NPFL123 Dialogue Systems

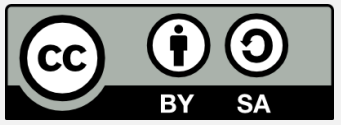
9. Dialog Authoring Tools

https://ufal.cz/npfl123

Ondřej Dušek, Patrícia Schmidtová, Vojtěch Hudeček & **Jan Cuřín**

17. 4. 2023

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



Short Intro of Jan Cuřín

Education

- IFAL, MFF UK – PhD in 2006 (Statistical Machine Translation)

Work Experience

- MAMA AI – 2021- Co-Founder, VP for NLP
  - Natural Language Processing, Conversational Agents, AI
  - https://themama.ai
- IBM – 2004-2021 – Research Scientist, Manager at IBM Watson R&D Lab
  - Conversational Systems, NLU Technologies, AI
- IFAL, MFF UK – 2002-2004 – Researcher, PhD Student
  - Machine Translation, NLU Technologies
- Schemantix – 2000-2001 – Software Engineer
  - Machine Translation, XML Technologies

IBM Watson Services
- Watson Assistant
- Watson Speech (STT/TTS)
- Watson Language Translator

Business Use Of Dialogue Systems

- Chat bots on web pages
  - Navigation through the content of the web pages (smart search)
  - Frequently asked question (FAQ)
- Mobile applications with open text input
  - Domain specific apps with chatting functionality, even banking apps
  - Intra company systems
- Assisting systems
  - Intra company “expert” system
  - Support for human operators
- Speech based systems – Voice Bots
  - Call centers automation – handling top x% of traffic
  - Outbound calls (to inform or collect feedback)
- Automotive applications
  - Search, calls, navigation, infotainment/entertainment (music, POIs)
- Infotainment systems
  - Infotainment systems for hotels, banks’ lobbies, home, games, VR etc …
- Healthcare/Society domain
  - Buddy to talk to, training buddy, elderly care

Challenges of Creating Good DS

- Data
  - Collection of human-to-human communication
  - Intracompany structured and unstructured data
  - No data, just ideas
  - No idea at all

- Scenarios, use-cases
  - Single domain
  - Single domain with chit-chat capability
  - Multi-domain
  - Open-domain

Authoring a Dialogue

Restaurant booking scenario

- System: Hello, this is Chez Pépé restaurant reservation system. How may I help you?
- User: I would like to reserve a table for tomorrow for 5 people
- System: From what time?
- User: From 7pm
- System: OK, I will make a reservation for tomorrow at 7pm, table for 5.
- User: Great!
- System: Looking forward to see you soon.

Authoring a Dialogues

Restaurant booking scenario

Intents

#reserve_table
- I would like to reserve a table for 5
- Can I make a reservation?
- I need a reservation for tomorrow

#opening_hours
- Until when are you open?
- What are the opening hours?

#cancel_reservation
- I made my reservation yesterday, I want to cancel it.
- We could not make it today, may I cancel the reservation?

Entities

I need reservation for 5 people for tomorrow at 7pm

Next Friday from 6pm

For two.

@date
@time
@number
@restaurant_location

Dialogue

Welcome
Book a table

Entity	Context variable	Req.
@date	$res_date	Y
@time	$res_time	Y
@number	$guests	Y

Opening hours
Cancel reservation
- Yes
- No

<default answer>

Intents

- Collection of example how users will trigger the intent
- Usually corresponds to the actions supported by the dialog

- Intent model can be trained even on a small set of examples
- Word and sentence embeddings, stemmer, lemmatizer
- Bigger data collection needed for production system

- Ordered n-best lists with confidences
- Use of intent n-bests in the dialog – disambiguation

Intents

#reserve_table
I would like to reserve a table for 5
Can I make a reservation?
I need a reservation for tomorrow

#opening_hours
Until when are you open?
What are the opening hours?

#cancel_reservation
I made my reservation yesterday,
I want to cancel it.
We could not make it today,
may I cancel the reservation?

Entities

~ Named entities recognition (NER)

- Different type of entities
  - Prebuilt (system) entities
    - Numbers, dates, time, GEO location, person names, units, currency
  - Domain catalogues
  - User defined entities
    - Gazetteers – fixed list of entities/synonyms
    - Regular expression based
    - Sequence labelling model based on sample annotations (contextual entities)

Entities

I need reservation for 5 people for tomorrow at 7pm

Next Friday from 6pm

For two.

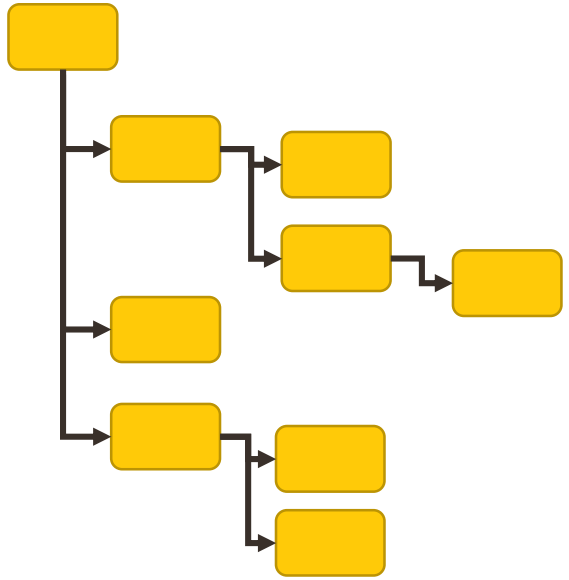
@date
@time
@number
@restaurant_location

Dialogue Flow/Tree

- Slot filling style (linear dialog)
  - Set of slots to fill is (required/optional)
  - Able to fill all slots partially or at once
  - Asking just for missing information
  - Ability to customize questions and answer for a particular slot
  - Ability to correct already filled information
  - Tight to user variables

Entity	Context variable	Req.
@date	$res_date	Y
@time	$res_time	Y
@number	$guests	Y

- Dialogues tree (non-linear dialog)
  - Dialogue flow driven by a tree or graph structure
  - Conditions to get to the individual nodes of the tree/graph
  - Fallback strategies (none of the conditions is specified)



Demo

- Sample chatbot in Watson Assistant

Restaurant booking scenario

http://www.bienvenuechezpepe.com/

Features used in runtime

- Dialogue context / history
  - Condition on context variables collected in previous turns
  - Reference/anaphora resolution using collected variables
- Fallback strategies / Digression
  - Allow "jumping" to different topic for a while and then return back
- Disambiguation support
  - Similar confidence of multiple choices – ask user to select
- Calling external APIs
  - Webhooks/Cloud functions ...

Deployment and Usage

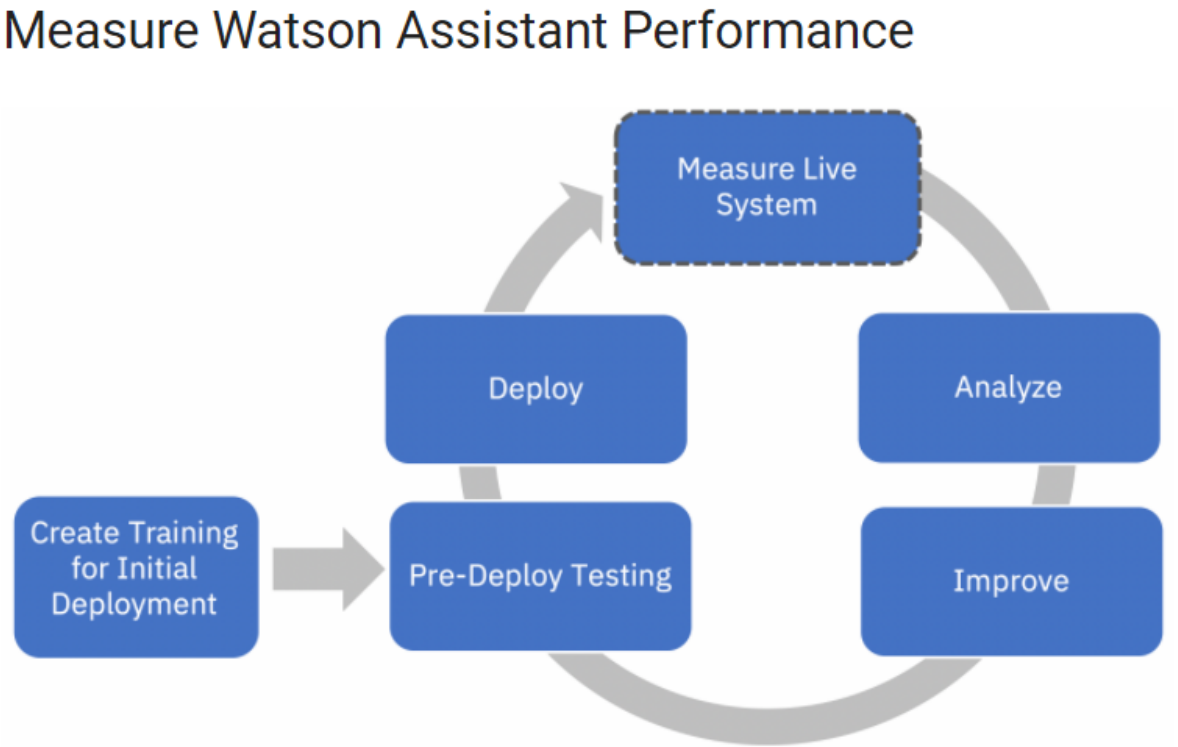
- Authoring tools usually go with an integration support
  - WebWidget - chatting console
  - Slack
  - Facebook
  - Intercom (voice) …
- APIs
  - To include it in customer apps, integration to other solutions
  - Using sessions or conversation ids to track context/history
  - REST API with JSON request/response
- Watson SDK
  - Python, Java, Node.js, .NET
  - https://github.com/watson-developer-cloud

Maintaining and improving chatbot in production

- Automatically
  - Learning from user selections
  - Statistics on user selections – automated "pre-selection" for next users
  - Boosting intent classification performance by generating “paraphrases” by LLM/GPT
- Semi-automatically or manually
  - Chat log analysis
  - Used Measures:
    - **Coverage** ... rate at which your chatbot is confident that it can address the user’s request (per dialogue turn)
    - **Containment** ... rate at which your chatbot can satisfy a user’s request without human intervention, i.e. connect to human agent not requested (per conversation)
  - Content updates
    - To increase the measures above
    - To cover new topics, entities, situations

Chat log analysis - IBM Watson Assistant example

- Python notebook provided to analyze chat log data
  - Covered – check the most frequent
  - Not Covered – extend the coverage
- Visualization of the statistics
  - Number of conversations
  - Conversation length (in turns) stats
  - Coverage and containment history
  - Most frequent intents and entities recognized
  - Low confident intents
  - …



Source: Measure Watson Assistant Performance Python notebook

20+ Metrics for Chatbot Analytics in 2021 by AI Multiple:
https://research.aimultiple.com/chatbot-analytics/

Authoring tools

- IBM Watson Assistant
  - https://www.ibm.com/cloud/watson-assistant/
  - Video tutorial: https://console.bluemix.net/docs/services/assistant/tool-overview.html
- Google Dialog Flow
  - https://dialogflow.com/
  - Video tutorials: https://cloud.google.com/dialogflow/docs/video
- Amazon Alexa Skills
  - https://developer.amazon.com/alexa-skills-kit
  - Video tutorial: https://www.alphavoice.io/video/alexa-developers/alexa-skills-kit-developer-console-build
- Microsoft Cortana Skills
  - https://developer.microsoft.com/en-us/cortana
- Apple SiriKit (Siri-enabled iOS apps)
  - https://developer.apple.com/sirikit/

Gartner Magic Quadrant for Enterprise Conversational AI Platforms 2022

- Evaluation of conversational AI platforms in Jan 2023
- IBM is historically performing well, still the best in the Completeness of vision

Source: Gartner – article by kore.ai

Figure 1. Magic Quadrant for Enterprise Conversational AI Platforms

Bots on Czech Market

- Vodafone CZ – Tobi
- Česká Spořitelna – George
- AirBank – Aneta

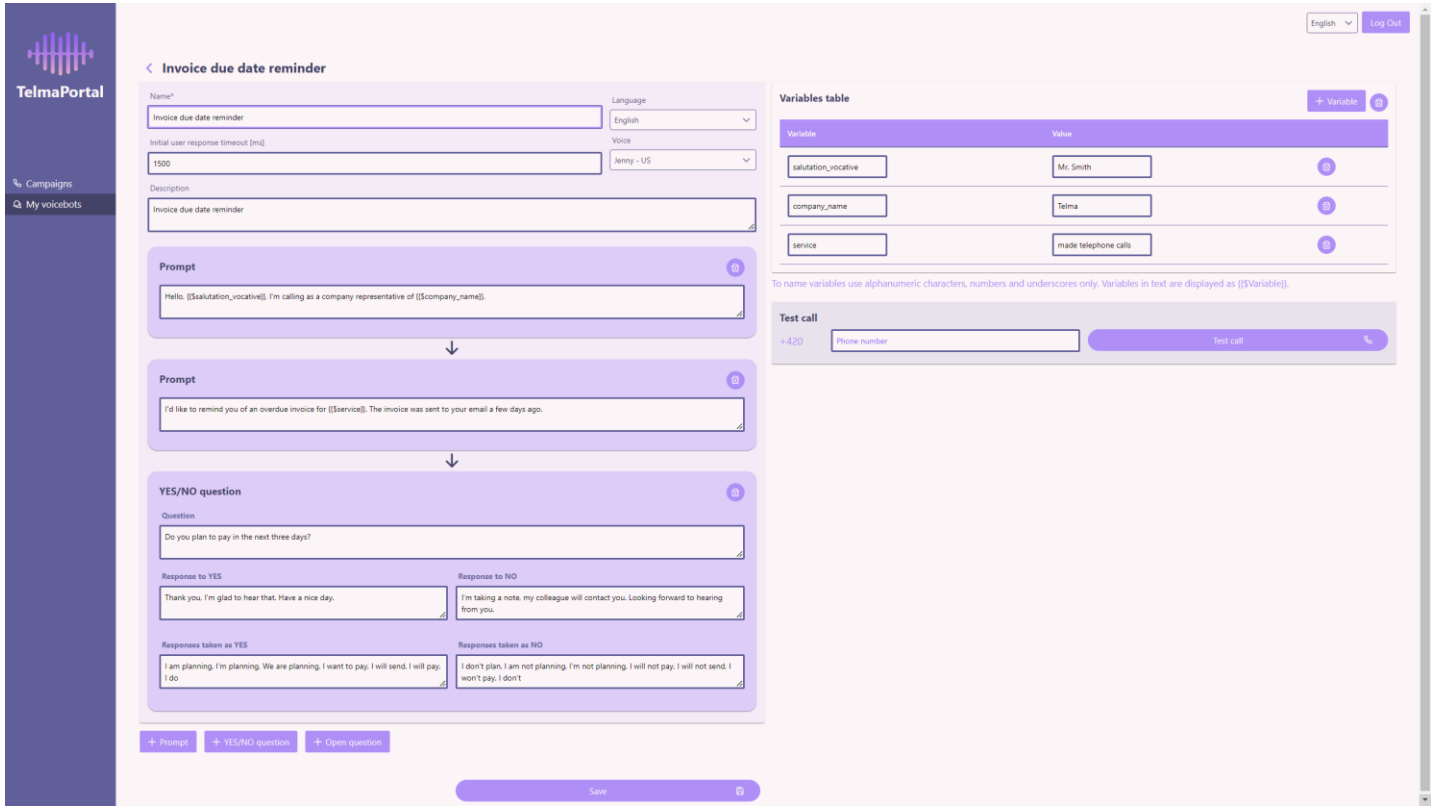
Past

- *Ministerstvo zdravotnictví, ČR – covid-bot Anežka*
- *ING – bot on mobile app*

Authoring tools for outbound calls by Mama AI/Telma AI

- Mama Telma AI tooling for outbound calls
  - Easy of use
  - Modularization
    - Yes/No
    - Rating
    - Open question
  - Language support
  - SMS integration

Examples: https://telma.ai/products/outbound



- Inbound call
  - Python implementation
  - Modules
    - (longer) Number dictation
    - Address dictation (RÚIAN)
  - Guess animal game on Alexa (see youtube)

Examples: https://telma.ai/products/inbound

LLM/GPT base dialog

- Fast growing area of Large Languages Models (LLMs), such as GPT, LLaMA, BART, …
- Generic chatbot/voicebot connected to GPT (info line)
  - Entertainment
  - Demonstration of AI power
  - Buddy for people who feel alone?
- Use of GPT in business – more tricky
  - Priming the model with company information
  - Use of GPT Plugins to connect to up-to-date info (internet, company backend ..)
  - Controlling

Thank you for you attention

Jan Cuřín <jan.curin@themama.ai>

https://themama.ai

https://telma.ai