

AI- Logo Detection and PDF Handling

Opportunities in processing e-paper PDFs

DataScan's way to extract articles from e-papers: eXpaperClip

Analyzing newspaper layouts – segmentating articles – releasing and pasting – collecting metadata – automating processes

Speaker:
Robert Fischer

Company:
DataScan Computersysteme

www.datascan.de



 WORLD MEDIA INTELLIGENCE
CONGRESS
BERLIN _____ 4-6 October 2017

@_FIBEP
#FIBEP
#WMIC17



DataScan:

Technologies for article extraction and press review production

newsWEB eXpaperClip - newsWEB press review editorial software

USPs

- World's changing From e-paper input and online sources to mobile output
- Reading clients' needs Flexibility in configuration & production. Your competitive advantage
- Faster and faster ! Client-/server separation in compiling/producing press reviews
- Avoiding Support Professional press clipping systems
- Saving money Parallel production of recipient specific editions
- We hate workarounds! Usability & automating processes



 **WORLD MEDIA INTELLIGENCE CONGRESS**
BERLIN _____ 4-6 October 2017

@_FIBEP
#FIBEP
#WMIC17



DataScan:

Technologies for article extraction and press review production

Scans + e-paper + online + database input \Rightarrow One format: XML \Rightarrow Multiple output

How to ??

- | | |
|---------------------|--|
| - No imaging on PDF | Direct search in and extract articles out of e-paper PDF without creating images |
| - No PDF | Using XML format for storage |
| - No disadvantage | Fast in building client setups |
| - No web production | Flexible in programming and configuration |
| - No speed limit | Fast and reliable in production |
| - No product limit | Creating output in different formats |
| - No errors | Automating recurring processes |
| - No ticketing | Instant & qualified support |

DataScan:

Technologies for article extraction and press review production

The challenge: Cutting out content from e-paper PDFs

newsWEB eXpaperClip

- E-paper instead of newspapers
- Indexing PDF text content for fulltext search
- Using PDF features for better segmentation
- Verification of segmentation results
- Cutting or clipping articles
- Keeping the original layout
- Different workflow options
- eXpaper distribution service ... into your production system

DataScan



WMIC
BERLIN
2017