

## Daftar Pustaka

- Abdelali, A., Guzman, F., Sajjad, H. dan Vogel, S. (2014), The AMARA Corpus: Building parallel language resources for the educational domain, *in* 'Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland'.
- Abdul, C. (2009a), *Pengantar Semantik Bahasa Indonesia*, Rineka Cipta.
- Abdul, C. (2009b), *Sintaksis Bahasa Indonesia (Pendekatan Proses)*, Rineka Cipta.
- Abualkishik, A. dan Omar, K. (2013), Framework for translating the Holy Quran and its reciting rules to Braille code, *in* 'Research and Innovation in Information Systems (ICRIIS), 2013 International Conference on', IEEE, hlm. 380–385.
- Adhoni, Z. dan Al Hamad, H. (2014), An API for Quran portal using Drupal technology, *in* 'Applications of Digital Information and Web Technologies (ICADIWT), 2014 Fifth International Conference on the', hlm. 160–164.
- Agarwal, A., Baechle, C., Behara, R. dan Zhu, X. (2018), 'A Natural language processing framework for assessing hospital readmissions for patients with COPD', *IEEE journal of biomedical and health informatics* **22**(2), 588–596.
- Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yilmazer, H., Atasoy, G., Öz, S., Yildiz, I. dkk. (2012), Construction of the Turkish National Corpus (TNC)., *in* 'LREC', hlm. 3223–3227.
- Al-Gharaibeh, A., Al-Taani, A. dan Alsmadi, I. (2011), 'The Usage of Formal Methods in Quran Search System', *ICICS 2011*.
- Al-Yahya, M., Al-Khalifa, H., Bahanshal, A., Al-Odah, I. dan Al-Helwah, N. (2010), 'An Ontological Model for Representing Semantic Lexicons: An Application on Time Nouns in the Holy Quran', *Arabian Journal for Science and Engineering* **35**(2), 21.
- AlAhmad, M. A., Alshaikhli, I. dan Jumaah, B. (2013), Protection of the Digital Holy Quran Hash Digest by Using Cryptography Algorithms, *in* 'Proceedings

of the 2013 International Conference on Advanced Computer Science Applications and Technologies', ACSAT '13, IEEE Computer Society, Washington, DC, USA, hlm. 244–249.

URL: <http://dx.doi.org/10.1109/ACSAT.2013.55>

Alqahtani, M. dan Atwell, E. (2017), 'Evaluation Criteria for Computational Quran Search', *International Journal on Islamic Applications in Computer Science And Technology* 5(1), 12–22.

Amrullah, A. Z., Hartanto, R. dan Mustika, I. W. (2017), A comparison of different part-of-speech tagging technique for text in Bahasa Indonesia, in '2017 7th International Annual Engineering Seminar (InAES)', hlm. 1–5.

Aryoyudanta, B., Adji, T. B. dan Hidayah, I. (2016), Semi-supervised learning approach for Indonesian Named Entity Recognition (NER) using co-training algorithm, in 'Intelligent Technology and Its Applications (ISITIA), 2016 International Seminar on', IEEE, hlm. 7–12.

Atkins, S., Clear, J. dan Ostler, N. (1992), 'Corpus design criteria', *Literary and linguistic computing* 7(1), 1–16.

Baker, P., Hardie, A., McEnery, T., Cunningham, H. dan Gaizauskas, R. J. (2002), EMILLE, A 67-Million Word Corpus of Indic Languages: Data Collection, Mark-up and Harmonisation., in 'LREC'.

Bentivogli, L., Girardi, C. dan Pianta, E. (2003), The MEANING Italian Corpus, in 'Proceedings of the Corpus Linguistics 2003 conference', hlm. 103–112.

Bentivogli, L. dan Pianta, E. (2005), 'Exploiting parallel texts in the creation of multilingual semantically annotated resources: the MultiSemCor Corpus', *Natural Language Engineering* 11(3), 247–261.

Biber, D., Conrad, S. dan Reppen, R. (1998), *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge Approaches To Linguistics, Cambridge University Press.

Bowes, D., Hall, T. dan Gray, D. (2012), Comparing the performance of fault prediction models which report multiple performance measures: recomputing the confusion matrix, in 'Proceedings of the 8th International Conference on Predictive Models in Software Engineering', ACM, hlm. 109–118.

BPPT (2001), *Pengembangan sistem korpus nasional dan sistem ekstraksi informasi bahasa Indonesia: laporan akhir tahun anggaran 2001 proyek Iptekda BPPT.*, Pusat Pengkajian dan Penerapan Teknologi Informasi dan Elektronika, Deputi Teknologi Informasi, Energi, Material, dan Lingkungan, BPPT.

- Brill, E. (1992), A simple rule-based part of speech tagger, *in* 'Proceedings of the workshop on Speech and Natural Language', Association for Computational Linguistics, hlm. 112–116.
- Briscoe, T., Carroll, J. dan Watson, R. (2006), The second release of the RASP system, *in* 'Proceedings of the COLING/ACL on Interactive presentation sessions', Association for Computational Linguistics, hlm. 77–80.
- Burnard, L. (1995), 'Users Reference Guide British National Corpus Version 1.0'.
- Bustami, B., Fadlisya, F. dan Mauliza, I. (2017), 'Sistem Pendekteksi Kesalahan Dalam Membaca Al-Qur'an Ayat 1-5 Menggunakan Metode Viterbi', *TECHSI-Jurnal Teknik Informatika* **9**(1), 30–44.
- Chen, K.-J., Huang, C.-R., Chang, L.-P. dan Hsu, H.-L. (1996), 'Sinica corpus: Design methodology for balanced corpora', *Language* **167**, 176.
- Chen, T. dan Kan, M.-Y. (2013), 'Creating a live, public short message service corpus: The NUS SMS corpus', *Language Resources and Evaluation* **47**(2), 299–335.
- Chong, F., Carraro, G. dan Wolter, R. (2006), 'Multi-tenant data architecture', *MSDN Library, Microsoft Corporation* hlm. 14–30.
- Choudhary, N., Singh, R., Rao, V. A. dan Shrivastava, M. (2018), Twitter corpus of Resource-Scarce Languages for Sentiment Analysis and Multilingual Emoji Prediction, *in* 'Proceedings of the 27th International Conference on Computational Linguistics', hlm. 1570–1577.
- Christodoulides, G. (2018), 'Praaline: An Open-Source System for Managing, Annotating, Visualising and Analysing Speech Corpora', *Proceedings of ACL 2018, System Demonstrations* hlm. 111–115.
- Christodoulopoulos, C., Goldwater, S. dan Steedman, M. (2010), Two decades of unsupervised POS induction: How far have we come?, *in* 'Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing', Association for Computational Linguistics, hlm. 575–584.
- Cohen, K. B., Ogren, P. V., Fox, L. dan Hunter, L. (2005), Corpus design for biomedical natural language processing, *in* 'Proceedings of the ACL-ISMB workshop on linking biological literature, ontologies and databases: mining biological semantics', Association for Computational Linguistics, hlm. 38–45.
- Cunningham, H., Tablan, V., Bontcheva, K. dan Dimitrov, M. (2003), Language engineering tools for collaborative corpus annotation, *in* 'Proceedings of Corpus Linguistics', Vol. 2003, Citeseer.

- Curtotti, M. dan McCreath, E. C. (2011), A Corpus of Australian Contract Language: Description, Profiling and Analysis, *in* 'Proceedings of the 13th International Conference on Artificial Intelligence and Law', ICAIL '11, ACM, New York, NY, USA, hlm. 199–208.  
**URL:** <http://doi.acm.org/10.1145/2018358.2018387>
- Das, D. dan Petrov, S. (2011), Unsupervised part-of-speech tagging with bilingual graph-based projections, *in* 'Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1', Association for Computational Linguistics, hlm. 600–609.
- Dash, N. (2005), *Corpus Linguistics and Language Technology: With Reference to Indian Languages*, Mittal Publications.  
**URL:** <https://books.google.co.id/books?id=zeeeVRq7gakC>
- Davies, M. (2005), 'The advantage of using relational databases for large corpora: Speed, advanced queries, and unlimited annotation', *International Journal of Corpus Linguistics* **10**(3), 307–334.
- De Pauw, G., Wagacha, P. W. dan De Schryver, G.-M. (2009), The SAWA corpus: a parallel corpus English-Swahili, *in* 'Proceedings of the First Workshop on Language Technologies for African Languages', Association for Computational Linguistics, hlm. 9–16.
- Dien, D. dan Kiem, H. (2003), POS-tagger for English-Vietnamese bilingual corpus, *in* 'Proceedings of the HLT-NAACL 2003 Workshop on Building and using parallel texts: data driven machine translation and beyond-Volume 3', Association for Computational Linguistics, hlm. 88–95.
- Dinakaramani, A., Rashel, F., Luthfi, A. dan Manurung, R. (2014), Designing an Indonesian part of speech tagset and manually tagged Indonesian corpus, *in* 'Asian Language Processing (IALP), 2014 International Conference on', IEEE, hlm. 66–69.
- Dipper, S., Faulstich, L., Leser, U. dan Lüdeling, A. (2004), Challenges in modelling a richly annotated diachronic corpus of german, *in* 'Workshop on XML-based richly annotated corpora, Lisbon, Portugal', hlm. 21–29.
- dos Santos, C. dan Milidiú, R. (2012), Part-of-Speech Tagging, *in* 'Entropy Guided Transformation Learning: Algorithms and Applications', SpringerBriefs in Computer Science, Springer London, hlm. 35–41.
- Dukes, K., Atwell, E. dan Habash, N. (2013), 'Supervised Collaboration for Syntactic Annotation of Quranic Arabic', *Lang. Resour. Eval.* **47**(1), 33–62.  
**URL:** <http://dx.doi.org/10.1007/s10579-011-9167-7>

- Dukes, K., Atwell, E. dan Sharaf, A. B. (2010), Syntactic annotation guidelines for the Quranic Arabic Dependency Treebank, *in* 'Proceedings of the Language Resources and Evaluation Conference (LREC)'.
- Dukes, K. dan Habash, N. (2010), Morphological annotation of quranic arabic, *in* 'Proceedings of the Language Resources and Evaluation Conference (LREC)'.
- Duong, L., Cook, P., Bird, S. dan Pecina, P. (2013), Simpler unsupervised POS tagging with bilingual projections., *in* 'ACL (2)', hlm. 634–639.
- Eisele, A. dan Chen, Y. (2010), MultiUN: A Multilingual Corpus from United Nation Documents., *in* 'LREC'.
- Erjavec, T., Kim, J.-D., Ohta, T., Tateisi, Y. dan Tsujii, J.-i. (2003), Encoding biomedical resources in TEI: the case of the GENIA corpus, *in* 'Proceedings of the ACL 2003 workshop on Natural language processing in biomedicine-Volume 13', Association for Computational Linguistics, hlm. 97–104.
- Fachrurrozi, M., Yusliani, N. dan Yoanita, R. U. (2013), Frequent Term based Text Summarization for Bahasa Indonesia.
- Farajian, M. A. (2011), Pen: Parallel english-persian news corpus, *in* 'Proceedings of the 2011th World Congress in Computer Science, Computer Engineering and Applied Computing'.
- Fillmore, C., Ide, N., Jurafsky, D. dan Macleod, C. (1998), An american national corpus: A Proposal, *in* 'Proc. First International Conference on Language Resources and Evaluation (LREC 1998)', hlm. 965–970.
- Galus, S. (2005), Dictionary-based part-of-speech tagging of Polish, *in* 'Intelligent Information Processing and Web Mining', Springer, hlm. 179–188.
- Garside, R., Leech, G. N., McEnery, T. dkk. (1997), *Corpus annotation: linguistic information from computer text corpora*, Longman London, New York.
- Georgila, K., Gordon, C., Choi, H., Boberg, J., Jeon, H. dan Traum, D. (2018), Toward Low-Cost Automated Evaluation Metrics for Internet of Things Dialogues, *in* 'Proceedings of the International Workshop on Spoken Dialogue Systems Technology (IWSDS)'.
- Graën, J. dan Clematide, S. (2015), 'Challenges in the Alignment, Management and Exploitation of Large and Richly Annotated Multi-Parallel Corpora', *Challenges in the Management of Large Corpora (CMLC-3)* p. 15.
- Granger, S. (2003), 'The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research', *Tesol Quarterly* **37**(3), 538–546.

- Guo, H., Na, X. dan Li, J. (2018), 'Qcorp: an annotated classification corpus of Chinese health questions', *BMC medical informatics and decision making* **18**(1), 16.
- Hamam, H., Othman, M. B., Kilani, A., Ammar, M. B. dan Ncibi, F. (2015), Data Mining in the Quran Using Aspects and Dependencies, in 'The 3rd International Conference on Islamic Applications in Computer Science and Technologies (IMAN 2015), Konya, Turkey, 1st-3rd October'.
- Hamzah, A. (2011), 'Penilaian kualitas terjemahan'.
- Heid, U., Schmid, H., Eckart, K. dan Hinrichs, E. W. (2010), A Corpus Representation Format for Linguistic Web Services: The D-SPIN Text Corpus Format and its Relationship with ISO Standards., in 'LREC'.
- Hijriyah, U. (2012), 'Metode dan Penilaian Terjemahan', *JURNAL AL BAYAN* **4**(1).
- Hong, C. (2018), Construction of Corpus in Artificial Intelligence Age, in 'MA-TEC Web of Conferences', Vol. 175, EDP Sciences, p. 03037.
- Ide, N., Bonhomme, P. dan Romary, L. (2000), An XML-based Encoding Standard for Linguistic Corpora, in 'Proceedings of the Second International Conference on Language Resources and Evaluation', hlm. 825–830.
- Ide, N. M. dan Sperberg-McQueen, C. M. (1995), 'The TEI: History, goals, and future', *Computers and the Humanities* **29**(1), 5–15.
- Ide, N. dan Macleod, C. (2001), The american national corpus: A standardized resource of american english, in 'Proceedings of Corpus Linguistics 2001', Vol. 3.
- Ide, N. dan Suderman, K. (2006), Integrating linguistic resources: The american national corpus model, in 'Proceedings of the 6th International Conference on Language Resources and Evaluation', Citeseer.
- Indurkha, N. dan Damerau, F. (2010), *Handbook of Natural Language Processing, Second Edition*, Chapman & Hall/CRC machine learning & pattern recognition series, CRC Press.  
**URL:** <https://books.google.co.id/books?id=nK-QYHZ0-gC>
- Isozaki, H. dan Kazawa, H. (2002), Efficient support vector classifiers for named entity recognition, in 'Proceedings of the 19th international conference on Computational linguistics-Volume 1', Association for Computational Linguistics, hlm. 1–7.



- Kamal Ismail, N., Abd Rahman, N. dan Abu Bakar, Z. (2007), Terms visualization for Malay translated Quran documents, in 'Proceedings of the International Conference on Electrical Engineering and Informatics'.
- Kennedy, G. (2014), *An Introduction to Corpus Linguistics*, Studies in Language and Linguistics, Taylor & Francis.  
**URL:** <https://books.google.co.id/books?id=qFuPBAAAQBAJ>
- Khotimah, T. (2014), 'Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means', *Jurnal Simetris* **5**(1), 83–88.
- Kim, J.-D., Ohta, T., Tateisi, Y. dan Tsujii, J. (2003), 'GENIA corpus—a semantically annotated corpus for bio-textmining', *Bioinformatics* **19**(suppl 1), i180–i182.
- Kiss, T. dan Strunk, J. (2006), 'Unsupervised Multilingual Sentence Boundary Detection', *Comput. Linguist.* **32**(4), 485–525.  
**URL:** <http://dx.doi.org/10.1162/coli.2006.32.4.485>
- Koyejo, O. O., Natarajan, N., Ravikumar, P. K. dan Dhillon, I. S. (2015), Consistent multilabel classification, in 'Advances in Neural Information Processing Systems', hlm. 3321–3329.
- Krause, T. dan Zeldes, A. (2014), 'ANNIS3: A new architecture for generic corpus query and visualization', *Literary and Linguistic Computing* p. fqu057.
- Kridalaksana, H. (1986a), *Kelas kata dalam bahasa Indonesia*, Gramedia Pustaka Utama.
- Kridalaksana, H. (1986b), 'Pendekatan Historis dalam Kajian Bahasa Melayu dan Indonesia', *makalah dalam Masyarakat Linguistik Indonesia Th 4*.
- Kriegel, A. (2011), *Discovering SQL: A Hands-On Guide for Beginners*, IT Pro, Wiley.  
**URL:** <https://books.google.co.id/books?id=HblnED6bKUgC>
- Krosing, H. dan Mlodgenski, J. (2013), *PostgreSQL Server Programming*, Community experience distilled, Packt Publishing.  
**URL:** <https://books.google.co.id/books?id=uZfeUAm-gG8C>
- Kucera, K. (2002), 'The Czech National Corpus: principles, design, and results', *Literary and linguistic computing* **17**(2), 245–257.
- Kupiec, J. (1992), 'Robust part-of-speech tagging using a hidden Markov model', *Computer Speech & Language* **6**(3), 225–242.
- Kurniawan, K. dan Aji, A. F. (2018), 'Toward a Standardized and More Accurate Indonesian Part-of-Speech Tagging', *arXiv preprint arXiv:1809.03391*.

- Kushartanti, U. Y. dan Lauder, M. R. (2005), 'Pesona bahasa: langkah awal memahami linguistik'.
- Landgrebe, J. dan Smith, B. (2019), 'Making AI Meaningful Again', *arXiv preprint arXiv:1901.02918*.
- Larasati, S. D. (2012), IDENTIC Corpus: Morphologically Enriched Indonesian-English Parallel Corpus., in 'LREC', hlm. 902–906.
- Larasati, S. D., Kuboň, V. dan Zeman, D. (2011), Indonesian morphology tool (morphind): Towards an Indonesian corpus, in 'Systems and Frameworks for Computational Morphology', Springer, hlm. 119–129.
- Lauder, M. R. (2008), Orientasi pengembangan kosakata dalam menyongsong masyarakat madani Indonesia, in 'Paper, Seminar Pemartabatan Bahasa Kebangsaan, Pusat Bahasa, Jakarta', hlm. 7–8.
- Lauder, M. R. dan Lauder, A. F. (2004), Upaya Pembuatan Korpus Linguistik: Persoalan dan Pemanfaatannya, in B. K. Purwo, ed., 'PELBBA 17', Yayasan Obor Indonesia.
- Leech, G. (1992), '100 million words of English: the British National Corpus (BNC)', *Language Research* **28**(1), 1–13.
- Leech, G. (1993), 'Corpus annotation schemes', *Literary and linguistic computing* **8**(4), 275–281.
- Leech, G., Garside, R. dan Bryant, M. (1994), CLAWS4: the tagging of the British National Corpus, in 'Proceedings of the 15th conference on Computational linguistics-Volume 1', Association for Computational Linguistics, hlm. 622–628.
- Lu, X. (2014), *Computational Methods for Corpus Annotation and Analysis*, Springer.  
**URL:** <https://books.google.co.id/books?id=UCu5BQAAQBAJ>
- Lubis, I. (2001), *Falsifikasi Terjemahan Al-Quran Departemen Agama Edisi 1990*, PT Tiara Wacana, Yogyakarta.
- Lubis, I. (2012), 'Ihwal penerjemahan bahasa Arab ke dalam bahasa Indonesia', *Jurnal Humaniora* **16**(1).
- Mangasi, T., Erwin, A. dan Ipung, H. P. (2014), Defined entity extraction based on Indonesian text document, in 'ICT For Smart Society (ICISS), 2014 International Conference on', IEEE, hlm. 61–65.



- Marcus, M. P., Marcinkiewicz, M. A. dan Santorini, B. (1993), 'Building a large annotated corpus of English: The Penn Treebank', *Computational linguistics* **19**(2), 313–330.
- Markhamah (2013), *Ragam dan Analis Kalimat Bahasa Indonesia*, Muhammadiyah University Press.
- McEnery, T. dan Wilson, A. (2001), *Corpus Linguistics: An Introduction*, Edinburgh University Press Series, Edinburgh University Press.  
**URL:** [https://books.google.co.id/books?id=nwmgdvN\\_akAC](https://books.google.co.id/books?id=nwmgdvN_akAC)
- McEnery, T., Xiao, R. dan Tono, Y. (2006), *Corpus-based Language Studies: An Advanced Resource Book*, Routledge applied linguistics, Routledge.  
**URL:** <https://books.google.co.id/books?id=SU8yRlBA9rAC>
- Meyer, C. F. (2002), *English corpus linguistics: An introduction*, Cambridge University Press.
- Moeliono, A. M. dan Dardjowidjojo, S. (1988), *Tata bahasa baku bahasa Indonesia*, Departemen Pendidikan dan Kebudayaan, Republik Indonesia.
- Muflikah, L., Marji, M. dan Liliana, D. Y. (2013), 'Penggalian Data dalam Penentuan Keterkaitan Topik pada Terjemahan Ayat-ayat Al-Qur'an', *Journal of Natural A* **1**(1), pp–40.
- Muljono, M., Afini, U., Supriyanto, C. dan Nugroho, R. A. (2017), 'The Development of Indonesian POS Tagging System for Computer-aided Independent Language Learning', *International Journal of Emerging Technologies in Learning (iJET)* **12**(11), 138–150.
- Mulyono, I. (2013), *Ilmu Bahasa Indonesia Morfologi Teori dan Sejumput Problematik Terapannya*, Yrama Widya.
- Muslich, M. (2010), *Garis-Garis Besar Tata Bahasa Baku Bahasa Indonesia*, Refika Aditama, Bandung.
- Mustapha, A. (2009), 'Dialogue-based visualization for quranic text', *European Journal of Scientific Research* **37**(1), 36–40.
- Nababan, M. (2008), 'Kompetensi Penerjemahan dan Dampaknya pada Kualitas Terjemahan'.
- Nancy, I. (1998), Encoding Linguistic Corpora, in 'Proceedings of the sixth Workshop on Very Large Corpora: 15th-16th August 1998, Université de Montréal, Montreal, Quebec, Canada', Université de Montréal, p. 9.

- Nazief, B. (2000), Development of computational Linguistics research: A challenge for Indonesia, in 'Proceedings of the 38th Annual Meeting on Association for Computational Linguistics', Association for Computational Linguistics, hlm. 1–2.
- Nugraheni, D., Bijaksana, M. dan Darwiyanto, E. (2017), 'Analisis Dan Implementasi Pencarian Kata Berbasis Konkordansi Dan N-gram Pada Terjemahan Al-quran Berbahasa Indonesia', *eProceedings of Engineering* 4(3).
- Ohta, T., Tateisi, Y. dan Kim, J.-D. (2002), The GENIA corpus: An annotated research abstract corpus in molecular biology domain, in 'Proceedings of the second international conference on Human Language Technology Research', Morgan Kaufmann Publishers Inc., hlm. 82–86.
- O'Keeffe, A. dan McCarthy, M. (2010), *The Routledge handbook of corpus linguistics*, Routledge.
- Oostdijk, N. H. (2006), 'A reference corpus of written Dutch', *Corpus design (D-coi 06-01)*. Persistent identifier: urn: nbn: nl: ui hlm. 22–2066.
- Palmer, D. D. dan Hearst, M. A. (1997), 'Adaptive multilingual sentence boundary disambiguation', *Computational Linguistics* 23(2), 241–267.
- Parera, J. D. (2010), *Morfologi Bahasa*, Gramedia Pustaka Utama.
- Pateda, M. (2010), *Semantik Leksikal*, Rineka Cipta.
- Pisceldo, F., Manurung, R. dan Adriani, M. (2009), Probabilistic part-of-speech tagging for bahasa indonesia, in 'Third International MALINDO Workshop, Colocated Event ACL-IJCNLP'.
- Przepiórkowski, A. dan Bański, P. (2011), Which XML standards for multilevel corpus annotation?, in 'Human Language Technology. Challenges for Computer Science and Linguistics', Springer, hlm. 400–411.
- Rafalovitch, A. dan Dale, R. (2009), United nations general assembly resolutions: A six-language parallel corpus, in 'Proceedings of the MT Summit', Vol. 12, hlm. 292–299.
- Raharjo, S. dan Hartati, S. (2014), 'Antarmuka Bahasa Alami Untuk Melakukan Query Terhadap Terjemahan Al-Quran', *Jurnal Teknologi* 7(1).
- Rahman, M. A. (2012), 'Pengelompokan Ayat Al-Qur'an Berdasarkan Kesamaan Kata Menggunakan Metode Association Rule', *Jurnal POINTER* 2(2), 101.
- Rashel, F., Luthfi, A., Dinakaramani, A. dan Manurung, R. (2014), Building an Indonesian rule-based part-of-speech tagger, in 'Asian Language Processing (IALP), 2014 International Conference on', IEEE, hlm. 70–73.

- Ratnaparkhi, A. dkk. (1996), A maximum entropy model for part-of-speech tagging, *in* 'Proceedings of the conference on empirical methods in natural language processing', Vol. 1, Philadelphia, USA, hlm. 133–142.
- Read, J., Dridan, R., Oepen, S. dan Solberg, L. J. (2012), Sentence Boundary Detection: A Long Solved Problem?, *in* 'COLING (Posters)', hlm. 985–994.
- Reppen, R. dan Ide, N. (2004), 'The american national corpus overall goals and the first release', *Journal of English Linguistics* **32**(2), 105–113.
- Reynar, J. C. dan Ratnaparkhi, A. (1997), A maximum entropy approach to identifying sentence boundaries, *in* 'Proceedings of the fifth conference on Applied natural language processing', Association for Computational Linguistics, hlm. 16–19.
- Rifai, M. A. (2013), Kemutlakan Pengadaan Korpus Demi Keberhasilan Pembinaan dan Pengembangan Bahasa Indonesia, *in* 'Kongres Bahasa Indonesia X'.
- Riset, K. N. (2006), 'INDONESIA 2005-2025 BUKU PUTIH Penelitian', *Pengembangan dan Penerapan Ilmu Pengetahuan dan Teknologi Jakarta*.
- Roberts, A., Gaizauskas, R., Hepple, M., Davis, N., Demetriou, G., Guo, Y., Kola, J. S., Roberts, I., Setzer, A., Tapuria, A. dkk. (2007), The CLEF corpus: semantic annotation of clinical text, *in* 'AMIA Annual Symposium Proceedings', Vol. 2007, American Medical Informatics Association, p. 625.
- Samokhvalov, N. (2007), 'XML Support in PostgreSQL.', *SYRCoDIS* **256**.
- Seah, Y. J. dan Bond, F. (2014), Annotation of Pronouns in a Multilingual Corpus of Mandarin Chinese, English and Japanese, *in* 'Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation', p. 82.
- Setiarini, N. L. P., Said, M. dan Imran, I. (2009), Kolokasi Bahasa Indonesia, *in* 'Proceeding PESAT (Psikologi, Ekonomi, Sastra, Arsitektur & Sipil)', Universitas Gunadarma.
- Sharaf, A.-B. dan Atwell, E. (2012a), QurAna: Corpus of the Quran annotated with Pronominal Anaphora, *in* 'Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)', European Language Resources Association (ELRA), Istanbul, Turkey.
- Sharaf, A.-B. dan Atwell, E. (2012b), QurSim: A corpus for evaluation of relatedness in short texts, *in* 'Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)', European Language Resources Association (ELRA), Istanbul, Turkey.

- Sharaf, A. M. dan Atwell, E. (2009), 'The Qur'an annotation for text mining', *First year transfer report. School of Computing, Leeds University. December*.
- Sokolova, M. dan Lapalme, G. (2009), 'A systematic analysis of performance measures for classification tasks', *Information Processing & Management* 45(4), 427–437.
- STM (2007), *Sistem Pengenal Akhir Kalimat Dengan Metode Pembelajaran Feed Forward Jaringan Syaraf Tiruan*.
- Stührenberg, M. (2012), 'The TEI and Current Standards for Structuring Linguistic Data. An Overview', *Journal of the Text Encoding Initiative* (3).
- Syahrullah (2012), 'Tarjamah Tafsiriah Terhadap Al- Qur'an: Antara Kontekstualisasi dan Distorsi ', *JOURNAL OF QUR'AN AND HADITH STUDIES (QUHAS)* 2(1).
- Tadjuddin, M. (2013), *Bahasa Indonesia: Bentuk dan Makna*, Alumni, Bandung.
- Tognini-Bonelli, E. (2001), *Corpus Linguistics at Work*, Linguistics Today, J. Benjamins.  
**URL:** <https://books.google.co.id/books?id=z0TZmK1YWTIC>
- Uliniansyah, T., Riza, H. dan Riandi, O. (2013), Developing corpus management system for Bahasa Indonesia the Perisalah project, in 'Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 2013 International Conference', hlm. 1–4.
- Utami, E. dan Raharjo, S. (2014), 'Database security model in the academic information system', *International Journal of Security and Its Applications* 8(170).
- van Halteren, H. (1999), *Syntactic Wordclass Tagging*, Text, Speech and Language Technology, Springer.  
**URL:** <https://books.google.co.id/books?id=6SmOssWntJMC>
- Walker, D. J., Clements, D. E., Darwin, M. dan Amtrup, J. W. (2001), Sentence boundary detection: A comparison of paradigms for improving MT quality, in 'Proceedings of the MT Summit VIII'.
- Widhiyanti, K. dan Harjoko, A. (2013), 'POS Tagging Bahasa Indonesia Dengan HMM dan Rule Based', *Jurnal Informatika* 8(2).
- Wijaya, M. T. (2014), 'Menakar Ulang Kualitas Buku-Buku Terjemahan di Indonesia', *AL-TURATS* 19(1).

- Wilcock, G. (2009), 'Introduction to linguistic annotation and text analytics', *Synthesis Lectures on Human Language Technologies* **2**(1), 1–159.
- Wong, D. F., Chao, L. S. dan Zeng, X. (2014), 'iSentenizer-: Multilingual Sentence Boundary Detection Model', *The Scientific World Journal* **2014**.
- Xiao, R. Z. (2008), Well-known and influential corpora., *in* 'Corpus linguistics. An international handbook, handbooks of linguistics and communication science ', Mouton de Gruyter.
- Yaacob, M. dkk. (2010), 'Dedicated graphical user interface system for the visually impaired users in learning Al-Quran'.
- Yong, C. Y., Sudirman, R., Chew, K. M. dan Salim, N. (2011), Comparison of Ontology Learning Techniques for Qur'anic Text, *in* 'Future Computer Sciences and Application (ICFCSA), 2011 International Conference on', hlm. 192–196.
- Yuningsih, Y. (2009), 'Penerjemahan dokumen resmi Arab-Indonesia (studi kritik gramatikal terjemahan penerjemah resmi Al-Hadi)'.