University of Glasgow | School of Computing Science

Assessed Coursework (and Demo)

Course Name	**Conversational Interfaces (M)**			
Coursework Number	**1**			
Deadline	**Time:**	**4.30pm**	**Date:**	**24 March 2023**
% Contribution to final course mark	**30% - 25% Report / 5% demo**			
Solo or Group ✓	**Solo**	✓	**Group**	
Anticipated Hours	**30 hours**			
Submission Instructions	**As per specification below.**			
Please Note: This Coursework cannot be Re-Assessed				

Code of Assessment Rules for Coursework Submission

Deadlines for the submission of coursework which is to be formally assessed will be published in course documentation, and work which is submitted later than the deadline will be subject to penalty as set out below.

The primary grade and secondary band awarded for coursework which is submitted after the published deadline will be calculated as follows:

(i) in respect of work submitted not more than five working days after the deadline
   a. the work will be assessed in the usual way;
   b. the primary grade and secondary band so determined will then be reduced by two secondary bands for each working day (or part of a working day) the work was submitted late.

(ii) work submitted more than five working days after the deadline will be awarded Grade H.

Penalties for late submission of coursework will not be imposed if good cause is established for the late submission. You should submit documents supporting good cause via MyCampus.

Penalty for non-adherence to Submission Instructions is 2 bands

You must complete an "Own Work" form via https://studentltc.dcs.gla.ac.uk/ for all coursework

Conversational Interfaces Coursework 2023

In this coursework, you must develop a task-oriented dialogue system that lets a user access various information about Premier League football teams and their personnel, using the implementation toolkit of your choice – the toolkit may be one of the ones covered in the labs, or if you would prefer, you can also choose another toolkit. A set of sample dialogues in a football information domain have been provided.

This assignment has two parts. In Part A, you will develop a system based on the sample dialogues, and will evaluate the NLU component of that system using a set of previously unseen dialogues from the same domain. In Part B, you will design an extended version of the system that goes beyond the sample dialogues. You will then have a choice: either you will implement the extended version, or you will carry out a small-scale user evaluation of the final system. For each part of the assignment, I have noted what should be included in the final report.

This is an **individual exercise**, and you should work independently. If you have questions concerning the specification, you can address these to the course lecturer or GTAs, or ask on the dedicated channel on Microsoft Teams. There will also be an FAQ document on Moodle that will be updated in response to students' questions.

Datasets[1]

On Moodle there is an accompanying zip file **examples.zip** that contains sample dialogues for Part A. The full list of intents and slots that are included in these dialogues are shown in the Appendix.

On 6 March 2023, an additional file of sample dialogues will be provided to allow you to complete Question 2.

Part A

Your aim in this task is to build a conversational agent that allows a user to access information about football teams, and to evaluate its NLU performance against a set of unseen test dialogues from the same domain. You should base your system on the provided dialogues, which include the core intents and entities that must be included.

Q1: Football agent development (6 marks)

Build an agent based on the provided dialogues. You can use the toolkit of your choice to build the agent.

The intents and slots that you should use are provided in the Appendix. To allow your system to be evaluated in Q2 below, you should be sure to use the indicated intents and slots exactly. You should use the sample conversations in the zip file as training examples. Depending on your preference, you could use the built-in training facility of your chosen tool, or you could just type them in directly as examples.

Your system must satisfy the user queries by making appropriate calls to a sports information API. There are several APIs available; we suggest using the API-Football API, which is documented here:

[1] These utterances are adapted from the Google "Taskmaster-2" dataset https://github.com/google-research-datasets/Taskmaster/tree/master/TM-2-2020.

https://www.api-football.com/

You can create a free account on API-Football through https://dashboard.api-football.com/register. The "Free" tier will allow up to 100 API calls per day; if this poses an issue, please get in touch as soon as possible.

There are links on Moodle to instructions on using the API-Football API, and also on accessing an external API using all of the main toolkits that were covered in the labs (Alexa, DialogFlow CX, Rasa, Furhat).

What to submit

For Q1 you must create a sample dialogue inspired by the samples provided, which should exercise all the slots provided in the schema. In your report you will provide an annotated conversation with the schema and slots corresponding to the training data. Use the test console provided with your selected tool and be sure to show the detected intents and entities for each turn. Discuss the implementation and final result.

Q2: NLU Evaluation (5 marks)

In this question you will run a set of test user utterances on your created agent to evaluate the NLU effectiveness.

On **7 March 2023**, a set of sample test dialogues from the same domain will be provided to allow you to test your system's NLU performance. The full details of the test procedure will be provided along with the dialogues. The NLU evaluation will focus on the utterances containing domain-relevant slots, rather than the more functional slots such as greetings and acknowledgements.

You are not required to update your system in any way in response to the results of this evaluation. You will be marked on whether you carried out the evaluation correctly and reported and discussed the results appropriately, not on the actual results themselves.

What to submit

The submission format for this part will be provided along with the test procedure and data.

Part B

Q3: Design an extended sports agent (5 marks)

You should **design** an extended version of the agent that you implemented in Part A. (Note that this task does not require you to implement the enhanced agent, only to design it.) Your design should include at least the following:

- A clear specification of the full behaviour of the system
- At least three target dialogues demonstrating the full system range
- A flow diagram showing the states in the extended dialog and the flow between them
- A detailed description of the advanced element (persona, card mockups, etc…)

Your agent design must incorporate **at least one** of the following advanced components:

- Multi-modal UI - display and render visual elements including cards, carousel, lists, etc…
- Detailed personality / persona responses

- Socialbot chat elements – in-depth discussion of the football results or standings, follow-up discussion, etc
- Advanced fulfillment – For example, discussing other aspects of sports, calling other APIs, etc
- … Or some other advanced feature – if you have an idea, feel free to check with us whether it would be appropriate

What to submit

Provide a full domain schema definition following the format and conventions in part A and in the appendix. Provide 3 sample aspirational target dialogues, including appropriate system responses and conversational components, along with a flow diagram or similar, and a detailed description of the advanced element. Discuss and analyse any assumptions and trade-offs made in the design of the system. You should reference the voice design principles discussed in class.

Q4: Extended implementation OR User evaluation (6 marks)

For the final part of the assignment, you have two options – you should choose **one** of the following options. Please make clear in your final report which of the options you have selected.

Option 1: Implementation of the extended agent

If you choose this option, you should implement the agent design from Q3 including the intents, entities, and extended fulfilment as required. **You are strongly recommended to save your Part A agent when it is complete and to do this task on a copy, to allow the NLU evaluation in Q2 to be carried out without interference from the extended functionality.**

What to submit: Show a fully annotated sample dialogue as in Part A, including the detected intents and slots and the actual system responses to the user turns. Discuss any decisions and tradeoffs made in the implementation – did you implement the full design outlined in Q3? Did any modifications need to be made? Be sure to detail the advanced feature you used and discuss how it changed the implementation of your agent.

Option 2: User evaluation of core agent

If you choose this option, you should implement and carry out a user evaluation of the core system from Part A, involving at least **5** users. (You are free to test on your friends/family/flatmates/fellow students on this course.) Be sure to use appropriate subjective and objective measures to assess the quality of the interactions. For subjective measures, you should select appropriate items from the SASSI questionnaire; depending on how you implemented your system, the items that are appropriate for your specific evaluation may vary. For objective measures, you should include measures of task success and dialogue efficiency. You are not required to compute a performance function, but you may choose to do so.

What to submit: A description of the evaluation set-up -- number of users, details of the task(s) that they are asked to perform. The summary results of the objective and subjective measures. A discussion of the results, including suggestions of how the system might be modified in response to the evaluation results (you do not need to do any modifications!).

Submission

Report – 25 marks total

In your **report** (up to **10 pages MAX**, plus appendices), provide the results for Part A & Part B. The report should include a title, your name, date and student number. In addition to the marks allocated for each part (as given above), an additional 3 marks will be allocated to the quality of the report (organisation, correct spelling, effective use of figures, presentation, etc).

Demo presentation video – 5 marks

You must record yourself giving a **short** (maximum 5 minute) presentation about your system, including a demo of the live system. The presentation should discuss:

- Important (and novel) design considerations
- Demo use case that shows the range of your agent's capabilities
- Reflection on the process as well as what / how you would change or improve the agent
- If you chose to implement the extended version for Question 4, your video should demonstrate the full version of the system; if you chose to carry out a user study in Question 4, your video should include a brief summary of the evaluation results

You do not need to record a video of yourself -- a screen recording with voice-over is sufficient. Guidance on the videos will be provided on the Moodle site.

Submission Process

Your report and video should be submitted through Moodle. The deadline is **Friday, 24 March @ 4:30pm**. You are encouraged to submit your video early if possible; if you give permission, I will add your video to a playlist which will be shared only with the rest of the class.

Please upload (a) **PDF report document** and (b) the **video of your presentation.**

Appendix 1: Slots and intents from the core football domain

Note that there is really only one main intent annotated in these dialogues, but it can have a range of slots as indicated. All other utterances are annotated as “Other” in the provided files; when you implement your system, you may want to use other intents for those, but for the NLU evaluation they can be thought of as “Other”.

Intent: GetInfo

Slots:

- team (required)
- slots (optional): must be one or more of ‘lastOpponent’, ‘lastScore’, ‘leaguePosition’, ‘manager’, ‘nextGameDate’, ‘nextOpponent’, ‘numGamesPlayed’, ‘playingNow’, ‘winLossRecord’