

Stalker vs. XWRAP

Comparison

Stefan Schönig

Marco Schönig

XWRAP Features

- supervised extraction uses 4 steps
 - data is cleaned from syntactical incorrectness
 - XWRAP parses the document
 - tree is parsed into a CSV file
 - XML templates are used to get XML output

STALKER Features

- Supervised extraction to markup content
 - Initial SLA candidates are being created
 - LearnDisjunct() creates set of refining terminals
 - Refinement
 - » Topology
and
 - » Landmark
 - If BestDisjunct = true -> return Extraction Rules

Differences

- Computing
 - STALKER uses “statemachines” to describe content
 - XWRAP uses a Tree to describe the content
- Output
 - STALKER generates Rules that a Wrapper can use to extract Data from a Document
 - XWRAP returns a XML file