

# Neuro-Symbolic Language Modeling with Automaton-augmented Retrieval

Appeared in ICML'2022

Uri Alon, Frank F. Xu, Junxian He, Sudipta Sengupta, Dan Roth, Graham Neubig

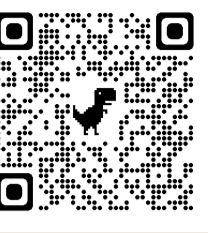
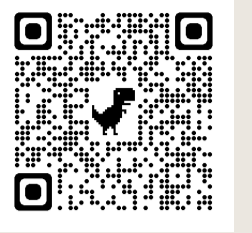


Carnegie Mellon University  
Language Technologies Institute



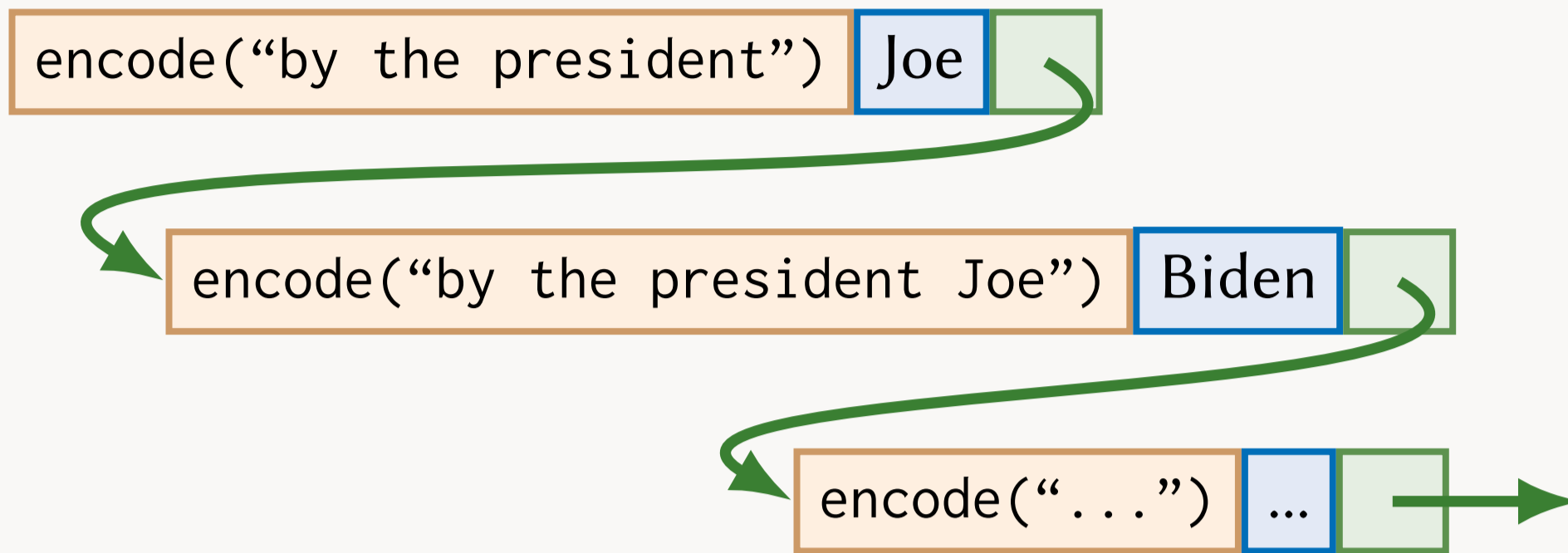
<https://github.com/neulab/retomaton>

<https://github.com/neulab/knn-transformers>



## Key Idea #1: Pointers Between Examples

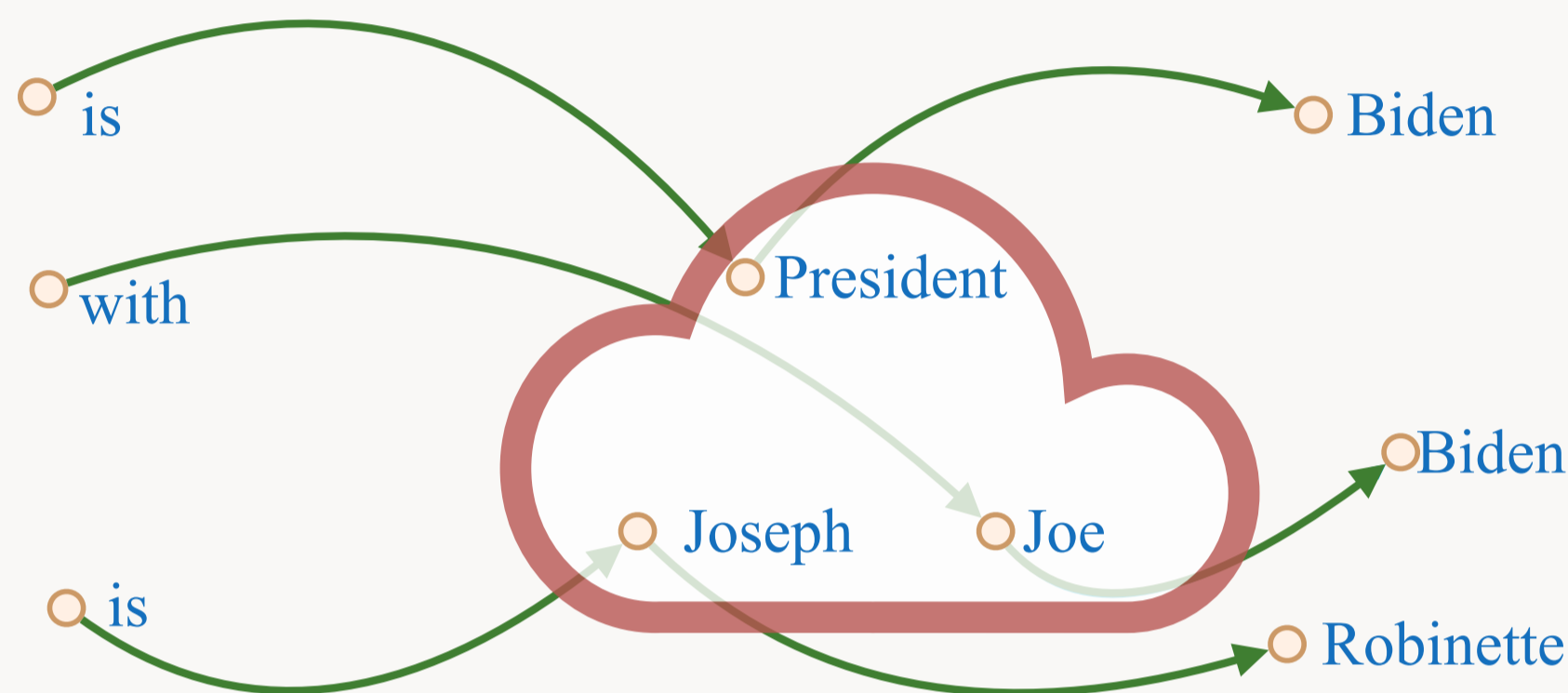
Encode the training set as linked lists of  $\langle \text{key}, \text{value}, \text{pointer} \rangle$ :



This extends  $k$ NN-LM (Khandelwal et al., '2020) by adding *pointers* between consecutive datastore entries.

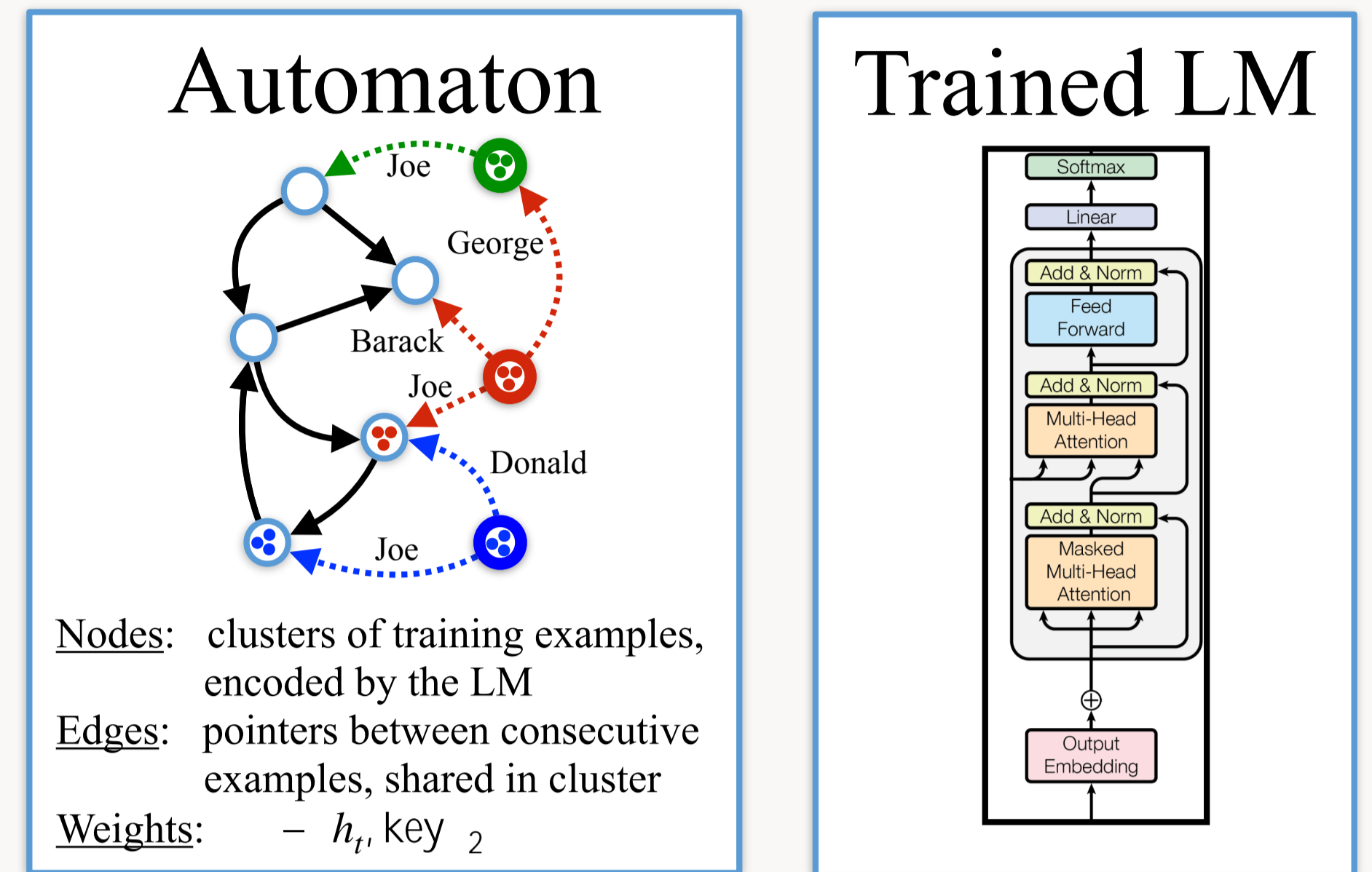
## Key Idea #2: Clustering Similar Keys

Cluster similar keys into automaton states:



## RETOMATON

Test context: The president is \_\_\_\_\_



**Nodes:** clusters of training examples, encoded by the LM

**Edges:** pointers between consecutive examples, shared in cluster

**Weights:**  $-h_t, \text{key}_2$

$$\lambda P_{\text{auto}} + (1 - \lambda) P_{\text{LM}}$$

$$P_{\text{auto}} = \sum_{s_1, s_2} \exp(-\lambda \|h_t - \text{key}_1\| - (1 - \lambda) \|h_t - \text{key}_2\|)$$

## Results

### In-domain Datastore

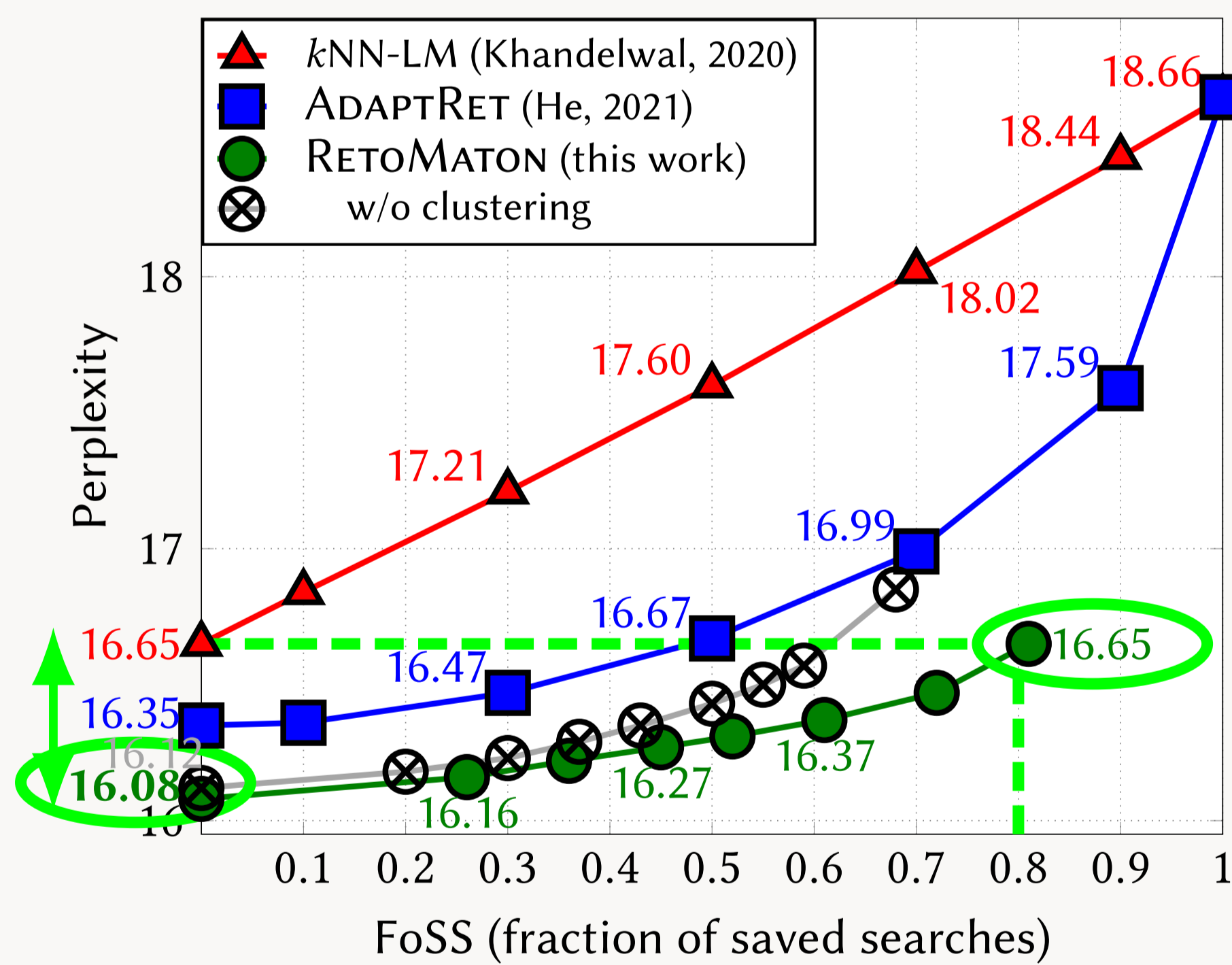


Figure: Experiments on WIKITEXT-103, where the datastore is created from the same training set that the base LM was trained on.

### Domain Adaptation

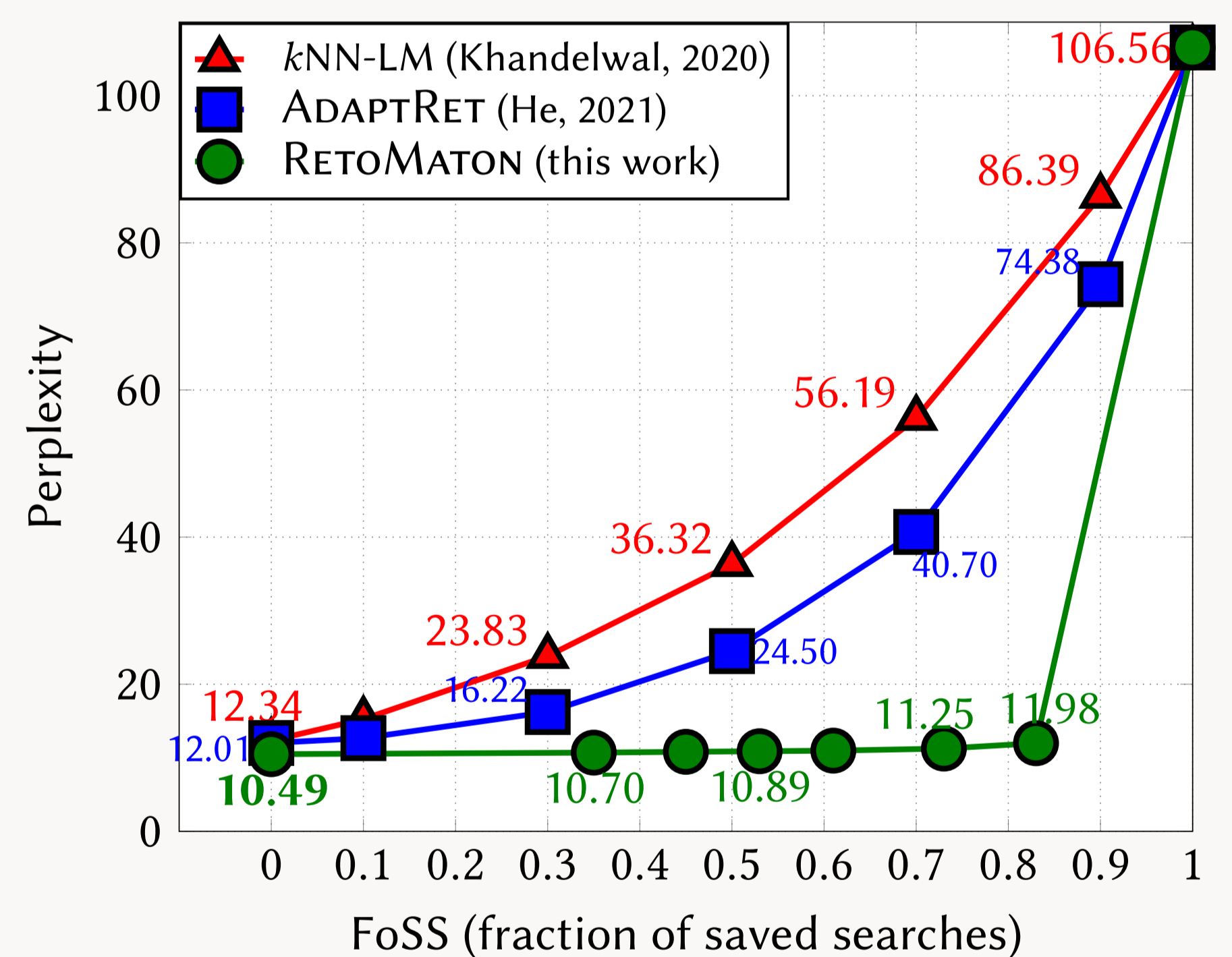


Figure: Domain adaptation experiments: the model was trained on News Crawl, and the datastore is constructed from Law-MT.

### Improving Fine-tuning

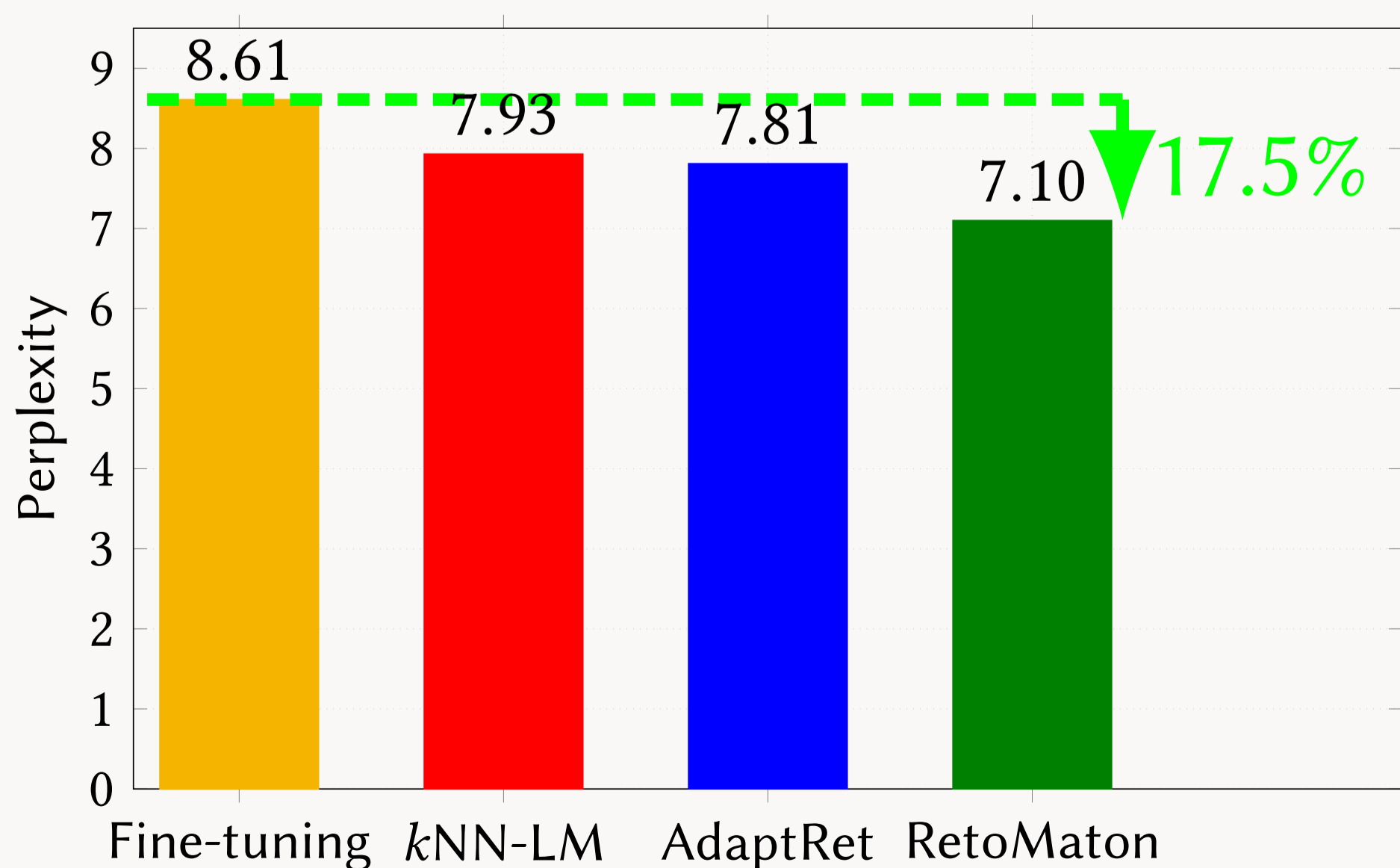


Figure: When constructing RETOMATON on top of a fine-tuned model, RETOMATON reduces perplexity by 17.5%.

### Sample

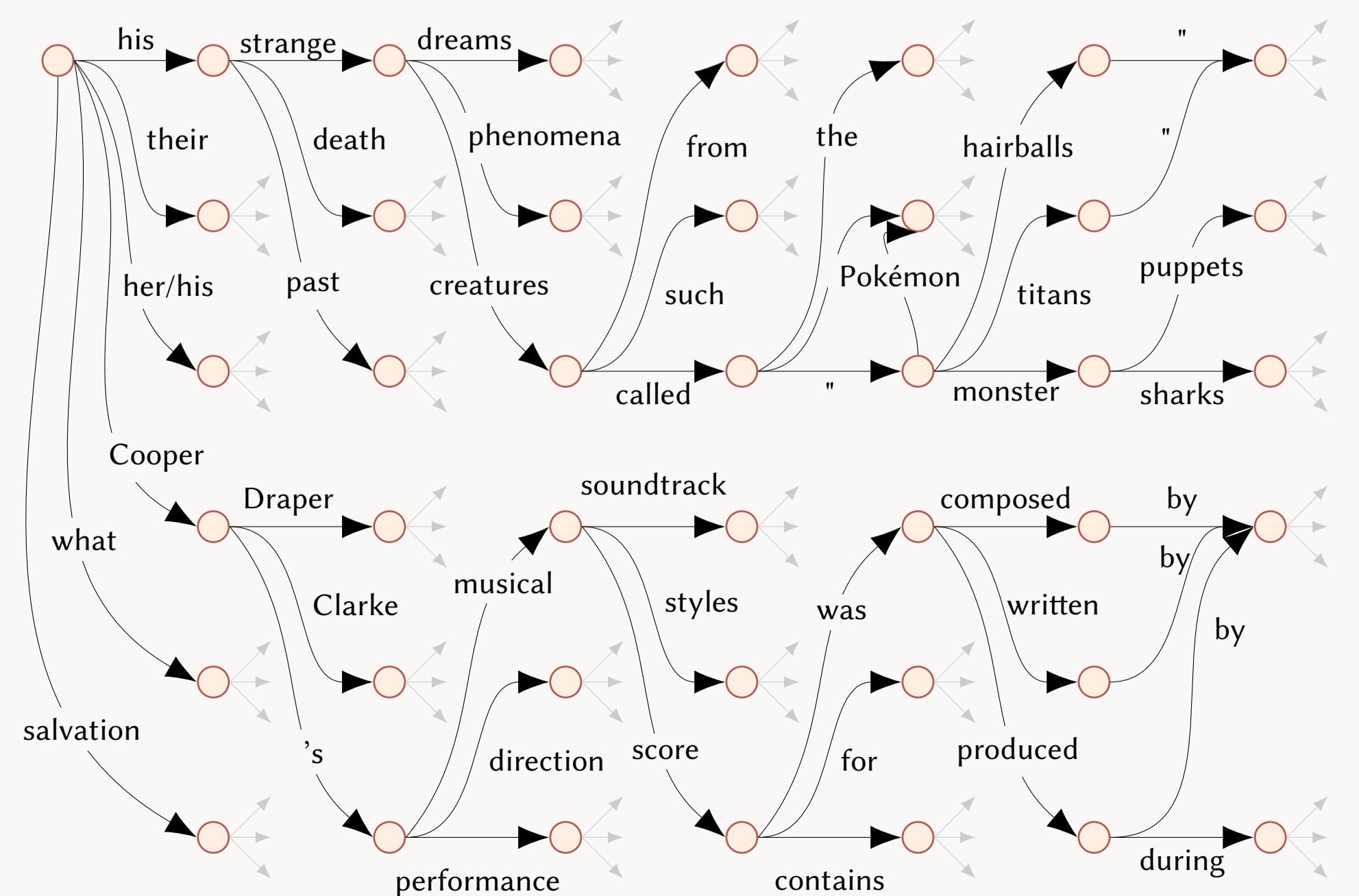


Figure: A sample of the automaton constructed from WIKITEXT-103