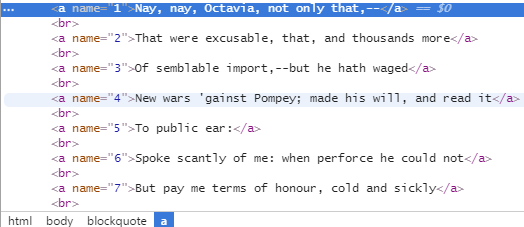
<http://www.jb51.net/article/60173.htm>

<http://kongtianyi.cn/2016/09/27/python/Scrapy-Lesson-2/>



<http://www.cnblogs.com/fnng/p/3576154.html>

爬图片

<http://blog.csdn.net/pleasecallmewhy/article/details/8929576>

正则表达式

正则表达式取网站

网站的分类

取内容 分组

<http://www.jb51.net/article/43999.htm>

结点

<http://www.runoob.com/xpath/xpath-syntax.html>

* Scrapy

['"news.html"',

"allswell/index.html"', '"asyoulikeit/index.html"', '"comedy\_er

rors/index.html"', '"cymbeline/index.html"', '"lll/index.html"', '"measure/index

.html"', '"merry\_wives/index.html"', '"merchant/index.html"', '"midsummer/index.

html"', '"much\_ado/index.html"', '"pericles/index.html"', '"taming\_shrew/index.h

tml"', '"tempest/index.html"', '"troilus\_cressida/index.html"', '"twelfth\_night/

index.html"', '"two\_gentlemen/index.html"', '"winters\_tale/index.html"',

'"1henr

yiv/index.html"', '"2henryiv/index.html"', '"henryv/index.html"', '"1henryvi/ind

ex.html"', '"2henryvi/index.html"', '"3henryvi/index.html"', '"henryviii/index.h

tml"', '"john/index.html"', '"richardii/index.html"', '"richardiii/index.html"',

'"cleopatra/index.html"', '"coriolanus/index.html"', '"hamlet/index.html"', '"j

ulius\_caesar/index.html"', '"lear/index.html"', '"macbeth/index.html"', '"othell

o/index.html"', '"romeo\_juliet/index.html"', '"timon/index.html"', '"titus/index

.html"', '"Poetry/sonnets.html"', '"Poetry/LoversComplaint.html"', '"Poetry/Rape

OfLucrece.html"', '"Poetry/VenusAndAdonis.html"', '"Poetry/elegy.html"']

<http://www.jb51.net/article/60173.htm>

<http://kongtianyi.cn/2016/09/27/python/Scrapy-Lesson-2/>

<http://www.cnblogs.com/fnng/p/3576154.html>

<http://blog.csdn.net/pleasecallmewhy/article/details/8929576>

<http://www.jb51.net/article/43999.htm>

<http://www.runoob.com/xpath/xpath-syntax.html>

<http://www.cnblogs.com/pzxbc/archive/2012/03/18/2404715.html>

《Natural Language Processing With Python》 Steven Bird，Ewan Klein，Edward Loper