Составление корпуса документов в соответствии с определенными предметными областями

1. Сбор документов. Документ должен быть уникально отнесен к той или иной предметной области. Не допускается, чтобы один документ относился к нескольким предметным областям
2. Преобразование документов в необходимый вид

Формирование словаря по корпусу документов выбранной предметной области

1. Формирование списка существующих терминов и определений, в явном виде присутствующих в корпусе
2. Проверка сформированного списка на коллизии в определениях
3. Экспертное согласование списка существующих терминов и выработка уникального определения для каждого термина
4. Формирование списка потенциальных терминов с использованием технологии частотного анализа
   1. Подготовка документов для частотного анализа
      1. Удаление разделов документа, содержащих определения терминов в явном виде
      2. Удаление служебной информации из документа
   2. Проведедение частотного анализа
   3. Экспертная оценка списка потенциальных терминов полученных с помощью частотного анализа
5. Экспертная работа в части получение списка терминов предметной области и их определений
   1. Сопоставление списков существующих терминов и потенциальных терминов и исключение существующих терминов, имеющих определение, но не встречающихся в результатах частотного анализа
   2. Исключение терминов, не имеющих отношения к выбранной предметной области
   3. Выработка уникального определения для терминов, не имеющих определений
6. Экспертная работа в части сопоставления полученного списка терминов с существующими системами классификации
   1. Сопоставление терминов с Единым классификатором строительной отрасли (ISO 81346)
   2. Сопоставление терминов с IFC
7. Экспертная работа в части выделения атрибутов терминов (характеристик понятия, которое обозначает термин)

Формирование единого словаря строительной отрасли

1. Объединение словарей по предметной области в единый словарь
2. Группировка терминов по
   1. Системам ISO 81346
   2. По родительскому термину (для терминов состоящих из существительного и прилагательного)
   3. По требованию ТР
   4. По алфавиту
3. Вёрстка единого словаря и/или разработка программной оболочки для навигации по словарю

Перевод НПА и НТД в цифровую форму

1. Используются те же корпуса документов, что и для формирования словаря
2. Подготовка документов
   1. Девариация и снятие неоднозначности в тексте документа в соответствии с правилами девариации.
3. Разметка предложений получившегося документа в соответствии с правилами разметки
4. Привязка терминов в предложениях размеченного документа с терминами единого словаря строительной отрасли
5. Экспертная работа в части классификации требований из размеченных документов в соответствии с правилами классификации (формализуемые/неформализуемые/частично формализуемые)
6. Экспертная работа по выделению параметризуемых требований из формализуемых требований на основе правил
7. Экспертная работа представлению параметризуемых требований в виде файлов, содержащих машиночитаемые инструкций
8. Экспертная работа по составлению непараметризуемых требований в виде файлов, содержащих машиночитаемые инструкций
9. Экспертная работа по выработке предложений по неформализуемый/частично формализуемым требованиям