

Language Models are Open Knowledge Graphs

Chenguang Wang, Xiao Liu, Dawn Song



Berkeley
UNIVERSITY OF CALIFORNIA

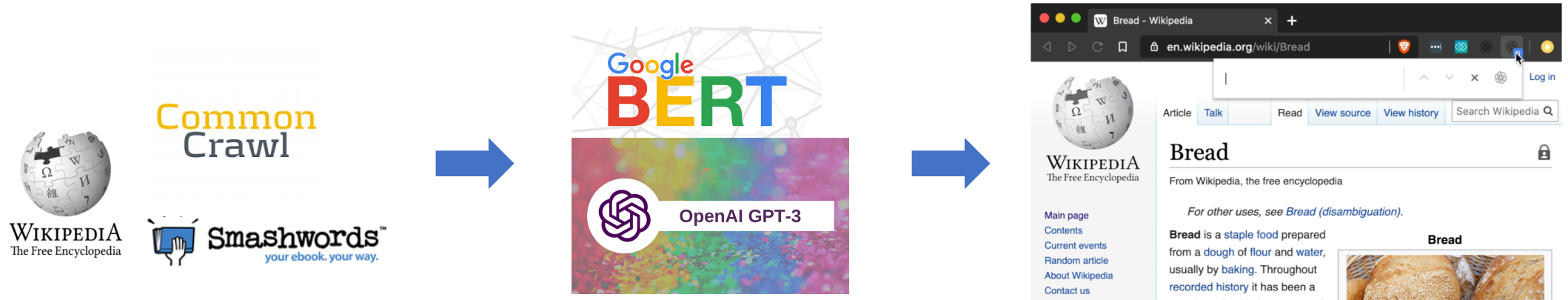


Problem

- Knowledge graph construction requires human supervision

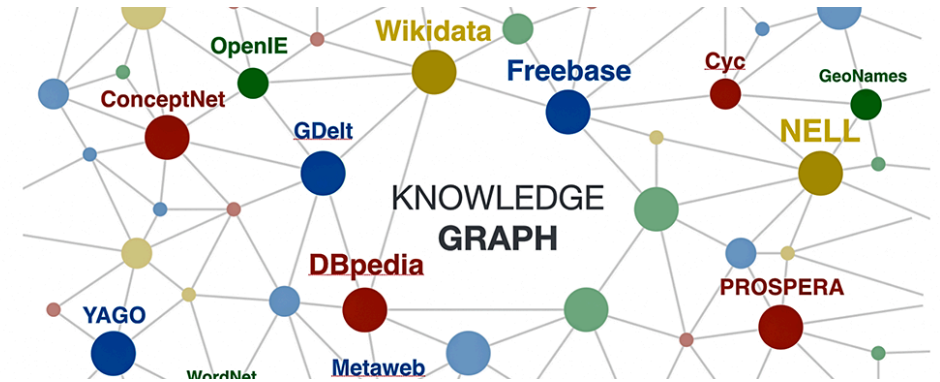
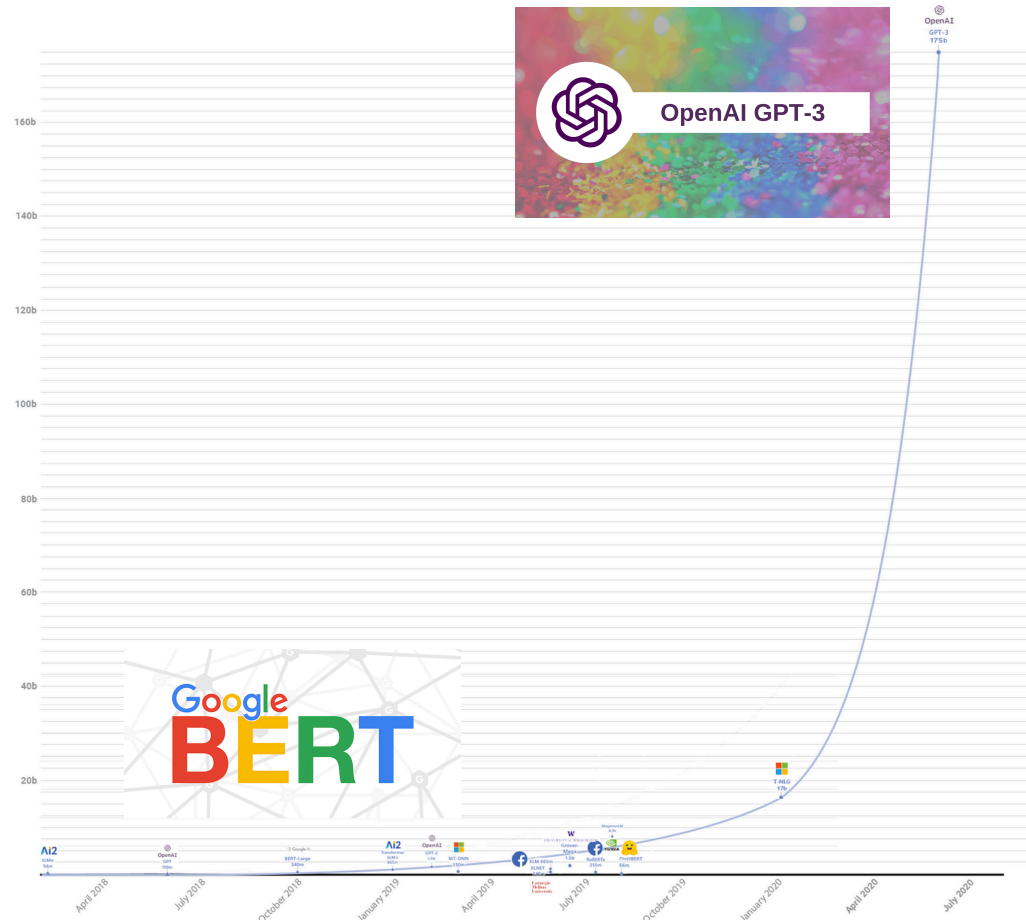


- Language models store knowledge




Problem

- How to use **language models** to construct **knowledge graphs**?



Challenges

Language model 



Target corpus 



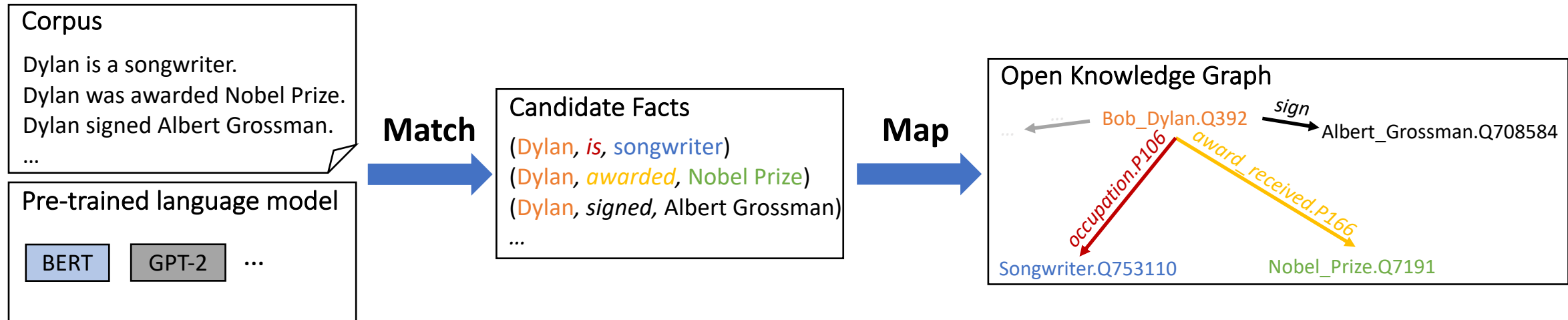
Knowledge graph 

General knowledge

Domain specific knowledge

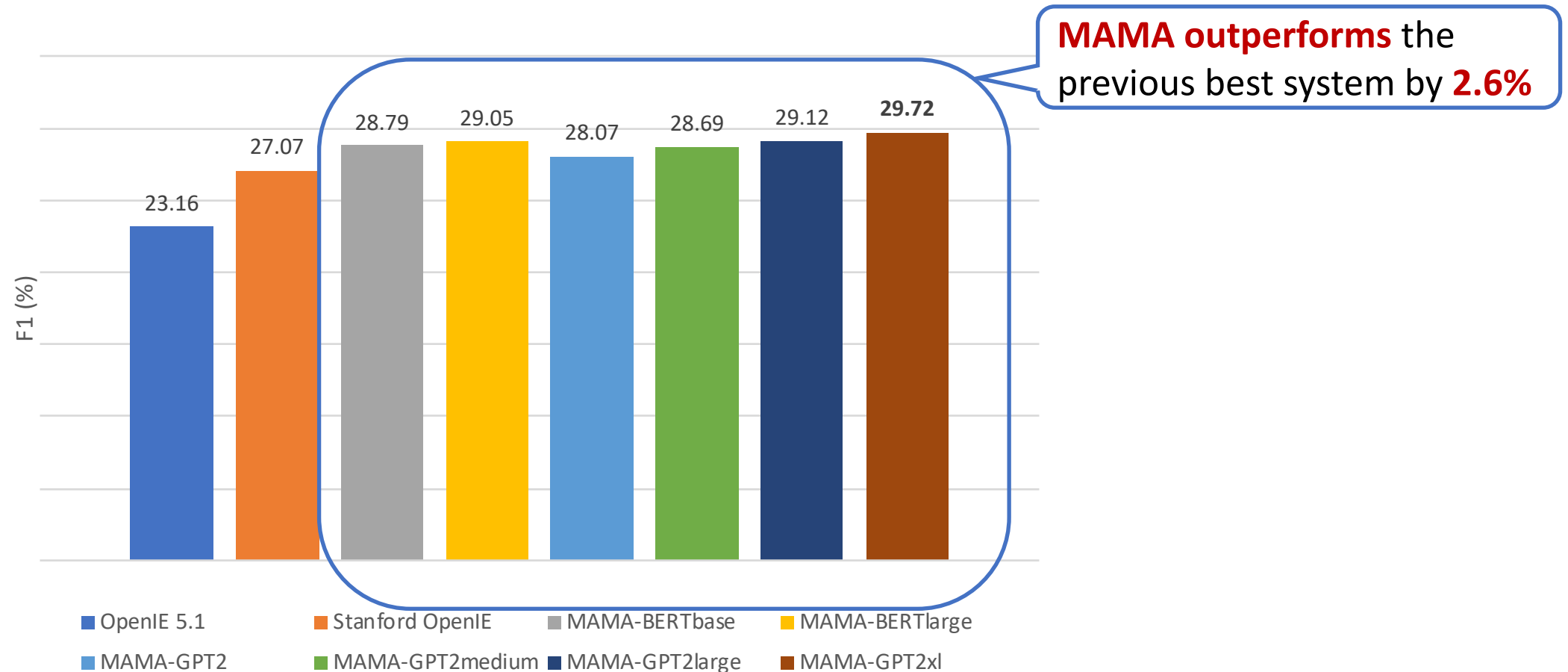
Structured representation

Proposed Approach

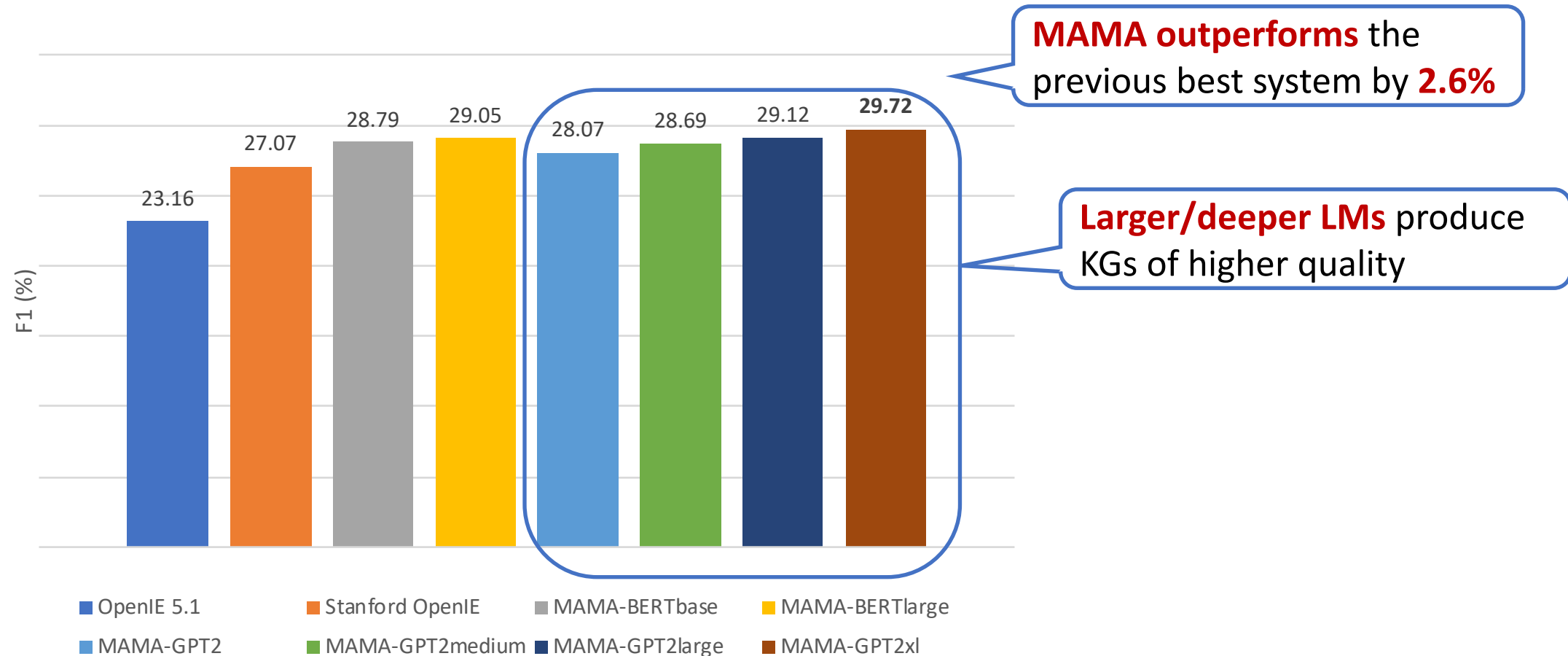


- MAMA constructs an open knowledge graph with **a single forward pass** of the language model (without fine-tuning) over the corpus
 - **Match**: generates a set of candidate facts from a textual corpus
 - **Map**: produces an open knowledge graph from the matched candidates

Results on TAC KBP



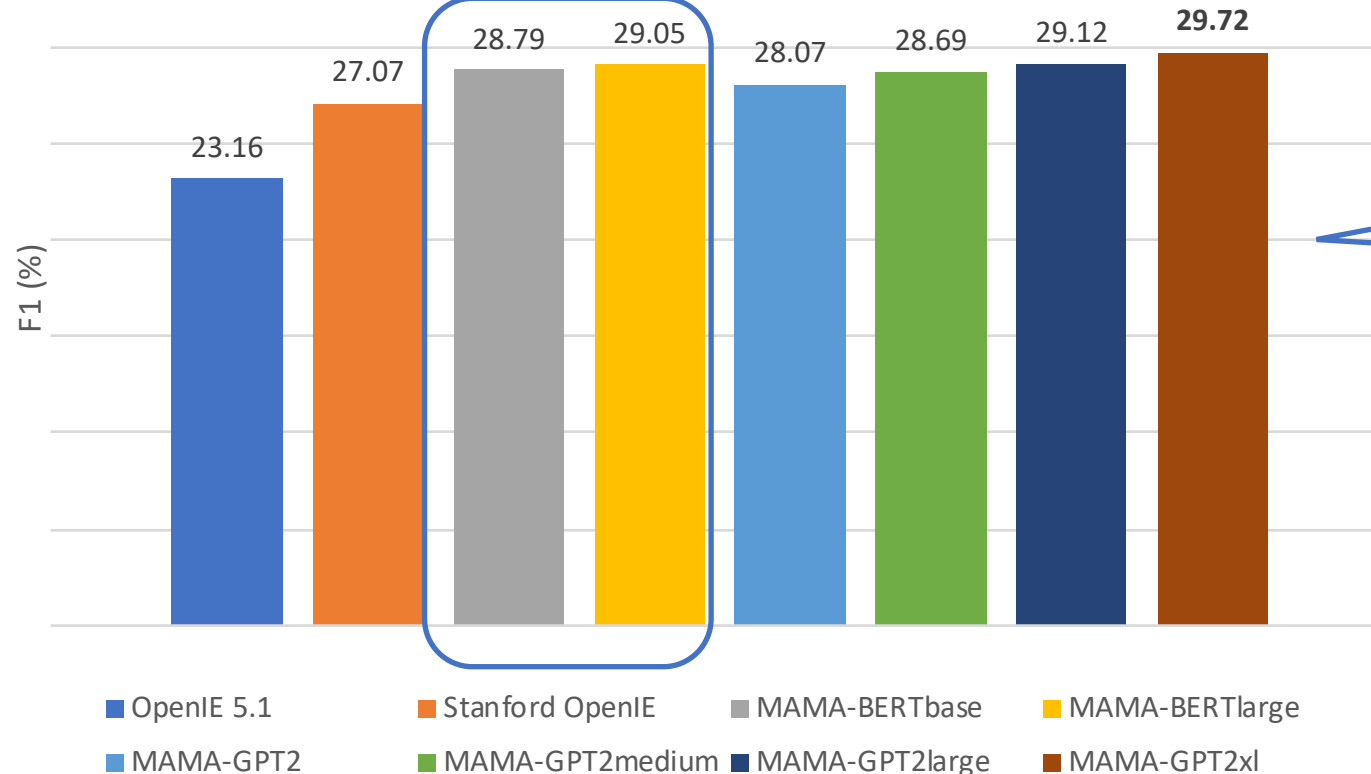
Results on TAC KBP



Results on TAC KBP

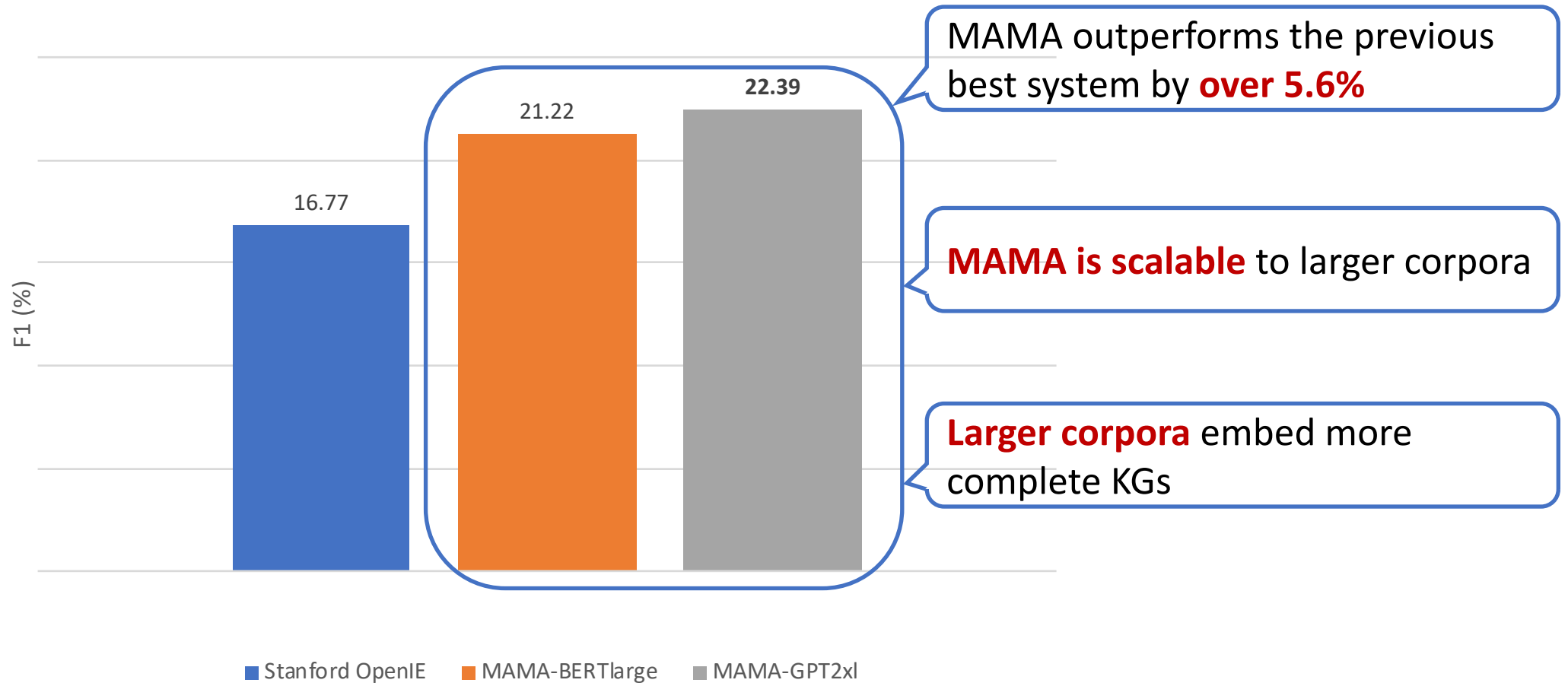
BERT LMs outperform GPT-2 LMs under similar model sizes

MAMA outperforms the previous best system by **2.6%**

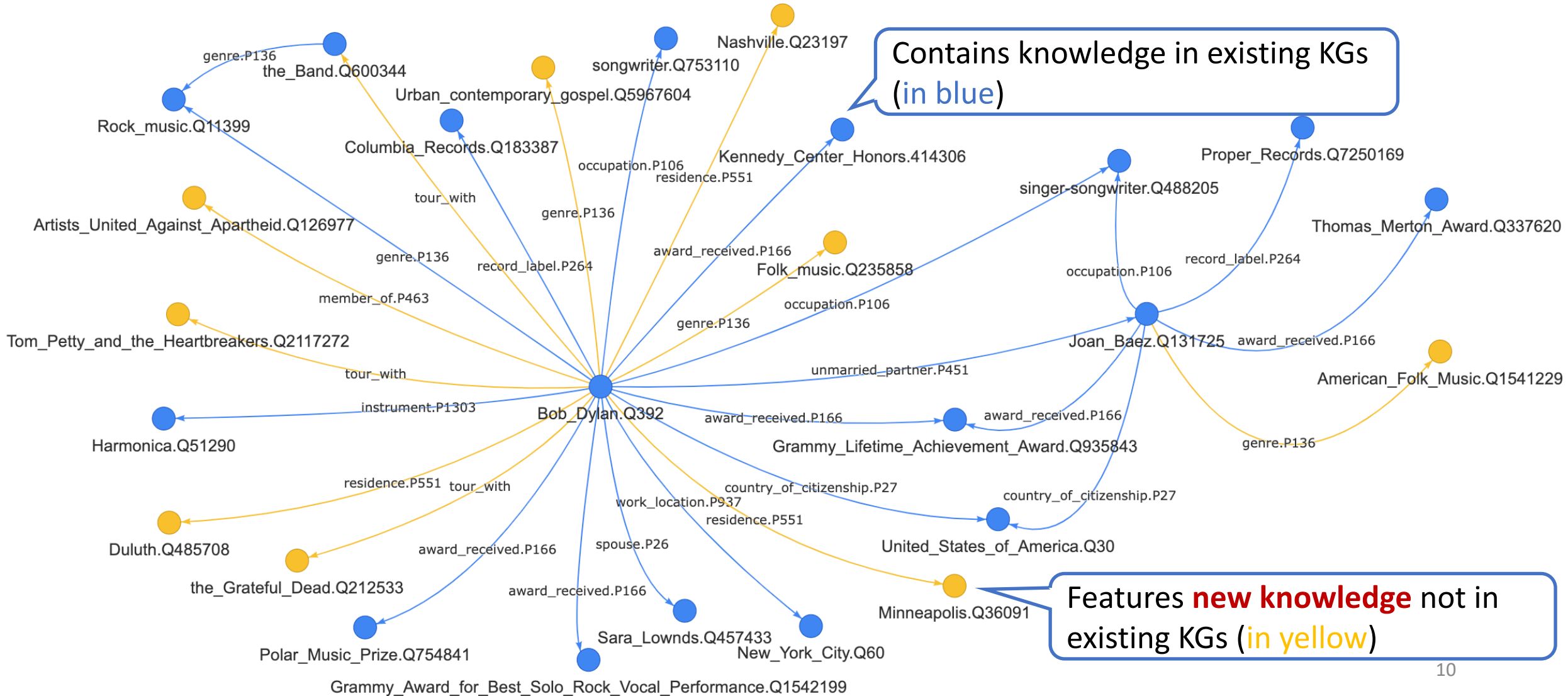


Larger/deeper LMs produce KGs of higher quality

Results on Wikidata



An Open KG Example



Conclusion

- **Problem:** How to construct knowledge graphs from pre-trained language models.
- **Approach:** An unsupervised two-stage approach that constructs knowledge graphs with a single forward pass of the pre-trained language models without fine-tuning over the textual corpora (outperforming compared methods by over 5.6% in F1 on Wikidata).
- **Result:** Open knowledge graphs not only cover the knowledge already in existing knowledge graphs (e.g., Wikidata), but also feature open factual knowledge that is new.