

## DAFTAR PUSTAKA

- Appelt, D.E., 1999. Introduction to information extraction. *Ai Communications*, 12(November), pp.161–172. Available at: <http://iospress.metapress.com/index/HQDWYFKJPM6B1G3A.pdf>.
- Bhumika, Sehra, S. & Nayyar, A., 2013. A Review Paper on Algorithms Used for Text Classification. *Ijaiem*, 2(3), pp.90–99. Available at: <http://ijaiem.org/Volume2Issue3/IJAIEM-2013-03-13-025.pdf>.
- Chengru, S., Shifeng, W. & Changshui, Z., 2013. Automatic Elements Extraction of Chinese Web News Using Prior Information of Content and Structure. *2013 2nd IAPR Asian Conference on Pattern Recognition*, pp.340–345. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6778337>.
- Cox, R., 2007. Regular Expression Matching Can Be Simple And Fast. Available at: <https://swtch.com/~rsc/regex/regex1.html>.
- Devid, S., Baizal, Z.K.A. & A.w, Y.F., 2010. *Ekstraksi Informasi Utama Halaman Web Berita Menggunakan Metode Hybrid*. Telkom University.
- Hu, Y., Li, H., Cao, Y., Meyerzon, D. & Zheng, Q., 2006. Automatic extraction of titles from general documents using machine learning. *Information Processing & Management*, 42(5), pp.1276–1293. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0306457305001639>.
- Ihlstrom, C. & Lundberg, J., 2003. The online news genre through the user perspective. *36th Hawaii International Conference on Systems Sciences*, p.10.
- Juwito, 2008. Menulis Berita Dan Feature ' S.
- Kohlschütter, C., Fankhauser, P. & Nejdl, W., 2010. Boilerplate Detection using Shallow Text Features. *Text*, pp.441–450. Available at: <http://portal.acm.org/citation.cfm?id=1718542>.
- Li, Y., X. Meng, Q. & Li, L.W., 2006. Hybrid Method for Automated News Content Extraction from the Web. *Proceedings of 7th Conference on Web Information System Engineering*, pp.327 – 338.
- Margianto, J.H. & Saefullah, A., 2012. *Media Online : Pembaca, Laba dan Etika. Business*.
- Moens, M.-F., 2006. Evaluation of Information Extraction. In *Information Extraction: Algorithms and Prospects in a Retrieval Context*. Springer Netherlands, pp. 179–197.
- NISO, 2004. Understanding Metadata. *National Information Standards*, (MD:NISO Press), p.20. Available at: <http://www.mendeley.com/catalog/understanding-metadata-3/%5Cnhttp://www.niso.org/publications/press/UnderstandingMetadata.pdf>.

- Putranto, D.D., 2008. Pengkajian masalah Longest Common Subsequence.
- Sano, D., 2015. Konsep dan Definisi Text Mining - Seri Text Mining dan Web Mining.
- Shirky, C., 2008. *Here Comes Here Comes Everybody*, Available at: <https://vpngate1.hrz.tu-chemnitz.de/+CSCO+0h756767633A2F2F767272726B63796265722E76727272E626574++/stamp/stamp.jsp?tp=&arnumber=4607925>.
- Siefkes, C. & Siniakov, P., 2005. An Overview and Classification of Adaptive Approaches to Information Extraction. *Lecture Notes in Computer Science*, 3730, pp.172–212. Available at: <http://www.springerlink.com/index/ppx0665435041w52.pdf>.
- Song, D., Sun, F. & Liao, L., 2013. A hybrid approach for content extraction with text density and visual importance of DOM nodes.
- Wagh, P.D., 2015. Self-directing Superimposed Text Detection and Removal from images using Inpainting for Region Filling with Smoothing. , pp.241–244.
- Weninger, T., Hsu, W.H. & Han, J., 2010. CETR - Content Extraction via Tag Ratios. , pp.971–980.
- Wikarsa, L. & Thahir, S.N., 2015. A text mining application of emotion classifications of Twitter's users using Na??ve Bayes method. *Proceeding of 2015 1st International Conference on Wireless and Telematics, ICWT 2015*.
- Wu, Y., 2016. Language independent web news extraction system based on text detection framework. , 342, pp.132–149.
- Xiao, L., Wissmann, D., Brown, M. & Jablonski, S., 2004. Information extraction from the web: System and techniques. , pp.195–224.