

Prodromos Malakasiotis
Curriculum Vitae

Last update: February 2023

1. CONTACT INFORMATION

Mobile phone: +30 697 2810683
Personal e-mail: rulller@gmail.com
Personal Web page: <http://pages.cs.aueb.gr/~rulller>

2. RESEARCH INTERESTS

Artificial Intelligence, especially **natural language processing**, **information retrieval**, and **information extraction** from **large document collections** and particularly:

- **information retrieval** and **question answering** for large document collections and the world wide web,
- **text classification** including **multi-label classification** with large label-sets hierarchically organized,
- **information extraction** and **opinion mining** including sentiment analysis,
- **machine learning** in **natural language processing**, especially **deep learning**, **transfer learning**, and **few/zero-shot learning**, and
- **natural language processing tools** for English and Greek, for example, part-of-speech taggers and pre-trained language models.

3. STUDIES

2006–11: Department of Informatics, Athens University of Economics and Business, Greece. **PhD in Artificial Intelligence**. Thesis title: *Paraphrase and textual entailment recognition and generation*. Greek PENED Project Scholarship.

2003–05: Department of Informatics, Athens University of Economics and Business, Greece. **MSc in Computer Science**. Grade: 8.08/10. Dissertation title: *Part-of-speech tagging in Greek texts using active learning techniques*. Dissertation grade: 9.5/10 (excellent). Scholarship (2nd highest grade) for the 2nd Semester (Spring 2004).

1999–03: Department of Informatics, Athens University of Economics and Business, Greece. **BSc in Informatics**, Computer Science theme. Grade: 7.10/10.

4. WORK EXPERIENCE

2022-present: Principal Machine Learning Engineer, Workable Software Limited.

2019-2022: Researcher / NLP Expert, NCSR “Demokritos”.

2018-present: Teaching Fellow, Department of Informatics, Athens University of Economics and Business (AUEB).

2018-2019: Data Scientist, Cognitiv+.

2012-present: Research Associate, Department of Informatics, Athens University of Economics and Business (AUEB).

2016-2018: Senior Researcher, StrainTek.

2014-2016: Research Associate, Institute for Language and Speech Processing (ILSP), “Athena” Research Center.

2014–2015: Research Associate in the European LangTERRA project, Institute for Language and Speech Processing (ILSP), “Athena” Research Center.

2014–2015: Post-doctoral Researcher in Action 2 of AUEB Research Support Program, Department of Informatics, Athens University of Economics and Business (AUEB).

2012–2014: Research Assistant in the European BioASQ project, Department of Informatics, Athens University of Economics and Business (AUEB).

2011–2012: Military service (9 months), Hellenic Army. Specialty: Computer Programmer.

2007–2010: Research Assistant in the European INDIGO project, Department of Informatics, Athens University of Economics and Business (AUEB).

5. TEACHING EXPERIENCE

5.1 Undergraduate courses

- *Machine Learning* (2018-21). Dept. of Informatics, AUEB.
- *Artificial Intelligence* (2008–09, as an assistant). Dept. of Informatics, AUEB.
- *Computer Programming in C++* (2005–06 to 2007–08, as an assistant) Dept. of Informatics, AUEB.
- *Logic* (2010–11, as an assistant). Dept. of Informatics, AUEB.
- *Programming Languages – Compilers* (2006–07, as an assistant). Dept. of Informatics, AUEB.

5.2 Postgraduate courses

- *Deep Learning* (2018-present). Dept. of Informatics, AUEB.
- *Logic and Artificial Intelligence* (2005–06 to 2010–11, as an assistant). Dept. of Informatics, AUEB.
- *Language Technology* (2013 – 14, as an assistant). Dept. of Informatics, AUEB.

6. PUBLICATIONS

6.1 Journal articles

J2. G. Tsatsaronis, G. Balikas, P. Malakasiotis, I. Partalas, M. Zschunke, M.R. Alvers, D. Weissenborn, A. Krithara, S. Petridis, D. Polychronopoulos, Y. Almirantis, J. Pavlopoulos, N. Baskiotis, P. Gallinari, T. Artieres, A. Ngonga, N. Heino, E. Gaussier, L. Barrio-Alvers, M. Schroeder, I. Androutsopoulos and G. Paliouras, "An Overview

of the BIOASQ Large-Scale Biomedical Semantic Indexing and Question Answering Competition". *BMC Bioinformatics*, 16:138, 2015.

J1. I. Androutsopoulos and P. Malakasiotis, "A Survey of Paraphrasing and Textual Entailment Methods". *Journal of Artificial Intelligence Research*, 38:135-187, 2010.

6.2 Conference articles

C30. S. Maroudas, S. Legkas, P. Malakasiotis and I. Chalkidis, "Legal-Tech Open Diaries: Lesson learned on how to develop and deploy light-weight models in the era of humongous Language Models". Proceedings of the *Workshop on Natural Legal Language Processing (NLLP) 2022* at the *2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Abu Dhabi, UAE, 2022.

C29. D. Pappas, P. Malakasiotis and I. Androutsopoulos, "Data Augmentation for Biomedical Factoid Question Answering". Proceedings of the *19th Workshop on Biomedical Natural Language Processing (BioNLP 2022)* at the *58th Annual Meeting of the Association for Computational Linguistics (ACL 2022)*, Dublin, Ireland, 2022.

C28. L. Loukas, M. Fergadiotis, I. Chalkidis, E. Spyropoulou, I. Androutsopoulos, P. Malakasiotis and G. Paliouras, "FiNER: Financial Numeric Entity Recognition for XBRL Tagging". Proceedings of the *58th Annual Meeting of the Association for Computational Linguistics (ACL 2022)*, Dublin, Ireland, 2022.

C27. L. Loukas, M. Fergadiotis, I. Androutsopoulos and P. Malakasiotis, "EDGAR-CORPUS: Billions of Tokens Make The World Go Round". Proceedings of the *Economics and Natural Language Processing Workshop (ECONLP 2021)* at the *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, Punta Cana, Dominican Republic, 2021.

C26. I. Chalkidis, M. Fergadiotis, D. Tsarapatsanis, N. Aletras, I. Androutsopoulos and P. Malakasiotis, "Paragraph-level Rationale Extraction through Regularization: A case study on European Court of Human Rights Cases". Proceedings of the *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2021)*, held on-line, 2021.

C25. I. Chalkidis, M. Fergadiotis, N. Manginas, E. Katakalous and P. Malakasiotis, "Regulatory Compliance through Doc2Doc Information Retrieval: A case study in EU/UK legislation where text similarity has limitations". Proceedings of the *16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, held on-line, 2021.

C24. I. Chalkidis, M. Fergadiotis, S. Kotitsas, P. Malakasiotis, N. Aletras and I. Androutsopoulos, "An Empirical Study on Large-Scale Multi-Label Text Classification Including Few and Zero-Shot Labels". Proceedings of the *Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, held on-line, 2020.

C23. I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras and I. Androutsopoulos, "LEGAL-BERT: The Muppets straight out of Law School". Findings of the *Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, 2020.

- C22.** N. Manginas, I. Chalkidis and P. Malakasiotis, "Layer-wise Guided Training for BERT: Learning Incrementally Refined Document Representations". Proceedings of the *Workshop on Structured Prediction for NLP* (SPNLP 2020) at the *Conference on Empirical Methods in Natural Language Processing* (EMNLP 2020), held on-line, 2020.
- C21.** J. Koutsikakis, I. Chalkidis, P. Malakasiotis and I. Androutsopoulos, "GREEK-BERT: The Greeks visiting Sesame Street". Proceedings of the *11th Hellenic Conference on Artificial Intelligence* (SETN 2020), held on-line, 2020.
- C20.** I. Chalkidis, M. Fergadiotis, P. Malakasiotis and I. Androutsopoulos, "Neural Contract Element Extraction Revisited". Proceedings of the *Document Intelligence Workshop* (DI 2019) of the *Thirty-third Conference on Neural Information Processing Systems* (NeurIPS 2019), Vancouver, Canada, 2019.
- C19.** S. Xenouleas, P. Malakasiotis, M. Apidianaki and I. Androutsopoulos, "SUM-QE: a BERT-based Summary Quality Estimation Model". Proceedings of the *Conference on Empirical Methods in Natural Language Processing* (EMNLP 2019), Hong Kong, China, 2019.
- C18.** I. Chalkidis, M. Fergadiotis, P. Malakasiotis and I. Androutsopoulos, "Large-Scale Multi-Label Text Classification on EU Legislation". Proceedings of the *57th Annual Meeting of the Association for Computational Linguistics* (ACL 2019), Florence, Italy, 2019.
- C17.** I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras and I. Androutsopoulos, "Extreme Multi-Label Legal Text Classification: A Case Study in EU Legislation". Proceedings of the *Workshop on Natural Legal Language Processing* (NLLP 2019) of the *Annual Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL-HLT 2019), Minneapolis, USA, 2019.
- C16.** J. Pavlopoulos, P. Malakasiotis and I. Androutsopoulos, "Deeper Attention to Abusive User Content Moderation". Proceedings of the *Conference on Empirical Methods in Natural Language Processing* (EMNLP 2017), Copenhagen, Denmark, pp. 1136-1146, 2017.
- C15.** J. Pavlopoulos, P. Malakasiotis, J. Bakagianni and I. Androutsopoulos, "Improved Abusive Comment Moderation with User Embeddings". Proceedings of the workshop *Natural Language Processing Meets Journalism*, at the *Conference on Empirical Methods in Natural Language Processing* (EMNLP 2017), Copenhagen, Denmark, pp. 51-55, 2017. [Slides, Dataset]
- C14.** J. Pavlopoulos, P. Malakasiotis and I. Androutsopoulos, "Deep Learning for User Comment Moderation". Proceedings of the *1st Workshop on Abusive Language Online* (ALW1), at the *55th Annual Meeting of the Association for Computational Linguistics* (ACL 2017), Vancouver, Canada, pp. 25-35, 2017.
- C13.** G. Brokos, P. Malakasiotis and I. Androutsopoulos, "Using Centroids of Word Embeddings and Word Mover's Distance for Biomedical Document Retrieval in Question Answering". Proceedings of the *15th Workshop on Biomedical Natural Language Processing* (BioNLP 2016), at the *54th Annual Meeting of the Association for Computational Linguistics* (ACL 2016), Berlin, Germany, pp. 114-118, 2016.

- C12.** S. Giorgis, A. Rousas, J. Pavlopoulos, P. Malakasiotis and I. Androutsopoulos, "aueb.twitter.sentiment at SemEval-2016 Task 4: A Weighted Ensemble of SVMs for Twitter Sentiment Analysis". *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*, San Diego, CA, USA, pp. 96-99, 2016.
- C11.** D. Xenos, P. Theodorakakos, J. Pavlopoulos, P. Malakasiotis and I. Androutsopoulos, "AUEB-ABSA at SemEval-2016 Task 5: Supervised Machine Learning for Aspect Based Sentiment Analysis". *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*, San Diego, CA, USA, pp. 324-329, 2016.
- C10.** P. Malakasiotis, E. Archontakis, I. Androutsopoulos, D. Galanis and H. Papageorgiou, "Biomedical question-focused multi-document summarization: ILSP and AUEB at BioASQ3". *Proceedings of the QA Lab of the Sixth Conference and Labs of the Evaluation Forum (CLEF 2015)*, Toulouse, France, 2015.
- C9.** R. M. Karampatsis, J. Pavlopoulos and P. Malakasiotis, "AUEB: Two Stage Sentiment Analysis of Social Network Messages". *Proceedings of the Eighth International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, pp. 114-118, 2014.
- C8.** A-C. Ngonga Ngomo, N. Heino, R. Speck and P. Malakasiotis, "A Tool Suite for Creating Question Answering Benchmarks". *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, 2014.
- C7.** P. Malakasiotis, R. M. Karampatsis, K. Makrynioti and J. Pavlopoulos, "nlp.cs.aueb.gr: Two Stage Sentiment Analysis". *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, Atlanta, Georgia, USA, pp. 96-106, 2013.
- C6.** P. Malakasiotis and I. Androutsopoulos, "A Generate and Rank Approach to Sentence Paraphrasing". *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2011)*, Edinburgh, UK, pp. 96-106, 2011.
- C5.** P. Malakasiotis, "AUEB at TAC 2009". *Text Analysis Conference (TAC 2009)*, Gaithersburg, Maryland, USA, 2009.
- C4.** P. Malakasiotis, "Paraphrase Recognition Using Machine Learning to Combine Similarity Measures". *Proceedings of the ACL-IJCNLP 2009 Student Research Workshop*, Suntec, Singapore, pp. 27-35, 2009.
- C3.** D. Galanis and P. Malakasiotis, "AUEB at TAC 2008". *Text Analysis Conference (TAC 2008)*, Gaithersburg, Maryland, USA, 2008.
- C2.** P. Malakasiotis and I. Androutsopoulos, "Learning Textual Entailment using SVMs and String Similarity Measures". *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, Prague, Czech Republic, pp. 42-47, 2007.
- C1.** V. Lin, V. Vassalos and P. Malakasiotis, "MiniCount: Efficient Rewriting of COUNT-Queries Using Views". *Proceedings of the 22nd International Conference on Data Engineering (ICDE 2006)*, Atlanta, GA, USA, 2006.

6.3 Conference demos

S1. S. Konstantopoulos, A. Tegos, D. Bilidas, I. Androutsopoulos, G. Lampouras, P. Malakasiotis, C. Matheson and O. Deroo, "Adaptive Natural Language Interaction". System demonstration, *12th Conference of the European Chapter of the Association for Computational Linguistics* (EACL 2009), Athens, Greece, 2009.

7. PARTICIPATION IN RESEARCH PROJECTS

“DICE” – Centre of Excellence which focuses on the development of novel AI methods for business documents.

“ML2P Filter – Machine Learning to Filter Profanity”. Google Digital News Initiative (DNI) project (2016–18). Senior researcher.

“BioQA – Biomedical Question Answering”. Action 2 AUEB Research Support Program (2014-15).

“LangTERRA – Enhancing the Research Potential of ILSP/”Athena” R.C. in Language Technology in the European Research ERA”. European FP7 REGPOT project (2012–15).

“BioASQ – A Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering”. European FP7 ICT project (2012–14).

“INDIGO – Interaction with Personality and Dialogue Enabled Robots”. European IST project (2007–10).

8. OTHER INFORMATION OF SCIENTIFIC ACKNOWLEDGEMENT

8.1 Patents

- E. P. Loukas, E. Spyropoulou, P. Malakasiotis, E. Fergadiotis, I. Chalkidis, I. Androutsopoulos, and G. Paliouras, "System and Method for Automatically Tagging Documents". Patent No: US 2023/0028664 A1, USA, 2023.

8.2 Program Committee in conferences and workshops

- ACL Rolling Review Initiative (reviewer, action editor), 2023
- 60th Annual Meeting of the Association for Computational Linguistics (ACL 2023), 2023
- ACL Rolling Review Initiative (reviewer, action editor), 2022
- Conference on Empirical Methods in Natural Language Processing (EMNLP 2022), 2022
- Natural Legal Language Processing Workshop (NLLP 2022), 2022
- 10th BioASQ workshop, Conference and Labs of the Evaluation Forum (CLEF 2022), 2022
- ACL Rolling Review Initiative (reviewer, action editor), 2021
- Conference on Empirical Methods in Natural Language Processing (EMNLP 2021), 2021
- 9th BioASQ workshop, Conference and Labs of the Evaluation Forum (CLEF 2021), 2021

- Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), 2021
- Deep Learning Inside Out (DeeLIO): The Second Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, 2021.
- Natural Legal Language Processing Workshop (NLLP 2021), 2021
- The 16th Conference of the European Chapter the Association for Computational Linguistics (EACL 2021), 2021
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2021), 2021
- Conference on Empirical Methods in Natural Language Processing (EMNLP 2020), 2020
- 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), Seattle, HPIA, 2020.
- 11th Hellenic Conference on Artificial Intelligence (SETN 2020), Αθήνα, Ελλάδα, 2020 (συν-επικεφαλής της Επιτροπής Προγράμματος).
- Deep Learning Inside Out (DeeLIO): The First Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, 2020.
- 8th BioASQ workshop, Conference and Labs of the Evaluation Forum (CLEF 2020), 2020.
- 2nd Workshop on Natural Legal Language Processing (NLLP), Knowledge Discovery and Data Mining (KDD 2020), 2020.
- 1st Joint Workshop on Narrative Understanding, Storylines, and Events (NUSE), 59th Annual Meeting of the Association for Computational Linguistics (ACL 2020), 2020.
- Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019), Hong Kong, China, 2019.
- 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), Florence, Italy, 2019.
- 3rd Workshop on Abusive Language Online (ALW3), 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), Florence, Italy, 2019.
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL – HLT 2019), Minneapolis, USA, 2019
- 57th Annual Meeting of the Association for Computational Linguistics (ACL 2018), Florence, Italy, 2019.
- 8th Joint Conference on Lexical and Computational Semantics (*SEM 2019), Minneapolis, USA, 2019.
- Conference on Empirical Methods in Natural Language Processing (EMNLP 2018), Brussels, Belgium, 2018.
- 2nd Workshop on Abusive Language Online (ALW2), Conference on Empirical Methods in Natural Language Processing (EMNLP 2018), Brussels, Belgium, 2018.
- 6th BioASQ workshop, Conference on Empirical Methods in Natural Language Processing (EMNLP 2018), Brussels, Belgium, 2018.
- 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018), Melbourne, Australia, 2018.

- 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), Vancouver, Canada, 2017.
- 5th BioASQ workshop (μέρος του BioNLP workshop), 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), Vancouver, Canada, 2017.
- 4th BioASQ workshop (μέρος του BioNLP workshop), 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016), Berlin, Germany, 2016.
- 3rd BioASQ workshop, Conference and Labs of the Evaluation Forum (CLEF 2015), Toulouse, France, 2015.
- 2nd BioASQ workshop, Conference and Labs of the Evaluation Forum (CLEF 2014), Sheffield, UK, 2014.
- 1st BioASQ workshop, post-conference workshop, Conference and Labs of the Evaluation Forum (CLEF 2013), Valencia, Spain, 2013.

8.3 Reviewing in scientific journals

- Natural Language Engineering (NLE)
- Computer Speech & Language
- Language Resources and Evaluation (LREV)

8.4 Citations

- 2671 according to Google Scholar, not excluding self-citations, H-index: 19 (February 2023).

9. TECHNICAL SKILLS

- **Programming languages:** Python (excellent), Java (excellent), C++ (very good), C (very good), R (good)
- **Deep Learning frameworks:** Keras (excellent), Tensorflow (excellent), Pytorch (very good).