

Round-Trip PDF Solutions

Dr. Tamir Hassan – Your Expert in Digital Documents

≡ Menu

ICDAR 2013 Table Competition

This page is archived from the competition website as of 2013. The competition datasets, complete with ground truth, are available [here](#).

The problems of table detection and table structure recognition in documents have attracted much interest not only from the document analysis community, but also from the database and information extraction (IE) communities. Whereas traditional methods have worked on scanned images of pages [2], a number of approaches from related fields have used natively-digital document formats [3, 4, 5]. The choice of born-digital PDF as the format for this competition is designed to unify both types of approaches, and the objective of this competition is to measure the state of the art and compare the relative advantages of both types of approach.

This competition is split up into two sub-competitions, **table detection** or **location** and **table structure recognition**. Entrants may choose to enter either one, or both, sub-competitions. In a previous publication [1], we proposed document-generic models for both of these procedures and generated a publicly available, ground-truthed [dataset](#). The format for the ground truth, as well as methods for numerically comparing an algorithm's result to the ground truth, are also described in this paper.

This dataset serves as the example dataset, against which all entries can be tested before submission. Entries will be evaluated using the methods described in the paper. Our paper also describes the **functional analysis** of tables; this does not form part of the present competition.

Table location sub-competition

In this sub-competition, entrants will be required to return the rectangular bounding-box coordinates of all tabular regions in the competition dataset. As with the example dataset, the competition dataset will be checked thoroughly to exclude documents where the

existence or boundary of any table is ambiguous, or where the smallest enclosing bounding box also includes content that is not part of the table.

A more detailed description of the submission format can be found [here](#).

Table structure recognition sub-competition

The aim of the table structure recognition sub-competition is to compare methods for determining the cell structure of tables *given correct information about their location*. It is therefore possible to participate only in this sub-competition, and not in the table location sub-competition. We strongly recommend entrants to use manually generated or corrected input regarding the table locations when generating their results, in order to avoid being unnecessarily penalized.

A more detailed description of the submission format can be found [here](#).

Submission procedure

Participants are requested to [register by email](#) by 31 March 2013, stating which one, or both, of the sub-competitions they will be taking part in. The competition dataset can be downloaded [here](#) [NB: see the [Downloads](#) page for the archive with ground truth].

Participants will have two weeks to run their system on this dataset and send us the results of their system in the same XML format and file structure as the ground-truth, along with a brief description of their system. The **new** deadline for the submission of results (by email) is **15 April 2013**.

Examples of the models for table detection and table structure recognition can be found [here](#). XML Schema definitions are now available for the [table detection](#) and [table structure recognition](#) entries. Submissions should be validated against these definitions. will be included shortly.

References

- [1] Göbel, M., Hassan, T., Oro, E., Orsi, G.: [A Methodology for Evaluating Algorithms for Table Understanding in PDF Documents](#), DocEng 2012
- [2] Shahab, A., Shafait, F., Kieninger, T., and Dengel, A.: An open approach towards the benchmarking of table structure recognition systems, DAS 2010
- [3] Oro, E., Ruffolo, M.: PDF-TREX: An approach for recognizing and extracting tables from PDF documents, ICDAR 2009
- [4] Silva, A.C.: Learning rich hidden markov models in document analysis: Table location, ICDAR 2009

[5] Krüpl, B., Herzog, M.: Visually guided bottom-up table detection and segmentation in web documents, WWW 2006

Competition organizers: [Max Göbel](#), [Tamir Hassan](#), [Ermelinda Oro](#) and [Giorgio Orsi](#).

[Return to top](#)

© 2019 Tamir Hassan • [Contact](#) • [Legal Notices \(Impressum\)](#)