



Conversational Recommender Systems

Tutorial at IJCAI 2021, Montreal/Online

Dietmar Jannach, University of Klagenfurt, Austria

dietmar.jannach@aau.at

Li Chen, Hong Kong Baptist University, China

lichen@comp.hkbu.edu.hk

Welcome

- Today's presenters
- Li Chen
 - Associate Professor at HKBU
 - *Research interests:* Recommender Systems, Artificial Intelligence, Human Computer Interaction
- Dietmar Jannach
 - Professor at University of Klagenfurt
 - *Research interests:* Recommender Systems, Software Engineering, Artificial Intelligence



Basic Reading

- Jannach, D., Manzoor, A., Cai, W. and Chen, L.:
"A Survey on Conversational Recommender Systems". ACM Computing Surveys, 2021
- Gao, C., Lei, W., He, X., de Rijke, M., Chua, T.:
"Advances and Challenges in Conversational Recommender Systems: A Survey",
<https://arxiv.org/abs/2101.09459>

About you



- In case of questions, feel free to raise your hand or write into the chat anytime ...

Agenda

1. Introduction
2. Characterization of Conversational Recommender System (CRS)
3. Interaction Modalities
4. Underlying Knowledge and Data
- Short break --
5. Computational Tasks
6. Evaluation of CRS
7. Wrap-up & Discussion

Agenda

1. **Introduction**
2. Characterization of Conversational Recommender System (CRS)
3. Interaction Modalities
4. Underlying Knowledge and Data
- Short break --
5. Computational Tasks
6. Evaluation of CRS
7. Wrap-up & Discussion

Towards Deep Conversational Recommendation (NeurIPS '19)

OURS: hi, how can i help you ?

SEEKER: 2001 : a space odyssey might be a great option. anything else that you would suggest ?

OURS: star wars : the force awakens is also a good one return of the jedi all good movies

SEEKER: those sound good ! i 'm going to look into those movies.

OURS: i hope i was able to help you find a good movie to watch

SEEKER: thank you for your help ! have a great night ! good bye

Already some history (1979)

I'D LIKE TO KNOW WHAT SORT OF PERSON YOU THINK YOU ARE. PLEASE TYPE IN A FEW SINGLE WORDS THAT YOU THINK CHARACTERIZE THE MOST IMPORTANT ASPECTS OF YOURSELF

unconventional open direct honest humorous persistent adventurous

HAVE YOU EVER READ EARTH ANGELS BY SUSAN CAHILL?

no

YOU MIGHT LIKE IT.

IT IS ABOUT A GIRL WHO GOES TO CATHOLIC ELEMENTARY SCHOOL, HIGH SCHOOL, AND COLLEGE, THEN GOES ON TO BECOME A NUN, BUT FINALLY BACKS OUT A FEW DAYS BEFORE FINAL VOWS.

DOES THAT SOUND GOOD?

no

WAS IT BECAUSE OF THE DEPTH OF THE CHARACTER DEVELOPMENT? (EITHER TOO MUCH OR NOT ENOUGH)

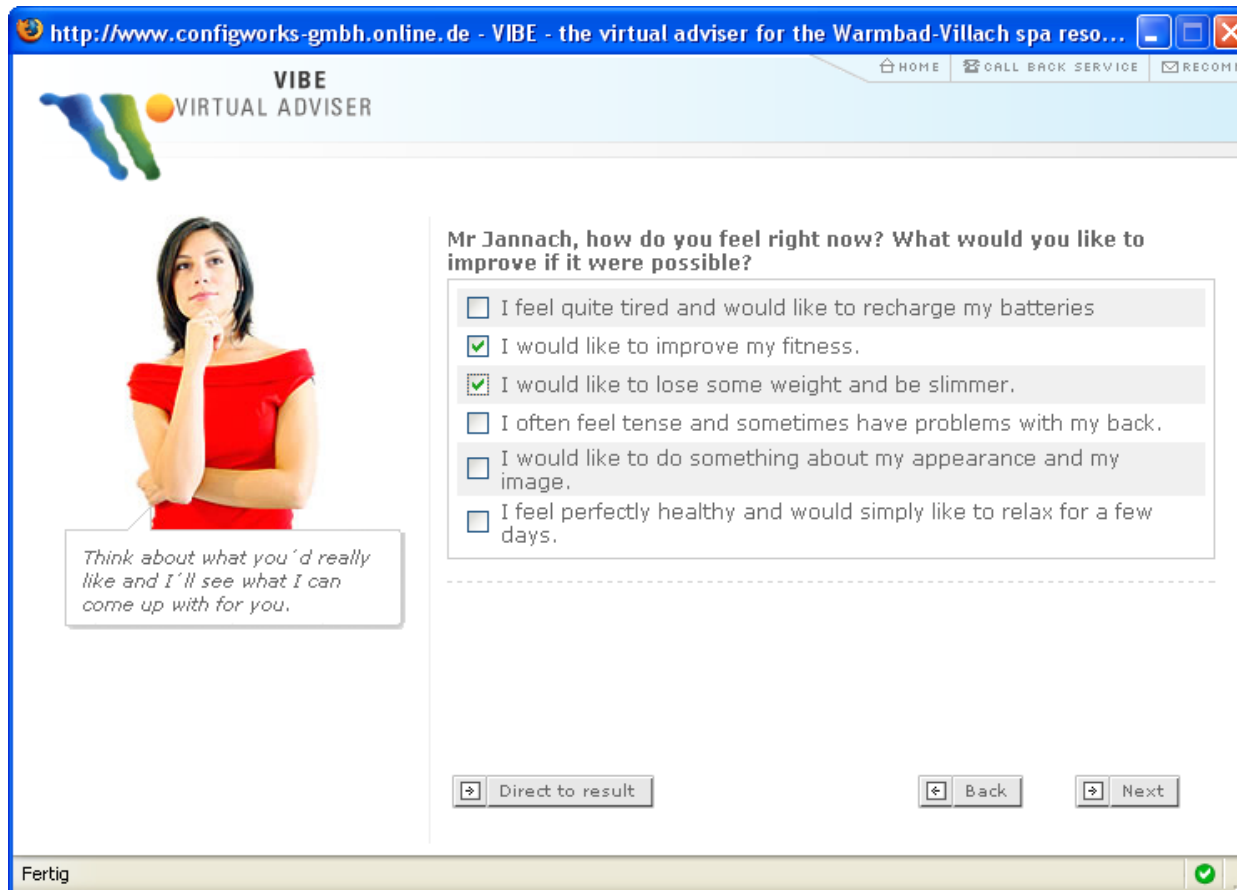
no

1997

The FindMe
System:
“Critiquing”

The screenshot shows the 'Entree Chicago' website interface. At the top, there is a logo featuring a silver dome with a city skyline inside, and the text 'Entree Chicago' in a cursive font. Below the logo, the text 'I would like to eat at a restaurant that has:' is displayed in a cursive font. Underneath this text are five input fields: 'Cuisine', 'Price', 'Style', 'Atmosphere', and 'Occasion'. Below these fields, the text 'I would like to eat at a restaurant just like:' is displayed in a cursive font. Underneath this text are two input fields: 'Chinois on Main' and 'Los Angeles'. At the bottom of the interface, there are two buttons: 'New Query' and 'Submit', both featuring a silver dome icon.

Dynamic form-based Interactions (early 2000s)




Form-based Interactions

http://www.configworks-gmbh.online.de - VIBE - the virtual adviser for the Warmbad-Villach spa reso...

VIBE
VIRTUAL ADVISER

HOME CALL BACK SERVICE RECOMMENDATIONS

Did you know that...



Wonderful, we've now got to your final selection. Here's my recommendation for you ...

❖ **Feel well week**

Length of stay:	per week (7 nights) per person
Meals:	Half board
Accommodation:	The Warmbaderhof
Dates:	At any season
Rate in single room:	from € 1595
Rate in double room:	from € 1595

[Details](#)
[Why?](#)

I can also recommend the following packages:

- You can book a personal massage or a whole massage programme for your stay at any time.

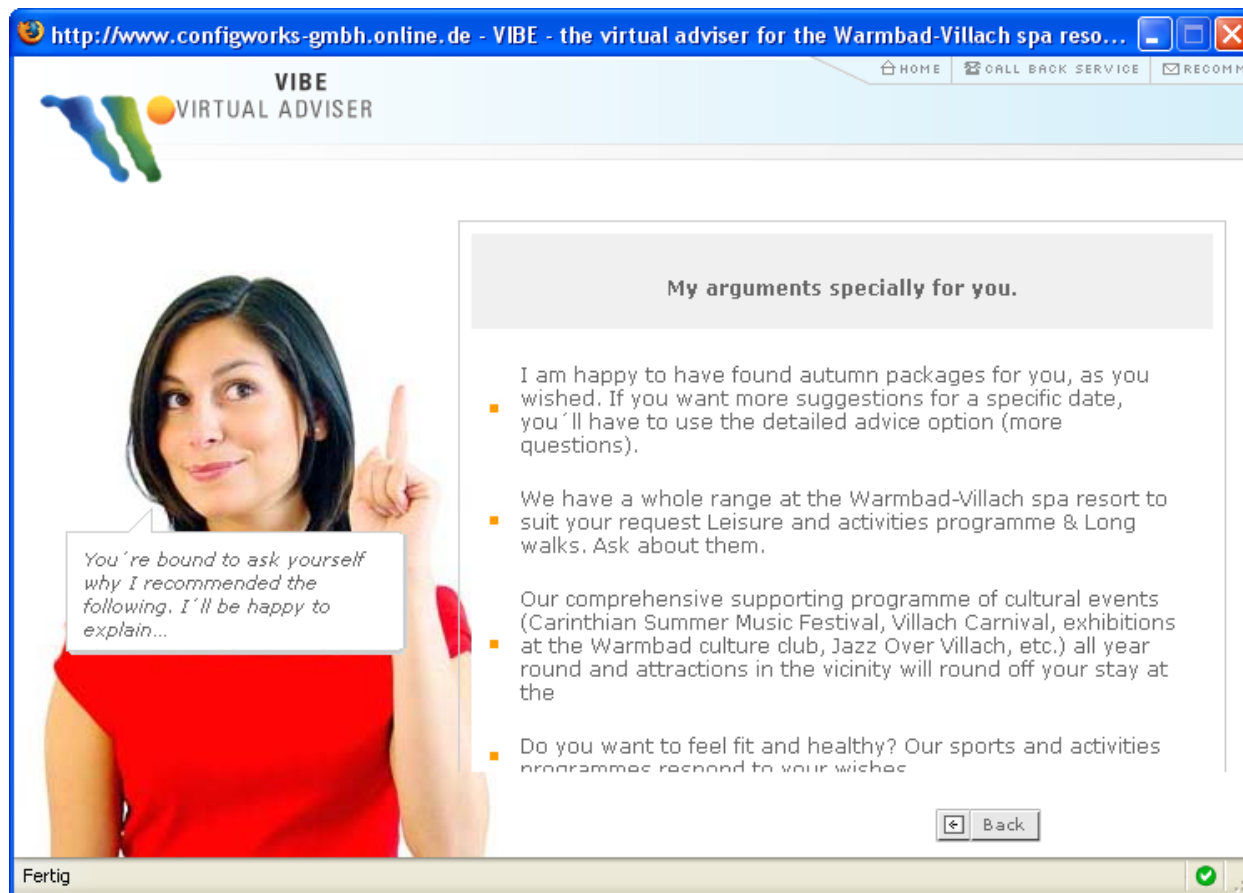
❖ **Golf & Spa**

Length of stay:	per week (7 nights) per person
Meals:	Half board
Accommodation:	The Warmbaderhof
Dates:	01.04.2008-31.10.2008

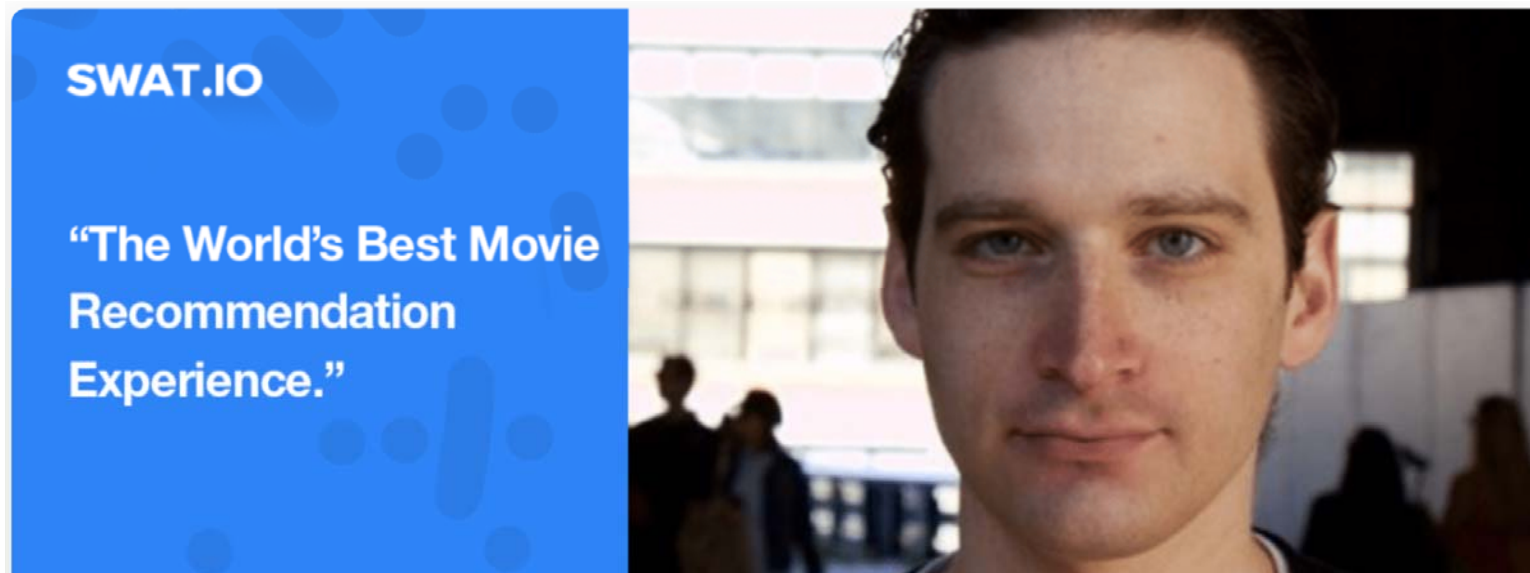
[Details](#)
[Why?](#)

Fertig

Form-based Interactions



In the real world in the early 2020s



And Chill



OK, Let's go! The floor is yours. Give me ONE movie you like, and why you like it.

I like rebecca, it is wiht a lot of suspense, black and white hitchcock-like

And Chill



You should try Band Of Angels!



Band of Angels -
Trailer

And Chill

why do you recommend
band of angels

Gotcha... that title,
right? There's only
1.8% of users that have
mentioned this one. I
am still generating AI-
powered movie recs
based on this film. Can
you tell me about a
different one you like?



And Chill

i'd rather like to understand more about why you recommended scarlett. what is so interesting about it

Darn, you outfoxed me & And Chill. Great request, but give me a raincheck on answering!



Towards Deep Conversational Recommendation - Supplemental

SEEKER: yes i loved that movie as well . are there any good movies without dinosaurs ?

OURS: i have n't seen that one . i will have to check it out . have you seen jurassic park

SEEKER: hi

OURS: what kind of movies do you like ? what kind of movies do you like ?

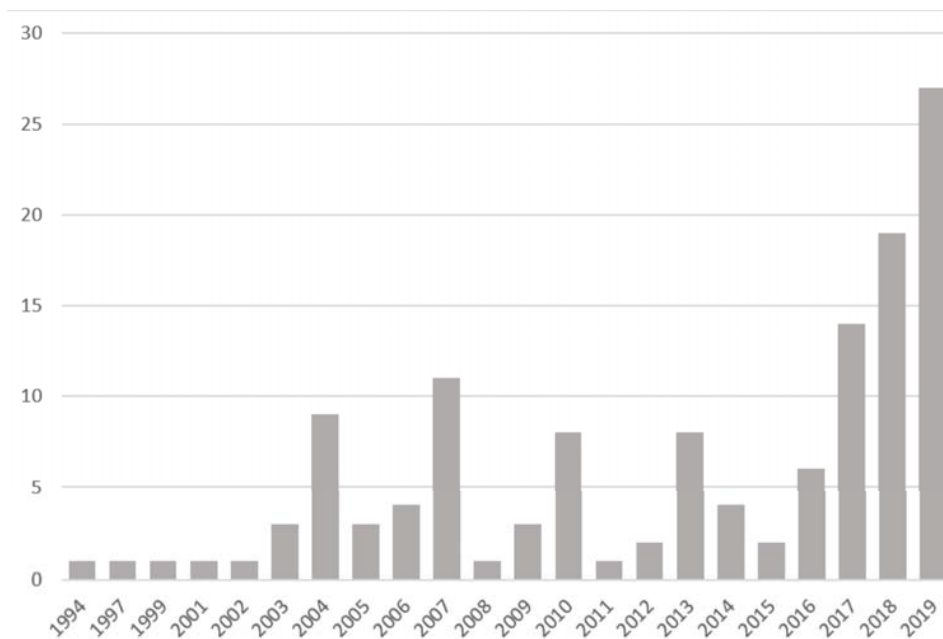
A Grand Challenge

	DeepCRS	KBRD
Number of dialogues	70	70
Generated sentences (overall)	399	359
Sentences labeled as meaningful	277 (69%)	209 (58%)
Sentences labeled as <i>not</i> meaningful	122 (31%)	150 (42%)
Dialogues without problems	5	5
Chit-chat sentences	132	88
Chit-chat labeled as meaningful	112 (85%)	77 (87%)
Number of recommendations	106	119
Recs. labeled as meaningful	63 (60%)	66 (55%)
Nb. dialogues with no meaningful recs.	25 (36%)	20 (28%)
Nb. dialogues with no rec. made.	7 (10%)	6 (8.5%)

Table 2: Analysis of Dialogue and Recommendation Quality

Growing Research Interest

- Advances in speech-enabled devices
- Advances in natural language processing
- Advances in machine learning in general



Agenda

1. Introduction
2. **Characterization of Conversational Recommender System (CRS)**
3. Interaction Modalities
4. Underlying Knowledge and Data
- Short break --
5. Computational Tasks
6. Evaluation of CRS
7. Wrap-up & Discussion

Characterization of a CRS

- “A CRS is a software system that supports its users in achieving recommendation-related *goals* through a *multi-turn dialogue*.”

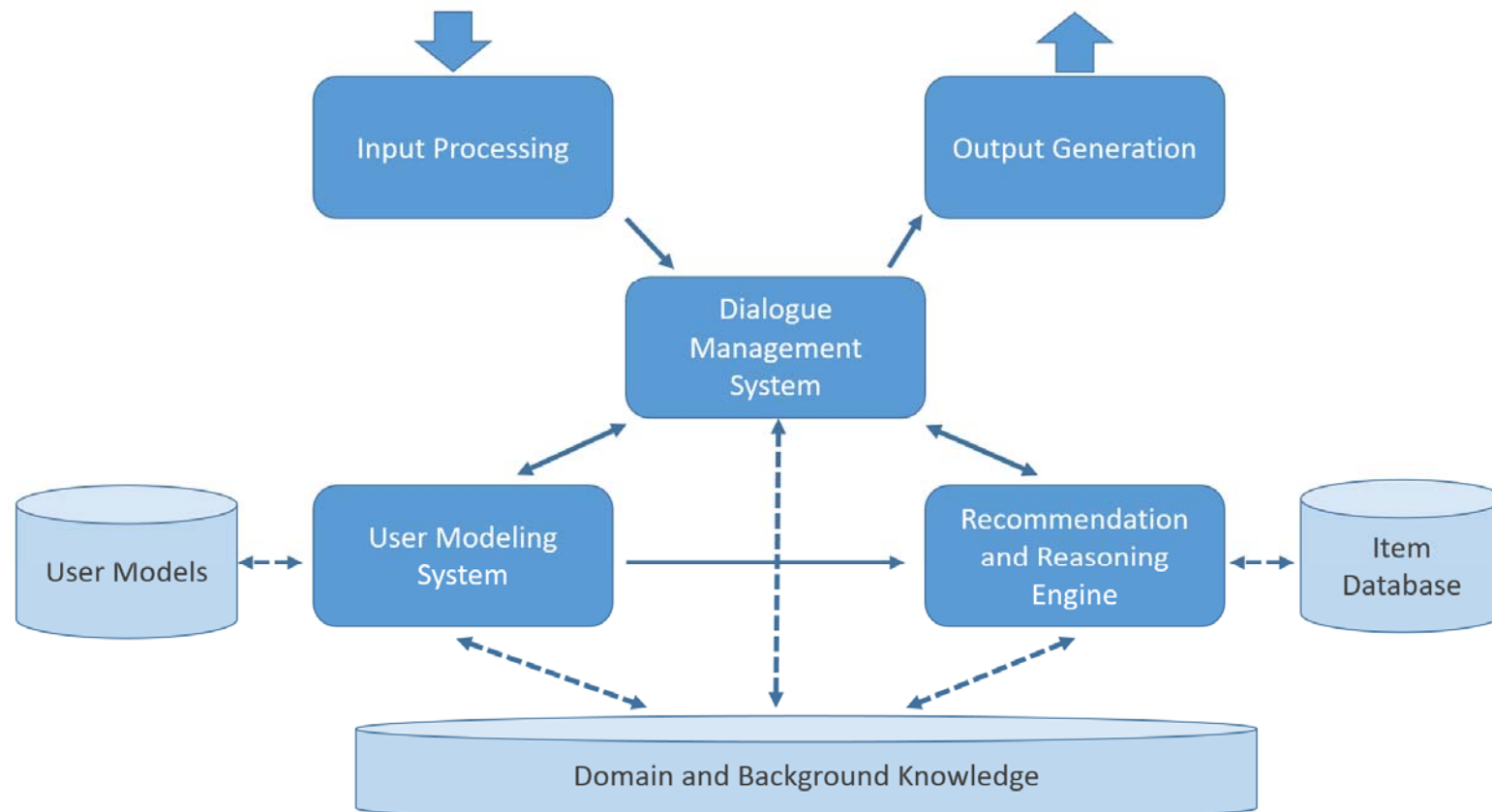
(Jannach et al., CSUR 2021)

- Characteristics
 - Task orientation
 - Recommendation, decision support, preference elicitation, explanation
 - Multi-turn interaction
 - Compare: Q&A systems, Apple’s Siri
 - Requires form of *state management*

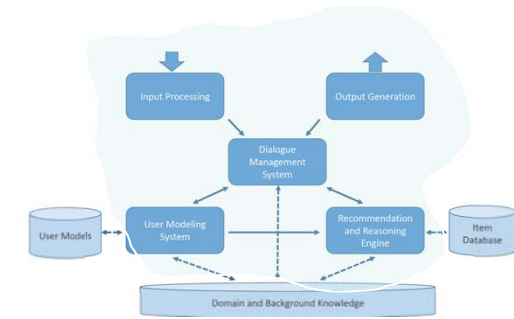
Characterization of a CRS

- Various interaction modalities possible
 - *Interaction forms*: voice input and output, forms and buttons, multimedia, gestures, facial expressions, ...
 - *Devices*: can be mobile, desktop, a robot, a kiosk, an interactive wall, a 3D environment, ...
- Relation to Conversational Search
 - Some similarities, blurry boundary
 - CRS: Often more complex approaches to model user (long-term) preferences and needs

Conceptual Architecture

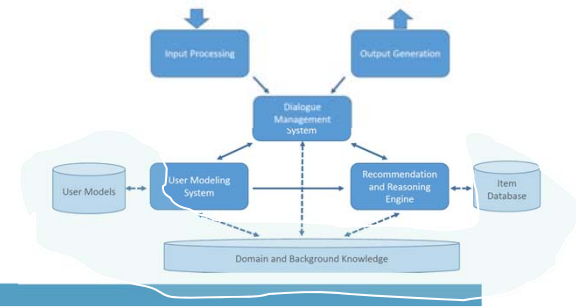


Conceptual Architecture



- Computational Elements
 - Dialogue Management (state tracker)
 - User Modeling System
 - Recommendation and Reasoning Engine
 - May include the generation of explanations or the decision about the next conversational move (including intent recognition)
 - Input and Output Processing
 - Speech-to-text conversion, voice output, named entity recognition, ...

Conceptual Architecture

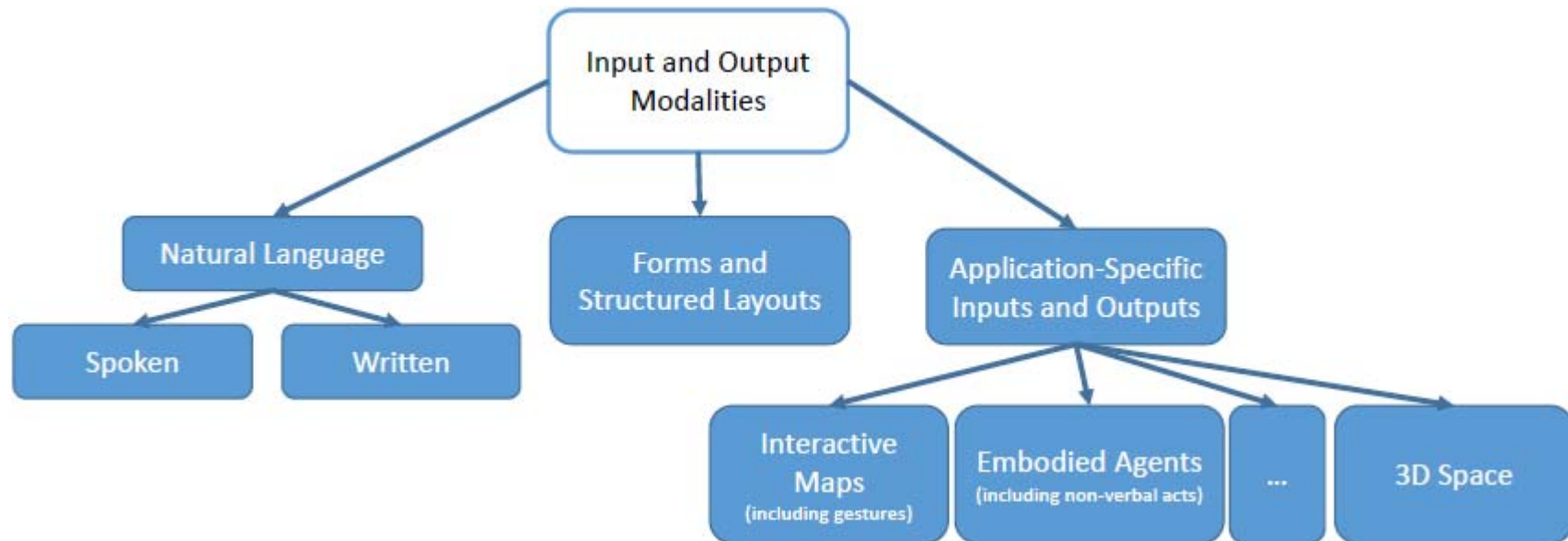


- Knowledge Elements
 - Item Database
 - Catalog of recommendable items, including meta-data
 - Domain & Background Knowledge
 - Dialogue Knowledge
 - Possible states and transitions, supported intents
 - World Knowledge
 - e.g., from DBPedia
 - Recorded dialogues for learning

Agenda

1. Introduction
2. Characterization of Conversational Recommender System (CRS)
- 3. Interaction Modalities**
4. Underlying Knowledge and Data
- Short break --
5. Computational Tasks
6. Evaluation of CRS
7. Wrap-up & Discussion

Interaction Modalities



Single Modality or Combined

- Forms (e.g., buttons, radio-buttons) and structured layouts
 - Common for critiquing-based CRS and web-based interactive advisory systems
- Natural language based interaction
 - Task-oriented dialogue systems and more recent (deep) learning-based approaches
 - Spoken-text-only approaches are often implemented on smart speakers (e.g., Amazon Alexa or Google Home)
- Hybrid approach
 - Combining natural language with other modalities

Example: FindMe

UKRAINIAN VILLAGE. TWO bedroom rehab garden apartment. Lr, Eurokitchen, hwrfl, excellent security, forced air, lots of closets, laundry in building. Garage space included. Dogs OK. Available immediately. \$600/ mo. 312-489-1554. / ;			
Phone: 312-489-1554	2-bedrooms	\$600	60622 (West Town Bucktown)

This apartment is OK, but make it...

bigger cheaper nicer safer

This neighborhood could be more...

convenient conservative dynamic


Burke, R., Hammond, K., and Young, B.: The FindMe Approach to Assisted Browsing. IEEE Expert: Intelligent Systems and Their Applications 12(4):32-40, 1997.

Example: Dynamic Critiquing


QWIKSHOP.COM[HOME](#) [ABOUT THIS PROJECT](#) [CONTACT](#)


» Digital Cameras

[Shop for:](#) [Digital Cameras](#) [Computers](#) [Holidays](#)



Product Found: Canon EOS 30
6.3 Megapixel CMOS sensor
7-point wide-area AF
High-performance DIGIC processor
100-1600 ISO speed range
Compatible with all Canon EF lenses and EX Speedlites
PictBridge, Canon Direct Print and Bubble Jet Direct compatible – no PC required

[I've found the Camera I want!](#) 

[No lets start again](#) 

Adjust your preferences to find the right camera for you

Manufacturer	X	Canon	X
Optical Zoom	↓	7x	↑
Memory (MB)	↓	512	↑
Weight (Grams)	↓	780	↑
Resolution	↓	6.2 M Pixels	↑
Size	X	Large	X
Case	X	Magnesium	X
Price	↓	995	↑

We have more matching cameras with the following:

1. Less Memory and Lower Resolution and Cheaper	EXPLAIN	PICK
2. Different Manufacturer and Less Zoom and Lighter	EXPLAIN	PICK
3. Lighter and Smaller and Different Case	EXPLAIN	PICK

Explain:

1. Less Memory and Lower Resolution and Cheaper

This Critique covers **153** other Digital Cameras

Less Memory
Current Value: 512 MB
Critique: Less Than
Remaining: (0 to 256 MB)

Lower Resolution
Current Value: 6.2 M Pixels
Critique: Less Than
Remaining: (1.4 to 5.9 M Pixels)

Cheaper
Current Value: 995 €
Critique: Less Than
Remaining: (75€ to 960€)

[PICK](#)

McCarthy, K., Reilly, J., McGinty, L., and Smyth, B.: Experiments in Dynamic Critiquing. In Proceedings of the Tenth International Conference on Intelligent User Interfaces, 175-182, New York: ACM Press, 2005.

Example: Example Critiquing

Compare

Would you like to compare

Apt 34: room in a house, 600 frs, 15 square meters, private bathroom, private kitchen, 15 minutes to your work place

with other apartments for

☐ Better Type ☐ Cheaper Price ☒ Bigger Area


☐ Better Bathroom ☐ Better Kitchen ☐ Closer Distance

You are willing to compromise on the following attributes:

☐ Type of Apartment ☐ Price ☐ Area

☐ Bathroom ☒ Kitchen ☒ Distance

To find similar products with better values than this one

 **Canon PowerShot S2 IS Digital Camera**
\$424.15
Canon, 5.3 M pixels, 12x optical zoom, 16 MB memory, 1.8 in screen size, 2.97 in thickness, 404.7 g weight. [detail](#)

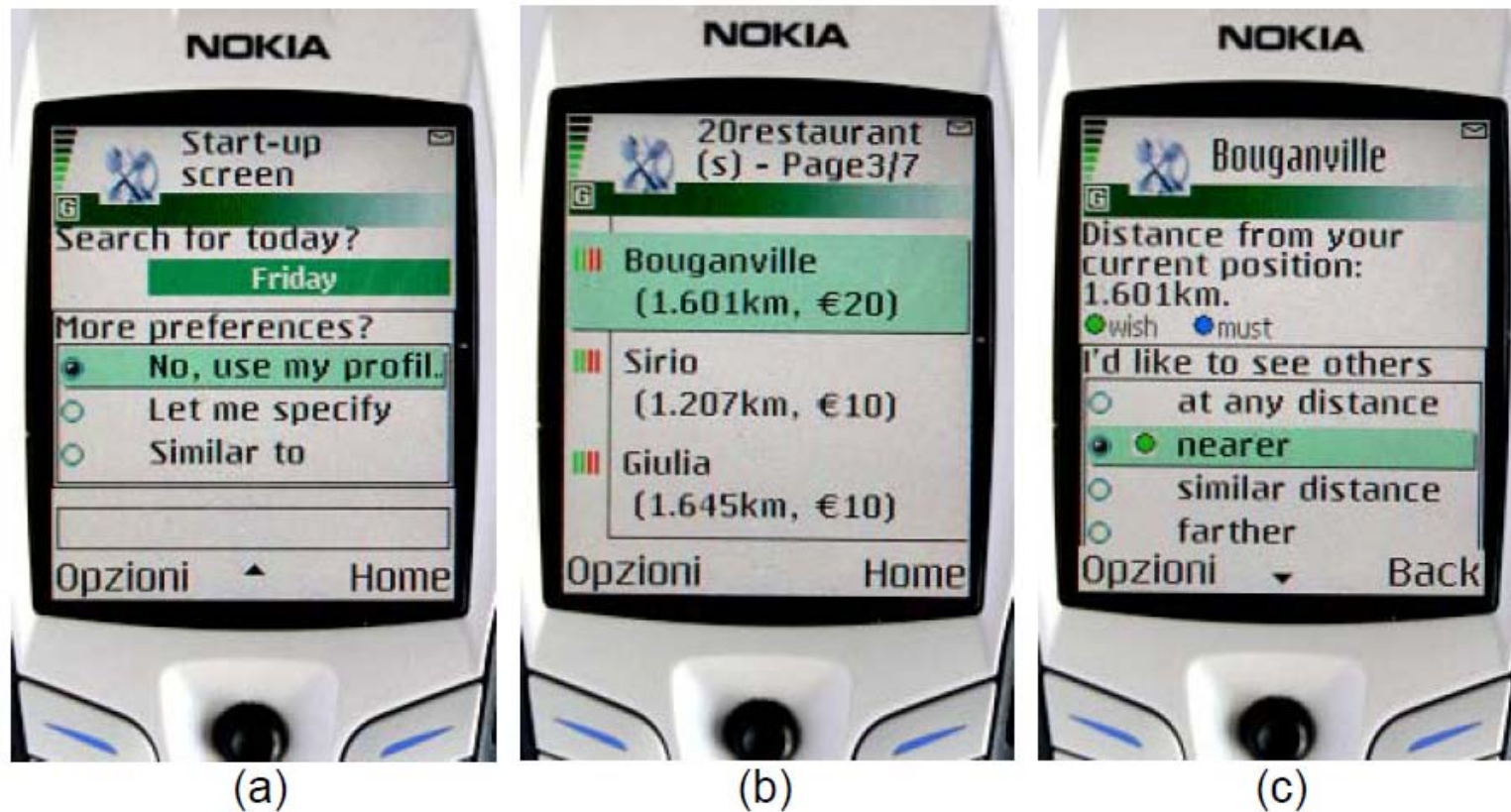
would you like to improve some values?

	Keep	Improve	Take any suggestion
Manufacturer	<input checked="" type="radio"/> Canon	<input type="radio"/> Sony	<input type="radio"/>
Price	<input type="radio"/> \$424.15	<input checked="" type="radio"/> less expensive	<input type="radio"/>
Resolution	<input checked="" type="radio"/> 5.3 M pixels	<input type="radio"/> \$100 cheaper	<input type="radio"/>
Optical Zoom	<input checked="" type="radio"/> 12x	<input type="radio"/> \$200 cheaper	<input type="radio"/>
Removable Flash Memory	<input checked="" type="radio"/> 16 MB	<input type="radio"/> \$300 cheaper	<input type="radio"/>
LCD Screen Size	<input checked="" type="radio"/> 1.8 in	<input type="radio"/> more memory	<input type="radio"/>
Thickness	<input checked="" type="radio"/> 2.97 in	<input type="radio"/> larger	<input type="radio"/>
Weight	<input checked="" type="radio"/> 404.7 g	<input type="radio"/> thinner	<input type="radio"/>
		<input type="radio"/> lighter	<input type="radio"/>

Pu, P. and Chen, L.: Integrating Tradeoff Support in Product Search Tools for E-Commerce Sites. In Proceeding of ACM Conference on Electronic Commerce (EC'05), pages 269-278, Vancouver, Canada, June 5-8, 2005.

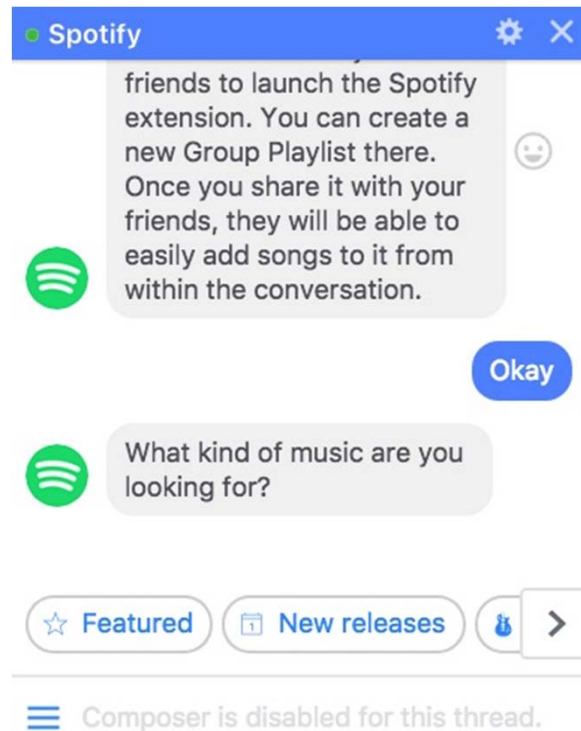
Pu, P. and Chen, L.: Evaluating Critiquing-based Recommender Agents. In Proceedings of Twenty-first National Conference on Artificial Intelligence (AAAI'06), pages 157-162, Boston, USA, July 16-20, 2006.

Example: Mobile Critiquing

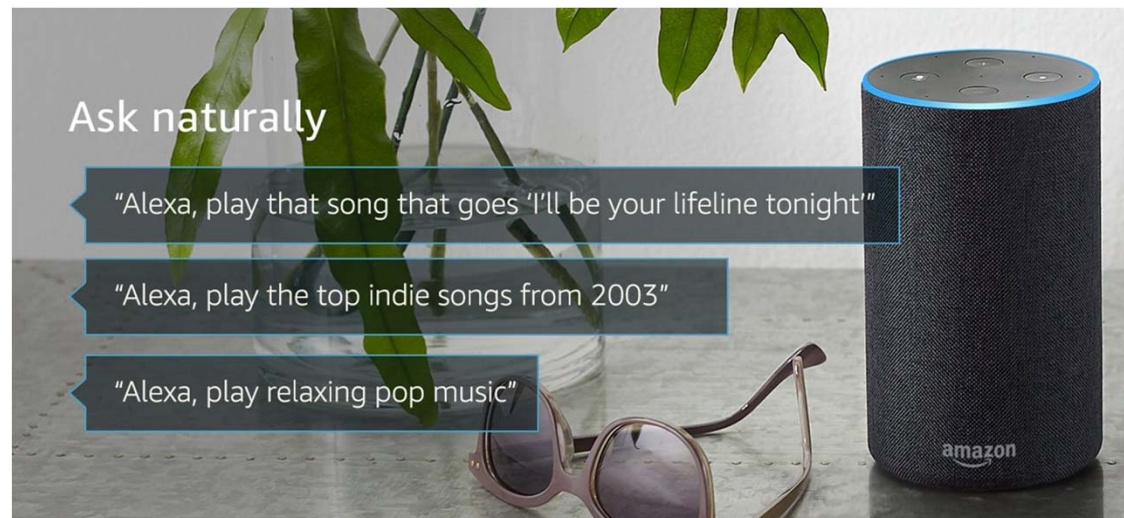


Ricci, F. and Nguyen, Q. N.: Acquiring and Revising Preferences in a Critique-Based Mobile Recommender System, in IEEE Intelligent Systems, vol. 22, no. 3, pp. 22-29, May-June 2007.

Natural Language Interaction

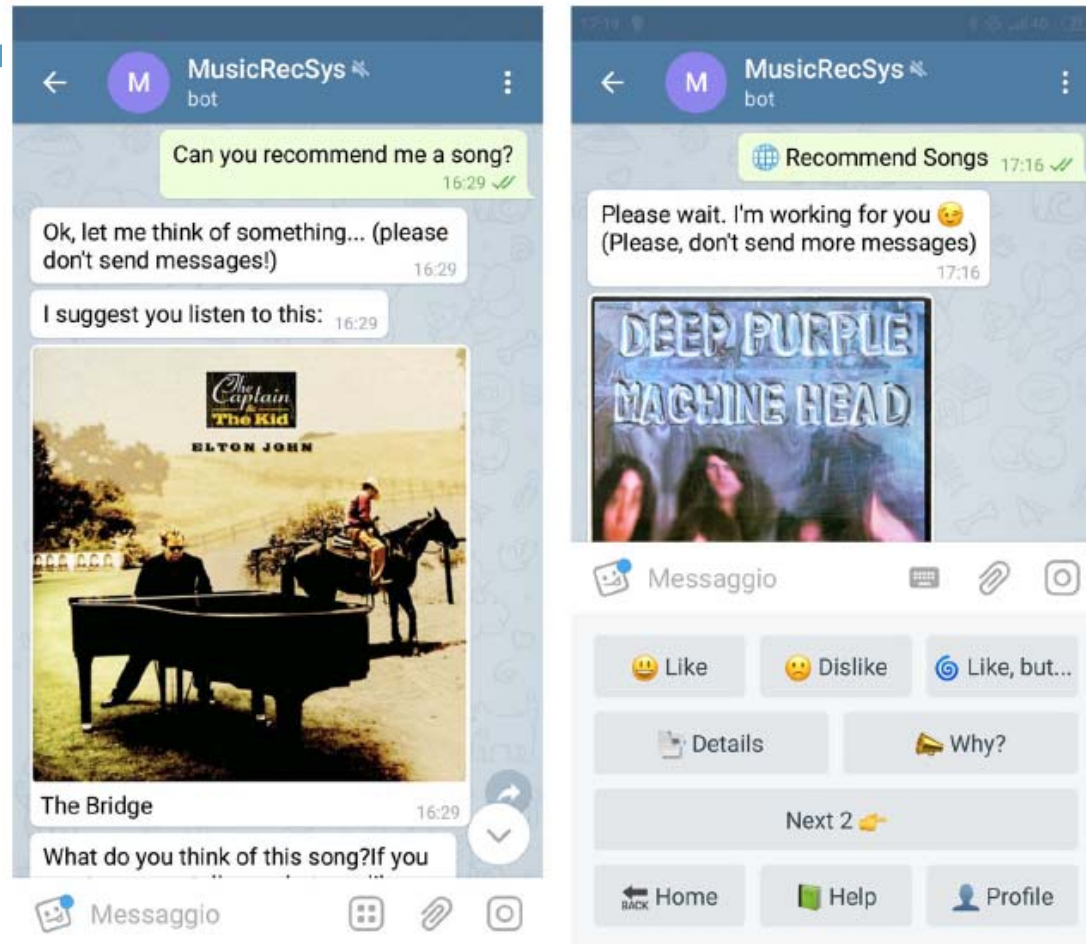


<https://www.poptin.com/blog/how-to-use-chatbots-drive-sales-engagement/>



<https://www.amazon.co.uk/b?ie=UTF8&node=11368385031>

Combined Modalities




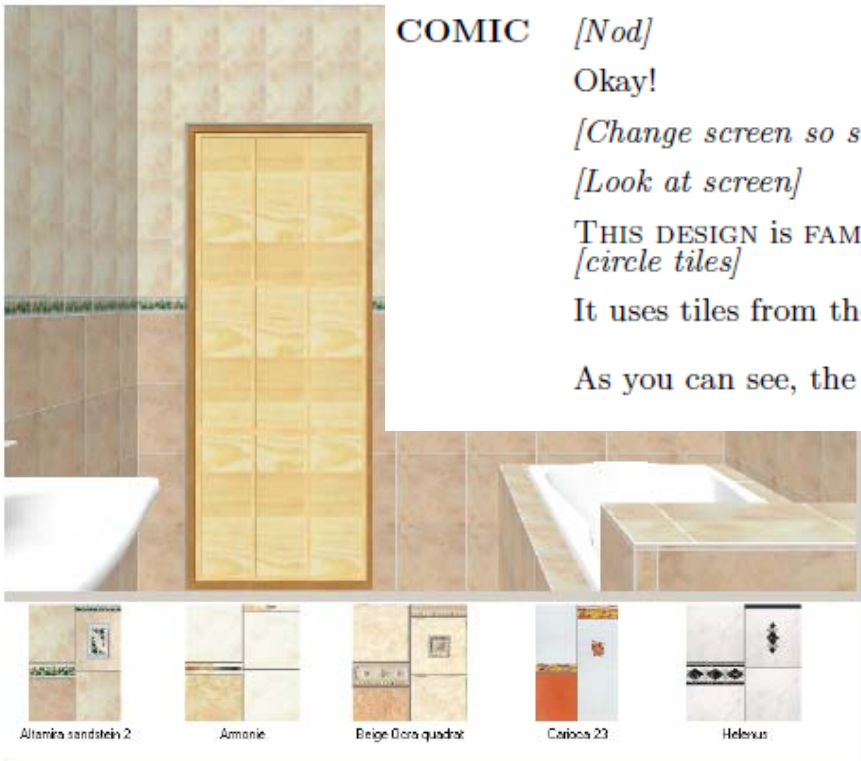
Narducci, F., Basile, P., de Gemmis, M., Lops, P., and Semeraro, G.: An investigation on the user interaction modes of conversational recommender systems for the music domain. UMUI '19, pages 1–34, 2019.

Application-specific Modalities

- Visual inputs on geographic map
- Pen gestures
- Handwritten input
- Body postures
- Gestures
- Facial expressions
- Speech prosody
- Etc.

Embodied Conversational Agents





User Tell me about this design *[circle Lollipop]*

COMIC *[Nod]*

Okay!

[Change screen so selected design is shown in room]

[Look at screen]

THIS DESIGN is FAMILY.
[circle tiles]

It uses tiles from the LOLLIPOP collection by AGROB BUCHTAL.
[point at manufacturer name]

As you can see, the colours are BLUE and GREEN.
[point at tiles]

Figure 1. Components of the COMIC system (tile-browsing phase)

Foster, M. E., and Oberlander, J.: User preferences can drive facial expressions: Evaluating an embodied conversational agent in a recommender dialogue system. *User Modeling and User-Adapted Interaction*, 20(4):341–381, 2010.

Virtual 3-D Space



(a) Three users are currently working individually, each one with its own panel (t_1 in Figure 4).



(b) Tester 2 likes the product on Tester 3's panel and presses "I like" button. Tester 3 and Tester 2 are collaborating and they use chat messages for deciding a critique (t_2 in Figure 4).



Fig. 3: Screenshot of the recommendation panel.

Contreras, D., Salamo, M., Rodriguez, I., and Puig, A.: Shopping decisions made in a virtual world: Defining a state-based model of collaborative and conversational user-recommender interactions. IEEE Consumer Electronics Magazine, 7(4): 260–35, 2018.

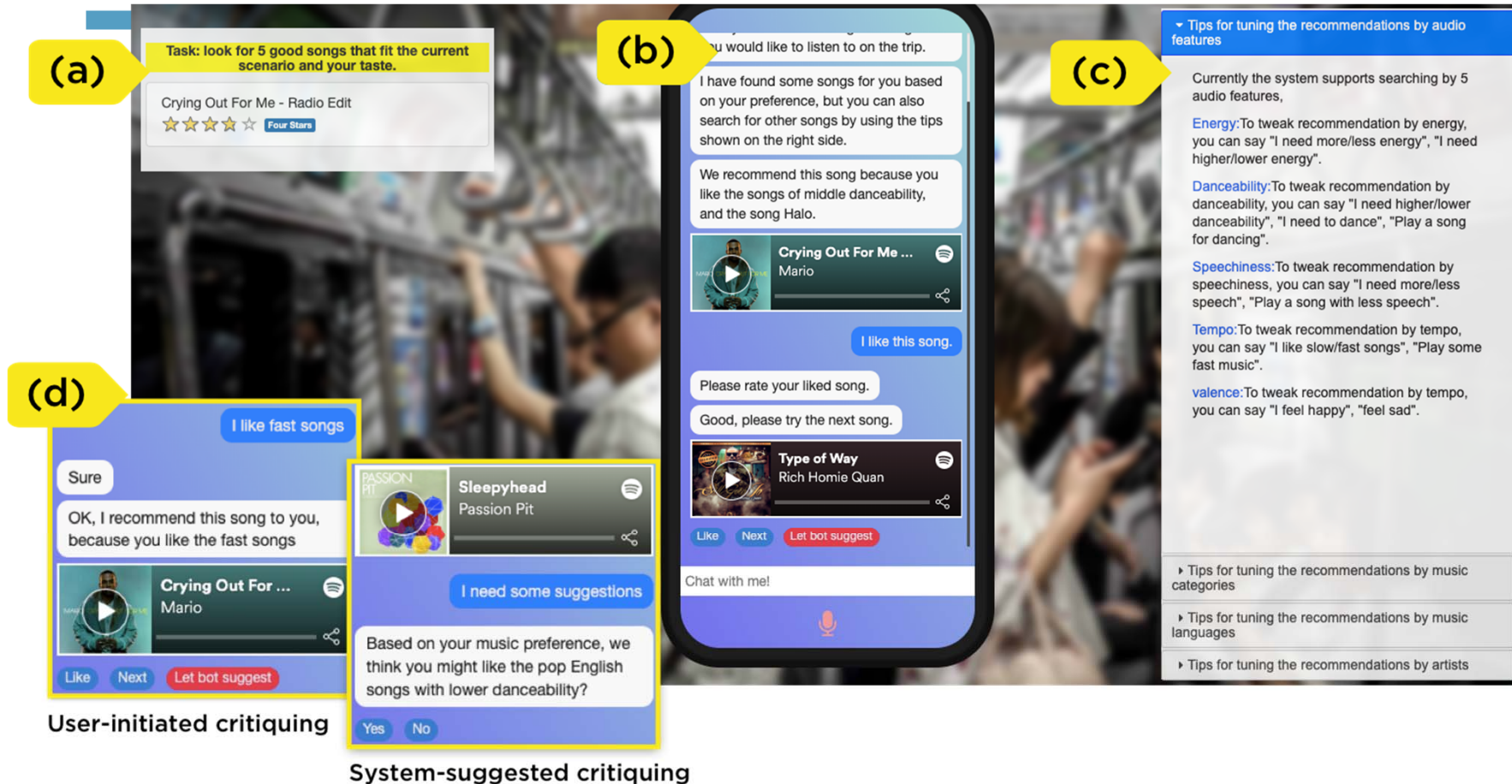
Application Environment

- Stand-alone vs. Embedded
 - Stand-alone
 - Recommendation is the central functionality of the system
 - E.g., mobile tourist guides, interactive e-commerce advisory systems
 - Embedded
 - Providing recommendations is only one of many functionalities the device is capable of
 - E.g., the use of a CRS on voice-based home assistants

Interaction Initiative

- System-driven
 - E.g., critiquing-based systems, form-based interactive advisory systems
- User-driven
 - “User-asks, system-responds”
 - Rarely entirely user-driven
- Mixed-initiative
 - Most of existing CRS

Mixed-initiative: Music Chatbot



Cai, W., Jin, Y., and Chen, L.: Critiquing for Music Exploration in Conversational Recommender Systems. In: Proceedings of 26th International Conference on Intelligent User Interfaces (UII'21), College Station, TX, USA, April 14–17, 2021.

Discussion

- Natural language interaction vs. form-based inputs
 - Pure natural language interfaces in principle provide the opportunity to elicit preferences in a more natural way.
 - Users might be better acquainted and feel more comfortable with more traditional interaction mechanisms (forms and buttons).
 - A mix of a natural language interface and buttons led to the best user experience (Iovine et al., 2020).

A. Iovine, F. Narducci, and G. Semeraro. Conversational recommender systems and natural language: A study through the Converse framework. *Decision Support Systems*, 131:113250–113260, 2020.

Discussion, cont.

- New application scenarios
 - Interactive wall
 - Service robot
 - In-car
 - Etc.
- General challenges
 - Privacy considerations
 - Aspects of technology acceptance
 - Application-specific ones (e.g., safety considerations in an in-car setting)

Agenda

1. Introduction
2. Characterization of Conversational Recommender System (CRS)
3. Interaction Modalities
- 4. Underlying Knowledge and Data**
-- Short break --
5. Computational Tasks
6. Evaluation of CRS
7. Wrap-up & Discussion

Types of Knowledge and Data

- Supported user intents
 - Users' particular information needs and intents that can occur in conversations
 - Pre-defined or automatic detection (e.g., in NLP-based systems)
- User modeling
- Dialogue states
- Background knowledge

High-level Overview

A high-level overview of domain-independent user intents

Intent Name	Intent Description
<i>Initiate Conversation</i>	Start a dialogue with the system.
<i>Chit-chat</i>	Utterances unrelated to the recommendation goal.
<i>Provide Preferences</i>	Share preferences with the system.
<i>Revise Preferences</i>	Revise previously stated preferences.
<i>Ask for Recommendation</i>	Obtain system suggestions.
<i>Obtain Explanation</i>	Learn more about why something was recommended.
<i>Obtain Details</i>	Ask about more details of a recommended object.
<i>Feedback on Recommendation</i>	Give feedback on the provided recommendation(s).
<i>Restart</i>	Restart the dialogue.
<i>Accept Recommendation</i>	Accept one of the recommendations.
<i>Quit</i>	Terminate the conversation.

Hierarchical Taxonomy of User Intents

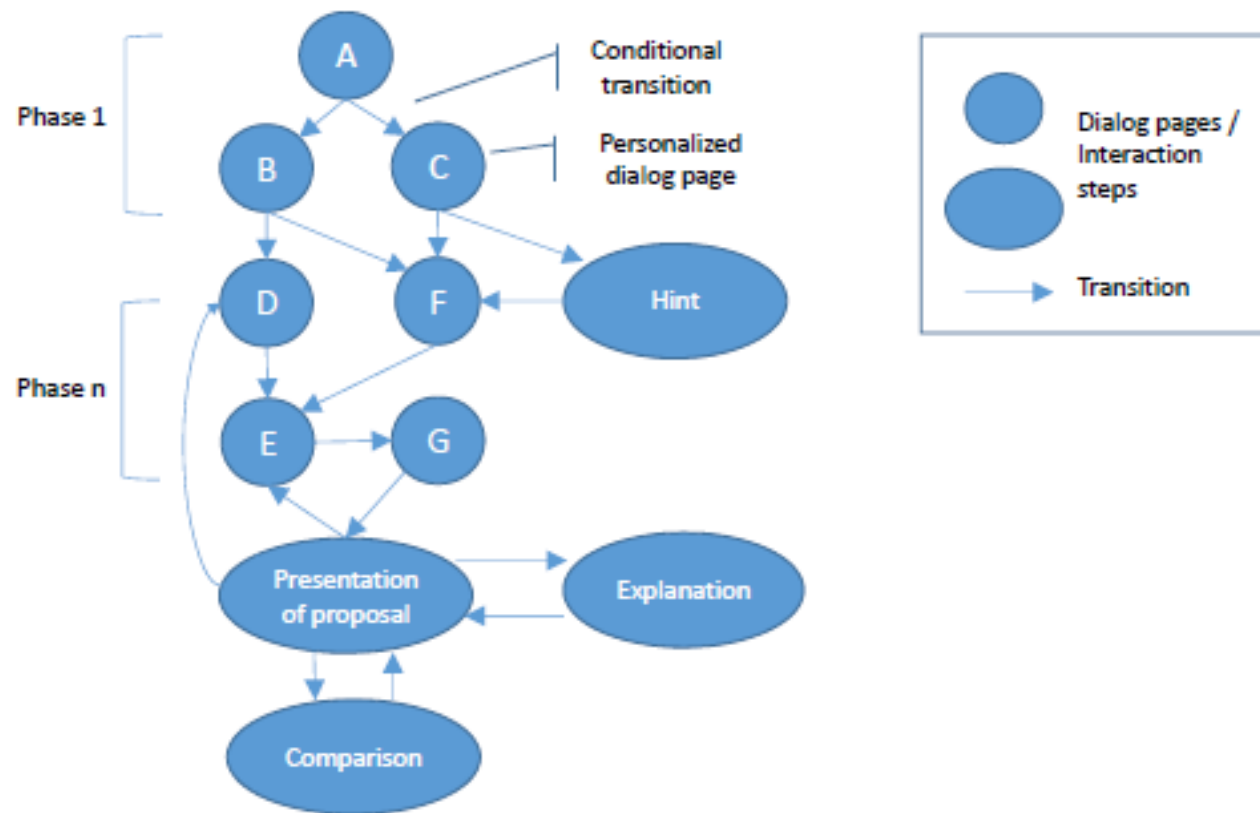
Intent (Code)	Description	Percentage
Ask for Recommendation		18.26%
Initial Query (IQU)	Seeker asks for a recommendation in the first query.	12.91%
Continue (CON)	Seeker asks for more recommendations in the subsequent query.	3.10 %
Reformulate (REF)	Seeker restates her/his query with or without clarification/further constraints.	1.50%
Start Over (STO)	Seeker starts a new query to ask for recommendations.	0.84%
Add Details		18.58%
Provide Preference (PRO)	Seeker provides specific preference for the item s/he is looking for.	12.30%
Answer (ANS)	Seeker answers the question issued by the recommender.	4.91%
Ask Opinion (ASK)	Seeker asks the recommender's personal opinions.	2.39%
Give Feedback		61.92%
Seen (SEE)	Seeker has seen the recommended item before.	21.14%
Accept (ACC)	Seeker likes the recommended item.	18.89%
Reject (REJ)	Seeker dislikes the recommended item.	11.50%
Inquire (INQ)	Seeker wants to know more about the recommended item.	6.55%
Critique-Feature (CRI-F)	Seeker makes critiques on specific features of the current recommendation.	6.50%
Critique-Add (CRI-A)	Seeker adds further constraints on top of the current recommendation.	5.35%
Neutral Response (NRE)	Seeker does not indicate her/his preference for the current recommendation.	4.29%
Critique-Compare (CRI-C)	Seeker requests sth similar to the current recommendation in order to compare.	1.55%
Others	Greetings, gratitude expression, or chit-chat utterances.	14.55%

Cai, W., and Chen, L.: Predicting User Intents and Satisfaction with Dialogue-based Conversational Recommendations. In Proceedings of 28th Conference on User Modeling, Adaptation and Personalization (UMAP'20), pages 33–42, July 14-17, 2020.

User Modeling

- User profile
 - Preference expressions or estimates regarding **individual items**, e.g., ratings, like and dislike statements
 - Preferences regarding **individual item facets**, e.g., the genre of a movie or the desired functionalities
- Long-term preference
 - Long-term and supposedly more stable preferences (e.g., for non-smoking rooms in restaurants) from multiple sessions

Dialogue States: Pre-defined



Jannach, D., and Kreutler, G.: Rapid development of knowledge-based conversational recommender applications with Advisor Suite. *Journal of Web Engineering*, 6(2):165–192, June 2007.

Dialogue States: NLP-based Systems

- Mainly two phases in NLP-based conversational preference elicitation systems
 - Phase 1: Asking questions
 - Phase 2: Presenting a recommendation list
- Implicit states
 - Based on the implemented intents
 - **Or** encoded in a complex neural model as trained on a corpus of recorded human conversations, e.g., in the end-to-end learning CRS

Background Knowledge: Item-related Knowledge

Domain	Description
Movies	Traditional movie rating databases from MovieLens, EachMovie, Netflix, used for example in [73, 166, 166].
Electronics	A product database with more than 600 distinct products was collected from various retailers [46].
	A smartphone database consisting of 1721 products with multiple features [34].
	Amazon electronics review dataset ⁵ containing millions of products, user reviews and product meta-data [164].
	A dataset consisting of 120 personal computers, each with 8 features [128].
Travel	More than 100 sightseeing spots in Japan with 25 different features [52].
	A database of restaurants in the San Francisco area covering 1,900 items with multiple features like cuisine, ratings, price, location, or parking [135].
	Search logs and reviews of 3,549 users of a restaurant review provider, focusing on locations in Cambridge [29].
	A travel destinations dataset, crawled from online platforms containing 5,723,169 venues in 180 cities around the globe [53].
	A restaurants dataset crawled for Dublin city, which consists of 632 restaurants with 28 different features [89].
Food Recipes	A food recipe dataset containing dishes and their ingredients [162].
E-commerce	A product database of 11M products and logged data from the search engine of an e-commerce website was collected. The logged data consists of 3,146,063 unique questions [156].
Music	A music dataset crawled from multiple online sources, containing 2,778 songs with 206k explanatory statements and 22 user tags [165].

Background Knowledge: Dialogue Corpora

Domain	Name	Description
Movies	ReDial	Crowdworkers from Amazon Mechanical Turk (AMT) were used to collect over 10,000 dialogues centered around the theme of providing movie recommendations [73]. A paired mechanism was used where one person acts as a <i>recommendation seeker</i> and the other as a <i>recommender</i> .
	CCPE-M	A Wizard-of-Oz (WoZ) approach ⁶ is taken to elicit movies preferences from crowdworkers within natural conversations. The dataset consists of over 500 dialogues that contain over 10,000 preference statements [111].
	GoRecDial	This dataset consists of 9,125 dialogue interactions and 81,260 conversation turns collected through pairs of human workers; here also one plays the role of a movie seeker and the other as a recommender [63].
	bAbI	In [95], the authors used a general movie dialogue dataset provided by Facebook Research [39] to build a CRS. The dataset contains task-based conversations in a question-answering style. It consists of 6,733 and 6,667 dialogue conversations for training and testing respectively.
Restaurants and Travel	CRM	An initial dataset containing 385 dialogues is collected using a pre-defined dialogue template through AMT [133]. Using this dataset, a larger synthetic dataset of 875,721 simulated dialogues is created.
	ParIAI	A goal-oriented, extended version of the bAbI dataset that was collected using a bot and users. It consists of three datasets (training, development and testing), each comprising 6,000 dialogues. [62].
	MultiWOZ	A large human-human dialogue corpus, which covers 7 domains and consists of 8,438 multi-turn dialogues around the themes of travel & planning recommendation [154].
Fashion	MMD	A dataset consisting of 150,000 conversations between shoppers and a large number of expert sales agents is collected. 9 dialogue states were identified in the resulting dataset [123].
Multi-domain	OpenDialKG	A dataset of chat conversations between humans is collected, consisting of 15,000 dialogues and 91,000 conversation turns on movies, books, sports, and music [93].

Background Knowledge: Logged Interaction Histories

Domain	Description
Movies	A dialogue dataset involving 347 users was collected in [64] during the experimental evaluation of a recommender system.
	A subset of the ReDial dataset was analyzed and annotated in [16] to classify the user feedback types in 200 dialogues at the utterance level.
	A dialogue corpus was collected in [146] for the purpose of dialogue quality analysis from experimental session logs consisting of 226 complete dialogue turns with 20 users.
	A user study was conducted in [147], where a <i>movie seeker</i> and a <i>human recommender</i> converse with each other. The dialogue corpus consists of 2,684 utterances and 24 complete dialogues.
Travel	A dataset containing preferences for hotel, flight, car rental searches was collected in [4] involving 200 users of a content-based recommender system that supports multiple tasks (i.e., hotel, car, flight booking) in the same dialogue.
Fashion	A user study was conducted using a virtual shopping system, where the goal was to find a wedding dress. A non-verbal feedback (e.g., gestures, facial expressions, voices) dataset involving 345 subjects was collected and then annotated for model training [18].
E-commerce	A dataset containing conversation logs of users with a chatbot of an online customer service center (Alibaba.com) was collected in [109]. It consists of over 91,000 Q&A pairs as a knowledge base used for the information retrieval task.

Background Knowledge: Lexicons and World Knowledge

Source Name	Description
Wikipedia	A dataset crawled from online sources (Wikipedia and Wikitravel) for the purpose of entity recognition in the travel domain [71].
WordNet	WordNet ⁷ is used in order to compute the semantic distance between entities and keywords mentioned in the conversation [71, 78].
Wikiquote	A quote dataset crawled from two online sources, Wikiquote ⁸ and the Oxford Concise Dictionary of Proverbs [68].
Citysearch	In [78], a dataset of 137,000 users reviews on 24,000 restaurants was harvested from two online sources (Citysearch ⁹ and MenuPages ¹⁰) to generate a dictionary of mappings between semantic representations of cuisines and dialogue concepts.

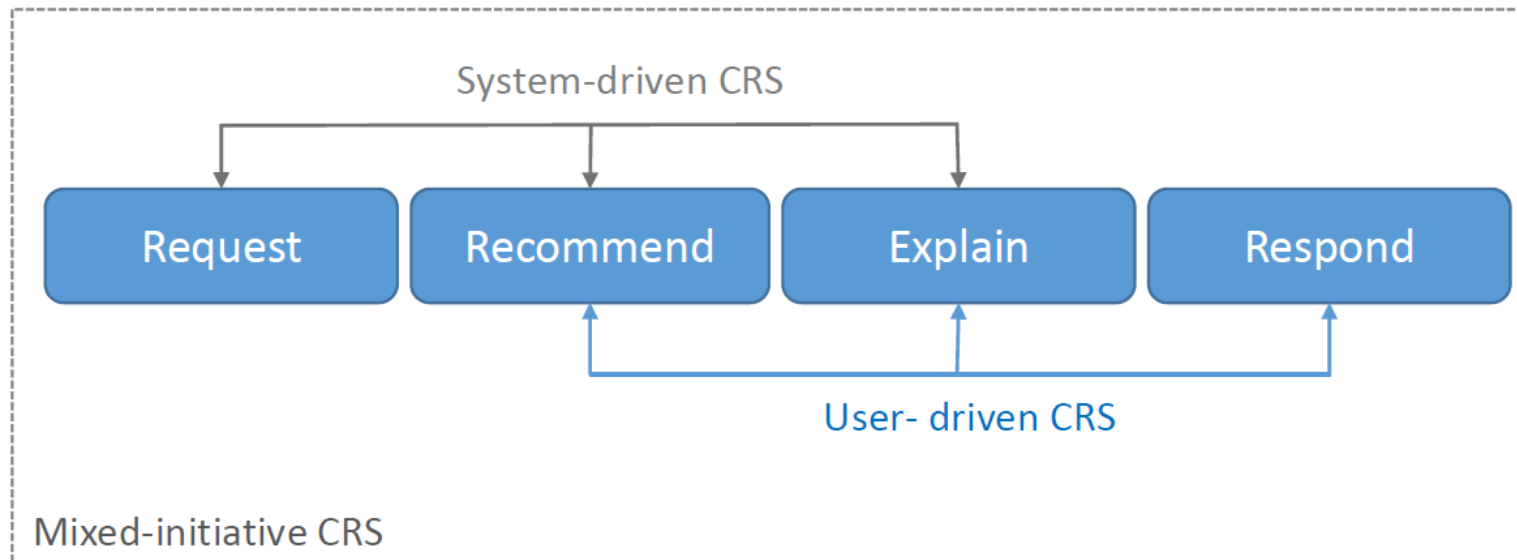
Discussion

- Pre-defined knowledge vs. Learning approaches
 - Form-based interaction
 - Typically pre-defined in terms of possible dialogue states, supported user intents, or user profile attributes to acquire
 - NLP-based interaction
 - More dynamic in terms of the possible dialogue flow, and relying on additional knowledge sources
- Intent engineering
 - The set of supported intents determines how rich and varied the resulting conversations can be
 - *Challenge*: how to anticipate or learn over time which intents the users might have

Agenda

1. Introduction
2. Characterization of Conversational Recommender System (CRS)
3. Interaction Modalities
4. Underlying Knowledge and Data
- Short break --
- 5. Computational Tasks**
6. Evaluation of CRS
7. Wrap-up & Discussion

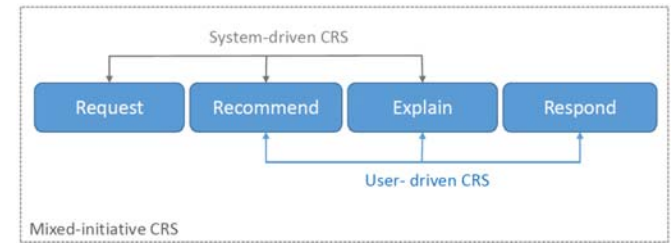
Main Computational Tasks



- A conceptual, generic view
 - Not all tasks (explicitly) supported in all systems

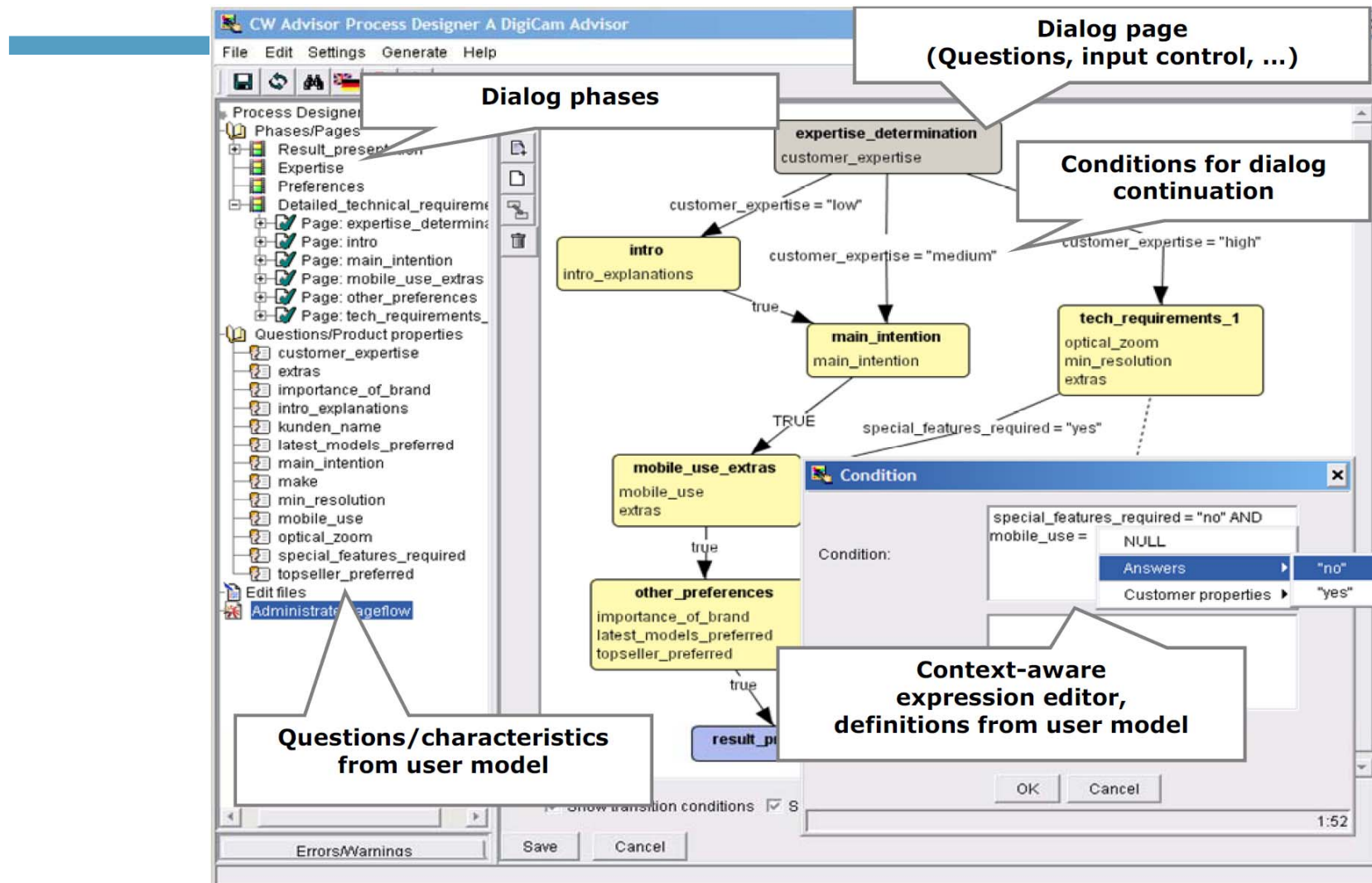
Main Tasks:

Request (examples)

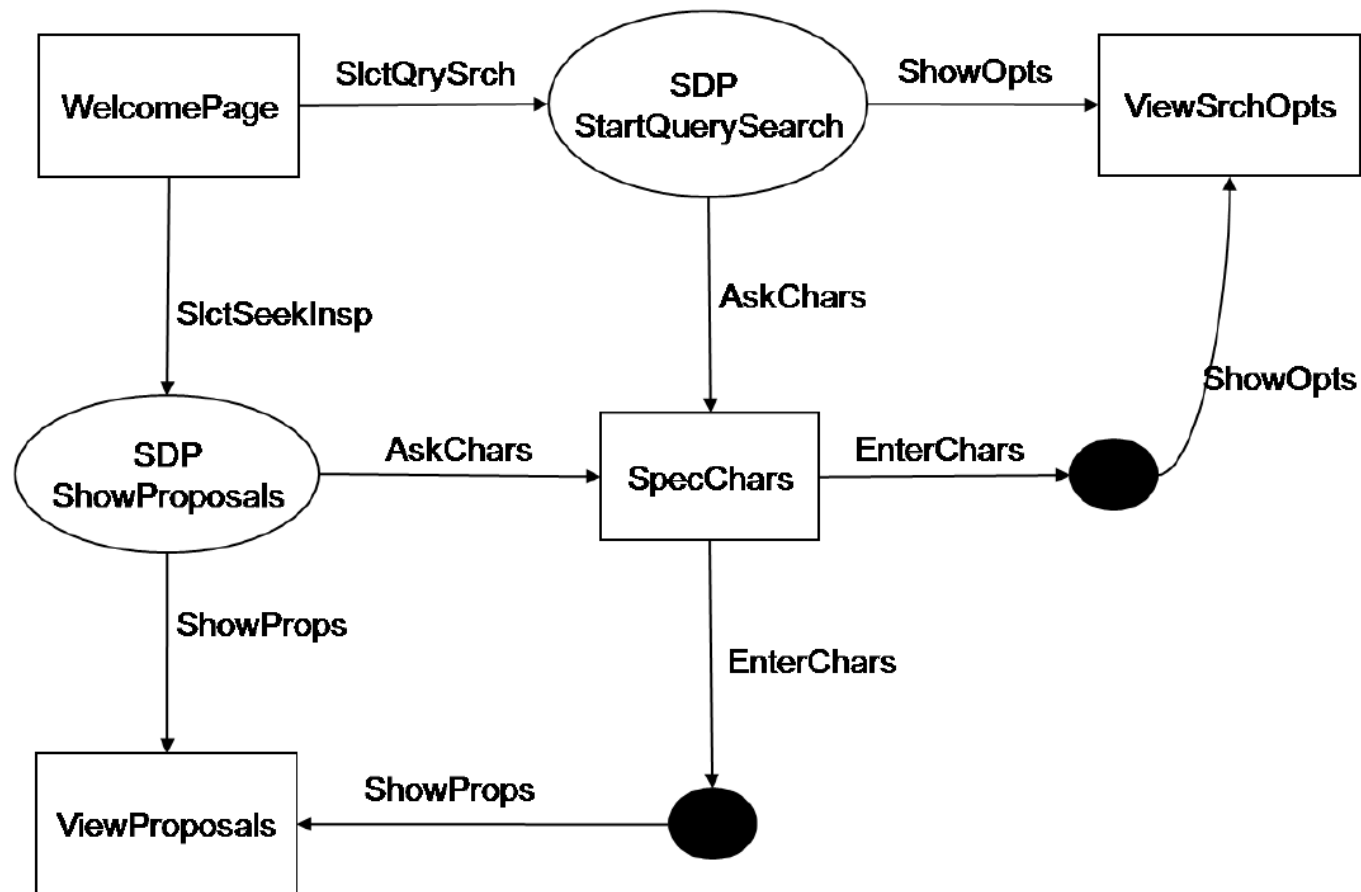


- Slot-filling
 - A common approach to acquire preferences
 - Determine next question to ask
 - Questions often relate to desired item features/attributes
 - Computation may be based, e.g., on set of remaining items in critiquing approaches, to reduce interaction time
- Reasoning about dialogue state
 - Determine next “conversational move” (e.g., ask more, explain, recommend)
 - e.g., based on reinforcement learning / bandit approaches
 - Or based on explicit rules

Request: A rule-based Approach

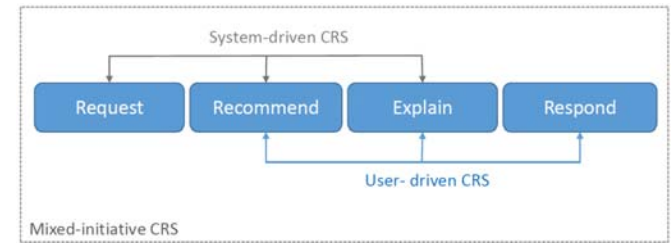


Request: Learning the next move



Main Tasks:

Recommend



- Various technical approaches possible:
 - Collaborative, content-based, hybrid, and in particular: **knowledge-based**
 - Knowledge-based:
 - Match elicited feature preferences with available items
 - Constraint-based, case-based, utility-based
 - Often only based on **short-term** preferences acquired during the usage session
 - Some models combine long-term and short-term user models

Main Tasks: Recommend

- Often only one recommendation at a time (in contrast to traditional recommenders)

And Chill

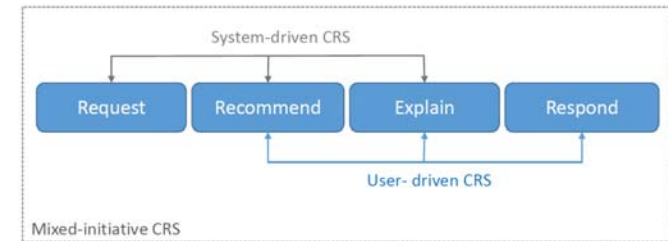
You should try Band Of Angels!



Band of Angels -
Trailer

Main Tasks:

Explain



- Explanations are widely recognized as a trust-building feature of recommenders
 - Little research in CRS, though
 - Practical systems also have their limitations

why do you recommend band of angels

Gotcha... that title, right? There's only 1.8% of users that have mentioned this one. I am still generating AI-powered movie recs based on this film. Can you tell me about a different one you like?



Explain: Trust-inspiring interfaces

- Critiquing-based Recommendation

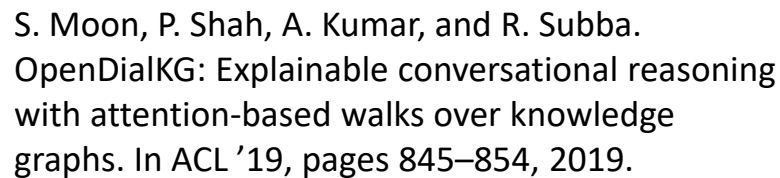
The most popular product								
	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
Ⓜ	—	\$2'095.00	1.67 GHz	4.5 hour(s)	512 MB	80 GB	38.6 cm	2.54 kg
We also recommend the following products because								
they are cheaper and lighter, but have lower processor speed								
	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
Ⓞ	—	\$1'499.00	1.5 GHz	5 hour(s)	512 MB	80 GB	33.8 cm	1.91 kg
Ⓞ	—	\$1'739.99	1.5 GHz	4.5 hour(s)	512 MB	80 GB	38.6 cm	2.49 kg
Ⓞ	—	\$1'625.99	1.5 GHz	5 hour(s)	512 MB	80 GB	30.7 cm	2.09 kg

Explain: User-tailored Descriptions

- Varying levels of conciseness

User	Z value	Output
CK	0.3	Bond Street has the best overall value among the selected restaurants. Bond Street has excellent food quality.
BA	0.3	Komodo has the best overall value among the selected restaurants. Komodo's price is \$29. It's a Japanese, Latin American restaurant.
VM	0.3	Komodo has the best overall value among the selected restaurants. Komodo's price is \$29 and it has very good food quality.

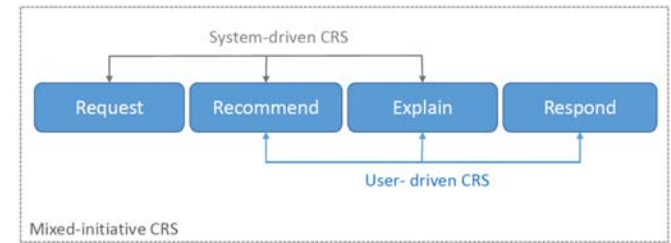
- (a) **Dialog**
- 1 Can you recommend any classic books like *Catcher in the Rye*?
- 2 Do you prefer books by the same **author** or same **genre**?
- 3 I am interested in reading classic examples of *American literature*.
- 4 *Literary realism* is a common genre in classic *American literature*.
- 5 Do you prefer *First-person* or *Third-person narrative*?
- 6 I mostly prefer *third-person narrative*.
- 7 Consider reading the *Scarlet Letter*: a novel by *Nathaniel Hawthorne*.
- ⋮



S. Moon, P. Shah, A. Kumar, and R. Subba.
OpenDialKG: Explainable conversational reasoning
with attention-based walks over knowledge
graphs. In *ACL '19*, pages 845–854, 2019.

Main Tasks:

Respond

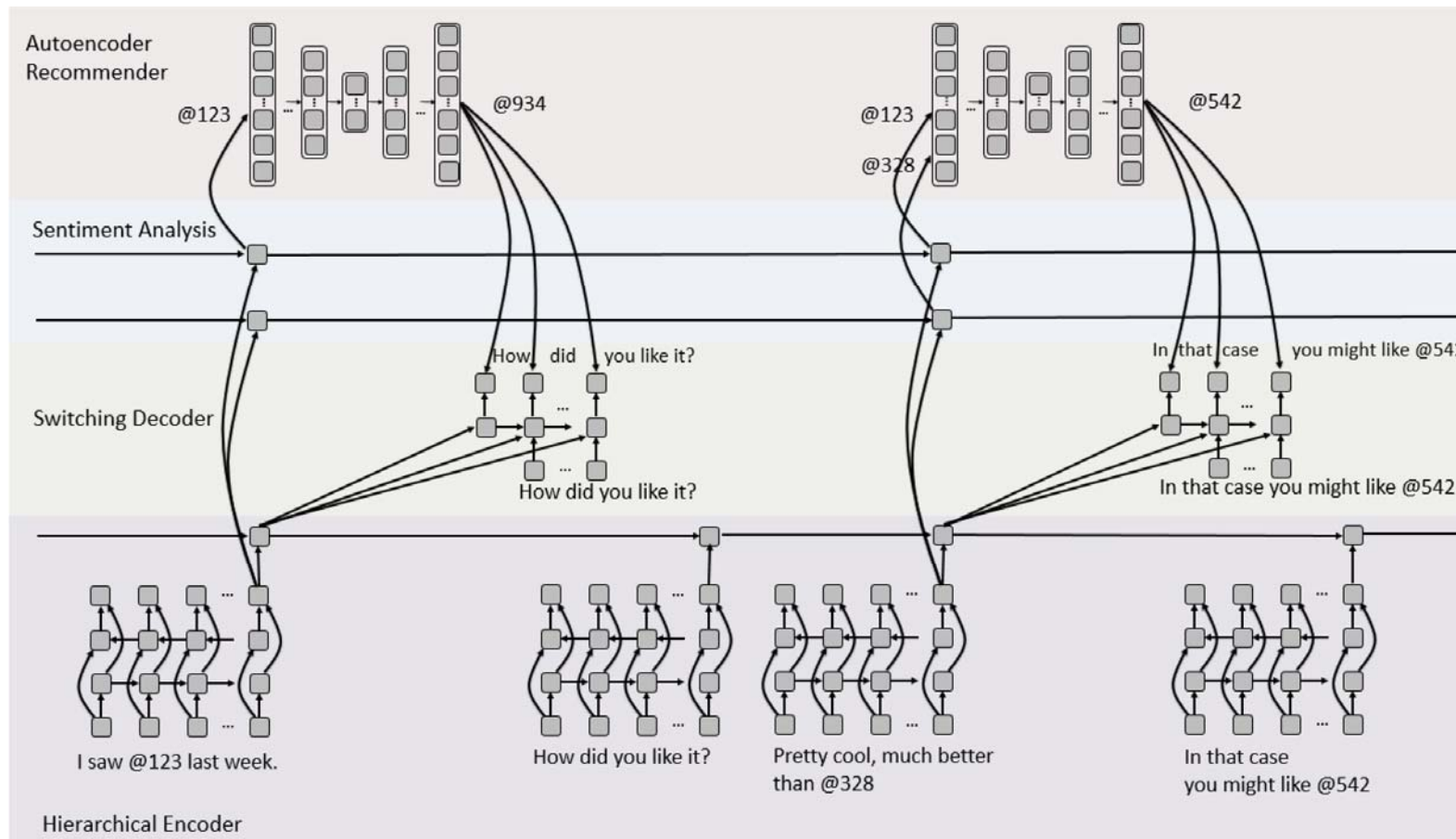


- A generic category
 - Relevant in user-driven and mixed-initiative CRS
 - React when user proactively leads the dialogue
- Properly respond to certain actions, e.g.,
 - when the user states a preference without being asked, or refines the requirements
 - e.g., “I like Pulp Fiction, but not Quentin Tarantino”
 - when user wants to restart the dialogue
 - e.g., because of a conversational breakdown
 - when the user asks for more information
 - e.g. “How about Huawei P9?”

Main Tasks: Respond

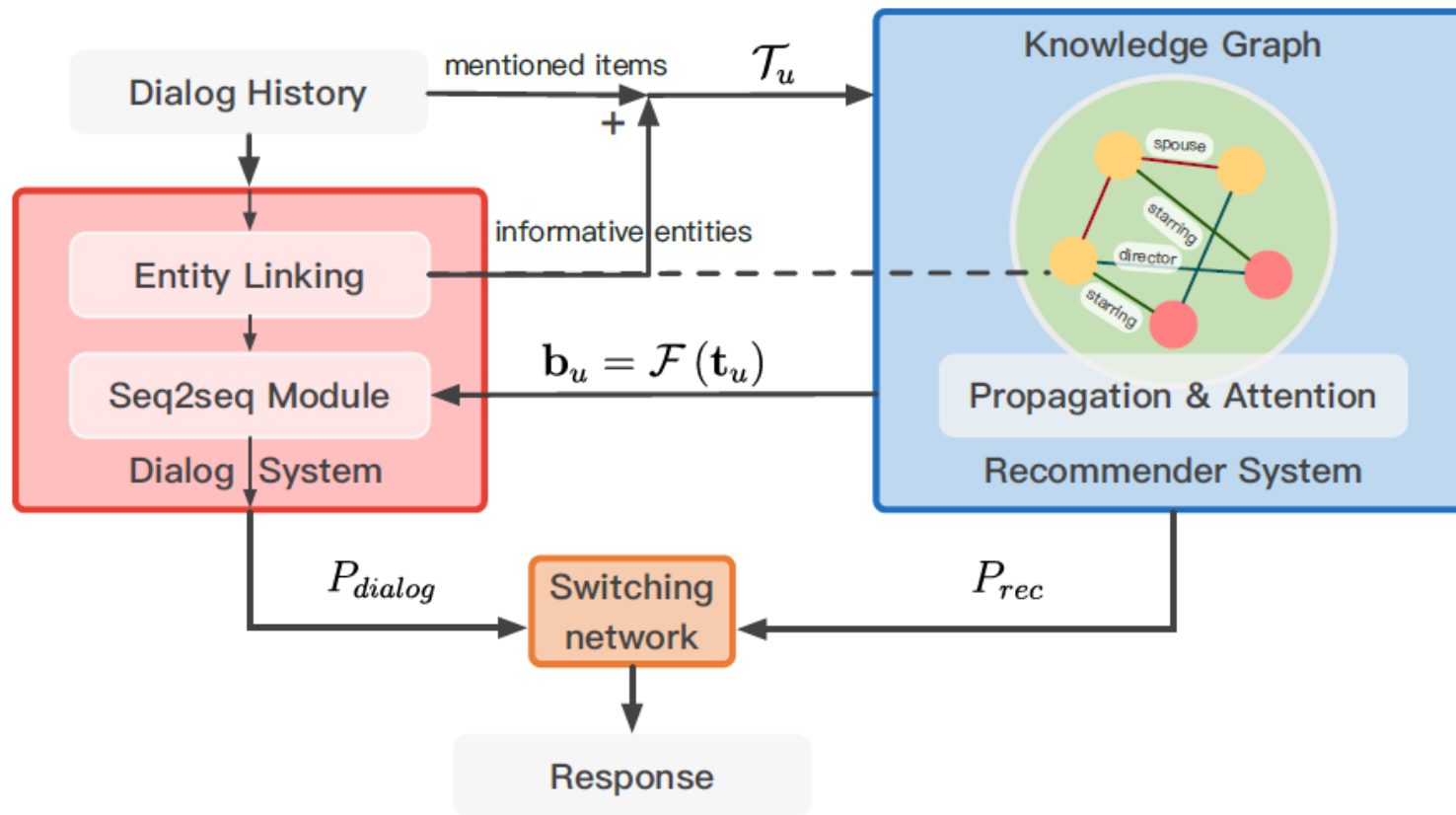
- Alternative technical approaches
- 1) Intent-detection / mapping
 - Based on a pre-defined list of supported user intents and an engineered or learned mapping logic
- 2) End-to-end learning
 - Automatically learn how to respond based on training data consisting of recorded dialogues
 - Retrieval and generation-based approaches; often using deep learning

Respond: Deep Learning



R. Li, S. E. Kahou, H. Schulz, V. Michalski, L. Charlin, and C. Pal. Towards deep conversational recommendations. In NIPS '18, pages 9725–9735, 2018.

Respond: Deep Learning



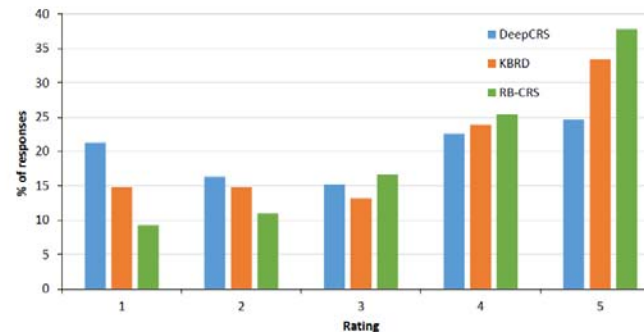
Respond: Retrieval-based Methods

- Language-generation-based techniques
 - Dominate the research landscape
 - Have the potential to creatively respond to user utterances
- Retrieval-based approaches
 - Idea is to retrieve (and adapt) existing utterances from a given corpus, given a current query
 - Potential plus: Retrieved sentences are semantically and logically correct
 - Not much research done

Respond: Retrieval-based Methods

- Recent user study
 - Comparing DeepCRS, KBRD, and simple retrieval-based system
- Let users (crowdworkers) assess different system responses

	DeepCRS	KBRD	RB-CRS
Avg. score	3.13	3.46	3.71
Std. deviation	1.49	1.45	1.32



A. Manzoor and D. Jannach. Generation-based vs. Retrieval-based Conversational Recommendation: A User-Centric Comparison. In RecSys '21, 2021

User Study

Dialog Situation

Situation#: 1

CHAT-BOT: Hello

USER: Hi. Can you recommend me some action movies?

CHAT-BOT: I recently saw "Black Panther (2018)" in the theater, it was great. "Taken (2008)", "Die Hard (1988)" and "The Matrix (1999)" are others I've really enjoyed

USER: Those sound really interesting. What are your favorite comedies?

CHAT-BOT: ...?

What should be the next 'CHAT-BOT's response ?

Please rate the following three chat-bot responses in the given dialog situation

Response 1

Or "The Grand Budapest Hotel (2014)"

Select Rating

Response 2

have you seen "john wick : chapter 2" ?

Select Rating

Response 3

i have n't seen that one . i will have to check it out .

Select Rating

Submit ratings

Agenda

1. Introduction
2. Characterization of Conversational Recommender System (CRS)
3. Interaction Modalities
4. Underlying Knowledge and Data
- Short break --
5. Computational Tasks
- 6. Evaluation of CRS**
7. Wrap-up & Discussion

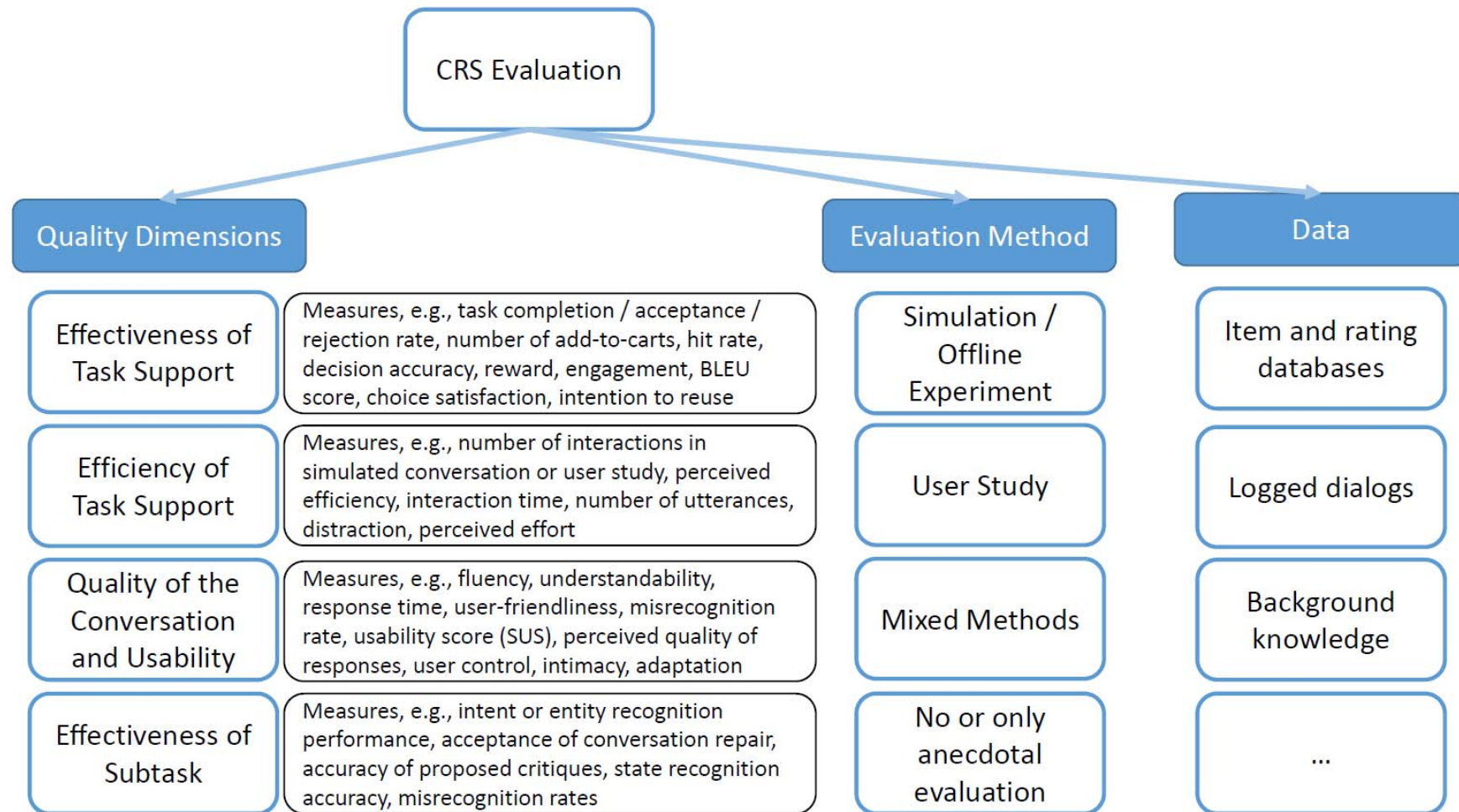
Evaluation

- General evaluation paradigms recommenders
 - Experimental research with users
 - Live (A/B) test of a deployed system
 - Controlled user studies with treatment and control groups
 - Research without users / simulations
 - “Offline” experiments regarding prediction or classification accuracy (e.g., for recommendations, entity recognition , etc.)
 - Non-experimental research
 - Quasi-experiments
 - Observational research
 - Qualitative research

Evaluation Dimensions

- Effectiveness of Task Support
 - To what extent does the system help users find something new or make good decisions?
- Efficiency of Task Support
 - To what extent does the system help users make decisions faster (without compromising their choice satisfaction etc.)?
- Quality of the Conversation and Dialogue, Usability
 - For example, is the conversation natural? Does the system respond quickly and fluently?
- Effectiveness of Subtask
 - For example, how accurate is the entity recognition module?

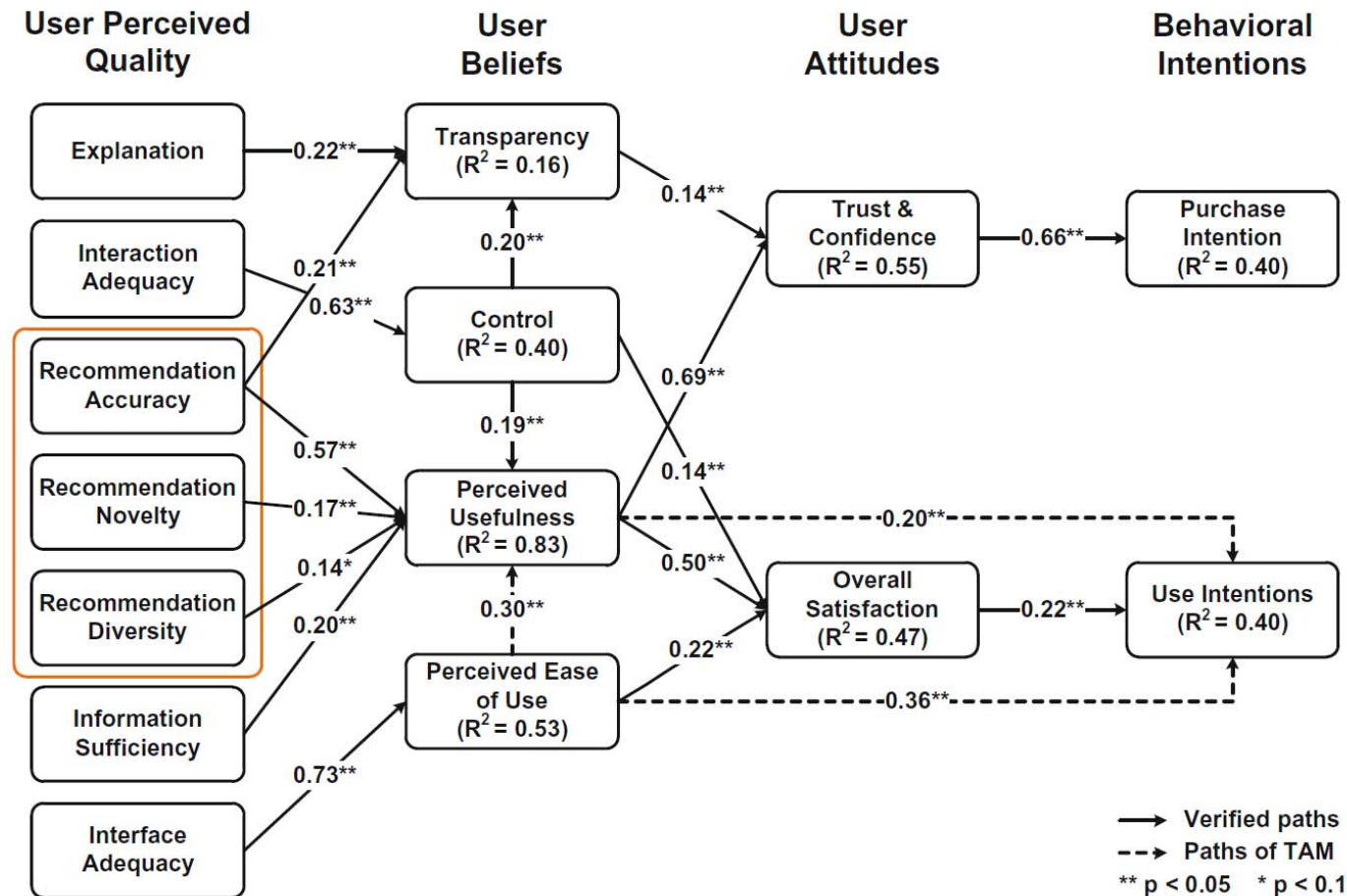
Overview of Evaluation Dimensions and Methods



Effectiveness of Task Support

- Objective Measures
 - **Offline:** Accuracy measures (RMSE, MAE, Precision, Recall, NDCG, MRR), proportion of simulated users accepting a recommendation
 - **With users:** Task success, adoption of recommendations, if users changed their mind afterwards
- Subjective Measures
 - Often based on the [ResQue](#) model, e.g., perceived accuracy, attractiveness, novelty, diversity, context compatibility, ...

Effectiveness: Beyond Recommendation Quality



Pu, P., Chen, L. & Hu, R. Evaluating recommender systems from the user's perspective: survey of the state of the art. User Model User-Adap Inter 22, 317–355 (2012)

Efficiency of Task Support

- Objective Measures
 - Number of interaction cycles (dialogue turns)
 - Often used in simulations and user studies, but not so frequently in language-based interaction approaches
 - Task completion time (time to make a decision)
 - Assumption is that shorter interactions are desirable; but may not always be the case
- Subjective Measures
 - Perceived effort, cognitive effort
 - Often as part of usability assessments

Quality of the Conversation

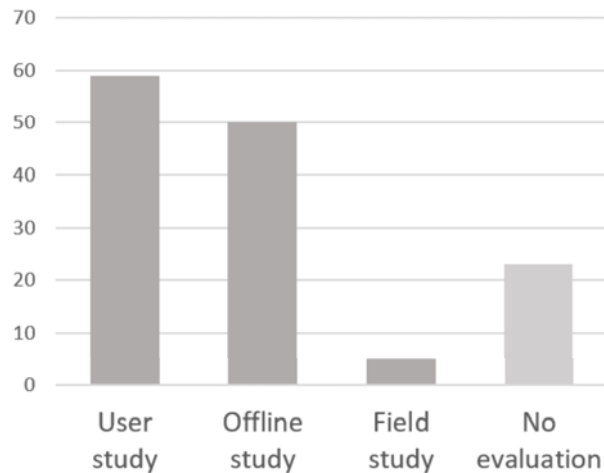
- .. and more usability
- Objective Measures
 - Linguistics: BLEU and NIST scores, perplexity, fluency
- Subjective Measures
 - Ease-of-use, task-ease, perceived level of user control, perceived transparency, adaptation of the system, expected behavior (intuitiveness), entertainment, mutual attentiveness, positivity, fluency, engagingness, overall dialogue quality, generation performance, ...

Effectiveness of Subtask

- Objective Measures
 - Regret/reward for RL-based selection of next action
 - Number of applied critiques (of those generated and proposed)
 - Entity and intent recognition accuracy
- Subjective Measures
 - Interpretation performance

Evaluation: Discussion

- Wide range of methods applied
- No standards exist, often only offline
 - Survey of papers



- Correspondence of offline metrics with user perceptions sometimes unclear (see BLEU)

Agenda

1. Introduction
2. Characterization of Conversational Recommender System (CRS)
3. Interaction Modalities
4. Underlying Knowledge and Data
- Short break --
5. Computational Tasks
6. Evaluation of CRS
- 7. Wrap-up & Discussion**

Future Directions, examples

- Modalities
 - Not entirely clear which modality is the best in a context
 - Understanding non-verbal communication acts
- Mixed-method Evaluation, Exploratory Research
 - Need better understanding of what users expect and how recommendation dialogues should be structured
- Explanations
 - Not much explored yet, but might be expected
- Business Value of CRS
 - Almost no studies

Thanks for the attention

- Time for questions
- Contact
 - dietmar.jannach@aau.at
 - lichen@comp.hkbu.edu.hk

