

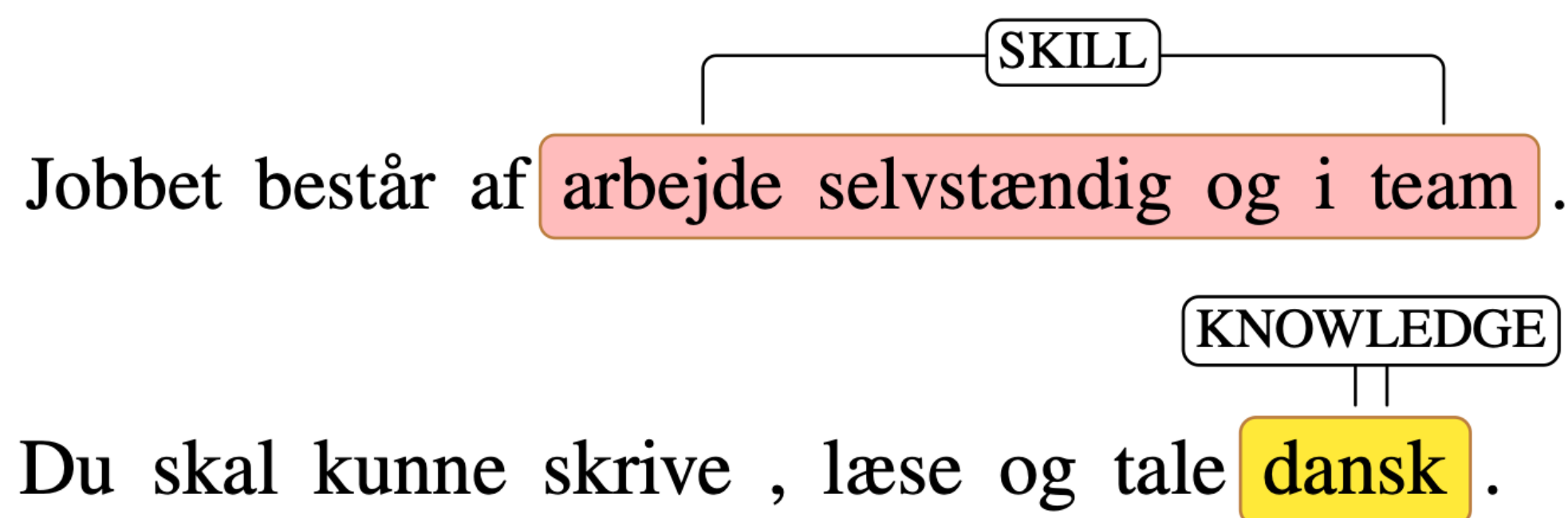
KOMPETENCER: Fine-grained Skill Classification in Danish Job Postings via Distant Supervision and Transfer Learning

Mike Zhang, Kristian Nørgaard Jensen, Barbara Plank
 mikz@itu.dk, krnj@itu.dk, b.plank@lmu.de

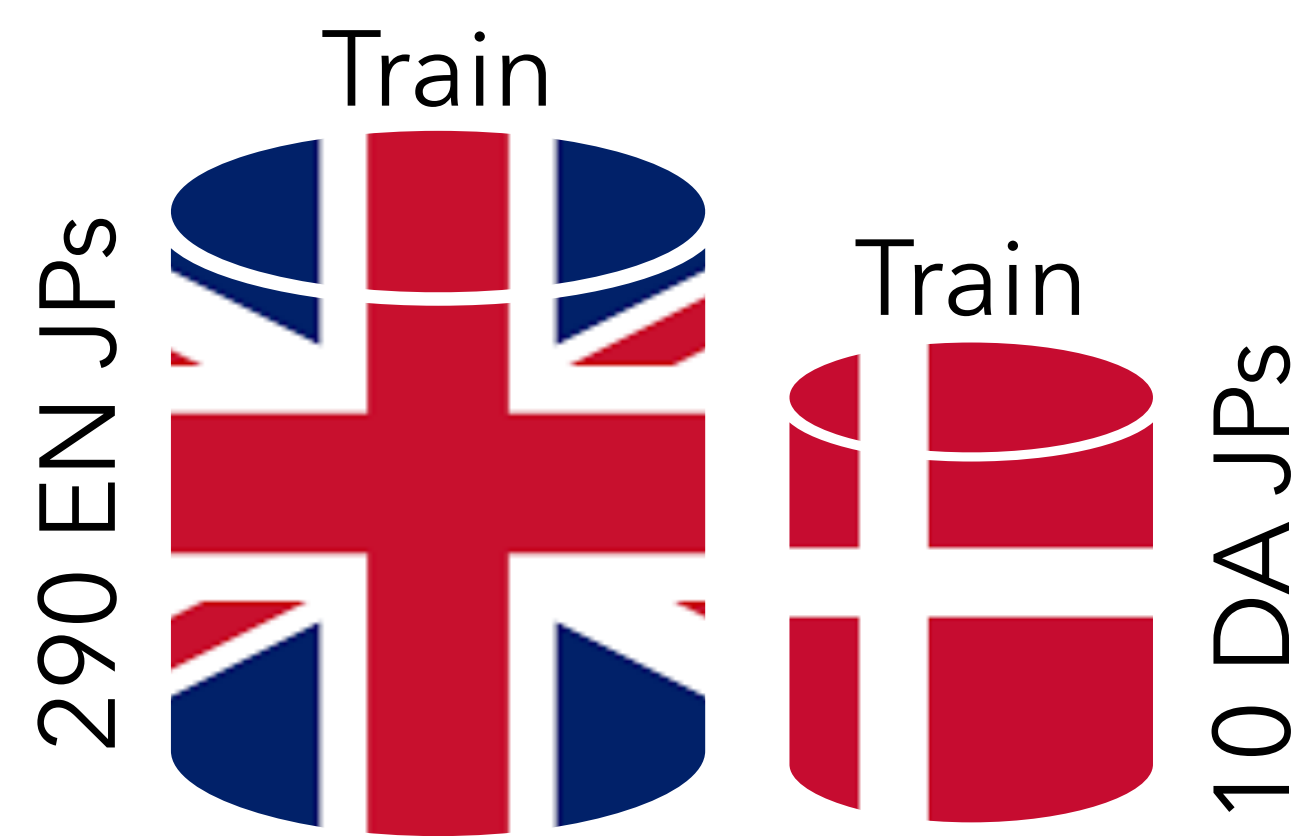
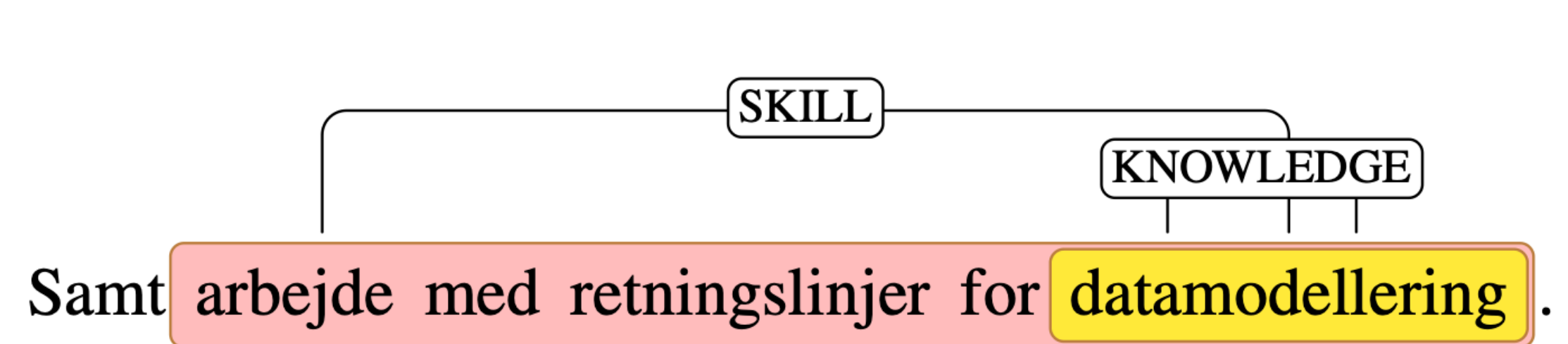
IT-UNIVERSITETET I KØBENHAVN



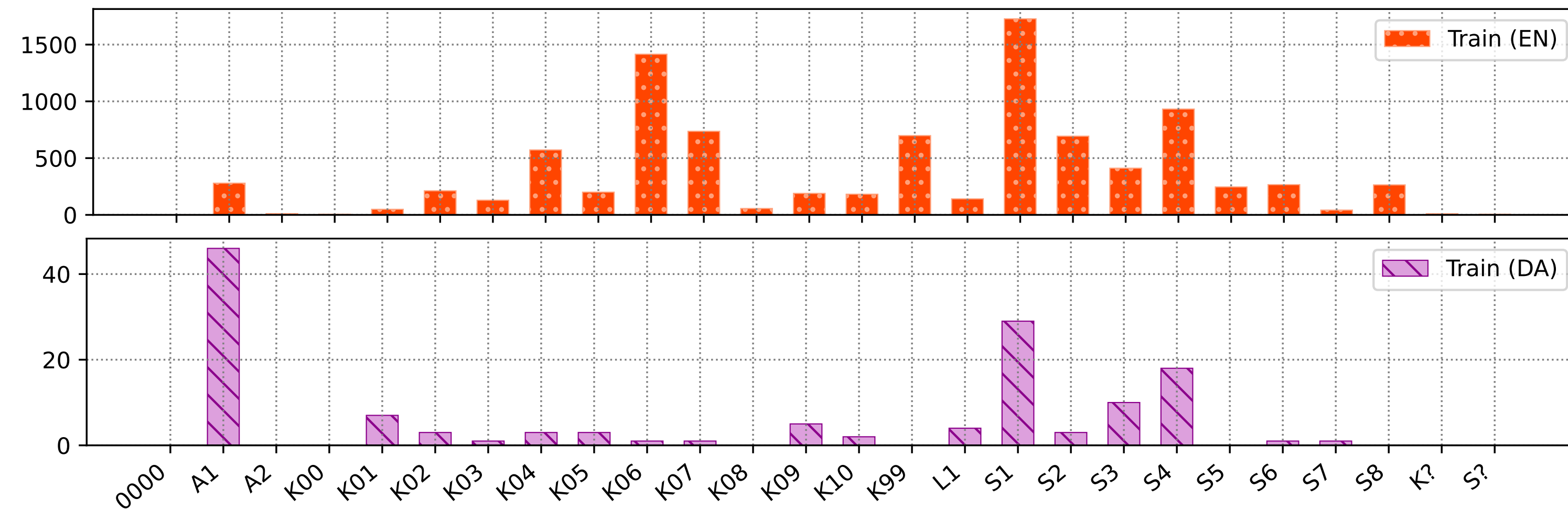
Skill & Knowledge components



Nested Skill & Knowledge components



Label Distribution



Legend

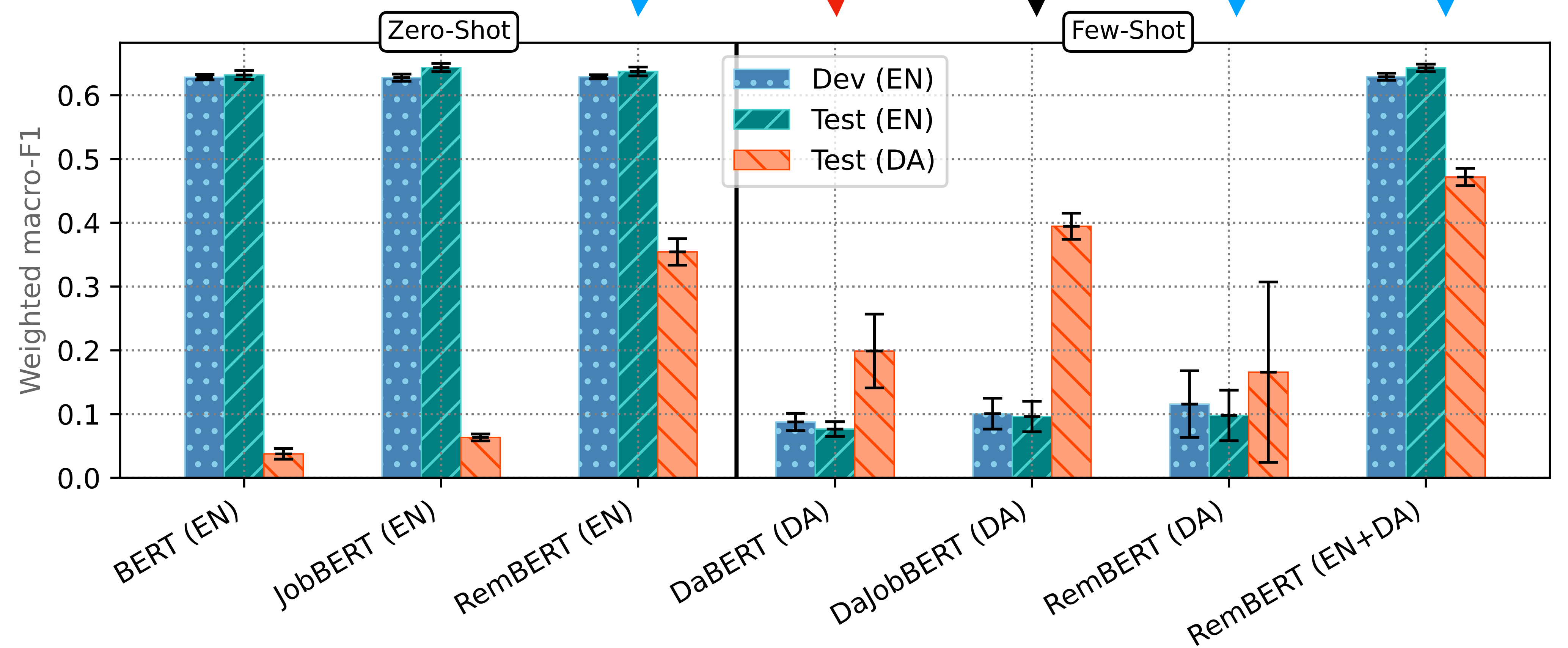
Code	Meaning	Code	Meaning
A1	Attitudes	L1	Languages
A2	Values	S1	Communication, collaboration and creativity
K00	Generic programmes and qualifications	S2	Information Skills
K01	Education	S3	Assisting and caring
K02	Arts and Humanities	S4	Management skills
K03	Social sciences, journalism and information	S5	Working with computers
K04	Business, administration and law	S6	Handling and moving
K05	Natural sciences, mathematics and statistics	S7	Constructing
K06	Information and communication technologies	S8	Working with machinery and special equipment
K08	Engineering, manufacturing and constructing		
K09	Health and welfare		
K10	Services		
K99	Field Unknown		



51 EN JPs Dev. - For model selection

50 EN JPs Test - Zero-shot setup

50 DA JPs - Few-shot setup



Contributions

- Releasing Kompetencer: The **first** Danish job posting dataset with nested annotations for skills and knowledge on the span-level and fine-grained distantly supervised skill codes;
- Empirical analysis of a zero-shot and few-shot classification setup on Danish job ads;
- Releasing a Danish job-ad-adapted BERT on HuggingFace: **DaJobBERT**.



Find the repo here!



[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).
 [2] Zhang, M., Jensen, K. N., Sonniks, S. D., & Plank, B. (2022). SkillSpan: Hard and Soft Skill Extraction from English Job Postings. arXiv preprint arXiv:2204.12811.
 [3] Chung, H. W., Fevry, T., Tsai, H., Johnson, M., & Ruder, S. (2020). Rethinking embedding coupling in pre-trained language models. arXiv preprint arXiv:2010.12821.