

Chatbol, a chatbot for the Spanish “La Liga”

Carlos Segura, Àlex Palau, Jordi Luque, Marta R.
Costa-Jussà and Rafael E. Banchs

International Workshop on Spoken Dialog System Technology, Singapore
May 15, 2018

Outline

- Introduction
- Architecture
 - NLU
 - Dialog Management
 - QA Model
 - Retrieval-based Model
- Evaluation
- Conclusions

01 >

Introduction

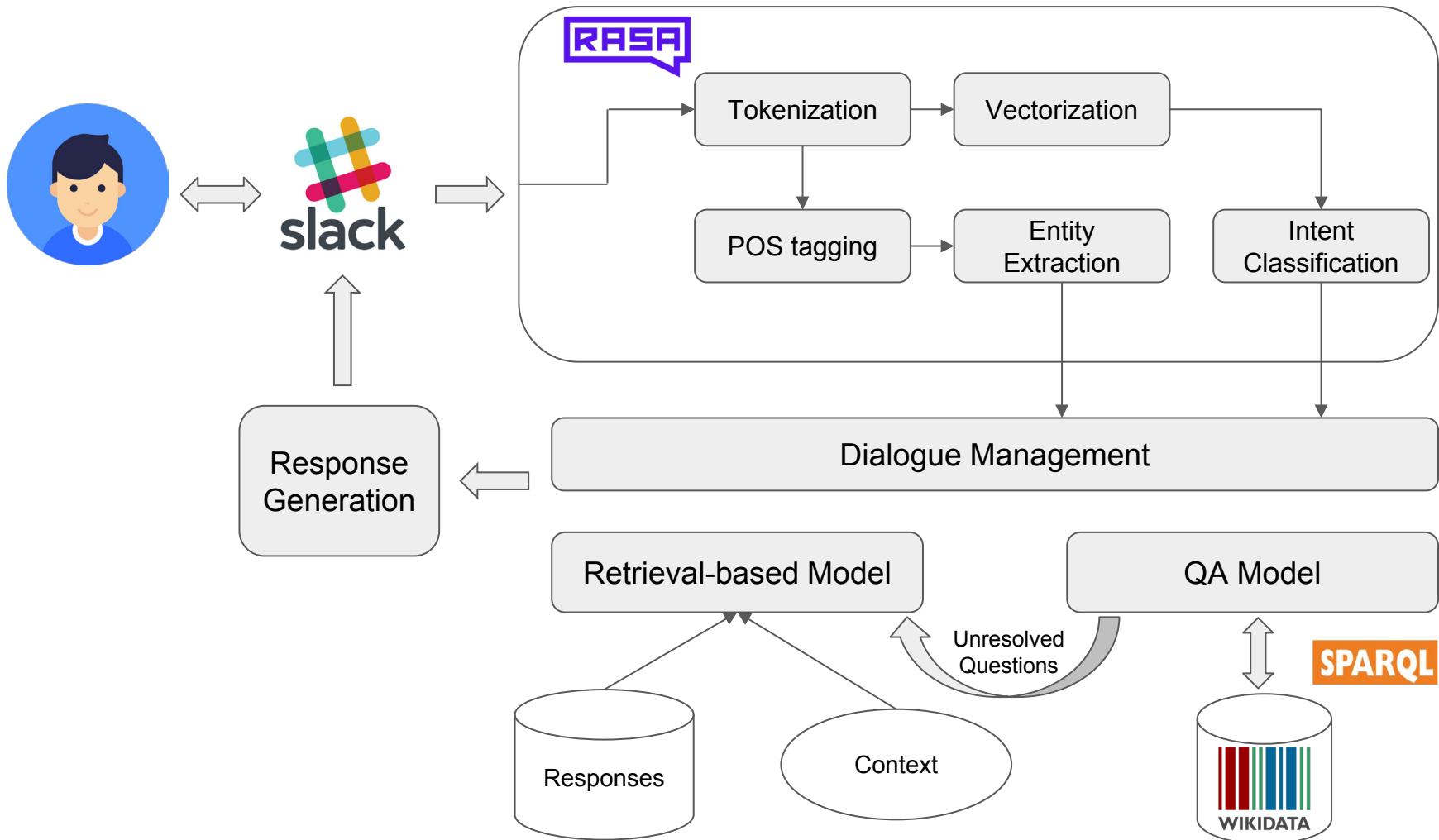
Chatbol - Social Chatbot for the football domain

Description of the development of Chatbol:

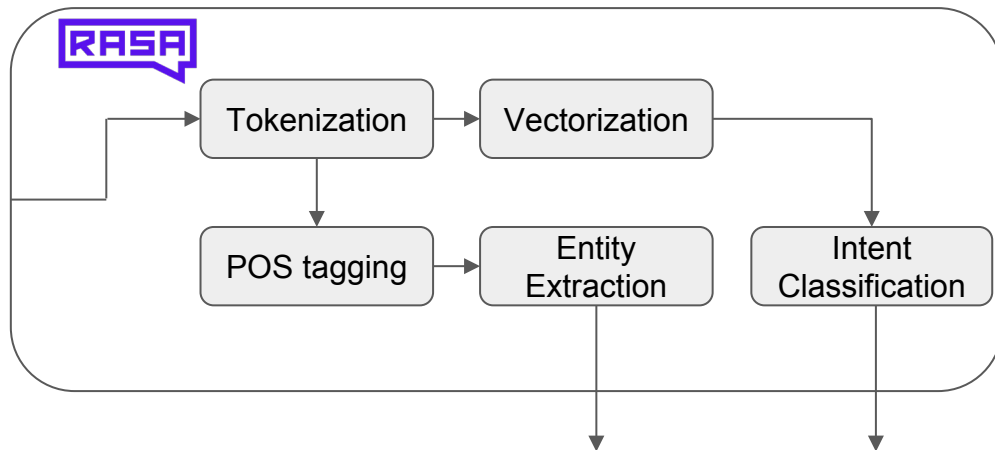
- Deployed in Slack, only text input
- Answers questions about football players, teams, trainers and next fixtures.
- Up-to-date information about entities obtained from Wikidata
- Fall-back retrieval-based conversational engine for a wider variety and freedom of responses

02 >

Architecture



NLU



- Rasa NLU

- Open source
- Modular design
- Relies on existing NLP and ML libraries
- Tokenization
- Vectorization, GloVe vectors
- Intent classification, multiclass SVM, (sk-learn)
- Parts of Speech tagging, (Spacy)
- Entity Extraction, Conditional Random Fields, (sklearn-crfsuite)

Intents and entities

Entities

- [player] Football player
- [team] Football team
- [coach] Coach
- [stadium] Stadium

Training:

- 10 Sentence templates for each intent
- Sentence expansion with list of entities

Intents

- What team is [player] playing for ?
- List of players of [team]
- What is the stadium of [team] ?
- What team is [stadium] of ?
- Who is the coach of [team] ?
- What team is coached by [coach] ?
- Get fixtures of [team]

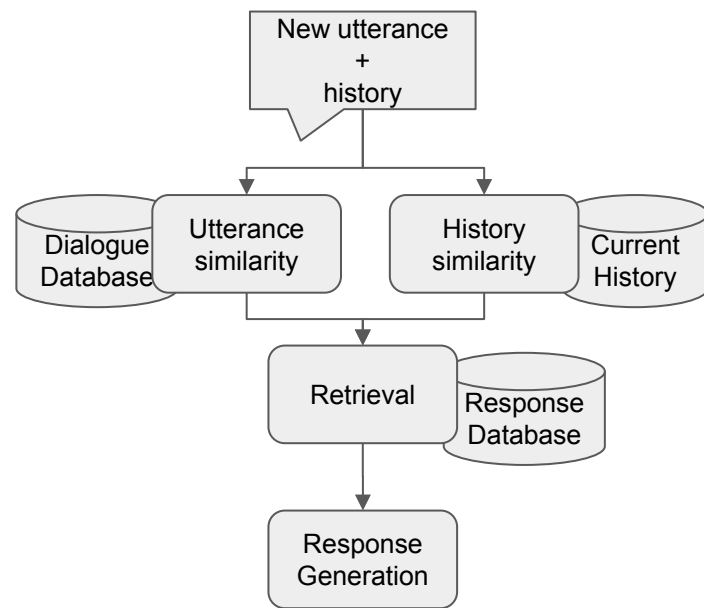
QA Model

- Wikibase entity search -> QNumber
 - Sensitive to the Exact name: Andres Iniesta not found -> Andrés Iniesta
 - Dictionary of alias and ways of calling different players:
 - CR7 -> Cristiano Ronaldo
 - Fuzzy Name search, (FuzzyWuzzy python package)
 - Depending on detected intent by NLU, find a match of the entity in the dictionary of players, coaches, teams or stadiums
- Wikidata query using SparQL
 - Sometimes wikidata incomplete and relation between entities not defined
 - Not all small football teams have defined a coach or stadium in wikidata
- Try the first 3 most probable Intents and entities from NLU until success

Chitchat - retrieval-based conversational engine

- Follows Informal Response Interactive System (IRIS) approach.
- Retrieval based on the semantic similarity scores at turn and dialogue history levels.
- Reranking favouring recall -> variability of responses
- History updated with every utterance and bot response
- Final score:

$$Score = 0.6 * u_c + 0.1 * u_p + 0.3 * h ,$$



Chitchat - Dialogue databases

Databases for training chitchat conversational model:

- Spanish movie data from Opensubtitles database
 - Filter conversations with high frequency (10+) of the word football in Spanish
 - 191,987 original files, 94 files after filtering
 - No explicit information about dialogue sessions
- #futbol channel in IRC-HISPANO
 - Mixed conversations
 - Filter out bad words
- Splitting dialogues into sessions with a 3 seconds silence threshold
- 3,638 sessions
- 148,893 utterances

03 >

Evaluation

Experimental framework

- Deployed in Slack
 - Users interact directly with the Chatbot app
 - Maintains a separate conversation context for each user
 - Convenient for evaluating chatbot responses
- Evaluation protocol follows WOCHAT Shared Task evaluation framework
 - Evaluation at Session level
 - User Satisfaction Survey after interacting with the bot
 - 8 Closed questions
 - 3 Open questions
 - Evaluation at Turn level
 - Annotate Chatbot responses:
 - Valid
 - Acceptable
 - Invalid
 - Session and Turn levels evaluated with 6 different users

Session level Evaluation

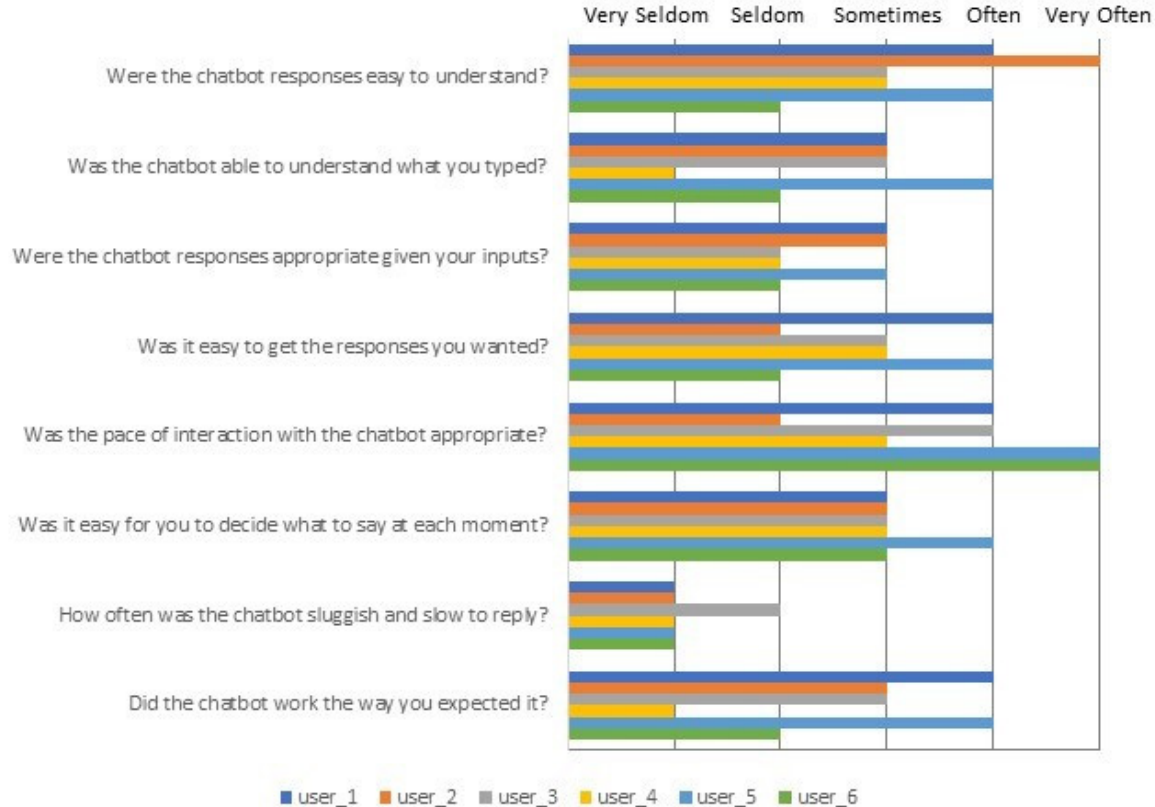
Open questions:

Strengths:

- “accurate information about Spanish soccer teams”
- “capable to understand incorrect words and very flexible with team names variations”
- “Fast way to get football data”

Weaknesses

- ”sometimes the question intents are not properly captured by the system”



Turn Level evaluation

System	Valid Responses	Acceptable	Invalid
QA Model	34 (53%)	12 (19%)	18 (28%)
Chichat Model	50 (43%)	28 (25%)	37 (32%)

Defining appropriate answer as a valid or acceptable:

- QA model generates appropriate answers 72% of the time
- Chitchat model generates appropriate answers 68% of the time

ChitChat - Cherry Picked successful examples

user: how's life ?

chatbol: I am a bit injured, I had an ankle sprain last Tuesday... How are you?

user: Which team do you support ?

chatbol: I support the rival for your team

user: I am an athletic fan

chatbol: athletic is a great team, good players, coacher and supporters, they are huge!

user: is the football match on TV now ?

chatbol: it iust started

user: how is the match going ?

chatbol: leading 1-0

user: the game ended in a tie

chatbol: was the goal during the penalty shoot-out?

user: yes, in the very last moment

chatbol: what were the scores?

user: 1-1

chatbol: cool, I won the sweepstakes! I think no one else bet for 1-1

user: bye

chatbol: take care :d

04 >

Conclusions

Conclusions

- Simple approach for combining a task-oriented bot with a chitchat engine
- Map natural language questions into SPARQL queries to retrieve data from Wikidata
- Robust against word errors, able to recognize variations of player and team names
- Sometimes fails to recognize intents correctly

Chatbol, a chatbot for the Spanish “La Liga”

Carlos Segura, Àlex Palau, Jordi Luque, Marta R.
Costa-Jussà and Rafael E. Banchs

International Workshop on Spoken Dialog System Technology, Singapore
May 15, 2018