	Per-document averages			Tables found (total=156)	
Participant	Recall	Precision	F_1-meas.	Complete	Pure
FineReader	*0.9971*	*0.9729*	***0.9848***	*142*	*148*
OmniPage	*0.9644*	*0.9569*	***0.9606***	*141*	*130*
Silva	0.9831	0.9292	**0.9554**	149	137
Nitro	*0.9323*	*0.9397*	***0.9360***	*124*	*144*
Nurminen	0.9077	0.9210	**0.9143**	114	151
Acrobat	*0.8738*	*0.9365*	***0.9040***	*110*	*141*
Yildiz	0.8530	0.6399	**0.7313**	100	94
Stoffel	0.6991	0.7536	**0.7253**	79	66
Liu et al. 2	0.3355	0.8836	**0.4864**	0	29
Hsu et al.	0.4601	0.3666	**0.4080**	39	95
Fang et al.	0.2697	0.7496	**0.3967**	28	41
Liu et al. 1	0.2207	0.8885	**0.3536**	0	25

TABLE II
RESULTS FOR THE TABLE LOCATION (LOC) SUB-COMPETITION

	Per-document averages		
Participant	Recall	Precision	F_1-measure
Nurminen	0.9409	0.9512	**0.9460**
Silva	0.6401	0.6144	**0.6270**
Hsu et al.	0.4811	0.5704	**0.5220**

TABLE III
RESULTS FOR THE TABLE STRUCTURE RECOGNITION (STR) SUB-COMPETITION (BASED ON CORRECT REGION INFORMATION)

	Per-document averages		
Participant	Recall	Precision	F_1-measure
FineReader	*0.8835*	*0.8710*	***0.8772***
OmniPage	*0.8380*	*0.8460*	***0.8420***
Nurminen	0.8078	0.8693	**0.8374**
Acrobat	*0.7262*	*0.8159*	***0.7685***
Nitro	*0.6793*	*0.8459*	***0.7535***
Silva	0.7052	0.6874	**0.6962**
Yildiz	0.5951	0.5752	**0.5850**

TABLE IV
TABLE STRUCTURE RECOGNITION RESULTS FOR THE COMPLETE PROCESS (COM) - BASED ON THE SYSTEM'S TABLE LOCATION RESULT

Our evaluation metrics were found to be a fair representation of the actual quality of the output from the various systems. The combination of completeness and purity with precision and recall on the character level gives a good overall picture of the region detection quality. Similarly, we have found that using cell adjacency relations to evaluate table structure detection enables us to obtain precision and recall measures which are repeatable and accurately reflect the quality of the result.

By calculating the results for each document first, we were able to reduce the bias of "data-heavy" tables on the overall result. A further improvement for the future would be to evaluate regions by calculating the area (in square points) of region overlap instead of counting characters, after "normalizing" each region first by shrinking it to the smallest region encompassing all characters within its bounds. This would avoid regions containing overprinted or non-printing characters skewing the result.

The structure results for the complete process (see Table IV) should also be treated with some caution. A number of systems

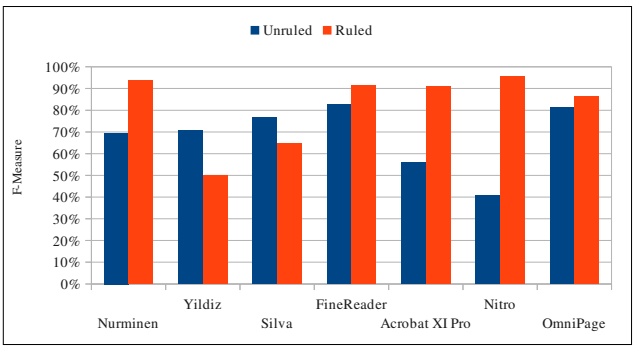

Fig. 1. Comparison of results with ruled versus unruled tables for the complete process sub-competition

returned large false positive regions, whose table structure consisted of only one cell. In many cases, this huge cell only neighboured one or two other cells, and therefore did not raise the overall false positive count significantly.

A further issue with our structure recognition metric is in the comparison of adjacency relations by their textual content. Although our normalization routine stripped or replaced most special characters, there were still some remaining encoding issues when evaluating certain approaches. This is a double-edged sword, as removing all non-alphanumeric characters would make it no longer possible to distinguish between cells that do not contain at least one letter or number, of which there were many in our dataset. In the future, we will therefore consider requiring further information about the cell, such as a bounding box, to enable its unique identification.

ACKNOWLEDGMENTS

This work has been supported by the EU FP7 Marie Curie Zukunftskolleg Incoming Fellowship Programme, University of Konstanz (grant no. 291784), the ERC grant agreement DIADEM (no. 246858) and by the Oxford Martin School (grant no. LC0910-019).

REFERENCES

[1] M. C. Göbel, T. Hassan, E. Oro, and G. Orsi, "A methodology for evaluating algorithms for table understanding in PDF documents," in *ACM Symposium on Document Engineering*, 2012, pp. 45–48.
[2] E. Oro and M. Ruffolo, "PDF-TREX: An approach for recognizing and extracting tables from PDF documents," in *Proc. of ICDAR*, 2009, pp. 906–910.
[3] B. Krüpl and M. Herzog, "Visually guided bottom-up table detection and segmentation in web documents," in *WWW*, 2006, pp. 933–934.
[4] A. C. e Silva, "Metrics for evaluating performance in document analysis: application to tables," *IJDAR*, vol. 14, no. 1, pp. 101–109, 2011.
[5] J. Fang, L. Gao, K. Bai, R. Qiu, X. Tao, and Z. Tang, "A table detection method for multipage PDF documents via visual seperators and tabular structures," in *ICDAR*, 2011, pp. 779–783.
[6] Y. Liu, K. Bai, P. Mitra, and C. L. Giles, "TableSeer: automatic table metadata extraction and searching in digital libraries," in *JCDL*, 2007, pp. 91–100.
[7] B. Yildiz, K. Kaiser, and S. Miksch, "pdf2table: A method to extract table information from pdf files," in *IICAI*, 2005, pp. 1773–1785.
[8] A. C. e Silva, "Parts that add up to a whole: a framework for the analysis of tables," Ph.D. dissertation, The University of Edinburgh, 2010.
[9] H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, D. A. Keim, and O. Deussen, "Document cards: A top trumps visualization for documents," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 1145–1152, 2009.
[10] A. Stoffel, D. Spretke, H. Kinnemann, and D. A. Keim, "Enhancing document structure analysis using visual analytics," in *SAC*, 2010, pp. 8–12.