

---

# A LITTLE CHAT GOES A LONG WAY

## HANDS-ON TUTORIAL ON CONVERSATIONAL INFORMATION ACCESS

Guglielmo **Faggioli**, Nicola **Ferro**, Simone **Merlo**



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA





# ABOUT US

---



Guglielmo Faggioli  
PostDoc Researcher  
[guglielmo.faggioli@unipd.it](mailto:guglielmo.faggioli@unipd.it)



Nicola Ferro  
Full Professor  
[ferro@dei.unipd.it](mailto:ferro@dei.unipd.it)



Simone Merlo  
PhD Student  
[simone.merlo@phd.unipd.it](mailto:simone.merlo@phd.unipd.it)



# WORKSHOP OUTLINE

## INTRODUCTION

Introduction on Conversational Search

Introduction on Conversational Recommendation

Joint Conversational Information Access

## HANDS-ON

Query Reformulation and Dense Encoding

CoSRec: building and using a conversational joint IR and RS dataset

---

# WHAT IS CONVERSATIONAL SEARCH NOT

# WHAT IS CONVERSATIONAL SEARCH

---



# WHAT IS CONVERSATIONAL SEARCH



# CONVERSATIONAL SEARCH & VIRTUAL ASSISTANTS

---

They are similar because:

They are different because:

# CONVERSATIONAL SEARCH & VIRTUAL ASSISTANTS

They are similar because:

- Both allow natural language interaction.
- Smart assistants' answers are designed to be enjoyed through different media (speech/text).

They are different because:

- The interaction is (mostly) one way.
- The interaction is (mostly) single turn.



# WHAT IS CONVERSATIONAL SEARCH

---

## Evaluating Retrieval over Sessions: The TREC Session Track 2011–2014

Ben Carterette  
University of Delaware,  
Newark, DE, USA  
[carteret@cis.udel.edu](mailto:carteret@cis.udel.edu)

Paul Clough  
University of Sheffield,  
Sheffield, UK  
[p.d.clough@sheffield.ac.uk](mailto:p.d.clough@sheffield.ac.uk)

Mark Hall  
Edge Hill University, Ormskirk,  
UK  
[hallmark@edgehill.ac.uk](mailto:hallmark@edgehill.ac.uk)

Evangelos Kanoulas  
University of Amsterdam,  
Amsterdam, The Netherlands  
[e.kanoulas@uva.nl](mailto:e.kanoulas@uva.nl)

Mark Sanderson  
RMIT University, Melbourne,  
Australia  
[mark.sanderson@rmit.edu.au](mailto:mark.sanderson@rmit.edu.au)



# WHAT IS CONVERSATIONAL SEARCH

## Evaluating Retrieval over Sessions: The TREC Session Track 2011–2014

Ben Carterette  
University of Delaware,  
Newark, DE, USA  
[carteret@cis.udel.edu](mailto:carteret@cis.udel.edu)

Paul Clough  
University of Sheffield,  
Sheffield, UK  
[p.d.clough@sheffield.ac.uk](mailto:p.d.clough@sheffield.ac.uk)

Mark Hall  
Edge Hill University, Ormskirk,  
UK  
[hallmark@edgehill.ac.uk](mailto:hallmark@edgehill.ac.uk)

Evangelos Kanoulas  
University of Amsterdam,  
Amsterdam, The Netherlands  
[e.kanoulas@uva.nl](mailto:e.kanoulas@uva.nl)

Mark Sanderson  
RMIT University, Melbourne,  
Australia  
[mark.sanderson@rmit.edu.au](mailto:mark.sanderson@rmit.edu.au)

# CONVERSATIONAL SEARCH & SESSION SEARCH

---

They are similar because:

They are different because:

# CONVERSATIONAL SEARCH & SESSION SEARCH

They are similar because:

- Both of them are expected to model the short-term search trajectory.

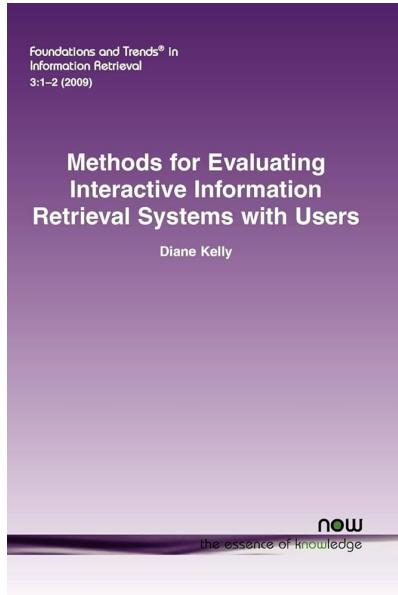
They are different because:

- Conversational search requires mixed initiative and the system should be active part of the conversation.



# WHAT IS CONVERSATIONAL SEARCH

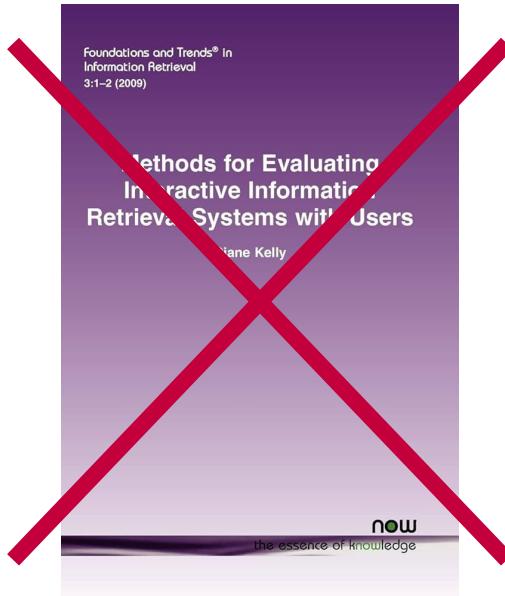
---





# WHAT IS CONVERSATIONAL SEARCH

---





# CONVERSATIONAL SEARCH & IIR

---

They are similar because:

They are different because:

# CONVERSATIONAL SEARCH & IIR

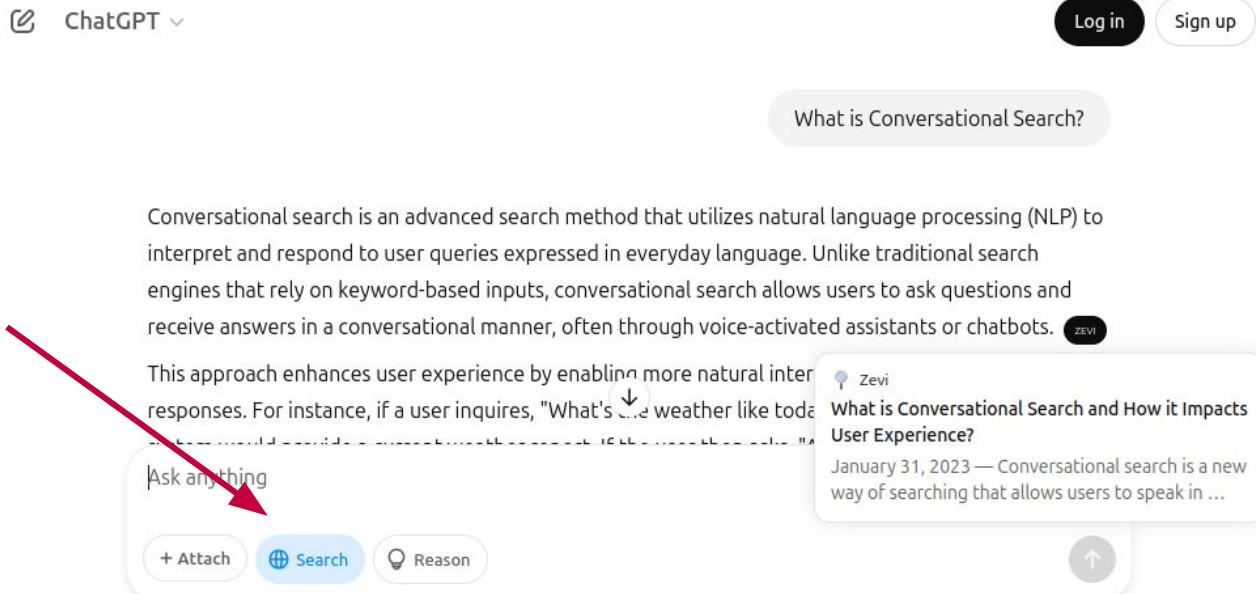
They are similar because:

- User interaction with the system is a focal aspect of both research areas.

They are different because:

- IIR focuses on how users interact with IR systems, and not on a specific modality (i.e., conversational).

# WHAT IS CONVERSATIONAL SEARCH



The screenshot shows the ChatGPT interface. At the top left is the ChatGPT logo and a dropdown menu. On the right are 'Log in' and 'Sign up' buttons. In the center, a search bar contains the query 'What is Conversational Search?'. Below the search bar, a text box explains what conversational search is. A red arrow points from this text box down to the 'Ask anything' input field. The input field has placeholder text 'Ask anything' and three buttons: '+ Attach', 'Search', and 'Reason'. To the right of the input field is a card with a user profile 'Zevi' and the title 'What is Conversational Search and How it Impacts User Experience?'. Below the title is a snippet of text: 'January 31, 2023 — Conversational search is a new way of searching that allows users to speak in ...'.

ChatGPT

Log in Sign up

What is Conversational Search?

Conversational search is an advanced search method that utilizes natural language processing (NLP) to interpret and respond to user queries expressed in everyday language. Unlike traditional search engines that rely on keyword-based inputs, conversational search allows users to ask questions and receive answers in a conversational manner, often through voice-activated assistants or chatbots. Zevi

This approach enhances user experience by enabling more natural interactions. For instance, if a user inquires, "What's the weather like today?"

Ask anything

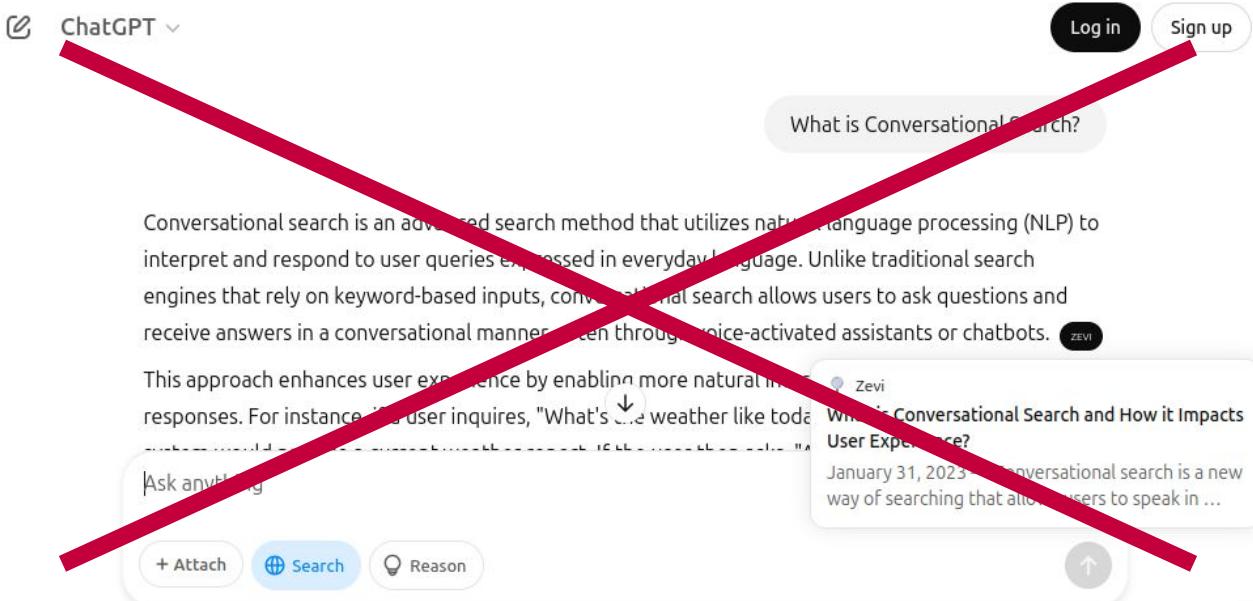
+ Attach Search Reason

Zevi

What is Conversational Search and How it Impacts User Experience?

January 31, 2023 — Conversational search is a new way of searching that allows users to speak in ...

# WHAT IS CONVERSATIONAL SEARCH



# CONVERSATIONAL SEARCH & LLMS

---

They are similar because:

They are different because:



# CONVERSATIONAL SEARCH & LLMS

They are similar because:

- Both allow natural language interaction.
- Both require a modeling the short-term context.

They are different because:

- Limited control over the output, trustworthiness problems.
- The interaction is (mostly) one way.
- No long-term representation of the user.
- *Bonus:* CS is a task, LLMs are tools.

# WHAT IS CONVERSATIONAL SEARCH



## Personalization & Google Search results

With personalization, you get Google Search results tailored for you based on [your activity](#). Personalization is only used if it can provide information that's likely to be more helpful to you. This page explains how Google shows personalized results and how to control them.

## Search Personalization at Netflix

[Recommendations](#) | [Personalization & Search](#)

[Publication](#) | [Paper](#)

April 30, 2023

### Abstract

At Netflix, personalization plays a key role in several aspects of our user experience, from ranking titles to constructing an optimal Homepage. Although personalization is a well established research field, its application to search presents unique problems and opportunities. In this paper, we describe the evolution of Search personalization at Netflix, its unique challenges, and provide a high level overview of relevant solutions.



## Authors

-  Vito Ostuni, Research Scientist
-  Christoph Kofler, Product Manager - Search Innovation
-  Manjesh Nilange, Software Engineer
-  Sudarshan Lamkhede, Senior Research Engineer
-  Dan Zylberglejd, Data Scientist

## Personalization and data



We want to provide you with the most interesting content when you use Pinterest. To do this, we use information like your activity on Pinterest (such as the Pins you click or the boards you create) and information from Pinterest site features like the Save button.

### Other articles

[Account privacy](#)

+

# WHAT IS CONVERSATIONAL SEARCH



## Personalization & Google Search results

With personalization, you get Google search results tailored for you based on [your activity](#). Personalization is only used if it can provide information that's likely to be more helpful to you. This page explains how Google shows personalized results and how to control them.

## Personalization and data

We want to provide you with the most interesting content while you use Pinterest. To do this, we use information like your activity on Pinterest (such as the Pins you click or the boards you create) and information from Pinterest's site features like the Save button.



Other articles  
Account privacy

## Authors

-  Vito Ostuni, Research Scientist
-  Christoph Kofler, Product Manager - Search Innovation
-  Manjesh Nilange, Software Engineer
-  Sudarshan Lamkhede, Senior Research Engineer
-  Dan Zylberglejd, Data Scientist



# CONVERSATIONAL SEARCH & PERSONALISED SEARCH

---

They are similar because:

They are different because:



# CONVERSATIONAL SEARCH & PERSONALISED SEARCH

---

They are similar because:

- Both types of search should account for long-term user's tastes and preferences.

They are different because:

- Personalised search assumes a one-shot interaction between the user and the system.



# WHAT IS CONVERSATIONAL SEARCH

---

**IMPORTANT:** this does not mean that CS cannot be studied as branch of IIR or that an LLM cannot be a part of a CS.

All these research areas must be integrated and combined to achieve full conversational search.

When studying CS, for abstraction and simplicity, we can focus on only some of these areas at once!

# WHAT IS CONVERSATIONAL SEARCH

---

**Conversational Search (CS)** systems are **IR** systems that allow for **multi-turn**, **mixed-initiative, natural language interactions**, capable of modeling and acting upon changes on the **short-term conversation state**, while adapting the conversation to the users' **long-term characteristics**. CS answers should be enjoyable via **different interaction channels** including, but not limited to, **speech**, short textual utterances, and classical search engines interfaces.

# WHAT IS CONVERSATIONAL SEARCH

## Session Search

**Conversational Search (CS)** systems are **IR** systems that allow for **multi-turn**,  
**NLP**

**mixed-initiative**, **natural language interactions**, capable of modeling and  
**LLMs**

acting upon changes on the **short-term conversation state**, while adapting

## Personalised IR

the conversation to the users' **long-term characteristics**. CS answers should

## Interactive IR

be enjoyable via **different interaction channels** including, but not limited to,  
**Virtual Assistants**

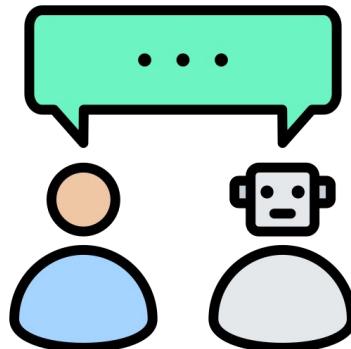
**speech**, short textual utterances, and classical search engines interfaces.

# WHY WE WANT CONVERSATIONAL SEARCH

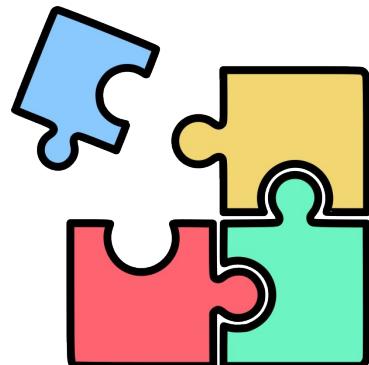
**Ease of use:**  
suited for children  
and elderly



**Natural language:** low  
learning curve and  
mental encumbering



Handles **complex**  
and **composite**  
**information needs**





# WHAT ARE THE CHALLENGES

---

There are several additional challenges compared to standard IR:

- Handle complex natural language constructs: **anaphoras, ellipses, co-references.**
- Model the **conversation: topic shifts, topic returns** and **contradictory inputs.**
- (In the future) model the long-term user's characteristics

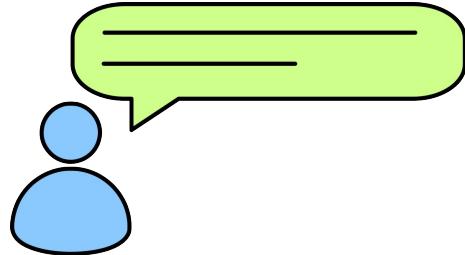


# A TYPICAL CONVERSATION

## User's Utterance

the first is (typically)  
self-contained.

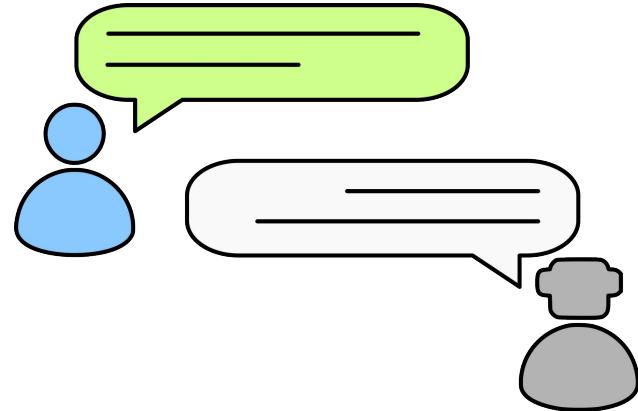
Sometimes also referred to as  
“turn”.





# A TYPICAL CONVERSATION

**User's Utterance**



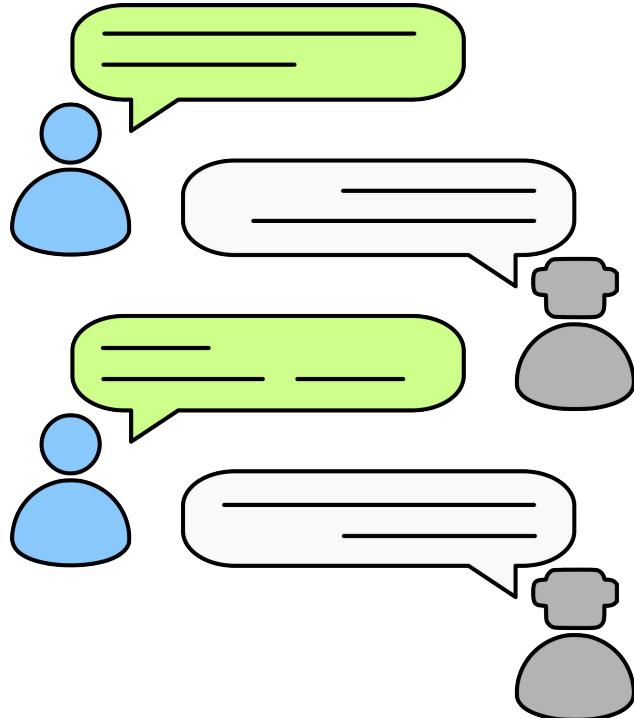
**System's Answer**

Must involve retrieval (otherwise, is not search 😊).

Could be a list of documents or a summary.



# A TYPICAL CONVERSATION



## Conversation

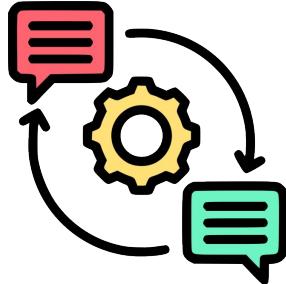
In a real world conversation, we can expect that:

- Utterances refer to each other, sometimes implicitly.
- Both parties take initiative.
- The topic might change.
- One of the parties might not have understood the other.

# WHAT MAKES A CONVERSATIONAL SEARCH SYSTEM

A CS system, typically, has one or more of the following components:

Query  
Reformulation



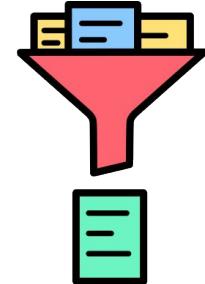
Question  
Clarification



Search



Response  
Generation





# QUERY REFORMULATION

---

One of the most important parts of a CS system consists of transforming an **utterance** in a **query**, to include the conversation context.

There are two main strategies:

- **query expansion**
- **query rewriting**



# QUERY EXPANSION: QuReTeC

Query Expansion is often based on **adding relevant terms** from the previous utterances/answers to the utterance.

Differently from query rewriting, query expansion produces queries that are **not well-formed natural language queries**.

A seminal approach for query expansion is **Query Resolution by Term Classification (QuReTeC)**.

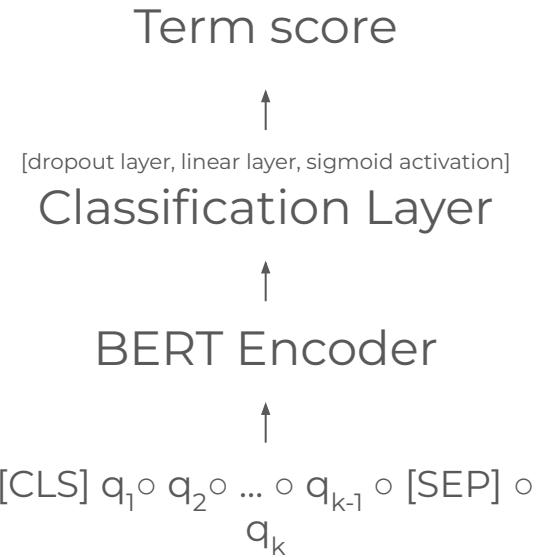
Nikos Voskarides, Dan Li, Pengjie Ren, Evangelos Kanoulas, Maarten de Rijke: Query Resolution for Conversational Search with Limited Supervision. SIGIR 2020: 921-930

# QUERY EXPANSION: QuReTeC

QuReTeC employs a **BERT-based classifier** to identify **context terms** that fully specify the query.

We call  $q_k$  the current utterance and  $q_1, \dots, q_{k-1}$  the previous utterances.

$q'_k$  is the self-contained query answering the same information need of  $q_k$ .





# QUERY EXPANSION: QuReTeC

To obtain the ground truth (i.e., a query with the , QuReTec employs a **weak supervision** approach.

relevant terms appear in any of the **previous utterances** and in a **relevant passage** in response to the training utterance, but not in the utterance itself.

 $q_1 : \dots : q_{k-1}$ 

Tell me about the neverending  
story film

 $q_k$ 

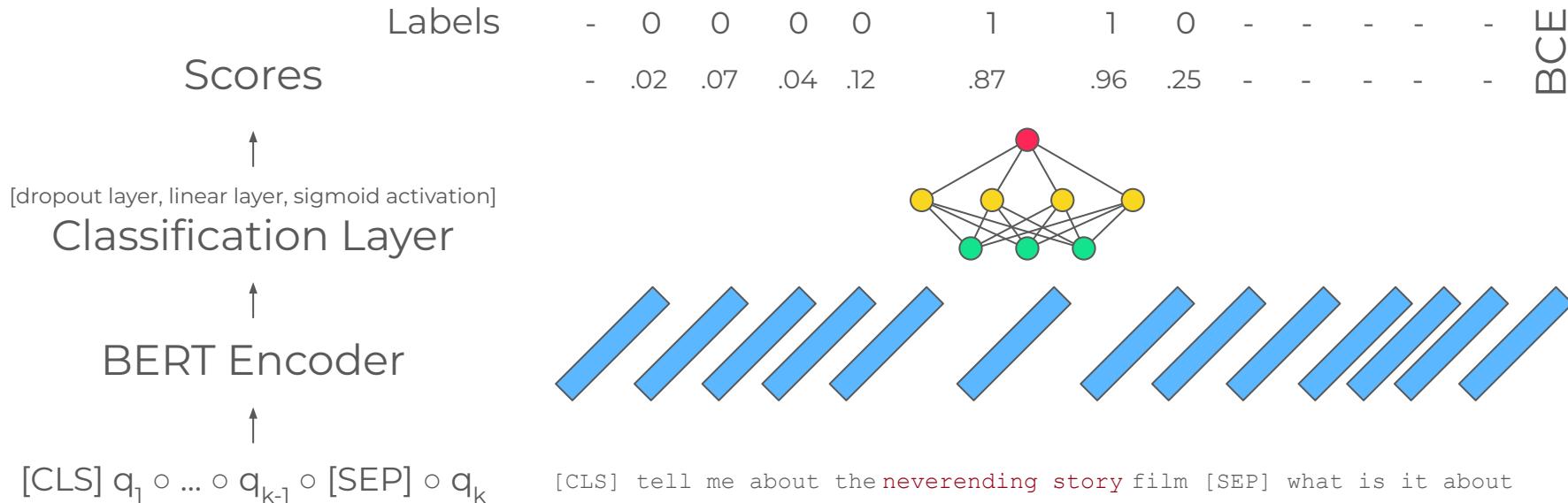
what is it about

relevant terms

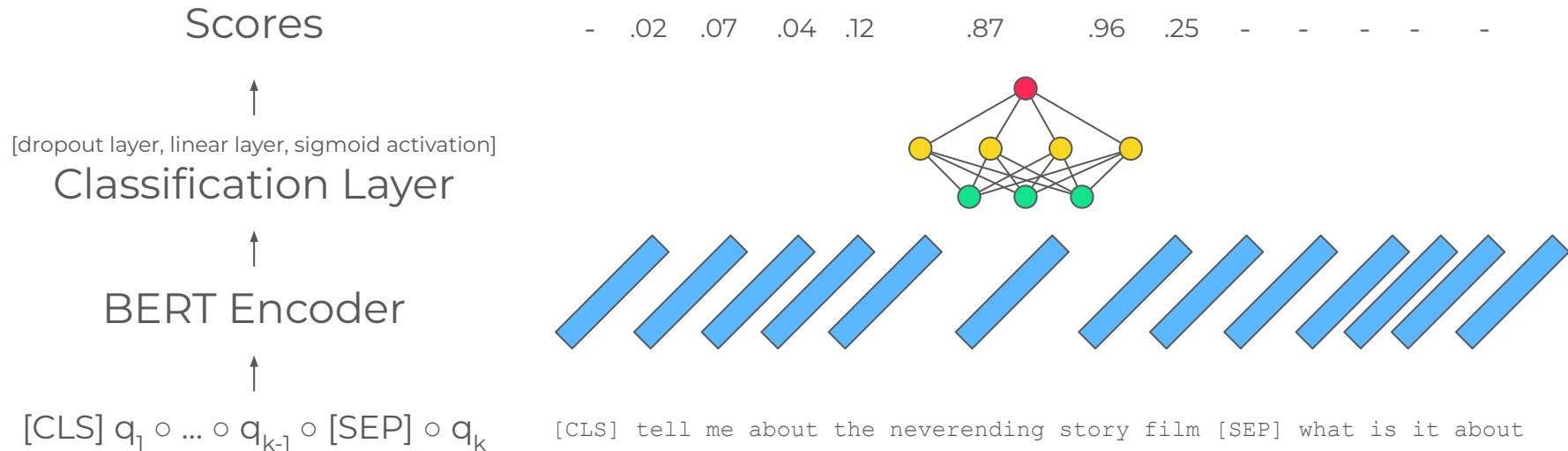
rel passage

The Neverending Story is about  
a boy who finds a magical book  
that tells of a young warrior ...

# QUERY EXPANSION: QuReTeC



# QUERY EXPANSION: QuReTeC





# QUERY REWRITING: CQR

Query Rewriting consists of **generating** a query from the utterance and the context.

A foundational approach is **Conversational Query Rewriting (CQR)**.

# QUERY REWRITING: CQR

CQR **fine-tunes a generative model** (GPT-2) to rewrite the utterances.

Given a transformer decoder function  $f$ ,  $w'_i$ , the  $i$ -th term of the rewritten query  $q'_k$  is:

$$w'_i = f(q_1 \circ [SEP] \circ \dots \circ [SEP] \circ q_k \circ [BOS] w'_1, \dots, w'_{i-1})$$

CQR uses **weak supervision** to fine-tune the model.

# QUERY REWRITING: CQR

Traditional search engines query logs contain a lot of search sessions.

In such sessions, **queries are related** but contain all the contextual information (i.e., **fully specified**).

To train CQR, we transform the fully specified queries of search sessions into conversational utterances as **weak supervision**.

Tell me about the neverending  
story film



what is the neverending story  
about

Tell me about the neverending  
story film

what is ~~it~~ the neverending story  
about



# QUERY REWRITING: CQR

CQR employs two strategies to generate the weak supervision:

## Rule based

Given a query, previously seen noun phrases are:

- omitted if after a **preposition (omission)**;
- otherwise, **replaced** with a **pronoun (coreference)**.

## Self-Learn

fine-tune a **GPT-2 model** to convert fully specified queries into conversational utterances.

Apply the model to MS MARCO sessions to generate conversations.



# QUESTION CLARIFICATION: TASKS

A should ask for clarification if the user's utterance is vague.

- Retrieve clarifying questions
  - Identify the best clarifying question from a pool
  - **PRO:** easy to evaluate; **CONS:** What happens in new domains?
- Generate clarifying questions
  - Use a generative model to create the clarifying question
  - **PRO:** should handle new domains; **CONS:** hard to evaluate
- Identify when a clarifying question is needed



# RETRIEVAL

---

The basic retrieval strategy **first rewrites the utterance** in a fully-specified query and **retrieves** the documents using **any retrieval approach**.

This approach is simple and effective but it does not allow **exploiting the conversation trajectory to guide the retrieval**.

We now discuss an end-to-end retrieval approach specifically designed to operate with conversational utterances.



# RETRIEVAL

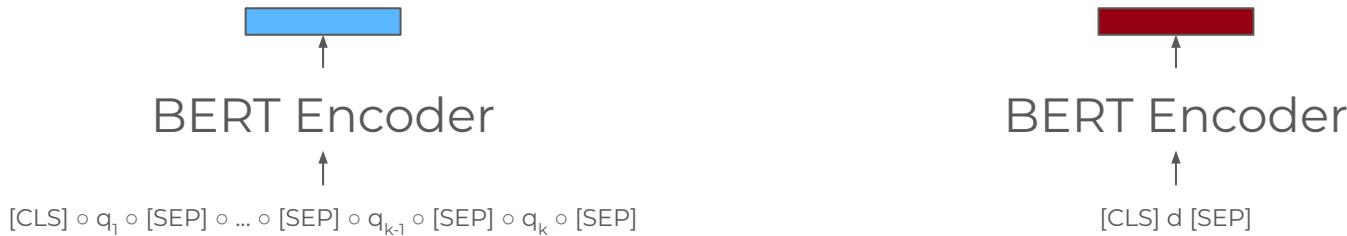
---

Conversational retrieval models mostly align with neural IR models:

- Dense approaches
  - Bi-encoder-based approaches:
    - ConvDR
    - CQE
  - Cross-encoder-based approaches:
    - ConvDR
  - Late-interaction-based approaches:
    - ZeCo<sup>2</sup>
- Sparse approaches:
  - CoSPLADE

# RETRIEVAL: ConvDR

ConvDR employs a BERT bi-encoder architecture to project utterances and documents in a latent space.



Given the k-th query  $q_k$ , a relevant and an irrelevant documents  $d^+$  and  $d^-$ , the BERT encoder is fine-tuned using negative log likelihood.

$$\mathcal{L}_{\text{Rank}} = -\log \frac{\exp(q'_k \cdot d^+)}{\exp(q'_k \cdot d^+) + \sum_{d^- \in D^-} \exp(q'_k \cdot d^-)}$$

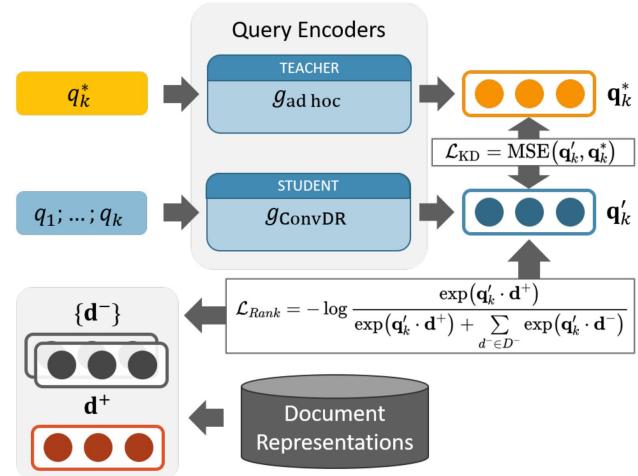
S. Yu, Z. Liu, C. Xiong, T. Feng, Z. Liu: Few-Shot Conversational Dense Retrieval. SIGIR 2021: 829-838

# RETRIEVAL: ConvDR

To address the limited supervision challenge, ConvDR training involves also a **teacher-student framework**.

Given the representation of an oracle manual rewritten query  $q_k^*$  according to a teacher model (ANCE), the **ConvDR** query encoder **is trained to mimic it**, producing similar representations.

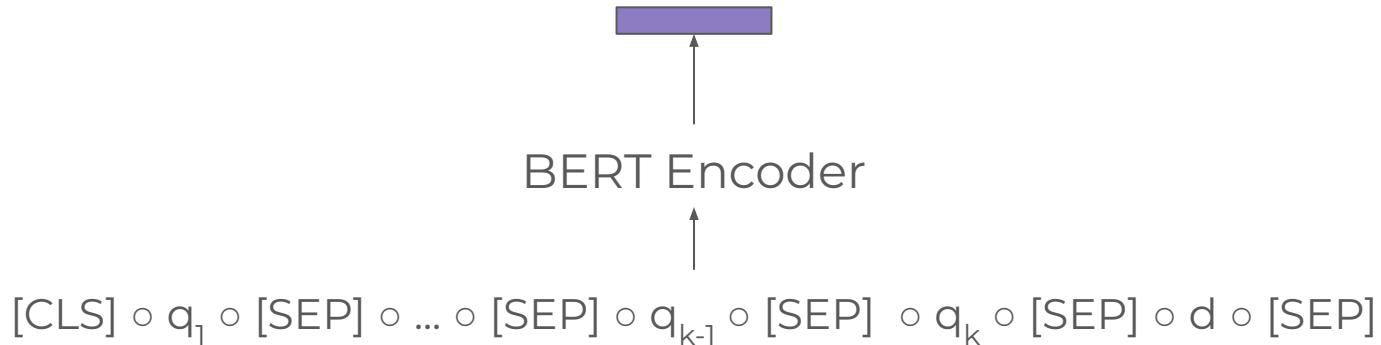
As document representation, ConvDR uses the teacher's one.





# RETRIEVAL: ConvDR RERANKING

ConvDR can be used also as reranker in a cross-encoder fashion:



In this case, the conversation's and document's representation is joint.

# ARE WE MISSING SOMEONE?

---





# EVALUATION!



# CaST 2019

---

**TREC CaST 2019** represents the first community effort in framing CS evaluation in the **Cranfield Paradigm**.

1. *Tell me about the Neverending Story film.*
  2. *What is **it** about?*
  3. *How was **it** received?*
  4. *Did **it** win any awards?*
  5. *Was **it** a book first?*
  6. *Who was **the author** and when what **it** published?*
  7. *What are **the main themes**?*
  8. *Who are **the main characters**?*
- ....

Conversations are simulated: how can we evaluate a system that would have lead to a different discourse with the user?



# CaST 2020 AND CaST 2021

---

**CaST 2020** built upon CaST 2019 introducing the concept of **Canonical Answer**.

**CaST 2021** further increased the conversation realism introducing two natural language constructs: **feedback** and **revealment**.

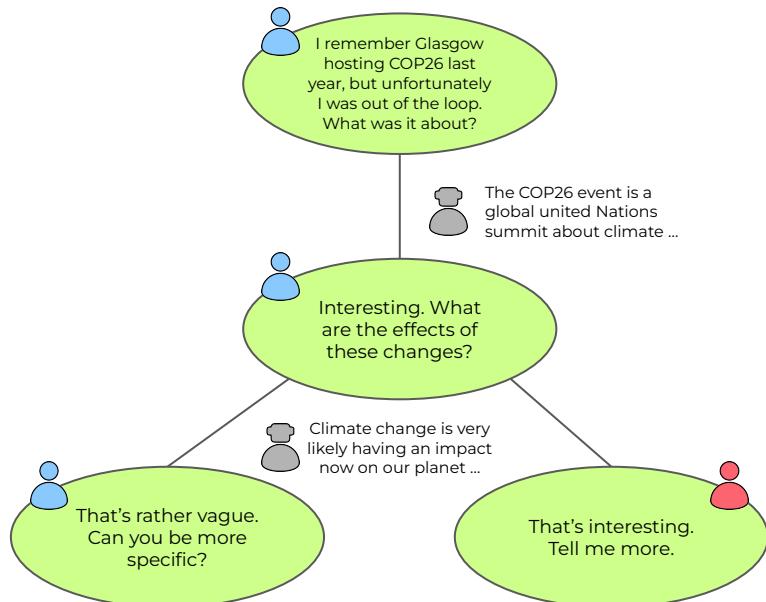
**Feedback:** the user expresses their opinion to steer the conversation.

**Revealment:** the user reveals information useful in subsequent turns.

# CaST 2022

TREC CaST 2022 used the **“dialogue tree”** as part of an evaluation campaign.

Additionally, for the first time, also **generated answers** and **mixed-initiative utterances** were considered.



P. Owoicho, J. Dalton, M. Aliannejadi, L. Azzopardi, J. R. Trippas, S. Vakulenko: TREC CAsT 2022: Going Beyond User Ask and System Retrieve with Initiative and Response Generation.



# iKAT

the interactive Knowledge Assistance Track (iKAT) moves forward and tries to operationalize the concept of “long-term” user characteristics, introducing the **Personal Text Knowledge Base** (PTKB)

```
"1": "I want to know about healthy cooking techniques.",  
"2": "I am lactose intolerant.",  
"3": "I'm looking for a speaker set to match my TV.",  
"4": "I'm willing to drive a long distance to find a cheaper TV.",  
"5": "I'm hoping to find some offers and discounts for TV.",  
"6": "I like to eat fruits and vegetables.",  
"7": "I don't read much.",  
"8": "I want to cook healthy and tasty recipes for my family.",  
"9": "I am on a diet and prefer low-calorie food.",  
"10": "I want to know about the nutritional value of the ingredients I use.",  
"11": "I'm looking for a new TV to replace my current one.",  
"12": "I want a TV that is okay for light and size of my living room."
```

iKAT considered three tasks:

- PTKB Statement Ranking
- Passage Ranking
- Response Generation

---

# WHAT IS CONVERSATIONAL RECOMMENDATION

# WHAT IS CONVERSATIONAL RECOMMENDATION

---

**Conversational Recommender Systems (CRS)** are software agents **that interactively support users in their decision-making processes**. The recommendation process is usually structured as a **multi-turn dialogue**, where the system tries to elicit the user's needs and preferences, makes recommendations, possibly provides explanations, and processes the users' feedback on the recommendations.

Ahtsham Manzoor, Dietmar Jannach, Conversational recommendation based on end-to-end learning: How far are we?, Computers in Human Behavior Reports, Volume 4, 2021.

# CONVERSATIONAL SEARCH VS RECOMMENDATION

---

**CS**

Conversational **Search**

**CRS**

Conversational **Recommender Systems**

# CONVERSATIONAL SEARCH VS RECOMMENDATION

## CS

### Conversational Search

- Based on “exploration”
- User asks → System responds (**often**)
- Open World
- Multiple goals
- No personalization (**currently**)

## CRS

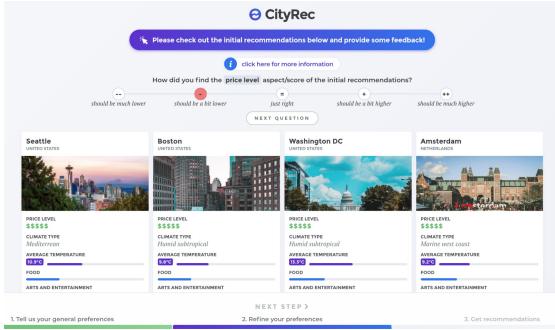
### Conversational Recommender Systems

- Based on “disambiguation”
- System ↔ User
- Closed World
- Single goal: recommend item(s)
- Personalization (**maybe**)



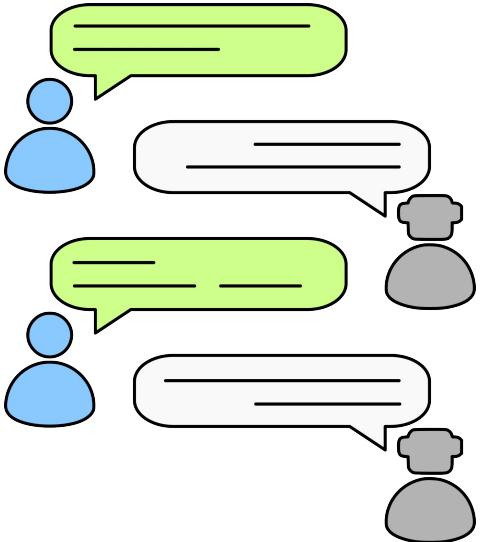
# CRS I/O TYPES

## Forms



The screenshot shows a user interface for a travel recommender system named "CityRec". At the top, there's a header with the logo and a button to "click here for more information". Below it, a question asks how the user found the price level aspect of the initial recommendations, with a scale from "should be much lower" to "should be much higher". There are four cards below, each representing a city: Seattle (United States), Boston (United States), Washington DC (United States), and Amsterdam (Netherlands). Each card displays the city name, country, price level (from \$1-\$5), climate type (Mediterranean), average temperature (15.0°C), food rating (4.0), and arts and entertainment rating (4.0). At the bottom, there are three navigation steps: "1. Tell us your general preferences", "2. Refine your preferences", and "3. Get recommendations".

## Dialogues



## Others

- Maps
- Gestures
- Writings

Dietz, L. W., Myftija, S., & Wörndl, W. (2019). Designing a Conversational Travel Recommender System Based on Data-Driven Destination Characterization. In RecTour@ RecSys (pp. 17-21).



# CRS APPROACHES

---

**Critique Based**

**Slot Filling**

**Natural Language**

# CRS APPROACHES

## Critique Based

## Slot Filling

## Natural Language

**CityRec**

Please check out the initial recommendations below and provide some feedback!

click here for more information

How did you find the **price level** aspect/score of the initial recommendations?

should be much lower      should be a bit lower      just right      should be a bit higher      should be much higher

[NEXT QUESTION](#)

**Seattle**  
UNITED STATES



**PRICE LEVEL**  
**\$\$\$\$\$**

**CLIMATE TYPE**  
Mediterranean

**AVERAGE TEMPERATURE**  
10.9°C

**FOOD**

**ARTS AND ENTERTAINMENT**

**Boston**  
UNITED STATES



**PRICE LEVEL**  
**\$\$\$\$\$**

**CLIMATE TYPE**  
Humid subtropical

**AVERAGE TEMPERATURE**  
9.8°C

**FOOD**

**ARTS AND ENTERTAINMENT**

**Washington DC**  
UNITED STATES



**PRICE LEVEL**  
**\$\$\$\$\$**

**CLIMATE TYPE**  
Humid subtropical

**AVERAGE TEMPERATURE**  
13.8°C

**FOOD**

**ARTS AND ENTERTAINMENT**

**Amsterdam**  
NETHERLANDS



**PRICE LEVEL**  
**\$\$\$\$\$**

**CLIMATE TYPE**  
Marine west coast

**AVERAGE TEMPERATURE**  
9.3°C

**FOOD**

**ARTS AND ENTERTAINMENT**

[NEXT STEP >](#)

1. Tell us your general preferences      2. Refine your preferences      3. Get recommendations

Dietz, L. W., Myftija, S., & Wörndl, W. (2019). Designing a Conversational Travel Recommender System Based on Data-Driven Destination Characterization. In RecTour@ RecSys (pp. 17-21).

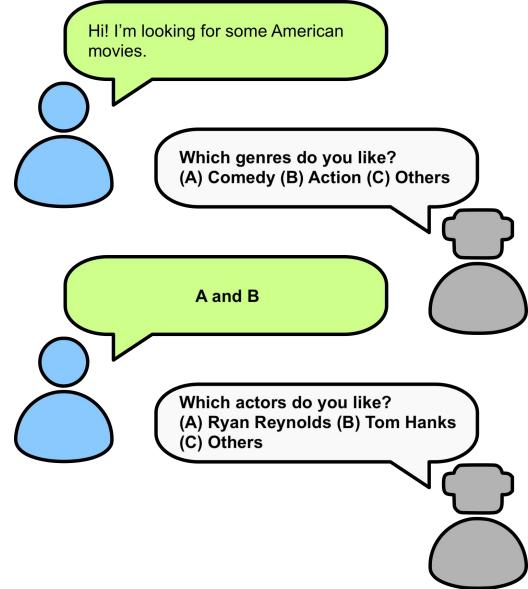


# CRS APPROACHES

Critique Based

Slot Filling

Natural Language



Y. Zhang, L. Wu, Q. Shen, Y. Pang, Z. Wei, F. Xu, B. Long, J. Pei: Multiple Choice Questions based Multi-Interest Policy Learning for Conversational Recommendation. WWW 2022: 2153-2162

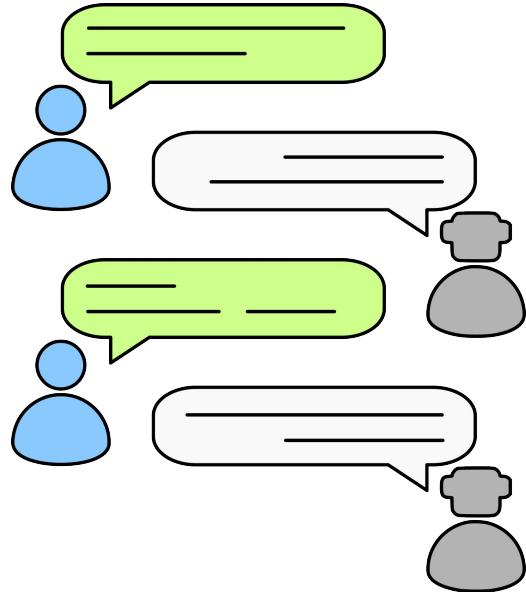


# CRS APPROACHES

Critique Based

Slot Filling

Natural Language



Chen, Q., Lin, J., Zhang, Y., Ding, M., Cen, Y., Yang, H., & Tang, J. (2019). Towards knowledge-based recommender dialog system.



# CRS ARCHITECTURES

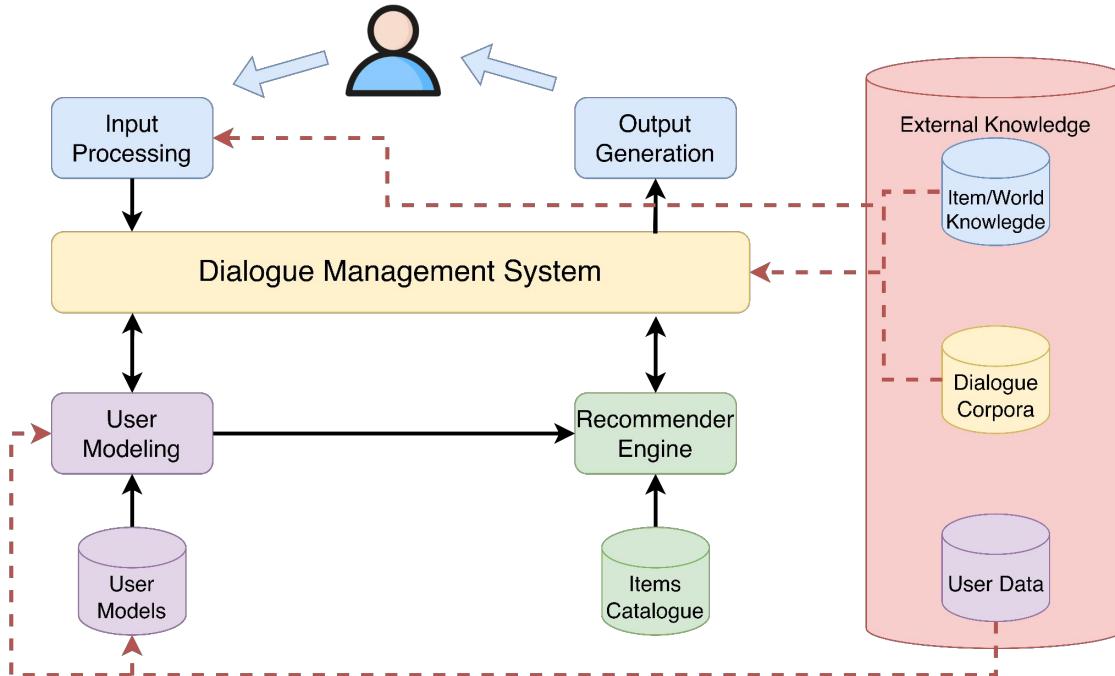
---

**Modular**

**Unified**

# CRS ARCHITECTURES

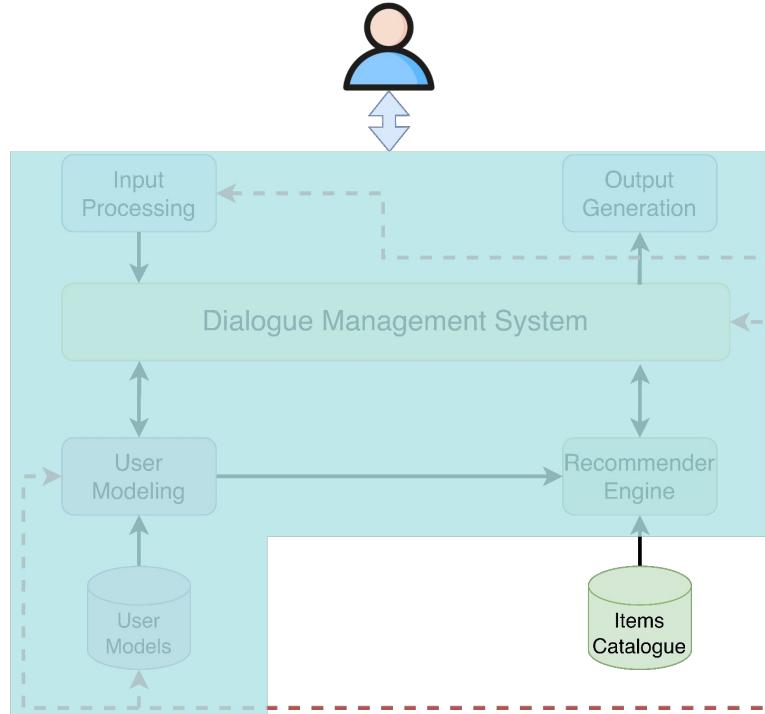
## Modular



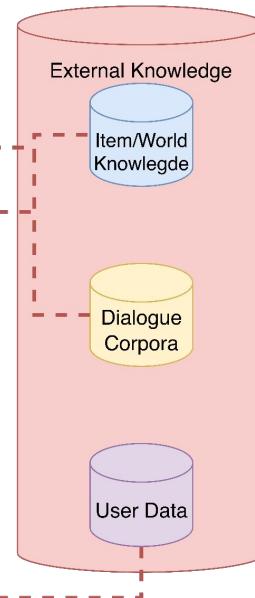
## Unified

# CRS ARCHITECTURES

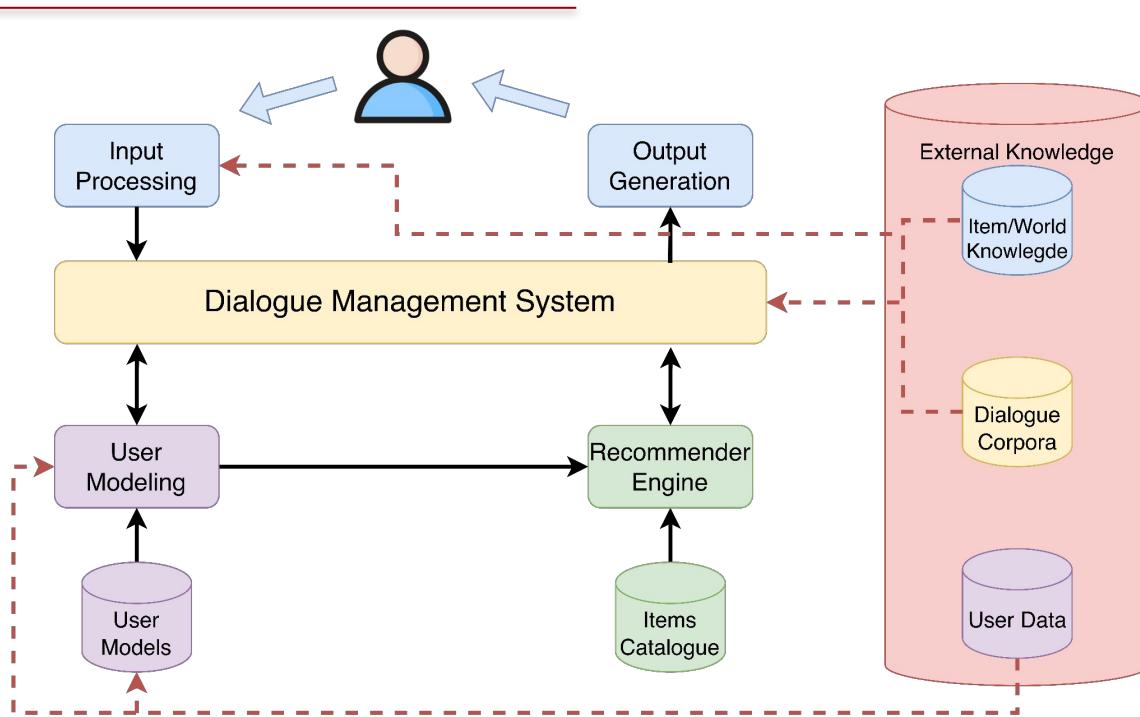
**Modular**



**Unified**



# MODULAR CRS



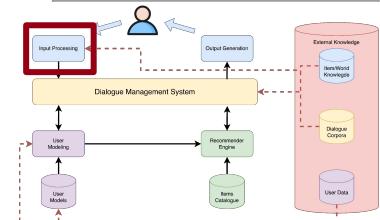
Jannach, D., Manzoor, A., Cai, W., & Chen, L. (2021). A survey on conversational recommender systems. ACM Computing Surveys (CSUR), 54(5), 1-36. Thompson, C. A., Goker, M. H., & Langley, P. (2004). A personalized system for conversational recommendations. Journal of Artificial Intelligence Research, 21, 393-428.

# INPUT PROCESSING

Processing the input includes:

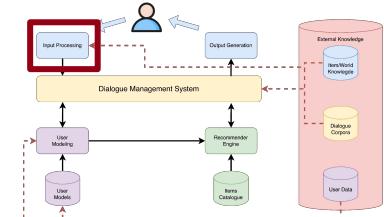
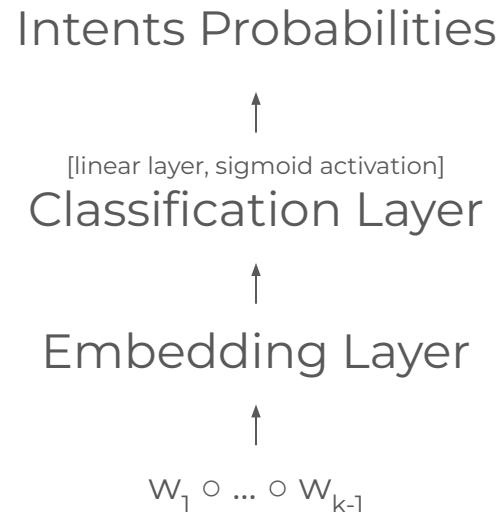
- Intent Detection
  - Can be seen as a classification task
- Named Entity Recognition and Entity Linking
  - Especially if external knowledge bases are used
- Sentiment Analysis

Intent Types
initiate, quit, restart
chit-chat
request for explanation
preference
recommendation
feedbacks



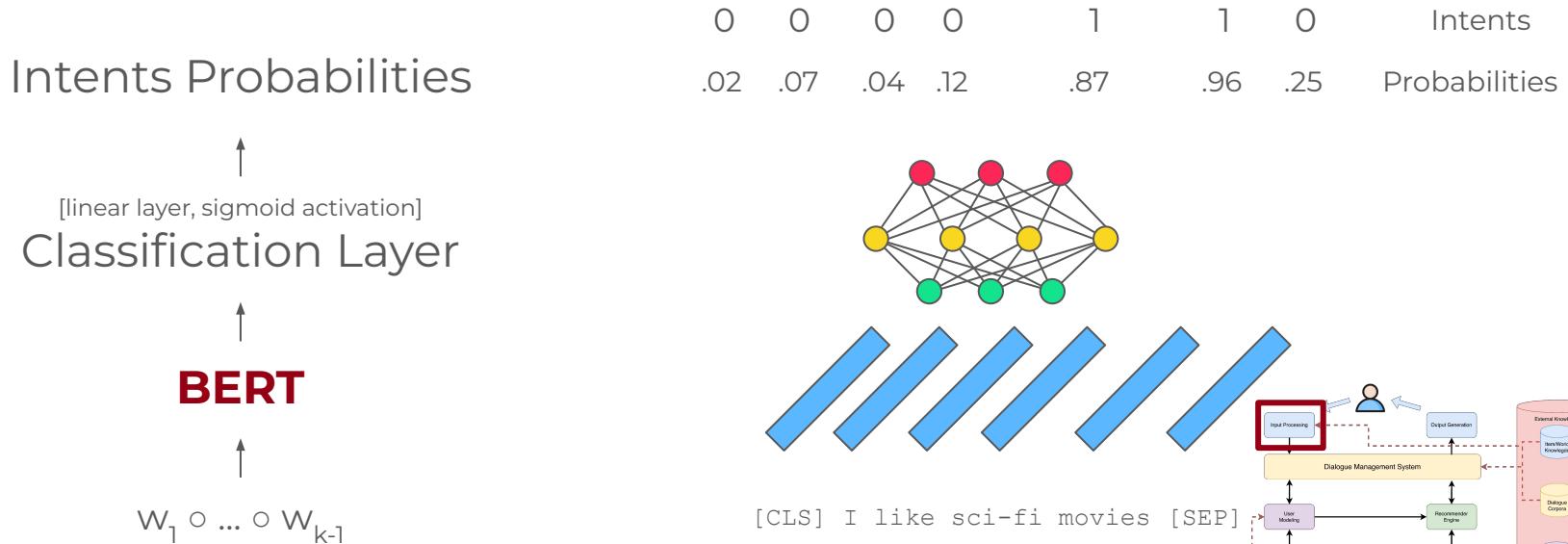
# INTENT DETECTION

Possible approach for intent detection:



# INTENT DETECTION

Possible approach for intent detection:



Farshidi, S., Rezaee, K., Mazaheri, S., Rahimi, A. H., Dadashzadeh, A., Ziabakhsh, M., ... & Jansen, S. (2024). Understanding user intent modeling for conversational recommender systems: a systematic literature review. *User Modeling and User-Adapted Interaction*, 1-64.

Moradizeyeh, S. (2022). Intent recognition in conversational recommender systems.

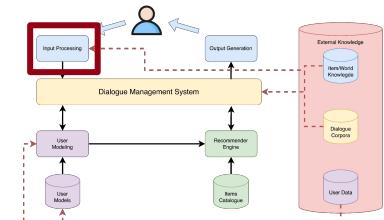
# NAMED ENTITY RECOGNITION

Possible approaches for NER:



Dingwall, N., & Gao, V. R. (2021). Enhancing Gazetteers for Named Entity Recognition in Conversational Recommender Systems (Long paper). In Joint Workshop Proceedings of the 3rd Edition of Knowledge-aware and Conversational Recommender Systems (KaRS) and the 5th Edition of Recommendation in Complex Environments (ComplexRec) co-located with 15th ACM Conference on Recommender Systems (RecSys 2021)

Iovine, A., Narducci, F., de Gemmis, M., Polignano, M., Basile, P., & Semeraro, G. (2020, September). A Comparison of Services for Intent and Entity Recognition for Conversational Recommender Systems. In IntRS@ RecSys (pp. 37-47).

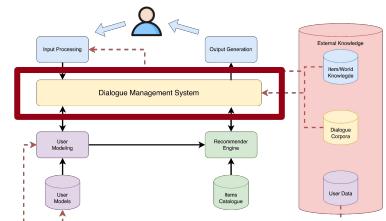




# DIALOGUE MANAGEMENT SYSTEM

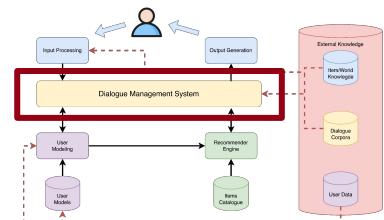
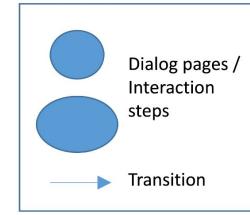
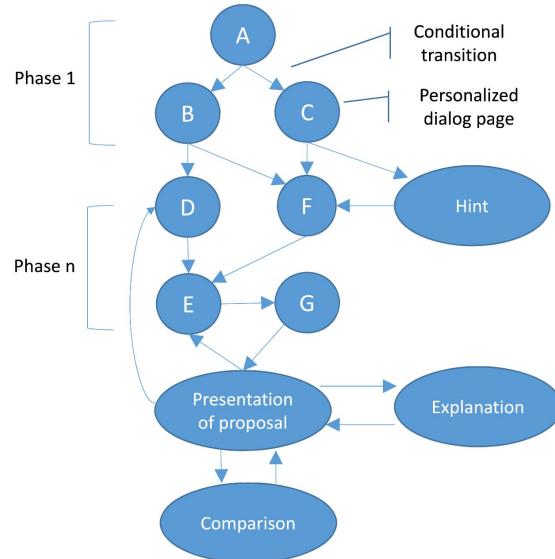
The **Dialogue Management System** is the core “controller”:

- Determines the **action** to be taken based on the intent(s)
- Keeps and updates dialogue **state**
- Exploits the external knowledge to enhance the recommendation



# STATE TRACKING

States are usually modelled through **finite state automata**:

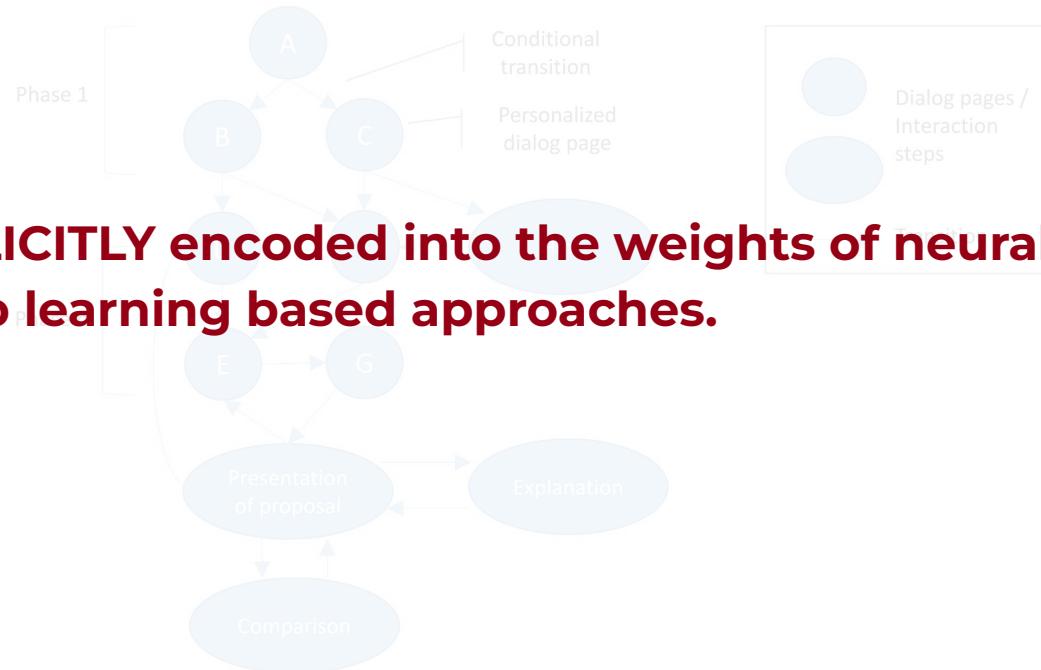


Jannach, D., Manzoor, A., Cai, W., & Chen, L. (2021). A survey on conversational recommender systems. ACM Computing Surveys (CSUR), 54(5), 1-36.

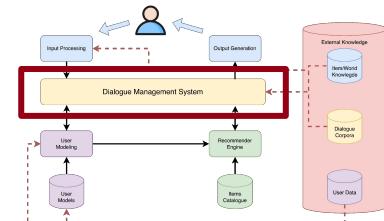


# STATE TRACKING

States are usually modelled through **finite state automata**:



This is **IMPLICITLY** encoded into the weights of neural networks in recent deep learning based approaches.

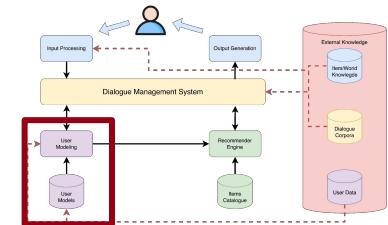


# USER MODELING

User Modeling strategies:

**Slot Filling** typical of critique & form based

1. **Item Attributes** preferences: e.g. product price, film genre, ...



# USER MODELING

User Modeling strategies:

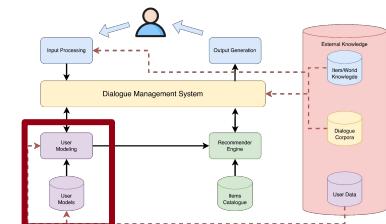
**Slot Filling** typical of critique & form based

1. **Item Attributes** preferences: e.g. product price, film genre, ...

Often used in deep learning approaches

2. **Items** preferences:

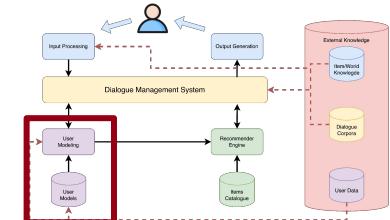
- a. Explicit feedbacks: ratings, reviews (like/dislike statements), ...
- b. Implicit feedbacks: purchases, clicks, ...



# USER MODELING

CRS can consider:

- **Short-term** user preferences (within the conversation).
  - **Cold-start** at the beginning of the conversation.
  - Most popular approach
- **Long-term** user preferences
  - Need of external data sources containing **user data**.



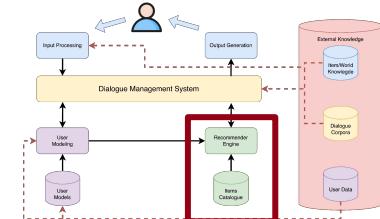


# RECOMMENDER ENGINE

The recommender engine computes the recommendations.

Following a **block based approach** it could be:

- **Matrix Factorization** like approaches
- Topic-sensitive **PageRank** like approaches
- Every other recommendation approach
- OR: Information Retrieval based approaches



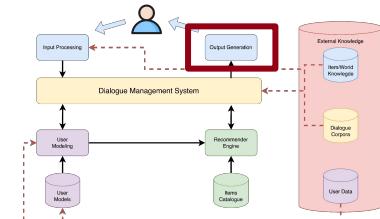


# OUTPUT GENERATION

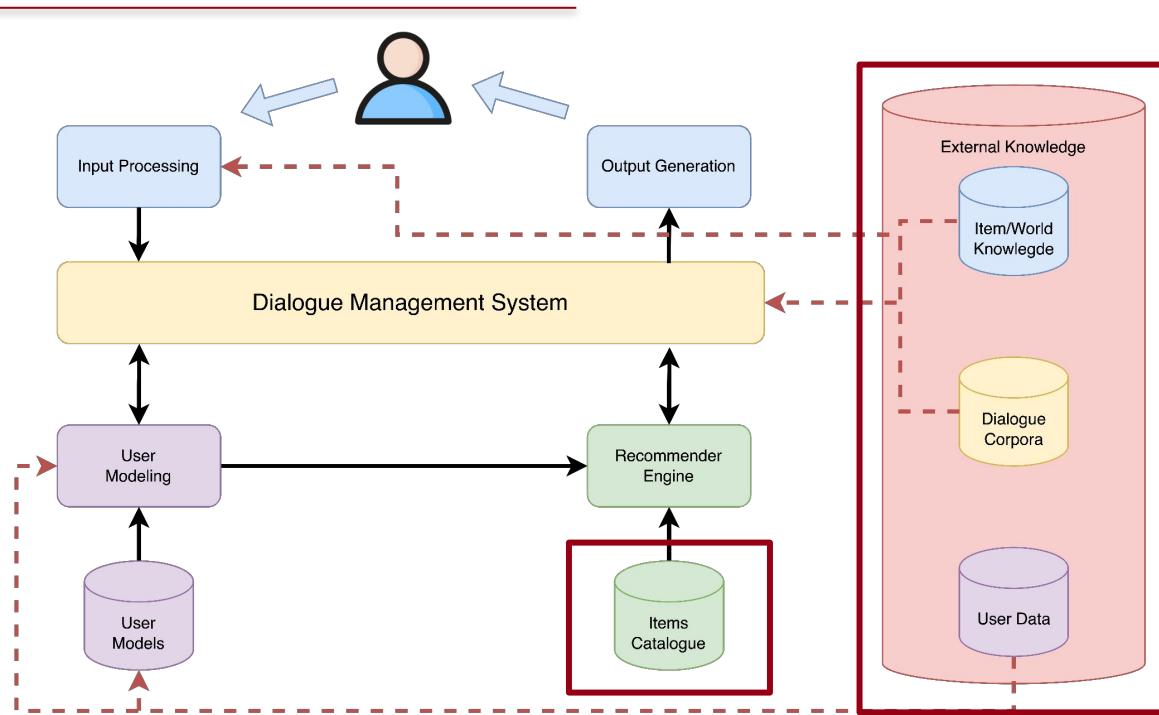
The output generated can be:

- Recommended items
- Questions for the user
- Additional information about the recommended items

The output format depends on the system's I\O type.



# DATA



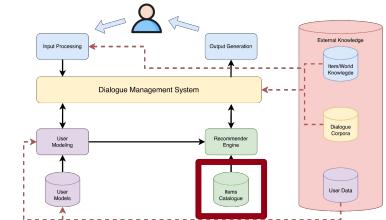
# DATA

---

## Items Catalogue:

- Set of recommendable items
- May correspond to traditional recommendation datasets

Dataset	Domain
MovieLens	Movies
AmazonReviews	Products
NetflixMovieRatings	Movies

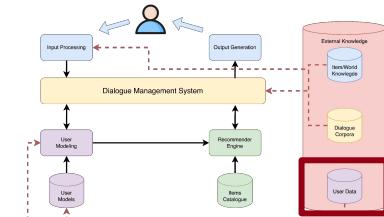




# DATA

## User Data (optional):

- Data containing implicit/explicit user feedbacks
- Used to model user (long-term) preferences
- Usually provided together with the item catalogue



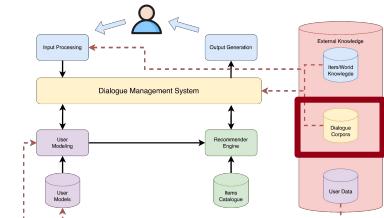
# DATA

---

## Dialogue Corpora (optional):

- Set of dialogues
- Can be used to train intent detection / sentiment analysis algorithms  
(If appropriately labelled)

Corpora	Domain
ReDial	Movies
MultiWOZ	Travel
OpenDialKG	Movies, Books, Sports, Music



# DATA

---

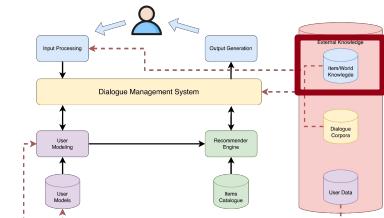
## Item/World Knowledge (optional):

- Usually represented by Knowledge Graphs

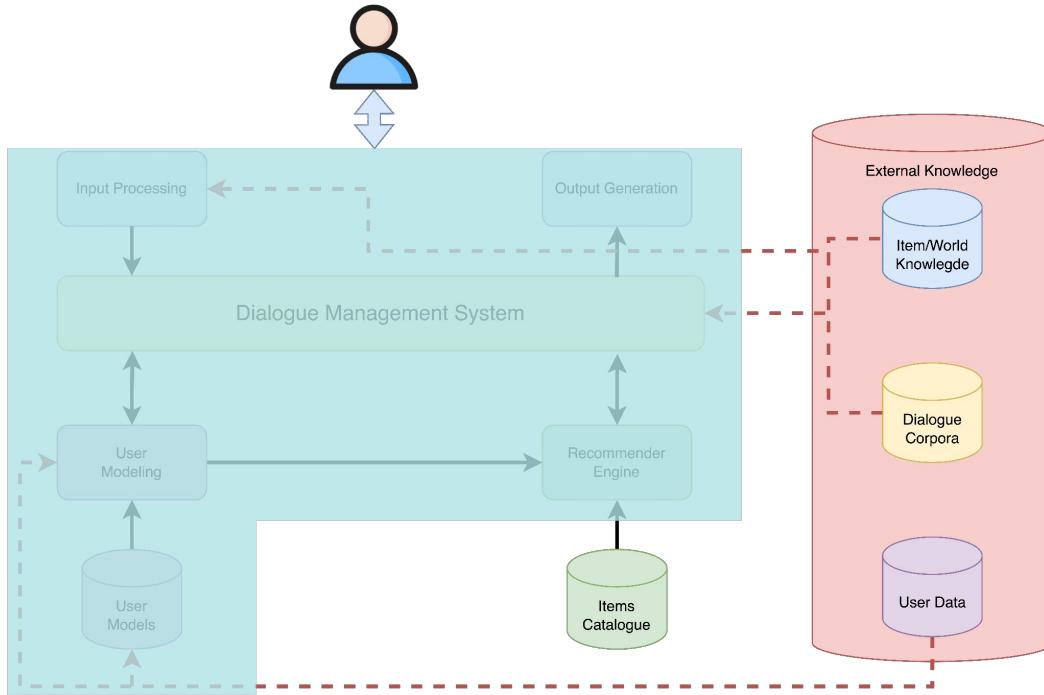
Useful for:

- Entity Recognition (“Entity Resolution”)
- Providing additional information in the recommendation process

Corpora	Domain
Wikidata	Multiple
DBpedia	Multiple
IMDb	Movies



# UNIFIED CRS





# UNIFIED CRS

---

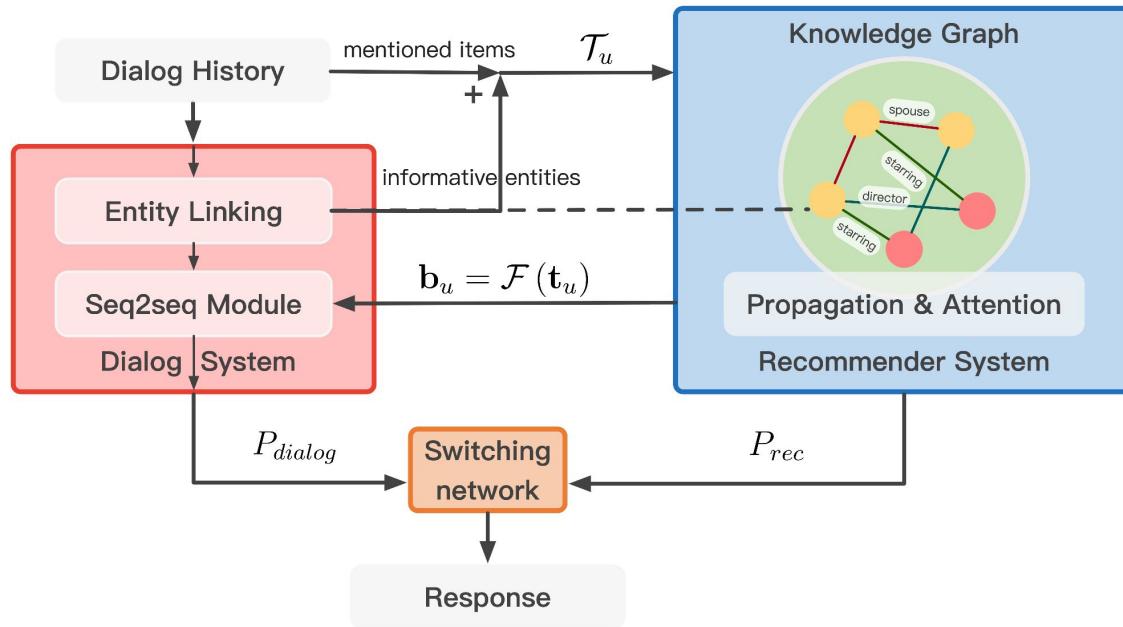
When modeling CRS with end-to-end trained models:

- Intent Detection / NER / Sentiment Analysis → **implicit**
- State Tracking → **implicit**
- Dialogue corpora (optional in modular CRS)



**Fundamental to train unified CRS systems.**

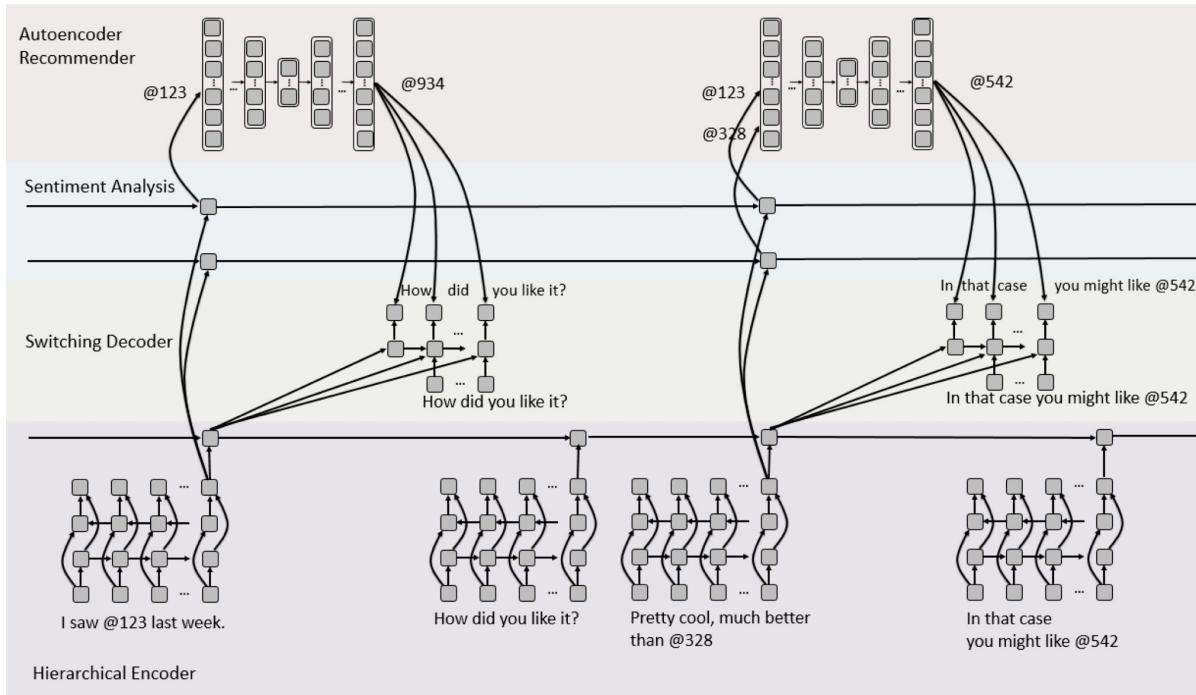
# UNIFIED CRS: KBRD



Chen, Q., Lin, J., Zhang, Y., Ding, M., Cen, Y., Yang, H., & Tang, J. (2019). Towards knowledge-based recommender dialog system.



# UNIFIED/HYBRID CRS



Same approach that introduced **ReDial**.

Li, R., Ebrahimi Kahou, S., Schulz, H., Michalski, V., Charlin, L., & Pal, C. (2018). Towards deep conversational recommendations. Advances in neural information processing systems, 31.

---

# THE DAGSTUHL CAFE FRAMEWORK FOR THE EVALUATION OF CONVERSATIONAL AGENTS

Bauer, C., Chen, L., Ferro, N., Fuhr, N., Anand, A., Breuer, T., Faggioli, G., Frieder, O., Joho, H., Karlgren, J., Kiesel, J., Knijnenburg, B. P., Lipani, A., Michiels, L., Papenmeier, A., Pera, M. S., Sanderson, M., Sanner, S., Stein, B., Trippas, J. R., Verspoor, K., and Willemsen, M. C. (2025). Manifesto from Dagstuhl Perspectives Workshop 24352 – Conversational Agents: A Framework for Evaluation (CAFE). *Dagstuhl Manifestos, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Germany.*

# CONVERSATIONAL AGENTS: A FRAMEWORK FOR EVALUATION (CAFE)



Jointly organized by

- Christine Bauer
- Li Chen
- Nicola Ferro
- Norbert Fuhr

# NOT ONLY SCIENCE BUT ALSO FUN



# A “SIMPLISTIC” EXAMPLE



I would like to buy a new swimming suit

Recommendation



I don't really know. What are the advantages of each material?



Search and Generate

Elastane has excellent UV-protective quality but it is not very breathable.



Although Polyester is not as elastic as elastane, swimwear made of it is highly durable.



Based on your previous purchases, you might prefer something more durable: why not going for polyester?

Personalized Recommendation



You are right! Which do you think are better, briefs or shorts?



What do you mean by briefs?



Sorry, I'm not a native speaker, I meant speedo



Looking at your preferences, you might like shorts more. Here few examples



Thank you! I like this one



Clarification

Personalized Recommendation

# A “SIMPLISTIC” EXAMPLE: GPT-3 VERSION

I would like to buy a new swimming suit

That's great! What kind of swimming suit are you looking for? There are many options depending on your preferences, such as:

- **Style:** One-piece, bikini, tankini, high-waisted, etc.
- **Purpose:** Casual swimming, athletic training, or beach lounging.
- **Fabric:** Some offer UV protection, while others are designed for durability in chlorinated water.
- **Fit:** Do you have a preference for more coverage or a certain cut (high-leg, low-back, etc.)?

Let me know what you're looking for, and I can help you narrow down the choices!



# A “SIMPLISTIC” EXAMPLE: GPT-3 VERSION

I don't really know. What are the advantages of each material?

Different swimsuit materials offer unique benefits depending on your needs. Here's a breakdown of common swimsuit fabrics and their advantages:

**1. Nylon/Polyamide**

- Advantages:**
  - Lightweight & Quick-drying:** Dries quickly, making it ideal for swimwear.
  - Comfortable Stretch:** Nylon is stretchy, providing a comfortable fit that moves with your body.
  - Smooth Feel:** It's soft and smooth against the skin.
- Cons:**
  - Tends to lose its shape or fade over time with prolonged exposure to chlorine and sunlight.

**2. Polyester**

- Advantages:**
  - Chlorine Resistance:** Polyester is very durable and maintains its shape and color even after multiple washes, making it great for frequent pool use.
  - UV Resistance:** Some polyester blends offer UV protection, which is good for outdoor swimming.
  - Durable:** Less likely to pill or fade over time.
- Cons:**
  - Less stretchy compared to nylon, so it may not fit as snugly.

**3. Lycra/Spandex/Elastane**

- Advantages:**
  - Super Stretchy:** Lycra or Spandex offers excellent elasticity, providing a snug and flexible fit.
  - Comfortable & Flexible:** Great for athletic swimsuits where mobility is key.
- Cons:**
  - Susceptible to chlorine and may lose elasticity over time without proper care.

**4. PBT (Polybutylene Terephthalate)**

- Advantages:**
  - High Chlorine Resistance:** Often used in competitive swimwear, PBT is one of the most durable options for chlorinated pools.
  - Quick-dry & Lightweight:** Stays light in the water and dries fast.
  - Resistant to Stretching:** Maintains its shape after repeated use.
- Cons:**
  - Less comfortable compared to fabrics with more stretch, often feels stiff.

**5. Neoprene**

- Advantages:**
  - Thermal Insulation:** Neoprene suits are used for wetsuits or swimwear designed for cold water as they provide warmth.
  - Buoyancy:** Helps with floating, making it ideal for water sports.
- Cons:**
  - Not as breathable and can feel heavier in warmer conditions.

**6. Cotton Blends**

- Advantages:**
  - Comfortable:** Cotton blends feel soft and natural on the skin.
  - Stylish:** Often used in fashion swimwear, though not ideal for performance.
- Cons:**
  - Absorbs water and takes longer to dry, making it heavier when wet and less ideal for swimming.

**Summary:**

- For casual swimming or beachwear, nylon/spandex blends offer comfort and style.
- For frequent pool use, polyester or PBT is ideal due to its chlorine resistance and durability.
- For water sports or cold water, consider neoprene.
- For flexibility and comfort, look for swimsuits with a higher percentage of Lycra/Spandex.

Do any of these stand out to you?

I don't really know. What are the advantages of each material?

Different swimsuit materials offer unique benefits depending on your needs. Here's a breakdown of common swimsuit fabrics and their advantages:

**1. Nylon/Polyamide**

- Advantages:**
  - Lightweight & Quick-drying:** Dries quickly, making it ideal for swimwear.
  - Comfortable Stretch:** Nylon is stretchy, providing a comfortable fit that moves with your body.
  - Smooth Feel:** It's soft and smooth against the skin.
- Cons:**
  - Tends to lose its shape or fade over time with prolonged exposure to chlorine and sunlight.

**2. Polyester**

What fabric lasts longest? Best for beachwear?

- Advantages:**
  - Chlorine Resistance:** Polyester is very durable and maintains its shape and color even after multiple washes, making it great for frequent pool use.
  - UV Resistance:** Some polyester blends offer UV protection, which is good for outdoor swimming.
  - Durable:** Less likely to pill or fade over time.
- Cons:**
  - Less stretchy compared to nylon, so it may not fit as snugly.

**3. Lycra/Spandex/Elastane**

- Advantages:**
  - Super Stretchy:** Lycra or Spandex offers excellent elasticity, providing a snug and flexible fit.
  - Comfortable & Flexible:** Great for athletic swimsuits where mobility is key.
- Cons:**
  - Susceptible to chlorine and may lose elasticity over time without proper care.

**4. PBT (Polybutylene Terephthalate)**

- Advantages:**
  - High Chlorine Resistance:** Often used in competitive swimwear, PBT is one of the most durable options for chlorinated pools.
  - Quick-dry & Lightweight:** Stays light in the water and dries fast.
  - Resistant to Stretching:** Maintains its shape after repeated use.
- Cons:**
  - Less comfortable compared to fabrics with more stretch, often feels stiff.

**5. Neoprene**

- Advantages:**
  - Thermal Insulation:** Neoprene suits are used for wetsuits or swimwear designed for cold water as they provide warmth.
  - Buoyancy:** Helps with floating, making it ideal for water sports.
- Cons:**
  - Not as breathable and can feel heavier in warmer conditions.

**6. Cotton Blends**

- Advantages:**
  - Comfortable:** Cotton blends feel soft and natural on the skin.
  - Stylish:** Often used in fashion swimwear, though not ideal for performance.
- Cons:**
  - Absorbs water and takes longer to dry, making it heavier when wet and less ideal for swimming.

**Summary:**

- For casual swimming or beachwear, nylon/spandex blends offer comfort and style.
- For frequent pool use, polyester or PBT is ideal due to its chlorine resistance and durability.
- For water sports or cold water, consider neoprene.
- For flexibility and comfort, look for swimsuits with a higher percentage of Lycra/Spandex.



# A “SIMPLISTIC” EXAMPLE: GPT-3 VERSION

You are right! Which do you think are better, briefs or shorts?

- ⓘ The choice between **briefs** and **shorts** for swimwear largely depends on your comfort, style preference, and intended use. Here's a comparison to help you decide:

## Briefs

- **Fit:** Tighter, more minimal coverage (resembling traditional men's underwear).
- **Mobility:** Offers maximum freedom of movement, especially for your legs. This makes them popular in competitive swimming or serious training.
- **Hydrodynamics:** The snug fit and minimal fabric create less drag in the water, improving speed.
- **Comfort:** Ideal for those who prefer a lightweight feel and don't want extra fabric while swimming.
- **Tan Lines:** Smaller coverage means fewer tan lines, which may appeal to some.
- **Style:** More revealing, offering a sleek, athletic look.
- **Best For:** Competitive swimmers, frequent pool users, or those who prioritize speed and freedom of movement.

## Shorts (Swim Trunks)

- **Fit:** More relaxed than briefs, providing moderate to full coverage down the thighs.
- **Versatility:** Can be worn not only for swimming but also for beach activities, volleyball, or lounging.
- **Comfort:** The looser fit can be more comfortable for all-day wear, especially outside the pool.
- **Style:** Available in a wider range of styles, lengths, and designs (e.g., board shorts, classic trunks), making them more versatile in terms of fashion.
- **Functionality:** Some shorts have inner liners for added support, though they can be less form-fitting than briefs.
- **Drag:** They create more drag in the water due to the extra fabric, which can slightly reduce speed for competitive swimming.
- **Best For:** Casual swimmers, beachgoers, or those looking for a more versatile swimwear option that can be worn beyond the pool.

## Quick Summary:

- **Briefs:** Better for competitive swimming, speed, and athleticism; more revealing and minimal.
- **Shorts:** Great for casual wear, more versatile, and stylish; provide more coverage but can create drag in the water.

Do you prefer a more functional, competitive option or something that gives you a bit more coverage and versatility?



# A “SIMPLISTIC” EXAMPLE: GPT-3 VERSION

Sorry, I'm not a native speaker, I meant speedo

No worries! When you say "speedo," you're likely referring to the classic **brief-style swimwear** made popular by the Speedo brand. These are tight-fitting and minimalist in design, often used by competitive swimmers. So now that we're talking specifically about **speedo-style briefs** versus **swim shorts**, here's the comparison again but with the right context!

#### Speedo (Brief-Style Swimwear)

- **Fit:** Snug, tight fit around the waist and hips with minimal fabric, offering **freedom of movement** for the legs.
- **Hydrodynamics:** Less drag in the water compared to swim shorts, making them ideal for competitive swimming or faster lap times.
- **Comfort:** Lightweight and dries quickly; ideal for **pool use** or serious swimming where you want to feel unrestricted.
- **Style:** Reveals more of the body, which can be great if you're confident in that style or prefer something sleek.
- **Mobility:** Allows maximum mobility, especially for your thighs and hips, making it popular for athletes and competitive swimmers.
- **Tan Lines:** Minimal coverage means fewer tan lines, if that's something you care about.
- **Best For:** Athletes, competitive swimmers, or those who prioritize speed and performance in the water.

#### Swim Shorts (Swim Trunks)

- **Fit:** Looser fit with more coverage, typically extending to the mid-thigh or knee.
- **Versatility:** More versatile for beach use, casual swimming, or outdoor activities. They can be worn for activities other than swimming.
- **Comfort:** The looser fit can feel more comfortable for all-day wear, especially when you're outside of the water.
- **Style:** Generally seen as more modest, covering more of the thighs and often featuring a drawstring for adjustment. Comes in a variety of styles and designs.
- **Drag:** The extra fabric creates more drag in the water, so they aren't as efficient for competitive swimming or fast laps.
- **Functionality:** Good for activities like beach sports, lounging, or general water play.
- **Best For:** Casual swimmers, those who want more coverage, or anyone who's looking for versatility in their swimwear.

#### Which to Choose?

- **Speedo:** If you're swimming for performance, training, or you prefer the sleek, minimal style.
- **Swim Shorts:** If you want more coverage, comfort outside of swimming, or a more versatile option for beach or casual use.

It really comes down to your preference for style and how you plan to use the swimwear! Which style sounds more like what you're after?

# A “SIMPLISTIC” EXAMPLE: GPT-4 VERSION



You

I would like to buy a new swimming suit



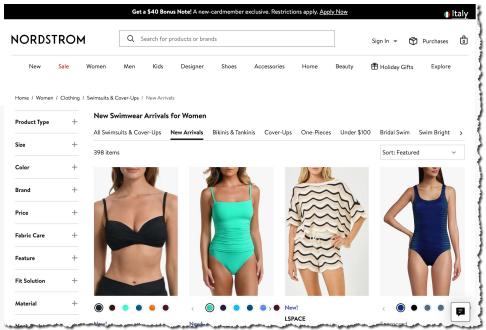
Copilot

Great! There are plenty of options available online. Here are a few places where you can find a variety of swimsuits:

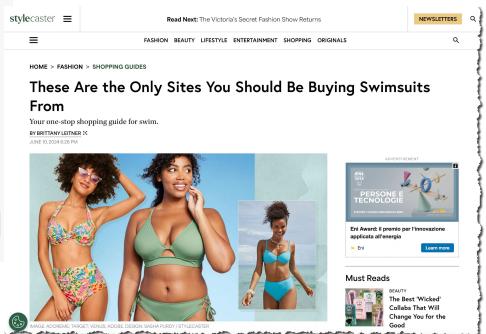
1. **Nordstrom**: They have a wide range of new arrivals, including bikinis, one-pieces, and cover-ups [1](#) [2](#).
2. **StyleCaster**: This site lists some of the best places to buy swimsuits online, catering to different budgets [3](#).
3. **DICK'S Sporting Goods**: They offer a variety of swimsuits from top brands like Calia, Roxy, and Nike [4](#) [5](#).

Do you have a specific style or brand in mind?

Learn more [1](#)  [nordstrom.com](#) [2](#)  [nordstrom.com](#) [3](#)  [stylecaster.com](#) [4](#)  [dickssportinggoods.com](#) [5](#)  [dickssportinggoods.com](#)



The screenshot shows the Nordstrom website with a search bar at the top. Below it, a navigation menu includes 'New', 'Sale', 'Women', 'Men', 'Kids', 'Designer', 'Shoes', 'Accessories', 'Home', 'Beauty', 'Holiday Gifts', and 'Explore'. A banner for 'New Swimwear Arrivals for Women' is displayed, featuring several models in different styles of swimsuits. On the left, there is a sidebar with filters for 'Product Type', 'Size', 'Color', 'Brand', 'Price', 'Fabric Care', 'Feature', 'Fit Solution', and 'Material'. The main content area shows a grid of swimwear products.

The screenshot shows the StyleCaster website under the 'FASHION > SHOPPING GUIDES' category. The main headline is 'These Are the Only Sites You Should Be Buying Swimsuits From'. It features a photo of two women in bikinis. To the right, there is a sidebar with sections for 'PERSONAL TECNOLOGIE' and 'Must Reads'.

# A “SIMPLISTIC” EXAMPLE: GPT-4 VERSION

I don't really know. What are the advantages of each material?

 Copilot

Choosing the right swimsuit material can make a big difference in comfort, durability, and performance. Here are some common materials used in swimsuits and their advantages:

1. Polyester:
  - **Durability:** Highly resistant to chlorine and UV rays, making it long-lasting [①](#).
  - **Colorfastness:** Retains color well even after multiple uses [②](#).
  - **Less Stretchy:** Not as stretchy as other materials, but often blended with spandex for better fit [③](#).
2. Nylon:
  - **Soft and Silky:** Known for its smooth, comfortable feel [④](#).
  - **Quick-Drying:** Dries faster than many other materials [⑤](#).
  - **Stretchy:** Offers good elasticity, but can lose shape over time with exposure to chlorine [⑥](#).
3. Spandex (Lycra/Elastane):
  - **High Stretchability:** Can stretch up to 600% of its original length, providing excellent flexibility and fit [⑦](#).
  - **Comfort:** Ensures a snug yet comfortable fit, allowing for freedom of movement [⑧](#).
  - **Less Durable:** Can degrade quickly when exposed to chlorine, so often blended with other materials [⑨](#).
4. Neoprene:
  - **Insulation:** Excellent for keeping warm in cold water, making it ideal for wetsuits [⑩](#).
  - **Durability:** Resistant to water and moisture, ensuring longevity [⑪](#).

Each material has its own strengths, so the best choice depends on your specific needs and preferences. Do you have any particular activities or conditions in mind for your new swimsuit?

Learn more [①](#)  [knowingfabric.com](#) [②](#)  [openwaterhq.com](#)

# A “SIMPLISTIC” EXAMPLE: GPT-4 VERSION

 You

You are right! Which do you think are better, briefs or shorts?

 Copilot

Both briefs and shorts have their own advantages, so it really depends on your personal preferences and intended use. Here's a quick comparison to help you decide:

**Briefs (Speedos)**

- **Fit:** Tight and body-hugging, providing minimal coverage [1](#).
- **Mobility:** Excellent for swimming as they offer less drag in the water [1](#).
- **Style:** Often considered more stylish and sleek, especially in competitive swimming [1](#).
- **Comfort:** Can be very comfortable if you prefer a snug fit [1](#).

**Shorts (Swim Trunks)**

- **Fit:** Looser and more relaxed, offering more coverage [2](#).
- **Versatility:** Great for both swimming and casual beachwear [2](#).
- **Comfort:** Generally more comfortable for lounging and casual activities [2](#).
- **Pockets:** Often come with pockets, which can be handy for carrying small items [2](#).

If you're planning to do a lot of swimming and want something that reduces drag, briefs might be the better choice. On the other hand, if you're looking for something versatile and comfortable for both swimming and hanging out at the beach, shorts could be more suitable.

Do you have a preference for a specific activity or style?

Learn more [1](#)  [2](#)  [realmenrealstyle.com](#)

# A “SIMPLISTIC” EXAMPLE: GPT-4 VERSION



You

Sorry, I'm not a native speaker, I meant speedo



Copilot

No problem at all! When you mentioned “briefs,” you were actually referring to Speedos. Here’s a quick recap of their advantages:

## Speedos (Briefs)

- **Fit:** Tight and body-hugging, providing minimal coverage.
- **Mobility:** Excellent for swimming as they offer less drag in the water.
- **Style:** Often considered more stylish and sleek, especially in competitive swimming.
- **Comfort:** Can be very comfortable if you prefer a snug fit.

Speedos are a great choice if you’re looking for something that enhances performance in the water. They are popular among competitive swimmers and those who prefer a streamlined look.

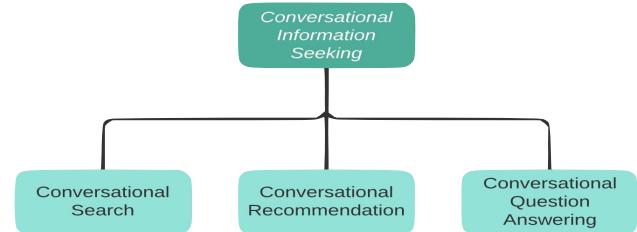
Do you think Speedos might be the right choice for you? Or would you like to explore other options?

# CONIAC AND CAFE

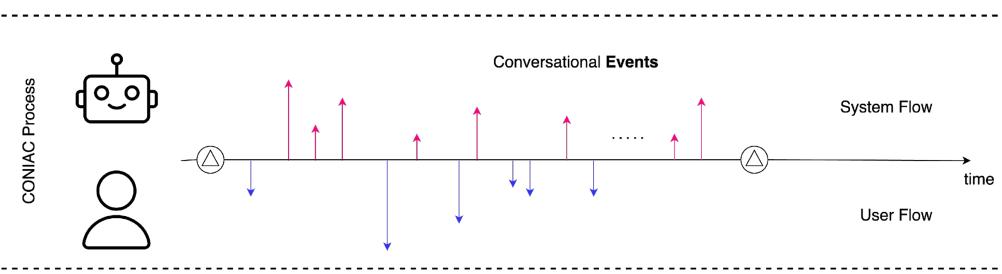
## CONIAC: CONversational Information ACcess

- Interact with users naturally and seamlessly.
- Guide a user through the process of refining and clarifying their needs.
- Aid decision-making by making personalized recommendations information while being able to explain them.
- Generate, retrieve and summarize relevant information.

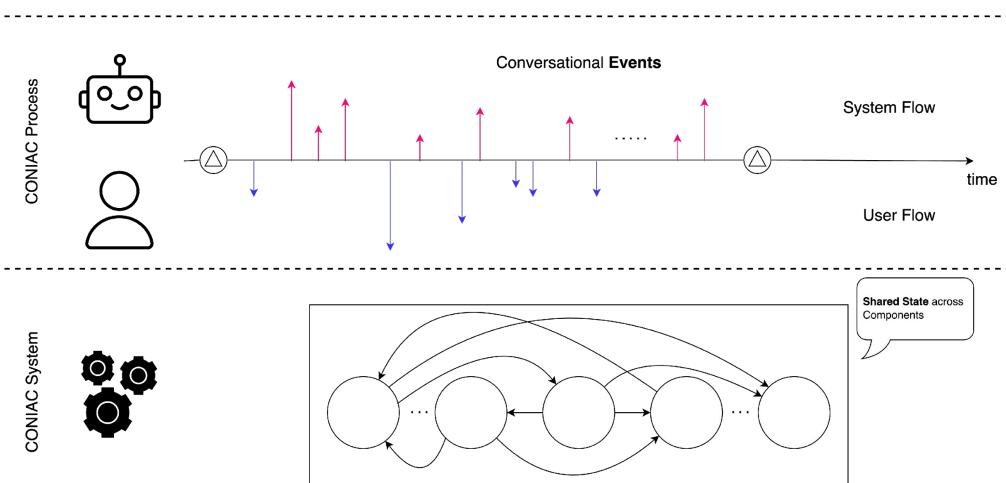
## CAFE: Conversational Agents Framework for Evaluation



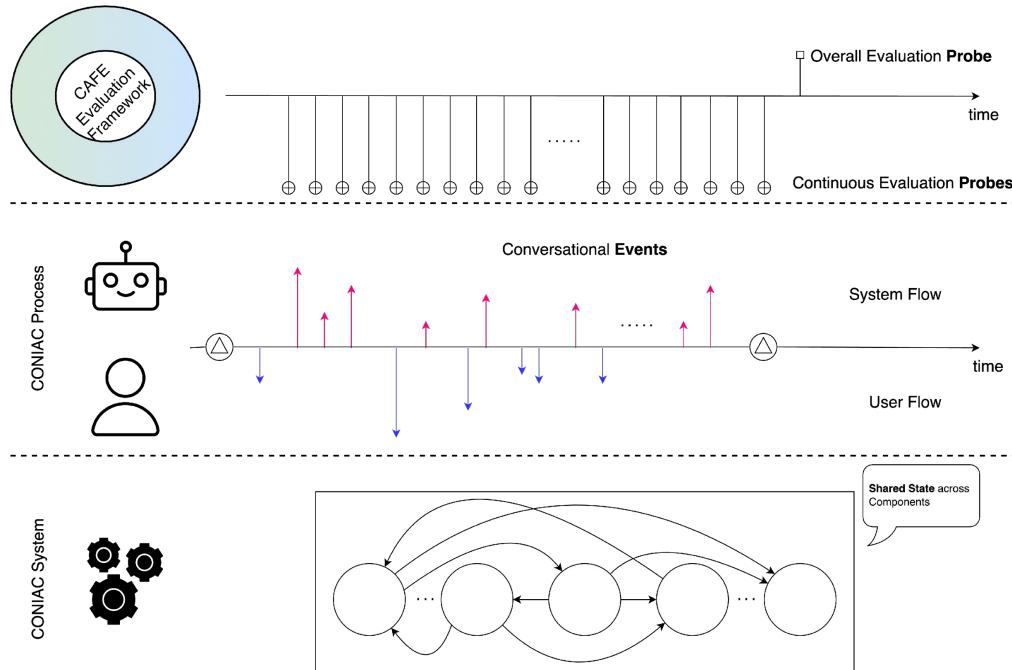
# CONIAC PROCESS



# CONIAC SYSTEM



# THE CAFE FRAMEWORK



# STAKEHOLDERS' GOALS

**Stakeholders** include users, system designers, app developers, distribution platforms, content creators, publishers, advertisers, and editors.

Each group has distinct **goals** that may overlap or conflict with others:

A successful system must address and balance differing objectives.

Key goals such as **relevance**, **truthfulness**, and **trustworthiness** are essential and are typically evaluated through established benchmarks.

Minimizing user effort, clarifying objectives through conversational interaction, and evolving goals over time are critical aspects of **user-system engagement**.

**Secondary goals**, such as relationship-building and trust assessment, which enhance the overall user experience.



# USER ASPECTS

---

A CONIAC system designed with principles of **Universal Design** would ensure accessibility for all users:

- impairments or modality, e.g., hands-free operation

Users vary in their **interaction preferences, domain knowledge, and cognitive needs.**

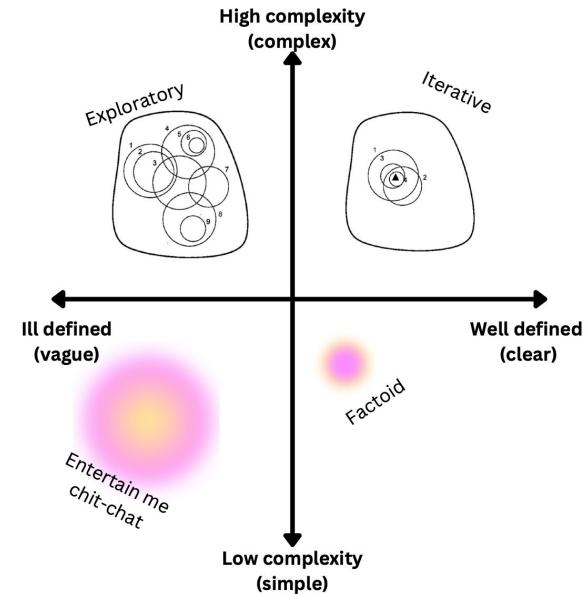
Individual differences such as **information-seeking behavior** or **maximization tendency** can affect how users interact with conversational systems.

Users' CONIAC tasks can also vary based on cultural **contexts.**

# TASKS

---

- Product search and recommendation
- Travel planning
- Hypothesis generation
- Health information seeking
- Enterprise & personal information management
- Children's Information Discovery
- Curate a bit of background entertainment
- Longitudinal learning
- ...

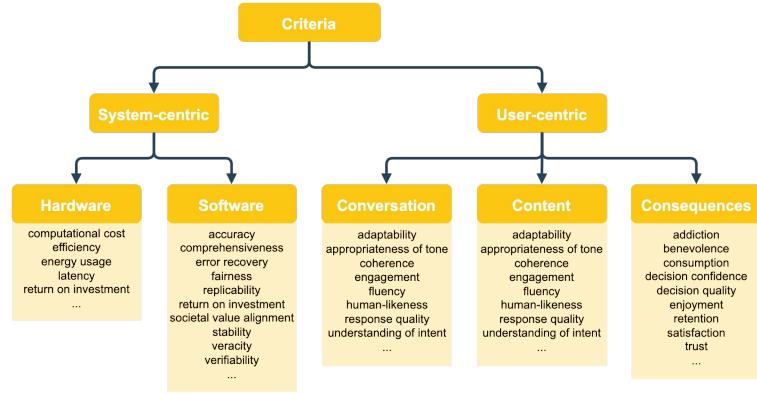


# CRITERIA

---

The **temporal dimension** of the interaction is particularly important in CONIAC systems.

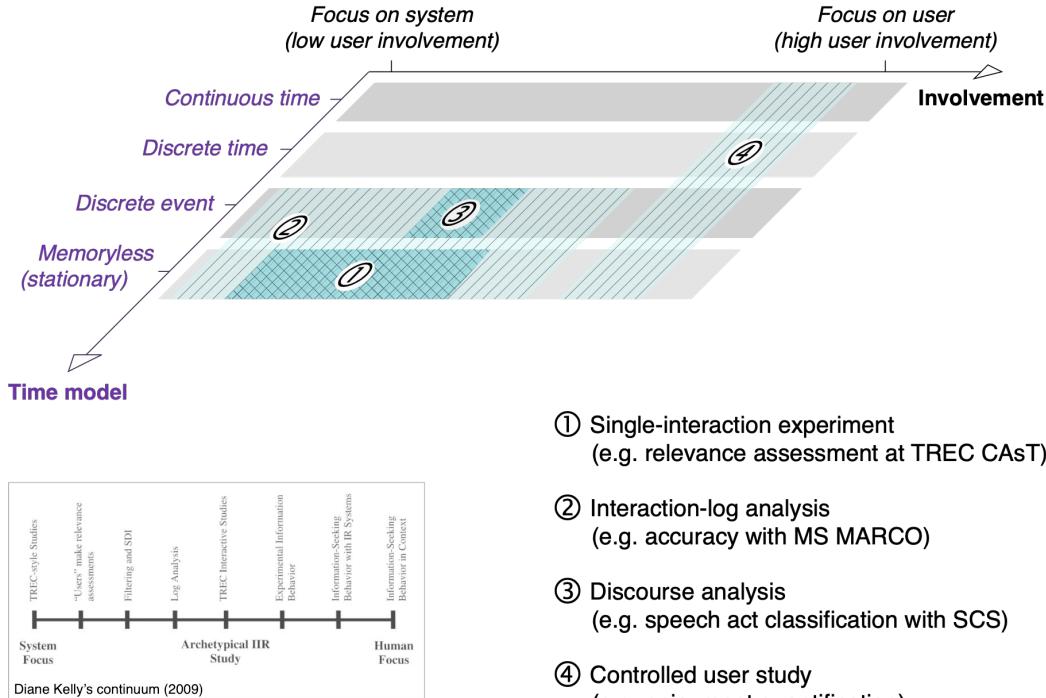
## Stationary and dynamic properties.



CONIAC systems must be evaluated at various temporal scopes:

- **Single turn** (i.e., the traditional evaluation metrics)
- **Conversation** (e.g., flow, continuance)
- Usage across **multiple conversations** (e.g., trust)

# METHODOLOGY AND CRITERIA



- ① Single-interaction experiment  
(e.g. relevance assessment at TREC CAsT)
- ② Interaction-log analysis  
(e.g. accuracy with MS MARCO)
- ③ Discourse analysis  
(e.g. speech act classification with SCS)
- ④ Controlled user study  
(e.g. enjoyment quantification)

# METHODOLOGY AND CRITERIA

---

Methodology	Criteria				
	System-centric		User-centric		
	Hardware	Software	Conversation	Content	Consequences
single-interaction experiment	✓	*			
offline simulation (incl. A/B)	✓	✓			
user simulation (incl. LLM-based)	✓	✓			
online A/B testing	*	✓			*
interaction-log analysis	✓	*	*		*
expert evaluation			*	*	
discourse analysis			✓	*	✓
data donation					*
wizard of oz / observational study			✓	*	*
controlled user study (incl. longitudinal)	*	*	✓	✓	✓
user survey (quantitative)			✓	✓	*
user interview (qualitative)			✓	✓	*
diary study / experience sampling			✓	✓	✓

---



# MEASURES

---





# STILL MANY OPEN QUESTIONS

---

- **Stakeholders' Goals**

- What do users want from CONIAC systems?
- How can CONIAC systems accurately deduce stakeholders' goals from interactions with users?
- How can we help users select the best CONIAC for a given task?

- **User Aspects**

- Which type of users can gain substantial benefits from CONIAC systems and how?
- What are good measures of individual and cultural differences for users of CONIAC systems?
- What individual and cultural differences are important moderators of the success of CONIAC systems?

- **Tasks**

- What common characteristics do complex tasks have that can benefit significantly from CONIAC systems?
- What temporal patterns can be observed in the level task complexity during interactions with a CONIAC system?
- How can a user and a CONIAC system efficiently agree that a given task is beyond the system's capabilities?
- What array of criteria and methods can capture, account for, and reflect the dynamic, evolving nature of conversations?
- Multi-modality for conversational systems

# STILL MANY OPEN QUESTIONS

- **Criteria**
  - What criteria make a difference in users' evaluation of CONIAC systems?
  - Can we foresee a future when we will be able to assess user-centric criteria (e.g., during system development) using LLMs?
  - How can we structurally report on multi-criterion evaluations that include both system-centric and user-centric criteria?
- **Methodology**
  - What methods can be developed to evaluate CONIAC systems to prevent test data leakage and ensure reliable evaluation metrics?
  - How can we accurately attribute the performance of a conversational model to its individual components?
  - What approaches can be used to evaluate the accuracy and relevance of responses from a CONIAC system, particularly when these responses are derived from the aggregation of multiple documents and the system's internal knowledge?
  - Can we reuse system-user interactions and corresponding labels to evaluate a different system without introducing bias?
- **Measures**
  - What offline evaluation measures can we develop to assess continuous criteria such as conversation flow and continuance?
  - What is the most effective and statistically sound approach to aggregate measures across different interactions? Is the average or should we develop protocols more aligned with the structure of a conversation (e.g., non-independent interactions, graph structure)?
  - Which measures are suited to evaluate the dynamic evolution of a conversation over time, including changes in mental models and user-centric evaluation criteria?

---

# HANDS-ON

# LINKS

---

Conversational search notebook:

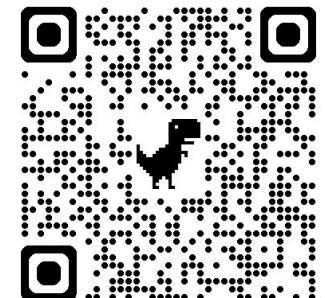
<https://colab.research.google.com/drive/1JcyIOd5RPCRAErFE0Jjg8TC4SP4-2Mu0>

CoSReC notebook:

<https://colab.research.google.com/drive/1BSwUk3qLq3cLTDrycQFWB4-Zx29738Td>

GitHub Repository:

<https://github.com/guglielmof/25-ECIR-CSR-TUTORIAL>



---

# A LITTLE CHAT GOES A LONG WAY

## HANDS-ON TUTORIAL ON CONVERSATIONAL INFORMATION ACCESS

Guglielmo **Faggioli**, Nicola **Ferro**, Simone **Merlo**



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



---

# BACKUP SLIDES

# QUESTION CLARIFICATION: DATASETS

## Qulac

Based on TREC Web Track 09-12 topic.

For each topic, **crowd** and **expert assessors** generated **12 clarification questions**, ensuring a question for each topic **facet**.

M. Aliannejadi, H. Zamani, F. Crestani, W. B. Croft: Asking Clarifying Questions in Open-Domain Information-Seeking Conversations. SIGIR 2019: 475-484

## ClariQ

**Crowd-sourced** datasets with three elements: **user requests**, corresponding **clarification questions**, **user answers**.

M. Aliannejadi, J. Kiseleva, A. Chuklin, J. Dalton, M. Burtsev: ConvAI3: Generating Clarifying Questions for Open-Domain Dialogue Systems (ClariQ). CoRR abs/2009.11352 (2020)

## ClariQ-FWk

Extract **facets** from ClariQ questions.

generate clarification questions corresponding to each facet for each query using a **fine-tuned** GPT-2.

I. Sekulic, M. Aliannejadi, F. Crestani: Towards Facet-Driven Generation of Clarifying Questions for Conversational Search. ICTIR 2021: 167-175



# RETRIEVAL: CoSPLADE

---

CoSPLADE is a refined version of the neural **sparse** approach SPLADE for **end-to-end conversational retrieval**.

CoSPLADE starts from the assumption that the documents' representation of the original model (SPLADE) is already good. Hence, it **focuses** exclusively **on the query representation**.

# RETRIEVAL: CoSPLADE

the k-th query encoding  $\mathbf{q}'_k$  is defined as follows:

$$\mathbf{q}'_k = \mathbf{q}_k^q + \mathbf{q}_{k,n}^a$$

where  $\mathbf{q}_k^q$  and  $\mathbf{q}_{k,n}^a$  are the contextual representations of the utterances and the system's answers, respectively.

We call SPLADE(tokens,  $\theta$ ) a SPLADE encoder that takes in input the tokens and produces their sparse representation:

SPLADE("the neverending story",  $\theta$ ) = [("neverending", 1.8), ("story", 1.7), ("film", 1.2), ("movie", 0.8), ("artax", 0.2), ...]



# RETRIEVAL: CoSPLADE

the k-th query encoding  $q'_k$  is defined as follows:

$$q'_k = q_k^q + q_{k,n}^a$$

the query encoder concatenates the current utterance with the previous ones

$$\text{SPLADE}([\text{CLS}] \circ q_k \circ [\text{SEP}] \circ q_1 \circ [\text{SEP}] \circ \dots \circ [\text{SEP}] q_{k-1}; \theta_q)$$

the answer encoder averages the SPLADE representation of the last n answers. n is usually 1 or k-1

$$\frac{1}{n} \sum_i^n \text{SPLADE}(q_k \circ [\text{SEP}] a_i; \theta_{a,n})$$

Each encoder is parametrized separately.

# RETRIEVAL: CoSPLADE

Like ConvDR, CoSPLADE employs a teacher-student approach where SPLADE is the teacher model.

We call  $\mathbf{q}_k^*$  the SPLADE representation of the oracle manual rewritten query.

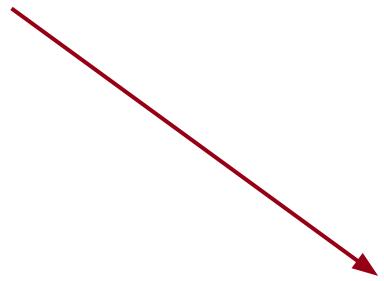
To learn the parameters  $\theta_q$  and  $\theta_{a,n}$ , the loss function is:

$$\text{Loss}(\mathbf{q}'_k, \mathbf{q}_k^*) = \text{MSE}(\mathbf{q}'_k, \mathbf{q}_k^*) + \max(\mathbf{q}_k^* - \mathbf{q}_{k,n}^a, 0)^2$$



# RETRIEVAL: CoSPLADE

Minimise the difference between contextual and gold truth representations of  $q_k$ .



$$\text{Loss}(q'_k, q_k^*) = \text{MSE}(q'_k, q_k^*) + \max(q_k^* - q_{k,n}^a, 0)^2$$

Force the answer encoder to take from answers terms in the oracle manual rewritten query.

