Comparison between two papers

A Flexible Learning System for Extracting Data From Tables and Lists in HTML documents

&

A Gateway From HTML to XML
Initial Comparison

- WL²: Wrapper learning system
- Gateway: stand-alone algorithm, outputs XML document extracted from HTML site
Extraction possibilities

- WL²: exploits several different representations of a document
- Gateway: uses DOM structure and layout elements from given document
Knowledge-based Systems

• WL²: based on KBS,
• Gateway: –
Structures

- WL²: doesn't use structures as output possibilities
- Gateway: returns XML document retrieved by transformation of HTML–DOM
Table transformation

- WL²: geometric relationship between data items, special builders
- Gateway: transforms multi-dimensional tables into one-dimensional rows
Software Products

• WL²: WhizBang Labs., software solutions at Inxight Inc.
• Gateway: available as web application at http://tao.my-net-space.net/h2x