Reading Reference Lists with Vision Models

A. Sample, B. Example · Demo Institute of Documents

Abstract. We compare template-based and learned approaches to pulling structured references out of scholarly PDFs, where layouts and citation styles vary from paper to paper.

1 Introduction

Deep networks now dominate document understanding (LeCun et al., 2015), and attention-based models [2] read long pages well. Earlier work framed extraction as a communication problem (Shannon, 1948). Residual networks [4] made very deep vision models trainable.

Dedicated tools exist: GROBID parses references with sequence labelling (Lopez, 2009), and CERMINE extracts article metadata and bibliographies (Tkaczyk et al., 2015). Mixed citation styles in one list, as here, are common in theses that combine sources.

References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

2. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 5998–6008.

3. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

4. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

5. Lopez, P. (2009). GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. In *Research and Advanced Technology for Digital Libraries* (pp. 473–474). Springer. https://doi.org/10.1007/978-3-642-04346-8_62

6. Tkaczyk, D., Szostek, P., Fedoryszak, M., Dendek, P. J., & Bolikowski, Ł. (2015). CERMINE: Automatic extraction of structured metadata from scientific literature. *International Journal on Document Analysis and Recognition, 18*(4), 317–335. https://doi.org/10.1007/s10032-015-0249-8