Правила:

- во всех ссылках сразу убираем www

- во всех ссылках сразу меняем https:// на http://

Основной класс проходит по todoTaskList, проверяет нет ли текущей ссылки в БД. Если нет, открывает ссылку, создает сущность для БД, заносит в неё укороченную ссылку без начальной ссылки класса, из открытой ссылки вырезает все элементы "a[href]" и передает их в класс-поток. То, что вернулось из потока, заносит в todoTaskList. После удаляет первый элемент в списке todoTaskList и идет на следующую итерацию цикла.

todoTaskList – список необработанных ссылок с WWW и может с HTTP или HTTPS, наполняется из класса-потока списком urlReadyToTodo

defaultUrl – текущая в цикле ссылка из списка todoTaskList

elements = doc.select("a[href]") - с WWW и может с HTTP или HTTPS

urlSite – начальная ссылка из сайта-аргумента на входе в класс, в самом начале из неё убирается WWW

класс-поток (на выходе список ссылок urlReadyToTodo с www и может с http или https) :

- если количество элементов на входе < 2 (только одна ссылка), делает «перебор» по списку.

1. Если ссылка есть в todoTaskList , пропускает итерацию (фактически выходит из потока, ничего не добавляя в todoTaskList)

2. Изменяет в ссылке https:// на http://

3. Если ссылка пустая или равна /, заменяет её на urlSite без www и начинающуюся на http://

4. Убирает из ссылки www

5. С помощью укороченной ссылки minusUrl проверяется, что ссылка начинается с начальной ссылки общего класса

6. С помощью minusResh проверяется, что не внутренняя на странице, т.е. не имеет #

7. Если minusUrl отсутствует в БД pages, добавляется urlStr в urlReadyToTodo. urlStr попадает в список без www и с http://

- если количество элементов на входе > 1, делим поток на два потока

urlReadyToTodo – список на выходе класса-потока, наполняется ссылками, прошедшими все проверки

elements – аргумент, равный elements из основного блока

urlStr = el.absUrl("href") – конкретная ссылка из elements, с WWW и может с HTTP или HTTPS

minusUrl – ссылка urlStr за вычетом начальной ссылки – для определения ссылки с рассматриваемого сайта