

Web Information Extraction

Automatic Data Extraction from Lists and Tables
in Web Sources

(Lerman, Knoblock, Minton)

vs.

Mining web pages for data records

(Liu, Grossman, Zhai)

Comparison

MDR

ADE

Comparison

MDR

- static

ADE

- unsupervised learning

Comparison

MDR

- static
- standalone system

ADE

- unsupervised learning
- DataPro, AutoClass, ALERGIA

Comparison

MDR

- static
- standalone system
- HTML tags

ADE

- unsupervised learning
- DataPro, AutoClass, ALERGIA
- HTML and text

Comparison

MDR

- static
- standalone system
- HTML tags
- every table-related region

ADE

- unsupervised learning
- DataPro, AutoClass, ALERGIA
- HTML and text
- only dynamic data regions

Comparison

MDR

- static
- standalone system
- HTML tags
- every table-related region
- handles non-contiguous streams

ADE

- unsupervised learning
- DataPro, AutoClass, ALERGIA
- HTML and text
- only dynamic data regions
- requires contiguous regions

Thank you for your attention.

Paul Bohunsky

0025058

Marian Schedenig

9725416