ARTICLE SEGMENTATION CHALLENGES AND NEW POSSIBILITIES

HOW ZISSOR USE TECHNOLOGY TO OVERCOME THE UNPREDICTABLE CHANGES IN PAGE LAYOUT

Ove Dirdal

ZISSOR

WHY CORRECT ARTICLE SEGMENTATION ARE IMPORTANT

▸ The quality in your offering and delivery to your clients depends on correct data and content

▸ Correct Article segmentation is essential for all your deliveries

▸ Correct segmented articles with its content and data are required to build up your data containers with quality data for analyzes and most of your other deliveries

THE LAYOUT IN THE PHYSICAL NEWSPAPER AND E-PAPER ARE GETTING MORE AND MORE COMPLEX AN CORRECT ARTICLE SEGMENTATION ARE IMPORTANT

▸ Publishers use more complex and creative layout to attract their readers

▸ The page structure is not standardized and not always logical

▸ More pictures - Text in pictures and text in color - Articles cross multipages – no standard headlinelayout

▸ The layout differ more and more from publication to publication

▸ New technologies and new ways of producing newspapers give less boundaries and restrictions for the publishers

▸ More manual work is required when extracting the articles

HOW WILL ZISSOR MEET THESE CHALLENGES

▸ Zissor already have a very advanced segmentation module based on sophisticated layout module and advanced mathematic algorithms for automatic finding the article structures.

▸ We are now adding more text analytic functionality plus more images classifications functionality, together with AI technics to our segmentation module

▸ Machine Learning will in addition be used to constantly update each AI model

▸ Introduction of more advanced functionality and AI will help us meeting the new layout challenges and requirements for more accurate content, and to reduce the need for rule based definitions and manual control.

CONCLUSIONS

▸ The automated segmentation module is one main component of the Zissor Media System and our focus is to keep it forefront to any possible complex page layouts that might occur

▸ Introduction of new technologies like AI and text analytic tools will make it easier for us to cope with the continuous ongoing unpredictable layout changes made by the publishers

We are now also introducing the segmentation module as a separate module to be implemented with or without the automated clipping module

You are welcome to come by us to have a more detailed look at our solutions