Chapter 63
From Image to XML:
Monitoring a Page Layout Analysis Approach for the Visually Impaired

Robert Keefer
Pomiet, LLC, USA

Nikolaos Bourbakis
Wright State University, USA

ABSTRACT

Page layout analysis and the creation of an XML document from a document image are useful for many applications including the preservation of archived documents, robust electronic access to printed documents, and access to print materials by the visually impaired. In this paper, the authors describe a document image process pipeline comprised of techniques for the identification of article headings and the related body text, the aggregation of the body text with the headings, and the creation of an XML document. The pipeline was developed to support multiple document images captured by the head-mounted cameras of a reading device for the visually impaired. Both automatic and manual adaptations of the pipeline processed a sample of 25 newspaper document images. By comparing the automatic and manual processes, we show that overall our approach generates high-quality XML encoded documents for use in further processing, such as a text-to-speech for the visually impaired.

1. INTRODUCTION

Page layout analysis and the creation of an XML document from a document image are useful for many applications including the preservation of archived documents (Wang, et al., 2009) and accessibility by those with visual impairments. TYFLOS (Keefer, et al., 2009a,b) is a prototype wearable mobile reading device for the visually impaired. TYFLOS is equipped with two web cameras mounted into a pair of glasses and the software for performing document image rectification and segmentation. Traditional document image analysis techniques play an important role in the operation of the TYFLOS prototype, including document image capture, binarization, page perspective correction in 3-dimensions, page curl correction, and page segmentation. In this paper

we describe techniques for headline identification, page segment aggregation, and the creation of an XML document from the document image. The XML document supports various forms of interaction with the text of the document, including a voice user interface (Keefer, et al., 2013).

Much work has been performed to identify headlines within web sites and document images. This work has been in the context of both improving access to documents for the visually impaired, as well as the digital access of archived documents. For example, Brudvick, et al. (2008) have developed a method to predict whether web page content semantically functions as a headline by considering the visual features of text when rendered in a browser. Similarly, Kohlschütter, et al. (2010) describe a method for identifying text elements within a web page.

Document segmentation has been of interest to the document image processing community for many years. O'Gorman's (1993) Docstrum method offered an original and well organized analysis of document layout analysis based on K-nearest neighbors to identify connected components and from these to identify regions of text. Akram et al. (2010) offer a review on the way to process a document and generally segment the layout area. In another approach, Winder et al. (2011) describe a method for page segmentation based on an analysis of the Voronoi zones of a histogram of the connected component heights of image segments. Similarly, Breuel et al. (2011) also patented a method for document image layout deconstruction. Finally, Ferilli, et al. (2011) apply supervised machine learning techniques to document image layout analysis.

For the purposes of supporting robust interaction with document images converted to XML, Ishitani (2003) proposed a method for transforming a document image into XML. This method extracts document elements such as title, headings, and body text from a document image. The hierarchical structure of the document is also extracted and described by a document object model (DOM).

The XML document is created through a set of transforms applied to the extracted document elements and the DOM.

WISDOM++ (Altamura, et al., 2001) is a document processing system that performs document analysis, classification, and text transformation to generate an XML document from a document image. Agrawal and Doermann (2010) also discuss a method for page segmentation that produces GEDI XML files.

To create an XML document from the document image, a document image segmentation method must separate images from text, identify headings within the document image, and identify article content within the document image. The methods described in (Ishitani, 2003), (Altamura, et al., 2001), (Agrawal and Doermann, 2010), (Antonacopoulos and Karatzas, 2004), and (Pletschacher and Antonacopoulos, 2010) all rely on robust document analysis methods to identify the structure and format of the document image, followed by an OCR step to convert the text within a segment to XML.

The method described here also relies on robust document image analysis in order to identify the structure of the document. However, an important feature of TYFLOS is that it guides the user in positioning the device, which facilitates the capture of an improved image. Thus, more than one image of the document may be captured and the guidance system of the device must track the location of each image. By utilizing this positional information, the text from multiple images can be combined to improve the OCR of document images.

Figure1 illustrates the TYFLOS document image processing pipeline. The first step in this pipeline is binarization, the process that transforms a color image into a black and white image. This step is followed by an image enhancement technique that removes artifacts in the high-resolution image and improves the OCR of the document image. (Keefer, et al., 2009a) The document image captured by a camera may suffer from both per-

Related Content

Text Preprocessing: A Tool of Information Visualization and Digital Humanities
Piotr Malak (2018). *Information Visualization Techniques in the Social Sciences and Humanities (pp. 86-104).*
www.irma-international.org/chapter/text-preprocessing/201305

Comparative Study of One Stage V/S Two Stage Surgery for Intra-Articular Distal Tibial Pilon Fracture
S. Patil Nitinand Paresh Patil (2023). *Recent Developments in Machine and Human Intelligence (pp. 108-118).*
www.irma-international.org/chapter/comparative-study-of-one-stage-vs-two-stage-surgery-for-intra-articular-distal-tibial-pilon-fracture/330323

Measuring Mobile Phone Technology Adoption in SMEs: Analysis of Metrics
Renatus Mushi, Said Jafariand Almar Ennis (2018). *Technology Adoption and Social Issues: Concepts, Methodologies, Tools, and Applications (pp. 1546-1557).*
www.irma-international.org/chapter/measuring-mobile-phone-technology-adoption-in-smes/196744

Technology Satisfaction in an Academic Context: Moderating Effect of Gender
A. Y. M. Atiquil Islam (2017). *Research Paradigms and Contemporary Perspectives on Human-Technology Interaction (pp. 187-211).*
www.irma-international.org/chapter/technology-satisfaction-in-an-academic-context/176116

Student Perspectives on Distraction and Engagement in the Synchronous Remote Classroom
Noah Q. Cowitand Lecia J. Barker (2022). *Digital Distractions in the College Classroom (pp. 243-266).*
www.irma-international.org/chapter/student-perspectives-on-distraction-and-engagement-in-the-synchronous-remote-classroom/296135