# Neuro-Symbolic Language Modeling with Automaton-augmented Retrieval

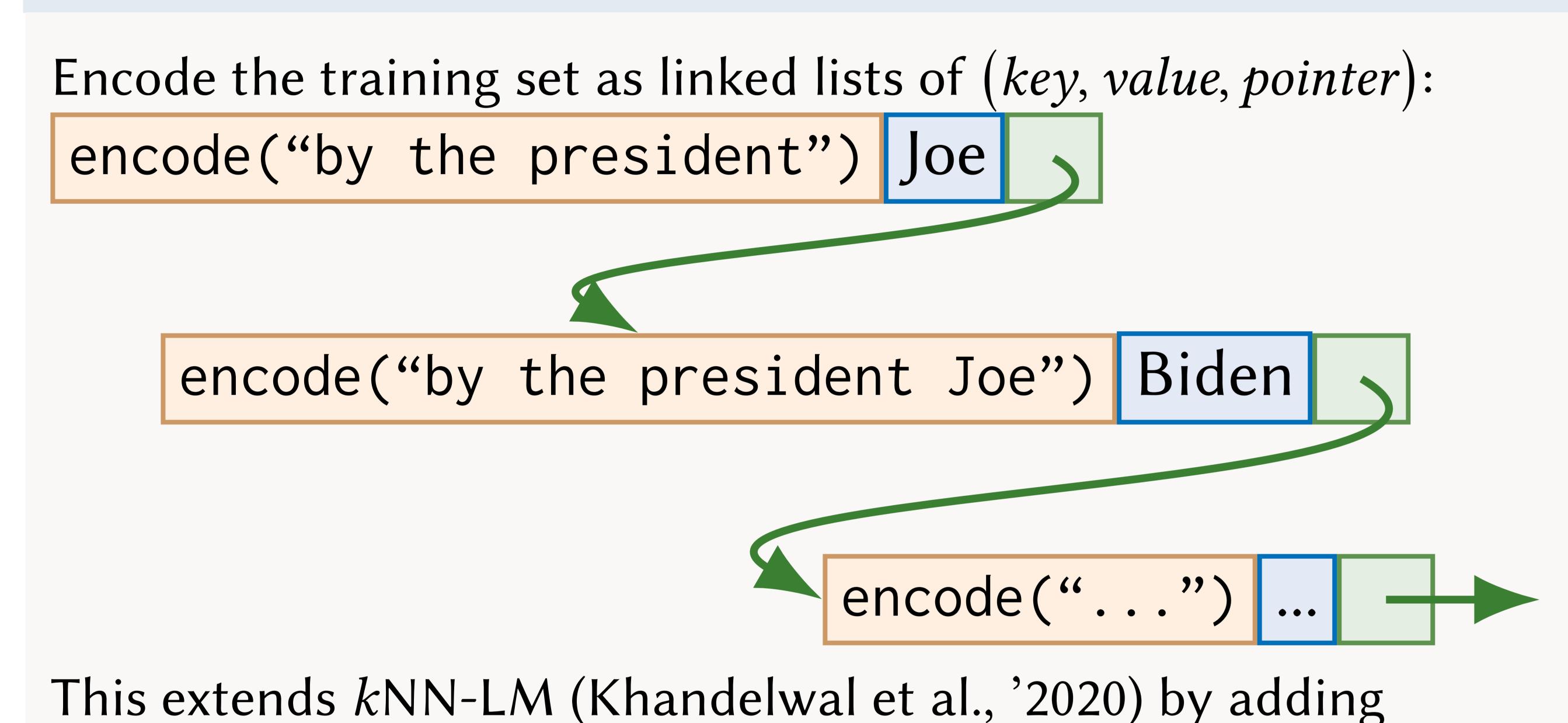




Uri Alon, Frank F. Xu, Junxian He, Sudipta Sengupta, Dan Roth, Graham Neubig https://github.com/neulab/retomaton https://github.com/neulab/knn-transformers



## Key Idea #1: Pointers Between Examples



# Key Idea #2: Clustering Similar Keys

pointers between consecutive datastore entries.

Cluster similar keys into automaton states:

Biden

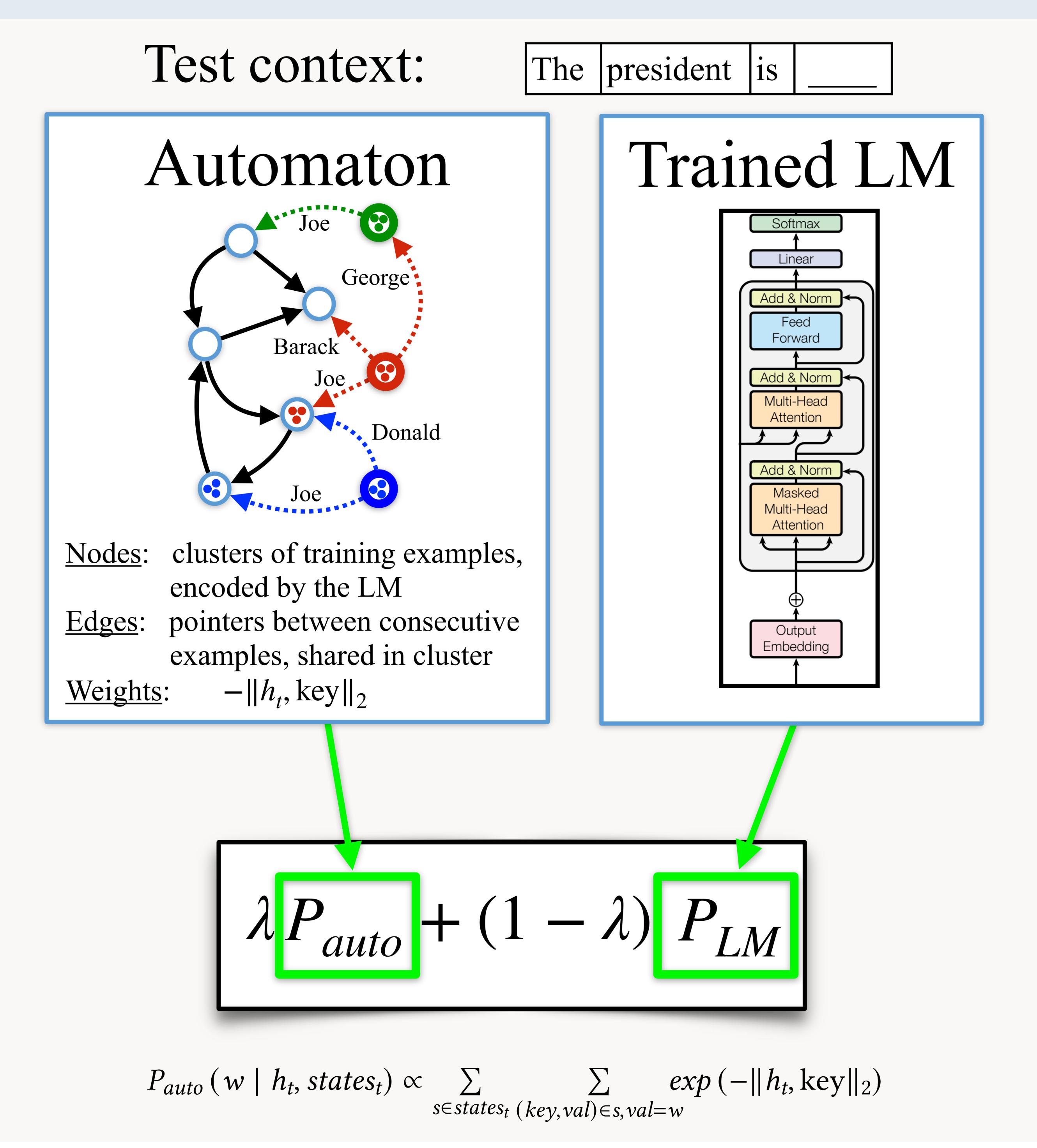
With

President

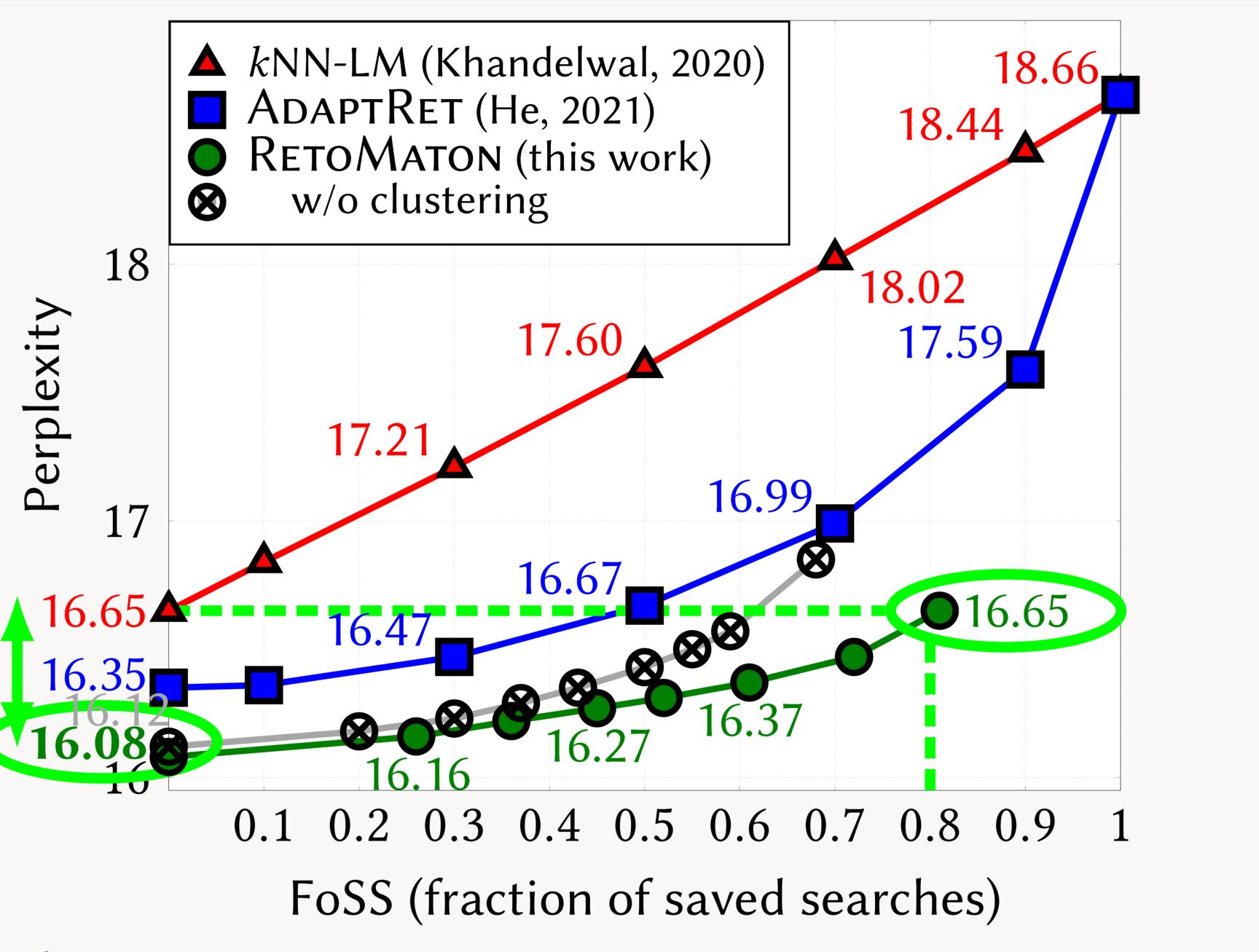
OBiden

Robinette

# RETOMATON

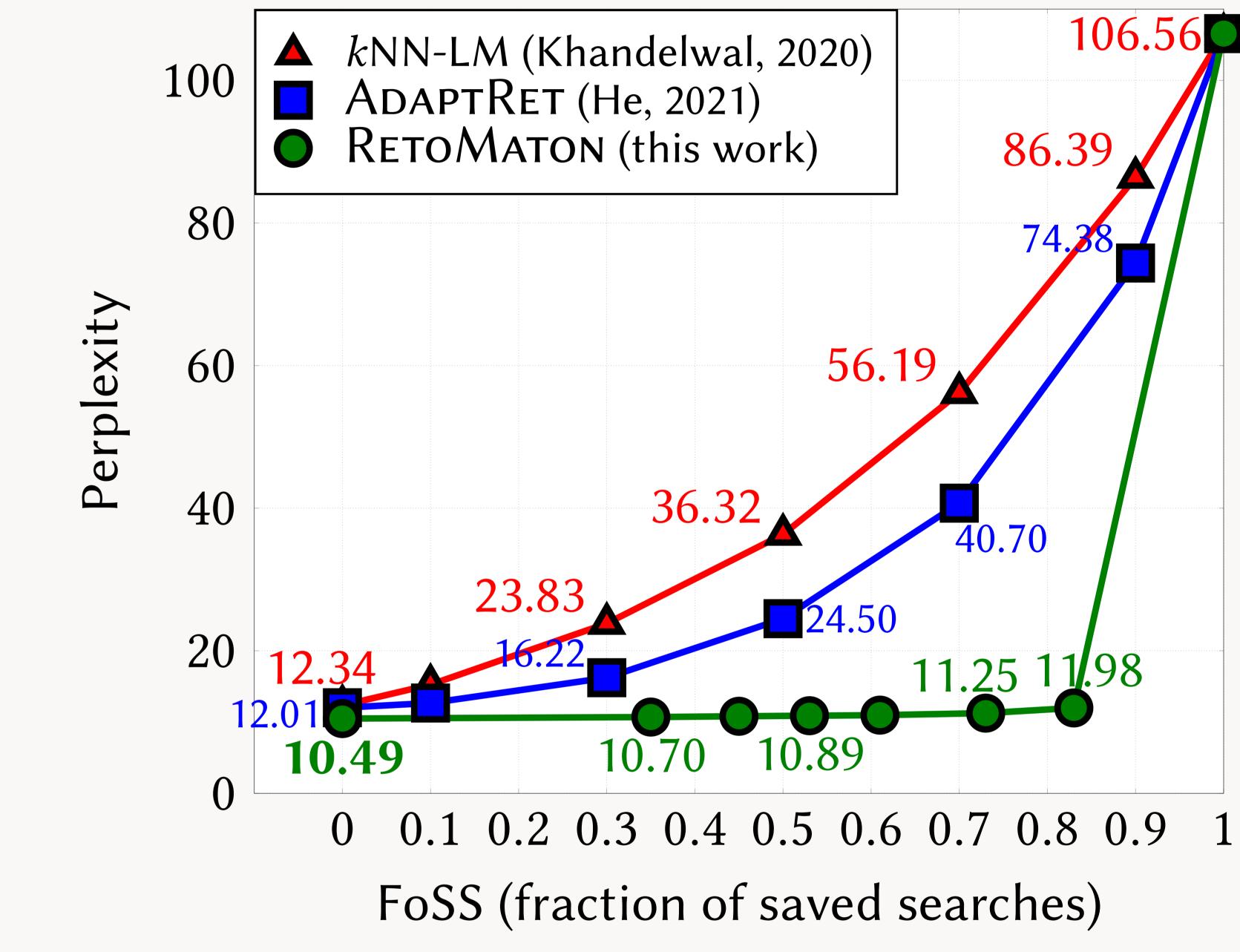


#### In-domain Datastore



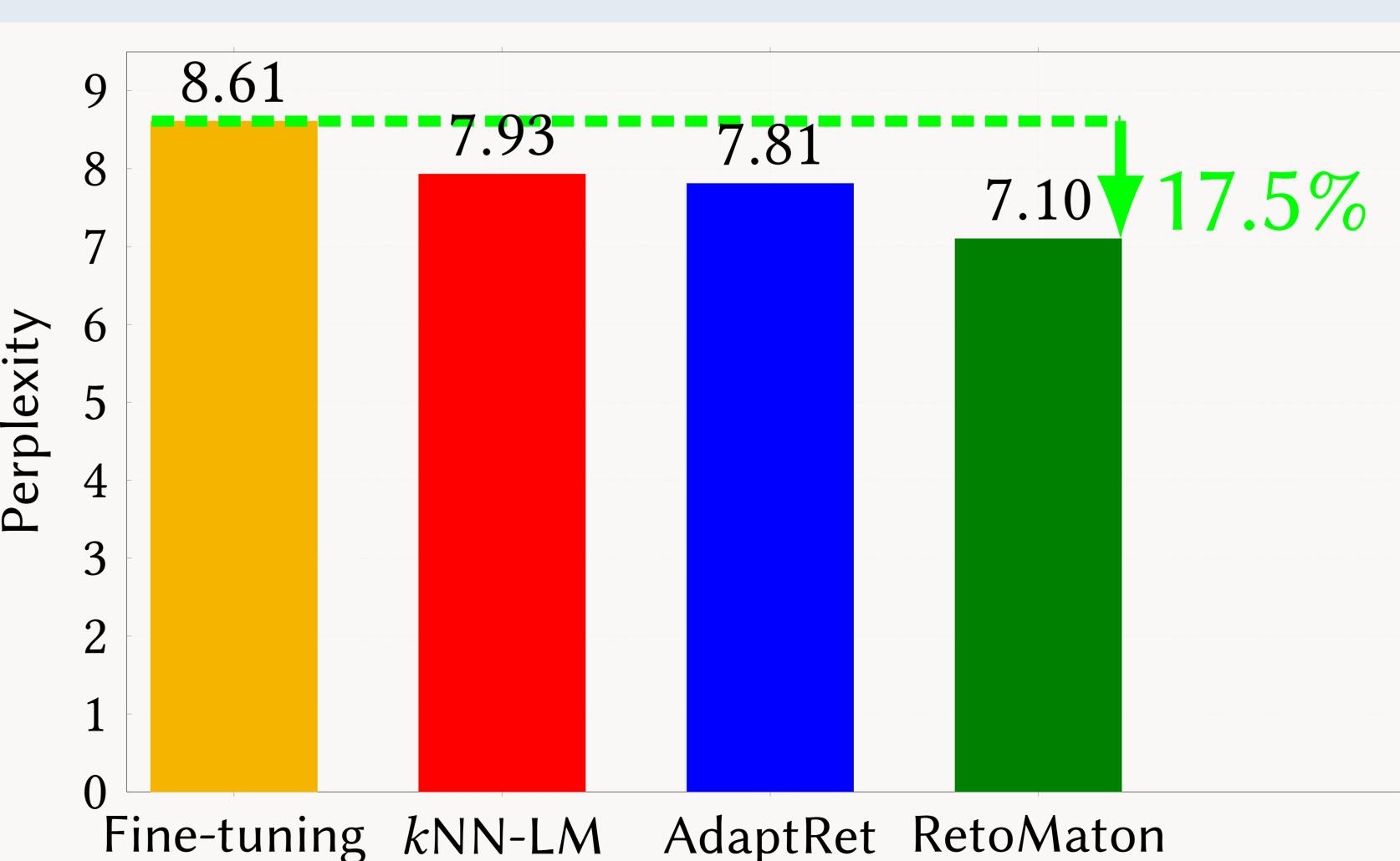
**Figure:** Experiments on WIKITEXT-103, where the datastore is created from the same training set that the base LM was trained on.

#### Domain Adaptation



**Figure:** Domain adaptation experiments: the model was trained on News Crawl, and the datastore is constructed from Law-MT.

## Improving Fine-Tuning



**Figure:** When constructing RETOMATON on top of a fine-tuned model, RETOMATON reduces perplexity by 17.5%.

## Sample



**Figure:** A random sample of the automaton constructed from the training set of Wikitext-103