# 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 games with their solution taken from the last 4 editions of the TV game as training data. The training data will be released in JSON format:

```
{
  "games":[
  {
        "id":1000,
        "clues":["uomo","cane","musica","casa","pietra"],
        "solution":"chiesa"
  },
  {
        "id":1001,
        "clues":["doppio","carta","soldi","pasta","regalo"],
        "solution":"pacco"
  }
  ]
}
```

The JSON file consists of a root element *games* which contains several *game* elements. Each game has five elements inside *clues* and one *solution*.

Participants are encouraged to integrate any knowledge resources in their systems except explicitly providing their systems with solutions from games not in the training set.

## 3 System evaluation

In order to evaluate the AI systems, we will 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.

We will provide detailed instructions on how to register a system to the RES, and will enable test functionalities to ensure that the system is setup correctly.

During the evaluation period, at **random intervals of time**, the RES will submit to the registered systems a *POST request* containing **a single game challenge**:

```
{
  "game_id": 1003,
  "clues": ["giro", "data", "buco", "religione", "locale"]
  "callback_url": https://unique-url-for-submitting-the-solution
}
```

The systems must **submit the solution** to the `callback_url` with a *POST request*:

```
{
  "uuid": user-id-obtained-during-the-registration-procedure
  "game_id": 1003,
  "solutions": ["solution1", "solution2", ... , "solution100"]
}
```

Where `solutions` is a ranked list of **maximum 100 tentative solutions** to the game.

### 3.1 Evaluation metric

As evaluation measure, we adopt the standard Mean Reciprocal Rank (MRR).

$$\frac{1}{|G|} \sum_{g \in G} \frac{1}{rank_g} \tag{1}$$

where $G$ is the set of games and $rank_g$ is the rank of the solution.

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

## 4 Useful Tips

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.

### 4.1 List of useful Resources

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-occurrencence 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

### 4.2 API System Setup

We will provide all technical details to the participants for them to setup their system correctly according to the API methodology illustrated in section 3.

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.

## 5 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.