

SEP	7
Title	ItemLoader processors library
Author	Ismael Carnales
Created	2009-08-10
Status	Draft

## SEP-007: ItemLoader processors library

This SEP proposes a library of `ItemLoader` processor to ship with Scrapy.

### date.py

#### `to_date`

Converts a date string to a YYYY-MM-DD one suitable for `DateField`

**Decision:** Obsolete. `DateField` doesn't exists anymore.

### extraction.py

#### `extract`

This adaptor tries to extract data from the given locations. Any `XPathSelector` in it will be extracted, and any other data will be added as-is to the result.

**Decision:** Obsolete. Functionality included in `XpathLoader`.

#### `ExtractImageLinks`

This adaptor may receive either `XPathSelectors` pointing to the desired locations for finding image urls, or just a list of `XPath` expressions (which will be turned into selectors anyway).

**Decision:** XXX

### markup.py

#### `remove_tags`

Factory that returns an adaptor for removing each tag in the `tags` parameter found in the given value. If no `tags` are specified, all of them are removed.

**Decision:** XXX

#### `remove_root`

This adaptor removes the root tag of the given string/unicode, if it's found.

**Decision:** XXX

#### `replace_escape`

Factory that returns an adaptor for removing/replacing each escape character in the `wich_ones` parameter found in the given value.

**Decision:** XXX

#### `unquote`

This factory returns an adaptor that receives a string or unicode, removes all of the CDATAs and entities (except the ones in CDATAs, and the ones you specify in the `keep` parameter) and then, returns a new string or unicode.

**Decision:** XXX

### misc.py

#### `to_unicode`

Receives a string and converts it to unicode using the given encoding (if specified, else utf-8 is used) and returns a new unicode object. E.g:

```
>> to_unicode('it costs 20\xe2\x82\xac, or 30\xc2\xa3')
[u'it costs 20\u20ac, or 30\xa3']
```

**Decision:** XXX

### **clean\_spaces**

Converts multispaces into single spaces for the given string. E.g:

```
>> clean_spaces(u'Hello  sir')
u'Hello sir'
```

**Decision:** XXX

### **drop\_empty**

Removes any index that evaluates to None from the provided iterable. E.g:

```
>> drop_empty([0, 'this', None, 'is', False, 'an example'])
['this', 'is', 'an example']
```

**Decision:** Obsolete. Functionality included in reducers.

### **delist**

This factory returns and adaptor that joins an iterable with the specified delimiter.

**Decision:** Obsolete. Functionality included in reducers.

### **Regex**

This adaptor must receive either a list of strings or an XPathSelector and return a new list with the matches of the given strings with the given regular expression (which is passed by a keyword argument, and is mandatory for this adaptor).

**Decision:** XXX