

# **CONCLUSION**

record boundaries

vs.

automatic segmentation

# SIMILARITIES

- unsupervised approach
  - no training, no user-input
  - can work automatically
- depend on special page / site structure
  - which ones depends on approach
  - let`s look at the differences

# DIFFERENCES

## record boundary

- plain lists
- single page
- 5 independent heuristics combined
- only find boundary

## automatic segmentation

- list / details page
- multiple page, generate template
- two different heuristics
- assumes about record structure

# AT A GLANCE

## Comparison of the Segmentation Algorithms:

<b>Algorithm</b>	<i>un-supervised</i>	<i>domain-independent</i>	<i>needs list-and detail-pages</i>	<i>infers Columns</i>	<i>uses tag-tree</i>	<i>allows inconsistencies in content</i>	<i>uses template finding</i>
5 Heuristics (#5)	X	1,2,3,4			X		
CSP (#6)	X	X	X				X
HMM (#6)	X	X	X	X		X	X

- both are useful and can coexist
- hardly overlap
- choice depends on page / site structure

**THANK YOU!**

...have fun with the next  
team.