ISEEQ: Information SEEking Question generation using Dynamic Meta-Information Retrieval and Knowledge Graphs

Manas Gaur[ŧ], Kalpa Gunaratna[§], Vijay Srinivasan[§], Hongxia Jin[§]

[ŧ]Artificial Intelligence Institute, University of South Carolina, Columbia, USA

[§]Samsung Research America, Mountain View, California, USA

AAAI-22

SAMSUNG

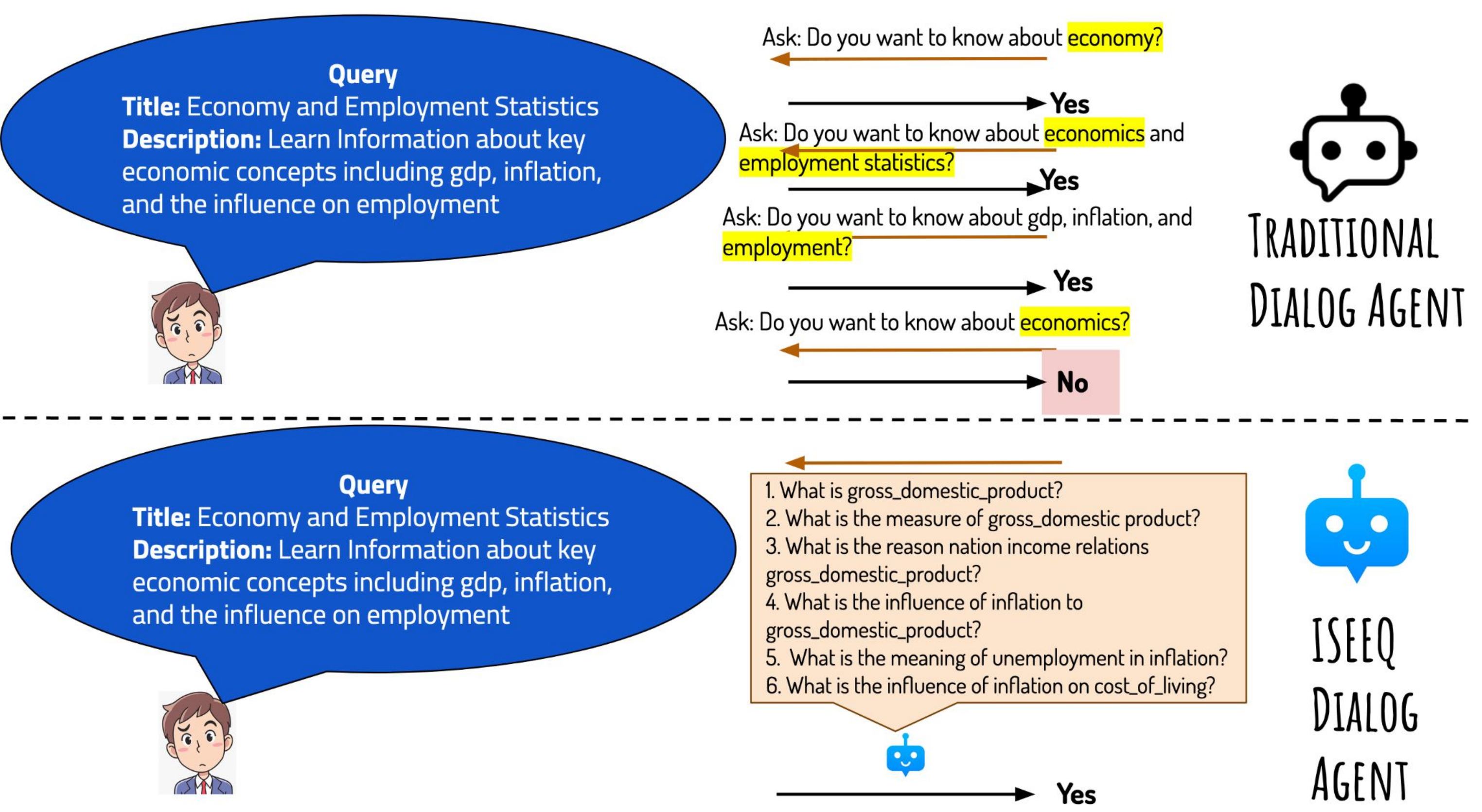
AI INSTITUTE #AIISC UNIVERSITY OF SOUTH CAROLINA

Question, Challenges, & Datasets

Research Question: Human information Seeking Behavior Can be learnt by a dialog agent to Support User Engagement and Shape a cohesive Response

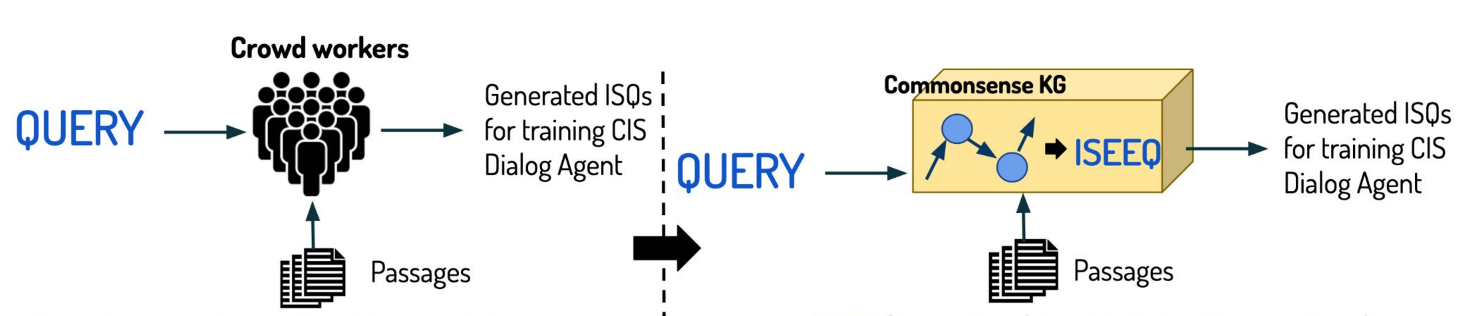
Information Seeking Questions (ISQs): ISQs differ from other question type (e.g Clarifying questions, Follow-up questions) by having a

- Structure : semantic relations between questions and logical coherence
- Cover objective details
- Expand on the breath of topic



I. Challenges in Traditional Dialog Agents in Conversational Information Seeking (CIS)

- Question Generation process lack curiosity in agent to learn more about the entities in the User Query
- Agent generated questions are redundant and lack diversity
- Multi-turn conversation often result in irrelevant question generation
- It is difficult to maintain flow of information when agent is generating questions at random.



Current process for creating dataset to train a Conversational Information Seeking **(CIS)** agent

ISEEQ for creating dataset to train a Conversational Information Seeking **(CIS)** agent

II. Challenges in Creating Datasets for CIS agents

- Tremendous amount of annotation effort
- Crowdworkers have to:
  - Search the Web
  - Creating good quality questions
  - Response Shaping
  - Maintaining the information flow
- Question from crowd workers **have curiosity** which is lacking in CIS agen's generated question
- **Need a data creator agent to assist annotators**

Benchmark Datasets

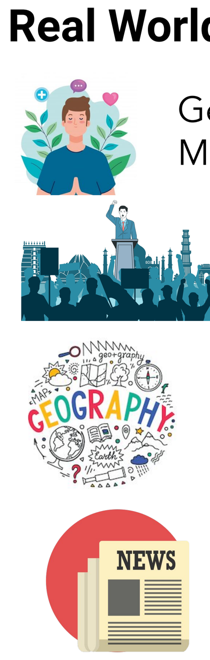
QADiscourse (QAD) (Pyatkin et al. 2020)
- Source for Passages: Wikipedia and WikiNews
- Training Samples: 125 User Queries with 25 ISQs per Query (125 * 25 = 3,125 Query-Question Pair)
- Testing Samples: 33 User Queries with 25 ISQs per Query
- ConceptNet KG hit percentage: 38.5%

Question Answer Meaning Representation (QAMR) (Michael et al. 2018)
- Source for Passages: WikiNews
- Training Samples: 395 User Queries with 63 ISQs per Query
- Testing Samples: 39 User Queries with 68 ISQs per Query
- ConceptNet KG hit percentage: 35.5%

Facebook Curiosity (FBC) (Rodriguez et al. 2020)
- Source for Passages: Geographic Wikipedia
- Training Samples: 8489 User Queries with 6 ISQs per Query
- Testing Samples: 2729 User Queries with 8 ISQs per Query
- ConceptNet KG hit percentage: 50%

Conversational Assistance Track Dataset (CAsT-19) (Dalton et al. 2020)
(Dataset only to test ISEEQ, train and test merged)
- Source for Passages: Microsoft MARCO (https://microsoft.github.io/msmarco/)
- Training Samples: 30 User Queries with 9 ISQs per Query
- Testing Samples: 50 User Queries with 10 ISQs per Query
- ConceptNet KG hit percentage: 57%

Real World Datasets



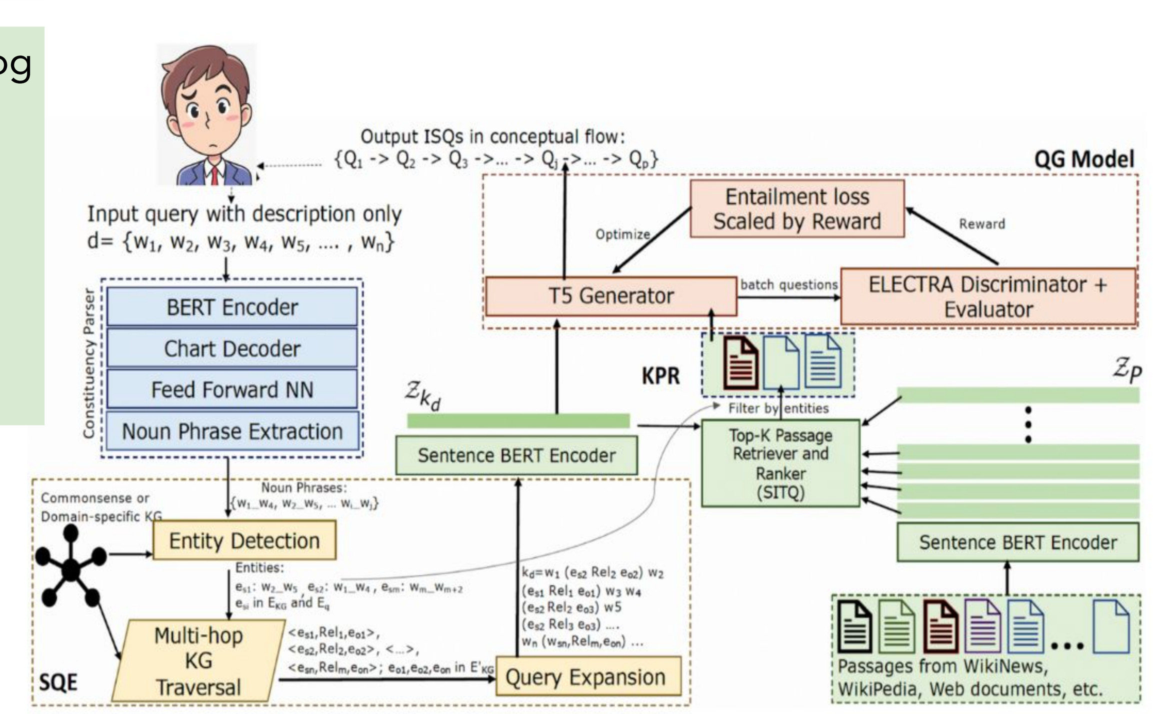
General Health and Mental Health

Public Policy and Politics

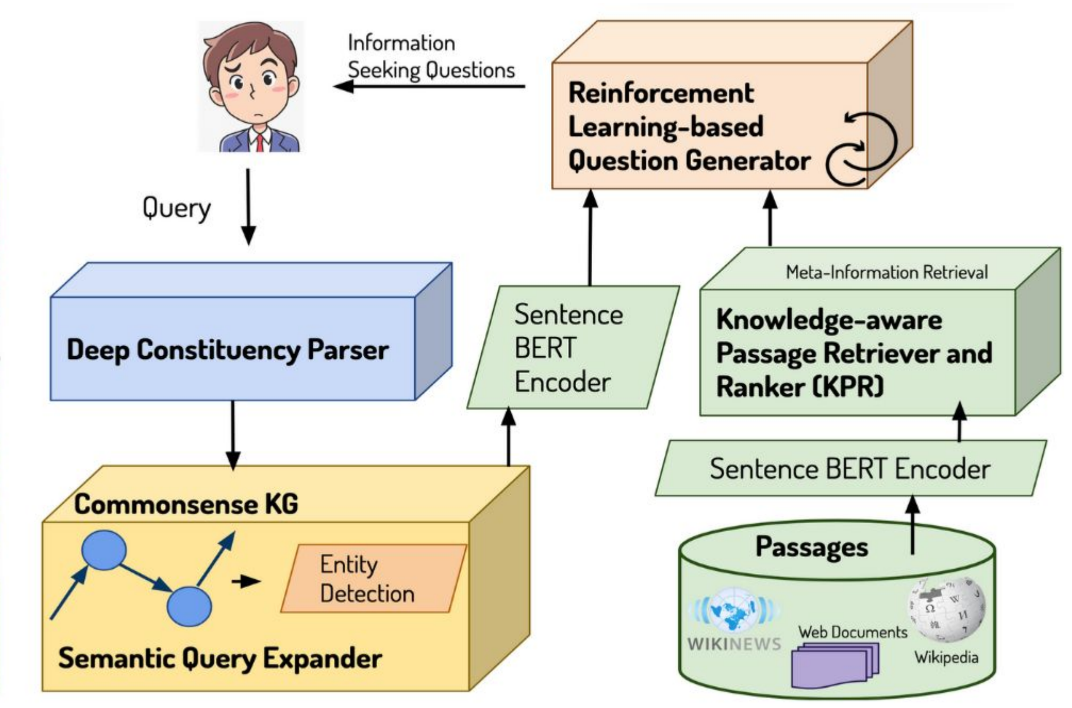
Geography (Travel Plan, etc.)

News (legal News, Stock News, etc.)

Architecture, Example, & ISEEQ's Approach



Finer Architecture of **ISEEQ**



Modular Architecture of **ISEEQ**

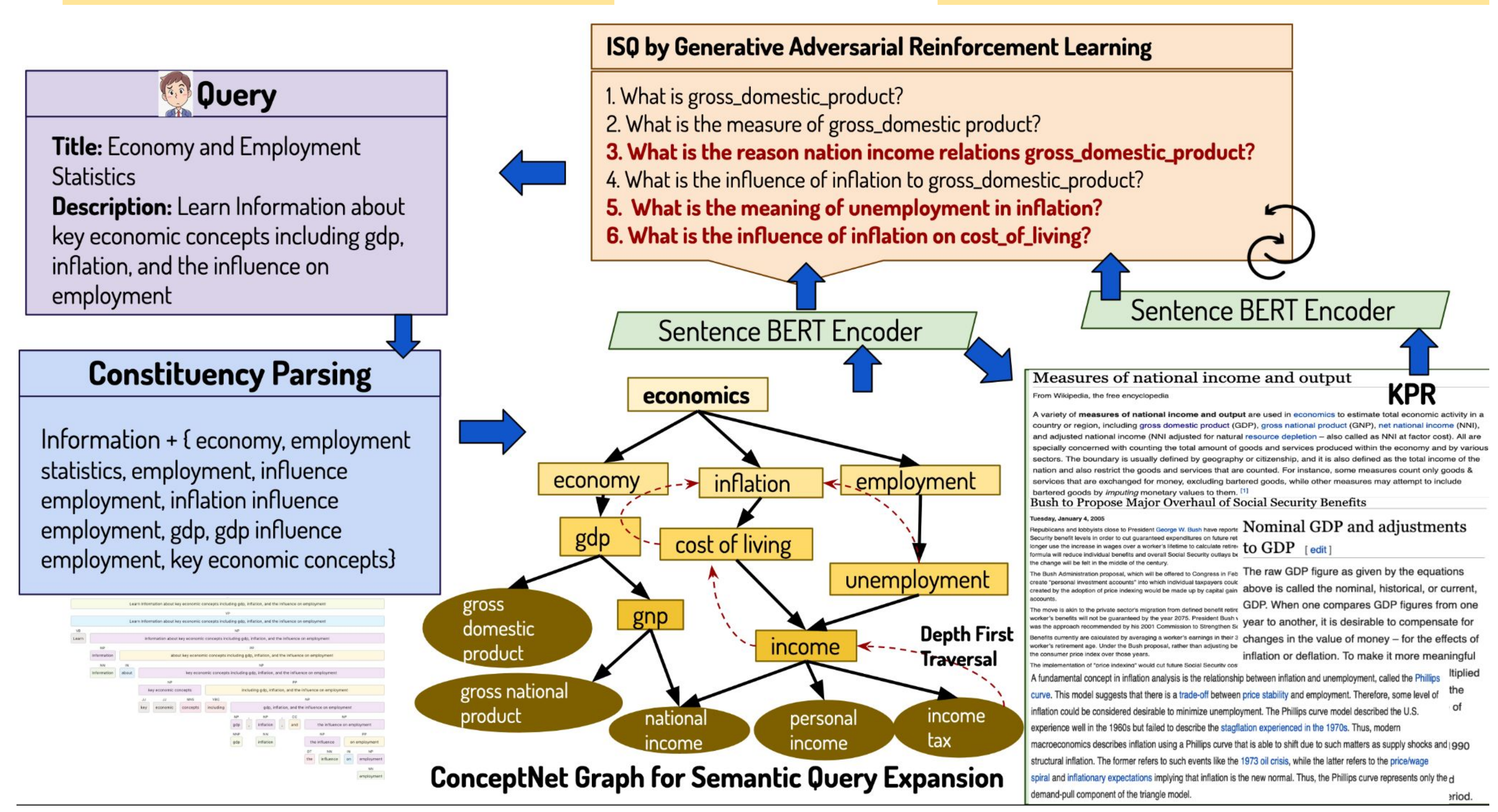


Illustration of **ISEEQ**

On topics on which people seek information on the Web like **mental health, public policy, ISEEQ's** generated questions were better than ground truth.

ISEEQ-RL model uses generative adversarial reinforcement learning framework to generate legitimate ISQs.

Reward Function:

$$R_i = \alpha\left[\frac{LCS(\hat{q}_i^m, q_i^n)}{|\hat{q}_i^m|}\right] + (1-\alpha)\left[\sum_{\hat{w}_{ij}\in\hat{q}_i^m} \max_{w_{ik}\in q_i^n} \mathrm{WMD}(\hat{w}_{ij}{}^T w_{ik})\right] \quad (a)$$

Loss Per Query with b ground-truth ISQs (since dataset have diverse queries):

$$\mathcal{L}(\hat{q}_{1:b}|q_{1:b},\theta) = \frac{-\sum_{i=1}^{b} R_i \cdot \mathbb{I}(q_i^n = \hat{q}_i^m)\cdot \mathbf{log} Pr(\hat{q}_i^m|\theta)}{b} \quad (b)$$

Loss Per Dataset Per Epoch:

$$\mathcal{L}(\hat{Q}|Q,\Theta)_t = \gamma\mathcal{L}(\hat{Q}|Q,\Theta)_{t-1} + (1-\gamma)\mathcal{L}(\hat{q}_{1:b}|q_{1:b},\theta) \quad (c)$$

ISEEQ-ERL model adds entailment constraints while learning to generate ISQs

$\hat{q}_{i|next}^m$: Next Generated Question after $\hat{q}_i^m$

We condition equation (b)

$$y_{max} = \mathrm{argmax}_Y\ \mathrm{RoBERTa}(\hat{q}_i^m, \hat{q}_{i|next}^m)$$

where $Y \in \{neutral, contradiction, entailment\}$

$$Pr(y_{max}) = \max_Y \mathrm{RoBERTa}(\hat{q}_i^m, \hat{q}_{i|next}^m)$$

$$\mathcal{L}(\hat{q}_{1:b}|q_{1:b},\theta) = \begin{cases} \textbf{if } y_{max} == \text{entailment } \textbf{then} \\ \quad \mathrm{CE} - Pr(y_{max}) \\ \textbf{else} \\ \quad \mathrm{RCE} - (1 - Pr(y_{max})) \\ \textbf{end if} \end{cases}$$

Reverse Cross Entropy Loss (RCE):
- Allows to place a thresholding strategy on adversarial generations
- Entailed questions are Non-Adversary
- Neutral and Contradictory questions are Adversar and handled by RCE

RoBERTa is pretrained on Stanford Natural Language Inference Dataset

$$\mathrm{RCE} = -\frac{\sum_{i=1}^{b} R_i(1-\mathbb{I}(q_i = \hat{q}_i))Pr(\hat{q}_i|\theta)}{b}$$

Results

Retrievers	HR@10	HR@20	MAP
TF-IDF + ECE (Clark et al. 2019)	0.31	0.45	0.16
BM25 + ECE*	0.38	0.49	0.23
DPR (Karpukhin et al. 2020)	0.44	0.61	0.31
KPR($\mathcal{Z}_{e_d}$)	0.47	0.66	0.35
KPR($\mathcal{Z}_{k_d}$)	0.49	0.70	0.38

Evaluating KPR on QAMR dataset

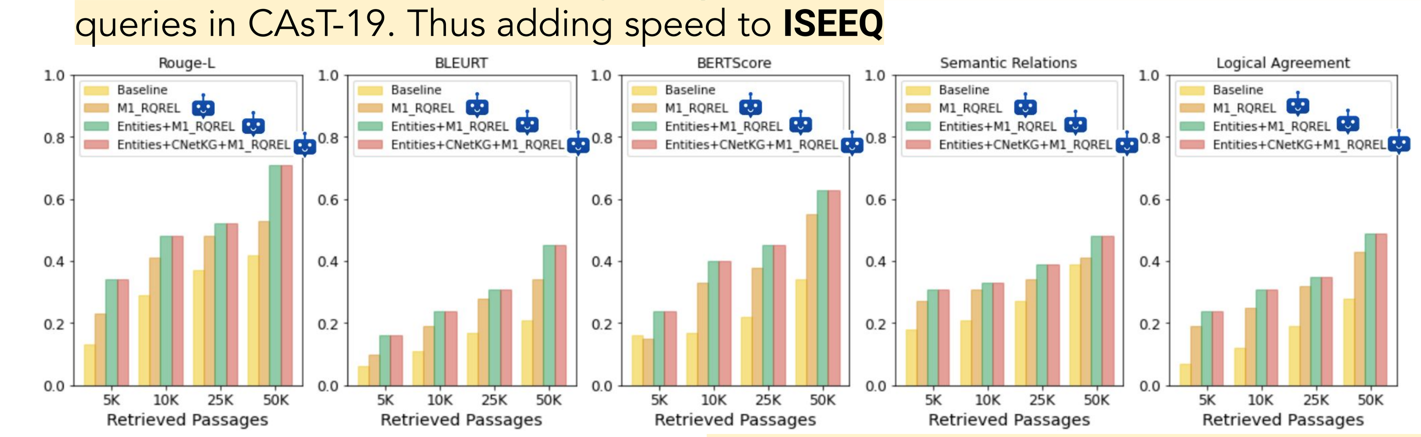
Ret.Pass.	DPR		KPR($\mathcal{Z}_{e_d}$)		KPR($\mathcal{Z}_{k_d}$)	
	Train	Test	Train	Test	Train	Test
5K	71	123	99	278	157	275
10K	96	133	154	301	194	316
25K	139	133	235	329	236	363
50K	173	144	269	358	269	402

Performance of KPR on MS-MARCO passages while retrieving atleast one passage per query in CAsT-19

✍ In terms of Relevant Passages retrieved KPR (entities in ConceptNet KG (k_d) and Noun Phrases (e_d)) outperformed Dense Passage Retrieval. (d: description)

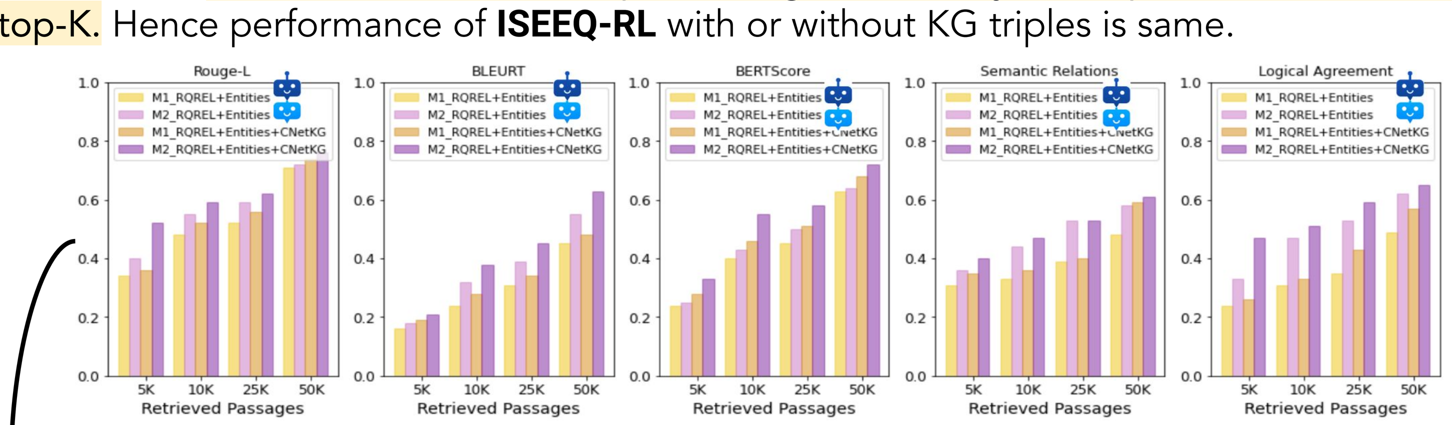
✍ [LEFT] In QAMR, ground-truth ISQs were about the entities in the query; Thus KPR outperformed DPR.

✍ [RIGHT] Minimal set of passages were retrieved to answer all the user queries in CAsT-19. Thus adding speed to **ISEEQ**



✍ With KG and minimal set of passages, **ISEEQ-RL** generated 27% more questions that are semantically similar to ground truth compared.

✍ Without entailment constraints, questions generated by KG triples never made to top-K. Hence performance of **ISEEQ-RL** with or without KG triples is same.



✍ Forcing the entailment constraints in **ISEEQ-ERL** yielded high scores on Semantic Relation and Logical Agreement -- Conceptual Flow.

Transferability Test was performed to show **ISEEQ-ERL's** ability to create new datasets for training and development o fnew CIS systems

Test → / Train ↓ (R-L/BRT/BScore/SR/LC(%))	QAD	QAMR	FBC	CAsT19
QAD	0.67/ 0.79/ 0.50/ 0.27/ 25.7	0.56/ 0.79/ 0.75/ 0.64/ 33.1	0.62/ 0.70/ 0.55/ 0.71/ 73.5	0.76/ 0.48/ 0.64/ 0.60/ 64.2
QAMR	0.73/ 0.89/ 0.62/ 0.28/ 27.7	0.57/ 0.83/ 0.77/ 0.68/ 37.[illegible]	0.74/ 0.89/ 0.67/ 0.75/ 77.8	0.67/ 0.41/ 0.57/ 0.57/ 58.6
FBC	0.70/ 0.73/ 0.56/ 0.31/ 33.0	0.61/ 0.85/ 0.72/ 0.67/ 35.8	0.79/ 0.89/ 0.66/ 0.78/ 79.4	0.75/ 0.37/ 0.76/ 0.67/ 66.5
CAsT-19	0.58/ 0.69/ 0.51/ 0.23/ 25.2	0.52/ 0.73/ 0.70/ 0.61/ 33.4	0.63/ 0.77/ 0.57/ 0.73/ 76.5	0.74/ 0.48/ 0.68/ 0.61/ 65.0

Closer

Summary

I. Imitating Human Information Seeking Behavior

✍ **ISEEQ-ERL** is the first CIS dialog agent in its class that infuses knowledge, entailment constraints, and self guides itself to generate diverse, conceptually coherent, and related questions.

✍ The generative adversarial training allows **ISEEQ-ERL** to balance the entities mentioned by users and entities from KG while generating questions.

✍ Through reinforcement learning with the reward on conceptual flow, **ISEEQ-ERL** can be trained to generate questions that are safety constrained and follow a specialized process knowledge.

II. Passing the Transferability test, ISEEQ proved to be assistive of annotators in NLP

✍ Dynamically retrieving passages containing suitable answers to users' queries.

✍ Generating numerous and diverse humanly understandable questions.

✍ Maintaining the flow of information through questions.

✍ Shaping the response after joining the response of independent questions.

More Details: https://github.com/manasgaur/AAAI-22