

References

- ARTEMENKO O., MANDL T., SHRAMKO M. and WOMSER-HACKER C. (2006), "Evaluation of a Language Identification System for Mono- and Multilingual Text Documents", in *Proceedings of the 13th Annual Workshop on Selected Areas in Cryptography*, Dijon.
- BEESLEY K. (1988), *Language identifier: A computer program for automatic natural-language identification on on-line text*.
- CAVNAR W. B. and TRENKLE J. M. (1994), "N-Gram-Based Text Categorization", in *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*: 161-175.
- DYEN I., KRUSKAL J. and BLACK P. (1992), "An Indo-European classification: A lexicostatistical experiment", in *Transactions of the American Philosophical Society*, 82:1-132.
- ELLISON T. M. and KIRBY S. (2006), "Measuring Language Divergence by Intra-Lexical Comparison", in *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Sydney.
- GIGUET E. (1995), "Multilingual Sentence Categorisation According to Language", in *Proceedings of the European Chapter of the Association for Computational Linguistics, SIGDAT Workshop, From Text to Tags: Issues in Multilingual Language Analysis*, Dublin.
- GORDON R. G. (2005), *Ethnologue: Languages of the World, Fifteenth edition (ed.)*, Online version: <http://www.ethnologue.com/web.asp>.
- INGLE N. C. (1976), "A Language Identification Table", in *The Incorporated Linguist*, 15(4).
- KIKUI G. (1996), "Identifying the Coding System and Language of On-line Documents on the Internet.", in *COLING*: 652-657.
- NERBONNE J. and HEERINGA W. (1997), "Measuring dialect distance phonetically", in *Proceedings of SIGPHON-97: 3rd Meeting of the ACL Special Interest Group in Computational Phonology*.
- PRAGER J. M. (1999), "Linguini: Language Identification for Multilingual Documents", in *Proceedings of the 32nd Hawaii International Conference on System Sciences*.
- SIBUN P. and REYNAR J. C. (1996), "Language Identification: Examining the Issues", in *In Proceedings of SDAIR-96, the 5th Symposium on Document Analysis and Information Retrieval*: 125-135.
- SINGH A. K. (2006), "Study of Some Distance Measures for Language and Encoding Identification", in *Proceedings of ACL 2006 Workshop on Linguistic Distance*, Sydney.
- STEPHEN J. (1993), "Solving the Problem of Language Recognition", in *Technical Report*, School of Computer Studies, University of Leeds.
- SWADESH M. (1952), "Lexico-dating of prehistoric ethnic contacts", in *Proceedings of the American philosophical society*, 96(4).
- TAN C., LEONG P. and HE S. (1999), "Language Identification in Multilingual Documents", in *Int'l Symp. Intelligent Multimedia and Distance Education*.
- TEAHAN W. J. and HARPER D. J. (2001), "Using Compression Based Language Models for Text Categorization", in J. Callan, B. Croft and J. Lafferty (eds.), *Workshop on Language Modeling and Information Retrieval*.

CLEANEVAL