Escape from Dungeon – Modelling User Intentions with Natural Language Processing Techniques

Stefan Toncu[1], Irina Toma[1], Mihai Dascalu[1,2], Stefan Trausan-Matu[1,2]

[1]University Politehnica of Bucharest, 313 Splaiul Independentei, 060042, Bucharest, Romania
[2]Academy of Romanian Scientists, Str. Ilfov, Nr. 3, 050044, Bucharest, Romania

stefan.toncu@stud.acs.pub.ro,
{irina.toma, mihai.dascalu, stefan.trausan}@upb.ro

Abstract. Educational games are a powerful solution for pedagogical problems, both from students' and teachers' points of view. Students may experience in a traditional learning environment the inability to focus on the lectures, as they cannot understand the lecture materials, are not motivated to study, or the subjects are not challenging enough. Research on learning strategies shows that students are more likely to remain focused and engaged in a smart learning environment that makes use of gamification, instead of a classical classroom scenario, where teachers present formal lectures. Our game, *Escape from Dungeon*, falls in the category of serious games for problem solving that integrate Natural Language Processing (NLP) techniques adopted to model user intentions. We focused on ensuring appealing graphics and ease of interaction, while relying on novel technologies. The main character of the game is controlled through vocal commands that are interpreted using NLP tools. The game was tested by ten users throughout a pilot test. Users considered the game innovative and entertaining. However, users suggested additional game scenes for an extended gameplay, as well as more actions and intents to be covered within the interaction with the character.

Keywords: Serious Games, Smart Learning Environment, Natural Language Processing, Voice Recognition, Virtual Reality.

1 Introduction

Games are one of the oldest means of relaxation and entertainment, as they offer participants the opportunity to enter a unique universe, transcending everyday reality [1]. Most games share common features, such as rules and goals: players agree upon game rules and respect them in order to achieve the predefined goals. The gaming process involves competition to a certain degree, while the outcome is uncertain. Some games include elements of chance, fiction, puzzle solving, but all games should consider leisure elements to ensure personal entertainment, satisfaction or self-fulfillment.

Even though games are designed to entertain players, their goals can include more than the act of play. Participants can exercise previous knowledge or acquire new

information during the gameplay, by being actively involved in a learning ecosystem. This process is more tiresome to be performed during classes or seminars. Game-based learning [2] is a method applied by teachers when presenting information to their students in a fun and interactive manner, while keeping students focused on the presented knowledge, rather than on the game core and mechanics. Thus, the traditional learning process, when theory is presented in a mechanical way, is transformed into an attractive activity, focused on understanding concepts while practicing theoretical notions [3].

This article focuses on serious games based on Natural Language Processing (NLP) for interactive storytelling [4] and dialogue systems [5]. NLP dialogue systems for serious games can be separated into Natural Language Understanding (NLU), which extracts intents and entities from data based on context and grammar, and Natural Language Generation (NLG), responsible for generating text based on structured data [6]. NLU is of interest for the scope of this study because our serious game – Escape from Dungeon – integrates speech recognition functionalities and NLP techniques for intent and entity identification.

The usage of intent identification is a growing market, as more products incorporate voice assistant features for everyday use. Intents vary from simple informative commands, such as "play music" or "call somebody", to more complex intents that require several connected systems in order to fulfill the request (e.g., "turn off the light", or "turn on heating system") [7]. For each spoken request, referred to as an utterance, an intent identification system performs the following tasks: domain classification, intent detection, and slot filling [8]. These three tasks are preceded by a preprocessing step that converts speech to text. Several models based on deep neural networks classifiers [9, 10] are trained for user intent detection. The problem arises when users express different intentions than the ones available in the training set. One solution to this problem is to simply ask users to rephrase their intent or tell them that the system does not understand the request. A second approach is to use an intent detection system that supports zero-shot learning [11]. The system goes beyond the manually labeled intentions and gathers data from external resources, such as labeled ontologies or manually defined attributes that describe intents, followed by associating existing or emerging intents with extra annotations [12].

Our serious game – Escape from Dungeon – aims to develop players' sense of observation and general knowledge. The game integrates speech recognition functionalities, NLP techniques for intent identification, and can render scenes into a virtual reality environment with the help of a VR headset.

2 Serious games based on NLP

Serious games represent a small percentage of the gaming market [13] when compared with the rest of the entertainment gaming industry [14]. The discrepancy is caused by the game objectives, as it is more likely for a user to play a game for entertainment than for educational purposes. Serious games based on NLP dialogue systems process and generate natural language to communicate with the players. The whole process is complex because the game must handle metaphors, bad spelling, grammar mistakes, and

contextual input. The following section details three existing games, Façade [15], LifeLine [16], and Crystal Island [17]. Each of the games is of interest for the current paper as they use NLP techniques for interacting with users, voice commands, or have a solid educative scenario. In addition, their specific features were further used to establish a starting point for developing our serious game.

2.1 Façade

Façade is a serious game introducing the players in the universe of a married couple. The game is considered an interactive drama: the player is invited by two friends, Trip and Grace, to their apartment for dinner. Here, he/she participate to a couple fight and interact with their virtual friends using natural language as input.

The story is not pre-scripted, but it instead is shaped based on the dialogue (see Fig. 1). Users can try to help their friends solve their conflict or, on the contrary, intensify it. Depending on what the player says to the two non-playable characters (NPCs), the story branches out. Thus, the game can have multiple endings: the marriage conflict ends with the spouses reunited or with their separation; the player is asked to leave the apartment, either Trip or Grace leaves the apartment, or their relationship can remain unchanged if the users interaction does not affect them negatively. The input provided by users is mapped into predefined sets of feelings, and the responses are selected accordingly, from a predefined pool of context-appropriate texts.

Fig. 1. Façade gameplay [15].

2.2 LifeLine

LifeLine (see Fig. 2) begins with several monsters attacking a space station where the main character named Rio lives. The purpose of the game is to guide and help Rio escape from the space station. The novelty of the game consists of using vocal commands to control the main character, instead of a standard keyboard. However, LifeLine does not use NLP to evaluate the commands, but maps the user's instructions to a list of approximately 5000 words and 100.000 phrases, such as: "walk", "run", "stop", "go back", "shoot", or questions – e.g., "what should we do?", "where am I?".

Fig. 2. Lifeline gameplay [16].

2.3 Crystal Island

Crystal Island (see Fig. 3) is a narrative-centered learning environment, developed specifically for eighth-grade microbiology students of the Center for Educational Informatics in North Carolina State University [17]. The purpose of the game is to identify the source of an infectious disease that spreads through a research center. The main character can gather and manipulate objects, take notes, view posters, read books, use laboratory equipment to test different items for pathogenic compounds, and keep a diagnosis worksheet of the gathered knowledge. At the end of the game, players must diagnose the source of the disease and synthesize a cure based on the knowledge acquired during gameplay. This serious game does not use NLP to interact with the characters, but it is of interest for the current work from design and object-interaction perspectives.

Fig. 3. Crystal Island gameplay[17].

3 Escape from Dungeon – A Serious Game using NLP

The first two games presented in the previous section, Façade and Lifeline, are designed to entertain users by controlling the protagonist using vocal commands in order to solve puzzles and win the game. One issue is that the commands are predefined, hardcoded, and contextualized for the current situation. The last game, Crystal Island is a serious game that teaches users microbiology in an interactive and entertaining way. Yet, the game is not flexible regarding the content of the story and it is limited to learning only one discipline, namely microbiology.

Escape from Dungeon is designed to overcome the above drawbacks. The game supports interactions in natural language and tests users through a predefined general knowledge questionnaire. The main advantage from other existing serious games is the possibility to change the content of the preloaded questions to match the targeted educative content.

3.1 Gameplay and Story

Escape from Dungeon is a first-person game. The graphical interface is viewed in a first-person view from the player's perspective; therefore, the environment is more realistic, users can better relate to the surroundings and feel as if they are part of the game. The purpose of the game is to escape from a dungeon by solving puzzles and answering to general knowledge questions. The solutions to puzzles are found by interacting with objects in each room, using only vocal commands. The background story is that you are only capable to observe the environment and ask a "butler" who is with you to perform actions on your behalf. Currently, players can explore two rooms. The first room contains a puzzle that must be solved to progress to the second room. Instructions on movement and interactions using voice commands are given at the start of the game using a computer-generated voice. Players are encouraged to explore and test different commands such as: "go", "stop", "go to the table", "pick up the key", etc.

Support for Google Virtual Reality was integrated into the game to enhance the visual experience. When deployed on Android, any Virtual Reality headset can be used as an extension to create a virtual world around the player, as seen in Fig. 4.

3.2 Integrated Technologies

Escape from Dungeon is based on Unity Game Engine [18] and integrates various technologies: Wit.ai [19] for extracting actions and intents from the text commands received from players, IBM Watson [20] for human voice recognition and Google VR [21] – for virtual reality support.

Unity Game Engine is a 2D and 3D system that allows developers to build game scenes, import and build assets and animations, interact with objects etc. The games developed using Unity Game Engine can be deployed across desktop, mobile, console or VR devices. This tool was chosen for development due to its ease of use.

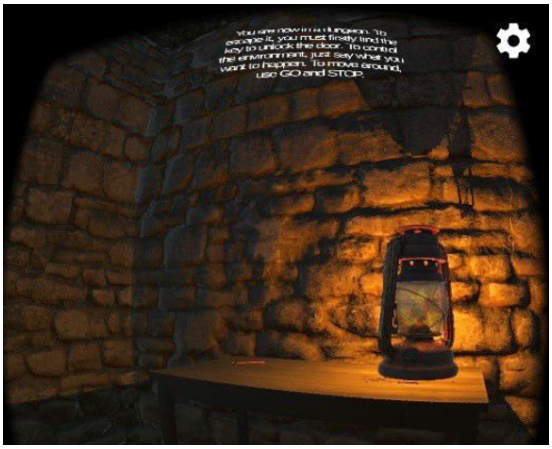

Fig. 4. Escape from Dungeon – Virtual Reality screenshot.

Wit.ai. is a tool that uses NLP techniques to transform text input into structured data**.** The user interface available to developers contains multiple menu items, but the most important one is the "Understanding" area, where developers use training examples to identify locations, emotions, or custom defined entities and intents. Intents are used to understand the meaning of a sentence, while entities are variables that define the details of a user task. Entities and intents define the application's behavior and shape how it interacts with users. The NLP model must cover all possible intentions and understand and predict the most likely intents and corresponding entities received as input. The recommendation from Wit.ai documentation is to use built-in entities when possible, as these are trained across the whole Wit.ai dataset.

An example of training using the web interface is shown in Fig. 5. Given the phrases: "I would like to be transported to the door. Thank you.", the engine considers that the input contains an intent for the action of "going" (from the verb "transported"). It also found a positive sentiment ("thank you") and a user-defined entity "door". Wit.ai is confident in the results, hence the high probabilities of 99.8%, 86.6% and 95.8%, respectively. The intents and entities predefined and user-defined in the application are displayed in Fig. 6.

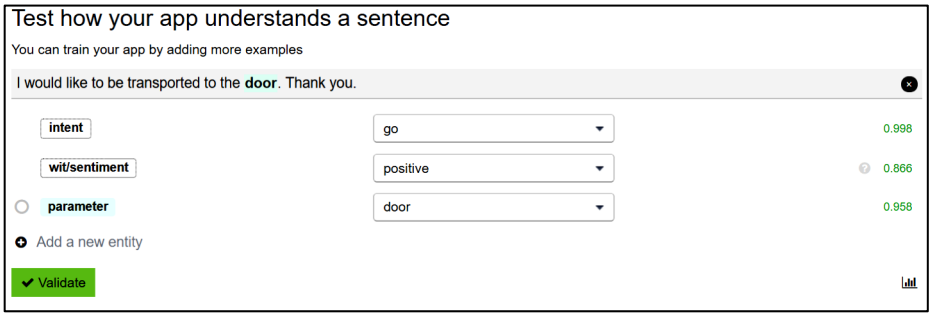

Fig. 5. Wit.ai web training interface [19].

Besides text processing, Wit.ai also offers speech understanding tools: audio clips encoded in wav, mpeg3, ulaw, or raw formats. The files can be sent to Wit.ai, converted to text, and then processed by the pre-trained model. However, a major drawback of using Wit.ai for speech-to-text is the long period of time between the moment of speech and the moment the actions are executed. Saving the audio clip to a file, importing it back into Unity, transferring it to Wit.ai, and deciphering the text out of the audio file takes several seconds, delay which is neither practical, nor acceptable, making the game flow slow-paced. The solution for this issue was to use a streaming speech recognition tool.

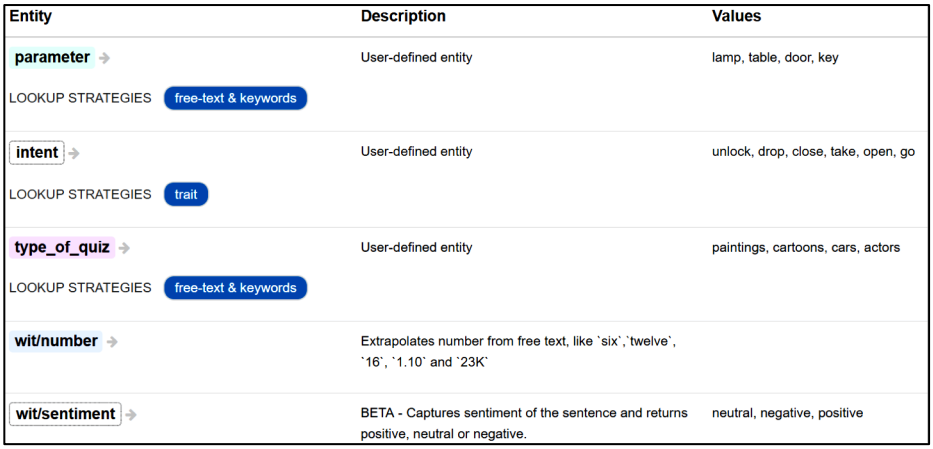

Fig. 6. Wit.ai user-defined entities and intents [19].

IBM Watson: speech-to-text and text-to-speech. This framework can recognize words while they are spoken and corrects misunderstood words, depending on context; after a short period of pause in speech, it automatically forms the sentence and returns it as a final JSON object, containing every word and its correctness probability. Watson speech-to-text also includes a Unity Game Engine plugin, making the integration process easier.

Google VR is a virtual reality platform developed by Google. In a VR world, players can interact with objects in a three-dimensions manner, creating the perception that they are part of the scene. Currently, three types of VR simulations are available: non -, semi -, or fully immersive. Fully-immersive simulations are commonly used for entertainment purposes and were selected for our game. In this category fall games requiring head-mounted displays, gloves, headphones, etc.. Google produces a head-mounted display called Google Cardboard [22], which is affordable and easy to assemble by end-users.

3.3 Game Flow and User Interface

The commands spoken by users are passed through IBM Watson to Wit.ai system. When responses are received from Wit.ai, all intents are searched through the implemented action methods and coupled with entities parameters. For example, a command like "Move to the door and open the door" will find "door" as a parameter entity, "move" and "open" as intents. For every intent, a corresponding method is applied for each parameter. If an intent cannot be identified, the player's natural language input is ignored, and he/she is asked to repeat the desired command.

Unity Game Engine supports tag definition and allows attaching tags to any virtual object rendered in the game. The link between parameter entities and game objects is made through tagging: each parameter is searched by its tag in the active scene. Similar to the case of intents, only the parameter with a corresponding tag is used, while the rest are ignored.

For example, the player wants to pick up the key and the lamp in Fig. 7. The entity is represented by the word “key”, while the object with the same tag name is searched in the room. If the object exists and is in the visual range of the character, within a reachable radius, the object is placed into the player’s inventory. In the current version, only one item is allowed in the inventory – when reaching for another item, the object currently stored in the inventory is dropped on the floor and the player is notified accordingly.

While the first room was oriented towards entertaining, familiarizing the player with possible actions and testing creativity, the second room requires correct answers to general knowledge questions (see Fig. 8). The player is asked to choose from a range of quiz domains, such as: paintings, cars, actors, cartoons, etc. Upon choosing a category, four images will appear one after the other in the picture frame.

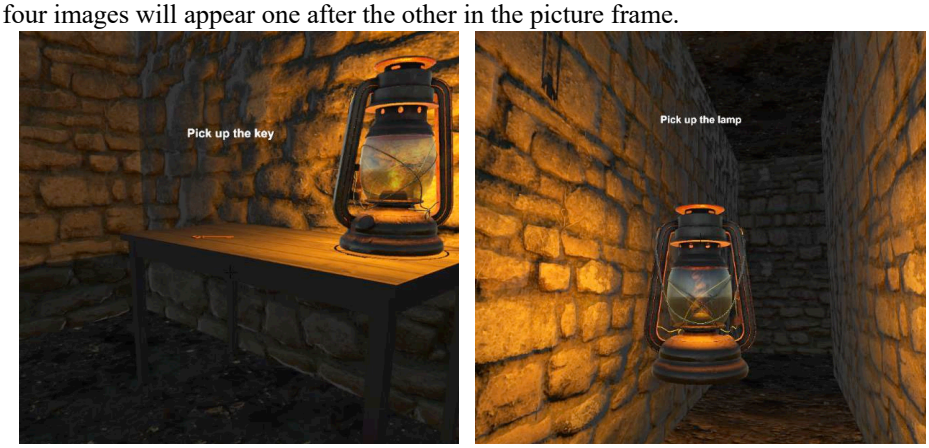

Fig. 7. Intents for picking up a key and a lamp.

The user must correctly guess what the image represents to escape the dungeon. Four choices are displayed next to the image and users must pick one of them and say the chosen number out loud. If the answer is correct, the next image will appear. In case the player makes a wrong choice, a message is shown, and the user is offered the

possibility to answer again. If the user has no knowledge of the content presented in the picture frame, he/she can guess all the possible answers until he/she learns the correct answer. This is a general example of learning while playing, and can be developed furthermore with other quiz examples, replacing the pictures with popular music, lines from classic movies, mazes of increasing complexities with potential traps, or specific information taught during class lectures.

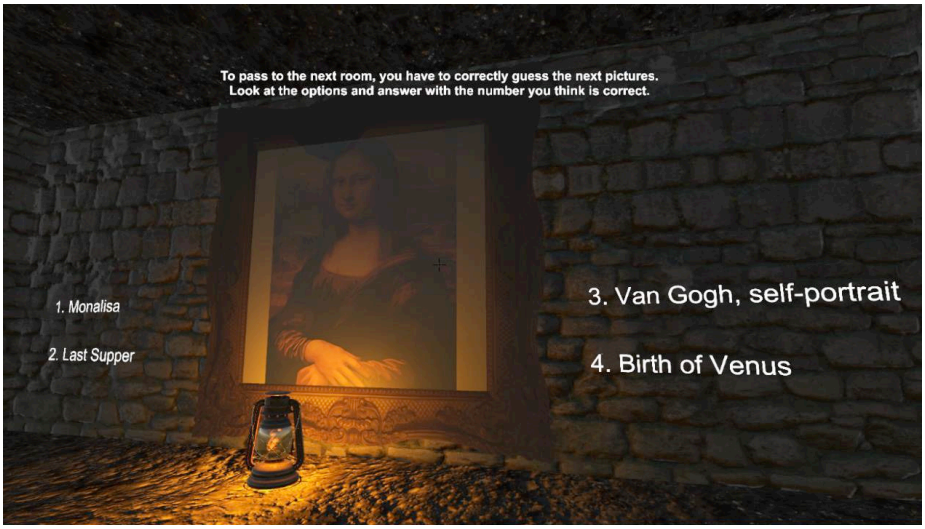

Fig. 8. Second challenge room.

4 Results

A survey was conducted with ten users, six males and four females, 3 with ages lower than 18 years old, 5 aged 18-25, while the rest were above 25 years old. The users were asked to respond to a 15-question survey with ratings on a 5-point Likert Scale (1–completely disagree; 5–completely agree) and three free input questions, covering their perspective on our serious game. Out of the first 15 questions, questions one to three ask for a general feedback: the impressions of the game, if the users understood the game's purpose, and if they fulfilled it. The next three questions refer to the graphics, VR experience, and the user interface. Questions seven to ten focus on NLP, action parsing and understanding, while the last questions concentrate on game content and future development. The three free input questions allowed users to express themselves about the game.

In terms of reliability statistics, Intraclass Correlation Coefficients [23] and Cronbach's Alpha [24] were computed based on the first 15 questions of the survey. The corresponding values, .437 for ICC and .625 for Cronbach's Alpha, denote a low level of agreement between the users, mostly regarding the accessibility of the game and game mechanics. The feedback gathered from the free input questions was positive. The users emphasized the fun and innovative aspects of the game, required more scenes

to play and increasing the difficulty level of the scenes. 50% of the users faced problems while interacting with the character, as the voice recognition software did not respond as expected. Users complained that their voice commands were not properly understood, or the character did not follow all commands.

5 Conclusion and future development

Compared to the entire gaming industry, the number of serious games is very small and the major drawbacks are represented by the limited user interface and interaction - lack of good animations or textures; lack of flexibility by focusing only on one educational field; the usage of predefined range of options, without giving users the opportunity to freely interact with the game. The current paper introduces our serious game – Escape from Dungeon – that uses Natural Language Processing techniques including speech recognition and speech generation, as well as Virtual Reality interaction using Unity. The game environment is represented from a first-person view, allowing users to immerse in the game. The main character is controlled through vocal commands, from which user intents and entities are extracted. These and later on mapped on the available scene elements, coupled with possible player actions.

The game was tested by 10 users throughout a pilot test and it was considered fun and innovative. From the user feedback, the game scenes require minor improvements: more objects should be available in the rooms and serve as escape tools, the escape scenario should be more complex, and should require the usage of multiple objects. Another user idea is adding a mechanism for combining two or more raw materials to create different objects. New actions could be introduced for a more realistic look, namely: "throw" – could be an action used to throw the object contained into inventory in front of the character or to a designated target; "break" – where the user asks to destroy an object in the scene using an item from his inventory or with his bare hands; "craft" – where the main character uses items in his inventory to create improved objects or repair already crafted ones.

The game can be enhanced by adding multiplayer support that encourages collaboration through the implementation of a virtual classroom, where teachers are offered the possibility to create personalized questions and quizzes, set a time limit for exiting the dungeon or disabling the use of some objects. This feature could also include template facilities to ease the online sharing of the content between teachers.

Acknowledgments
This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-III 72PCCDI/2018, ROBIN – "Roboții și Societatea: Sisteme Cognitive pentru Roboți Personali și Vehicule Autonome" and by the Operational Programme Human Capital of the Ministry of European Funds through the Financial Agreement 51675/09.07.2019, SMIS code 125125.

References

1. Susi, T., Johannesson, M. & Backlund, P., S. of H. and I.: Serious Games – An Overview: Technical report, HS-IKI-TR-07-001. , Sweden (2006).
2. Eck, R. Van: Digital Game-Based Learning : It ' s Not Just the Digital Natives Who Are Restless …. Educause Review. 41, 1–16 (2006). https://doi.org/10.1145/950566.950596.
3. Westera, W., van der Vegt, W., Bahreini, K., Dascalu, M., van Lankveld, G.: Software Components for Serious Game Development. In: Connolly, T. and Boyle, L. (eds.) 10th European Conf. on Games Based Learning. pp. 765–772. Reading UK, Paisley, Scotland (2016).
4. Kampa, A., Haake, S., Burelli, P.: Storytelling in serious games. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 521–539. Springer Verlag (2016). https://doi.org/10.1007/978-3-319-46152-6_19.
5. Merdivan, E., Singh, D., Hanke, S., Holzinger, A.: Dialogue Systems for Intelligent Human Computer Interactions. In: Electronic Notes in Theoretical Computer Science. pp. 57–71. Elsevier B.V. (2019). https://doi.org/10.1016/j.entcs.2019.04.010.
6. Hennigan, B.: Making the case for NLP in dialogue systems for serious games. In: 8th international conference on natural language processing (JapTAL), 1st workshop on games and NLP (2012).
7. Hoy, M.B.: Alexa, Siri, Cortana, and more: an introduction to voice assistants. Medical reference services quarterly. 37, 81–88 (2018).
8. Tur, G., De Mori, R.: Spoken language understanding: Systems for extracting semantic information from speech. John Wiley & Sons (2011).
9. Zhang, C., Fan, W., Du, N., Yu, P.S.: Mining user intentions from medical queries: A neural network based heterogeneous jointly modeling approach. In: Proceedings of the 25th International Conference on World wide web. pp. 1373–1384 (2016).
10. Xu, P., Sarikaya, R.: Convolutional neural network based triangular crf for joint intent detection and slot filling. In: 2013 ieee workshop on automatic speech recognition and understanding. pp. 78–83. IEEE (2013).
11. Lampert, C.H., Nickisch, H., Harmeling, S.: Attribute-based classification for zero-shot visual object categorization. IEEE transactions on pattern analysis and machine intelligence. 36, 453–465 (2013).
12. Bucher, M., Herbin, S., Jurie, F.: Improving semantic embedding consistency by metric learning for zero-shot classiffication. In: European Conference on Computer Vision. pp. 730–746. Springer (2016).
13. Sonawane, K.: Serious Games Market Outlook: 2023, https://www.alliedmarketresearch.com/serious-games-market, last accessed 2020/10/02.
14. Wijman, T.: The Global Games Market Will Generate $152.1 Billion in 2019 as the U.S. Overtakes China as the Biggest Market, https://newzoo.com/insights/articles/the-global-games-market-will-generate-152-1-billion-in-2019-as-the-u-s-overtakes-china-as-the-biggest-market/, last accessed 2020/10/20.
15. Mateas, M., Stern, A.: Façade: An experiment in building a fully-realized interactive drama. In: Game developers conference. pp. 4–8 (2003).
16. Konami: LifeLine, https://www.playstation.com/en-us/games/lifeline-ps2/, last accessed 2020/02/10.
17. Rowe, J., Mott, B., McQuiggan, S., Robison, J., Lee, S., Lester, J.: CRYSTAL ISLAND : A Narrative-Centered Learning Environment for Eighth Grade Microbiology. In: Education. pp. 11–20 (2009).

18. Brodkin, J.: How Unity3D Became a Game-Development Beast, http://insights.dice.com/2013/06/03/how-unity3d-become-a-game-development-beast/.
19. Wit.ai Documentation, https://wit.ai/docs, last accessed 2020/02/05.
20. High, R.: The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works. International Business Machines Corporation. 1, 1–14 (2012).
21. LLC, G.: Google VR, https://arvr.google.com/vr, last accessed 2020/05/02.
22. Yoo, S., Parker, C.: Controller-less interaction methods for Google cardboard. In: Proceedings of the 3rd ACM Symposium on Spatial User Interaction. p. 127 (2015).
23. Koch, G.G.: Intraclass correlation coefficient. In: Kotz, S. and Johnson, N.L. (eds.) Encyclopedia of Statistical Sciences. pp. 213–217. John Wiley & Sons, New York, NY (1982).
24. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. Psychometrika. 16, 297–334 (1951).