Answering Complex Open-domain Questions Through Iterative Query Generation

Peng Qi, Xiaowen Lin*, Leo Mehr*, Zijian Wang*, Christopher D. Manning
Stanford University
(* = Equal Contribution)

👋 Introduction

Example complex open-domain question:

> "How many people does the EMNLP-IJCNLP 2019 venue hold?"

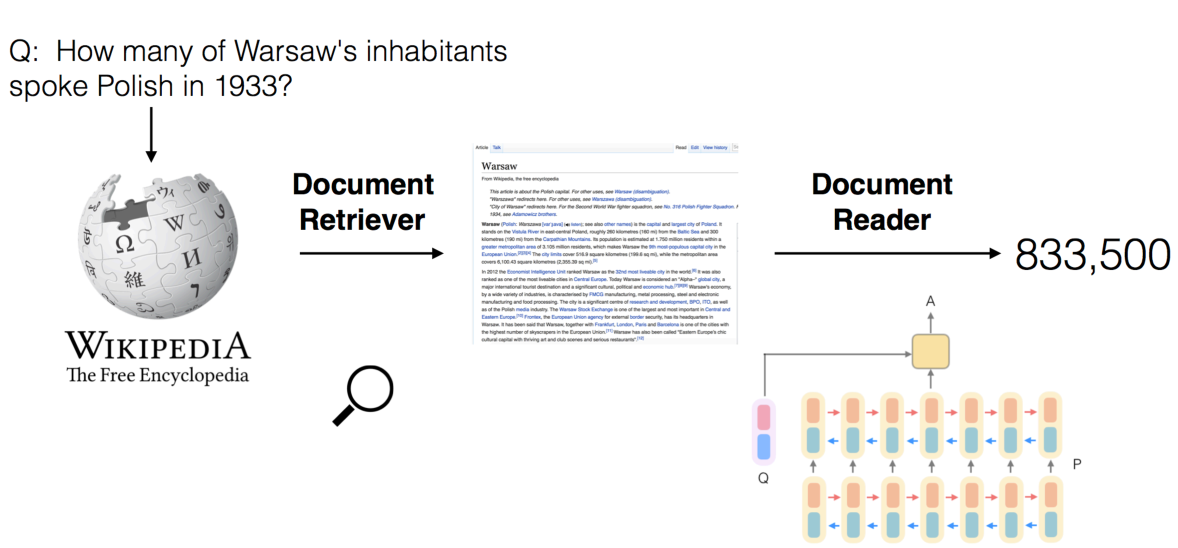
We present

- GoldEn (Gold Entity) Retriever, an *iterative retrieve-and-read* model that performs explainable open-domain multi-step reasoning
- An *efficient* method for training components that generate natural language queries to retrieve supporting facts
- *Competitive performance* without using powerful pretrained neural models like BERT

	Efficient	Multi-hop	Explainable
Retriever & Read	✓	✗	✓
End-to-end	✗	✓?	✗
GoldEn Retriever	✓	✓	✓

📚 Previous Work on Open-domain QA

- One-step retrieve-and-read (e.g., DrQA)



- More end-to-end models that replace the retriever and reader models with neural components that can be optimized jointly

🌐 Language Note

The HotpotQA dataset we tested on is available in the English language only, thus that is the only natural language our experiments are in. But our principle of semantic overlap is applicable to answering open-domain complex questions in other languages than English if suitably augmented with lemmatization for highly inflected languages.

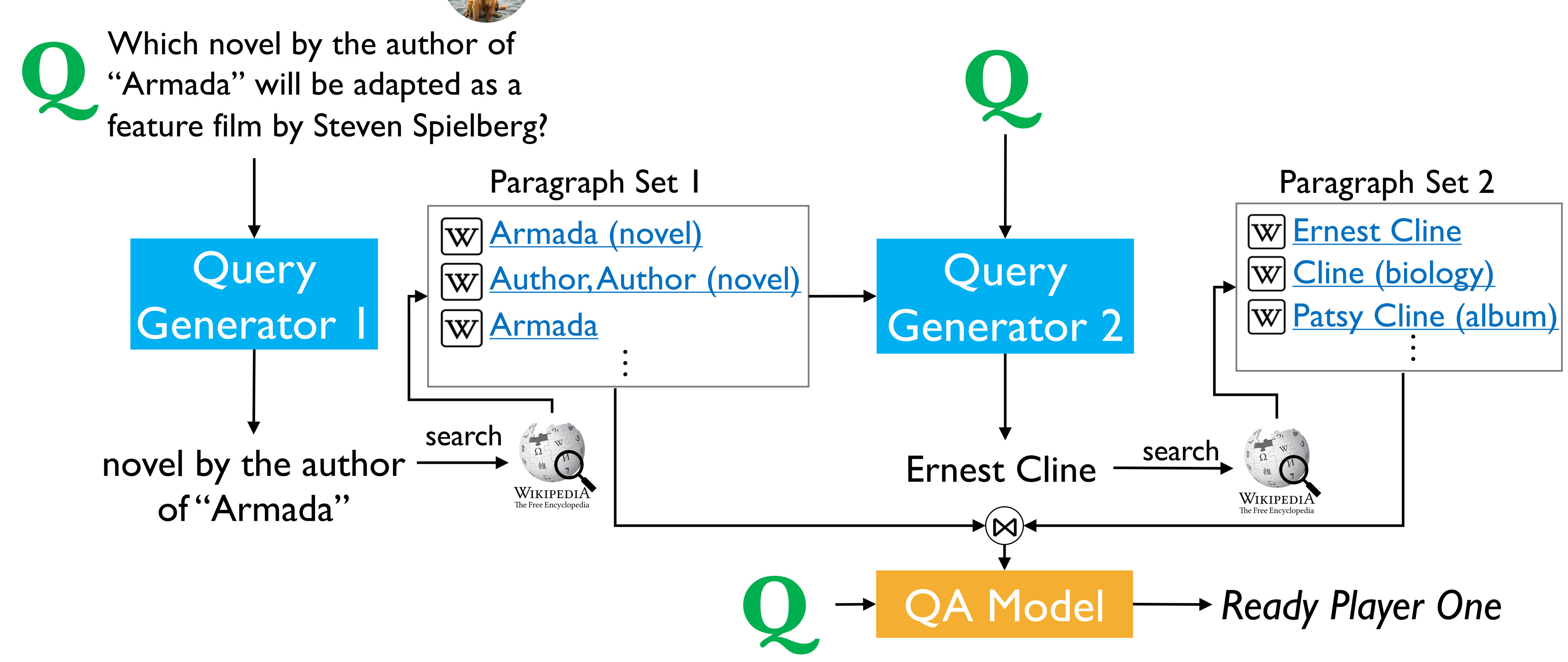
🔍 Where's GoldEn Retriever

- Github repo (scan 👉): https://github.com/qipeng/golden-retriever
- Contains all code, scripts to download models, & prediction files to use with *your own* QA models

🙏 Acknowledgements

This research is funded in part by Samsung Electronics Co., Ltd. and in part by the SAIL-JD Research Initiative.

GoldEn Retriever Model



- Iterates between "reading" and retrieving more evidence to find supporting facts
- Casts query generation as a problem of extractive question answering to reduce search space, then employ the DrQA Document Reader model (explainable!)

$$\textit{Query} = \mathbf{DrQA}(\textit{Question}; \mathbf{concat}([\textit{Question}, \textit{Retrieved documents}]))$$

- Employs modified BiDAF++ as final QA component

🔗 Deriving Supervision to Generate Queries

Challenges

- Human annotations are slow and expensive
- Search engines are not differentiable
- Huge search space

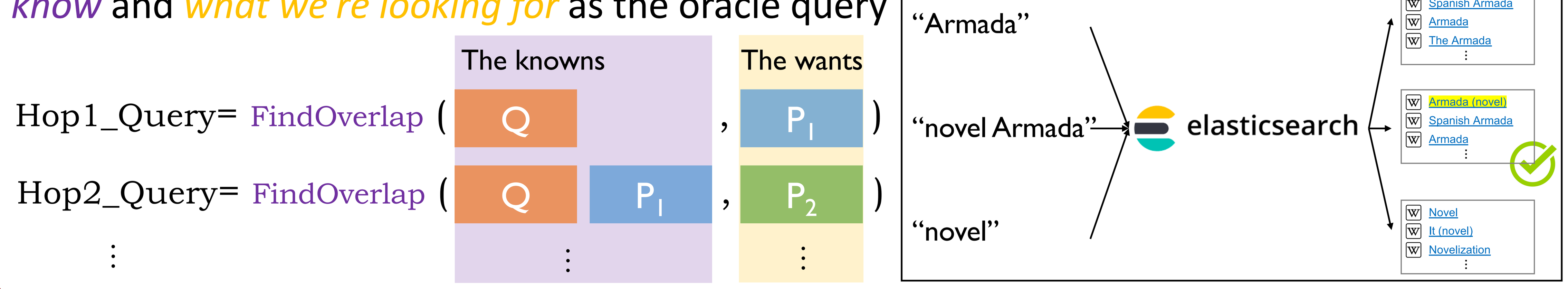
Solution

Derive oracle queries automatically that encode search engine preferences as strong supervision

Main Observation

Supporting facts used to answer a complex question are connected in a chain (or web) of reasoning by *strong semantic overlaps*, which can be used to find them efficiently.

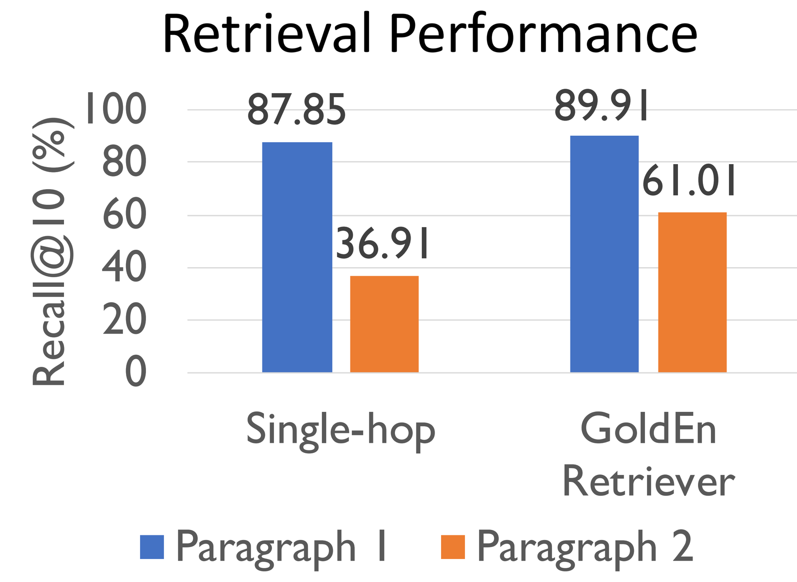
We find this semantic overlap between *what we know* and *what we're looking for* as the oracle query

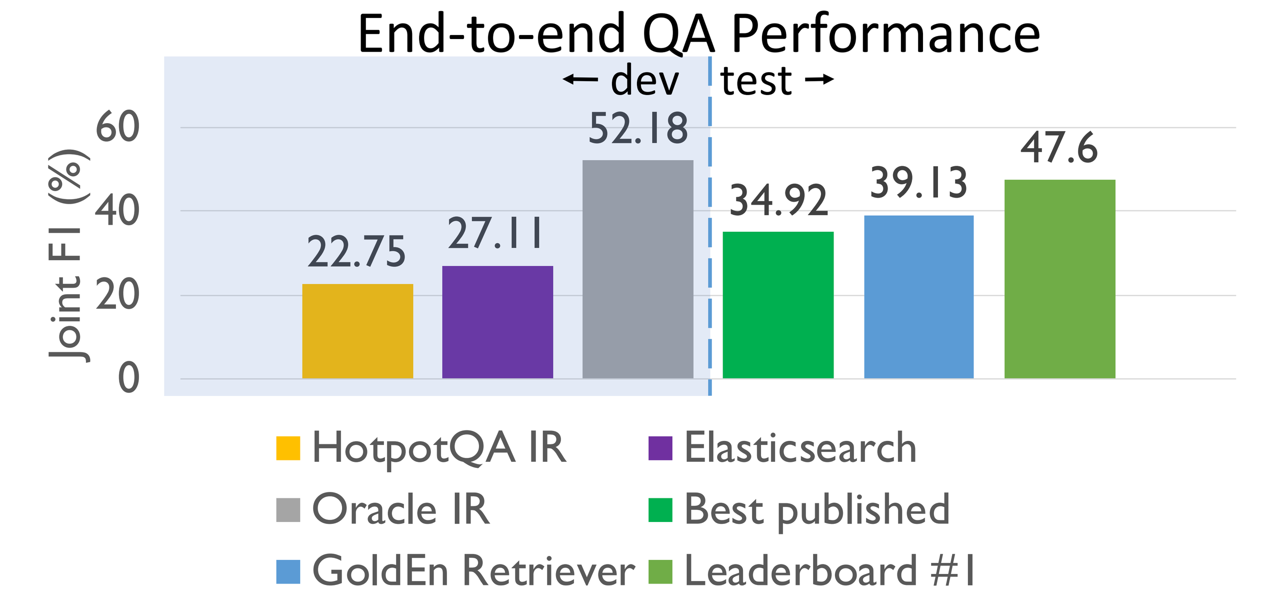


🧪 Experimental Results

We test our GoldEn Retriever system on the HotpotQA dataset, which contains ~113k questions requiring two Wikipedia articles to answer. We focus on the *fullwiki* setting, where the QA system is given a question and ~5 million Wikipedia articles to answer from.

We retrieve 10 paragraphs in total for each question (5 in each retrieval step).





Question	Hop 1 Predicted	Hop 2 Predicted
What video game character did the voice actress in the animated film Alpha and Omega voice?	voice actress in the animated film Alpha and Omega *(animated film Alpha and Omega voice)*	Hayden Panettiere
Yau Ma Tei North is a district of a city with how many citizens?	Yau Ma Tei North	Yau Tsim Mong District of Hong Kong *(Hong Kong)*