User Manual

New Map

New Subtitle

Contents

ICDAR 2013 Table Competition

Abstract—Table understanding is a well studied problem in document analysis, and many academic and commercial approaches have been developed to recognize tables in several document formats, including plain text, scanned page images and born-digital, object-based formats such as PDF. Despite the abundance of these techniques, an objective comparison of their performance is still missing. The Table Competition held in the context of ICDAR 2013 is our first attempt at objectively evaluating these techniques against each other in a standardized way, across several input formats. The competition independently addresses three problems: (i) table location, (ii) table structure recognition, and (iii) these two tasks combined. We received results from seven academic systems, which we have also compared against four commercial products. This paper presents our findings

The problem of analysing tables in documents has been dealt with in many academic publications over the previous two decades. The systems reported in the literature vary according to the input document format (ASCII text, HTML, image, PDF) and the type of structure that is recognized and output by the system. Table understanding has been gaining traction since the beginning of the big data era due to the massive amounts of tabular data in documents on the Web. In addition, private, public and governmental institutions often publish reports in PDF format. Such data is immensely valuable for decision support and object search, but without reliable table understanding techniques, it cannot be easily indexed by search engines or used by automatic data processing applications. In our previous work [1], we proposed a methodology for evaluating these approaches independently from the format of the input and of the output. In particular, the problem of table understanding was split up into three tasks:

Index (Fig.)

Footnotes

n/a

Index