References

[1] A. Fader, L. Zettlemoyer, and O. Etzioni, "Open question answering over curated and extracted knowledge bases," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '14. New York, NY, USA: Association for Computing Machinery, 2014. doi: 10.1145/2623330.2623677. ISBN 9781450329569 p. 1156–1165. [Online]. Available: https://doi.org/10.1145/2623330.2623677

[2] X. Wang, C. Macdonald, and I. Ounis, "Deep reinforced query reformulation for information retrieval," 2020.

[3] Z. Yan, N. Duan, J. Bao, P. Chen, M. Zhou, Z. Li, and J. Zhou, "DocChat: An information retrieval approach for chatbot engines using unstructured documents," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016. doi: 10.18653/v1/P16-1049 pp. 516–525. [Online]. Available: https://www.aclweb.org/anthology/P16-1049

[4] A. Gupta, A. Agarwal, P. Singh, and P. Rai, "A deep generative framework for paraphrase generation," *arXiv preprint arXiv:1709.05074*, 2017.

[5] Q. Yang, D. Shen, Y. Cheng, W. Wang, G. Wang, L. Carin *et al.*, "An end-to-end generative architecture for paraphrase generation," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 3123–3133.

[6] Z. Li, X. Jiang, L. Shang, and H. Li, "Paraphrase generation with deep reinforcement learning," *arXiv preprint arXiv:1711.00279*, 2017.

[7] J. Tse, D. Schrader, D. Ghosh, Liao, and D. Lundie, "A bibliometric analysis of privacy and ethics in ieee security and privacy," *Ethics and Information Technology*, vol. 17, pp. 153–163, 08 2015. doi: 10.1007/s10676-015-9369-6

[8] D. Hovy and S. L. Spruit, "The social impact of natural language processing," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016. doi: 10.18653/v1/P16-2096 pp. 591–598. [Online]. Available: https://www.aclweb.org/anthology/P16-2096

[9] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014.

[10] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015. doi: 10.18653/v1/D15-1166 pp. 1412–1421. [Online]. Available: https://www.aclweb.org/anthology/D15-1166

[11] T. Dozat and C. D. Manning, "Deep biaffine attention for neural dependency parsing," 2016.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[13] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," 2018.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018.

[15] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," 2019.

[16] D. R. So, C. Liang, and Q. V. Le, "The evolved transformer," 2019.

[17] J. M. Zurada, *Introduction to artificial neural systems*. West St. Paul, 1992, vol. 8.

[18] S. Marsland, *Machine learning: an algorithmic perspective*. CRC press, 2015.

[19] D. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[21] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 07 2011.

[22] G. Hinton, "Neural networks for machine learning," Coursera, 2012.

[23] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," 2015.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[25] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 12 1997. doi: 10.1162/neco.1997.9.8.1735

[27] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013.

[28] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," 2014.

[29] C. Doersch, "Tutorial on variational autoencoders," 2016.

[30] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79 – 86, 1951. doi: 10.1214/aoms/1177729694. [Online]. Available: https://doi.org/10.1214/aoms/1177729694

[31] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[32] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013.

[33] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014. doi: 10.3115/v1/D14-1162 pp. 1532–1543. [Online]. Available: https://www.aclweb.org/anthology/D14-1162

[34] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *arXiv preprint arXiv:1607.04606*, 2016.

[35] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Łukasz Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," 2016.

[36] B. Turovsky, "Found in translation: More accurate, fluent sentences in google translate," 11 2016.

[37] ——, "Ten years of google translate," 10 2016.

[38] *Cloud Translation documentation*.

[39] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, "Bleu: a method for automatic evaluation of machine translation," 10 2002. doi: 10.3115/1073083.1073135

[40] A. Lavie and A. Agarwal, "Meteor: An automatic metric for mt evaluation with high levels of correlation with human judgments," pp. 228–231, 07 2007.

[41] C. J. V. Rijsbergen, *Information Retrieval*, 2nd ed. USA: Butterworth-Heinemann, 1979. ISBN 0408709294

[42] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul, "A study of translation edit rate with targeted human annotation," in *In*

Proceedings of Association for Machine Translation in the Americas, 2006, pp. 223–231.

[43] K. McKeown, "Paraphrasing questions using given and new information," *Am. J. Comput. Linguistics*, vol. 9, pp. 1–10, 1983.

[44] I. Bolshakov and A. Gelbukh, "Synonymous paraphrasing using wordnet and internet," 01 2004. ISBN 3-540-22564-1 pp. 312–323.

[45] C. Fellbaum, *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.

[46] S. Zhao, X. Lan, T. Liu, and S. Li, "Application-driven statistical paraphrase generation," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ser. ACL '09. USA: Association for Computational Linguistics, 2009. ISBN 9781932432466 p. 834–842.

[47] A. Prakash, S. A. Hasan, K. Lee, V. Datla, A. Qadir, J. Liu, and O. Farri, "Neural paraphrase generation with stacked residual lstm networks," 2016.

[48] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," 2014.

[49] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," 2014.

[50] X. Li and X. Wu, "Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition," 2014.

[51] O. Vinyals, L. Kaiser, T. Koo, S. Petrov, I. Sutskever, and G. Hinton, "Grammar as a foreign language," 2014.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[53] Z. Cao, C. Luo, W. Li, and S. Li, "Joint copying and restricted generation for paraphrase," 11 2016.

[54] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014.

[55] S. Ma, X. Sun, W. Li, S. Li, W. Li, and X. Ren, "Query and output: Generating words by querying distributed word representations for paraphrase generation," 2018.

[56] S. Huang, Y. Wu, F. Wei, and M. Zhou, "Dictionary-guided editing networks for paraphrase generation," 2018.

[57] J. Ganitkevitch, B. Van Durme, and C. Callison-Burch, "PPDB: The paraphrase database," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia: Association for Computational Linguistics, Jun. 2013, pp. 758–764. [Online]. Available: https://www.aclweb.org/anthology/N13-1092

[58] M. Iyyer, J. Wieting, K. Gimpel, and L. Zettlemoyer, "Adversarial example generation with syntactically controlled paraphrase networks," 2018.

[59] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Baltimore, Maryland: Association for Computational Linguistics, Jun. 2014. doi: 10.3115/v1/P14-5010 pp. 55–60. [Online]. Available: https://www.aclweb.org/anthology/P14-5010

[60] S. Wang, R. Gupta, N. Chang, and J. Baldridge, "A task in a suit and a tie: paraphrase generation with semantic augmentation," 2018.

[61] M. Ringgaard, R. Gupta, and F. C. N. Pereira, "Sling: A framework for frame semantic parsing," 2017.

[62] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *arXiv preprint arXiv:1511.06349*, 2015.

[63] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in Neural*

Information Processing Systems 28, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 3483–3491. [Online]. Available: https://proceedings.neurips.cc/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf

[64] M. Aggarwal, N. Kumari, A. Bansal, and B. Krishnamurthy, "Redecode framework for iterative improvement in paraphrase generation," *arXiv preprint arXiv:1811.04454*, 2018.

[65] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," 2015.

[66] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," 2016.

[67] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky, "Adversarial learning for neural dialogue generation," 2017.

[68] Y. Zhang, Z. Gan, K. Fan, Z. Chen, R. Henao, D. Shen, and L. Carin, "Adversarial feature matching for text generation," 2017.

[69] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, and J. Wang, "Long text generation via adversarial training with leaked information," 2017.

[70] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, no. 3–4, p. 229–256, May 1992. doi: 10.1007/BF00992696. [Online]. Available: https://doi.org/10.1007/BF00992696

[71] Z. An and S. Liu, "Towards diverse paraphrase generation using multi-class wasserstein gan," *arXiv preprint arXiv:1909.13827*, 2019.

[72] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. International Convention Centre, Sydney, Australia: PMLR, 06–11 Aug 2017, pp. 214–223. [Online]. Available: http://proceedings.mlr.press/v70/arjovsky17a.html

[73] N. D. Ratliff, J. A. Bagnell, and M. A. Zinkevich, "Maximum margin planning," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: Association for Computing Machinery, 2006. doi: 10.1145/1143844.1143936. ISBN

1595933832 p. 729–736. [Online]. Available: https://doi.org/10.1145/1143844.1143936

[74] A. B. Siddique, S. Oymak, and V. Hristidis, "Unsupervised paraphrasing via deep reinforcement learning," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, ser. KDD '20. New York, NY, USA: Association for Computing Machinery, 2020. doi: 10.1145/3394486.3403231. ISBN 9781450379984 p. 1800–1809. [Online]. Available: https://doi.org/10.1145/3394486.3403231

[75] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," 2015.

[76] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A Python natural language processing toolkit for many human languages," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2020. [Online]. Available: https://nlp.stanford.edu/pubs/qi2020stanza.pdf

[77] L. Borin, M. Forsberg, L.-J. Olsson, and J. Uppström, "The open lexical infrastructure of språkbanken," in *Proceedings of the 8th International Conference on Language Resources and Evaluation : May 23-25, 2012 / eds. Nicoletta Calzolari*, 2012. ISBN 978-2-9517408-7-7 pp. 3598–3602.

[78] A. Gatt and E. Krahmer, "Survey of the state of the art in natural language generation: Core tasks, applications and evaluation," 2018.

[79] J. Randolph, "Free-marginal multirater kappa (multirater free): An alternative to fleiss fixed-marginal multirater kappa," vol. 4, 01 2010.

[80] G. Lembersky, N. Ordan, and S. Wintner, "Language models for machine translation: Original vs. translated texts," *Computational Linguistics*, vol. 38, no. 4, pp. 799–825, 2012. doi: 10.1162/COLI_a_00111. [Online]. Available: https://doi.org/10.1162/COLI_a_00111

[81] E. A. Avner, N. Ordan, and S. Wintner, "Identifying translationese at the word and sub-word level," *Digital Scholarship in the Humanities*, vol. 31, no. 1, pp. 30–54, 2016.

Appendix A

Evaluation Instructions To The Human Judges

I detta frågeformulär består uppgiften av att utvärdera frågepar utifrån kriteriena grammatiskt korrekthet och hur likvärdighet mellan frågorna, dvs om de har samma innebörd. Båda kriterier bedöms på en skala 0-3 där 3 innebär att du instämmer fullt med påståendet och 0 att du inte alls instämmer med påståendet. Exempel på hur de olika poängen defineras ges nedan:

Grammatisk korrekthet av fråga X:
Påstående: Fråga X är grammatiskt korrekt.

(0) Instämmer inte alls. Fråga X är grammatiskt inkorrekt och innehåller så grova fel att den är svår att läsa.

(1) Instämmer lite grann. Fråga X är ej grammatiskt korrekt och innehåller flera mindre fel.

(2) Instämmer till viss del. Fråga X är nästan grammatiskt korrekt men innehåller något mindre fel.

(3) Instämmer fullt. Fråga X är helt grammatiskt korrekt.

Likvärdighet mellan fråga X och fråga Y
Påstående: Fråga X har samma betydelse som fråga Y.

(0) Instämmer inte alls. Fråga X skiljer sig markant från fråga Y både vad