MARRS: Multi-Modal Reference Resolution System

Joel Moniz, Apple

CRAC 2023 | December 6, 2023

https://machinelearning.apple.com/research/marrs

Agenda

Why context matters?
Different types of applications and context.
Our proposed system: an overview
    Query Rewriting
    Reference Resolution

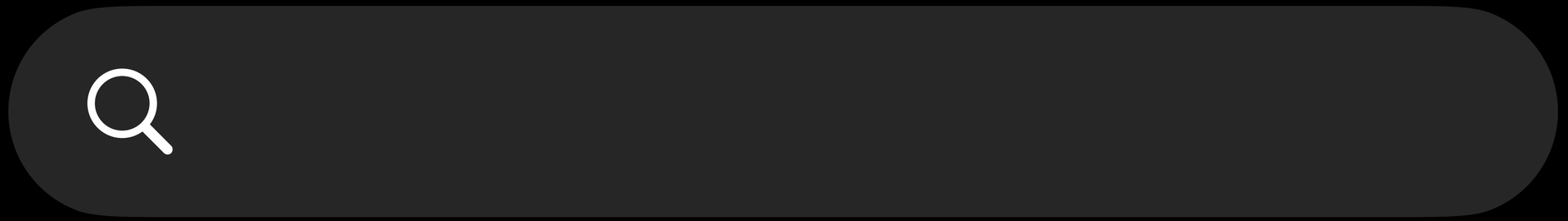
Agenda

Why does context matter?

Different types of applications and context

Our proposed system: an overview

- Query Rewriting
- Reference Resolution



how many people like roses?

Bouquet - EP

The Chainsmokers

DANCE · 2015

A colossal blast of radio-ready pop hooks and steep drops—featuring "Roses."...

MORE

Lossless

Play | Shuffle | + Add | •••

Roses (feat. ROZES)		3:46
2	New York City	3:51
3	Until You Were Gone (feat. Emily Warren) The Chainsmokers, Tritonal	3:37

Agenda

Why does context matter?

Different types of applications and context

Our proposed system: an overview

- Query Rewriting
- Reference Resolution

Search Engines

Virtual Assistants

Chatbots

IVR

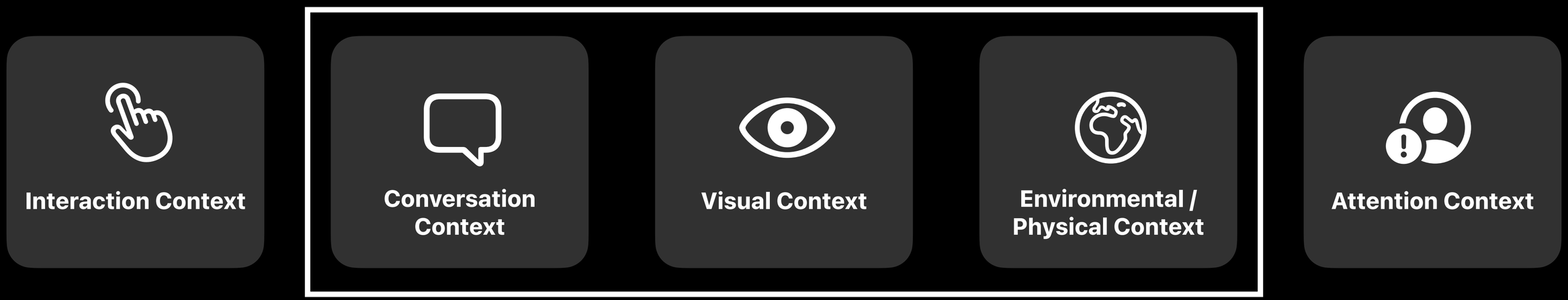
Interaction Context

Conversation Context

Visual Context

Environmental / Physical Context

Attention Context


Interaction Context
Conversation Context
Visual Context
Environmental / Physical Context
Attention Context

Agenda

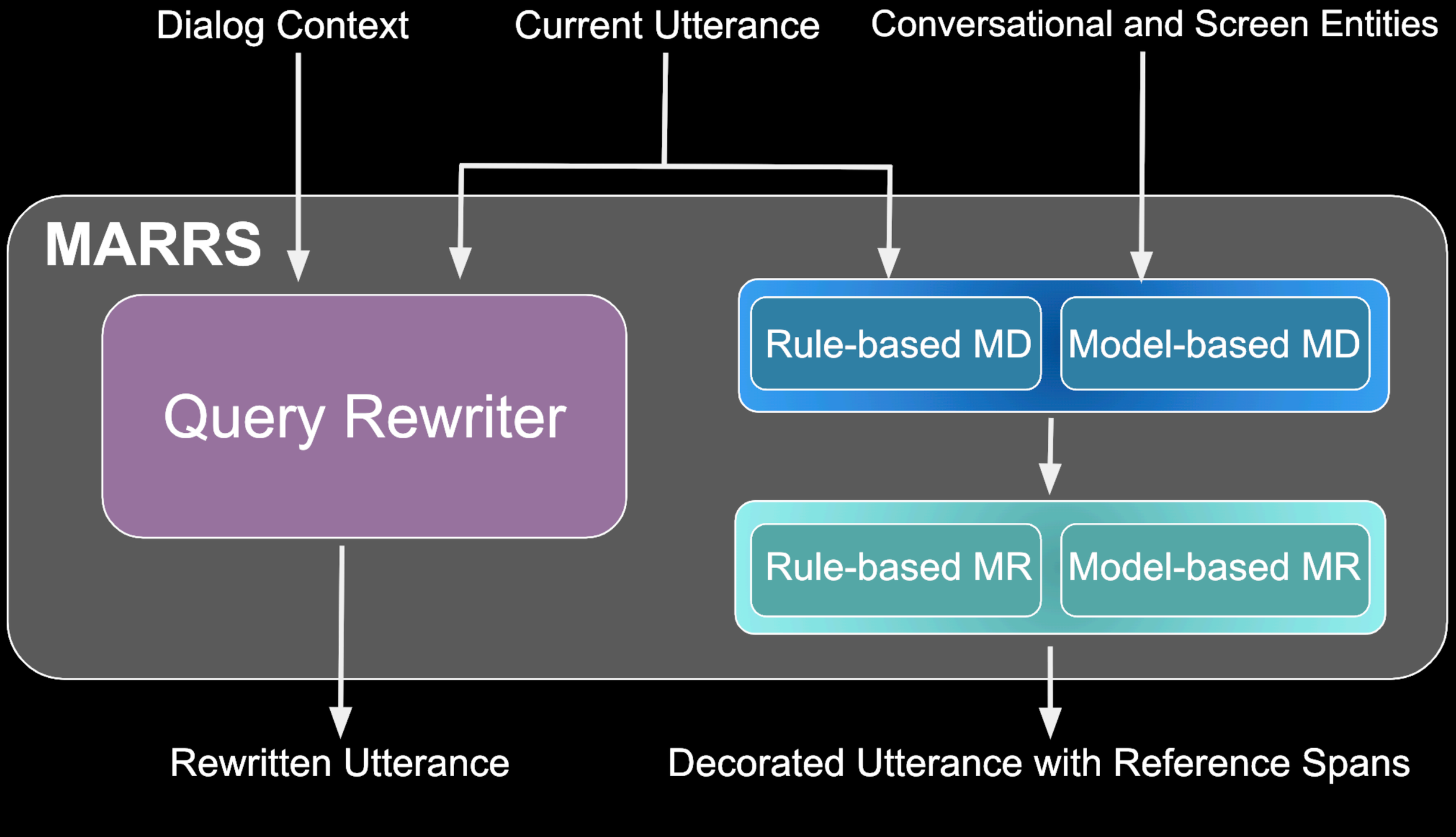
Why context matters?

Different types of applications and context

MARRS: an overview

- Query Rewriting
- Reference Resolution

Our proposed system: MARRS


Dialog Context
Current Utterance
Conversational and Screen Entities
MARRS
Query Rewriter
Rule-based MD
Model-based MD
Rule-based MR
Model-based MR
Rewritten Utterance
Decorated Utterance with Reference Spans

Query Rewriting

QR

Use cases

QR

Use cases

Conversational Queries

Q1: Play Sweeny Todd

Q2: On my living room HomePod

QR

Use cases

Conversational Queries

Q1: Play Sweeny Todd

Q2: On my living room HomePod

Equivalent Query

Q: Play Sweeny Todd on my living room HomePod

QR

Use cases

Conversational Queries

Q1: Play Sweeny Todd

Q2: On my living room HomePod

Query Rewrite

Equivalent Query

Q: Play Sweeny Todd on my living room HomePod

QR

Use cases

Conversational Queries

Q1: How old is Barack Obama

Q2: What about Joe Biden

Query Rewrite

Equivalent Query

QR

Use cases

Conversational Queries

Q1: How old is Barack Obama

Q2: What about Joe Biden

Query Rewrite

Equivalent Query

Q: How old is Joe Biden

QR

Use cases

Conversational Queries

Q1: When does Rocket Sushi close

Q2: How long does it take to drive there

Query Rewrite

Equivalent Query

QR

Use cases

Conversational Queries

Q1: When does Rocket Sushi close

Q2: How long does it take to drive there

Query Rewrite

Equivalent Query

Q: How long does it take to drive to Rocket Sushi

QR

Use cases

Conversational Queries

Q1: How tall is Barack Obama

Q2: What about his wife

Query Rewrite

Equivalent Query

QR

Use cases

Conversational Queries

Q1: How tall is Barack Obama

Q2: What about his wife

Query Rewrite

Equivalent Query

Q: How tall is Barack Obama's wife

QR

Data Collection

QR

Data Collection

- Propose a query rewrite data collection strategy by mining “improvement opportunities”

QR

Data Collection

- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:

QR

Data Collection

- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses

QR

Data Collection

- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
  - correction by repetition

QR

Data Collection: Anaphora and Ellipses

- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses

QR

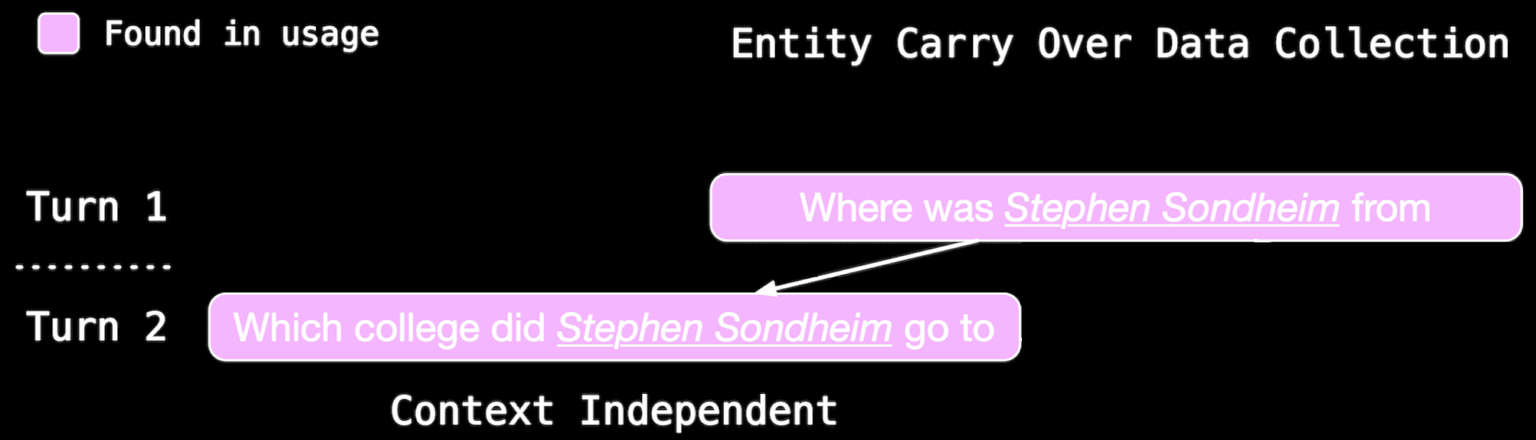
Data Collection: Anaphora and Ellipses

- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
    - identify user queries discussing the same entity in two consecutive turns without the use of any referring expressions (context-free query)

QR

Data Collection: Anaphora and Ellipses

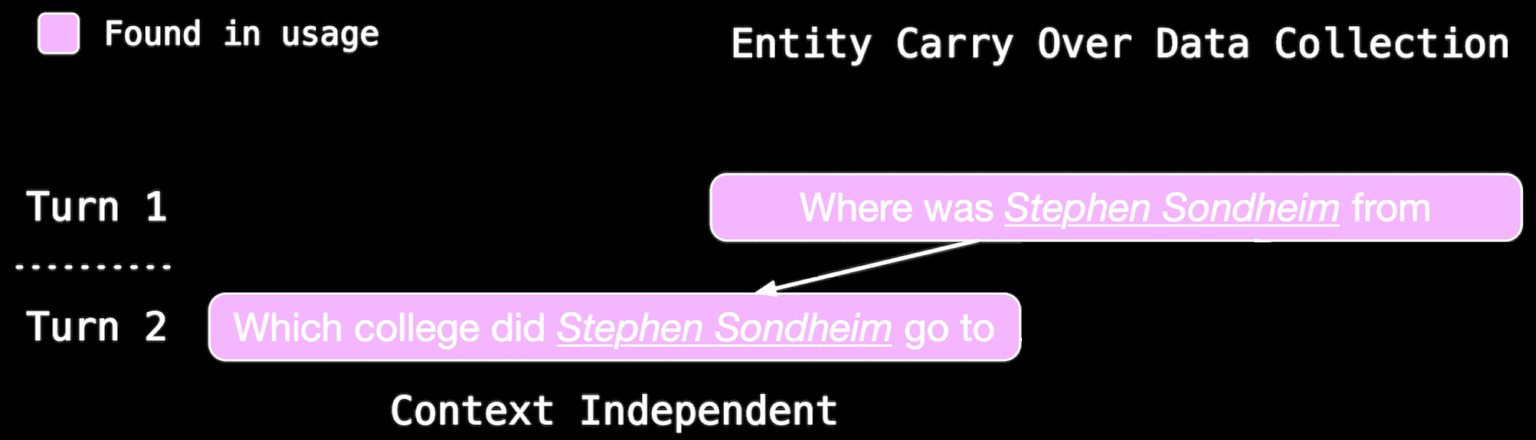
- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
    - identify user queries discussing the same entity in two consecutive turns without the use of any referring expressions (context-free query)



QR

Data Collection: Anaphora and Ellipses

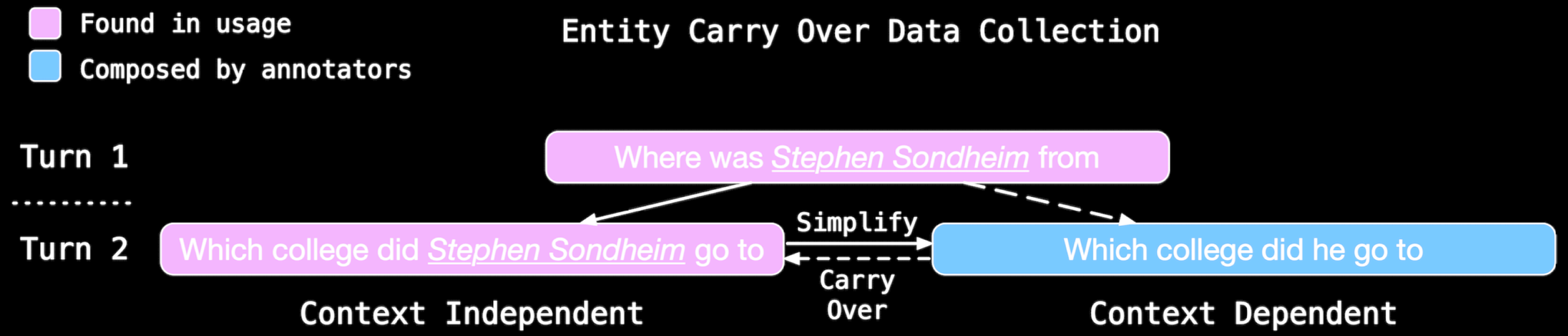
- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
    - identify user queries discussing the same entity in two consecutive turns without the use of any referring expressions (context-free query)
    - ask annotators to simplify these complex queries to provide queries in a more natural way (context-dependent query)



QR

Data Collection: Anaphora and Ellipses

- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
    - identify user queries discussing the same entity in two consecutive turns without the use of any referring expressions (context-free query)
    - ask annotators to simplify these complex queries to provide queries in a more natural way (context-dependent query)



QR

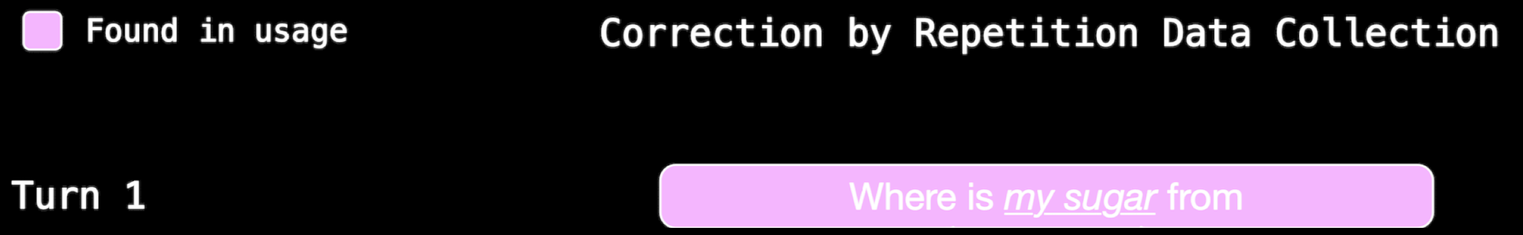
Data Collection: Correction by Repetition

- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
  - correction by repetition

QR

Data Collection: Correction by Repetition

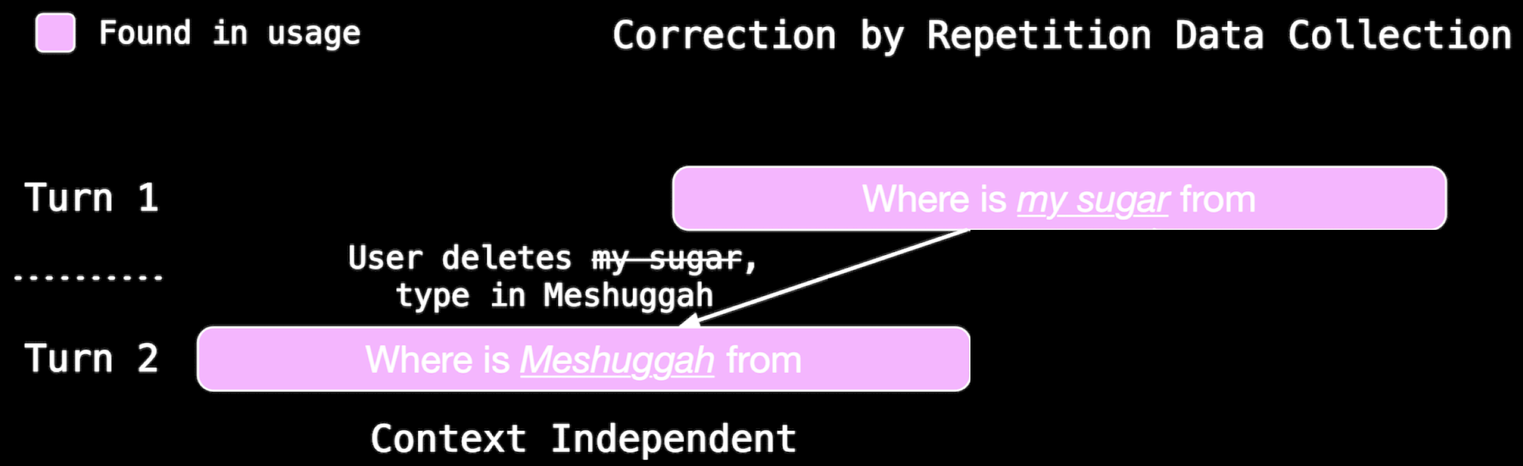
- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
  - correction by repetition
    - recognize queries where the user tapped on the transcribed prompt to edit the query into something else (context-free query)



QR

Data Collection: Correction by Repetition

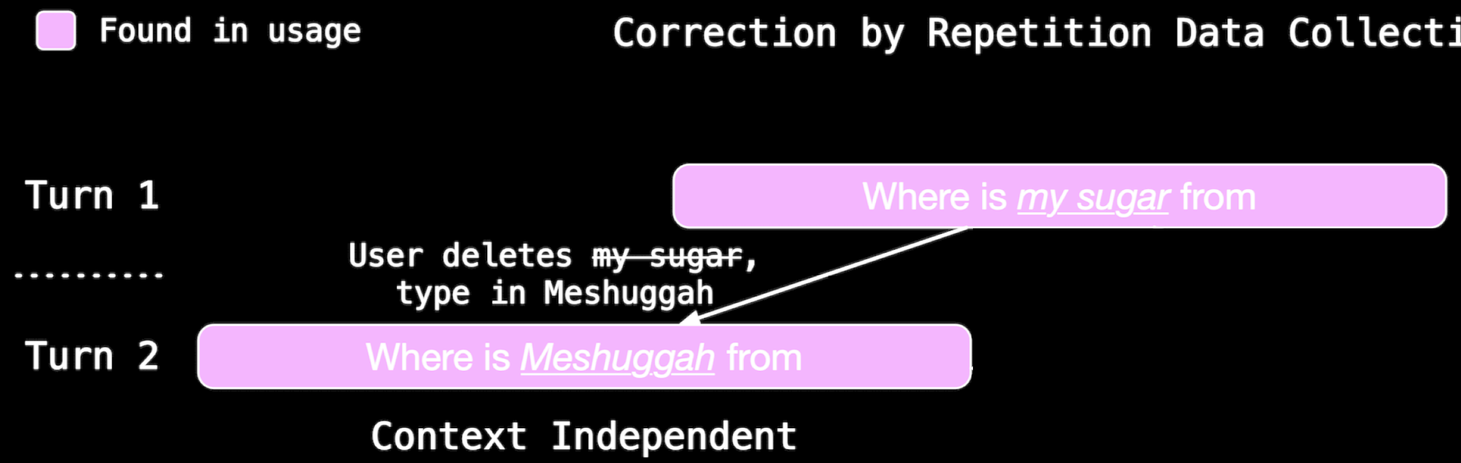
- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
  - correction by repetition
    - recognize queries where the user tapped on the transcribed prompt to edit the query into something else (context-free query)



QR

Data Collection: Correction by Repetition

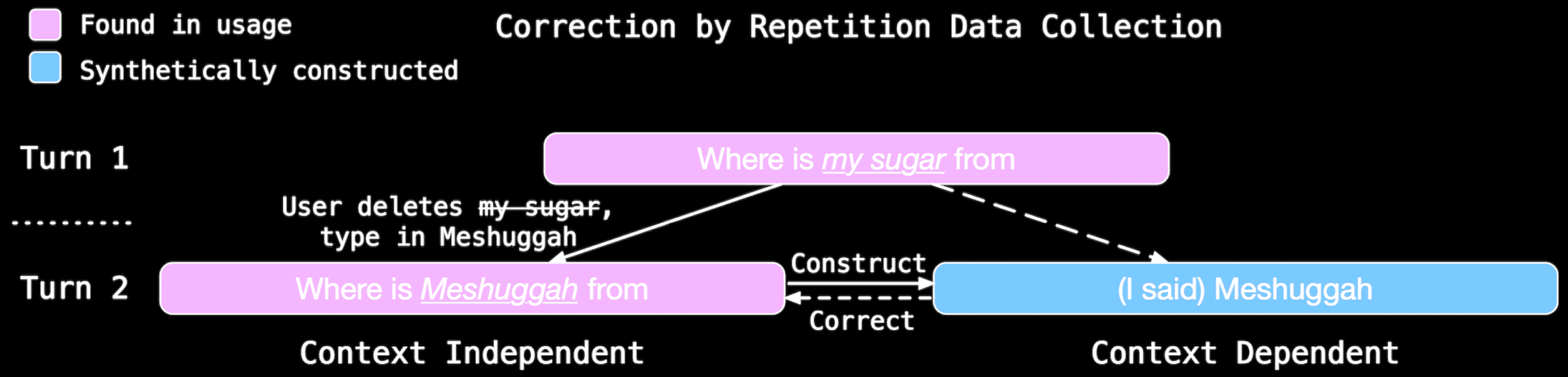
- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
  - correction by repetition
    - recognize queries where the user tapped on the transcribed prompt to edit the query into something else (context-free query)
    - prepend edited parts with common prefixes like “*I said*” to synthesize the context-dependent query



QR

Data Collection: Correction by Repetition

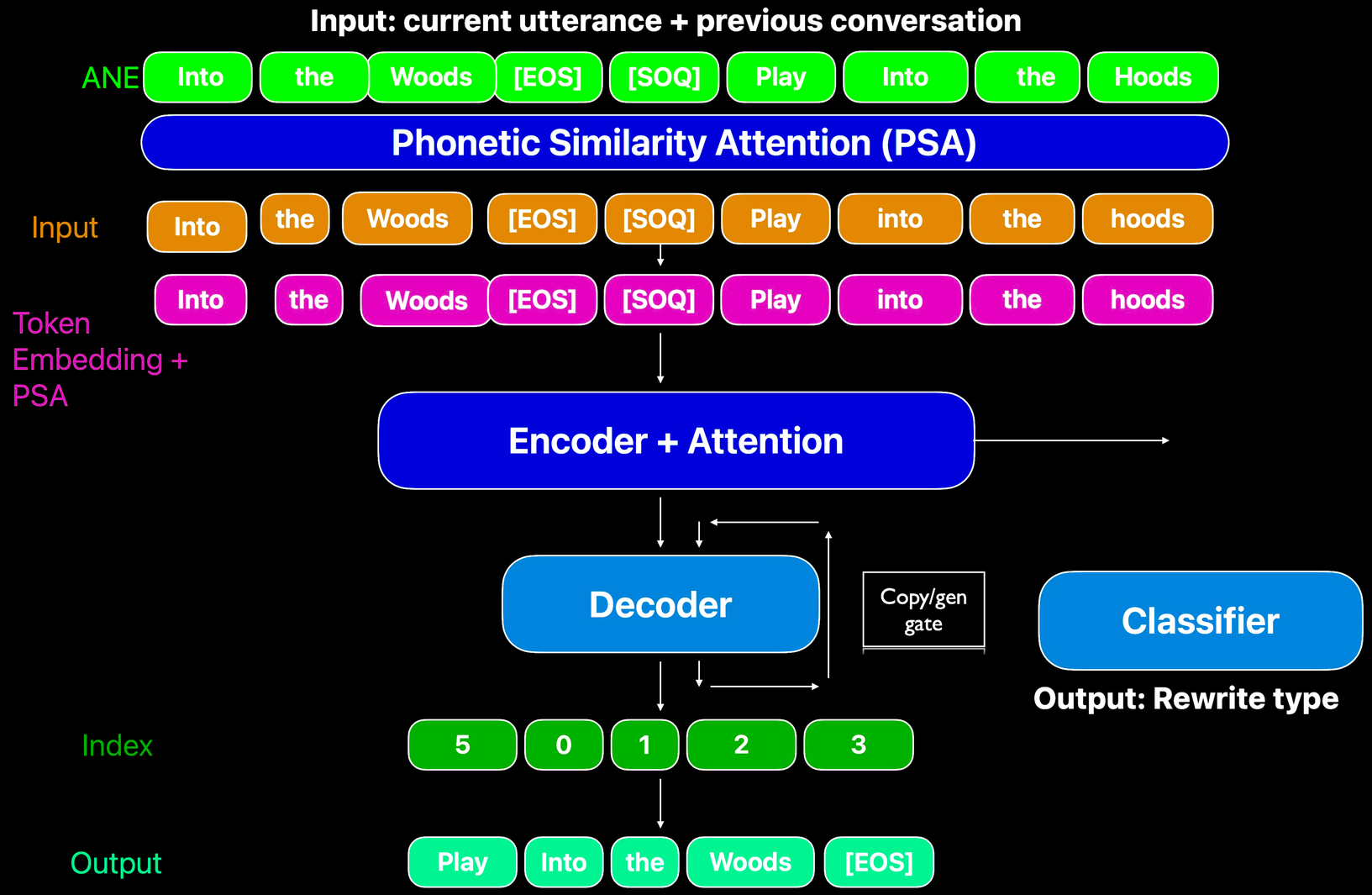
- Propose a query rewrite data collection strategy by mining “improvement opportunities”
- For the use cases of:
  - anaphora and ellipses
  - correction by repetition
    - recognize queries where the user tapped on the transcribed prompt to edit the query into something else (context-free query)
    - prepend edited parts with common prefixes like “*I said*” to synthesize the context-dependent query



QR

Model

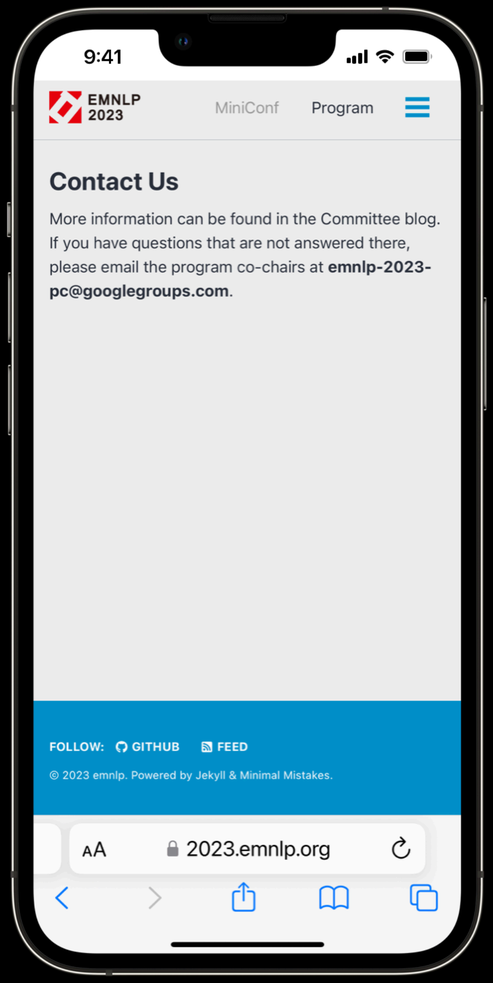
- Joint QR model for 2 conversational use cases, entity carry over and correction by repetition
- Joint learning shows significant latency and accuracy improvement for correction by repetition use case



Reference Resolution

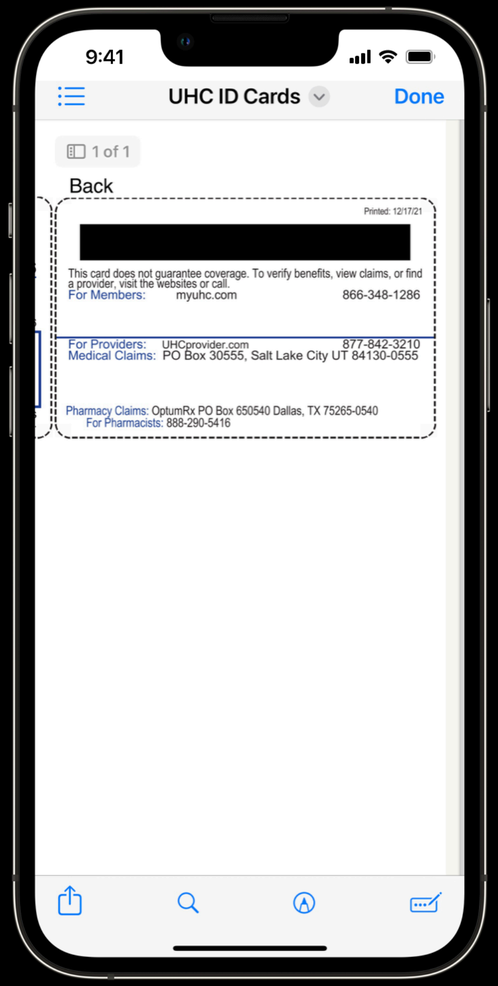
Reference Resolution

Use cases



Save an email on screen

Send message to physical card



Dial a number in document

Reference Resolution

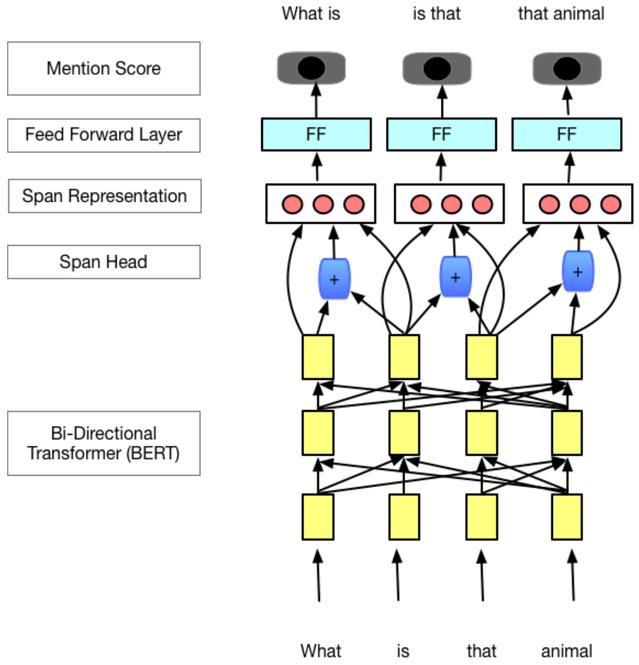
Data

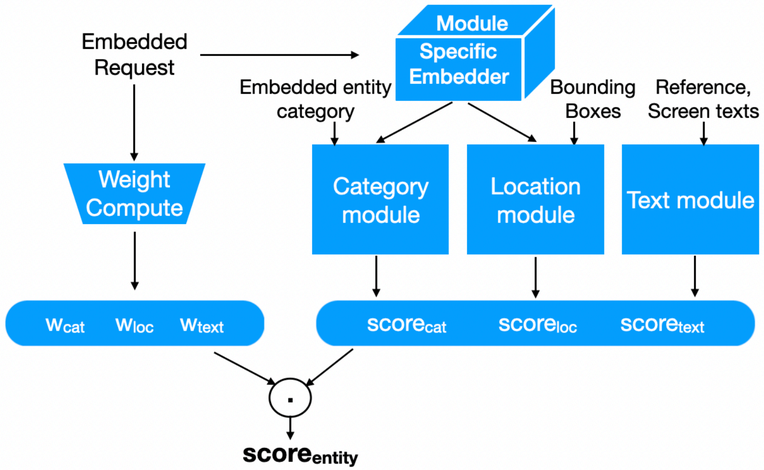
- Screen and Conversational Entity Resolution Data
  - Show annotators a screenshots and ask them to refer to entities as a user would
- Synthetic Data
  - Generated via templates

Reference Resolution

Model

- Proposed a lightweight mention detector and modular reference resolution model
- Flexible to support varied use cases - screen referencing, visual and text referencing
- Scores are computed for the mention against each entity independently
- Reuses upstream feature extractors

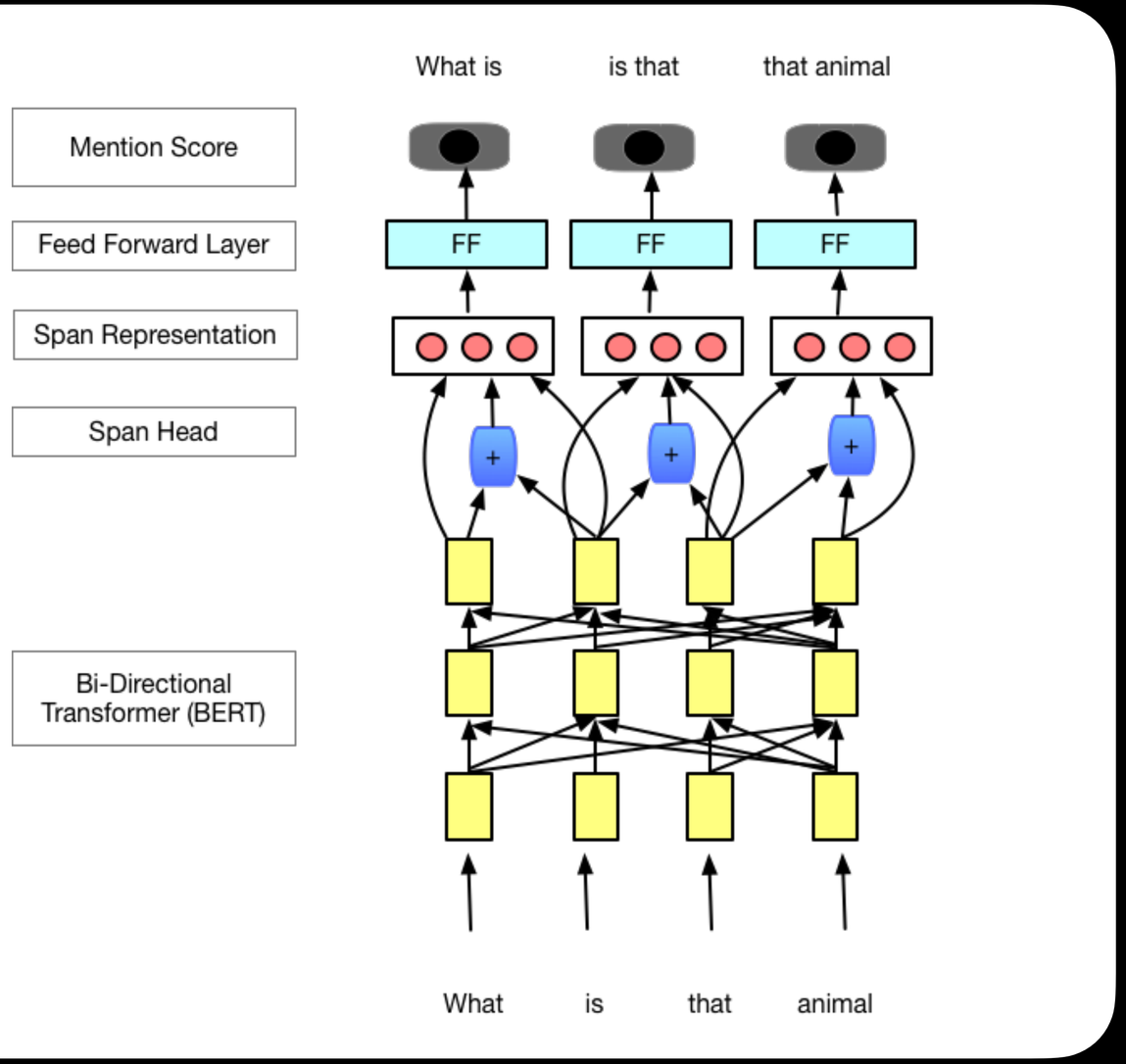




Reference Resolution

Model: MD

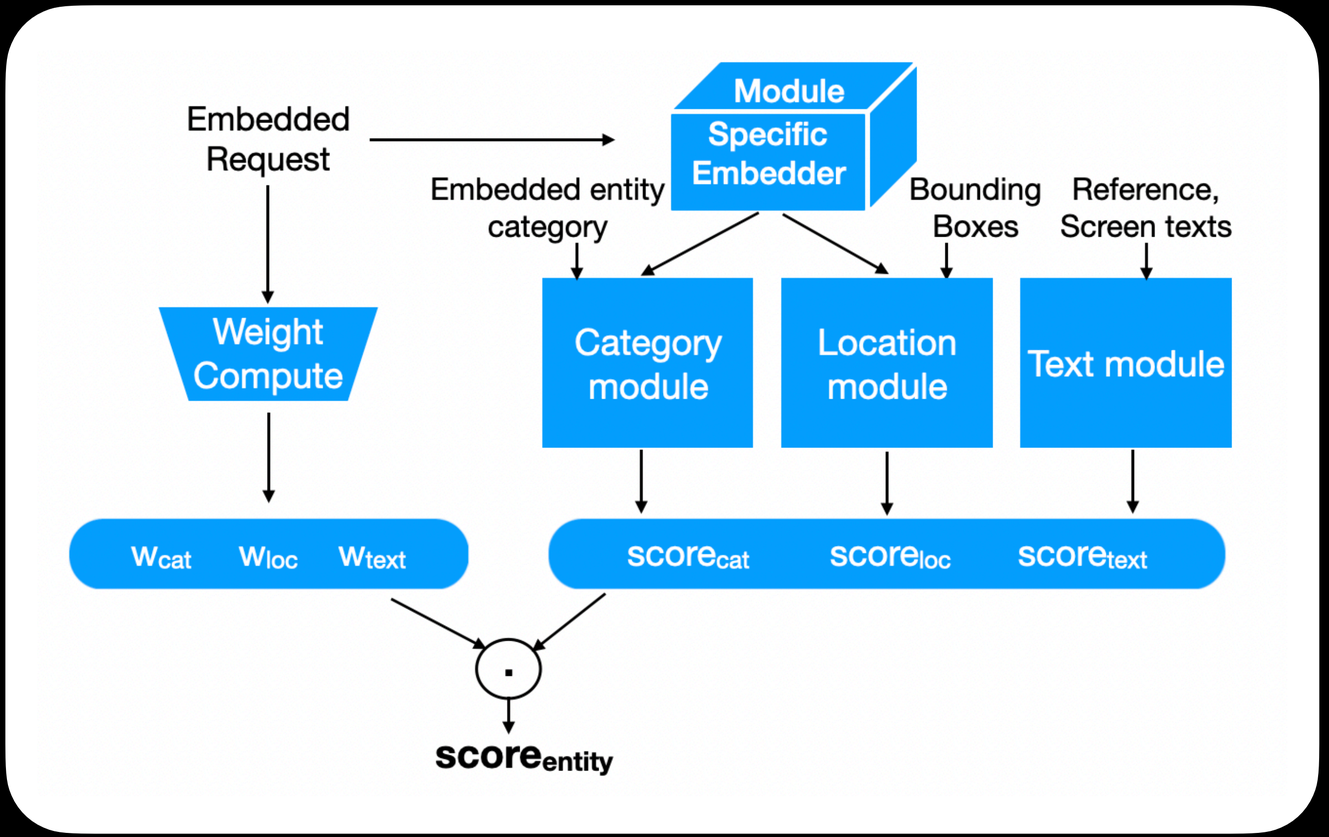
- Lightweight mention detector performs span-level classification
- Identifies relevant mention spans that need resolution



Reference Resolution

Model: MR

- Modular Mention Resolver identifies entities relevant to the mention using the mention span, along with other information such as:
  - The entity category
  - The location of the entity on the screen
  - The entity text
  - The location and text of entities surrounding the current entity



TM and © 2023 Apple Inc. All rights reserved.