TidyMarkup

TidyMarkup

TidyMarkup is a Data Format that uses the TagSoup to tidy up HTML. It can be used to parse ugly HTML and return it as pretty wellformed HTML.

> ✅ **Camel eats our own -dog food- soap**
>
> We had some issues in our pdf Manual where we had some strange symbols. So Jonathan used this data format to tidy up the wiki html pages that are used as base for rendering the pdf manuals. And then the mysterious symbols vanished.

TidyMarkup only supports the **unmarshal** operation as we really don't want to turn well formed HTML into ugly HTML 🙂

Java DSL Example

An example where the consumer provides some HTML

```
from("file://site/inbox").unmarshal().tidyMarkup().to("file://site/blogs");
```

Spring XML Example

The following example shows how to use TidyMarkup to unmarshal using Spring

```
<camelContext id="camel" xmlns="http://camel.apache.org/schema/spring">
  <route>
    <from uri="file://site/inbox"/>
    <unmarshal>
      <tidyMarkup/>
    </unmarshal>
    <to uri="file://site/blogs"/>
  </route>
</camelContext>
```

Dependencies

To use TidyMarkup in your camel routes you need to add the a dependency on **camel-tagsoup** which implements this data format.

If you use maven you could just add the following to your pom.xml, substituting the version number for the latest & greatest release (see the download page for the latest versions).

```
<dependency>
  <groupId>org.apache.camel</groupId>
  <artifactId>camel-tagsoup</artifactId>
  <version>x.x.x</version>
</dependency>
```