

# Edubot: a proposed modeling approach for a chatbot system for student support in distance education

N. Ntaliakouras<sup>1</sup>, I. Moustaka<sup>2</sup>, G. Vonitsanos<sup>3</sup>

<sup>1</sup>University of Patras / Department of Computer Engineering & Informatics  
ntaliakour@ceid.upatras.gr

<sup>2</sup>University of Nicosia / Department of Languages and Literature  
School of Humanities and Social Sciences  
moustaka.i@live.unic.ac.cy

<sup>3</sup>Hellenic Open University / School of Humanities  
mvonitsanos@eap.gr

## Abstract

Computer Science is ever becoming a matter of grave concern in various aspects of everyday life. One of the most important fields of Computer Science is Artificial Intelligence, widely known as “AI”. AI refers to the intelligence added to machines mainly via smart programming and engineering. In the area of Education, we can clearly notice the increasing penetration of Computer Science and especially AI, which provides students with the framework to interact with the learning environment and knowledge on its whole. Also, the huge data size needed to be processed led to the use of NoSQL databases that have a lot more to offer than just presenting solutions to scale problems. Moreover, AI tools are capable of improving educational procedures in the form of chatbots that interact with the users via text or vocal interfaces. In this proposed work we present a novel chatbot system for student service in distance education using modern machine learning techniques. The training data was acquired via a web scraping method out of student forums. A NoSQL database was used for the storage and manipulation of our data. The proposed system is expected to be tested on a sample of students to evaluate its functionality.

**Keywords:** Computer Science, Distance Education, Artificial Intelligence, Machine Learning, NoSQL Databases, Chatbots.

## *1. Introduction*

Education, being a vital part of every person’s life and simultaneously the basic mean leading to the improvement of people’s living standard, is an excellent barometer that indicates the progress degree of each society. Nowadays, the most popular form of education is distance education. According to Keegan (1980), Holmberg, Loi, Peters, and Moore gave the definition for distance education, portraying that it relates to

miscellaneous modes of multilevel study not continuously supervised by teachers in specific classrooms, but rather taking benefit of a tutorial organization planning and guidance via electronic, mechanic or other devices. Artificial Intelligence is the sector of Computer Science that contributes to the improvement of distance education as it provides teachers and students with the framework to interact and form high-quality studies. AI ideally provides chatbots which are actually computer programs that mimic conversation with people (Wong, 2016). Chatbots are a novel technology ranging their use in a varicoloured area from information retrieval and answering questions, shopping as an assistant and in museums as a guide, to language learning as a partner and in education (Shawar and Atwell, 2007). Their most popular use being an irreplaceable tool facilitating the learning and practicing process in a language, renders chatbots an excellent such choice. Since distance education has improved, chatbots can also contribute by creating tailor made tests for every student level and interacting with them. On the other hand, the absence of a teacher's presence as the main characteristic of distance education, creates the need to provide students with services that fill them with confidence and ensure them that their teacher's assistance will be at hand, any time.

Many researchers have raised the issue of the use of chatbots in the field of education by combining chatbot technology with machine learning algorithms, techniques and methodologies, e-learning environments and generally educational applications and services. In recent years, an AIML (Artificial Intelligence Markup Language) based chatbot was developed to answer FAQs relevant to universities in a most accurate and efficient way via 24/7 service to users interested in a specific university (Ranoliya et al., 2017). Addressing this need, Robert Schmidt built the Yellow Circle learning platform in 2014. This free online learning platform for cloud computing uses a text-based Amazon chatbot with a voice-based chatbot to supply students with a unique teaching environment; they receive personalized answers according to their individually distinct teaching and learning style. Another impressive work is NLAST (Natural Language Assistant For Students), an assistant for students using natural language. NLAST refers to a chatbot that interacts with students providing them with the ability to carry out several actions relevant to their studies using natural language (Fonte et al., 2016). Research on the use of chatbots in education continues in the same year when Carmen Holotescu built MoocBuddy. MoocBuddy is a chatbot for Facebook Messenger that recommends MOOCs (Massive Open Online Courses) to users, according to their social media profile and interests. In this way, users have an interactive and friendly tool that helps them find the appropriate MOOC and related to it information. Also, there is research on the use of educational chatbots for visually impaired people that can utter a question in spoken natural language using

Google Voice Search, find the answer and return the result as both text and voice (Kumar et al., 2016).

Our research is primarily concentrated on the development of Edubot, a novel chatbot system for student service in distance education. Edubot comes to empower students of distance learning, granting them certainty that the received service is of high-quality and most productive. It can be utilized by any distance learning university to address students' ever increasing needs effectually. The above-mentioned technology is particularly significant for this type of education as it meets different requirements of a large number of students, who might either be in a working schedule with but limited time to study or maintain an irregular daily program. These restrictions call for a handy tool that will provide continuous support on their study. The remainder of the paper is organized as follows. Section II presents the definition of the problem while Section III briefs about the preliminaries in particular artificial intelligence, machine learning, NoSQL databases and distance learning. Section IV presents the proposed chatbot model design. Finally, Section V presents our conclusions while Section VI draws directions for future work.

## ***2. The definition of the problem***

Distance instruction is mainly characterized by the detachment between professor and learner as mentioned above. Therefore, the main problem for distance learning students is having to cope with a variety of problems without the instant help of their professor. Let us bring forward a simple example of a working student who studies at late hours. The scenario requires that this student face an assortment of difficulties and is in need of immediate help from the professor. Unfortunately, these needs cannot be met as the professor has specific, and many times strict, communicating hours. In such cases, it is crucial for students to have an all-day support that our proposed system will provide at any time. Thus, the obstacle of professor-student dissociation can be tackled by a system that “understands” the student and provides user-friendly answers.

All existing literature focuses on supporting online learning. The foremost difference with our proposed system is that it harnesses the techniques and means of online learning to support distance education, which is extremely difficult as distance learning needs special instructive material and a distance learning institution has to deploy all available mechanisms to support this type of education. Clarifying instructions intended for students need to be present and professors are vital to be specially trained so as to support distance education. All these factors provide a framework, where the educative process must be continuously supported with the conditions leading to the indispensability of a system like Edubot.

### ***3. Preliminaries***

#### ***3.1 Artificial Intelligence***

Artificial Intelligence, widely known as "AI", is a significant sector of Computer Science referring to the intelligence added to machines mainly via smart programming and engineering (McCarthy, 2007). It comprises the science of creating computer systems that can understand their environment and react according to its variables and the feedback they receive. Russell and Norvig (2016) define AI as “the study of agents that receive percepts from the environment and perform actions”. A function that maps percept sequences to specific actions is implemented by each agent. The agents can have the roles of reactive agents, real-time planners or decision-theoretic systems based on the different ways their functionality is represented.

#### ***3.2 Machine Learning***

The definition of machine learning describes a computer sciences field starting from pattern recognition study and computational learning theory and proceeding to artificial intelligence. It declares the process of algorithm construction that can learn and make predictions after the elaboration of data sets. The main idea is the construction of a data-driven model that predicts and make choices according to the input data and not in a static way. This model has the ability to be adjusted to each situation prognosticating and deciding the following (Simon et al., 2016). The machine learning processes can be interpreted by using paradigms of a pattern recognition including identifying speech processes, human handwriting or distinguishing images through computer procedures. Artificial intelligence tools are capable of improving educational methods in the form of chatbots. These applications are systems that interact with the users via text or vocal interfaces.

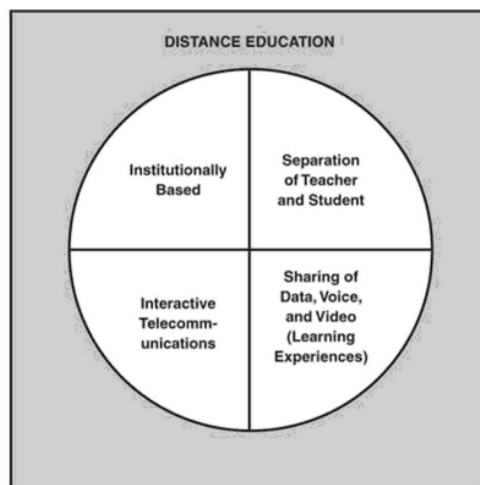
#### ***3.3 NoSQL Databases***

The real meaning of this generic term NoSQL, is that this type of databases does not follow the principles of the traditional relational databases. The data they are dealing with is not relational, and they are superior to traditional databases which encounter scalability and availability problems because of the data size. NoSQL defines a filter to determine precisely the databases that fulfill the requirements mentioned before (Vaish, 2013). Nowadays, it is stated that NoSQL databases have a lot more to present than just offering solutions to scale problems.(Corbellini et al., 2017). Apache Cassandra is an extensively scalable NoSQL database. Some of its characteristics are the provision of availability which is continuous, its scalability which is linear and the

simplicity in operating on multiple servers without any single point of failure (Datastax, 2018).

### 3.4 Distance learning

Distance education is the scholastic process described by geographical distance, time distance and even intellectual distance between professor-student or teacher-learner. The main difference between traditional education and distance education is that traditional education takes place in the same time and place, but distance education can take place either in a different place and different time called asynchronous distance learning, or different place and same time known as synchronous distance learning (Simonson, Smaldino, & Zvacek, 2015). In both forms of distance learning, synchronous and asynchronous, the educator advises, supports and encourages the learner to achieve the goals of the learning process (Eumorfopoulou & Lionarakis, 2015).



**Figure 1.** The four components that define distance education (Simonson, Smaldino, & Zvacek, 2015)

Figure 1 depicts the main components that comprise the definition of distance learning. Notice that it is institutionally based, which differentiates it from self-study. The second component is the actual distance that separates professors-students referring not only to geographic terms but also to time. The intellectual separation of teachers and students is also important and its reduction is of the main goals of each distance education system. Observe that the interactive telecommunications refer to the learners' ability to interact with each other, with instructional resources and with

their teachers. It is a very important component that must be available, commonplace and relevant. Last but not least, the learning experience is consisted of all these means and resources that are properly designed and organized in order to promote learning through interaction, which is a pivotal objective of education (Simonson, Smaldino, & Zvacek, 2015).

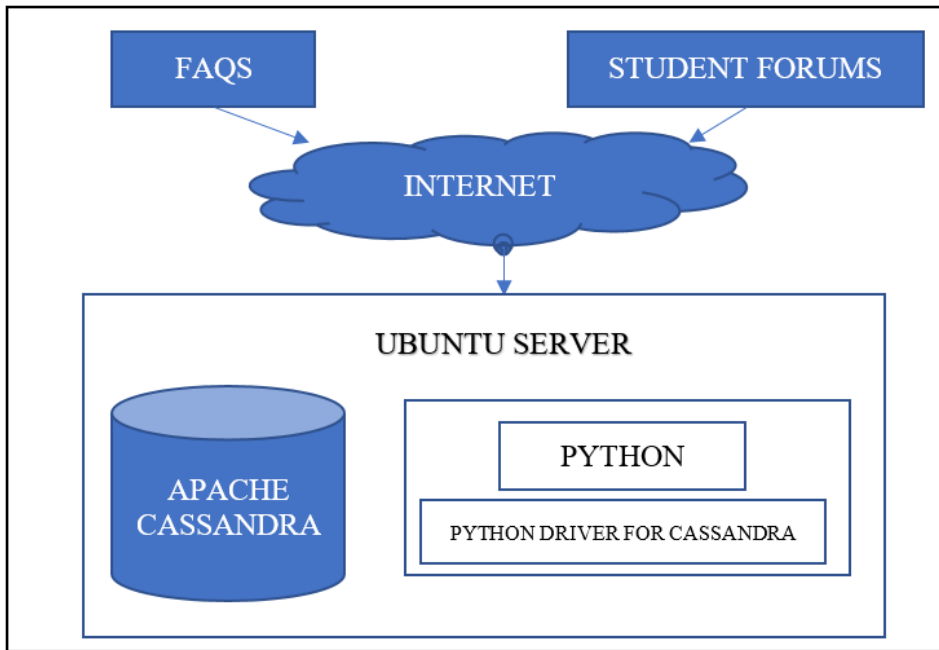
## ***4. Implementation***

### ***4.1 Dataset Description***

In the context of this project, a set of training data was constructed to perform the experiments for a specific case study. The training data was acquired from the web, and mainly from several student forums regarding universities that use distance learning methods. The data was collected in multiple files by using a web scraping method. It was also structured and formatted in input-output pairs to provide examples for the training phase. A large number of files and the heterogeneity between them led to the conclusion that they had to be preceded by an initialization, be divided into small groups and then processed more efficiently (Kan, 2014). The files were then grouped into different categories, whether they contained data about educational or student service purposes. In the current development, we focused on the application itself and not the data and their relational theories. The method of modeling by query was used and led us to the creation of database tables, containing all the needed data, giving the ability to make the required aggregations. Also, a table containing only the data needed for our project was created, complying exactly with Cassandra result oriented model and acting as a pre-built result set (Kan, 2014).

### ***4.2 Model and System Description***

Our proposed system is based on a single Ubuntu server running a single node Cassandra cluster. The virtual machine was created using Oracle VirtualBox, which allows the user to save a particular state in the form of a snapshot. By using the snapshot features, a user can roll back the Virtual machine in a previous state in case of a failure. The development tool Eclipse, which is commonly used for creating Java applications is also installed. The system overview is illustrated in Figure 1, including all major components and their implementation.



*Figure 2. The overview of the system*

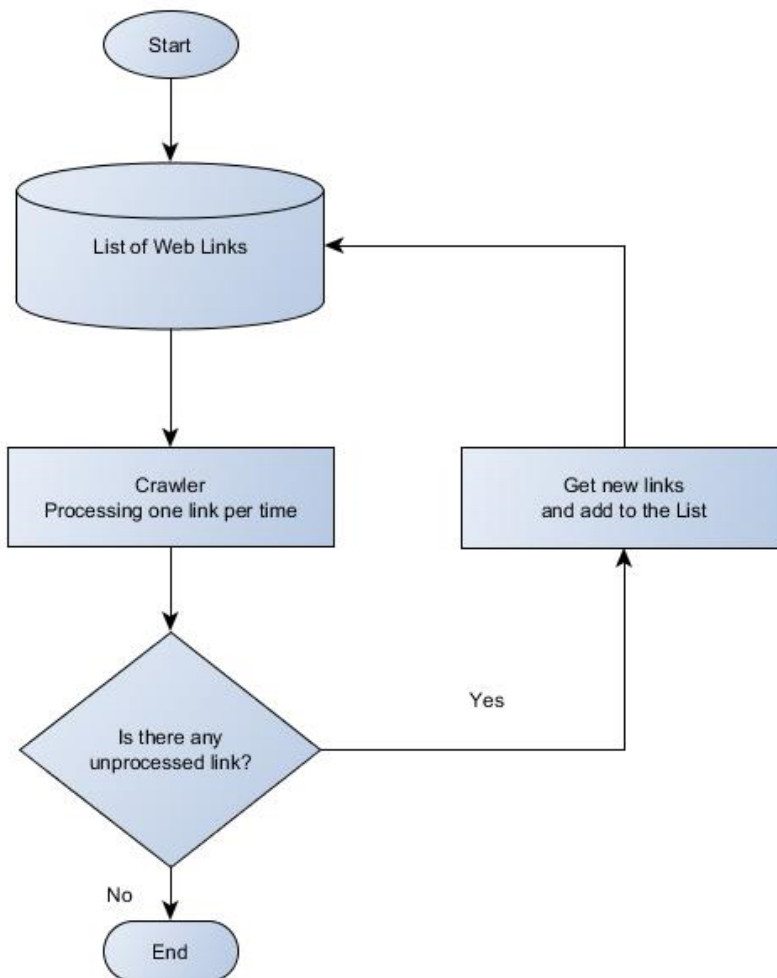
To test the system more efficiently, all the latest versions of the programs were downloaded. The specific versions of the applications are shown in Table 1.

*Table 1. Installed programs and versions*

Application	Version
Ubuntu Server	16.04.3
Java Virtual Machine	8
Apache Cassandra	3.11.1

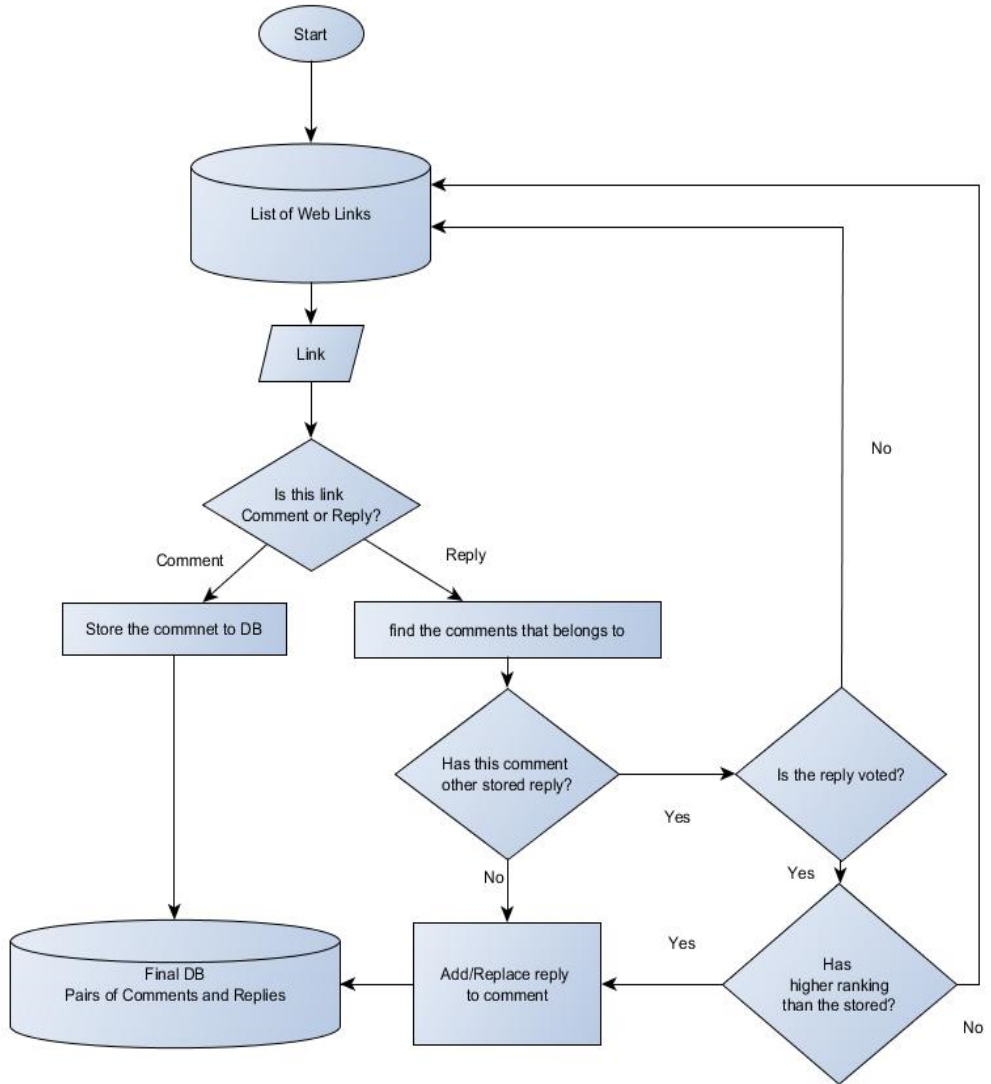
For a robust chatbot platform, we need to separate data into comments and replies. The comment is the input, and the reply is the output. One main problem that occurred is the difficulty of scraping web forums. First of all, some comments may have no replies and also a comment may be reiterated in different threads and so a mechanism must convert these comments into one, the most representative. Also, some threads may be very old and the replies deferred. Of course, in student forums,

not all comments have at least one reply. We can even encounter cases where comments have many replies, and therefore we have to opt for a single reply. Also, not all replies are correct, and we should take into account the upvoted ones which are marked as appropriate by other users. If the forum doesn't provide a voting system, we can iterate the scraped file many times to find more appropriate replies in other parts of the data. In Figure 3 and 4 below, we can see the web crawling process and the scraping method diagrams that depict the flow of the data and the processes. Finally, not all forums have the same structure which makes it difficult to analyze the required data to make the comment-reply pairs.



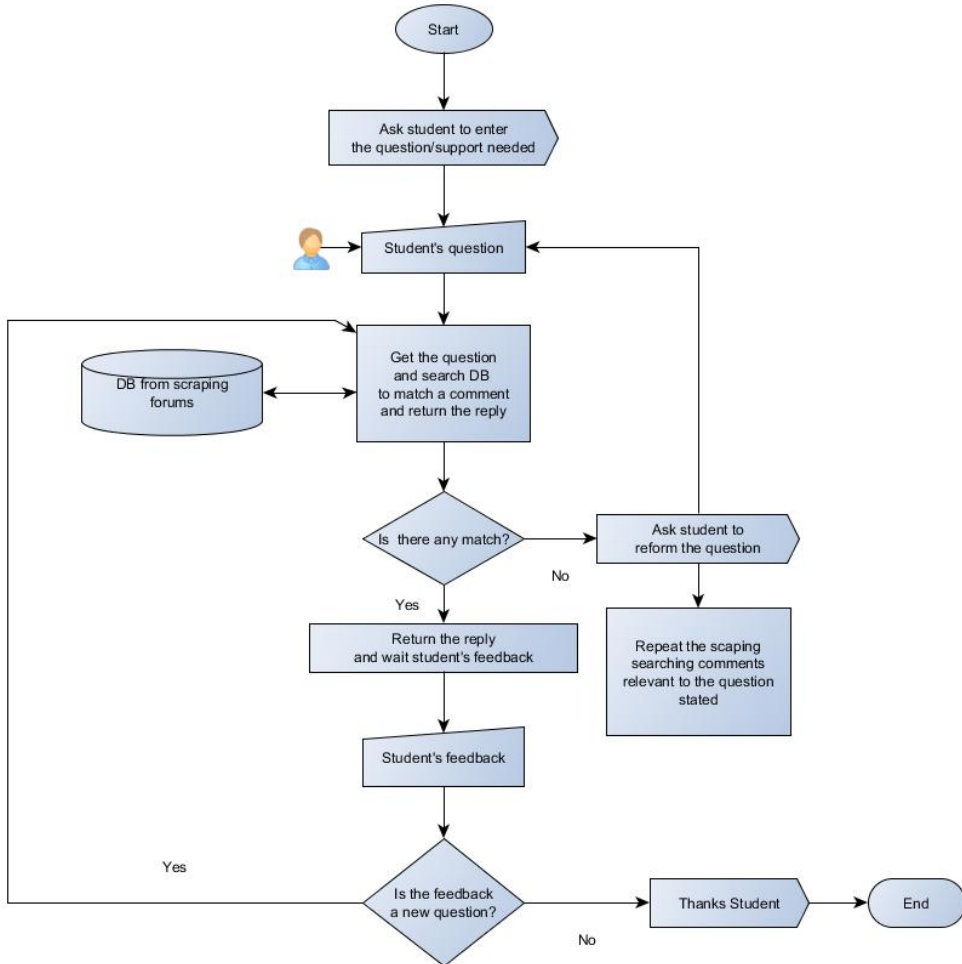
**Figure 3.** The web crawling process





**Figure 4.** The forum scraping method

Figure 5 below shows the processes and the interaction of Edubot with the student.



*Figure 5. The Edubot flowchart*

## 5. Conclusions

Artificial Intelligence tools like chatbots are already very popular for web applications and scientific intelligent systems. Some other possible applications in the academic section could be the digital indexing of teaching material, and direct feedback on newly available teaching material. An intelligent calendar could also be a useful application. In this paper, we have proposed a model for an interactive chatbot

for universities in a distance learning environment. Intelligent automation is needed in administration and teaching activities throughout the university structure. The results of implementing these applications could be the enhancement of productivity, increase in the reputation of the university, and improved quality of service. The reduction of costs could also lead to the reduction of tuition fees. Of course, in such situation some risks can occur, so the right tool should be previously identified and accepted first by all faculty, staff, and then by students. The upgradability and interoperability should be taken into account, and of course, there should be ways to manage such autonomous systems dependently. The existing administration should always use a step by step approach. The aim of this paper was to investigate the potential of the development of an educational chatbot providing high-quality services to students and professional development of the faculty and staff. In distance education, the need for continuous support is vital as its processes that concerns students does not happen in a specific place and time. It depends on each student's program and way of living. Edubot can easily cover this need.

## **6. Future Work**

In such an innovative area of educational chatbots, there is a lot of possible future research to suggest. First of all, we should consider the criteria that define the proper implementation of a chatbot and more particularly its role as a tutor, a student helper, or a front office gateway. An extended examination can be carried out to determine if a chatbot can become a proper pedagogical gateway or a substitute for student services office. It is known that every university providing distance education studies operate in a different way to provide quality services to the students. This indicates that each case will also need a separate chatbot system to provide the required services and the diversity of the artificial technologies should be further researched.

## **References**

Corbellini, A., Mateos, C., Zunino, A., Godoy, D., & Schiaffino, S. (2017). Persisting big-data: The NoSQL landscape. *Information Systems*, 63, 1-23.

Datastax, (2018a, 07 20). *About Cassandra*, Retrieved from Datastax: <http://docs.datastax.com/en/archived/Cassandra/3.x/Cassandra/cassandraAbout.html>

Eumorfopoulou, E., Lionarakis, A. (2015). Teacher's role in school distance education using models of conventional and distance learning methods. *Proceedings of the 8th International Conference on Open & Distance Learning "Innovation and Research", 7-8 November 2015*, (Volume 8, No. 1, page 1-15). Athens: Hellenic Open and Distance Learning Network.

- Fonte, F., Mikic, A., Nistal, M., Rial, H., & Rodriguez, M., (2016). NLAST: A natural language assistant for students. In *Global Engineering Education Conference (EDUCON), 2016 IEEE*. IEEE, pp. 709–713.
- Holotescu, C. (2016). MOOCBuddy: a Chatbot for personalized learning with MOOCs. In *RoCHI* (pp. 91-94).
- Kan, C. (2014). *Cassandra Data Modeling and Analysis*. Packt Publishing Ltd.
- Keegan, D., (1980) On defining distance education. In *Distance Education*, 1:1, pp. 13-36.
- Kumar, N., Chandar, P., Prasad, A., & Sumangali, K., (2016). Android-based educational Chatbot for visually impaired people. In *International Conference on Computational Intelligence and Computing Research (ICCIC), 2016 IEEE*. IEEE, pp. 1–4.
- McCarthy, J. (2007). What is Artificial Intelligence? Retrieved from <http://www-formal.stanford.edu/jmc/whatisai.pdf>
- Ranoliya, R., Raghuvanshi, N., & Singh, S., (2017). Chatbot for University Related FAQs. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Udupi, pp. 1525–1530.
- Russell, S., & Norvig, P. (2016), *Artificial Intelligence, A Modern Approach (Third Edition)*. England: Pearson Education Limited Editions
- Schmidt, R. (2014). *Learning System Customer Service Chatbot*. Georgia Institute of Technology.
- Shawar, Bayan Abu and Eric Atwell (2007). “Chatbots: are they really useful?” In: *Ldv forum*. Vol. 22. 1, pp. 29–49.
- Simonson, M., Smaldino, S., & Zvacek, S. (2015). *Teaching and Learning at a Distance: Foundations of Distance Education*, 6th Edition. Charlotte, North Carolina: Information Age Publishing, Inc.
- Vaish, G. (2013). *Getting started with NoSQL*. Packt Publishing Ltd.
- Wong J. (2016). What is a chat bot, and should I be using one? Retrieved from <https://www.theguardian.com/technology/2016/apr/06/what-is-chat-bot-kik-bot-shop-messaging-platform>

## Περίληψη

Η Επιστήμη των Υπολογιστών ενσωματώνεται όλο και περισσότερο σε διάφορες πτυχές της καθημερινής μας ζωής. Ένα από τα πιο σημαντικά πεδία της Επιστήμης των Υπολογιστών είναι η Τεχνητή Νοημοσύνη, ευρέως γνωστή ως "TN". Η TN αναφέρεται στην ευφυΐα που προστίθεται στις μηχανές κυρίως μέσω του έξυπνου προγραμματισμού και της μηχανικής. Στον τομέα της Εκπαίδευσης, μπορούμε να παρατηρήσουμε με σαφήνεια την αυξανόμενη διείσδυση της Επιστήμης των Υπολογιστών και ειδικότερα της TN, η οποία παρέχει στους φοιτητές το πλαίσιο αλληλεπίδρασης με το μαθησιακό περιβάλλον και τη γνώση στο σύνολό της. Επίσης, ο τεράστιος όγκος δεδομένων που απαιτείται για επεξεργασία, οδήγησε στη χρήση των βάσεων δεδομένων τύπου NoSQL που έχουν πολλά περισσότερα να προσφέρουν από την απλή παρουσίαση λύσεων σε περιπτώσεις μεγάλων δεδομένων. Επιπλέον, τα εργαλεία TN όπως είναι οι έξυπνοι βοηθοί αυτοματοποιημένης γραπτής επικοινωνίας (chatbots) είναι ικανά να βελτιώσουν τις εκπαιδευτικές διαδικασίες μέσω της αλληλεπίδρασης με τους χρήστες με χρήση κειμένου ή φωνητικών εντολών. Σε αυτή την εργασία προτείνουμε ένα καινοτόμο σύστημα που παρέχει υπηρεσίες σε φοιτητές που συμμετέχουν σε εκπαίδευση εξ' Αποστάσεως χρησιμοποιώντας σύγχρονες τεχνικές μηχανικής μάθησης. Τα στοιχεία της εκπαίδευσης του συστήματος αντλήθηκαν μέσω μιας μεθόδου ανάκτησης δεδομένων (web scraping) από φοιτητικά ιστολόγια. Μια βάση δεδομένων τυπου NoSQL χρησιμοποιήθηκε για την αποθήκευση και τη διαχείριση των δεδομένων μας. Το προτεινόμενο σύστημα αναμένεται να αξιολογηθεί η λειτουργικότητά του.

**Λέξεις Κλειδιά:** Επιστήμη των Υπολογιστών, Εξ' Αποστάσεως Εκπαίδευση, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, NoSQL Βάσεις Δεδομένων, Chatbots