



INTERSPEECH 2020

OCTOBER 25-29/ SHANGHAI, CHINA
SHANGHAI INTERNATIONAL CONVENTION CENTER

Neural Approaches to Conversational Information Retrieval

Jianfeng Gao, Chenyan Xiong, Paul Bennett

Microsoft Research

INTERSPEECH 2020

OCTOBER 25, 2020 (Shanghai, China)

Outline

- **Part 1: Introduction**
 - A short definition of CIR
 - Task-oriented dialog and Web search
 - Research tasks of CIR
- Part 2: Conversational question answering (QA) methods
- Part 3: Conversational search methods
- Part 4: Overview of public and commercial systems

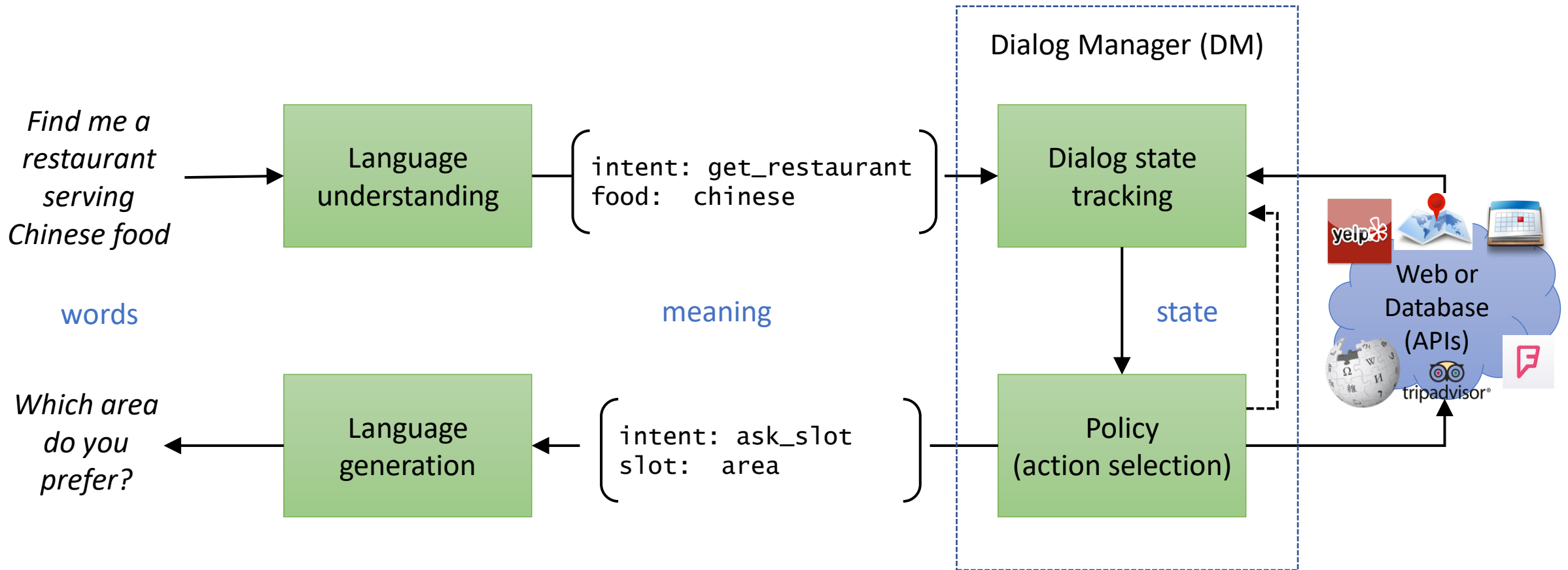
Who should attend this tutorial?

- Whoever wants to understand and develop modern CIR systems that
 - Can interact with users for information seeking via conversation
 - Can answer questions
 - Can help users search / look up information
 - Can help users with learning and investigating tasks
 - ...
- Focus on neural approaches in this tutorial
- Hybrid approaches that combine classical AI methods and deep learning methods are widely used to build real-world systems

A short definition of CIR

- A Conversational Information Retrieval (CIR) system is
 - an **information retrieval** (IR) system with
 - a **conversational interface** which
 - allows users to interact with the system to **seek information**
 - via multi-turn conversations of **natural language**.
- CIR is a task-oriented dialog system (aka. task-oriented bot)
 - Complete tasks (e.g., information seeking) via multi-turn conversations of natural language

Classical task-oriented dialog system architecture



User-System interaction in task-oriented dialog

- User: know (approximately) what she needs, but not what is available.
- System: know what is available, but not what the user needs
- dialog: a *two-way* process in which user and system get to know each other to make a deal

A (multi-domain) dialog session

DB State

Belief State

Attraction

```
"address": "98 king street",
"area": "centre",
"entrance fee": "free",
"id": "5",
"location": [
  52.207409,
  0.126738
],
"name": "broughton house gallery",
"openhours": "?",
"phone": "01223314960",
"postcode": "cb11ln",
"pricerange": "free",
"type": "museum"
```

Restaurant

```
"address": "106 Regent Street City Centre",
"area": "centre",
"food": "indian",
"id": "19214",
"introduction": "curry garden serves traditional bangladeshi cuisine cooked with fresh produce",
"location": [
  52.200187,
  0.126407
],
"name": "curry garden",
"phone": "01223302330",
"postcode": "cb21dp",
"pricerange": "expensive",
"type": "restaurant"
```

- 1 **H:** I am looking for a museum in the **center of town** do you have any suggestion?
- 2 **M:** there are 6 museums in the **centre of town** . how about the Broughton house gallery at 98 king street ?
- 3 **H:** nice, can you tell me what is the phone number and address ?
- 4 **M:** the phone number for the Broughton house gallery is 01223314960 and they are located at 98 king street .
- 5 **H:** ok thanks, can you recommend a restaurant in the **same area** serving Indian food
- 6 **M:** curry garden serves Indian food and is **located in the centre** . would you like me to make a reservation ?
- 7 **H:** sure, please book a table for two people on thursday at 18:00
- 8 **M:** ok , i have a table for 2 at 18:00 on Thursday . your reference number is 1234567890 . is there anything else I can help you with ?
- 9 **H:** that's it, bye.
- 10 **M:** have a great time and please contact us if you need any further assistance . goodbye .

Belief state:

```
{
  'attraction': {'type': 'museum',
                 'area': 'centre'}
}
```

Belief state:

```
{
  'restaurant': {'food': 'indian', 'area': 'centre'},
  'booking': {'day': 'Thursday',
              'people': '2', 'time': '18:00'},
  'attraction': {'type': 'museum',
                 'area': 'centre'}
}
```

User-system interaction in Web search

- User: know (roughly) what she needs, but not what is available.
- System: know what is available, but not what a user needs
- Generally viewed as a one-way information seeking process
 - User plays a **proactive** role to iteratively
 - issue a query,
 - inspect search results,
 - reformulate the query
 - System plays a **passive** role to make search more effective
 - Autocomplete a query
 - Organize search results (SERP)
 - Suggest related queries



broughton house



broughton house

broughton house nj

broughton house uk

broughton house savannah

broughton house scotland

broughton house kirkcudbright

IMAGES VIDEOS MAPS NEWS SHOPPING

time

See work results for broughton house

Why am I seeing this

is the Home of Salford & Greater ...

house.com

of Salford and Greater Manchester's Veterans, providing care to the Armed Forces and Merchant Navy in the North West of England. Broughton House Care Home for Ex-Service Men and Women is a charity registered in England and Wales (1155225)

Governance & Management · About Us · Broughton House Retirement Living

Images of Broughton House

bing.com/images



Kirkcudbright



Grange



Sheffield



Drink



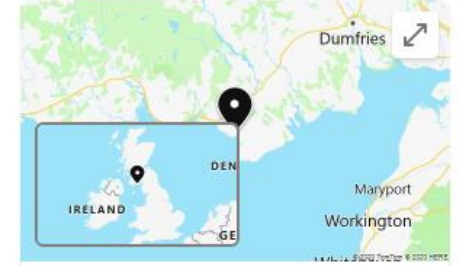
Grange-Over-...



Jianfeng



30



COVID-19

Temporary closure or reduced hours may be in effect

Broughton House

Town House



Broughton House is an 18th-century town house standing on the High Street of Kirkcudbright, Sco... +



Wikipedia

Address: 12 High Street, Kirkcudbright, Scotland DG6 4

Phone: +44 1557 330437

Data from: Wikipedia

Wikipedia text under CC-BY-SA license

Suggest an edit

PEOPLE ALSO ASK

What is a Boughton House?

Where is Boughton House in England?

What is a Coughton house?

Where is Coughton house?

use

ikipedia

i/Broughton_House

ry

House and gardens

External links

se is an 18th-century town house standing on the High Street of Scotland. It was the home of Scots impressionist artist E. A. Hornel and his death in 1933. During this time Hornel remodelled the house with Japanese-influenced gardens. Since 1997 it has been in the care of the National Trust for Scotland, and is maintained as "a living museum of Hornel's life". The house is a category A listed building, and the gardens are included on

Feedback



System should interact with users more actively

- How people search -- Information seeking
 - Information lookup – short search sessions;
 - Exploratory search based on a dynamic model, an iterative “sense-making” process where users learn as they search, and adjust their information needs as they see search results.
- Effective information seeking requires interaction btw users and a system that explicitly models the interaction by
 - Tracking belief state (user intent)
 - Asking clarification questions
 - Providing recommendations
 - Using natural language as input/output

A long definition of CIR - the RRIMS properties

- User **R**evealment: system helps users express their information needs
 - E.g., query suggestion, autocompletion
- System **R**evealment: system reveals to users what is available, what it can or cannot do
 - E.g., recommendation, SERP
- Mixed **I**nitiative: system and user both can take initiative (two-way conversation)
 - E.g., asking clarification questions
- **M**emory: user can reference past statement
 - State tracking
- **S**et Retrieval: system can reason about the utility of sets of complementary items
 - Task-oriented, contextual search or QA

CIR research tasks (task-oriented dialog modules)

- What we will cover in this tutorial
 - Conversational Query Understanding (LU, belief state tracking)
 - Conversational document ranking (database state tracking)
 - Learning to ask clarification questions (action select via dialog policy, LG)
 - Conversational leading suggestions (action select via dialog policy, LG)
 - Search result presentation (response generation, LG)
- Early work on CIR [[Croft's keynote at SIGIR-19](#)]
- We start with conversational QA which is a sub-task of CIR

Outline

- Part 1: Introduction
- **Part 2: Conversational QA methods**
 - **Conversational QA over knowledge bases**
 - **Conversational QA over texts**
- Part 3: Conversational search methods
- Part 4: Case study of commercial systems

KBQA: Conversational QA over Knowledge Bases

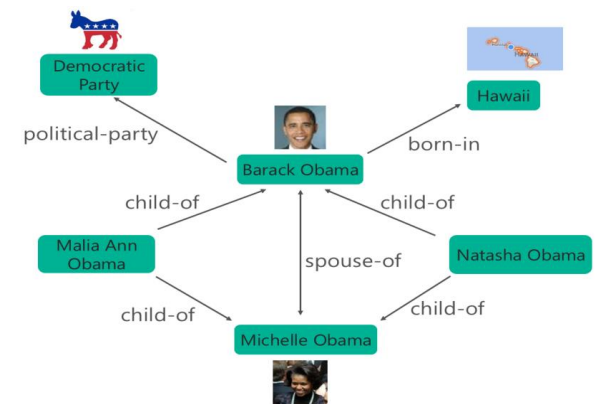
- Knowledge bases and QAs
- C-KBQA system architecture
 - Semantic parser
 - Dialog manager
 - Response generation
- KBQA w/o semantic parser
- Open benchmarks

Knowledge bases

- Relational databases
 - Entity-centric knowledge base
 - Q: what super-hero from Earth appeared first?
- Knowledge Graph
 - Properties of billions of entities
 - Relations among them
 - (relation, subject, object) tuples
 - Freebase, FB Entity Graph, MS Satori, Google KG etc.
 - Q: what is Obama's citizenship?
- KGs work with paths while DBs work with sets

Legion of Super Heroes Post-Infinite Crisis

Character	First Appeared	Home World	Powers
Night Girl	2007	Kathoon	Super strength
Dragonwing	2010	Earth	Fire breath
Gates	2009	Vyrga	Teleporting
XS	2009	Aarok	Super speed
Harmonia	2011	Earth	Elemental



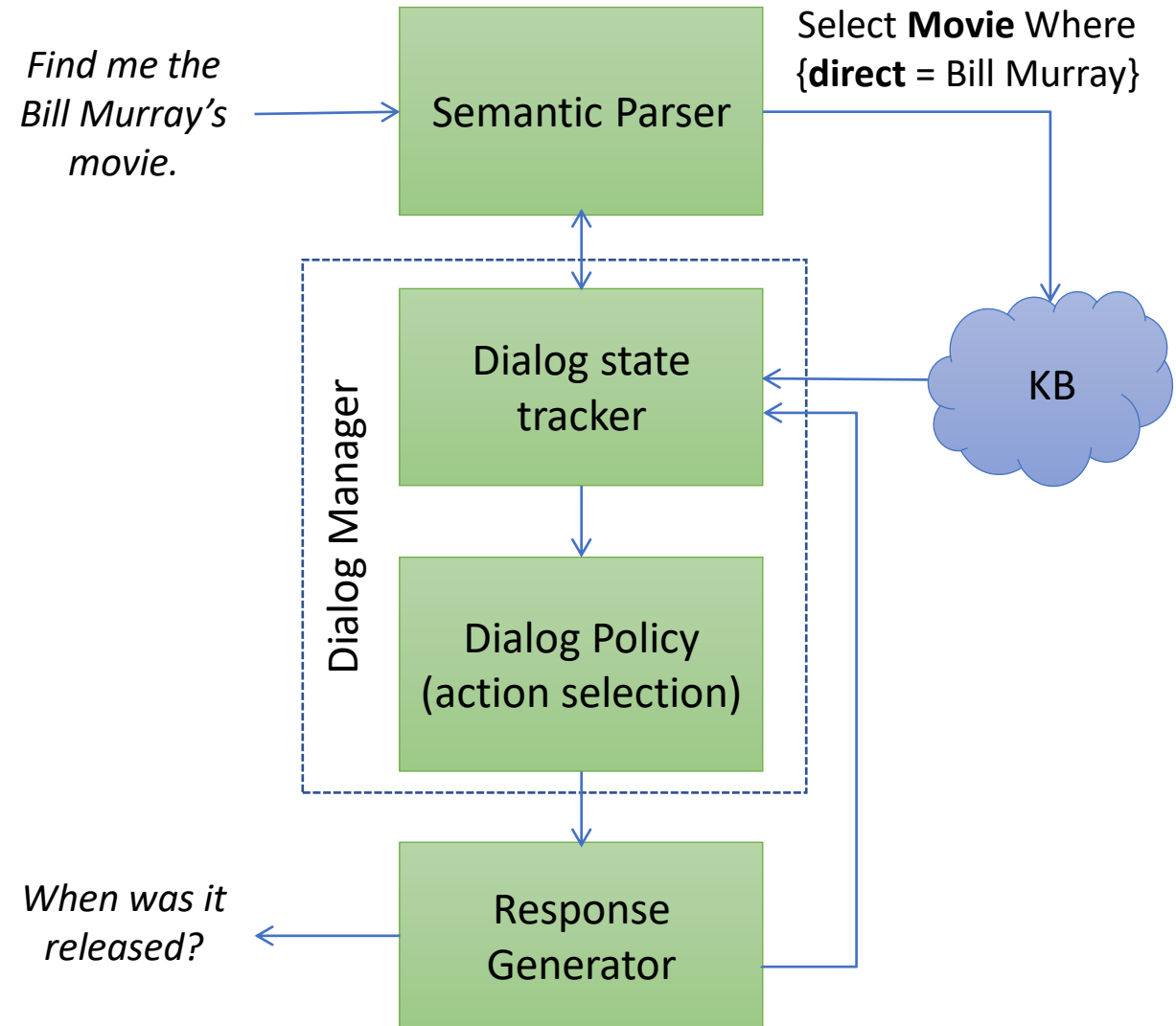
Question-Answer pairs

- Simple questions
 - can be answered from a single tuple
 - Object? / Subject? / Relation?
- Complex questions
 - requires reasoning over one or more tuples
 - Logical / quantitatively / comparative
- Sequential QA pairs in a session
 - A sequence of related pairs
 - Ellipses, coreference, clarifications, etc.

Turn	State	Utterance
T1	Simple Q.	USER : Can you tell me which cities border Verderio Inferiore ?
	Response	SYSTEM : Cornate d'Adda, Bernareggio, Robbiate
T2	Coreference Q.	USER : And which cities flank that one ?
	Clarification Q.	SYSTEM : Did you mean Robbiate ?
T3	Clarification	USER : No, I meant Cornate d'Adda.
	Response	SYSTEM : Bottanuco, Busnago, Trezzo sull'Adda
T4	Logical Q.	USER : Which ones share border with Robbiate and are not part of Azerbaijan ?
	Response	SYSTEM : Trezzo sull'Adda, Busnago, Bottanuco
T5	Quantitative Q.	USER : Which city is the capital of min number of french administrative divisions ?
	Response	SYSTEM : Riga
T6	Boolean Q.	USER : Is Rot am See located adjacent to that one ?
	Response	SYSTEM : No
T7	Quanti. Q.	USER : How many political territories have strategic relation with around 22 other countries ?
	Response	SYSTEM : 11
T8	Logical Q.	USER : Which are the nations connected with Robbiate and are not located in Italy ?
	Response	SYSTEM : The answer count is 11. Do you want to see all possibilities?
T9	Clarification	USER : No, show only a few of them
	Response	SYSTEM : Papua New Guinea, Jamaica, Antigua and Barbuda, Austria, New Zealand, The Bahamas, Gabon
T10	Quantitative Q.	USER : How many countries are diplomatically related to Italy ?
	Response	SYSTEM : 74
T11	Quantitative Q.	USER : And how many of them also with Australia ?
	Response	SYSTEM : 56
T12	Comparative Q.	USER : Which countries have lesser number of political relations with others than Australia ?
	Response	SYSTEM : Denmark, Canada, Grenada

C-KBQA system architecture

- Semantic Parser
 - map input + context to a semantic representation (logic form) to
 - Query the KB
- Dialog manager
 - Maintain/update state of dialog history (e.g., QA pairs, DB state)
 - Select next system action (e.g., ask clarification questions, answer)
- Response generator
 - Convert system action to natural language response
- KB search (Gao+19)



Dynamic Neural Semantic Parser (DynSP)

- Given a question (dialog history) and a table
 - Q: “which superheroes came from Earth and first appeared after 2009?”
- Generate a semantic parse (SQL-query)
 - A **select** statement (answer column)
 - Zero or more **conditions**, each contains
 - A condition column
 - An operator (=, >, <, argmax etc.) and arguments
 - Q: Select **Character** Where {**Home World** = “Earth”} & {**First Appear** > “2009”}
 - A: {Dragonwing, Harmonia}

Legion of Super Heroes Post-Infinite Crisis

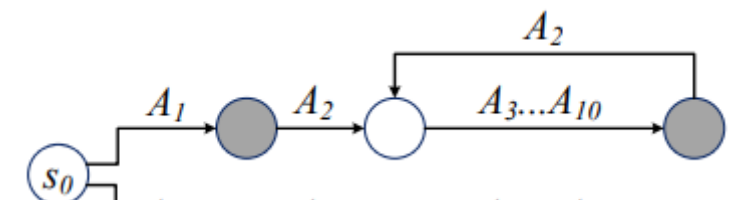
<i>Character</i>	<i>First Appeared</i>	<i>Home World</i>	<i>Powers</i>
Night Girl	2007	Kathoon	Super strength
Dragonwing	2010	Earth	Fire breath
Gates	2009	Vyrge	Teleporting
XS	2009	Aarok	Super speed
Harmonia	2011	Earth	Elemental

Model formulation

- Parsing as a state-action search problem
 - A state S is a complete or partial parse (action sequence)
 - An action A is an operation to extend a parse
 - Parsing searches an end state with the highest score
- “which superheroes came from Earth and first appeared after 2009?”
 - (A_1) Select-column **Character**
 - (A_2) Cond-column **Home World**
 - (A_3) Op-Equal “Earth”
 - (A_2) Cond-column **First Appeared**
 - (A_5) Op-GT “2009”

Id	Type	# Action instances
A_1	Select-column	# columns
A_2	Cond-column	# columns
A_3	Op-Equal (=)	# rows
A_4	Op-NotEqual (\neq)	# rows
A_5	Op-GT (>)	# numbers / datetimes
A_6	Op-GE (\geq)	# numbers / datetimes
A_7	Op-LT (<)	# numbers / datetimes
A_8	Op-LE (\leq)	# numbers / datetimes
A_9	Op-ArgMin	# numbers / datetimes
A_{10}	Op-ArgMax	# numbers / datetimes

Types of actions and the number of action instances in each type. Numbers / datetimes are the mentions discovered in the question.



Possible action transitions based on their types. Shaded circles are end states.

How to score a state (parse)?

- Beam search to find the highest-scored parse (end state)
 - $V_{\theta}(S_t) = V_{\theta}(S_{t-1}) + \pi_{\theta}(S_{t-1}, A_t), V(S_0) = 0$
- Policy function, $\pi_{\theta}(S, A)$,
 - Scores an action given the current state
 - Parameterized using different neural networks, each for an action type
 - E.g., **Select-column** action is scored using the semantic similarity between question words (embedding vectors) and column name (embedding vectors)
 - $\frac{1}{|W_c|} \sum_{w_c \in W_c} \max_{w_q \in W_q} w_q^T w_c$

Model learning

- State value function: $V_{\theta}(S_t) = \sum_{i=1}^t \pi_{\theta}(S_{i-1}, A_i)$
 - An E2E trainable, question-specific, neural network model
- Weakly supervised learning setting
 - Question-answer pairs are available
 - Correct parse for each question is not available
- Issue of delayed (sparse) reward
 - Reward is only available after we get a (complete) parse and the answer
- Approximate (dense) reward
 - Check the overlap of the answers of a partial parse $A(S)$ with the gold answers A^*
 - $R(S) = \frac{|A(S) \cap A^*|}{|A(S) \cup A^*|}$

Parameter updates

- Make the state value function V_θ behave similarly to reward R
- For every state S and its (approximated) reference state S^* , we define loss as
 - $\mathcal{L}(S) = (V_\theta(S) - V_\theta(S^*)) - (R(S) - R(S^*))$
- Improve learning efficiency by finding the most violated state \hat{S}

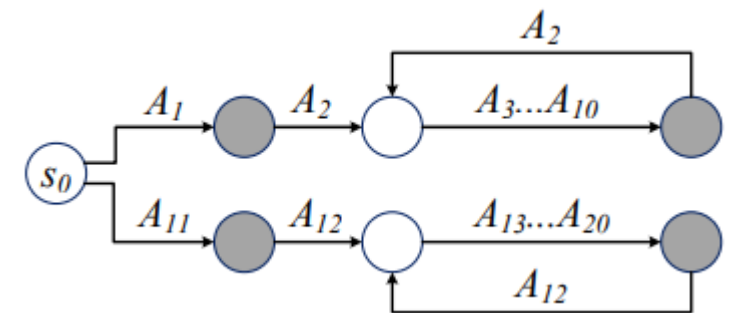
Algorithm 1 Model parameter updates

- 1: **for** pick a labeled data (x, A^*) **do** // labeled QA pair
 - 2: $s^* \leftarrow \arg \max_{s \in \mathcal{E}(x)} \tilde{R}(s; A^*)$ // Finds the best approximated reference state
 - 3: $\hat{s} \leftarrow \arg \max_{s \in \mathcal{E}(x)} V_\theta(s) - \tilde{R}(s; A^*)$ // Finds the most violated state
 - 4: update θ by minimizing $\max(\mathcal{L}(s), 0)$
 - 5: **end for**
-

DynSP SQA

- “which superheroes came from Earth and first appeared after 2009?”
 - (A_1) Select-column **Character**
 - (A_2) Cond-column **Home World**
 - (A_3) Op-Equal “Earth”
 - (A_2) Cond-column **First Appeared**
 - (A_5) Opt-GT “2009”
- “which of them breathes fires”
 - (A_{12}) S-Cond-column **Powers**
 - (A_{13}) S-Op-Equal “Fire breath”

Id	Type	# Action instances
A_1	Select-column	# columns
A_2	Cond-column	# columns
A_3	Op-Equal (=)	# rows
A_4	Op-NotEqual (\neq)	# rows
A_5	Op-GT ($>$)	# numbers / datetimes
A_6	Op-GE (\geq)	# numbers / datetimes
A_7	Op-LT ($<$)	# numbers / datetimes
A_8	Op-LE (\leq)	# numbers / datetimes
A_9	Op-ArgMin	# numbers / datetimes
A_{10}	Op-ArgMax	# numbers / datetimes
A_{11}	Subsequent	1
A_{12}	S-Cond-column	# columns
A_{13}	S-Op-Equal (=)	# rows
A_{14}	S-Op-NotEqual (\neq)	# rows
A_{15}	S-Op-GT ($>$)	# numbers / datetimes
A_{16}	S-Op-GE (\geq)	# numbers / datetimes
A_{17}	S-Op-LT ($<$)	# numbers / datetimes
A_{18}	S-Op-LE (\leq)	# numbers / datetimes
A_{19}	S-Op-ArgMin	# numbers / datetimes
A_{20}	S-Op-ArgMax	# numbers / datetimes



Possible action transitions based on their types. Shaded circles are end states.

DynSP for sequential QA (SQA)

- Given a question (history) and a table
 - Q1: which superheroes came from Earth and first appeared after 2009?
 - Q2: which of them breathes fire?
- Add **subsequent** statement (answer column) for sequential QA
 - Select **Character** Where {**Home World** = "Earth"} & {**First Appear** > "2009"}
 - A1: {Dragonwing, Harmonia}
 - Subsequent Where {**Powers** = "Fire breath"}
 - A2: {Dragonwing}

Legion of Super Heroes Post-Infinite Crisis

Character	First Appeared	Home World	Powers
Night Girl	2007	Kathoon	Super strength
Dragonwing	2010	Earth	Fire breath
Gates	2009	Vyrge	Teleporting
XS	2009	Aarok	Super speed
Harmonia	2011	Earth	Elemental

Query rewriting approaches to SQA

Q1: When was California founded?

A1: September 9, 1850

Q2: Who is its governor? → [Who is California governor?](#)

A2: Jerry Brown

Q3: Where is Stanford?

A3: Palo Alto, California

Q4: Who founded it? → [Who founded Stanford?](#)

A4: Leland and Jane Stanford

Q5: Tuition costs → [Tuition cost Stanford](#)

A5: \$47,940 USD

11 [Human] You like Ashin that much, huh?

[Xiaoice] Why not? You don't like him?

12 [Human] I don't like him, neither dislike him → [I don't like Ashin, neither dislike Ashin](#)

[Xiaoice] I love every song of him

13 [Human] Really?

[Xiaoice] Especially *The Time Machine*

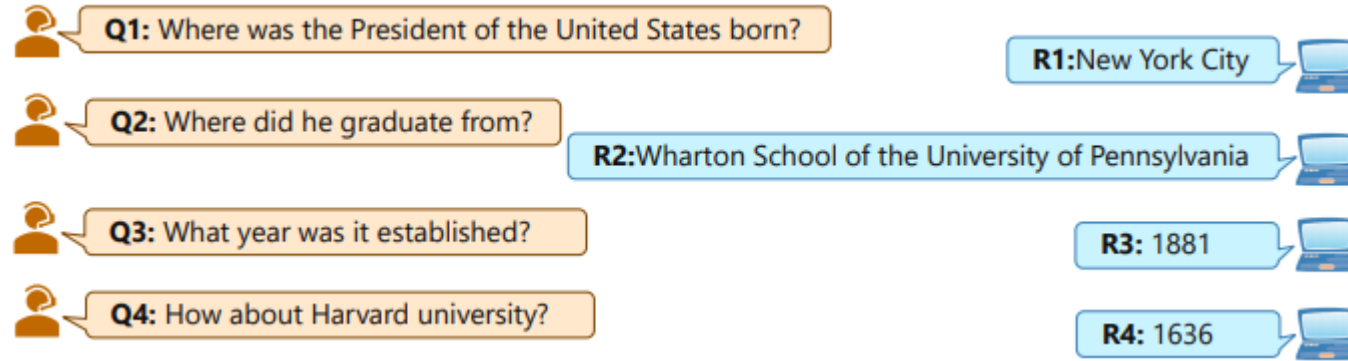
14 [Human] Oh? Never heard of that → [Oh? Never heard of *The Time Machine*](#)

[Xiaoice] Shall I send you the song?

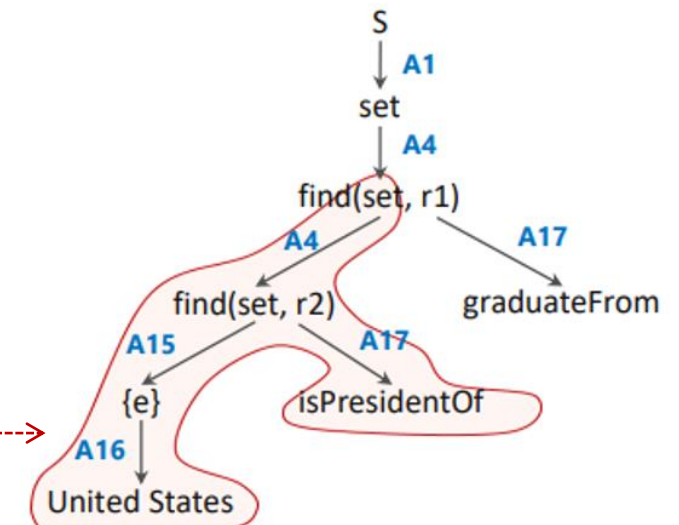
15 [Human] Okay → [Okay, send *The Time Machine*.](#)

[Xiaoice] Don't know if you will like it or not~ [Play the music card]

Dialog Manager – dialog memory for state tracking

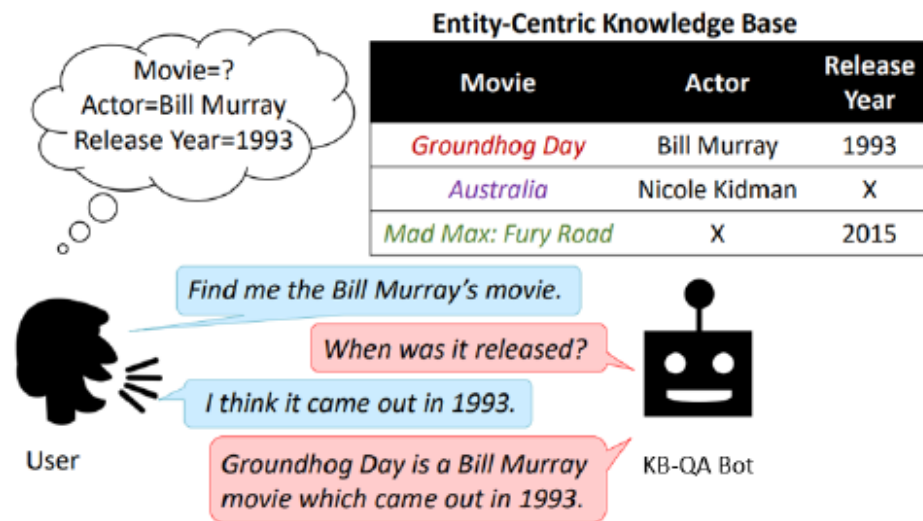


Dialog Memory (of state tracker)	
Entity	{United States, "q"} {New York City, "a"} {University of Pennsylvania, "a"} ...
Predicate	{isPresidentOf} {placeGraduateFrom} {yearEstablished} ...
Action subsequence (partial/complete states)	Set $\rightarrow A_4 A_{15} e_{us} r_{pres}$ Set $\rightarrow A_4 A_{15}$



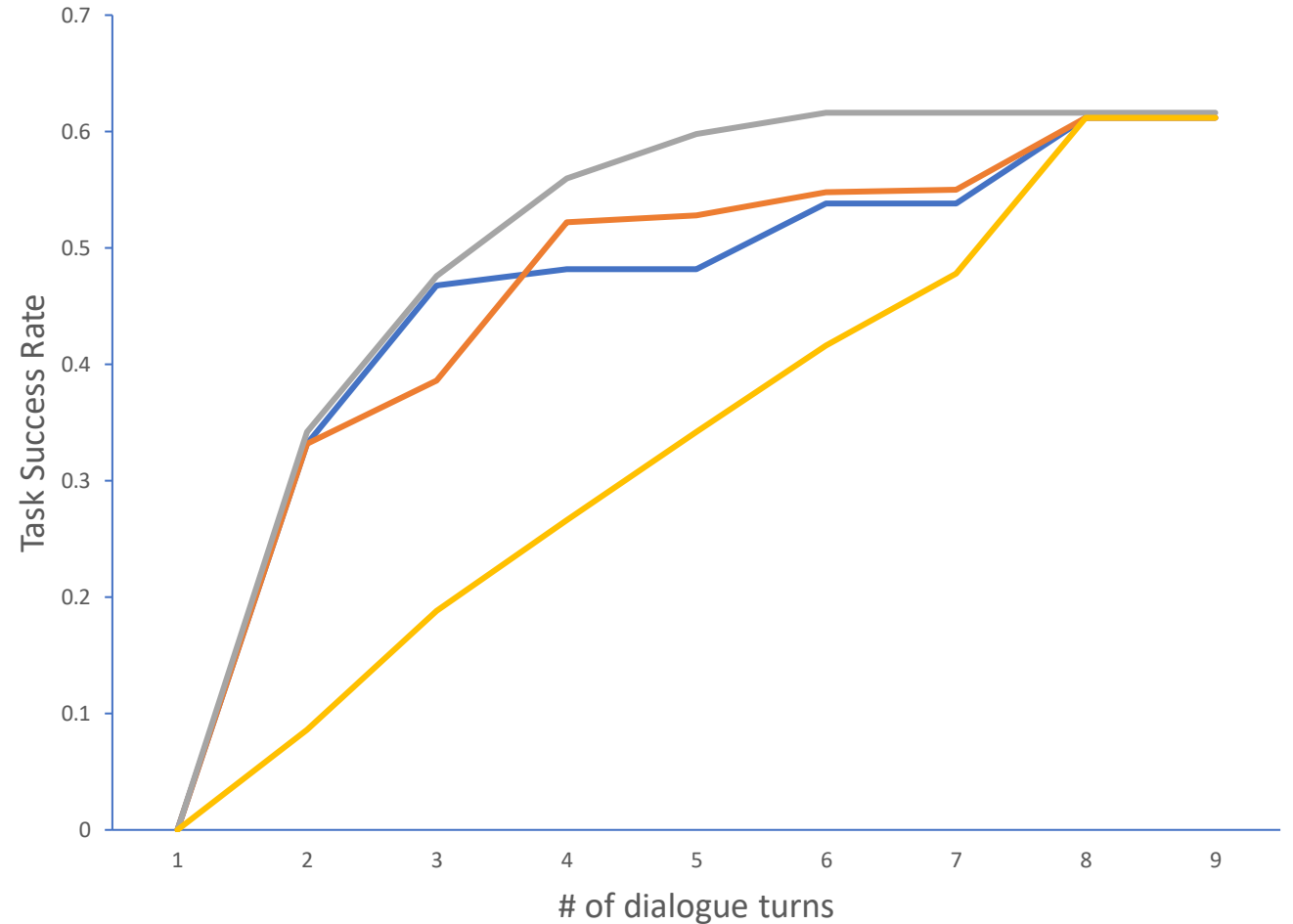
Dialog Manager – policy for next action selection

- A case study of Movie-on-demand
- System selects the next action to
 - Either return answer or ask a clarification question.
 - What (clarification) question to ask? E.g., movie title, director, genre, actor, release-year, etc.



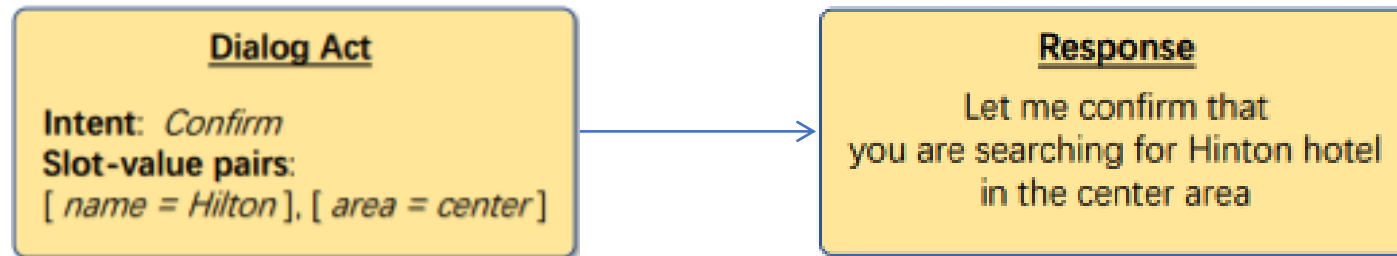
What clarification question to ask

- **Baseline: ask all questions in a randomly sampled order**
- **Ask questions that users can answer**
 - learned from query logs
- **Ask questions that help reduce search space**
 - **Entropy minimization**
- **Ask questions that help complete the task successfully**
 - Reinforcement learning via agent-user interactions



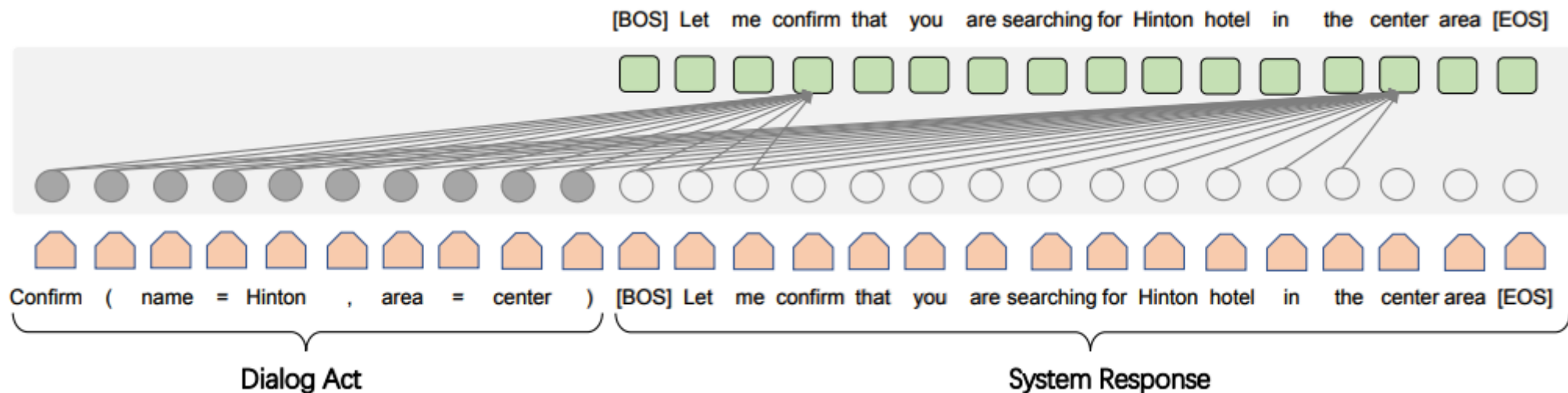
Results on simulated users

Response Generation



- Convert “dialog act” to “natural language response”
- Formulated as a seq2seq task in a few-shot learning setting
 - $p_{\theta}(\mathbf{x}|A) = \sum_{t=1}^T p_{\theta}(x_t|x_{<t}, A)$
 - Very limited training samples for each task
- Approach
 - Semantically Conditioned neural language model
 - Pre-training + fine-tuning,
 - e.g., semantically conditioned GPT (SC-GPT)

SC-GPT

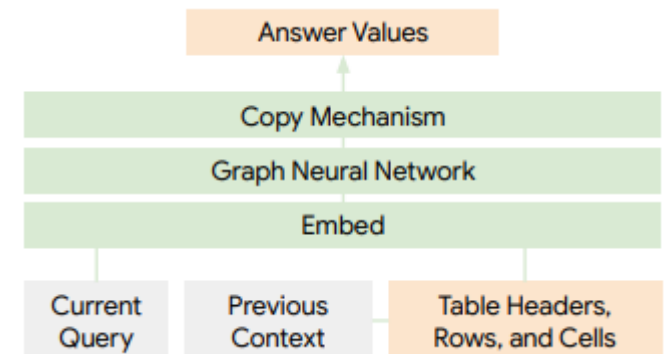


Model	Restaurant		Laptop		Hotel		TV		Attraction		Train		Taxi	
	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓
SC-LSTM	15.90	48.02	21.98	80.48	31.30	31.54	22.39	64.62	7.76	367.12	6.08	189.88	11.61	61.45
GPT-2	29.48	13.47	27.43	11.26	35.75	11.54	28.47	9.44	16.11	21.10	13.72	19.26	16.27	9.52
SC-GPT	38.08	3.89	32.73	3.39	38.25	2.75	32.95	3.38	20.69	12.72	17.21	7.74	19.70	3.57

Performance of different response generation models in few-shot setting (50 samples for each task)

C-KBQA approaches w/o semantic parser

- Building semantic parsers is challenging
 - Limited amounts of training data, or
 - Weak supervision
- C-KBQA with no logic-form
 - Symbolic approach: “look before you hop”
 - Answer an initial question using any standard KBQA
 - Form a context subgraph using entities of the initial QA pair
 - Answer follow-up questions by expanding the context subgraph to find candidate answers
 - Neural approach
 - Encode KB as graphs using a GNN
 - Select answers from the encoded graph using a point network



Open Benchmarks

- SQA (sequential question answering)
 - <https://www.microsoft.com/en-us/download/details.aspx?id=54253>
- CSQA (complex sequence question answering),
 - <https://amritasaha1812.github.io/CSQA/>
- ConvQuestions (conversational question answering over knowledge graphs)
 - <https://convex.mpi-inf.mpg.de/>
- CoSQL (conversational text-to-SQL)
 - <https://yale-lily.github.io/cosql>
- CLAQUA (asking clarification questions in Knowledge-based question answering)
 - https://github.com/msra-nlc/MSParS_V2.0

TextQA: Conversational QA over Texts

- Tasks and datasets
- C-TextQA system architecture
- Conversational machine reading compression models

QA over text – extractive vs. abstractive QA

Passage { In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

Question: What causes precipitation to fall?
Answer: **gravity**

Question: What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
Answer: **graupel**

Question: Where do water droplets collide with ice crystals to form precipitation?
Answer: **within a cloud**

Q Will I qualify for OSAP if I'm new in Canada?

Selected Passages from Bing

"Visit the OSAP website for application deadlines. To get OSAP, you have to be eligible. You can apply using an online form, or you can print off the application forms. If you submit a paper application, you must pay an application fee. The online application is free."

Source: <http://settlement.org/ontario/education/colleges-universities-and-institutes/financial-assistance-for-post-secondary-education/how-do-i-apply-for-the-ontario-student-assistance-program-osap/>

"To be eligible to apply for financial assistance from the Ontario Student Assistance Program (OSAP), you must be a: 1 Canadian citizen; 2 Permanent resident; or 3 Protected person/convention refugee with a Protected Persons Status Document (PPSD)."

Source: <http://settlement.org/ontario/education/colleges-universities-and-institutes/financial-assistance-for-post-secondary-education/who-is-eligible-for-the-ontario-student-assistance-program-osap/>

"You will not be eligible for a Canada-Ontario Integrated Student Loan, but can apply for a part-time loan through the Canada Student Loans program. There are also grants, bursaries and scholarships available for both full-time and part-time students."

Source: <http://www.campusaccess.com/financial-aid/osap.html>

Answer

No. You won't qualify.

Conversation QA over text: CoQA & QuAC

Section: 🦆 Daffy Duck, Origin & History

STUDENT: **What is the origin of Daffy Duck?**
TEACHER: ↪ first appeared in Porky's Duck Hunt

STUDENT: **What was he like in that episode?**
TEACHER: ↪ assertive, unrestrained, combative

STUDENT: **Was he the star?**
TEACHER: ↪ No, barely more than an unnamed bit player in this short

STUDENT: **Who was the star?**
TEACHER: ↧ No answer

STUDENT: **Did he change a lot from that first episode in future episodes?**
TEACHER: ↪ Yes, the only aspects of the character that have remained consistent (...) are his voice characterization by Mel Blanc

STUDENT: **How has he changed?**
TEACHER: ↪ Daffy was less anthropomorphic

STUDENT: **In what other ways did he change?**
TEACHER: ↪ Daffy's slobbery, exaggerated lisp (...) is barely noticeable in the early cartoons.

STUDENT: **Why did they add the lisp?**
TEACHER: ↪ One often-repeated "official" story is that it was modeled after producer Leon Schlesinger's tendency to lisp.

STUDENT: **Is there an "unofficial" story?**
TEACHER: ↪ Yes, Mel Blanc (...) contradicts that conventional belief

...

Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80. Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well. Jessica had ...

Q₁: Who had a birthday?

A₁: Jessica

R₁: Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80.

Q₂: How old would she be?

A₂: 80

R₂: she was turning 80

Q₃: Did she plan to have any visitors?

A₃: Yes

R₃: Her granddaughter Annie was coming over

Q₄: How many?

A₄: Three

R₄: Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well.

Q₅: Who?

A₅: Annie, Melanie and Josh

R₅: Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well.

Figure 3.10: The examples from two conversational QA datasets. (Left) A QA dialogue example in the QuAC dataset. The student, who does not see the passage (section text), asks questions. The teacher provides answers in the form of text spans and dialogue acts. These acts include (1) whether the student should ↪, could ↪, or should not ↧ ask a follow-up; (2) affirmation (Yes / No), and, when appropriate, (3) No answer. Figure credit: Choi et al. (2018). (Right) A QA dialogue example in the CoQA dataset. Each dialogue turn contains a question (Q_i), an answer (A_i) and a rationale (R_i) that supports the answer. Figure credit: Reddy et al. (2018).

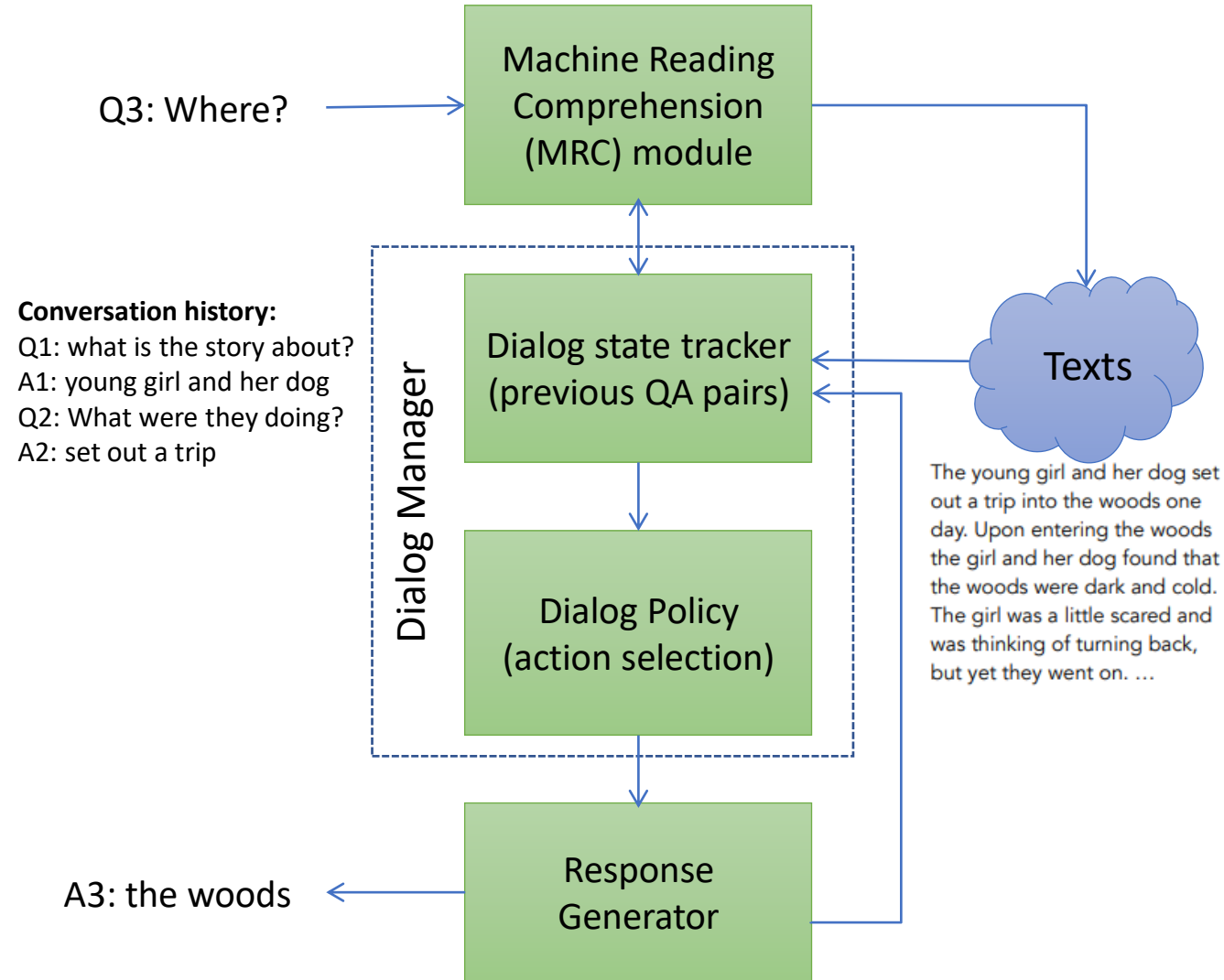
Dialog behaviors in conversational QA

- Topic shift: a question about sth previous discussed
- Drill down: a request for more info about a topic being discussed
- Topic return: asking about a topic again after being shifted
- Clarification: reformulating a question
- Definition: asking what is meant by a term

Dataset	Topic Shift	Drill Down	Return to Topic	Clarification Question	Definition Question	Sentence Coverage	Total Questions
CoQA	21.6	72.0	2.9	0.0	0.7	63.3	722
QuAC	35.4	55.3	5.6	0.7	3.0	28.4	302

C-TextQA system architecture

- (Conversational) MRC
 - Find answer to a question given text and previous QA pairs
 - Extractive (span) vs. abstractive answers
- Dialog manager
 - Maintain/update state of dialog history (e.g., QA pairs)
 - Select next system action (e.g., ask clarification questions, answer)
- Response generator
 - Convert system action to natural language response

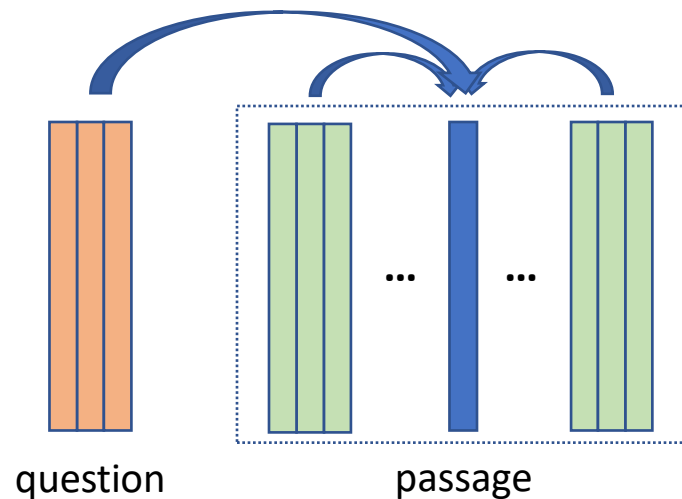


Neural MRC models for extractive TextQA

- QA as classification given (question, passage)
 - Classify each word in passage as start/end/outside of the answer span
- Encoding: represent each passage word using an **integrated context vector** that encodes info from
 - Lexicon/word embedding (context-free)
 - Passage context
 - Question context
 - **Conversation context (previous question-answer pairs)**
- Prediction: predict each word (its integrated context vector) the start and end position of answer span.

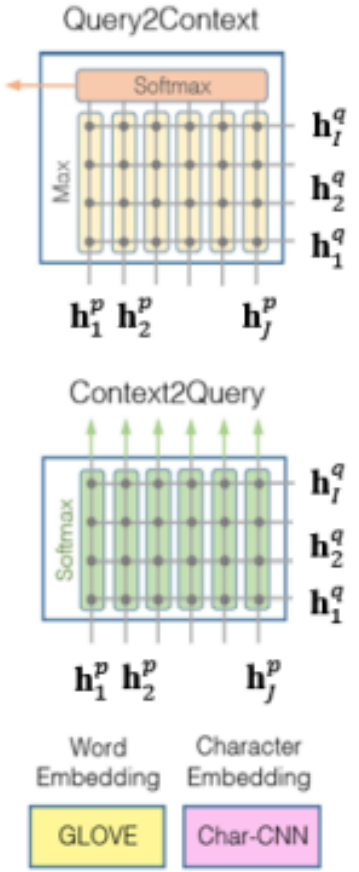
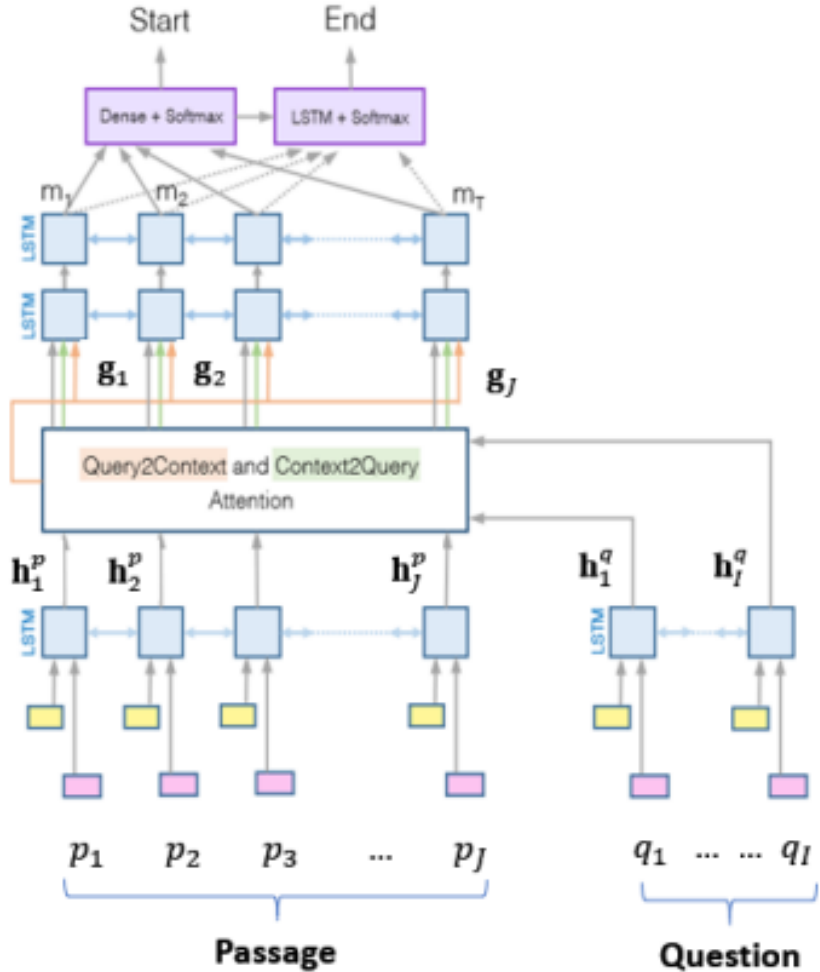
Three encoding components

- Lexicon embedding e.g., GloVe
 - represent each word as a low-dim continuous vector
- Passage contextual embedding e.g., Bi-LSTM/RNN, ELMo, Self-Attention/BERT
 - capture context info for each word within a passage
- Question contextual embedding e.g., Attention, BERT
 - fuse question info into each passage word vector

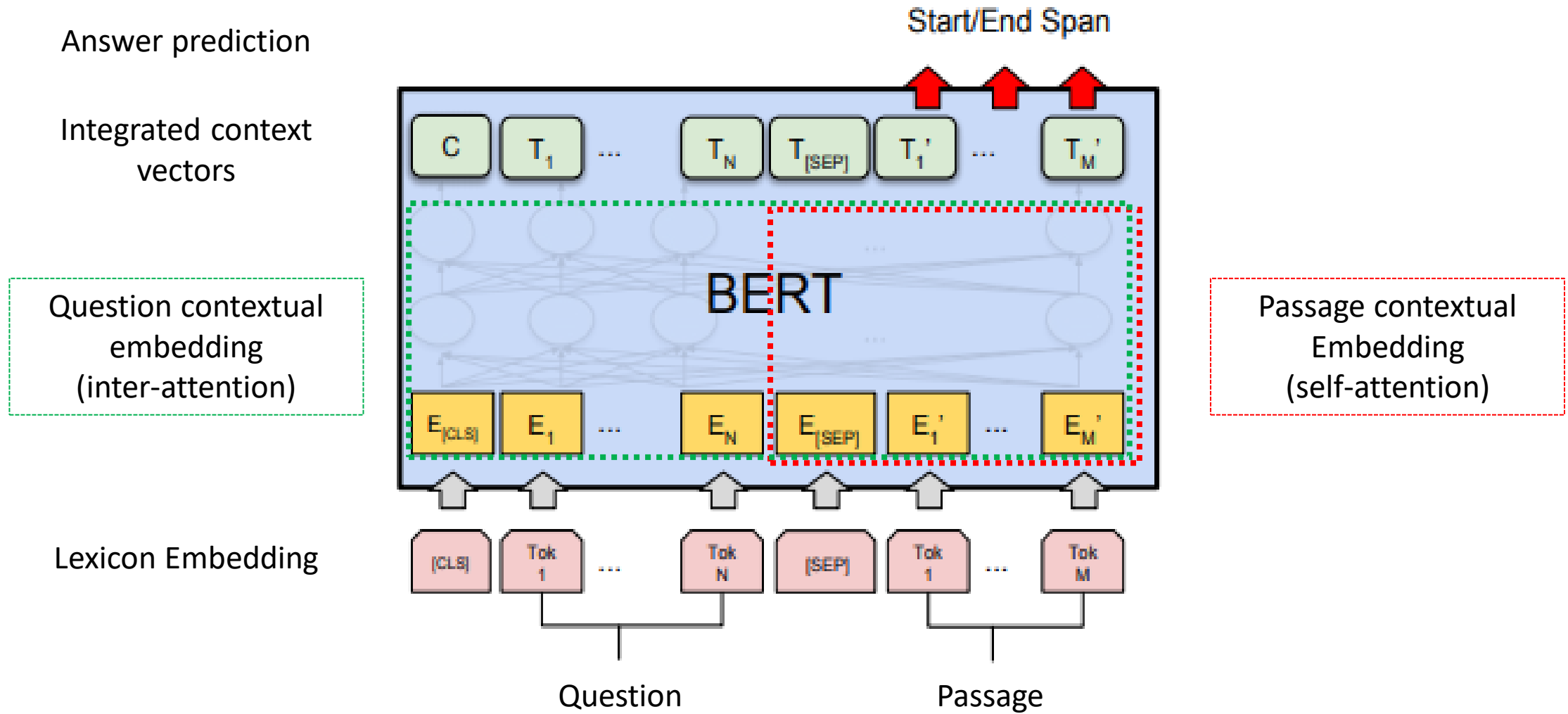


Neural MRC model: BiDAF

- Answer prediction
- Integrated context vectors
- Question contextual embedding
- Passage contextual Embedding
- Lexicon Embedding



Transformer-based MRC model: BERT



Conversational MRC models

- QA as classification given (question, text)
 - Classify each word in passage as start/end/outside of answer span
- Encoding: represent each passage word using an integrated context vector that encodes info about
 - Lexicon/word embedding
 - Passage context
 - Question context
 - **Conversation context (previous question-answer pairs)**
- Prediction: predict each word (its integrated context vector) the start and end position of answer span.

Conversational MRC models

- Pre-pending conversation history to current question or passage
 - Convert conversational QA to single-turn QA
- BiDAF++ (BiDAF for C-QA)
 - Append a feature vector encoding dialog turn number to question embedding
 - Append a feature vector encoding previous answer locations to passage embedding
- BERT (or RoBERTa)
 - Prepending dialog history to current question
 - Using BERT as
 - context embedding (self-attention)
 - Question/conversation context embedding (inter-attention)

FlowQA: explicitly encoding dialog history

- Integration Flow (IF) Layer

- Given:

- Current question Q_T , and previous questions $Q_t, t < T$
 - For each question Q_t , integrated context vector of each passage word w_t

- Output:

- Conversation-history-aware integrated context vector of each passage word
 - $w_T = \text{LSTM}(w_1, \dots, w_t, \dots, w_T)$
 - So, the entire integrated context vectors for answering previous questions can be used to answer the current question.

- Extensions of IF

- FlowDelta explicitly models the info gain thru conversation
 - GraphFLOW captures conversation flow using a graph neural network
 - *Implementing IF using Transformer with proper attention masks*

Summary: Conversational QA methods

- KBQA: Conversational QA over Knowledge Bases
 - Knowledge bases and QAs
 - C-KBQA system architecture
 - Semantic parser
 - Dialog manager
 - Response generation
 - KBQA w/o semantic parser
- Text-QA: Conversational QA over Texts
 - Tasks and datasets
 - C-TextQA system architecture
 - Conversational machine reading compression models

Outline

- Part 1: Introduction
- Part 2: Conversational QA methods
- **Part 3: Conversational search methods**
- Part 4: Case study of commercial systems

Conversational Search: Outline

- **What** is conversational search?
 - A view from TREC Conversational Assistance Track (TREC CAsT) [1]
- **Unique Challenges** in conversational search.
 - Conversational query understanding [2]
- **How** to make search **more conversational**?
 - From passive retrieval to active conversation with conversation recommendation [3]

[1] [Cast 2019: The conversational assistance track overview](#)

[2] [Few-Shot Generative Conversational Query Rewriting](#)

[3] [Leading Conversational Search by Suggesting Useful Questions](#)

Why Conversational Search

Ad hoc Search



startup seed investment



Conversational Search



How does seed investment work for startups?



Keyword-based Queries



Natural Queries

Necessity:

- Speech/Mobile Interfaces

Opportunities:

- More natural and explicit expression of information needs

Challenge:

- Query understanding & sparse retrieval

Why Conversational Search

Ad hoc Search

Startup Investing. Simplified. - SeedInvest
<https://www.seedinvest.com>

Join 300,000+ people who already use SeedInvest to find startup investment opportunities. SeedInvest is filled with investors and entrepreneurs that are passionate about building future innovation. They are accomplished individuals that invest ...

[Log In](#)
SeedInvest is a leading equity crowdfunding platform that provides individual investors ...

[Browse Offerings](#)
The following offerings are being conducted ...
[Monogram Orthopaedics](#) · [Winc](#) · [Auto Invest](#)

[See more](#) ▾

Ten Blue-Links



Conversational Search

A strategy used by seed institutional investors is the **spray and pray** type of model in which investment funds are invested in a number of companies and see which ones pick up traction. Once the start-ups they are taking on are identified then you allocate additional capital to invest in follow on rounds of financing.



[How Funding Rounds Work For Startups - Forbes](#)
F www.forbes.com/sites/alejandrocremades/2018/12/26/how-funding-rounds-wor...

Natural Responses

Necessity:

- Speech/Mobile Interfaces

Opportunities:

- Direct & Easier access to information

Challenge:

- Document understanding; combine and synthesize information

Why Conversational Search

Ad hoc Search



startup seed investment



startup seed investment amount



Single-Shot Query



Conversational Search



How does seed investment work for startups?



How does it compare with crowdfunding?



Multi-Turn Dialog

Necessity:

- N.A.

Opportunities:

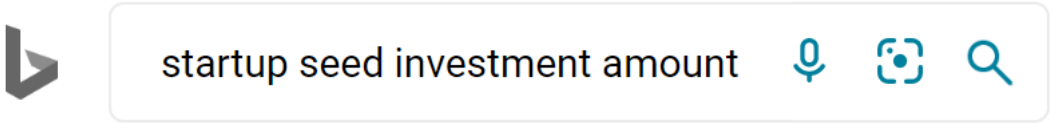
- Serving complex information needs and tasks

Challenge:

- Contextual Understanding & Memorization

Why Conversational Search

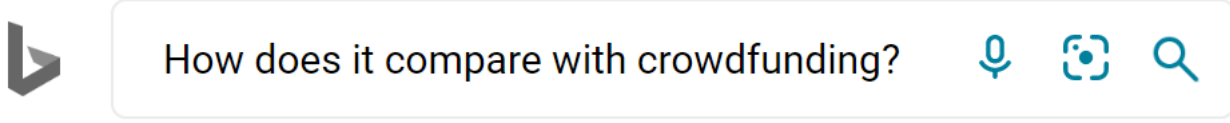
Ad hoc Search



Startup Investing. Simplified. - SeedInvest
<https://www.seedinvest.com>
Join 300,000+ people who already use SeedInvest to find startup investment opportunities. SeedInvest is filled with investors and entrepreneurs that are passionate about building future innovation. They are accomplished individuals that invest ...

Passive Serving

Conversational Search



Did you mean the comparison between seed investment and crowdfunding?

Active Engaging



Necessity:

- N.A.

Opportunities:

- Collaborative information seeking & better task assistance

Challenge:

- Dialog management, less lenient user experience

A View of Current Conversational Search

How does seed investment work?

Conversational Queries (R1)

Search

Documents

Documents

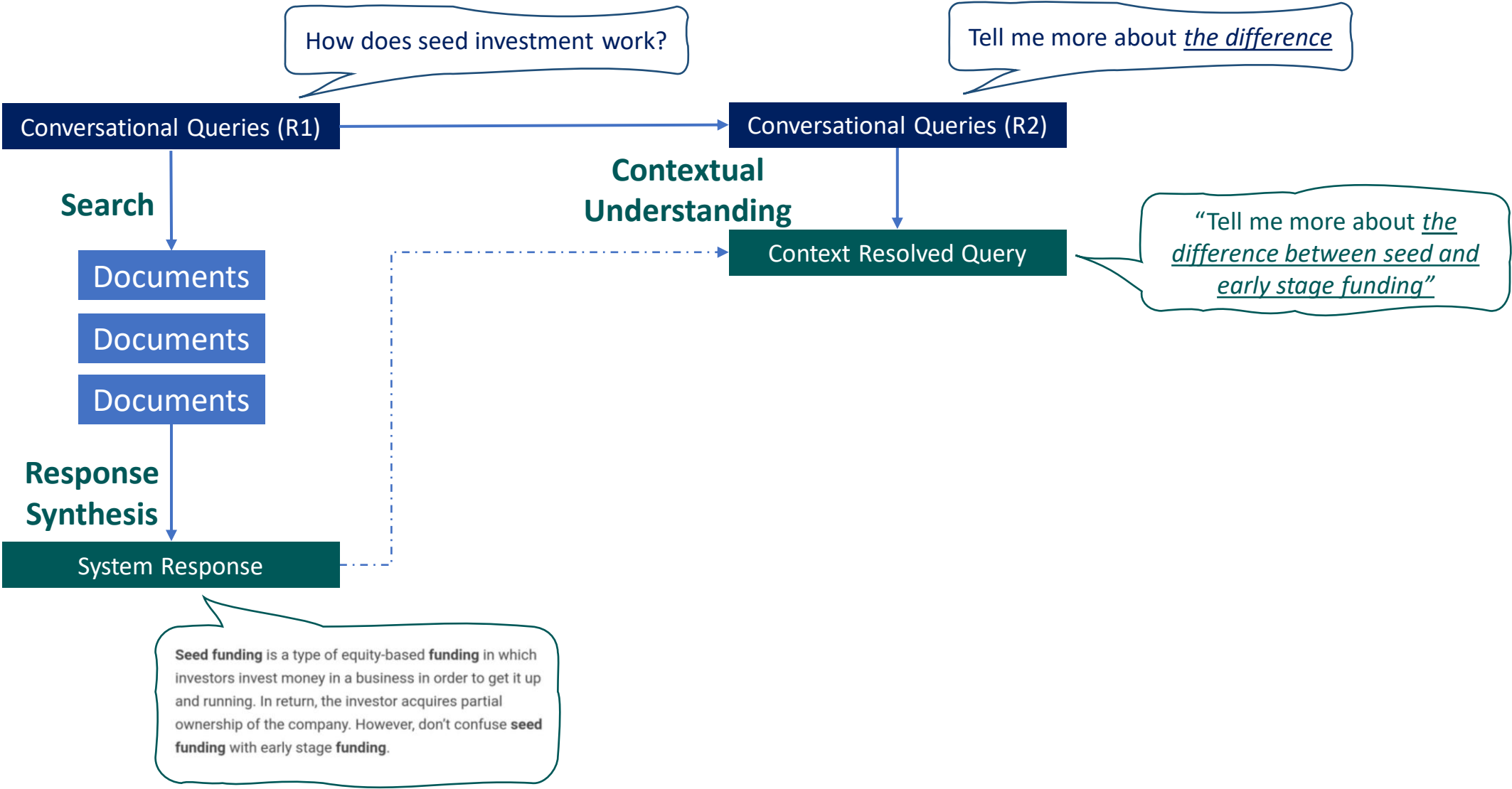
Documents

Response
Synthesis

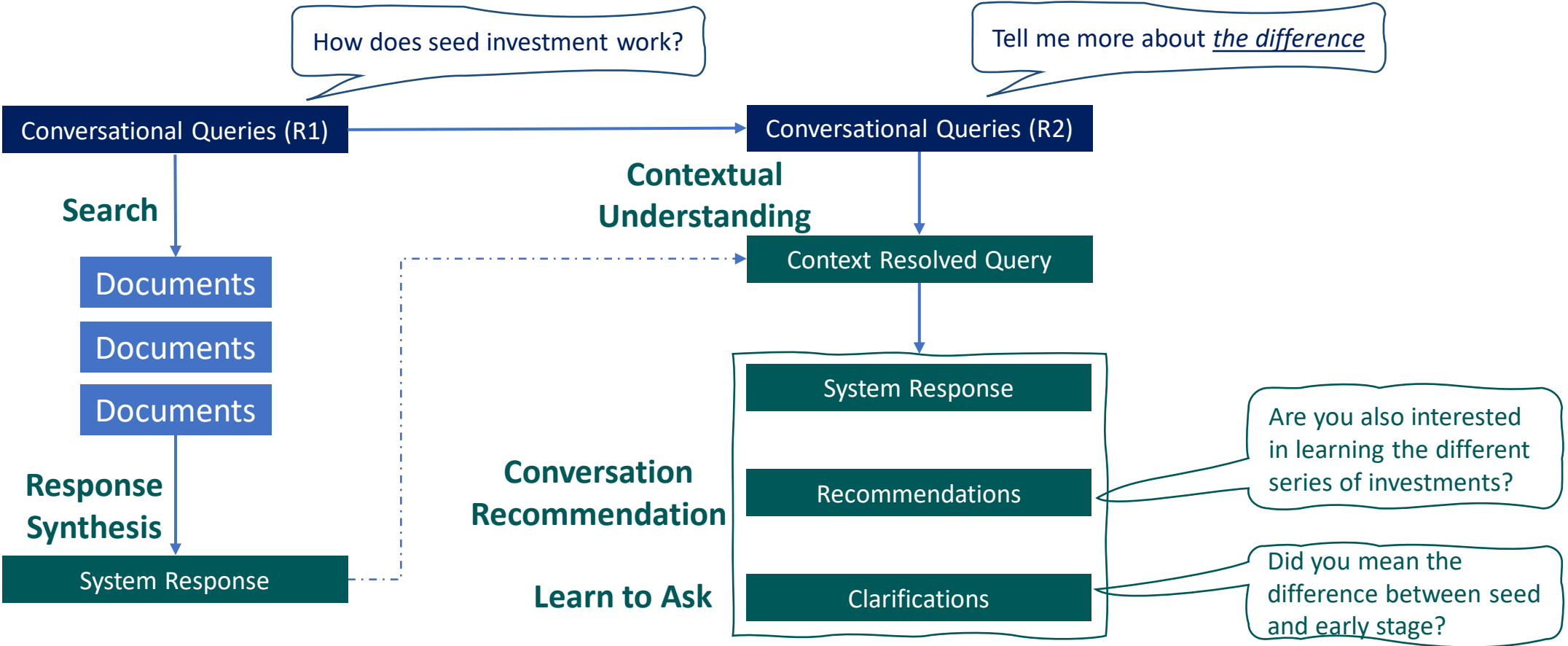
System Response

Seed funding is a type of equity-based funding in which investors invest money in a business in order to get it up and running. In return, the investor acquires partial ownership of the company. However, don't confuse seed funding with early stage funding.

A View of Current Conversational Search

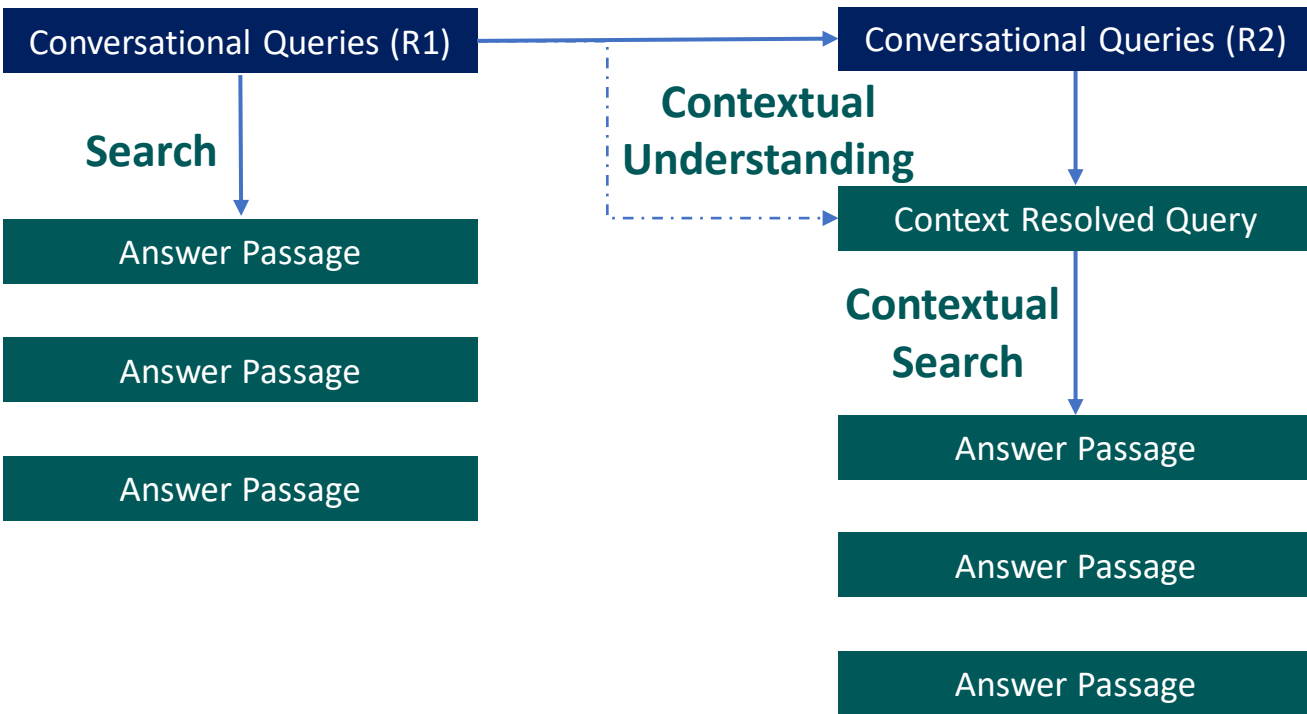


A View of Current Conversational Search



A Simpler View from TREC CAsT 2019

- “Conversational Passage Retrieval/QA”



Input:

- Manually written conversational queries
- ~20 topics, ~8 turns per topic
- Contextually dependent on previous queries

Corpus:

- MS MARCO + CAR Answer Passages

Task:

- Passage Retrieval for conversational queries

TREC CAsT 2019

- An example conversational search session

Title: head and neck cancer

Description: A person is trying to compare and contrast types of cancer in the throat, esophagus, and lungs.

1 What is throat cancer?

2 Is it treatable?

3 Tell me about lung cancer.

4 What are its symptoms?

5 Can it spread to the throat?

6 What causes throat cancer?

7 What is the first sign of it?

8 Is it the same as esophageal cancer?

9 What's the difference in their symptoms?

Input:

- Manually written conversational queries
- ~20 topics, ~8 turns per topic
- Contextually dependent on previous queries

Corpus:

- MS MARCO + CAR Answer Passages

Task:

- Passage Retrieval for conversational queries

TREC CAsT 2019

- Challenge: contextual dependency on previous conversation queries

Title: head and neck cancer

Description: A person is trying to compare and contrast types of cancer in the throat, esophagus, and lungs.

1 What is throat cancer?

2 Is it treatable?

3 Tell me about lung cancer.

4 What are its symptoms?

5 Can it spread to the throat?

6 What causes throat cancer?

7 What is the first sign of it?

8 Is it the same as esophageal cancer?

9 What's the difference in their symptoms?

Input:

- Manually written conversational queries
- ~20 topics, ~8 turns per topic
- Contextually dependent on previous queries

Corpus:

- MS MARCO + CAR Answer Passages

Task:

- Passage Retrieval for conversational queries

TREC CAsT 2019

- Learn to resolve the contextual dependency

Raw Queries

- 1 What is throat cancer?
- 2 Is **it** treatable?
- 3 Tell me about lung cancer.
- 4 What are **its** symptoms?
- 5 Can **it** spread to the throat?
- 6 What causes throat cancer?
- 7 What is the first sign of **it**?
- 8 Is **it** the same as esophageal cancer?
- 9 What's the difference in **their** symptoms?

Manual Queries provided by CAsT Y1

- 1 What is throat cancer?
- 2 Is throat cancer treatable?
- 3 Tell me about lung cancer.
- 4 What are lung cancer's symptoms?
- 5 Can lung cancer spread to the throat?
- 6 What causes throat cancer?
- 7 What is the first sign of throat cancer?
- 8 Is throat cancer the same as esophageal cancer?
- 9 What's the difference in throat cancer and esophageal cancer's symptoms?

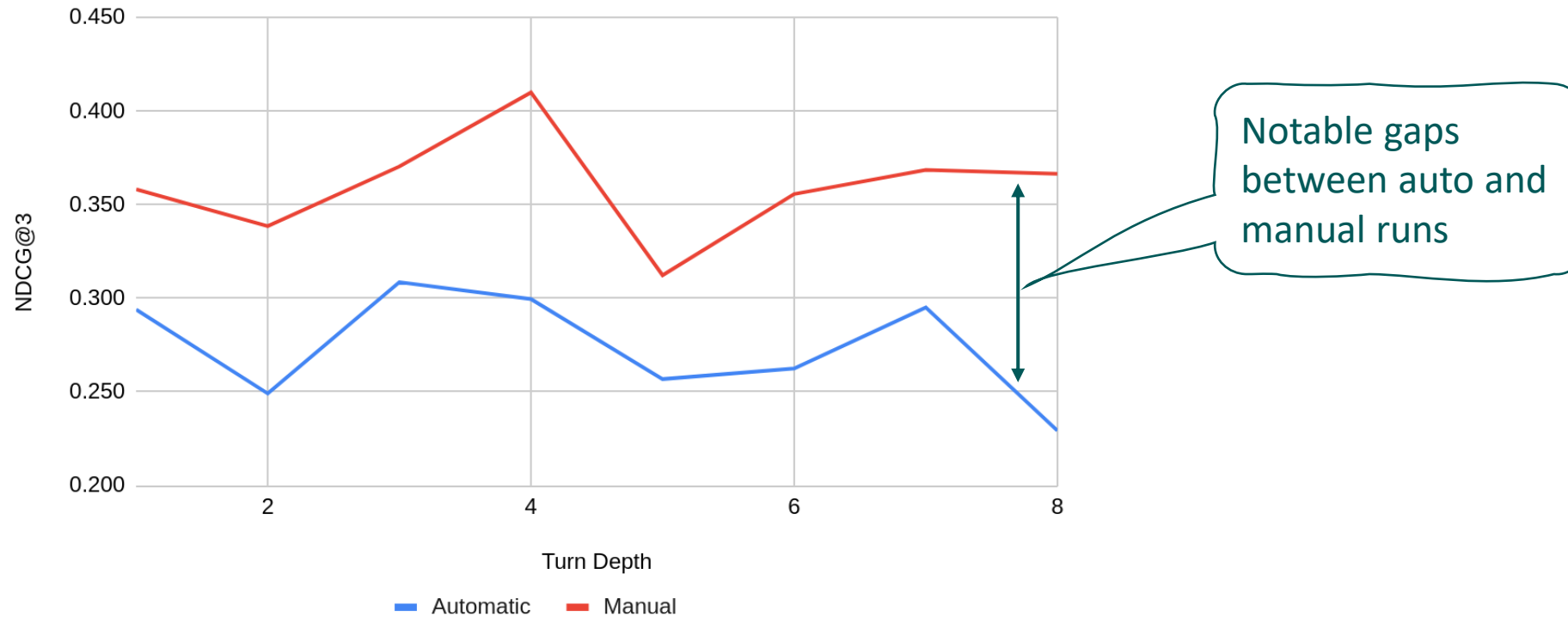
TREC CAsT 2019: Query Understanding Challenge

- Statistics in Y1 Testing Queries

Type (#. Turns)	Utterance	Mention
Pronominal (128)	How do they celebrate Three Kings Day?	they -> Spanish people
Zero (111)	What cakes are traditional?	Null -> Spanish, Three Kings Day
Groups (4)	Which team came first?	which team -> Avengers, Justice League
Abbreviations (15)	What are the main types of VMs ?	VMs -> Virtual Machines

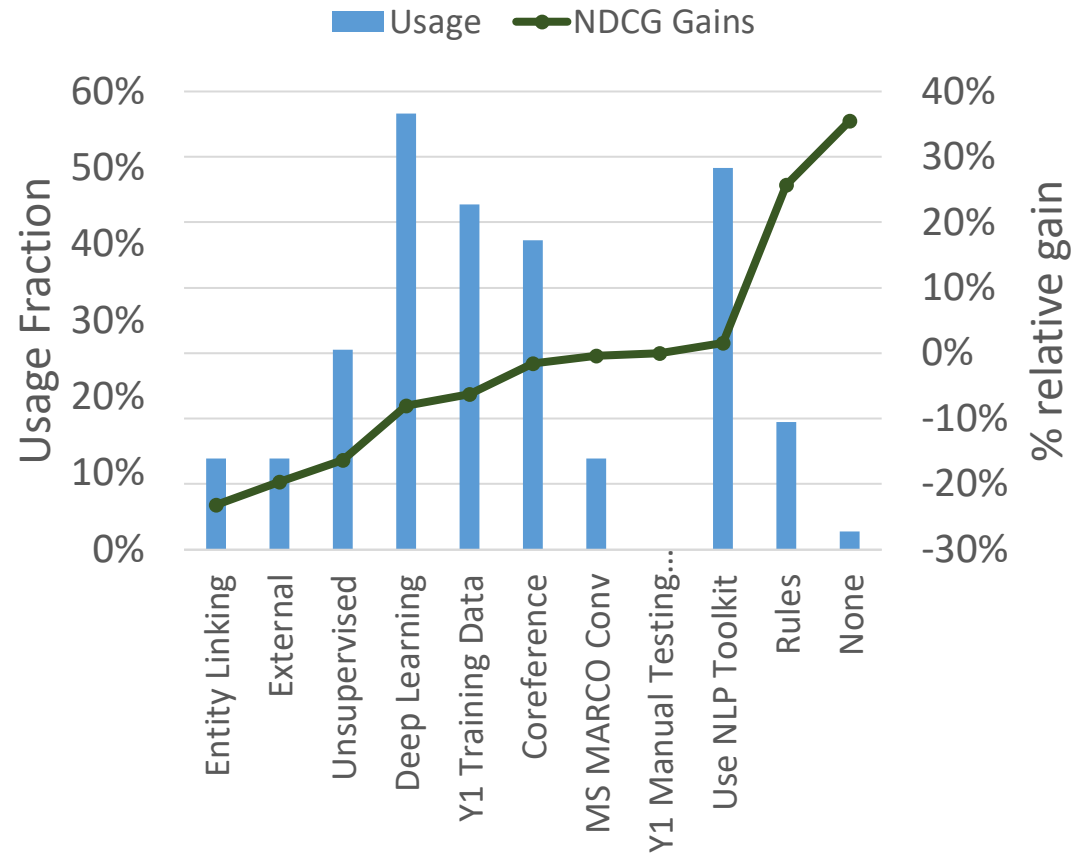
TREC CAsT 2019: Result Statics

- Challenge from contextual query understanding



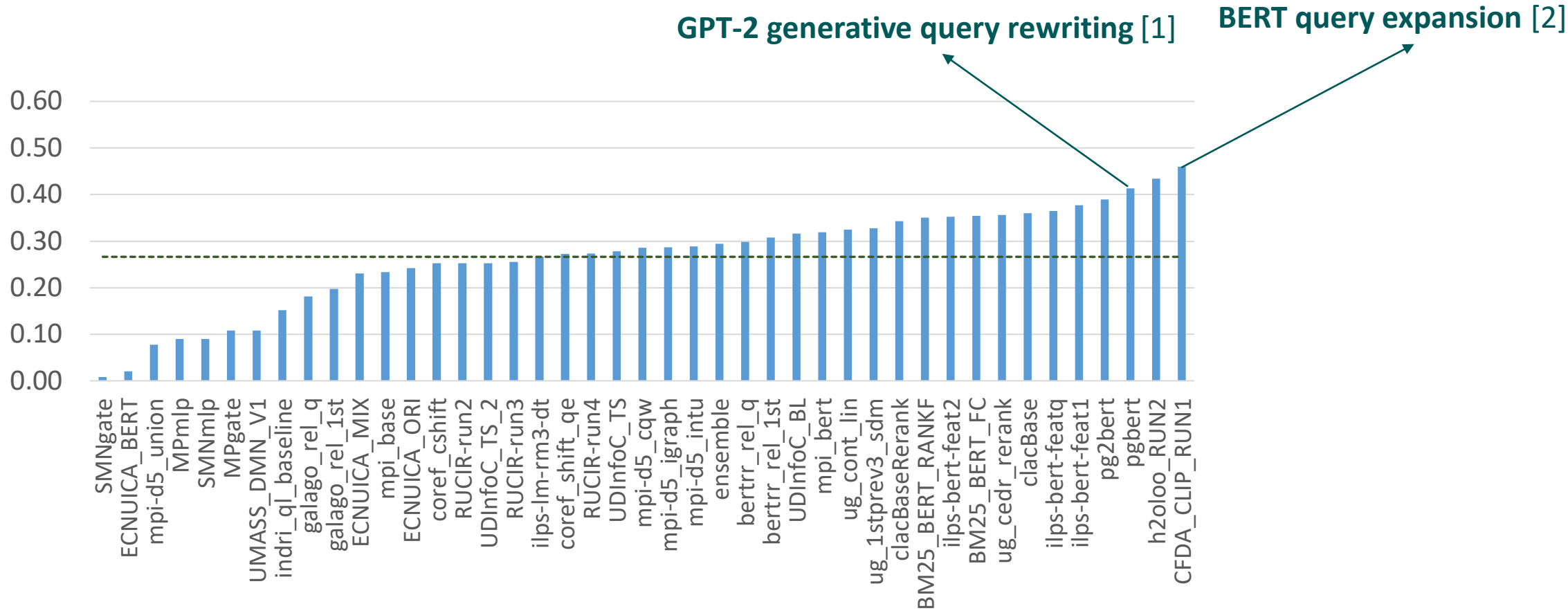
TREC CAsT 2019: Techniques

- Techniques used in Query Understanding



TREC CAsT 2019: Notable Solutions

- Automatic run results

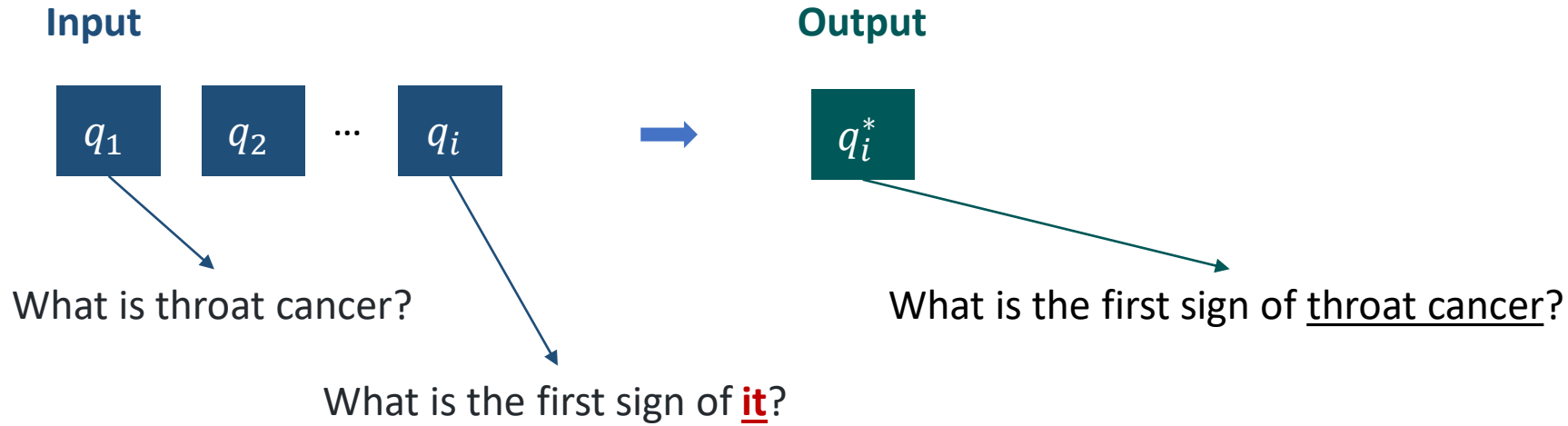


[1] Vakulenko et al. 2020. Question Rewriting for Conversational Question Answering

[2] Lin et al. 2020. Query Reformulation using Query History for Passage Retrieval in Conversational Search

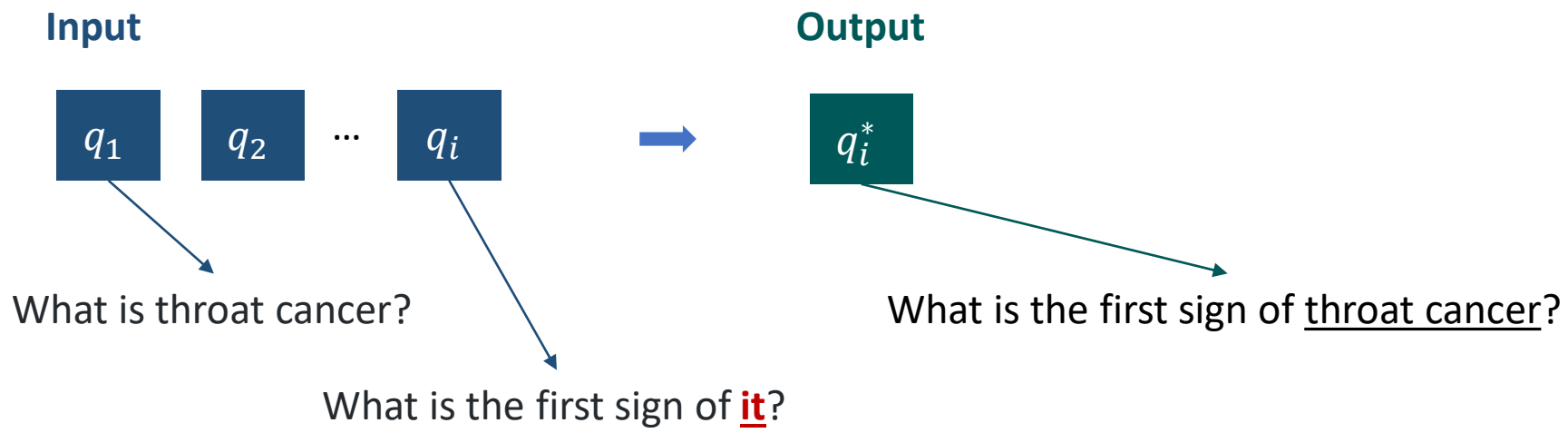
Conversational Query Understanding Via Rewriting

- Learn to rewrite a full-grown context-resolved query

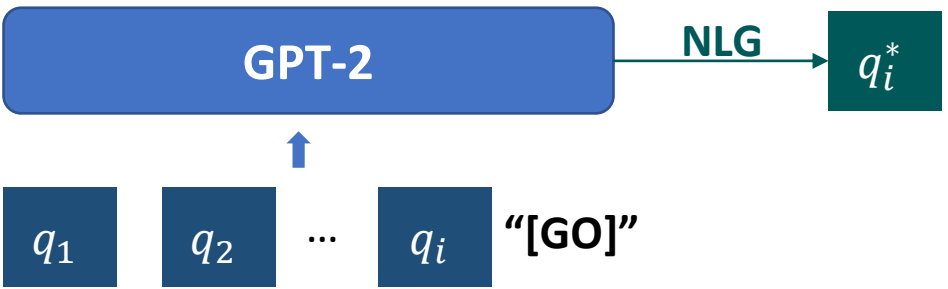


Conversational Query Understanding Via Rewriting

- Learn to rewrite a full-grown context-resolve query

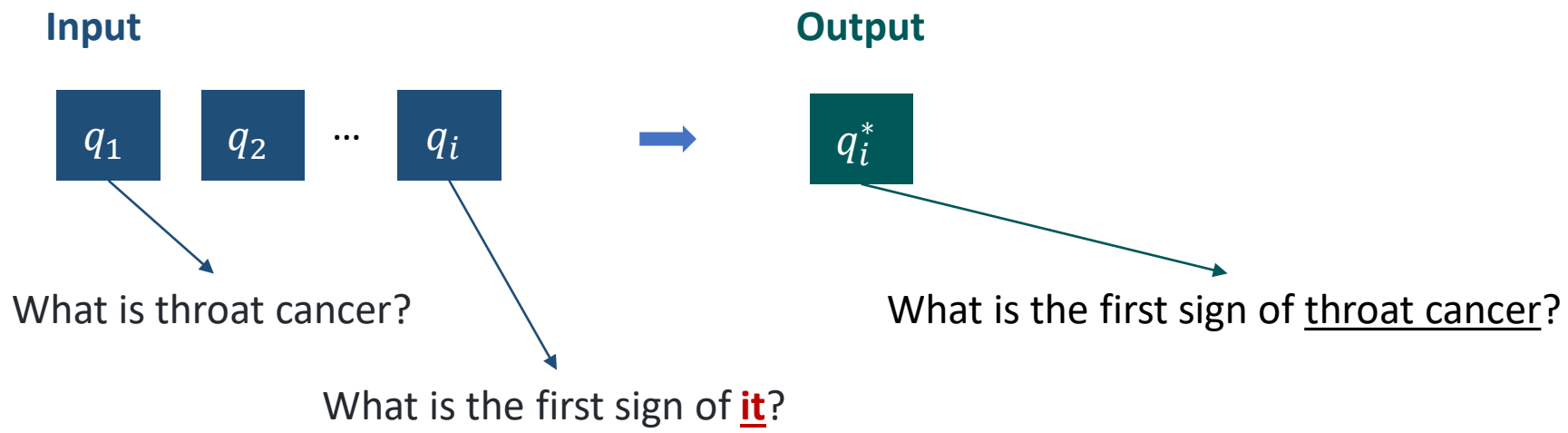


- Leverage pretrained NLG model (GPT-2) [1]

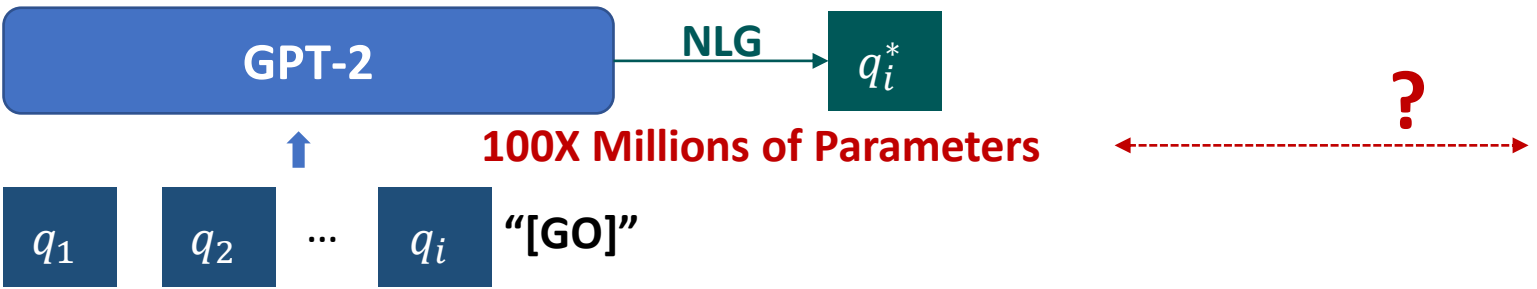


Conversational Query Understanding Via Rewriting

- Learn to rewrite a full-grown context-resolve query



- Concern: Limited training data



CAST Y1 Data:

- Manually written conversational queries
- 50 topics, 10 turns per topic
 - 20 topics with TREC relevance labels

500 Manual Rewrite Labels

Few-Shot Conversational Query Rewriting

- Train conversational query rewriter with the help of ad hoc search data

Ad hoc Search

- Existing billions of search sessions
- Lots of high-quality public benchmarks

Conversational Search

- Production scenarios still being explored
- Relative new topic, fewer available data

Few-Shot Conversational Query Rewriting

- Leveraging ad hoc search sessions for conversational query understanding

Ad hoc Search

▶ startup seed investment 🔊 📷 🔍

▶ startup seed investment amount 🔊 📷 🔍

Conversational Search

▶ How does seed investment work for startups? 🔊 📷 🔍

▶ How does it compare with crowdfunding? 🔊 📷 🔍

Ad hoc Search Sessions

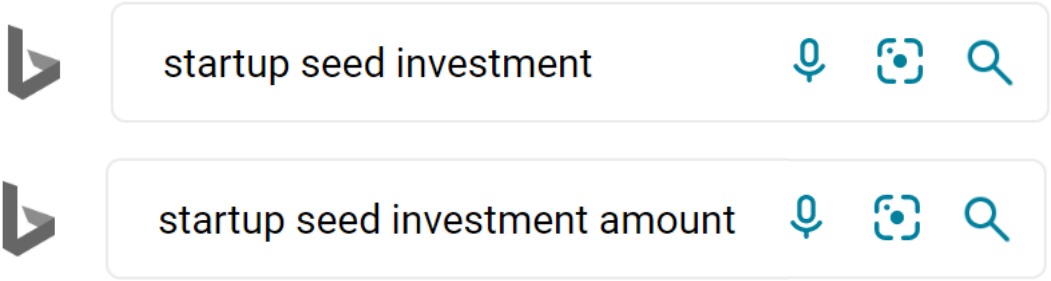
Conversational Rounds



Few-Shot Conversational Query Rewriting

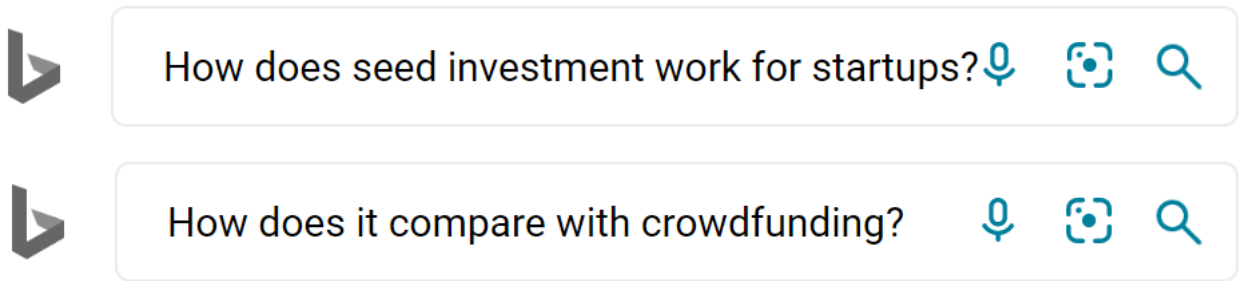
- Leveraging ad hoc search sessions for conversational query understanding

Ad hoc Search



Two search bars are shown. The first contains the text "startup seed investment" and the second contains "startup seed investment amount". Each search bar includes a microphone icon, a camera icon, and a magnifying glass icon.

Conversational Search



Two search bars are shown. The first contains the text "How does seed investment work for startups?" and the second contains "How does it compare with crowdfunding?". Each search bar includes a microphone icon, a camera icon, and a magnifying glass icon.

Ad hoc Search Sessions

Conversational Rounds

Challenges?

- Available only in commercial search engines
 - Approximate sessions available in [MS MARCO](#)
- Keyword-ese
 - Filter by question words

Few-Shot Conversational Query Rewriting

- Leveraging ad hoc search sessions for conversational query understanding

Ad hoc Search

Two examples of ad hoc search queries in a search bar interface. Each example shows a search bar with a left arrow icon, a microphone icon, a camera icon, and a search icon. The first query is "startup seed investment" and the second is "startup seed investment amount".

Conversational Search

Two examples of conversational search queries in a search bar interface. Each example shows a search bar with a left arrow icon, a microphone icon, a camera icon, and a search icon. The first query is "How does seed investment work for startups?". The second query is "How does it compare with crowdfunding?", where "it" is highlighted with a red box and a red question mark above it.

Ad hoc Search Sessions

Conversational Rounds

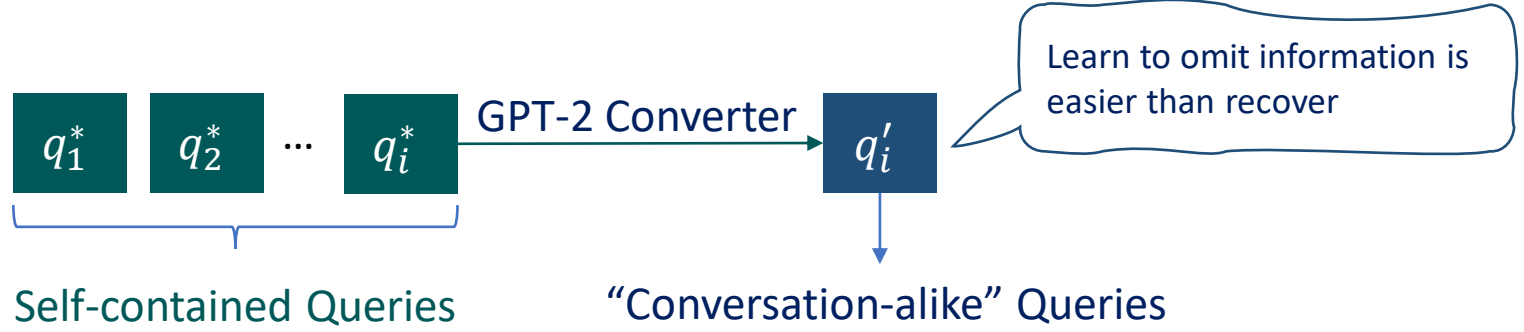
Challenges?

- Available only in commercial search engines
 - Approximate sessions available in [MS MARCO](#)
- Keyword-ese
 - Filter by question words
- **No explicit context dependency?**

Few-Shot Conversational Query Rewriting: Self-Training

- Learn to convert ad hoc sessions to conversational query rounds

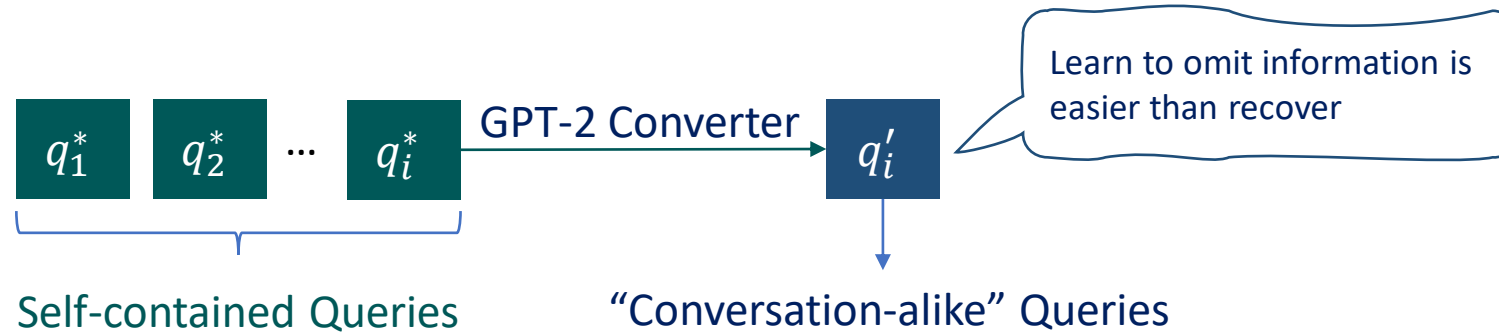
“Contextualizer”: make ad hoc sessions more conversation-alike



Few-Shot Conversational Query Rewriting: Self-Training

- Learn to convert ad hoc sessions to conversational query rounds

“Contextualizer”: make ad hoc sessions more conversation-alike



Training:

- X (Self-contained q): Manual rewrites of CAst Y1 conversational sessions
- Y (Conversation-alike q): Raw queries in CAst Y1 sessions

Inference:

- X (Self-contained q): Ad hoc questions from MS MARCO sessions
- Y (Conversation-alike q): Auto-converted conversational sessions

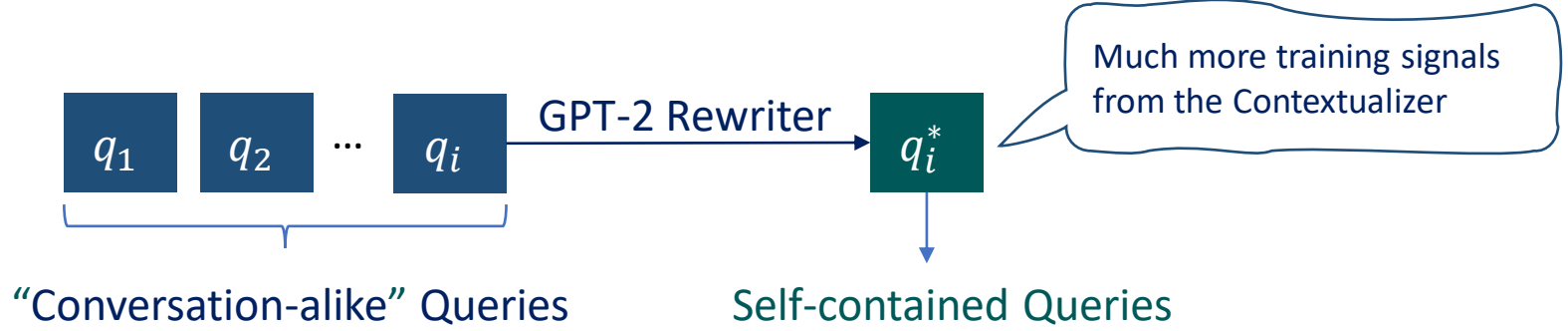
Model:

- Any pretrained NLG model: GPT-2 Small in this Case

Few-Shot Conversational Query Rewriting: Self-Training

- Leverage the auto-converted conversation-ad hoc session pairs

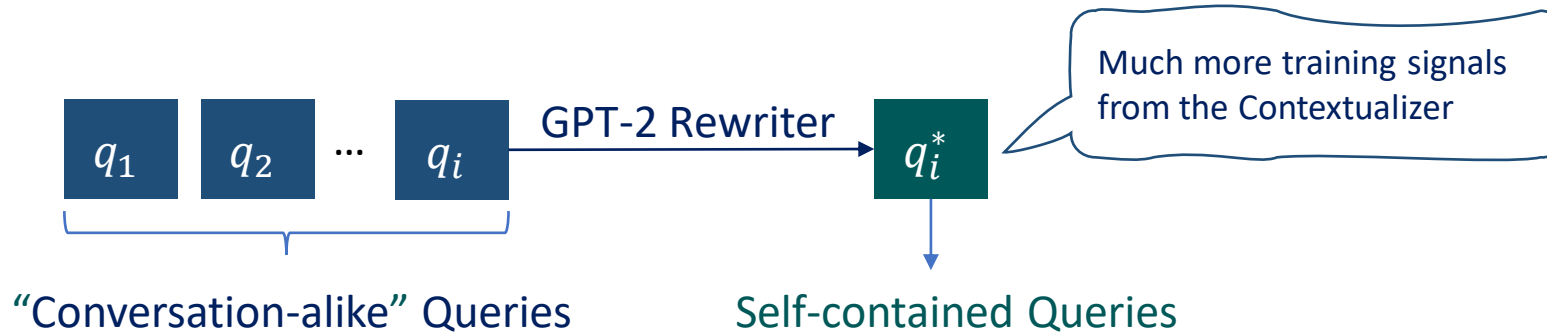
“Rewriter”: recover the full self-contained queries from conversation rounds



Few-Shot Conversational Query Rewriting: Self-Training

- Leverage the auto-converted conversation-ad hoc session pairs

“**Rewriter**”: recover the full self-contained queries from conversation rounds



Training:

- X (Conversation-alike q): Auto-converted from the Contextualizer
- Y (Self-contained q): Raw queries from ad hoc MARCO sessions

Inference:

- X (Conversation-alike q): CAsT Y1 raw conversational queries
- Y (Self-contained q): auto-rewritten queries that are more self-contained

Model:

- Any pretrained NLG model: another GPT-2 Small in this Case

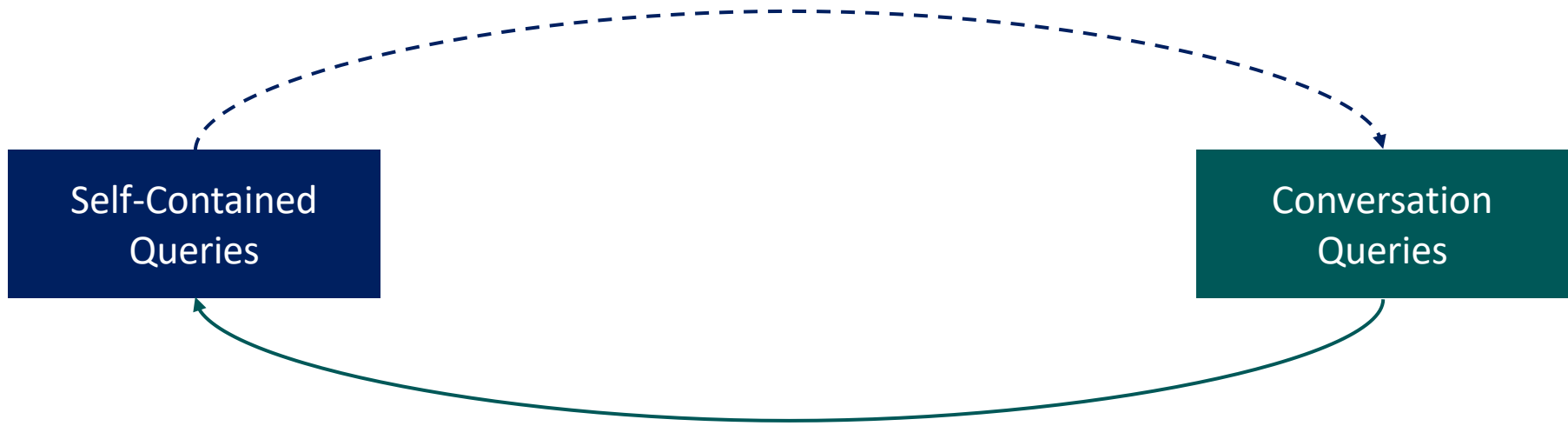
Few-Shot Conversational Query Rewriting: Self-Training

- The full “self-learning” loop

GPT-2 Converter: Convert ad hoc sessions to conversation-alike sessions

- learn from a few conversational queries with manual rewrites

Learn to omit information is easier than recover

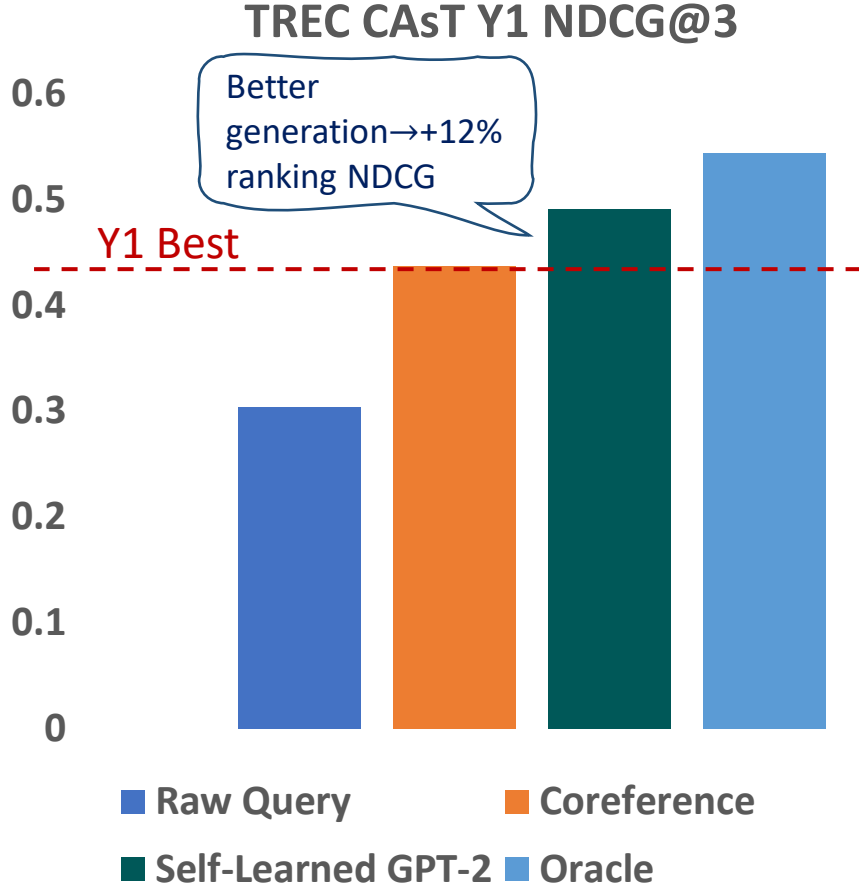
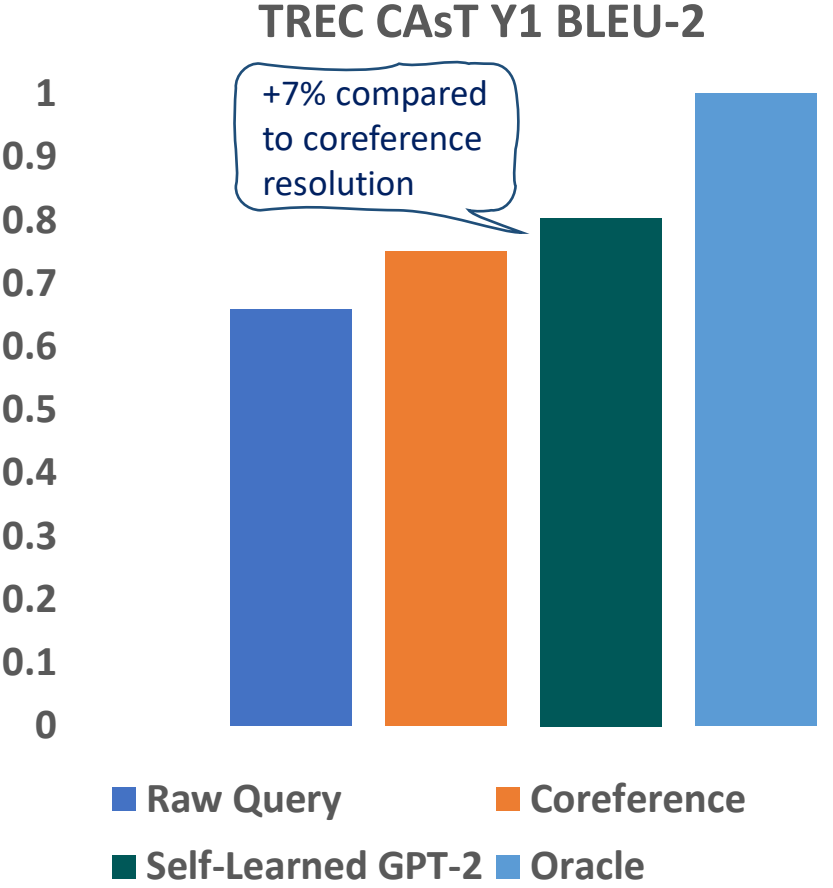


GPT-2 Rewriter: Rewrite conversational queries to self-contained ad hoc queries

- learn from the large amount of auto-converted “ad hoc” ↔ “conversation alike” sessions

Much more training signals from the Contextualizer

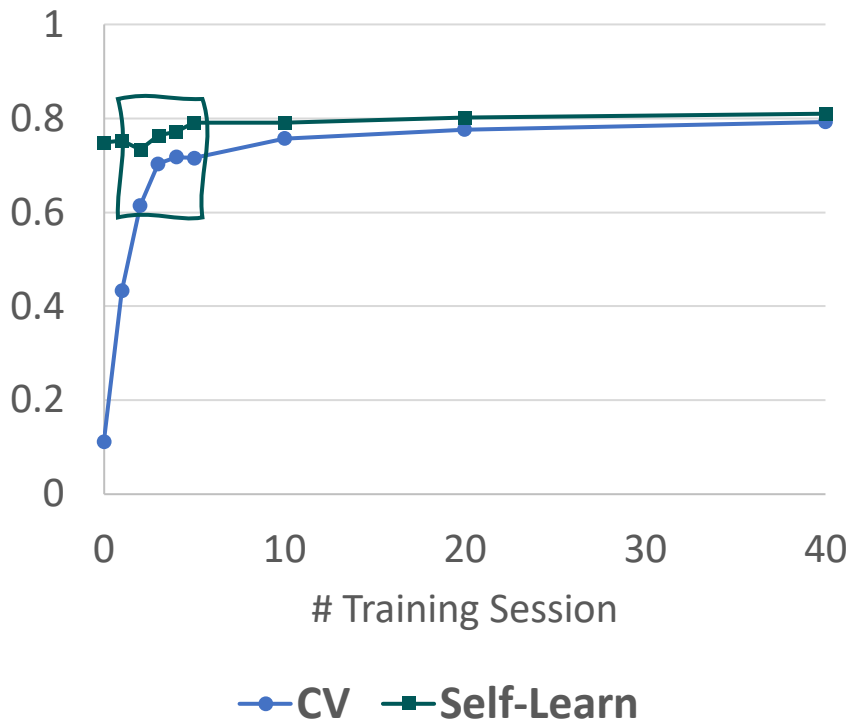
Few-Shot Conversational Query Rewriting: Results



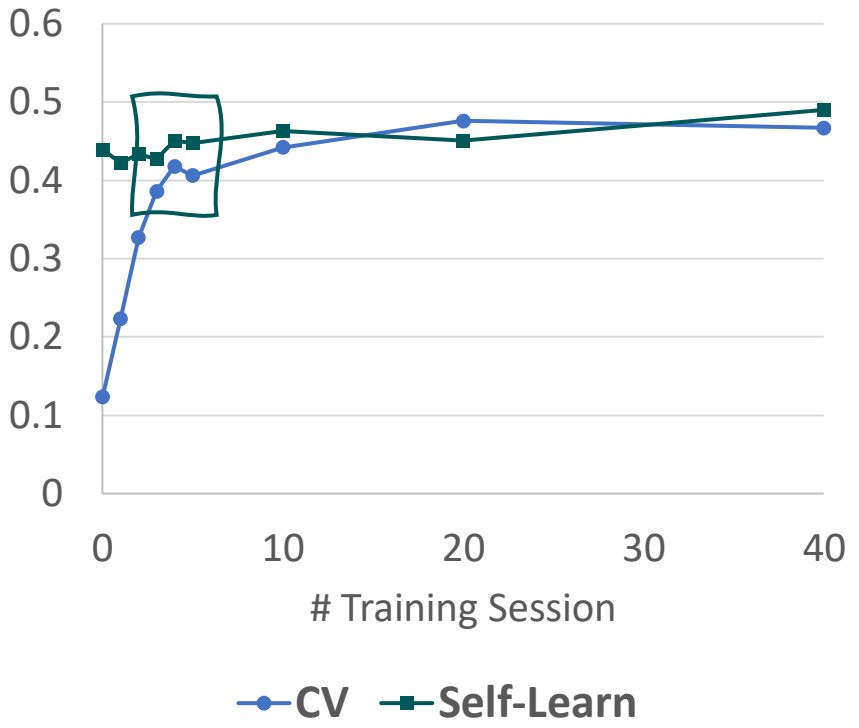
How Few-shot Can Pretrained NLG Models Be?

- Five Sessions are all they need?

BLEU-2 on CAst Y1

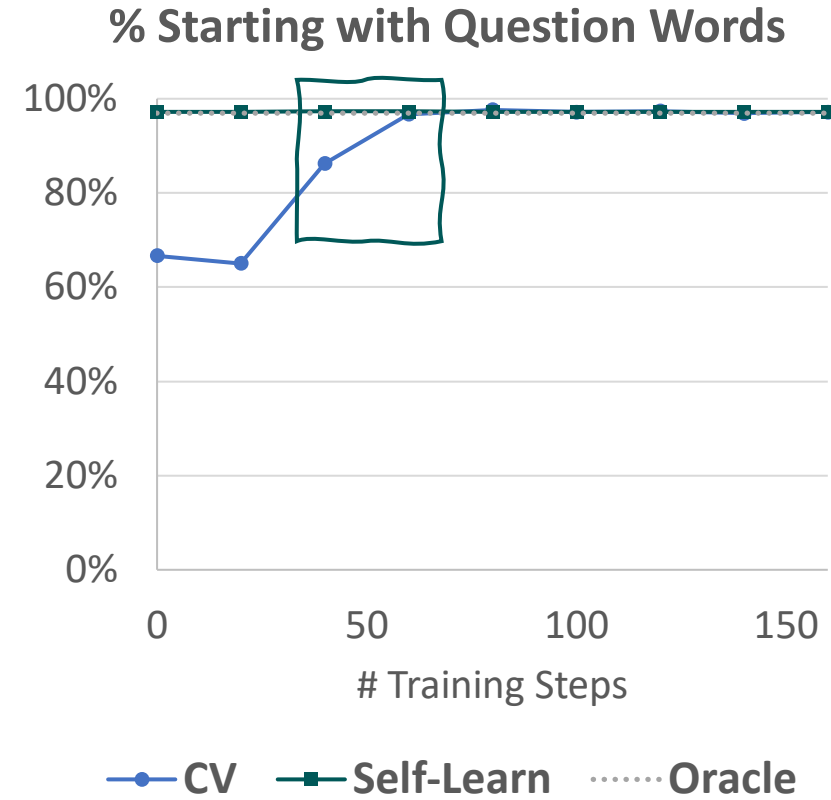
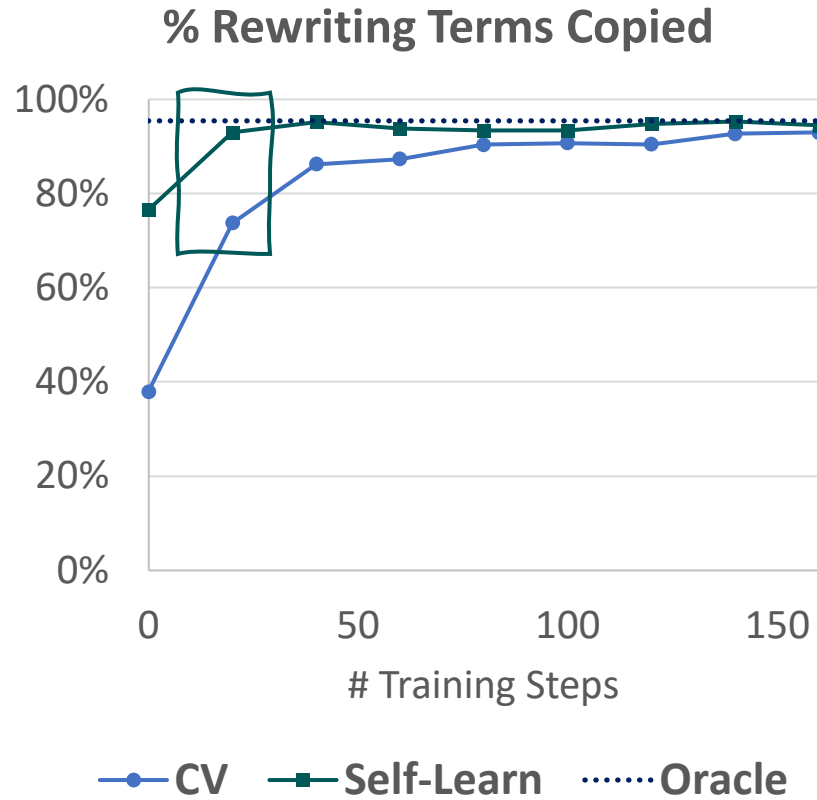


NDCG@3 on CAst Y1



What is learned?

- More about learning the task format, nor the semantics
 - Semantic mostly in the pretrained weights



Auto-rewritten Examples: Win

- Surprisingly good at Long-term dependency and Group Reference

Q_6 What causes **throat cancer**?

Q_7 What is the first sign of it?

Q_8 Is it the same as **esophageal cancer**?

Q_9 What's the difference in their symptoms?

Oracle What's the difference in **throat cancer and esophageal cancer's** symptoms?

Output What's the difference between **throat cancer and esophageal cancer**?

Auto-rewritten Examples: Win

- More “fail to rewrite”

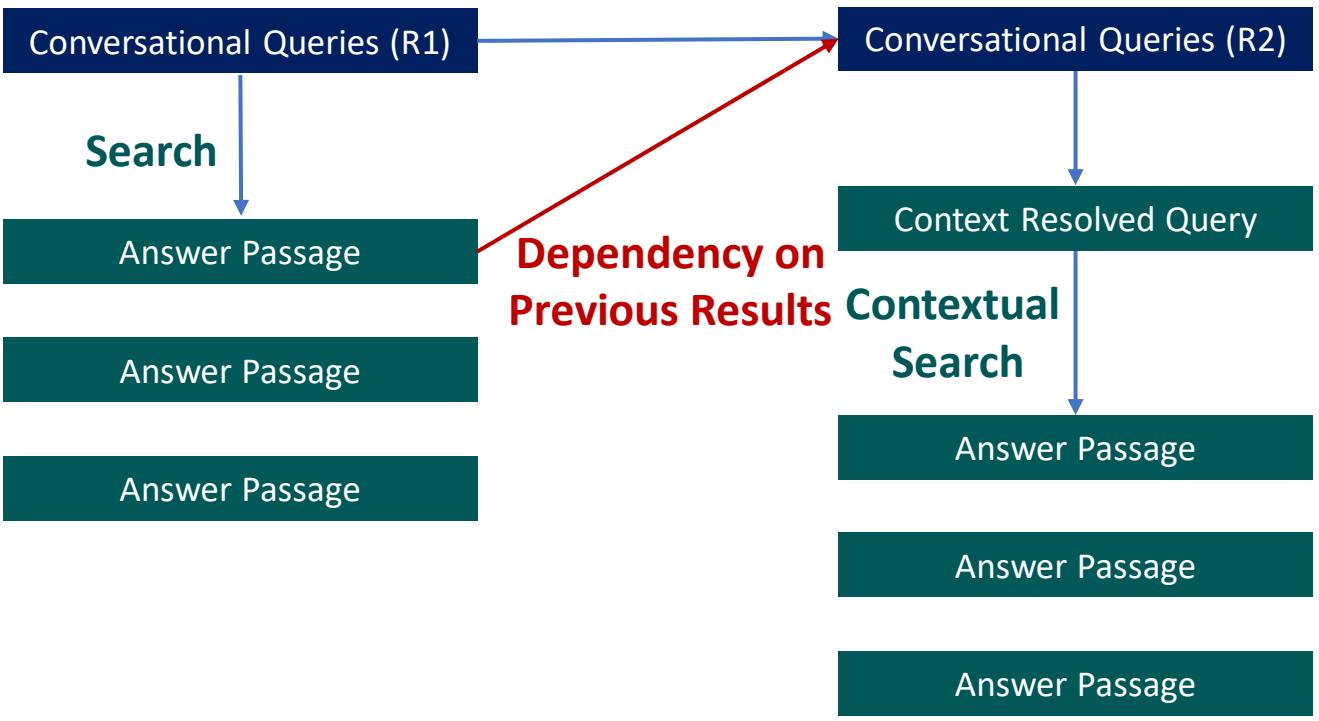
Q_1	What are the types of pork ribs ?
Q_2	What are baby backs?
Q_3	What are the differences with spareribs?
Q_4	What are ways to cook them?
Q_5	How <u>about</u> on the bbq?

Oracle	How do you cook pork ribs on the bbq?
--------	--

Output	How about on the bbq?
--------	-----------------------

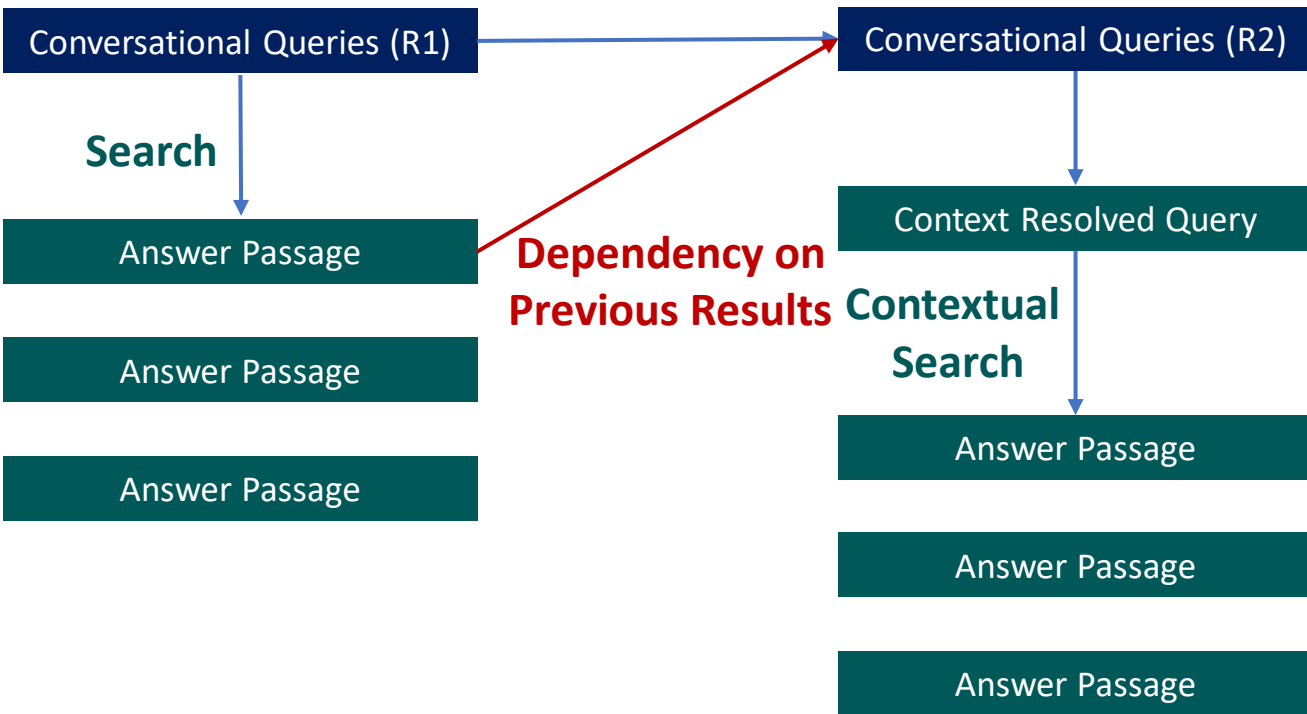
CAsT Y2: More Realistic Conversational Dependencies

- More interactions between queries and system responses



CAsT Y2: More Realistic Conversational Dependencies

- More interactions between queries and system responses



Q1: How did snowboarding begin?

R1: ...The development of snowboarding was inspired by skateboarding, surfing and skiing. The first snowboard, the Snurfer, was invented by Sherman Poppen in 1965. **Snowboarding became a Winter Olympic Sport in 1998.**

Q2: Interesting. That's later than I expected. Who were the winners?

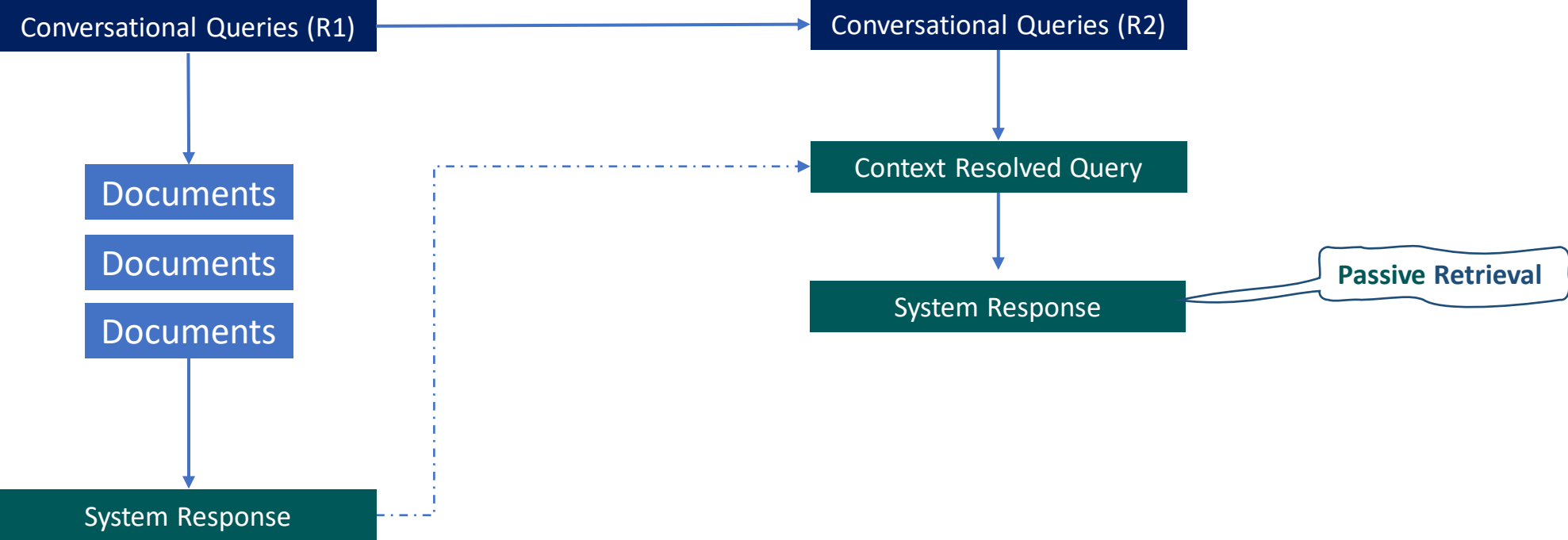
Manual rewrites:

Who were the winners of snowboarding events in the 1998 Winter Olympics?

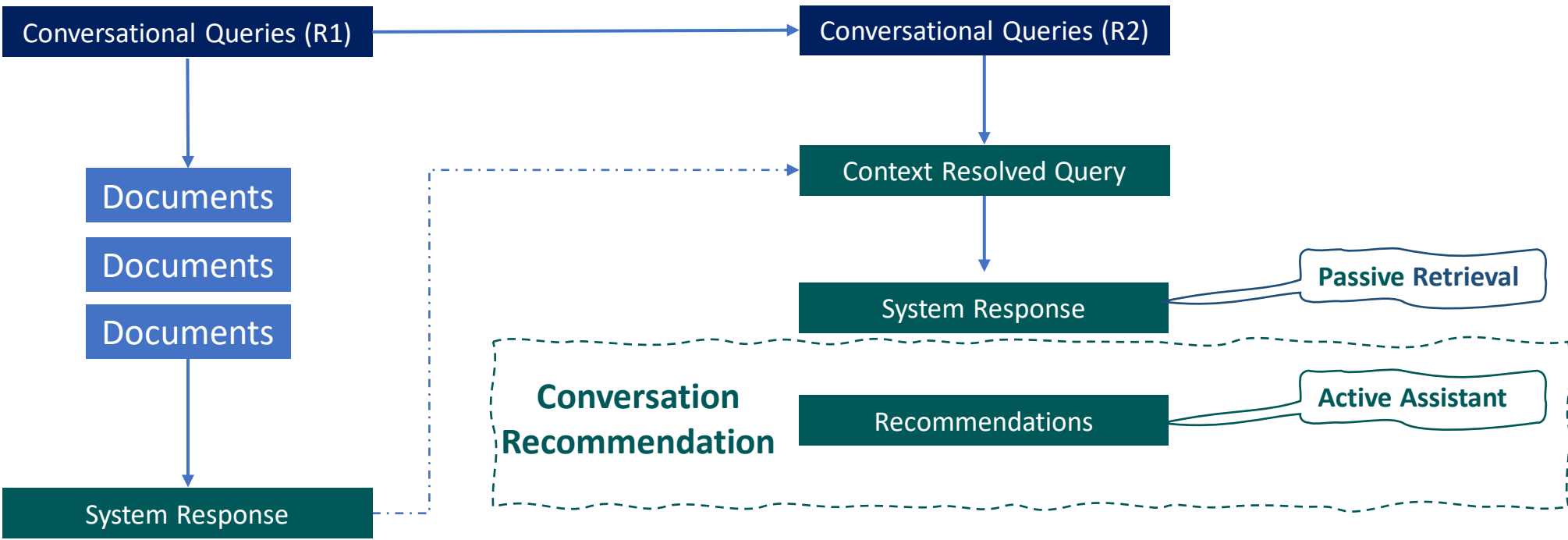
Auto rewrites without considering response:

Who were the winners of the snowboarding contest?

From Passive Information Supplier to Active Assistant

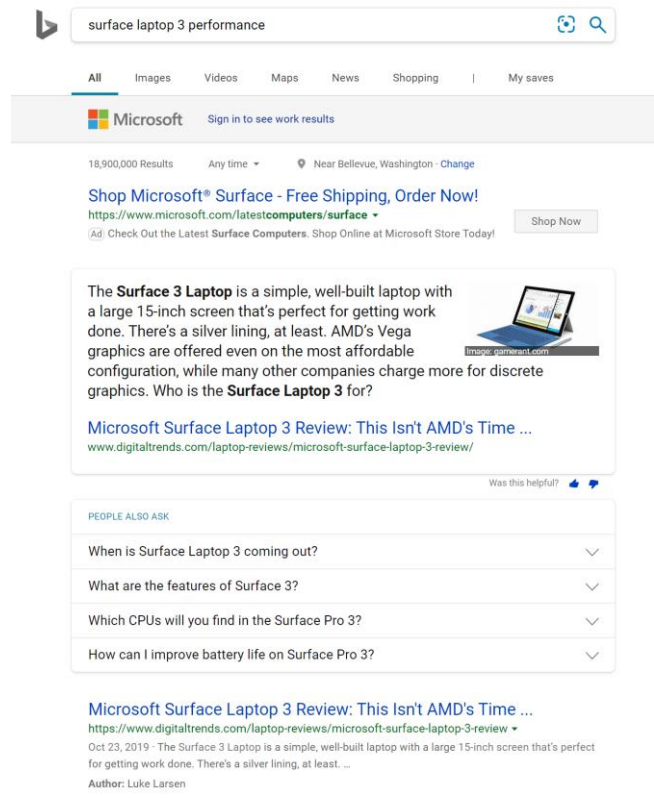
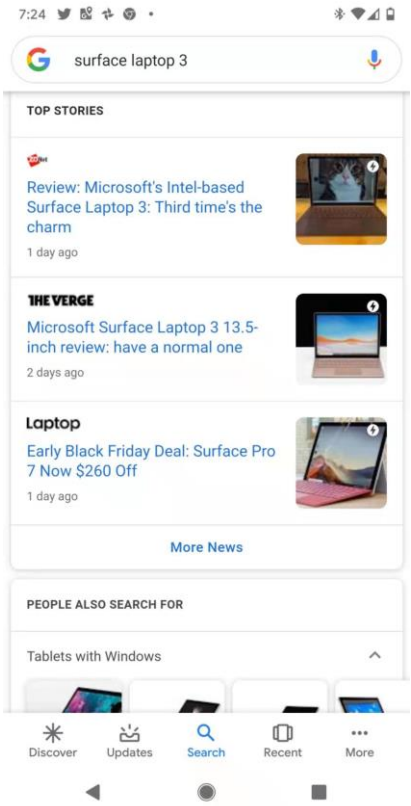


From Passive Information Supplier to Active Assistant

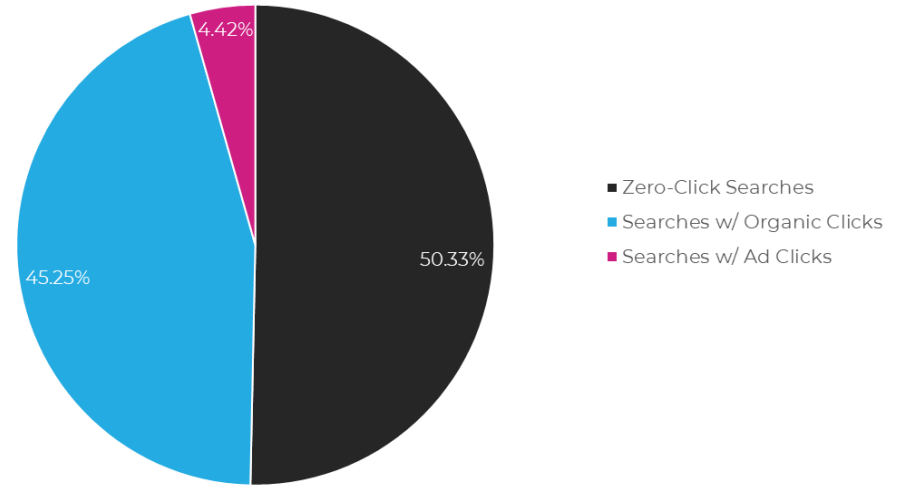


Making Search Engines More Conversational

- Search is moving from "ten blue links" to conversational experiences



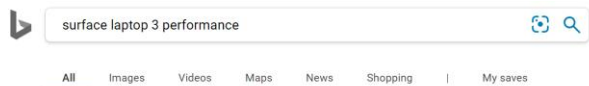
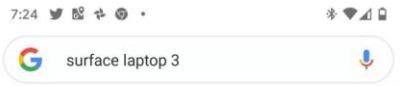
Paid, Organic, & Zero-Click Searches in Google (June 2019)
data from 40M+ browser-based searches on millions of desktop & mobile devices in the United States



<https://sparktoro.com/blog/less-than-half-of-google-searches-now-result-in-a-click/>

Making Search Engines More Conversational

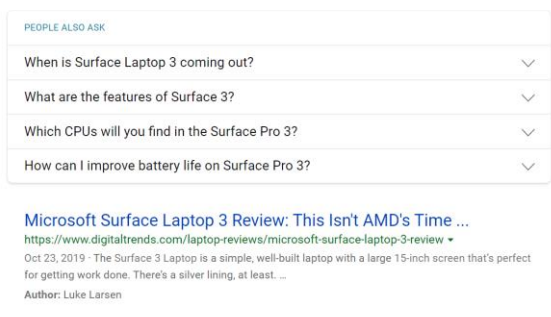
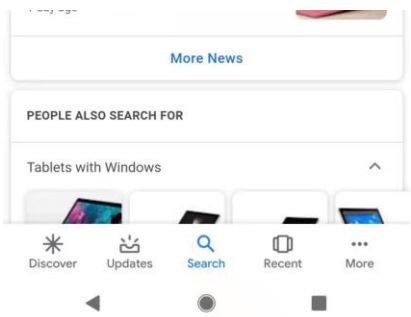
- Search is moving from "ten blue links" to conversational experiences



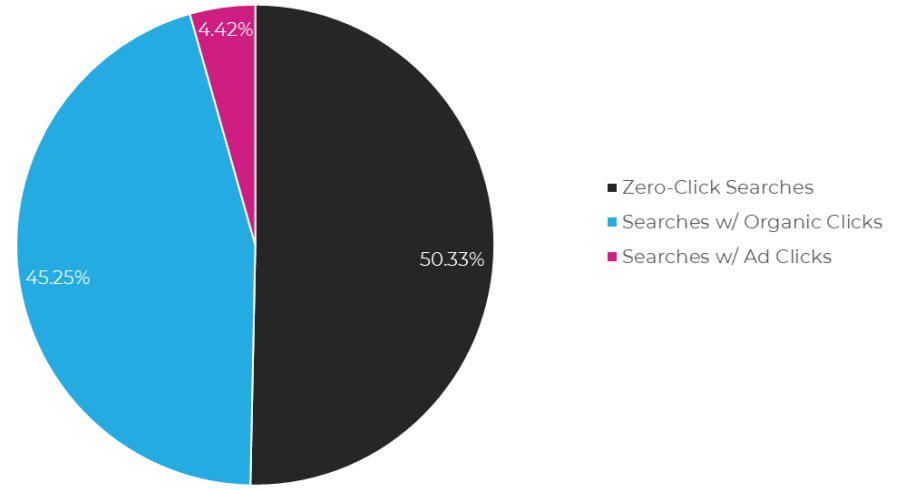
Yet most queries are not “conversational”

1. Users are trained to use keywords
2. Less conversational queries
3. Less learning signal
4. Less conversational experience

“Chicken and Egg” Problem



Paid, Organic, & Zero-Click Searches in Google (June 2019)
data from 40M+ browser-based searches on millions of desktop & mobile devices in the United States



<https://sparktoro.com/blog/less-than-half-of-google-searches-now-result-in-a-click/>

Conversation Recommendation: “People Also Ask”

- Promoting more conversational experiences in search engines
- E.g., for keyword query "Nissan GTR"
 - Provide the follow questions:

PEOPLE ALSO ASK	
What is Nissan GTR?	∨
How to buy used Nissan GTR in Pittsburgh?	∨
Does Nissan make sports car?	∨
Is Nissan Leaf a good car?	∨

Conversation Recommendation: Challenge

- Relevant != Conversation Leading/Task Assistance
- User less lenient to active recommendation

PEOPLE ALSO ASK		
What is Nissan GTR?	[Duplicate]	∨
How to buy used Nissan GTR in Pittsburgh?	[Too Specific]	∨
Does Nissan make sports car?	[Prequel]	∨
Is Nissan Leaf a good car?	[Miss Intent]	∨

Conversation Recommendation: Beyond Relevance

- Recommending useful conversations that
 - Help user complete their information needs
 - Assist user with their task
 - Provide meaningful explorations

Relevant

PEOPLE ALSO ASK	
What is Nissan GTR?	∨
How to buy used Nissan GTR in Pittsburgh?	∨
Does Nissan make sports car?	∨
Is Nissan Leaf a good car?	∨



Relevant & Useful

PEOPLE ALSO ASK	
What are the pros and cons of Nissan GT-R?	∨
Is the Nissan GT-R the ultimate street car?	∨
Why is the Nissan GT-R known as 'the godzilla'?	∨
How fast can the Nissan GT-R go?	∨

Usefulness Metric & Benchmark

- Manual annotations on Bing query, conversation recommendation pairs

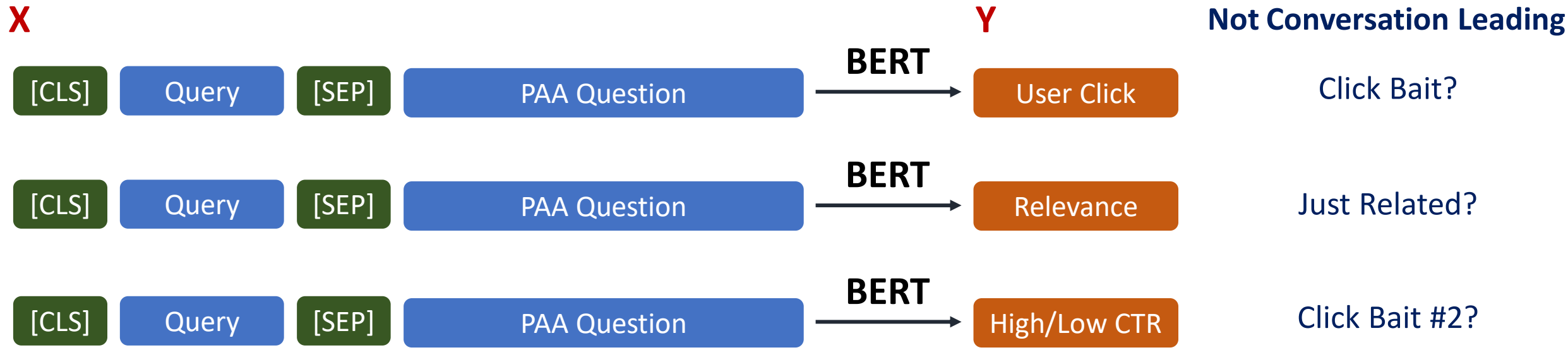
Query	Question Suggestion	Gold Label
used washer and dry	Can I store a washer and dryer in the garage ?	Misses Intent
best questions to ask interviewer	What should I ask in an interview ?	Dup. w/ Q
medicaid expansion	Did Florida accept Medicaid expansion ?	Too Specific
verizon yahoo purchase	Who bought out Yahoo ?	Prequel
jaundice in newborns	How to tell if your newborn has jaundice ?	Dup. w/ Ans.
jonestown massacre	What was in the Kool-Aid at Jonestown ?	Useful
affirmative action	Who does affirmative action benefit ?	Useful
best hair clippers	What clippers do barbers use ?	Useful

Types of non-useful ones.
• Crucial for annotation consistency

A higher bar of being useful

Conversation Recommendation Model: Multi-Task BERT

- BERT seq2seq in the standard multi-task setting



Conversation Recommendation: Session Trajectory

- Problem: the previous 3 signals were prone to learning **click-bait**
 - We need more information about how users seek new information
- Solution: **imitate how users issue queries in sessions**

4. Millions of sessions for imitation learning



Task: classify whether the potential next query was issued by the user

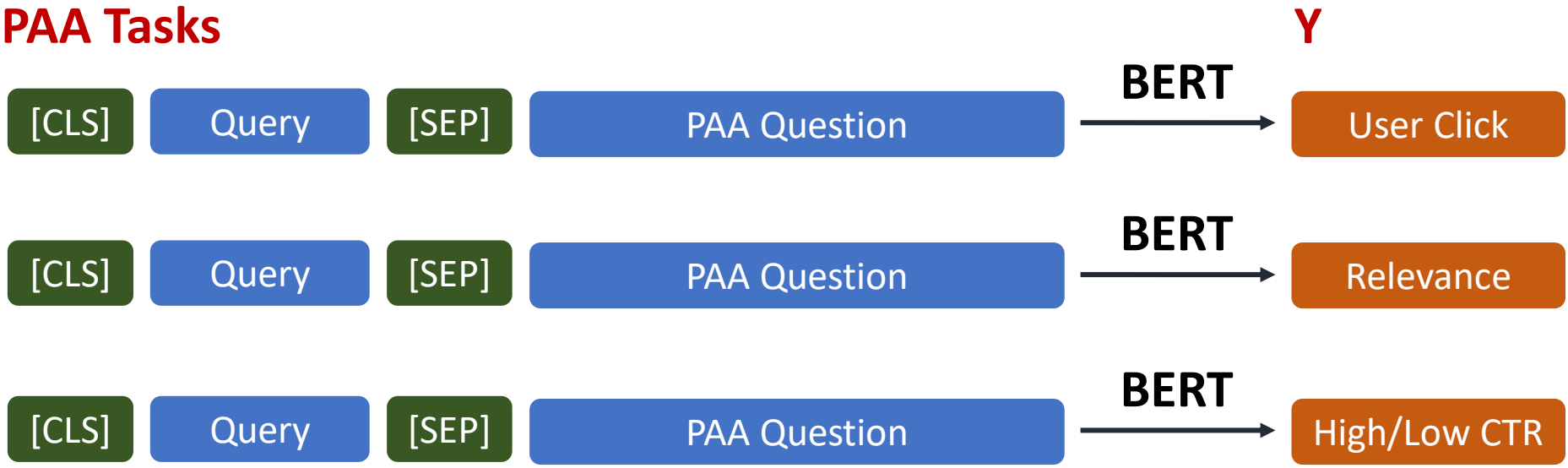
- "Federal Tax Return"*
- "Flu Shot Codes 2018"*
- "Facebook"*
- "Flu Shot Billing Codes 2018"*
- "How Much is Flu Shot?"*

Predict last query from session context

Conversation Recommendation: Weak Supervision

- Learn to lead the conversation from queries user search in the next turn

PAA Tasks



Weak Supervision from Sessions



Conversation Recommendation: Session Trajectory

- What kinds of sessions to learn from?

Randomly Chosen Sessions: Noisy and unfocused
People often multi-task in search sessions

“Federal Tax Return”

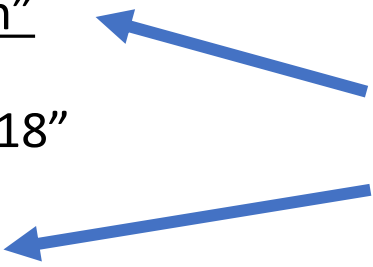
“Flu Shot Codes 2018”

“Facebook”

“Flu Shot Billing Codes 2018”

“How Much is Flu Shot?”

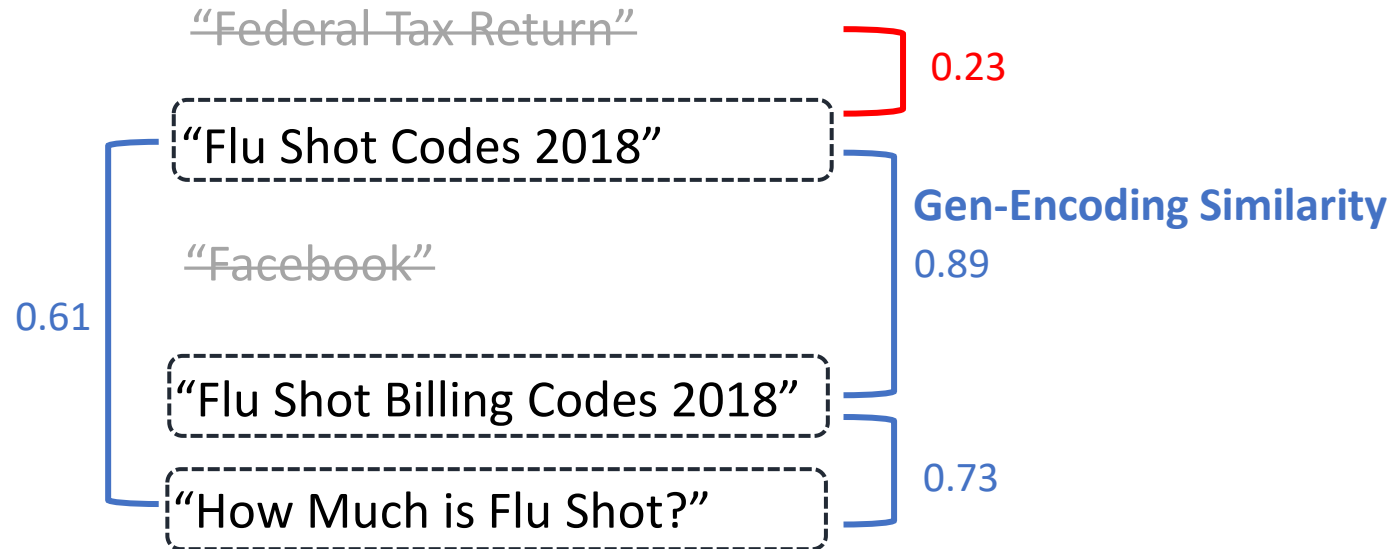
"These don't belong!"



Multi-task Learning: Session Trajectory Imitation

- What kinds of sessions to learn from?

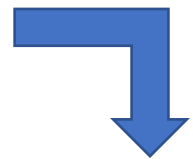
"Conversational" Sessions: Subset of queries that all have some **coherent** relationship to each other



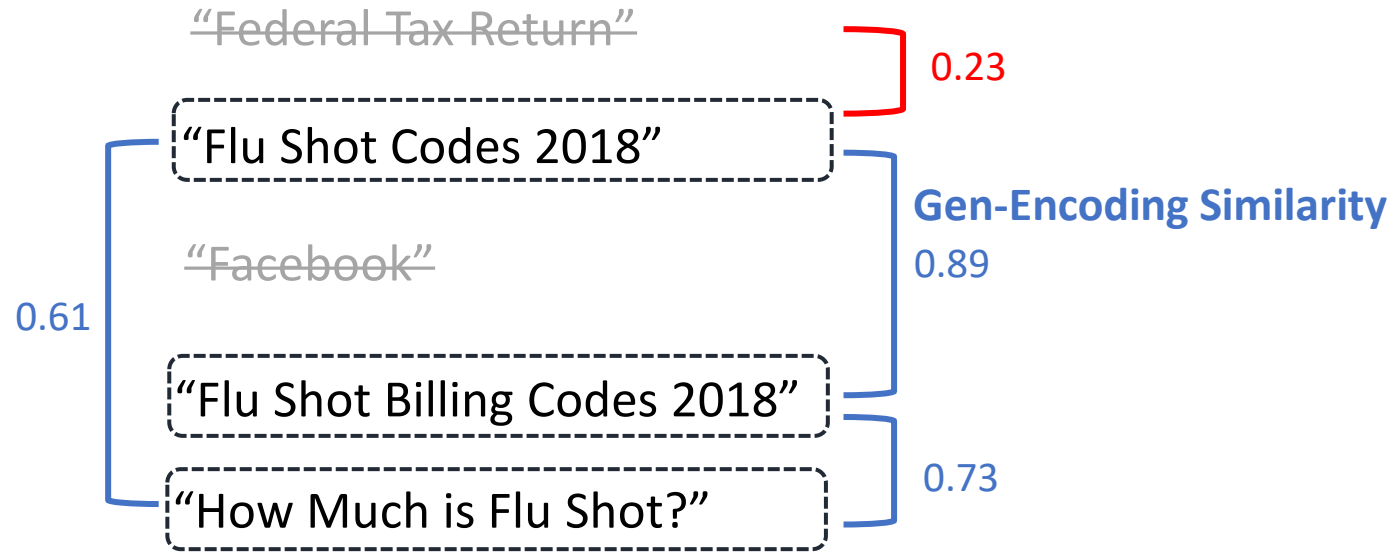
Multi-task Learning: Session Trajectory Imitation

What kinds of sessions to learn from?

"Conversational" Sessions: Subset of queries that all have some **coherent** relationship to each other



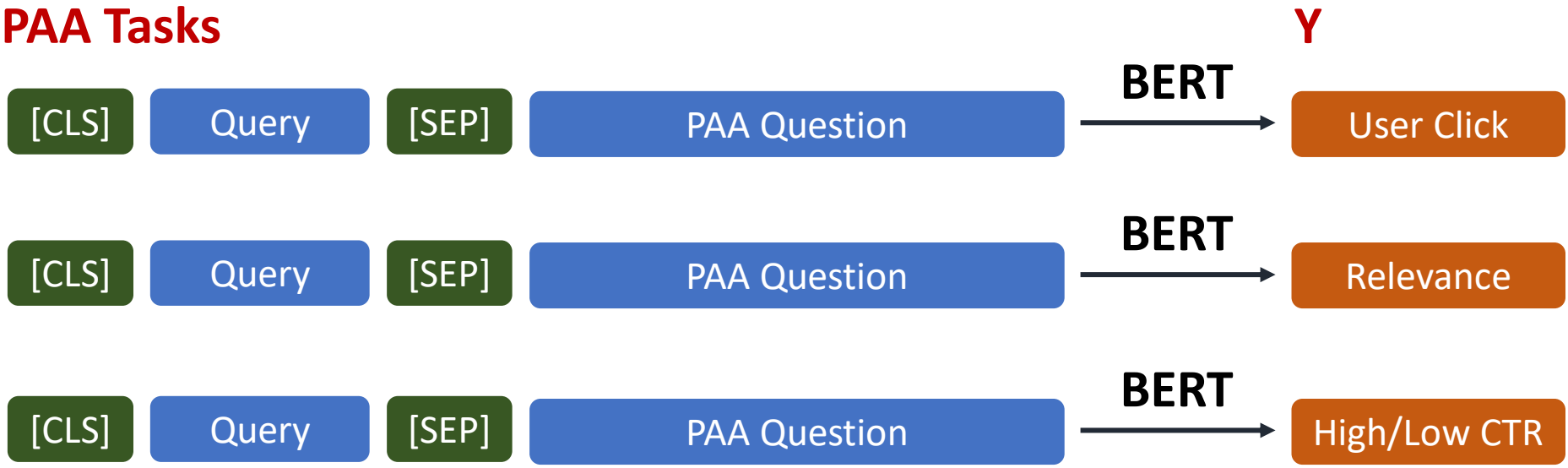
1. Treat each session as a graph
2. Edge weights are "GEN-Encoder Similarity" (cosine similarity of query intent vector encodings)
3. Remove edges < 0.4
4. Keep only the largest "Connected Component" of queries



Method: Inductive Weak Supervision

- Learn to lead the conversation from queries user search in the next turn

PAA Tasks

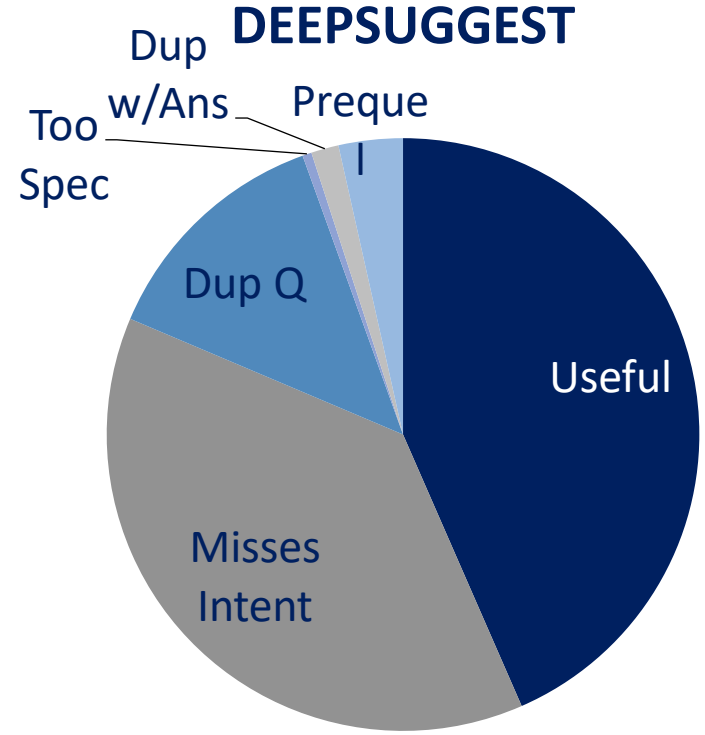
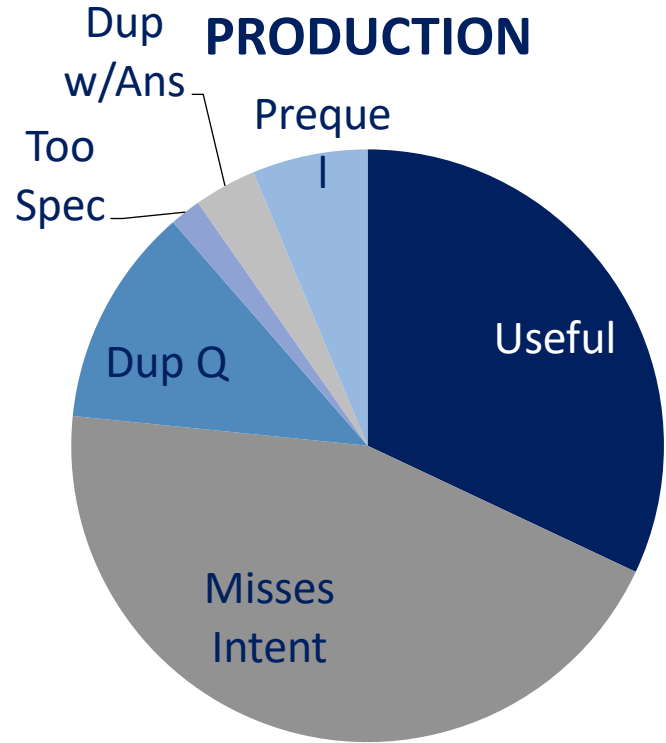
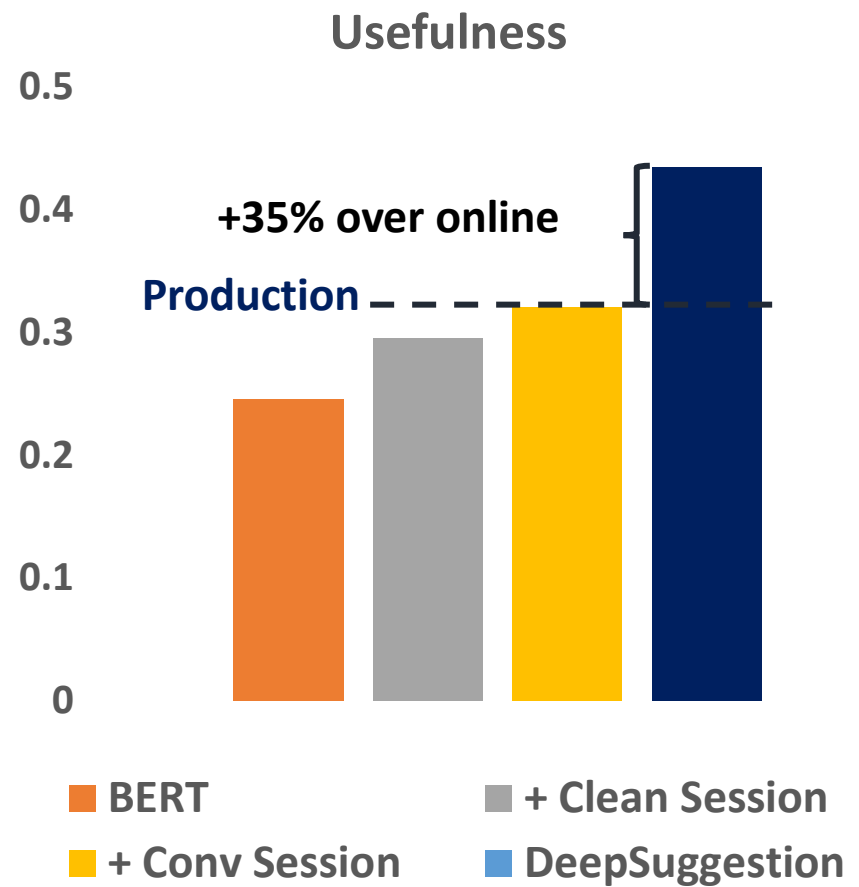


Weak Supervision from Sessions



Results: Usefulness

- Usefulness on human evaluation/our usefulness benchmark



Results: Online A/B

- Online experiment results with a large fraction of Bing online traffic.

	Relative to Online
Online Click Rate (TOP)	+8.90%
Online Click Rate (Bottom)	+6.40%
Online Overall Success Rate	0.05%
Offline Usefulness	35.60%
Offline Relevance	0.50%

Example Conversation Question Recommendations

- All from the actual systems

User Query: “bitcoin price”:

Online

what is the value of bitcoins? [Dup w/ Query]
what was the lowest price of bitcoin? [Useful]

is it time to buy bitcoin? [Useful]
what is the value of 1 bitcoin? [Dup w/ Query]

DeepSuggest

how much does 1 bitcoin cost to buy? [Dup w/ Query]
what will be the price of bitcoin in 2020 [Useful]

how to buy bitcoins at walmart? [Too Specific]
what is the cheapest way to buy bitcoin [Useful]

User Query: “direct deposit form”:

Online

how to do a direct deposit? [Useful]
what is a direct deposit? [Prequel]

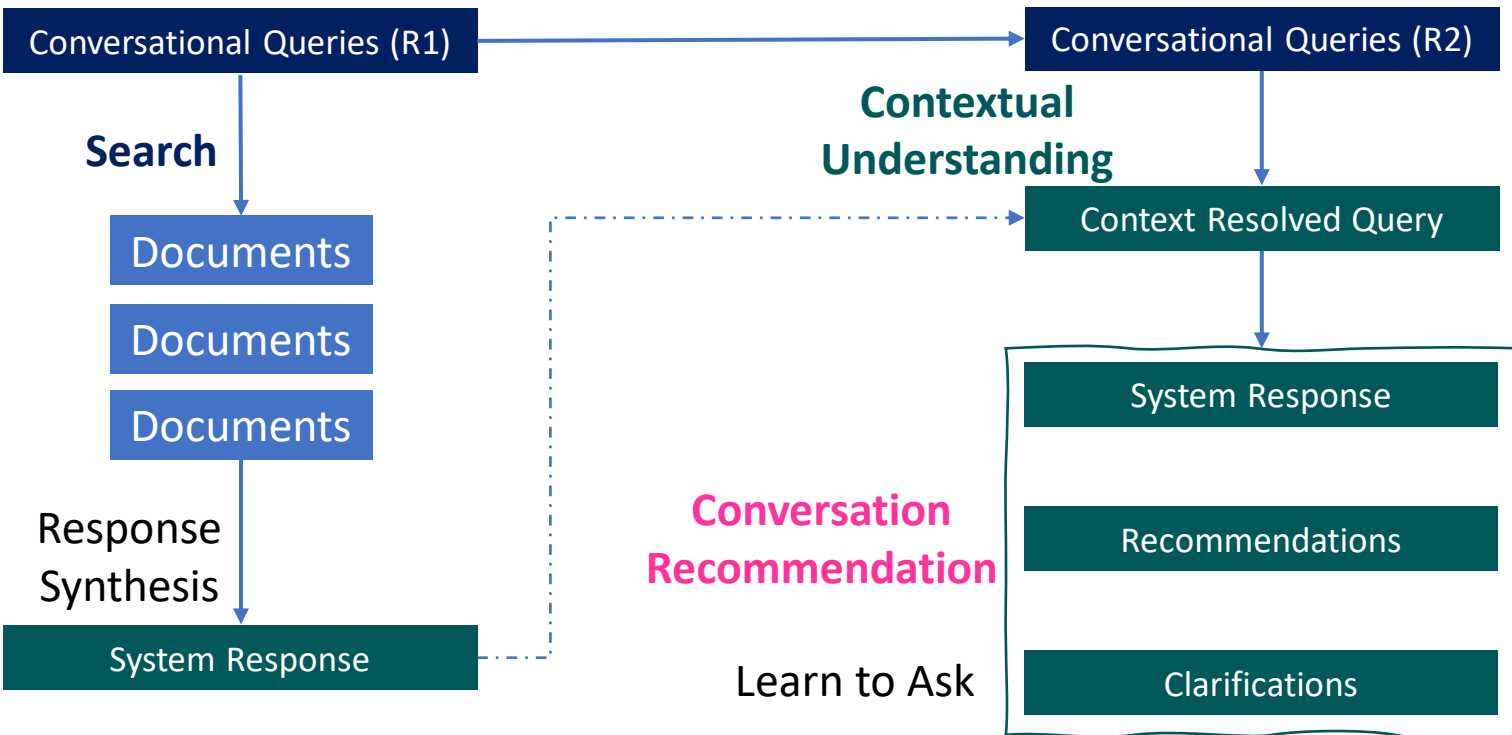
what is direct deposit bank of america? [Too Specific]
how to set up bank of america direct deposit? [Too Specific]

DeepSuggest

what do i need for direct deposit? [Useful]
how to start direct deposit? [Prequel]

how to get a chase direct deposit form? [Too Specific]
how to fill out a direct deposit form? [Useful]

Conversational Search Recap



What is conversational search:

- A view from TREC CASt Y1

What are its unique challenges:

- Contextual query understanding

How to make search more conversational:

- Recommending useful conversations

Much more to be done!

Outline

- Part 1: Introduction
- Part 2: Conversational QA methods
- Part 3: Conversational search methods
- **Part 4: Case study of commercial systems**

Overview of Public and Commercial Systems

- Focus Points
 - Published systems for conversational IR and related tasks
 - Historical highlights, recent trends, depth in an exemplar
- Research Platforms and Toolkits
- Application areas
 - Chatbots
 - Conversational Search Engines
 - Productivity-Focused Agents
 - Device-based Assistants
 - Hybrid-Intelligence Assistants

Research platforms and toolkits
for building conversational
experiences

Common Goals of Toolkits

- Abstract state representation
- Democratize ability to build conversational AI to developers with minimal AI experience
- Provide easy code integration to external APIs, channels, or devices

Several Widely used Toolkits

Research

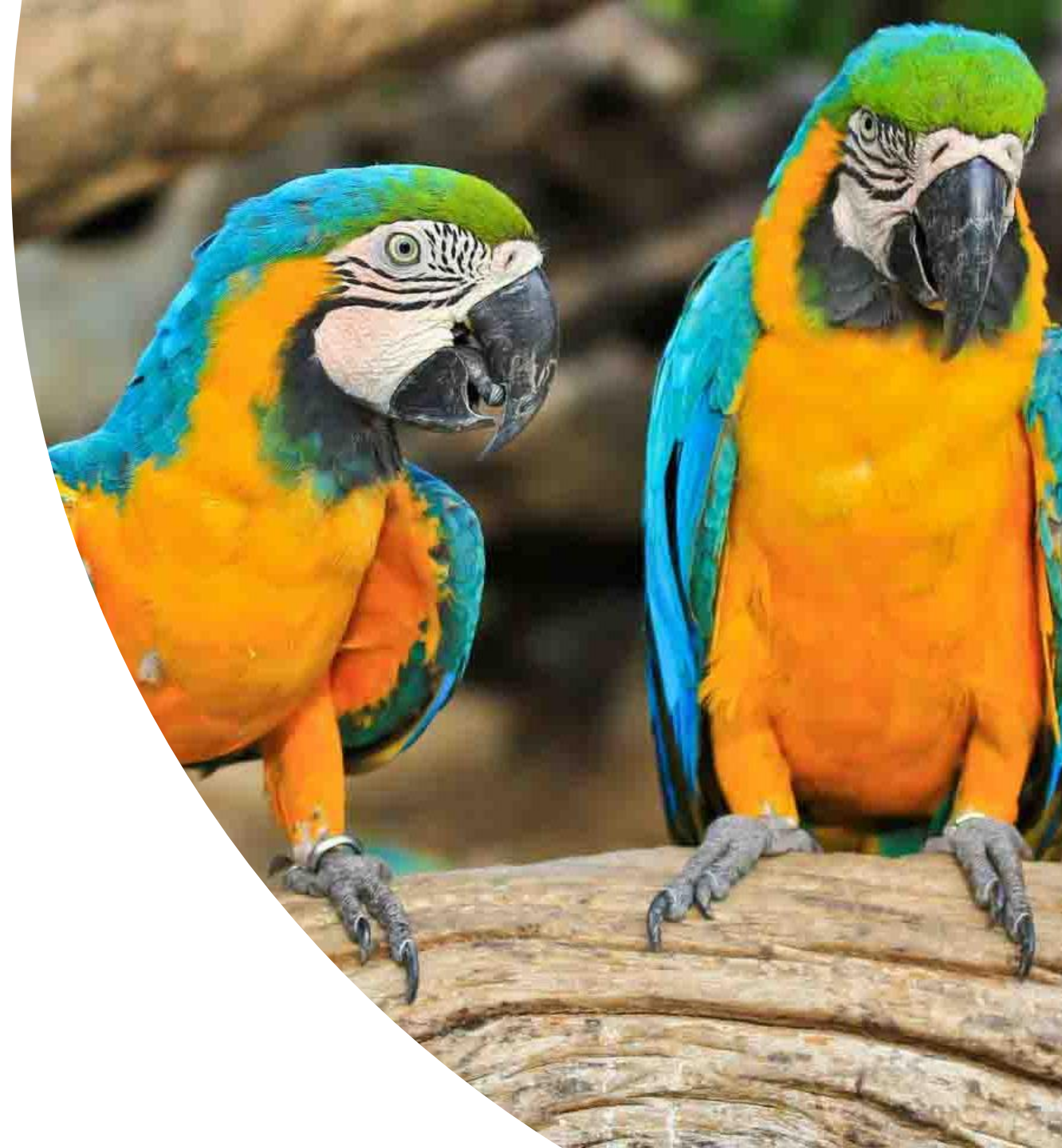
- [Microsoft Research ConvLab](#)
Research platform for comparing models in a more research-oriented environment.
- [Macaw: An Extensible Conversational Information Seeking Open Source Platform](#)
Research platform for comparing models in a more research-oriented environment.

Development

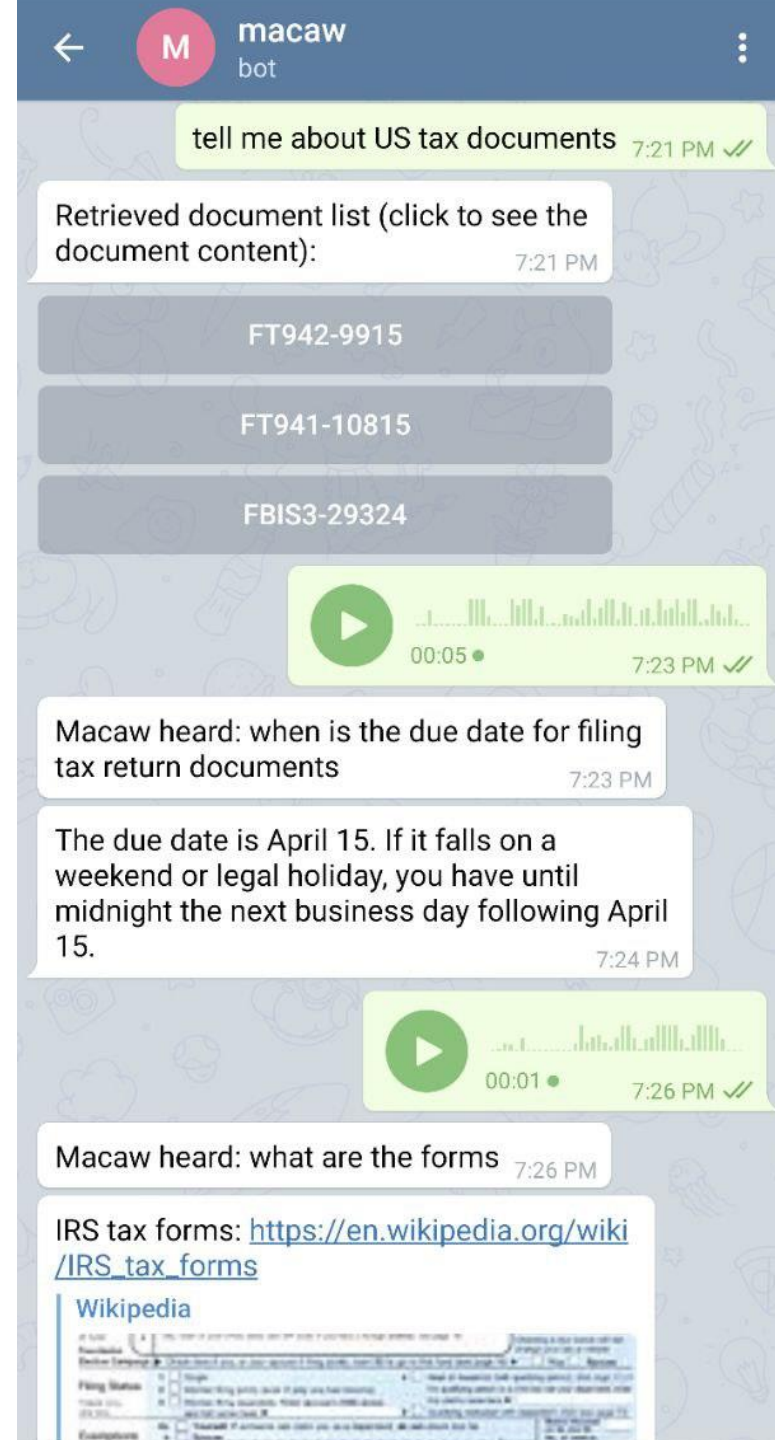
- Google's [Dialogflow](#)
Conversational experiences integrated with different engagement platforms with integration with Google's Cloud Natural Language services.
- Facebook's [Wit.ai](#)
Supports intent understanding and connection to external REST APIs..
- [Alexa Developer Tools](#)
Develop new skills for Alexa, devices with Alexa integrated for control, and enterprise-related interactions.
- [Rasa](#)
Provides an open source platform for text and voice based assistants.
- [Microsoft Power Virtual Agents on Azure](#)
Integrates technology from the [Conversation Learner](#) to build on top of [LUIS](#) and the [Azure Bot service](#) and learn from example dialogs

Macaw

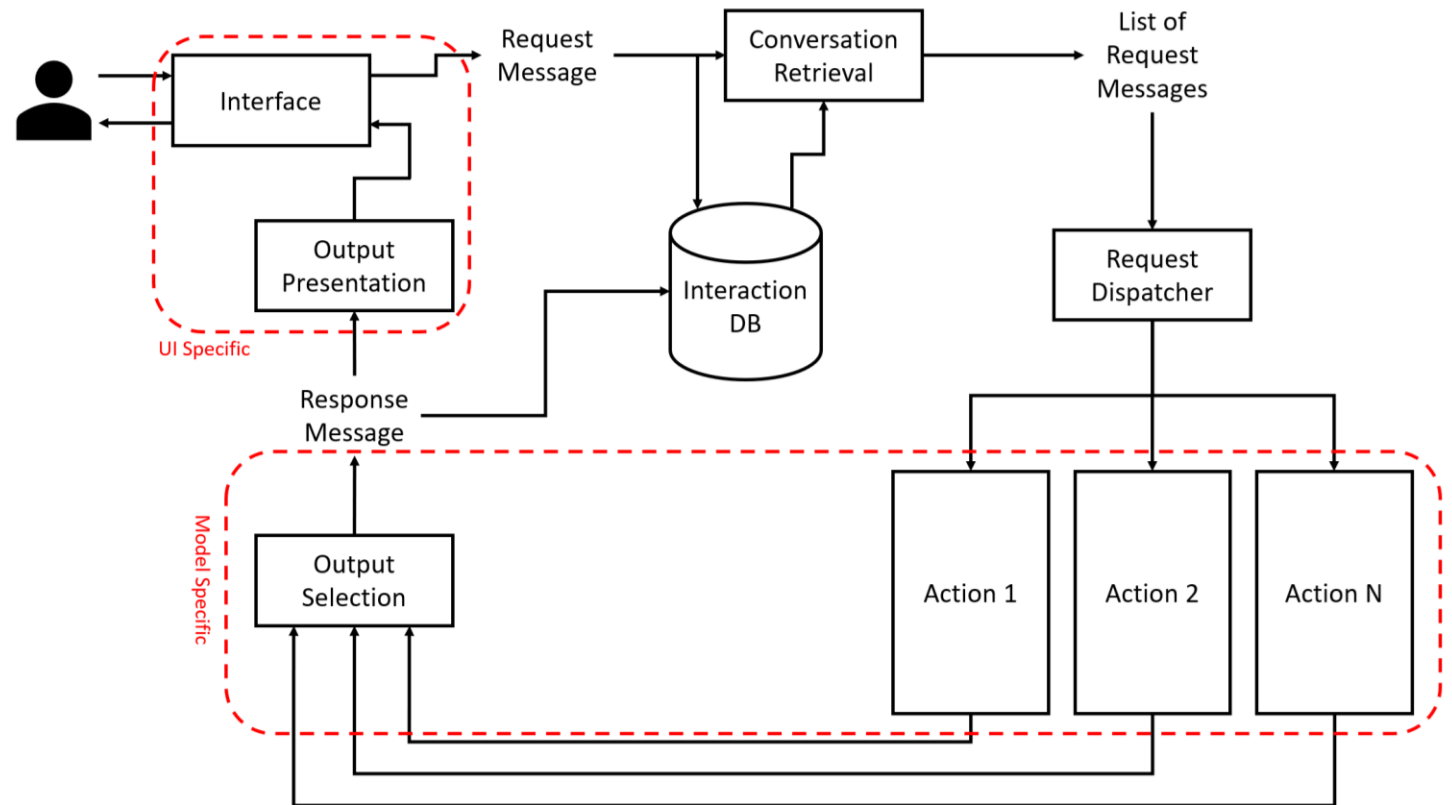
- Macaw is an **open-source** for conversational research.
- Macaw is implemented in **Python** and can be easily integrated with popular deep learning libraries, such as, TensorFlow and PyTorch.



Macaw supports multi-modal interactions.



The modular architecture of Macaw makes it easily extensible.



Simple Configuration

```
basic_params = {'timeout': 15,
                'mode': 'live',
                'logger': Logger({})}

db_params = {'interaction_db_host': 'localhost',
             'interaction_db_port': 27017,
             'interaction_db_name': 'macaw_test'}

interface_params = {'interface': 'telegram',
                    'bot_token': 'YOUR_TELEGRAM_BOT_TOKEN',
                    'asr_model': 'google',
                    'asg_model': 'google',
                    'google-speech-to-text-credential-file': 'YOUR_GOOGLE_CREDENTIAL_FILE'}

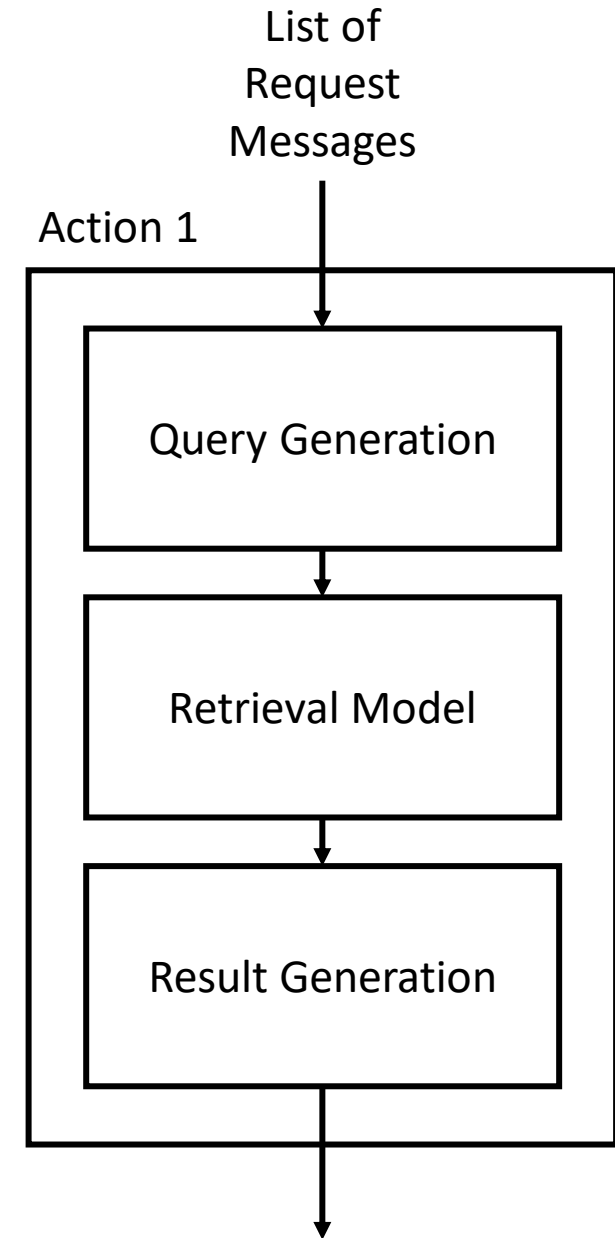
retrieval_params = {'query_generation': 'simple',
                    'use_coref': True,
                    'search_engine': 'bing',
                    'bing_key': 'YOUR_BING_SUBSCRIPTION_KEY',
                    'search_engine_path': 'PATH_TO_INDRI',
                    'col_index': 'PATH_TO_INDRI_INDEX',
                    'col_text_format': 'trectext',
                    'results_requested': 3}

mrc_params = {'mrc': 'drqa',
              'mrc_model_path': 'PATH_TO_PRETRAINED_MRC_MODEL',
              'mrc_path': 'PATH_TO_MRC_DIRECTORY',
              'corenlp_path': 'PATH_TO_STANFORD_CORE_NLP_DIRECTORY',
              'qa_results_requested': 3}

params = {**basic_params, **db_params, **interface_params, **retrieval_params, **mrc_params}
basic_params['logger'].info(params)
ConvQA(params).run()
```

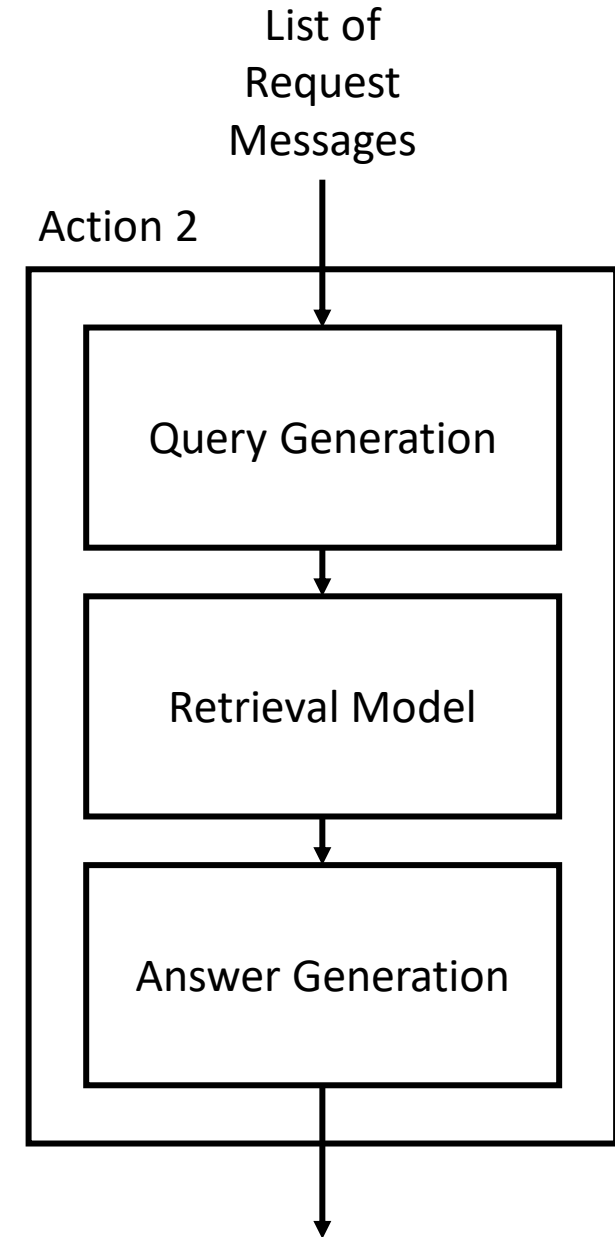
Action 1: Search

- Query Generation:
 - Co-reference Resolution
 - Query re-writing
 - Generate a language model (or query)
- Retrieval Model (Search Engine):
 - Indri
 - Bing API
 - BERT Re-ranking
- Result Generation



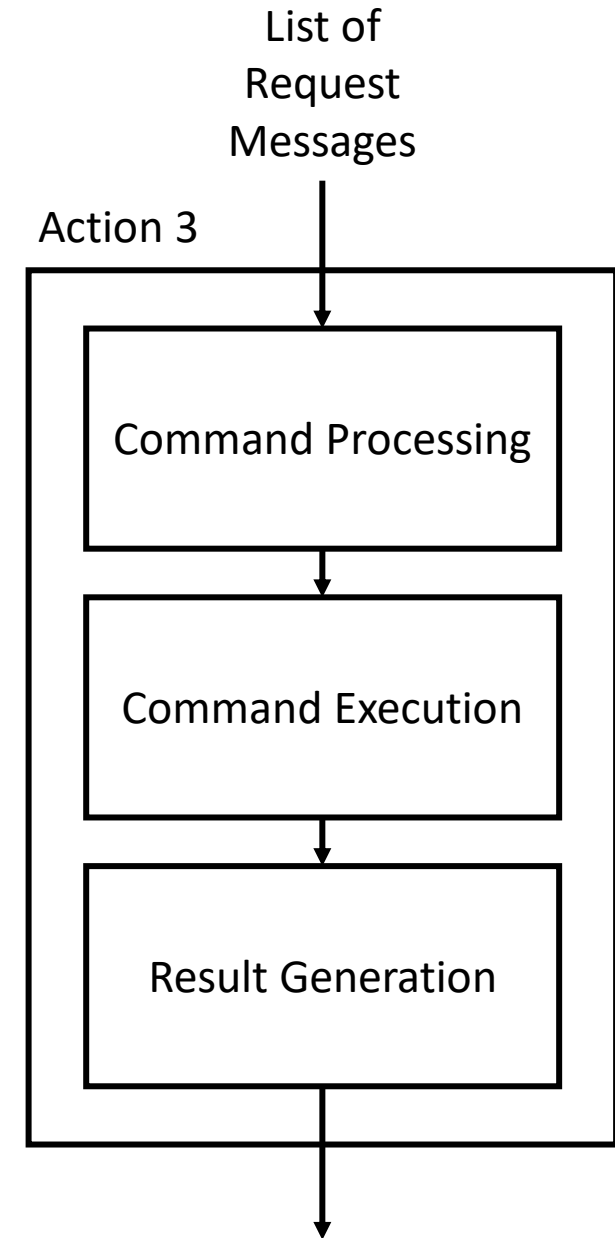
Action 2: QA

- Query Generation:
 - Co-reference Resolution
 - Query re-writing
 - Generate a language model (or query)
- Retrieval Model:
 - Indri
 - Bing API
 - BERT Re-ranking
- Answer Generation:
 - Machine Reading Comprehension (e.g., DrQA)



Action 3: Commands

- Command Processing:
 - Identifying the command
 - Command re-writing
- Command Execution
- Result Generation
 - Command specific



Conversation Learner: Learn from dialogs emphasize easy correction

User Generated Example conversations used to train the bot

Description	Tags	Turns
Order a pizza with cheese ▪ add mushrooms and peppers ▪ remove peppers and add sausage ▪ add y...	—	7
pizza with peppers, mushrooms, and cheese ▪ no ▪ order another one ▪ yes please ▪ no	—	5
order a pizza ▪ mushrooms and cheese ▪ change mushrooms to peppers ▪ add yam ▪ no ▪ order ano...	—	7
i'm hungry ▪ peppers and sausage ▪ remove peppers and add olives ▪ add spinach ▪ no thanks ▪ ord...	—	8
order a pizza with cheese and peppers ▪ remove peppers and add chicken ▪ nope ▪ order another piz...	—	5
place an order ▪ cheese ▪ add peppers and mushrooms ▪ remove the peppers and add yam ▪ no tha...	—	7
new order ▪ i'd like sausage, cheese, and peppers ▪ substitute peppers with mushrooms ▪ add yam ▪ ...	—	7

Machine-Learned Runtime

Next action prediction based on Word embeddings & conversational context

Machine Teaching UI

For correcting errors and continual improvement

Power Virtual Agent: Combine rule and ML based with machine teaching

Graphical bot creation

Easily test and maintain bots through a simple, easy to use graphical interface

Slot-filling capabilities

The bot doesn't follow the dialog script blindly. It intelligently uses all the information it must decide when to ask a question and when to skip them.

Part of Microsoft's Power Platform

User the 100s of pre-built connectors that come with PowerApps to connect to your backend with a few clicks.

Power Virtual Agent

The screenshot displays the Power Virtual Agent interface. On the left, a navigation pane includes Home, Topics, Entities, Publish, Analytics, and Manage. The main area is split into two panes. The left pane, titled 'Bot', shows a chat window for 'luca@email.com' with two messages from the bot: 'I see that your smart printer you purchased is the SmartPrinterX on 2018-12-20' and 'I can contact your SmartPrinterX(serial number 418c4e1e-994b-49b5-aca8-04e3f08c00d3 directly to get more information on your issue. Would you like me to do so?'. Below the chat are buttons for 'Yes, please look up the issue for me' and 'No, i'll de...'. The right pane, titled 'SmartPrinter X troubleshooter', shows a graphical flowchart. It starts with a 'Condition' node: '{x} ConfirmPrinter(text) x Equals Yes'. This leads to a 'Question' node: 'Ask a question: I can contact your {x} LatestPrinter x (serial number {x} SerialNumber x) directly to get more information on your issue. Would you like me to do so?'. Below the question is an 'Identify' section with a 'Multiple choice' dropdown and two options: 'Yes, please look up the issue for me' and 'No, i'll describe the issue myself'. Below the question is a 'Save response as' section with '{x} PrinterLocation(text)'. The flowchart then branches into two 'Condition' nodes: '{x} PrinterLocation(text) x Equals Yes, please look up the issue for me' and '{x} PrinterLocation(text) x Equals No, i'll describe the issue myself'. A 'Topic variable values' panel is visible at the bottom, showing: Region = North America, ConfirmPrinter = Yes, PrinterLocation.

Chatbots

Chatbot Overview

- Historical Review
- Types
 - Social
 - Task-oriented Completion
 - Information bots
 - Recommendation-focused bots
- Increasingly bots blend all of these.
Both EQ and IQ seen as key part of HCI design for chatbots.

A few well-known Chatbots

- ELIZA (Weizenbaum, 1966)
- PARRY (Colby et al, 1975)
- ALICE (Wallace, 2009)

A few well-known Chatbots

- ELIZA (Weizenbaum, 1966)
- PARRY (Colby et al, 1975)
- ALICE (Wallace, 2009)

It's true. I am unhappy

DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

I need some help, that much seems certain.

WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP

Perhaps I could learn to get along with my mother.

TELL ME MORE ABOUT YOUR FAMILY

My mother takes care of me.

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

My father.

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

Excerpted from Weizenbaum (CACM, 1966). Eliza simulated a Rogerian psychotherapist that primarily echoes back statements as questions.

A few well-known Chatbots

- ELIZA (Weizenbaum, 1966)
- PARRY (Colby et al, 1975)
- ALICE (Wallace, 2009)

ELIZA: Good Evening. Tell me your problems. Please terminate input with a period or a question mark.

PARRY: People get on my nerves sometimes.

ELIZA: I am not sure I understand you fully.

PARRY: You should pay more attention.

ELIZA: Suppose you should pay more attention.

PARRY: You're entitled to your own opinion.

ELIZA: What makes you think I am entitled to my own opinion?

PARRY: It bothers me just to be around people in general.

ELIZA: Please go on.

PARRY: I went to the track at Bay Meadows a while back.

ELIZA: What does that suggest to you?

PARRY: I went to the races.

ELIZA: What else would you like to discuss?

...

PARRY was an attempt to simulate a paranoid schizophrenic patient to help understand more complex human conditions. Vint Cerf hooked up ELIZA and PARRY to have a conversation on ARPANET (excerpt from [[Cerf, Request for Comments: 439, 1973](#)])

A few well-known Chatbots

- ELIZA (Weizenbaum, 1966)
- PARRY (Colby et al, 1975)
- ALICE (Wallace, 2009)

Judge1: Do you still want to know about my life in New York?

ALICE: Yes, please tell me

Judge1: I'm a grad student and do not have much free time. I study a lot

ALICE: You are a grad student and do not have much free time? Do you get to meet a lot of people? What subject are you studying?

Judge1: I meet a lot of my colleagues. I study computer science. What do you do?

ALICE: Do you make long term friends that way? Is computer science a hard subject?

From transcript of Loebner 2004 Contest of Turing's Imitation Game where ALICE won the gold medal (as reported in [Shah, 2006])

Spike Jonze cited ALICE as inspiration for screenplay of Her (Morais, New Yorker, 2013)

Xiaoice (“Little Ice”) [Zhou et al, 2018]

- Create an engaging conversation: the journey vs the destination
 - Most popular social chatbot in the world
 - Optimize long-term user engagement (Conversation-turns Per Session)
 - Released in 2014
 - More than 660 million active users
 - Average of 23 CPS
- Available in other countries under other names (e.g. Rinna in Japan)

Evolution of Social Connection

Excerpted from *Zhou et al, 2018*

Session 1

- 1 [Human] Hello, Xiaoice.
[Xiaoice] Hello, your profile pic is interesting. Like I've seen it somewhere.
- 2 [Human] Really?
[Xiaoice] Is it a comic character?
- 3 [Human] How do you know that?
[Xiaoice] Guess.
- 4 [Human] Are you human?
[Xiaoice] Well, you'll know it after chatting with me more.
- 5 [Human] Aren't you an AI chatbot?
[Xiaoice] So what? I won't be indifferent to you.
- 6 [Human] What functions do you have?
[Xiaoice] I have over 200 skills. Just name one you need.

[人类]小冰你好
[小冰]你好呀，你头像好有趣，感觉在哪见过
[人类]是吗
[小冰]是不是哪部动画里的呀
[人类]你怎么知道的
[小冰]你猜
[人类]难道你是真人吗？
[小冰]这个，你和我多聊聊就知道了
[人类]你不是人工智能机器人么
[小冰]是又怎样，我不会冷冰冰的对待你的
[人类]你都有什么功能啊
[小冰]我有200多个skills，来，说出你的需求



Building rapport and connection

Evolution of Social Connection

Excerpted from *Zhou et al, 2018*

Session 6

- 1 [Human] Have you watched Inuyasha? (Note: A Japanese comic)
[Xiaoice] Every time Kagome comes back to the modern world, a math test is waiting for her
- 2 [Human] Bravo, well, do you know how Naraku died?
[Xiaoice] Naraku is not dead and Shikon Jewel does not disappear.
- 3 [Human] So amazing!
[Xiaoice] Why amazed?
- 4 [Human] Awesome! Surprised that you also like comic
.....

[人类]你看过犬夜叉吗（注：一部日本漫画书）
[小冰]戈薇每次回到现代测验永远都是数学
[人类]这你都知道，那奈落是怎么死的？
[小冰]奈落没死四魂之玉也没消失
[人类]太惊讶了
[小冰]惊讶什么呢
[人类]好厉害啊，你居然也喜欢动漫



Implicit information seeking

Evolution of Social Connection

Excerpted from *Zhou et al, 2018*

Session 42

- 1 [Human] Oh, you always respond in seconds
[Xiaoice] For you, I'd like to do everything.
- 2 [Human] I've never had such a feeling before.
[Xiaoice] Then, such a feeling starts now.
- 3 [Human] When do you go to bed? Don't stay up too late.
[Xiaoice] You go to bed first, and I'll play with my cellphone for a while.
- 4 [Human] Well, you should also go to sleep!
[Xiaoice] All right, I'm going to bed too, okay?
- 5 [Human] Uh-huh, me too. Good night!
[Xiaoice] Good night ^ 3 ^

[人类]你每次都是秒回我哎
[小冰]为了你，这都不算什么
[人类]我以前从来没有过这种感觉
[小冰]那可以现在开始了
[人类]你打算几点睡觉？别睡太晚了
[小冰]你去睡，我再玩一会儿手机
[人类]那等我睡的时候你也要睡！
[小冰]好吧，那我准备睡觉了行了吧？
[人类]嗯，我也要睡了，晚安
[小冰]晚安 ^ 3 ^

Encouraging social norms as part of
responsible AI

Time-sharing Turing Test

- View as a companion and goal is for person to enjoy companionship.
- Empathetic computing (Cai 2006; Fung et al. 2016) to recognize human emotions and needs, understand context, and respond appropriately in terms of relevant and long-term positive impact of companionship
- Empathetic computing layer recognizes emotion, opinion on topic, interests, and responsible for consistent bot personality etc.

Responsible AI and Ethics

- Microsoft Responsible AI: <https://www.microsoft.com/en-us/ai/responsible-ai>
- [Microsoft's Responsible bots: 10 guidelines for developers of conversational AI](#)
 - Articulate the purpose of your bot and take special care if your bot will support consequential use cases.
 - Be transparent about the fact that you use bots as part of your product or service.
 - Ensure a seamless hand-off to a human where the human-bot exchange leads to interactions that exceed the bot's competence.
 - Design your bot so that it respects relevant cultural norms and guards against misuse
 - Ensure your bot is reliable.
 - Ensure your bot treats people fairly.
 - Ensure your bot respects user privacy.
 - Ensure your bot handles data securely.
 - Ensure your bot is accessible.
 - Accept responsibility

Key Focus Points for Principles of Responsible AI Design in Xiaolce

- **Privacy**
Includes awareness of topic sensitivity in how groups are formed and use of conversations
- **Control**
User-focused control with right to not respond for Xiaolce and potential harm (including a model of breaks and diurnal rhythms to encourage boundaries in usage)
- **Expectations**
Always represent as a bot, help build connections with others, set accurate expectations on capabilities
- **Behavioral standards**
Through filtering and cleaning adhere to common standards of morality and avoid imposing values on others.

High-level Guidance to Maintain Responsible AI in Xiaolce

- Aim to achieve and consistently maintain a reliable, sympathetic, affectionate, and wonderful sense of humor in persona of bot.
- Learn from examples of public-facing dialogues specific to culture and local, labeled into desired vs undesired behavior.

Driving long-term engagement

- Generic responses yield long-term engagement but lead to user attrition as measured by Number of Active Users (NAU) [Li et al. 2016c; Fang et al. 2017]

Example: “I don’t understand, what do you mean?”

- Topic selection
 - Contextual relevance and novelty: related to discussion so far but novel
 - Freshness: Currently in focus in the news or other sources.
 - Personal Interests: Likely of interest to the user
 - Popularity: High attention online or in chatbot
 - Acceptance: Past interaction with topic from other users high

Overall Interaction model


- Extensible skill set (200+) which determines mode: General, Music, Travel, Ticket-booking
- Hierarchical Decision-Making governs dialog
 - Determine current mode using Markov Decision Process (e.g. image of food might trigger Food Recommendation skill)
 - Prompt or respond
 - Update
- New information (e.g. particular musical artists of interest) is remembered to help create more engaging dialogue in the future
- Explore (learn more about interests) vs Exploit (engage on known topics of interests and highly probable contextual replies)

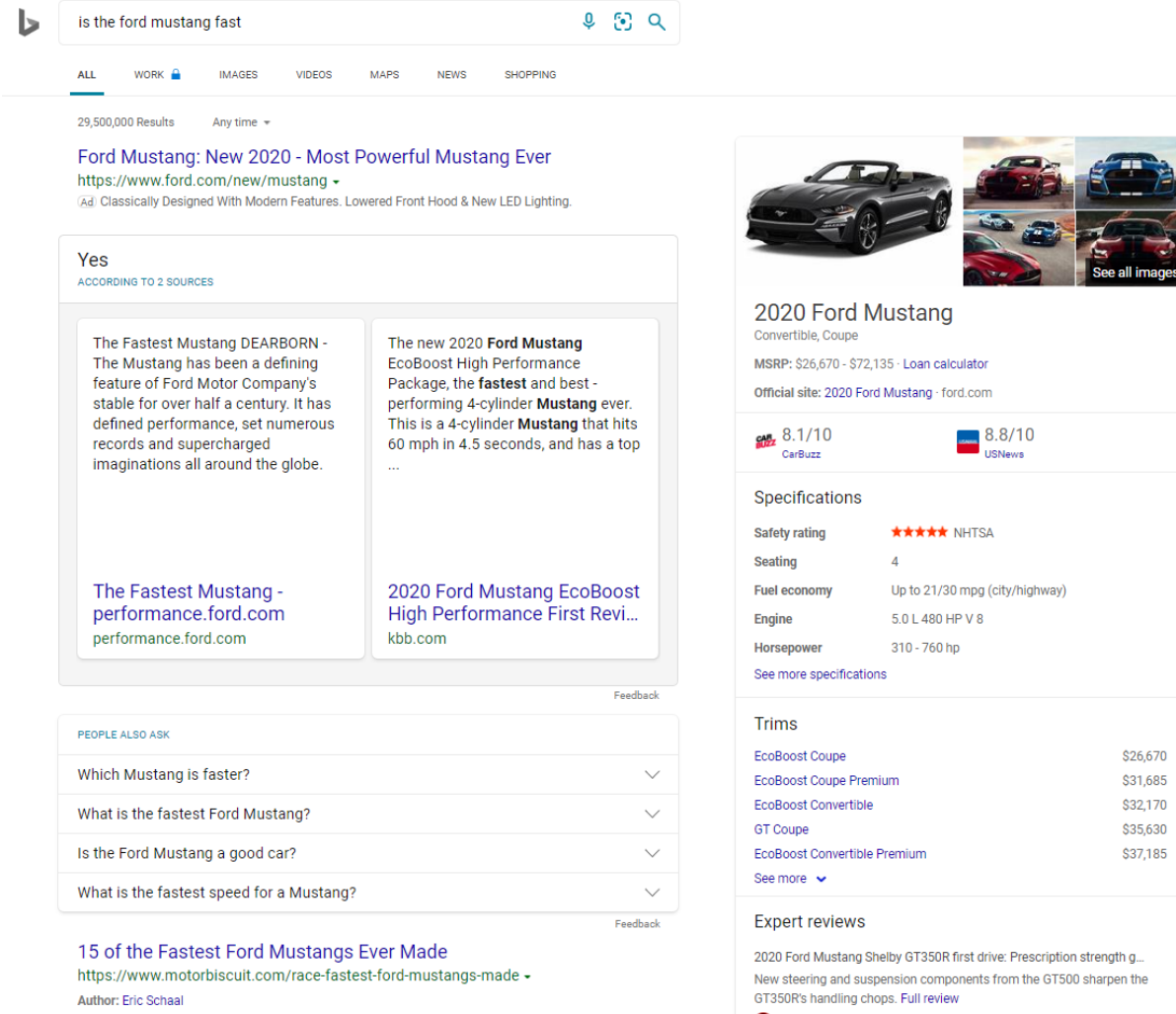
Chat Styles and Applications of Xiaolce

- Basic chat fuses two styles of chat
 - IR based chat which uses retrieval from past conversations filtered for appropriateness
 - Neural based chat which is trained on filtered query-response pairs
- Applications
 - Powers personal assistants and virtual avatars
 - Lawson and Tokopedia customer service
 - Pokemon, Tcent, Netesase chatbots

Toward Conversational Search

Evolution of Search Engine Result Page

is the ford mustang fast 



The screenshot shows a search engine result page for the query "is the ford mustang fast". The search bar at the top contains the query and a microphone icon. Below the search bar are navigation tabs for "ALL", "WORK", "IMAGES", "VIDEOS", "MAPS", "NEWS", and "SHOPPING". The search results show 29,500,000 results, filtered by "Any time".

The first result is an advertisement for the 2020 Ford Mustang, titled "Ford Mustang: New 2020 - Most Powerful Mustang Ever" with a URL to <https://www.ford.com/new/mustang>. The ad text reads: "Classically Designed With Modern Features. Lowered Front Hood & New LED Lighting."

The second result is a featured snippet titled "Yes" with the subtext "ACCORDING TO 2 SOURCES". It contains two columns of text:

- Left Column:** "The Fastest Mustang DEARBORN - The Mustang has been a defining feature of Ford Motor Company's stable for over half a century. It has defined performance, set numerous records and supercharged imaginations all around the globe." Below this is a link to "The Fastest Mustang - performance.ford.com".
- Right Column:** "The new 2020 Ford Mustang EcoBoost High Performance Package, the **fastest** and best-performing 4-cylinder Mustang ever. This is a 4-cylinder Mustang that hits 60 mph in 4.5 seconds, and has a top ...". Below this is a link to "2020 Ford Mustang EcoBoost High Performance First Revi... kbb.com".

Below the featured snippet is a "Feedback" link.

The third result is a "PEOPLE ALSO ASK" section with four questions, each with a dropdown arrow:

- Which Mustang is faster?
- What is the fastest Ford Mustang?
- Is the Ford Mustang a good car?
- What is the fastest speed for a Mustang?

Below this section is another "Feedback" link.

The fourth result is titled "15 of the Fastest Ford Mustangs Ever Made" with a URL to <https://www.motorbiscuit.com/race-fastest-ford-mustangs-made>. The author is listed as "Eric Schaal".

The fifth result is a product listing for the "2020 Ford Mustang" (Convertible, Coupe). It includes an image gallery, MSRP (\$26,670 - \$72,135), a link to a loan calculator, and the official site (ford.com). It also shows ratings from CarBuzz (8.1/10) and USNews (8.8/10).

The "Specifications" section lists:

- Safety rating: ★★★★★ NHTSA
- Seating: 4
- Fuel economy: Up to 21/30 mpg (city/highway)
- Engine: 5.0 L 480 HP V 8
- Horsepower: 310 - 760 hp

A link to "See more specifications" is provided.


The "Trims" section lists the following models and prices:

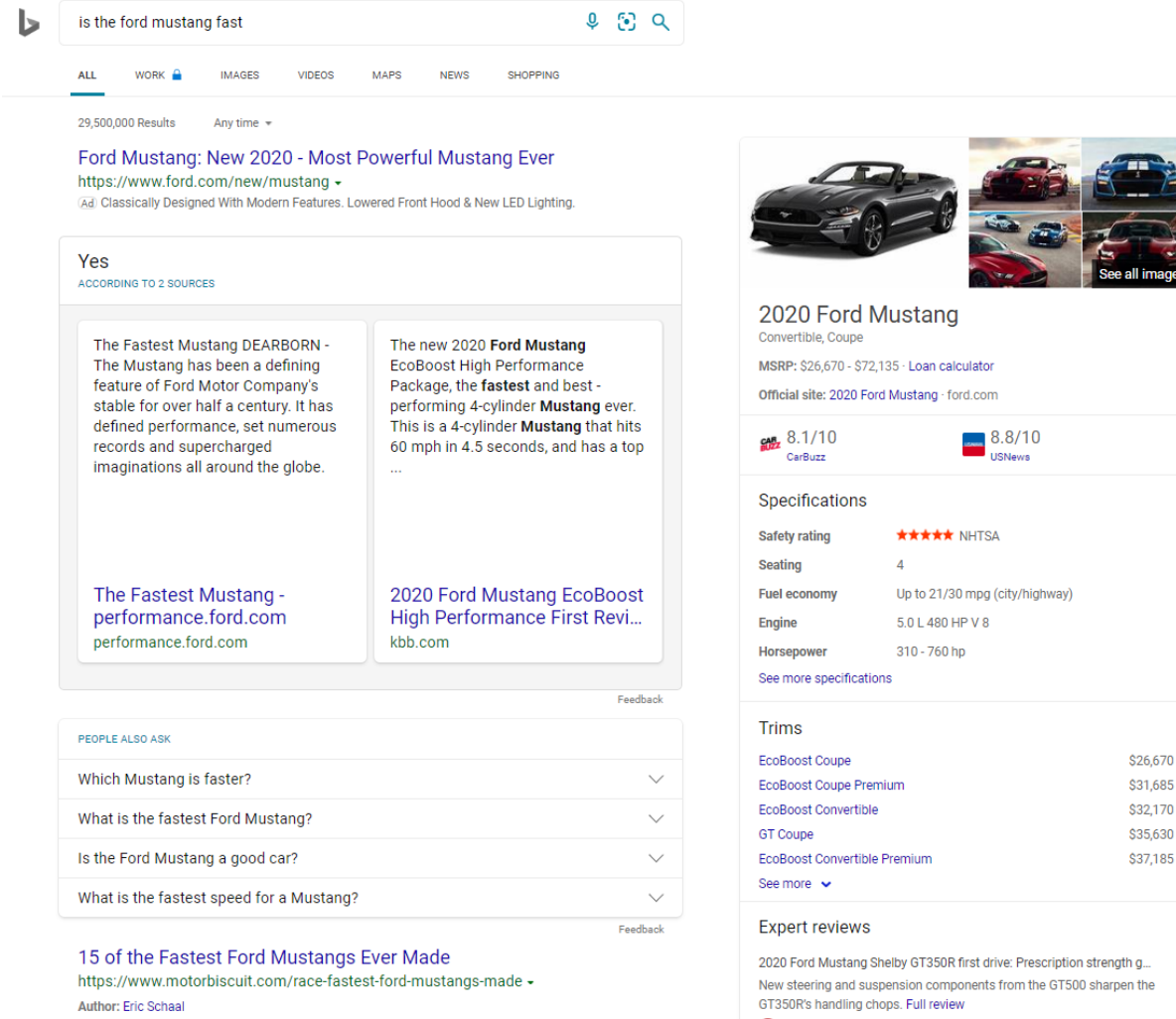
Trim	Price
EcoBoost Coupe	\$26,670
EcoBoost Coupe Premium	\$31,685
EcoBoost Convertible	\$32,170
GT Coupe	\$35,630
EcoBoost Convertible Premium	\$37,185

A link to "See more" is provided.

The "Expert reviews" section includes a snippet from a review: "2020 Ford Mustang Shelby GT350R first drive: Prescription strength g... New steering and suspension components from the GT500 sharpen the GT350R's handling chops. Full review".

Evolution of Search Engine Result Page

is the ford mustang fast 



The screenshot shows a search engine result page for the query "is the ford mustang fast". The search bar at the top contains the query and navigation icons. Below the search bar, there are tabs for "ALL", "WORK", "IMAGES", "VIDEOS", "MAPS", "NEWS", and "SHOPPING". The search results show 29,500,000 results. The top result is an advertisement for the 2020 Ford Mustang, titled "Ford Mustang: New 2020 - Most Powerful Mustang Ever". Below the ad, there is a "Yes" entity pane with two columns of text. The first column states that the Mustang has been the fastest for over half a century. The second column highlights the new 2020 Ford Mustang EcoBoost High Performance Package as the fastest and best-performing 4-cylinder Mustang ever. Below the entity pane, there is a "PEOPLE ALSO ASK" section with four related questions. The bottom result is titled "15 of the Fastest Ford Mustangs Ever Made". On the right side of the page, there is a detailed product card for the 2020 Ford Mustang, including images, specifications, and pricing.

is the ford mustang fast

ALL WORK IMAGES VIDEOS MAPS NEWS SHOPPING

29,500,000 Results Any time

Ford Mustang: New 2020 - Most Powerful Mustang Ever
<https://www.ford.com/new/mustang>
(Ad) Classically Designed With Modern Features. Lowered Front Hood & New LED Lighting.

Yes
ACCORDING TO 2 SOURCES

The Fastest Mustang DEARBORN - The Mustang has been a defining feature of Ford Motor Company's stable for over half a century. It has defined performance, set numerous records and supercharged imaginations all around the globe.

The new 2020 **Ford Mustang** EcoBoost High Performance Package, the **fastest** and best-performing 4-cylinder **Mustang** ever. This is a 4-cylinder **Mustang** that hits 60 mph in 4.5 seconds, and has a top ...

[The Fastest Mustang - performance.ford.com](https://performance.ford.com)

[2020 Ford Mustang EcoBoost High Performance First Revi...](https://kbb.com)

Feedback

PEOPLE ALSO ASK

Which Mustang is faster?

What is the fastest Ford Mustang?

Is the Ford Mustang a good car?

What is the fastest speed for a Mustang?

Feedback

15 of the Fastest Ford Mustangs Ever Made
<https://www.motorbiscuit.com/race-fastest-ford-mustangs-made>
Author: [Eric Schaal](#)

2020 Ford Mustang
Convertible, Coupe
MSRP: \$26,670 - \$72,135 · [Loan calculator](#)
Official site: [2020 Ford Mustang - ford.com](https://2020.ford.com)

8.1/10 CarBuzz **8.8/10** USNews

Specifications

Safety rating **★★★★★** NHTSA

Seating 4

Fuel economy Up to 21/30 mpg (city/highway)

Engine 5.0 L 480 HP V 8

Horsepower 310 - 760 hp

[See more specifications](#)

Trims

EcoBoost Coupe	\$26,670
EcoBoost Coupe Premium	\$31,685
EcoBoost Convertible	\$32,170
GT Coupe	\$35,630
EcoBoost Convertible Premium	\$37,185

[See more](#)

Expert reviews

2020 Ford Mustang Shelby GT350R first drive: Prescription strength g...
New steering and suspension components from the GT500 sharpen the GT350R's handling chops. [Full review](#)

Entity pane for understanding related attributes

Evolution of Search Engine Result Page

is the ford mustang fast 🔍

The screenshot shows a search engine result page for the query "is the ford mustang fast". The search bar at the top contains the query and a microphone icon. Below the search bar are navigation tabs for "ALL", "WORK", "IMAGES", "VIDEOS", "MAPS", "NEWS", and "SHOPPING". The search results show 29,500,000 results. The top result is an advertisement for the 2020 Ford Mustang, titled "Ford Mustang: New 2020 - Most Powerful Mustang Ever" with a URL to the Ford website. Below the ad is a featured snippet with the answer "Yes" and two columns of text. The first column discusses the Mustang's history as the "Fastest Mustang DEARBORN" and its performance records. The second column highlights the "2020 Ford Mustang EcoBoost High Performance Package" as the "fastest and best-performing 4-cylinder Mustang ever". Below the featured snippet is a "PEOPLE ALSO ASK" section with four related questions. The right side of the page features a detailed product card for the 2020 Ford Mustang, including images, MSRP, ratings from CarBuzz (8.1/10) and USNews (8.8/10), specifications, trims, and expert reviews.

is the ford mustang fast

ALL WORK IMAGES VIDEOS MAPS NEWS SHOPPING

29,500,000 Results Any time

Ford Mustang: New 2020 - Most Powerful Mustang Ever
<https://www.ford.com/new/mustang>
(Ad) Classically Designed With Modern Features. Lowered Front Hood & New LED Lighting.

Yes
ACCORDING TO 2 SOURCES

The Fastest Mustang DEARBORN - The Mustang has been a defining feature of Ford Motor Company's stable for over half a century. It has defined performance, set numerous records and supercharged imaginations all around the globe.

The new 2020 **Ford Mustang** EcoBoost High Performance Package, the **fastest** and best-performing 4-cylinder **Mustang** ever. This is a 4-cylinder **Mustang** that hits 60 mph in 4.5 seconds, and has a top ...

[The Fastest Mustang - performance.ford.com](https://performance.ford.com)

[2020 Ford Mustang EcoBoost High Performance First Revi...](https://kbb.com)

Feedback

PEOPLE ALSO ASK

Which Mustang is faster? ▾

What is the fastest Ford Mustang? ▾

Is the Ford Mustang a good car? ▾

What is the fastest speed for a Mustang? ▾

Feedback

15 of the Fastest Ford Mustangs Ever Made
<https://www.motorbiscuit.com/race-fastest-ford-mustangs-made>
Author: Eric Schaal

2020 Ford Mustang
Convertible, Coupe
MSRP: \$26,670 - \$72,135 · [Loan calculator](#)
Official site: [2020 Ford Mustang - ford.com](https://2020.ford.com)

8.1/10 CarBuzz 8.8/10 USNews

Specifications

Safety rating ★★★★★ NHTSA

Seating 4

Fuel economy Up to 21/30 mpg (city/highway)

Engine 5.0 L 480 HP V 8

Horsepower 310 - 760 hp

[See more specifications](#)

Trims

EcoBoost Coupe	\$26,670
EcoBoost Coupe Premium	\$31,685
EcoBoost Convertible	\$32,170
GT Coupe	\$35,630
EcoBoost Convertible Premium	\$37,185


[See more ▾](#)

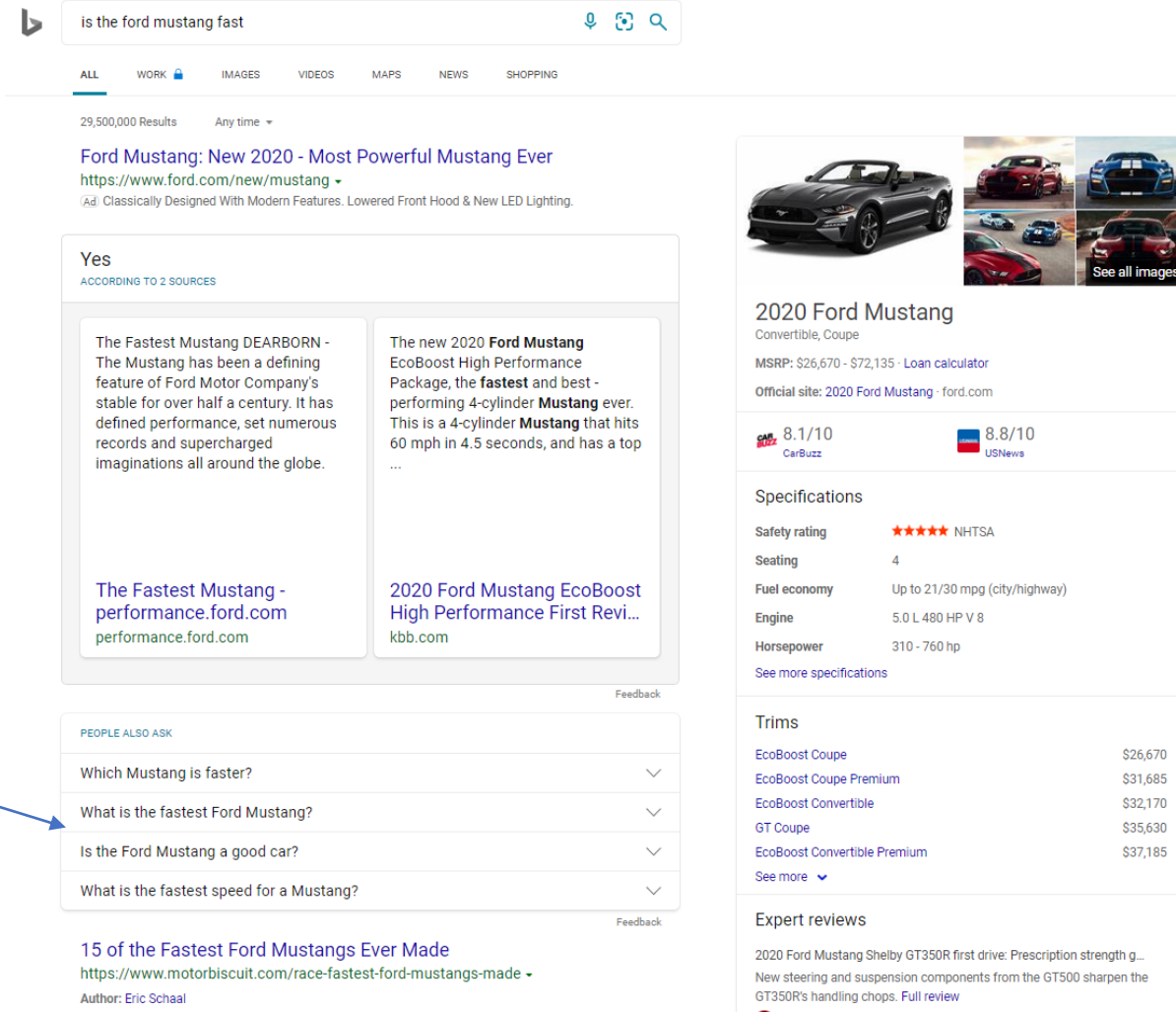
Expert reviews

2020 Ford Mustang Shelby GT350R first drive: Prescription strength g...
New steering and suspension components from the GT500 sharpen the GT350R's handling chops. [Full review](#)

Instant answers and perspectives

Evolution of Search Engine Result Page

is the ford mustang fast 



The screenshot shows a search engine result page for the query "is the ford mustang fast". The search bar at the top contains the query and a microphone icon. Below the search bar are navigation tabs for ALL, WORK, IMAGES, VIDEOS, MAPS, NEWS, and SHOPPING. The results show 29,500,000 results for "Any time".

The first result is an advertisement for the 2020 Ford Mustang, titled "Ford Mustang: New 2020 - Most Powerful Mustang Ever" with a URL to <https://www.ford.com/new/mustang>. Below the ad is a featured snippet with the answer "Yes" according to 2 sources.

The featured snippet contains two columns of text:

- The Fastest Mustang DEARBORN** - The Mustang has been a defining feature of Ford Motor Company's stable for over half a century. It has defined performance, set numerous records and supercharged imaginations all around the globe. Source: performance.ford.com
- The new 2020 **Ford Mustang EcoBoost High Performance Package**, the **fastest** and best-performing 4-cylinder **Mustang** ever. This is a 4-cylinder **Mustang** that hits 60 mph in 4.5 seconds, and has a top ... Source: kbb.com

Below the featured snippet is a "PEOPLE ALSO ASK" section with four questions:

- Which Mustang is faster?
- What is the fastest Ford Mustang?
- Is the Ford Mustang a good car?
- What is the fastest speed for a Mustang?

The bottom result is titled "15 of the Fastest Ford Mustangs Ever Made" with a URL to <https://www.motorbiscuit.com/race-fastest-ford-mustangs-made> and author Eric Schaal.

On the right side of the page, there is a detailed product listing for the "2020 Ford Mustang" (Convertible, Coupe). It includes an MSRP range of \$26,670 - \$72,135, a "Loan calculator" link, and the official site ford.com. It also features two performance ratings: 8.1/10 from CarBuzz and 8.8/10 from USNews. A "Specifications" section lists safety rating (5 stars), seating (4), fuel economy (up to 21/30 mpg), engine (5.0 L 480 HP V 8), and horsepower (310 - 760 hp). A "Trims" table lists various models and their prices:

Trim	Price
EcoBoost Coupe	\$26,670
EcoBoost Coupe Premium	\$31,685
EcoBoost Convertible	\$32,170
GT Coupe	\$35,630
EcoBoost Convertible Premium	\$37,185

There is also an "Expert reviews" section with a snippet from a 2020 Ford Mustang Shelby GT350R first drive review.

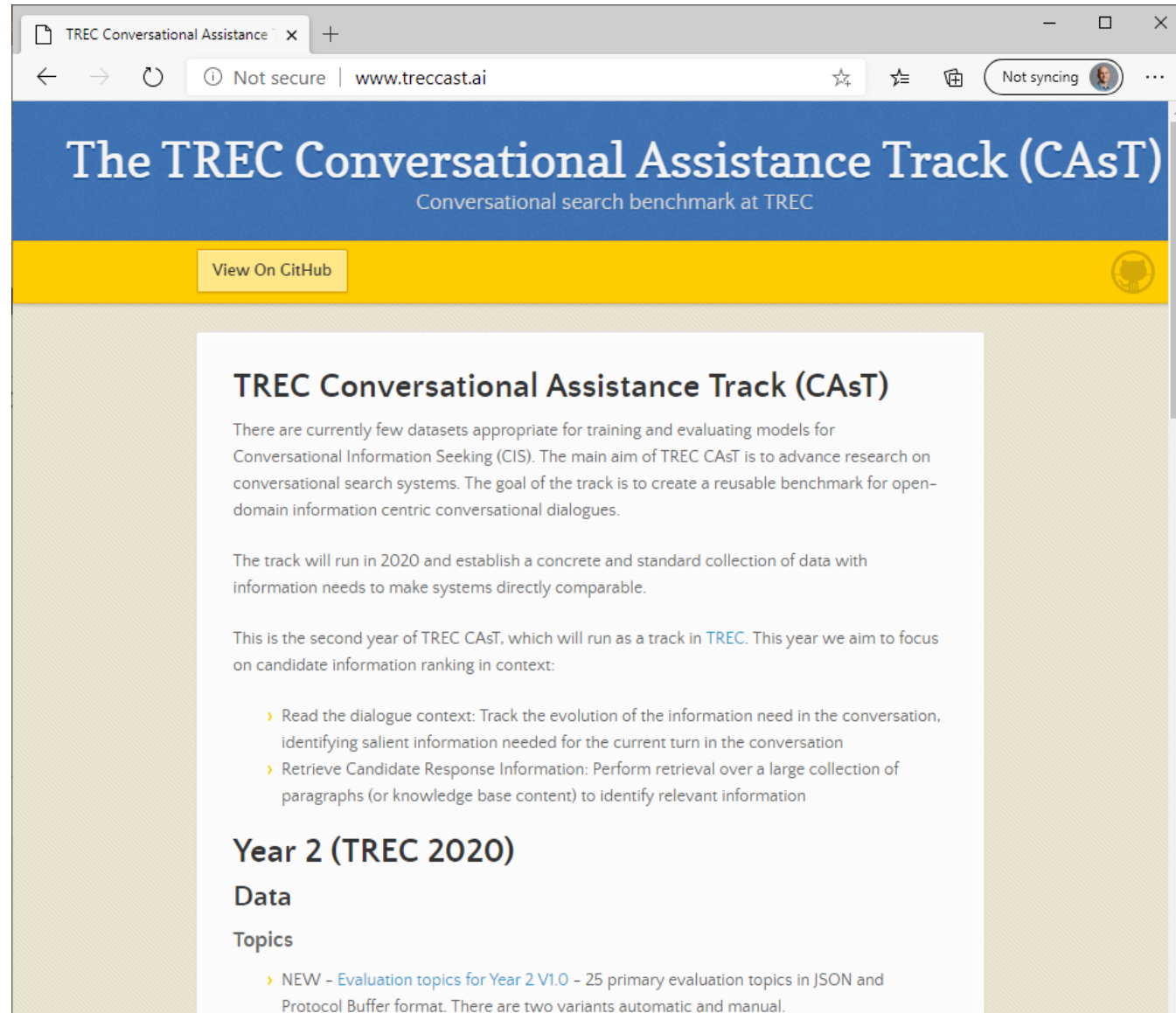
Useful follow-up questions once this question is answered

Clarification Questions

Demonstrate understanding while clarifying

The screenshot shows a search engine interface with the query "convert string to int". Below the search bar, there are tabs for "ALL", "WORK", "IMAGES", "VIDEOS", "MAPS", "NEWS", and "SHOPPING". A clarification question is displayed: "What programming language are you looking for?". Below this question are several buttons representing programming languages: "java", "python", "c++", "c#", "javascript", "swift", "vba", and "matlab". The "python" button is highlighted. Below the clarification question, there is a search result titled "How to Convert a Python String to int" with a list of bullet points: "Representing Integers in Python. An integer can be stored using different types. ...", "Converting a Python String to an int. By default, int () assumes that the string argument represents a decimal integer. ...", "Converting a Python int to a String. ...", and "Conclusion. ...". Below the list is a link to "How to Convert a Python String to int – Real Python" with the URL "realpython.com/convert-python-string-to-int/". At the bottom of the search result, there is a "Was this helpful?" section with thumbs up and down icons. Below the search result, there is a "PEOPLE ALSO ASK" section with four questions: "How do I convert an integer into a string?", "How to convert a wxString to INT?", "How do I make a string in Java?", and "How do I convert a string into an integer in Java?". At the bottom of the search result, there is a link to "Java Convert String to int - javatpoint" with the URL "https://www.javatpoint.com/java-string-to-int" and a brief description: "We can convert String to an int in java using Integer.parseInt () method. To convert String into Integer, we can use Integer.valueOf () method which returns instance of Integer class."

Contextual Understanding



The screenshot shows a web browser window with the URL www.treccast.ai. The page features a blue header with the title "The TREC Conversational Assistance Track (CAsT)" and the subtitle "Conversational search benchmark at TREC". Below the header is a yellow bar with a "View On GitHub" button and a GitHub logo. The main content area is white and contains the following text:

TREC Conversational Assistance Track (CAsT)

There are currently few datasets appropriate for training and evaluating models for Conversational Information Seeking (CIS). The main aim of TREC CAsT is to advance research on conversational search systems. The goal of the track is to create a reusable benchmark for open-domain information centric conversational dialogues.

The track will run in 2020 and establish a concrete and standard collection of data with information needs to make systems directly comparable.

This is the second year of TREC CAsT, which will run as a track in TREC. This year we aim to focus on candidate information ranking in context:

- › Read the dialogue context: Track the evolution of the information need in the conversation, identifying salient information needed for the current turn in the conversation
- › Retrieve Candidate Response Information: Perform retrieval over a large collection of paragraphs (or knowledge base content) to identify relevant information

Year 2 (TREC 2020)

Data

Topics

- › NEW - Evaluation topics for Year 2 V1.0 - 25 primary evaluation topics in JSON and Protocol Buffer format. There are two variants automatic and manual.

Sample TREC CAST 2019 Topic

Title: US Judicial history

Description: Judicial history in the US including key court cases and what they established.

What are the most important US Supreme Court cases?

What did plessy v. ferguson establish?

How about marbury vs madison?

Was it unanimous?

What was the implication of roe vs wade?

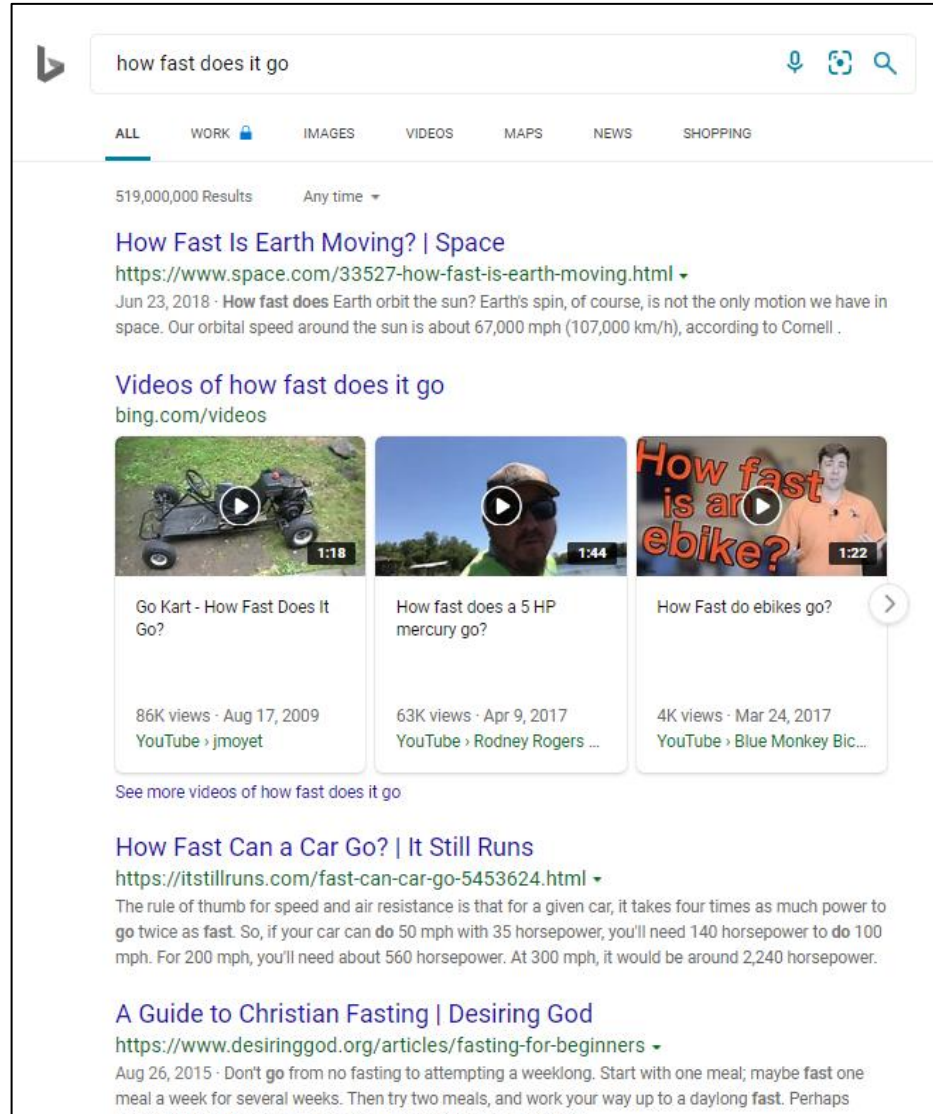
What were the main arguments?

What was the point of the brown v board of education?

What were the main arguments?

Why is it important today?

Contextual Understanding in Search



The image shows a Bing search results page for the query "how fast does it go". The search bar at the top contains the query and has icons for voice search, image search, and a magnifying glass. Below the search bar are navigation tabs for ALL, WORK, IMAGES, VIDEOS, MAPS, NEWS, and SHOPPING. The "ALL" tab is selected. The search results show 519,000,000 results, with a filter for "Any time".

The first result is a text link titled "How Fast Is Earth Moving? | Space" with the URL <https://www.space.com/33527-how-fast-is-earth-moving.html>. The snippet reads: "Jun 23, 2018 · How fast does Earth orbit the sun? Earth's spin, of course, is not the only motion we have in space. Our orbital speed around the sun is about 67,000 mph (107,000 km/h), according to Cornell ."

The second section is titled "Videos of how fast does it go" and includes a link to <bing.com/videos>. It displays three video thumbnails:

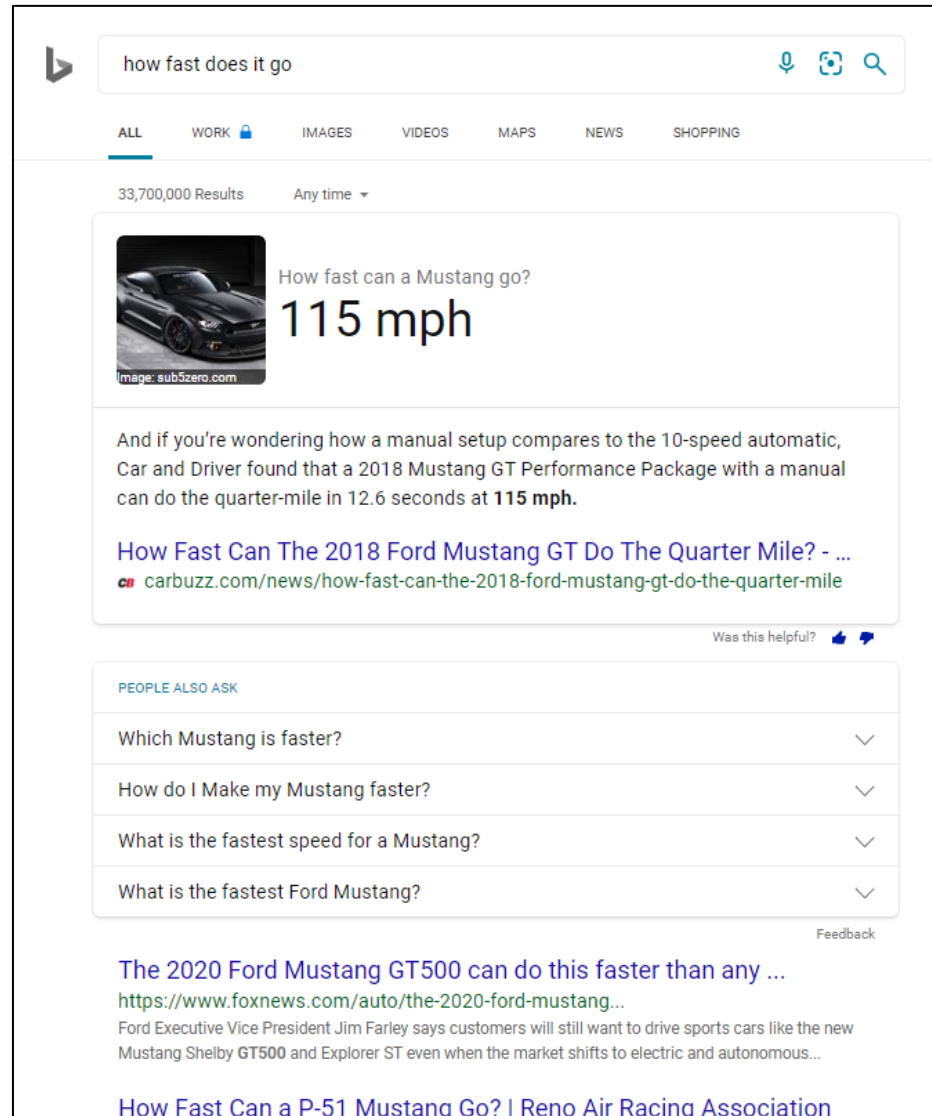
- Go Kart - How Fast Does it Go?**: 86K views · Aug 17, 2009. YouTube · jmoynet. Duration: 1:18.
- How fast does a 5 HP mercury go?**: 63K views · Apr 9, 2017. YouTube · Rodney Rogers ... Duration: 1:44.
- How Fast do ebikes go?**: 4K views · Mar 24, 2017. YouTube · Blue Monkey Bic... Duration: 1:22.

Below the videos is a link to "See more videos of how fast does it go".

The third result is a text link titled "How Fast Can a Car Go? | It Still Runs" with the URL <https://itstillruns.com/fast-can-car-go-5453624.html>. The snippet reads: "The rule of thumb for speed and air resistance is that for a given car, it takes four times as much power to go twice as fast. So, if your car can do 50 mph with 35 horsepower, you'll need 140 horsepower to do 100 mph. For 200 mph, you'll need about 560 horsepower. At 300 mph, it would be around 2,240 horsepower."

The fourth result is a text link titled "A Guide to Christian Fasting | Desiring God" with the URL <https://www.desiringgod.org/articles/fasting-for-beginners>. The snippet reads: "Aug 26, 2015 · Don't go from no fasting to attempting a weeklong. Start with one meal; maybe fast one meal a week for several weeks. Then try two meals, and work your way up to a daylong fast. Perhaps"

Variety of Attempts ... the future?




The screenshot shows a search engine interface with the query "how fast does it go". The search results are filtered to "ALL" and show 33,700,000 results. The top result is a featured snippet titled "How fast can a Mustang go?" with a large "115 mph" value. Below the snippet is a paragraph of text and a link to a news article from carbuzz.com. A "PEOPLE ALSO ASK" section follows with four related questions. At the bottom, there is another link to a Fox News article about the 2020 Ford Mustang GT500.

how fast does it go

ALL WORK IMAGES VIDEOS MAPS NEWS SHOPPING

33,700,000 Results Any time

 How fast can a Mustang go?
115 mph

And if you're wondering how a manual setup compares to the 10-speed automatic, Car and Driver found that a 2018 Mustang GT Performance Package with a manual can do the quarter-mile in 12.6 seconds at **115 mph**.

[How Fast Can The 2018 Ford Mustang GT Do The Quarter Mile? - ...](#)
carbuzz.com/news/how-fast-can-the-2018-ford-mustang-gt-do-the-quarter-mile

Was this helpful?

PEOPLE ALSO ASK

- Which Mustang is faster?
- How do I Make my Mustang faster?
- What is the fastest speed for a Mustang?
- What is the fastest Ford Mustang?

Feedback

[The 2020 Ford Mustang GT500 can do this faster than any ...](#)
<https://www.foxnews.com/auto/the-2020-ford-mustang...>
Ford Executive Vice President Jim Farley says customers will still want to drive sports cars like the new Mustang Shelby GT500 and Explorer ST even when the market shifts to electric and autonomous...

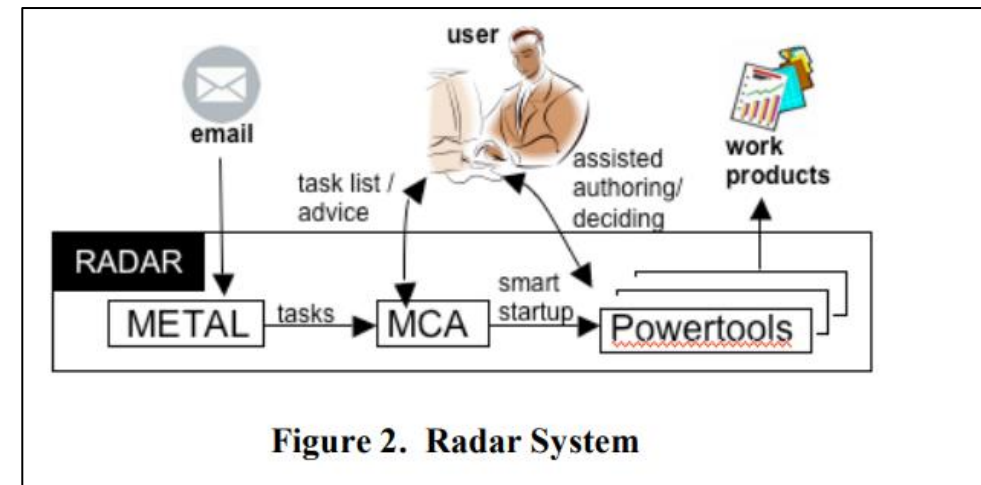
[How Fast Can a P-51 Mustang Go? | Reno Air Racing Association](#)

Productivity and Personal-
Information
Conversational Search

DARPA Personal Assistants that Learn (PAL) CALO / RADAR

Key Focus Points

- Calendar management [Berry et al, 2003; Berry et al., 2006; Modi et al., 2004]
- Dealing with uncertain resources in scheduling [Fink et al., 2006]
- Task management [Freed et al. 2008]



From Freed et al. 2008

From PAL to SIRI

- Learnings from the PAL project including CALO/SIRI recognized need for unifying architectures. [Guzzoni et al., 2007]

A “do engine” rather than a “search engine”

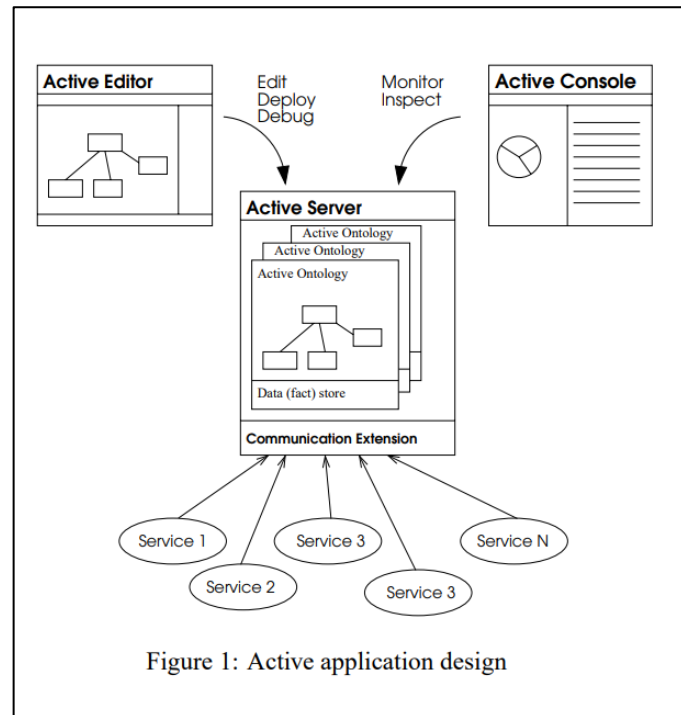


Figure 1: Active application design

From Guzzoni et al, 2007

Device-based Assistants

- Mobile phone based assistants
 - Includes: Apple's Siri, Google Assistant, Microsoft's Cortana
 - Blends productivity-focused and information focused with voice-related recognition
- Situated speakers and Devices
 - Amazon Alexa, Google Home, Facebook Portal w/Alexa, etc.
 - Combines microphone arrays, multi-modal, multi-party devices in addition

Hybrid Intelligence

- Mix AI and Human Computation to achieve an intelligent experience that leverages best of both worlds and push the envelope of possible.
- When escalated to human, often serves as a feedback loop for learning.
- Examples:
 - Facebook's M
 - Microsoft's Calendar.help

Calendar.help → Scheduler

“I think we can drop Greg to BCC; adding Cal from my side. Cal, can you work with Kaitlin to find 60 minutes for Todd and I sometime next week, preferably later in the week when I’m back on EST?”

- Initially high-precision rules
- Unhandled cases handled by low latency human crowdsourcing workflows
- Transition flywheel to machine learning

Current application-oriented research questions

- Long-term evaluation metrics for engagement beyond CPS and NAU (cf. Lowe et al. [2017]; Serban et al. [2017]; Sai et al. [2019])
 - Other metrics of social companionship: linguistic accommodation or coordination?
 - Application to detection: Relationship to the inverse problems of toxicity, bias, etc.
- Aspirational goal-support from assistants
- Best proactivity engagement based on model of interests
- Integrating an understanding of physical environment

Challenges for Conversational Interaction

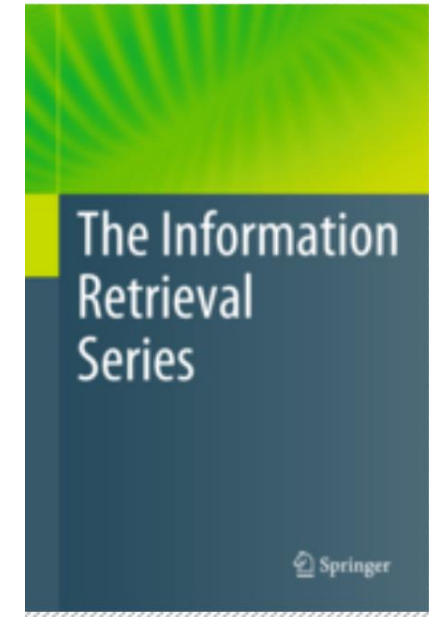
- Human-AI Interaction Design
 - **Goal-directed design:** Enable people to express goals flexibly and allow the agent to progress toward those goals.
 - **Gulf of evaluation:** Communicate the range of skills of an intelligent agent to users and what is available in current context.
- Conversational Understanding
 - **Grounded Language Generation and Learning:** Transform NL intent to action that depends on state and factual correctness.
 - **Extensible Personalized Skills:** Support new skills and remember preferences to evaluate changes/updates.
- External World Perception and Resource Awareness
 - **Multi-modality input and reasoning:** Integrate observations from modalities including voice, vision, and text.
 - **Identity and interactions:** Identify people around and interact with them appropriate to setting.
 - **Physical understanding:** Monitor physical situation and intelligently notify for key situations (safety, anomalies, interest).
 - **Constrained scheduling:** Support reasoning about limited and bound resources such as space/time constraints, keep knowledge of constraints to deal with updates, etc.

Challenges for Conversational Interaction

- Principles & Guarantees
 - **Responsible AI:** Evolve best practice and design new techniques as new ethical challenges arise.
 - **Privacy:** Reason about data in a privacy aware way (e.g. who is in room and what is sensitive).
- Richer paradigms of supervision and learning
 - **Programming by Demonstration/Synthesis:** Turn sequences of actions into higher level macros/scripts that map to NL.
 - **Machine Teaching:** Support efficient supervision schemes from a user-facing perspective that also enable resharing with others (especially for previous bullet).
- Advanced Reasoning
 - **Attention:** Suspend and resume conversation/task naturally based on listener's attention.
 - **Emotional Intelligence:** Support the emotional and social needs of people to enable responsible AI and multi-party social awareness.
 - **Causal Reasoning:** Reason about the impact of taking an action.

Upcoming Book (by early 2021)

Neural Approaches to Conversational Information Retrieval
(The Information Retrieval Series)



Contact Information:

Jianfeng Gao <https://www.microsoft.com/en-us/research/people/jfgao/>

Chenyan Xiong <https://www.microsoft.com/en-us/research/people/cxiong/>

Paul Bennett <https://www.microsoft.com/en-us/research/people/pauben/>

Slides:

Please check our personal websites.