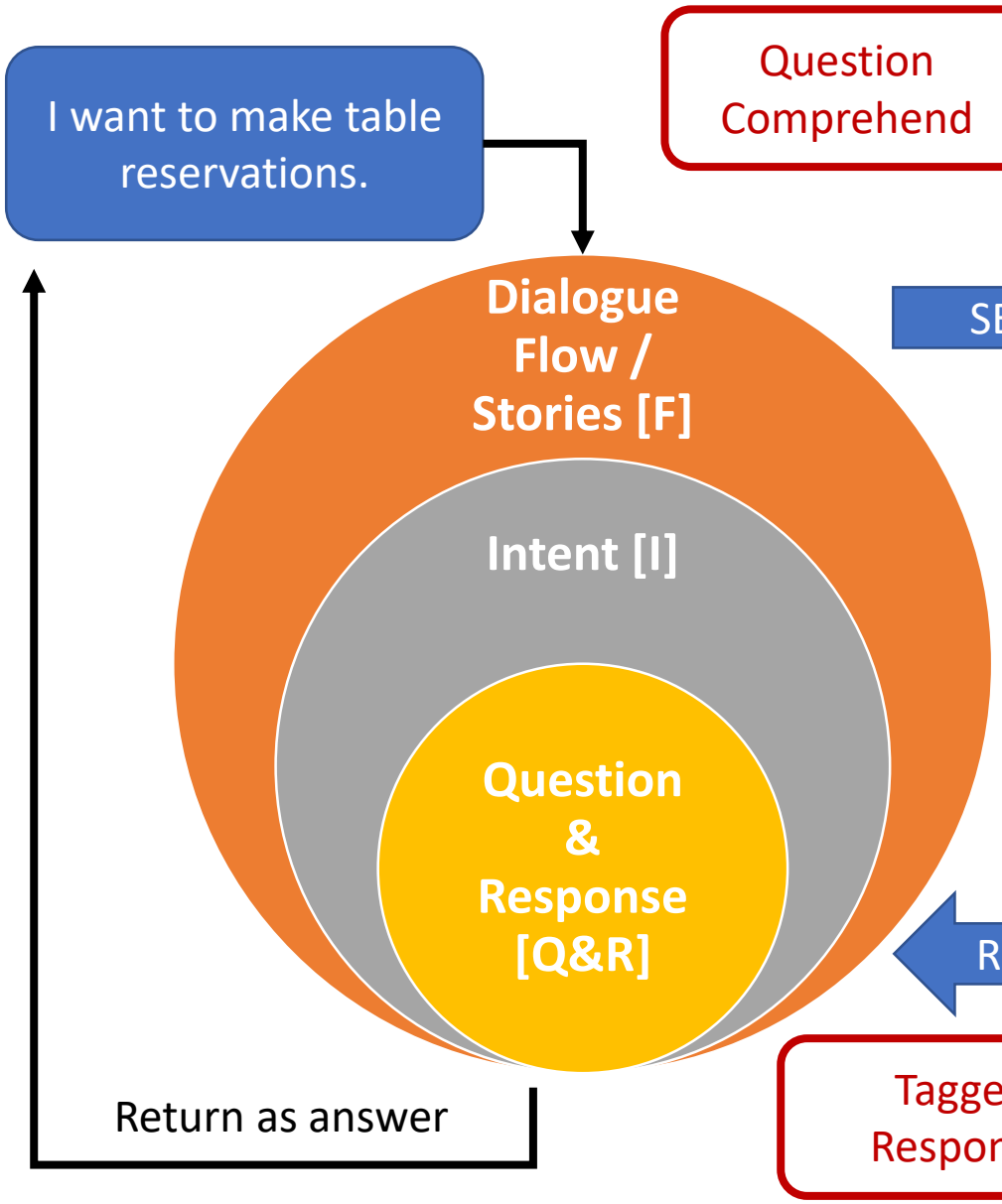
LTA Data Science X NTU EDGE

Chatbot Project

Common Basic Chatbot Workings



[F] Dinner Reservations

= = = = = = = = = = = = = = = = = = = GUIDED = = = = = = = = = = = = = = = = = = =

[Q] I want to make table reservations.
→ [R1] For how many would you like to make reservations for?
→ [R2] At what time would you like to reserve for?
→ [R3] …

= = = = = UNGUIDED: QUESTION MATCHING TO CODED QUESTIONS = = = = =

[I] Reservation Time & Size
[Q] Can I increase the table size after confirmation?
→ [R] Yes you may subject to availability.
[Q] Can I reschedule the time slot?
→ [R] Yes you may subject to availability.

[I] Menu Choice
[Q] What food choices do you have?
→ [R] We have spicy, vegetarian…
[Q] What lunch sets to you have?
→ [R] Fish or chicken set…

[I] Miscellaneous
[Q] Any card promotion and discounts?
→ [R] Yes we have discounts for seafood during this Spring
[Q] Is there a cost to last minute cancellation?
→ [R] Yes, you will be charged for advance food preparations

Knowledge Base Maintenance

Internship Scope

Project Objectives

- User question comprehension
- Knowledge base of chatbot represented in graphs
- Response generation
- Auto-generation of question intent topics

Mountains to Climb

- Format questions to database searchable
- Structuring knowledge in graphs
- Returning human-like answers
- Flow & Intent recommendation/creation

Decomposing Questions to Chatbot

(Annex A)

Aligning questions to database searchable format

- Text normalisation
  - Out-of-vocabulary words, acronyms, contractions, emoticons, etc.
  - Localised English (Singlish), slangs, jargons

- 5W1H decompose to relation triplets <subject, predicate, object>
  - Mary (subj) had a (pred) little lamb (obj)
  - =< 2 of the 3 items will be extracted from the user's question to chatbot

Knowledge Structure in Graphs

(Annex B)

Training knowledge/fact extraction and representation

- Text normalisation
- Content summarisation
  - Extract salient topics/information

Knowledge auto-structure, modelled after Open Knowledge Bases (DBPedia)

- Custom Entity-linkers to disambiguate polysemes
- Heuristic-dependency parser (?) into relation triplets

Completeness of knowledge graph relationships

- Undiscovered relationships between nodes (subj/obj) and edges (pred)

Natural Language Responses

(Annex C)

Graph Traversal for Relevant Information

- Augmentation of original question – better search
  - i.e. What does the NATO military stand for?
  - *North Atlantic Treaty Organisation + Security Force + Member Countries + Strategic Military*
- Relevance of traversed path between nodes and edges
- Relevance to original question

Translating structured inputs into natural languages

- Transform relevant relation triplet sets into human-sentences

AI-assisted Dialogue Flow Creations

(Annex D)

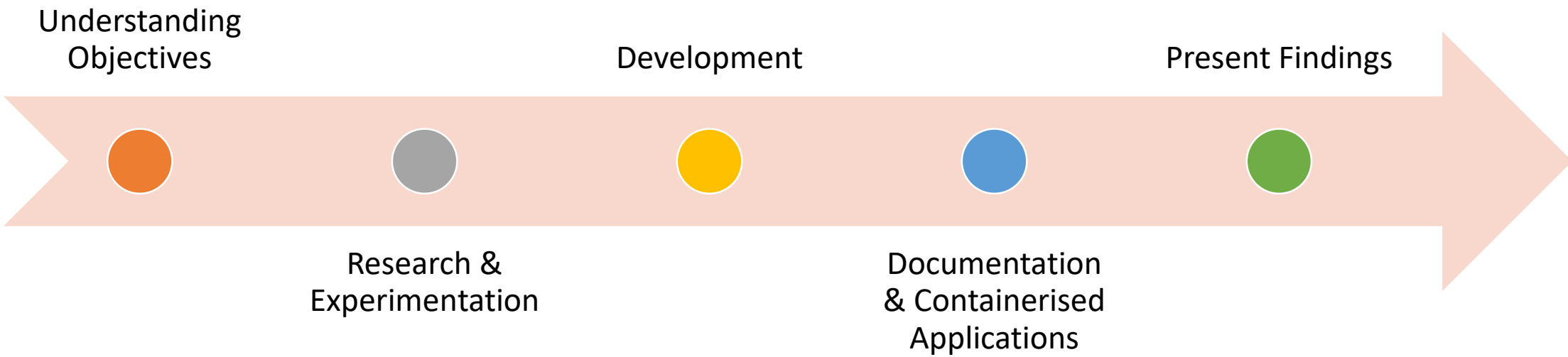
Training Dataset Linked Topics

- Precedent Topic → Consequent Topic
  - (Conseq.) User request for computers
  - (Preced.) Enough funding? Obtain approval from?
- Grouping related training data sub-topics together to create flows

Intent-chaining

- Question similarities
  - Possibly recommender systems - series of questions (topics)
  - Populate question-respond chains based on inferred “recommended questions”

General Flow of Things



Thank You!

Annexes

Annex A: Question Decomposition

Normalise localised Singlish to English

- Machine Translation
  - Paper, project on translating Singapore SMS Singlish to proper English using MT
  - C++ Tool, Moses MT tool used by paper
- Leverage word embedding to normalise words
  - Python Tool, word-mover distance based on word embedding to find expanded word
  - Paper, word-mover distance

OOV word syntactic understanding

- Paper, language model to infer dependencies/POS of words in sentences

Heuristic information extraction

- Web blog, overview of information extraction via spaCy

Annex B: Knowledge Structuring

Extractive & Abstractive Text Summarisation

- Paper, text summarisation overview
- Paper, unsupervised summarisation
- Web Blog, text rank for extractive summary
- Web Blog, sequence model with attention mechanism

Knowledge relation & fact extraction

- Web Blog, overview relation extraction
- Paper, overview of tools for relation extraction

Creating knowledge in graph

- Paper, procedure to creating knowledge graph
- Web Blog, heuristic fact extraction, creating knowledge graph

Graph completion

- Web Blog, graph embedding with Deep Walk
- Web Blog, overview of graph completion algorithms, TransE/H/R/D
- Paper, TransR
- Paper, extends TransE models with Linear Programming

Annex C: Question-Knowledge Match
Annex D: Novel Dialogue-Flow Generation

Annex C

Knowledge retrieval

- Paper, candidate ranking of relation units in graph DB

Question-answer model

- Web Blog, supervised SQuAD with bidirectional/attention sequence model

Annex D

Content clustering & visualisation

- Paper, visualising LDA topics by relevance
- Paper, visualise LDA topics
- Python Tool, scikit-learn example on text clustering

Recommender with Word Embedding

- Web Blog, utilised method of training Word2Vec to recommend products

Works beyond the horizon for thought…

Classify type of question

- FAQ, transactional in nature or complex queries

Master & slave bots

- Traffic direction chatbot vs. domain specific

Forms in chatbot

- Receive inputs to make transactions: i.e. book appointment

Sentiment detection

- Moderate responses into cordial mode when user exhibits frustration