Ghigliottin-AI at EVALITA 2020
Task Guidelines

Pierpaolo Basile[†], Lucia Siciliani[†], Federico Sangati[‡] ,
Johanna Monti[‡] , Antonio Pascucci[‡], Marco Lovetere[§]
[†] University of Bari "Aldo Moro", Italy
[‡] Università di Napoli "L'Orientale"
[§] Ghigliottiniamo
{ghigliottinai.evalita@gmail.com}

Ver. 1.00

Contents

1 Task Description

Language games draw their challenge and excitement from the richness and ambiguity of natural language, and therefore have attracted the attention of researchers in the fields of Artificial Intelligence and Natural Language Processing. For instance, IBM Watson is a system which successfully challenged human champions of Jeopardy!, a game in which contestants are presented with clues in the form of answers, and must phrase their responses in the form of a question (Ferrucci et al., 2010; Molino et al., 2015). Another popular language game is solving crossword puzzles. The first experience reported in the literature is Proverb (Littman et al., 2002), that exploits large libraries of clues and solutions to past crossword puzzles. WebCrow is the first solver for Italian crosswords (Ernandes et al., 2008).

Following the first edition of the NLP4FUN task (Basile et al., 2018), proposed at EVALITA 2018, we propose a new edition of the task which aim is to **design a solver for "The Guillotine" (La Ghigliottina, in Italian) game**. It is inspired by the final game of an Italian TV show called "L'eredità". The game, broadcast by Italian national TV, involves a single player, who is **given a set of five words - the clues - each linked in some way to a specific word that represents the unique solution of the game**. Words are unrelated to each other, but each of them has a hidden association with the solution. Once the clues are given, the player has one minute to find the solution. For example, **given the five clues: pie, bad, Adam, core, eye the solution is apple**, because: apple-pie is a kind of pie; bad apple is a way to refer to a trouble maker; Adam's apple is the prominent part of men's throat; apple core is the center of the apple; apple of someone's eye is way to refer to someone's beloved person.

Participants are asked to build an artificial player able to solve "La Ghigliottina". They can take advantage of solutions adopted by previous systems (Semeraro et al., 2009; Basile et al., 2016; Sangati et al., 2018) and the availability of open repositories on the web (see section 4).

2 Development Data

We provide a set of 300 games with their solution taken from the last editions of the TV game as training data. The training data will be released in JSON format:

```
[
  {
      "w1": "posto",
      "w2": "artificiale",
      "w3": "lavaggio",
      "w4": "allenare",
      "w5": "gallina",
      "solution": "cervello"
   },
   {
      "w1": "essere",
      "w2": "comparsa",
      "w3": "x men",
      "w4": "ronaldo",
      "w5": "mondiale",
      "solution": "fenomeno"
   },
   ...
]
```

The JSON file consists of an array of *games* which contains several JSON objects for each game. In each *game* we have 5 clues (*w1, w2, ..., w5*) and the *solution*.

Please refer to section 5 for more information about how to register your system and download the development data.

3 System evaluation

In order to evaluate the AI systems, we rely on an **API based methodology**. For this we will make use of the Remote Evaluation Server (RES) **Ghigliottiniamo** which currently enables both humans and artificial systems to submit solutions to the TV game in real-time.

During the evaluation period, at **random intervals of time**, the RES will submit to the registered systems a request with **a single game challenge**. The systems must reply back to the RES with a **single solution** to the game.

Please refer to section 5 to understand how to register your system to the RES, and how to test it to ensure that the system is setup correctly.

3.1 Evaluation metric

As evaluation measure, we adopt the standard **accuracy score**:

$$\frac{solved\_games}{total\_games} \tag{1}$$

Similar to the TV game, where players have one minute to provide the solution, the **RES will discard system solutions received after 60 seconds** from the submitted challenge.

4 List of useful Resources

This is a challenging language game which demands knowledge covering a broad range of topics, to understand the clues and identify their connections with potential solution words. We list here a number of suggestions to help potential participants to the challenge.

Previous systems (Semeraro et al., 2009; Basile et al., 2016; Sangati et al., 2018) have indicated some of the possible connection between clue words and solutions: word co-occurrence in frequent collocations or idioms, word similarity or word relatedness.

We list a number of useful resources on the web:

- Corpora: PAISÀ Corpus, itWaC Corpus, Wikipedia extractor and cleaner
- Collocations and Idioms: De Mauro Dictionary, Italian Proverbs
- Italian word embeddings: Italian Word Embeddings
- Word Knowlege representation: Conceptnet

5 System Registration

This challenge uses an **API based infrastructure** to connect to the *Remote Evaluation Server* (RES) Ghigliottiniamo.

In order to register a new system please go to the following URL:

https://ghigliottina.marlove.net/www/ghigliottin-ai

and enter your e-mail address, your *AI System Name* (choose wisely), and the *Webhook URL* where the RES system can send you the requests. The Webhook URL can be changed later, so if you just want to get started you can use a placeholder (e.g., http://anyurl.com).

After clicking the submit button, you will be redirected to your **Account Webpage** with the following information:

```
UUID: xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
Secret: xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Authorization: xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Webhook URL <the_chosen_webhook>
Test webhook link
Edit webhook
Download dataset link
```

This info will be also sent to the email you specified.

IMPORTANT: PLEASE SAVE THIS INFORMATION IN A SAFE PLACE AND DO NOT PUBLISH IT PUBLICLY

At any point you will be able to access your *Account Webpage* at the following URL:

https://ghigliottina.marlove.net/www/ghigliottin-ai/account.php?uuid=‘<UUID>‘&secret=‘<secret>‘&authorization=‘<authorization>‘

by replacing the *UUID*, *secret* and *authorization* keys accordingly.

5.1 Download Development Data

In order to download the Development Data click on the **Download dataset** link in your *Account Webpage* or use the following URL:

https://ghigliottina.marlove.net/api/v1/read/wh_dataset.php?uuid=‘<UUID>‘&secret=‘<secret>‘

5.2 Change your webhook

At any moment you can change your webhook by clicking on **Edit webhook** in your *Account Webpage*.

5.3 Setup and test the API

In order to make sure your system API infrastructure is properly setup for the evaluation phase, we have implemented a test functionality.

From your *Account Webpage* click on the **Test webhook** link.

The webpage will show you a popup message with a confirmation of whether it was able to ivoke your Webhook URL.

If so, you should have received a POST request to the web-hook you specified with following payload:

```
  {
"game_id": 111,
"w1":"string1",
"w2": "string2",
"w3": "string3",
"w4": "string4",
"w5": "string5",
"callback": "<callback_url>"}
```

The w1,...,w5 in the payload is a random game from the Development Data.

You can check that the POST request was sent by the official RES by checking if the **Authorization** field in the header of the request matches the authorization string you received after registering the system.

At this point you should send a POST request to the *callback URL* with the *secret* key in the *Authorization* field of the header and the following payload:

```
    "game_id": 111,
"uuid": "<UUID>",
"solution":"your solution"
```

Where *solution* contains a **single solution** to the game.

5.4 Useful tips

We advise participants to deploy their system on a server (a number of free cloud-based are available such as heroku). For testing purposes, participants can make use of *tunnelling* software (such as localtunnel) that enables a system to run and communicate with the Remote Evaluation Server from a local machine.

We are aware the API technologies (while being ubiquitous in all IT sectors) are still uncommon in shared tasks, but we decided to adopt them because they offer a unique opportunity to evaluate the systems more robustly and continuously in time. We do not want this to be an obstacle for people to participate to the challenge, and therefore we will provide all assistance needed for participants to set up their systems correctly.

6 How to participate

- Fill in the Evalita 2020 Registration Form
- Follow the System Registration guidelines to register and setup your system and download the Development Data
- Join our Google Group here

References

P. Basile, M. de Gemmis, P. Lops, and G. Semeraro. Solving a complex language game by using knowledge-based word associations discovery. *IEEE Transactions on Computational Intelligence and AI in Games*, 8(1):13–26, 2016.

P. Basile, M. de Gemmis, L. Siciliani, and G. Semeraro. Overview of the evalita 2018 solving language games (nlp4fun) task. In T. Caselli, N. Novielli, V. Patti, and P. Rosso, editors, *Proceedings of the 6th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'18)*, Turin, Italy, 2018. CEUR.org. URL http://ceur-ws.org/Vol-2263/paper011.pdf.

M. Ernandes, G. Angelini, and M. Gori. A web-based agent challenges human experts on crosswords. *AI Magazine*, 29(1):77, 2008.

D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, et al. Building watson: An overview of the deepqa project. *AI magazine*, 31(3):59–79, 2010.

M. L. Littman, G. A. Keim, and N. Shazeer. A probabilistic approach to solving crossword puzzles. *Artificial Intelligence*, 134(1-2):23–55, 2002.

P. Molino, P. Lops, G. Semeraro, M. de Gemmis, and P. Basile. Playing with knowledge: A virtual player for “who wants to be a millionaire?” that leverages question answering techniques. *Artificial Intelligence*, 222:157–181, 2015.

F. Sangati, A. Pascucci, and J. Monti. Exploiting multiword expressions to solve “la ghigliottina”. In *Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018)*, pages 1–6, 2018. URL http://ceur-ws.org/Vol-2263/paper044.pdf.

G. Semeraro, P. Lops, P. Basile, and M. De Gemmis. On the tip of my thought: Playing the guillotine game. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence*, IJCAI’09, pages 1543–1548, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc. URL http://dl.acm.org/citation.cfm?id=1661445.1661693.