

SEP	06
Title	Extractors
Author	Ismael Carnales and a bunch of rabid mice
Created	2009-07-28
Status	Obsolete (discarded)

SEP-006: Rename of Selectors to Extractors

This SEP proposes a more meaningful naming of XPathSelectors or "Selectors" and their `x` method.

Motivation

When you use Selectors in Scrapy, your final goal is to "extract" the data that you've selected, as the [\https://docs.scrapy.org/en/latest/topics/selectors.html XPath Selectors documentation] says (bolding by me):

When you're scraping web pages, the most common task you need to perform is to **extract** data from the HTML source.

Scrapy comes with its own mechanism for **extracting** data. They're called `XPath` selectors (or just "selectors", for short) because they "select" certain parts of the HTML document specified by `XPath` expressions.

To actually **extract** the textual data you must call the selector `extract()` method, as follows

Selectors also have a `re()` method for **extracting** data using regular expressions.

For example, suppose you want to **extract** all `<p>` elements inside `<div>` elements. First you get would get all `<div>` elements

Rationale

As and there is no `Extractor` object in Scrapy and what you want to finally perform with `Selectors` is extracting data, we propose the renaming of `Selectors` to `Extractors`. (In Scrapy for extracting you use selectors is really weird :))

Additional changes

As the name of the method for performing selection (the `x` method) is not descriptive nor mnemonic enough and clearly clashes with `extract` method (`x` sounds like a short for extract in english), we propose to rename it to `select`, `sel` (is shortness if required), or `xpath` after [lxml's](http://lxml.de/xpathxslt.html) `xpath` method.

Bonus (ItemBuilder)

After this renaming we propose also renaming `ItemBuilder` to `ItemExtractor`, because the `ItemBuilder/Extractor` will act as a bridge between a set of `Extractors` and an `Item` and because it will literally "extract" an item from a webpage or set of pages.

References

1. XPath Selectors (<https://docs.scrapy.org/topics/selectors.html>)
2. XPath and XSLT with lxml (<http://lxml.de/xpathxslt.html>)