭМ СО | Уровень обязательности | Требование | Правило проверки требования | Предельное значение | Единица измерения предельного значения |
| ствол мусоропровода | внутренний диаметр | rubbish chute.diameter |  |  | Обязательно | должен быть не менее 400 мм | Меньше или равно | 400 | мм |
| ковш загрузочного клапана | диагональ | garbage\_ chute\_ ladle.size |  |  | Обязательно | не должна превышать величину 0,9 от размера внутреннего диаметра труб ствола | Меньше или равно | 0,9 \* rubbish chute.diameter | мм |