

PARSER GENERATOR

Оглавление

Глава 1. Интерфейсы

Tree	3
TreePath	3
Markup	4
Algorithm	5

Глава 2. Реализации

Глава 1. Интерфейсы

Tree

Интерфейс деревьев, для парсинга которых генерируется парсер. Задаёт три абстрактных метода:

```
def get_value(self, tree_path: TreePath) -> str
```

Метод возвращает значение элемента дерева, на который указывает *tree_path*.

```
def get_elements(self, tree_path: TreePath) -> List[Tree]
```

Метод возвращает список поддеревьев, которые удовлетворяют пути *tree_path*.

```
def get_iter(self)
```

Переопределение базового метода. Возвращает итератор по всем элементам дерева.

TreePath

Интерфейс путей в дереве. Задаёт основные операции, которые можно производить с путями:

```
def get_relative_path(self, tree_path) -> TreePath
```

Метод возвращает путь относительно пути *tree_path*. Необходима поддержка лишь случая, когда *tree_path* является префиксом текущего пути. Без обобщения на пути вида «../..».

```
def get_common_prefix(self, tree_path, in_block=False) -> TreePath
```

Метод возвращает наибольший общий префикс путей. Если флаг *in_block* выставлен, то для путей производится их обобщение по индексам. Иначе определяется точное совпадение.

```
def len(self) -> int
```

Метод возвращает длину пути – целое число.

```
def drop_for_len(self, len) -> TreePath
```

Обрезание пути до заданной длины.

```
def concat(self, tree_path) -> TreePath
```

Конкатенация путей. Входной параметр должен содержать относительный путь.

Markup

Интерфейс разметки. Необходимо определить три поля и один метод:

```
file
```

Относительный путь до файла, которому соответствует разметка.

```
type
```

Тип разметки, поддерживаемый классом *MarkupTypeRegistry*.

```
components
```

Список элементов разметки, определённых в заданном файле.

```
def add(self, component)
```

Метод добавления новой компоненты в список *components*.

Algorithm

Интерфейс основного алгоритма. Задаёт два метода, которые определяют схему взаимодействия с ним:

```
def learn(self, markup_list: List[Markup])
```

Метод обучения парсера на списке разметок. На вход принимает список разметок одного типа.

```
def parse(self, raw_page: str) -> ParserResult:
```

Метод для парсинга страницы. Следует вызывать после вызова обучения парсера. Возвращает элемент типа *ParserResult*.

Глава 2. Реализации