ThreatFactor NSIA - Feature #307

Support Parsing of Content with jQuery-like Syntax

01/07/2011 08:25 AM - Luke Murphey

Status:	Closed		**Start date:**	01/12/2011
Priority:	Normal		**Due date:**	01/15/2011
Assignee:	Luke Murphey		**% Done:**	100%
Category:	Scan Engine		**Estimated time:**	0.00 hour
Target version:				

Description

Parsing content with jQuery-like syntax would make it easy to parse content. See
http://jsoup.org/cookbook/extracting-data/selector-syntax

History

#1 - 01/07/2011 08:28 AM - Luke Murphey

- *Assignee set to Luke Murphey*

- *Target version set to 1.0.3*

#2 - 01/07/2011 08:28 AM - Luke Murphey

- *Category set to Scan Engine*

#3 - 01/12/2011 07:31 AM - Luke Murphey

- *Due date set to 01/15/2011*

- *Start date set to 01/12/2011*

- *% Done changed from 0 to 40*

JSoup has been added. However. the parsed content is not stored which means that definitions will have to reparse the content each time (slowing down definitions). An alternative would be to:

1. Cache the jsoup document in HttpResponseData
2. Replace the older HTML parsing library completely for jsoup

#4 - 02/12/2011 07:51 AM - Luke Murphey

Two options are available for this:

- Could add the ability to store items about the specimen in the specimen itself (like decorations)
- Could use the ability provide in r1157 to store variables for the current URL evaluation (this would have to be done by convention, not something enforced by the scan engine)

#5 - 02/19/2011 11:20 PM - Luke Murphey

- *Target version deleted (1.0.3)*

#6 - 05/28/2011 10:11 PM - Luke Murphey

- *Status changed from New to Closed*

- *% Done changed from 40 to 100*

Having all of the content pre-processed should not be be necessary.