Chatbots: An Analysis of the State of Art of Literature

Andréia Ana Bernardini[1], Arildo A. Sônego[2], Eliane Pozzebon[2]

[1] Escola Superior de Criciúma (ESUCRI) – Criciúma, SC - Brazil

[2] Universidade Federal de Santa Catarina (UFSC) – Araranguá, SC – Brazil

andreia@esucri.com.br, asonego@gmail.com, eliane.pozzebon@ufsc.br

Abstract. *Considering the context of Computer Science, Chatbots are computer programs that, through techniques of Artificial Intelligence, aim to simulate human behavior when at the moment of a dialogue through text messages. Despite being a relatively new area of study, the application of this concept has increased considerably, whether in academic settings or in commercial applications, such as situations involving assistants to clarify doubts on websites of manufacturers of electronic products or even in services such as telephony. Thus, this article proposes an analysis of the state of the art of research through a quantitative bibliometric analysis, supported by the Scopus database.*

1. Introduction

Artificial Intelligence (AI) denotes as a science that is dedicated to the study of systems that, from the perspective of any observer, act intelligently [Coppin 2015]. This is a new field of research, when compared to other sciences, since it began in 1956 [Russel and Norvig 2013]. The most accepted concept is of an intelligent agent, in which the symbolic and connectionist approaches can act in a collaborative way aiming at solving problems through a computational system [Franco 2014].

The study of AI is interdisciplinary, since it makes use of a large amount of knowledge from other areas, such as mathematical calculation and logic [Artero 2009]. Artificial Intelligence brings together a considerable variety of areas, such as learning and perception to more specific tasks such as games, demonstration of mathematical theorems, and diagnosis of diseases among others, being considered relevant to any intellectual task [Russel and Norvig 2013].

A Chatbot is a computer program that works with programming languages oriented to the manipulation of natural human language [Lima 2014]. The possibility of training a computer to establish dialogues with humans dates back to the 1950s, with Alan Turing. In his essay titled "Computing Machinery and Intelligence," Turing proposed a test consisting of a program that developed a text message conversation with an interrogator for five minutes [Coppin 2015]. The person should then guess if he was talking to a human being or a computer and if the machine tricked the interrogator for 30% of the time, the program would pass the test.

The basic principle employed in a Chatbot consists of an environment that receives questions in natural human language, associates these questions with a knowledge base, and then issues an answer [Fryer and Carpenter 2006]. Usually, a Chatbot works from the insertion by the user of a question or comment, and from this

moment the program answers the question, makes a comment or starts a new topic [Huang, Zhou and Yang 2007].

Examples of Chatbots [Russel and Norvig 2013] are the ELIZA, MGONZ, NATACHATA and CYBERLOVER projects. In addition to these, [Huang, Zhou and Yang 2007] indicates the PARRY environment. This list [Lima 2014] is complemented with the ALICE (Artificial Internet Computer Entity) system, activated in the year of 1995.

In this way, this article aims to demonstrate, based on the information reported by the Scopus platform, the state of the art of research involving the theme Chatbots.

2. Methods, Results and Discussions

In order to strengthen the credibility of the research regarding its results, we opted for the adoption of a systematic literature review and quantitative analysis. This process consists of a data survey in which rigorous reviews of academic publications are demanded for signs that may lead to the identification of evidences regarding a research topic or even a topic in a desired area [Freire 2014].

In this way, for the elaboration of this research, the Scopus platform (www.scopus.com) was chosen, international base responsible for scientific publications of interdisciplinary nature.

Initially, the term "chatbot" was inserted in the search window of the platform. This procedure reported the existence of 314 documents, distributed as shown in Figure 1. The originality of the theme is confirmed by the results presented, since the first publication registered dates from the year of 2002. In the same way, the total amount of documents about the subject can be considered small when compared to other topics, such as Internet of Things (IoT) with about 32.949 publications at the time of this research. It is also observed the linearity of these quantities over the years, and in the year 2016 a slight increase occurred and in 2017 this number presented the highest indices. The year 2018, being the current year, still presents little expressive numbers when faced with the others.
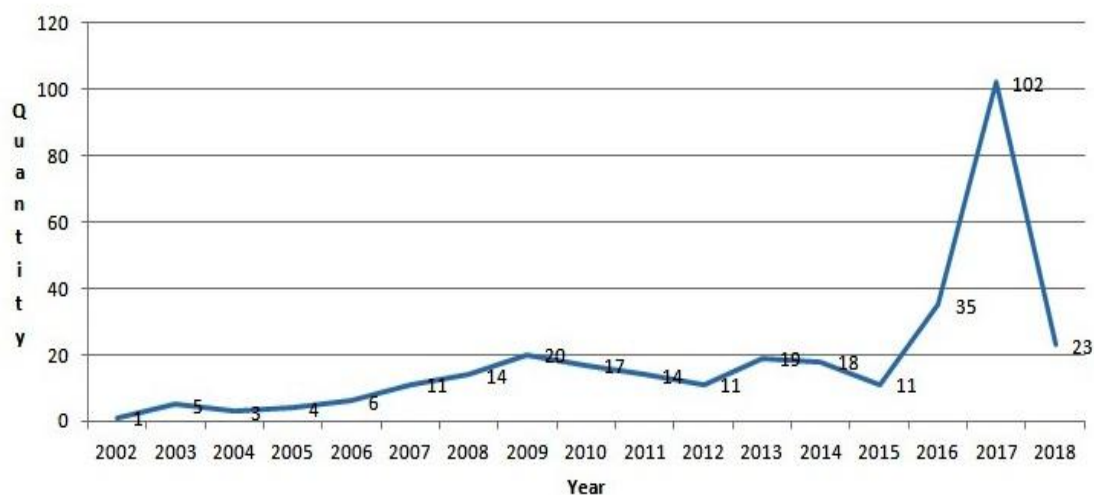

Figure 1. Evolution of polls related to Chatbots

Regarding the publications according to the country of origin, Figure 2 presents the results obtained. The ten countries that most published during the period chosen were considered. We can see the predominance of the United States, followed by Italy. In this context, the two countries hold 41.93% of all material produced. An interesting point to consider is the position of China, ranking only sixth in this classification given its traditional technological nature.
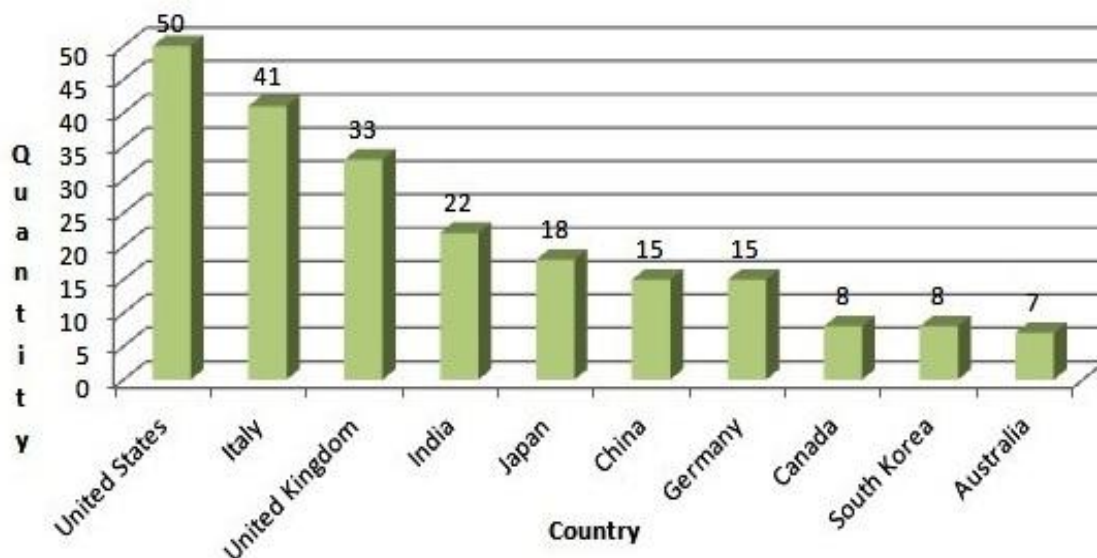

Figure 2. Publications by country

The next item, shown in Figure 3 and reported by the Scopus database, relates to documents classified according to the subject area.
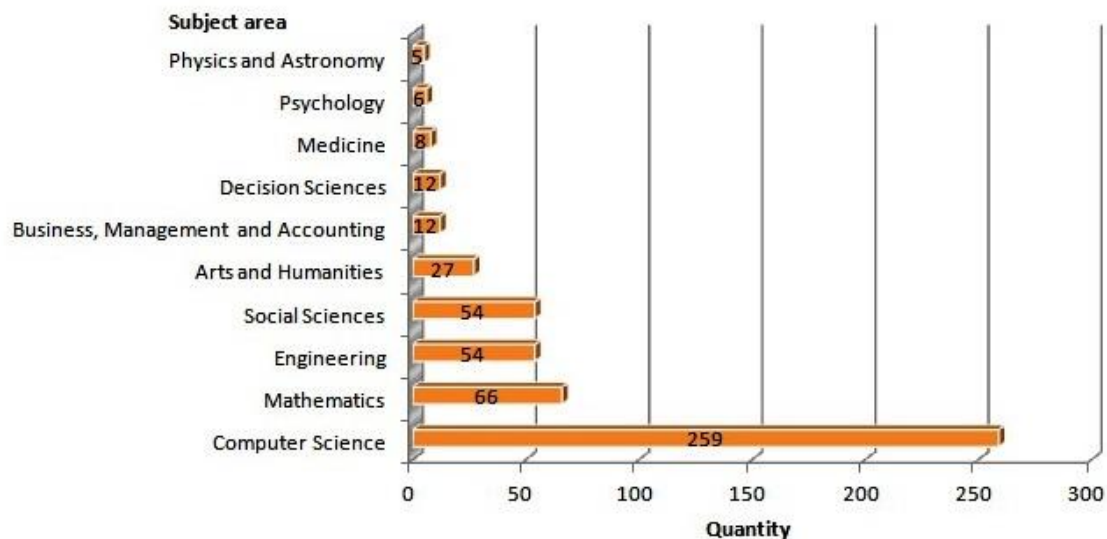

Figure 3. Publications by subject area

The absolute majority of publications are related to the area of Computer Science, as would be expected, given the affinity with the main theme. Accordingly, the universe of contexts to which the term chatbots is associated is broad and diversified, confirming its interdisciplinarity.

For documents of the type, the search results are shown in Figure 4. Conference papers, categorized as papers published in seminars and international conferences, aggregate the largest amount with 63.38% of publications, followed by articles, classified as papers published in journals and specialized websites, with 22.29%. The conference revisions occupy third place with 7.64%, book chapters in fourth with 1.91% and publications of other types 4.78%, in fifth place.
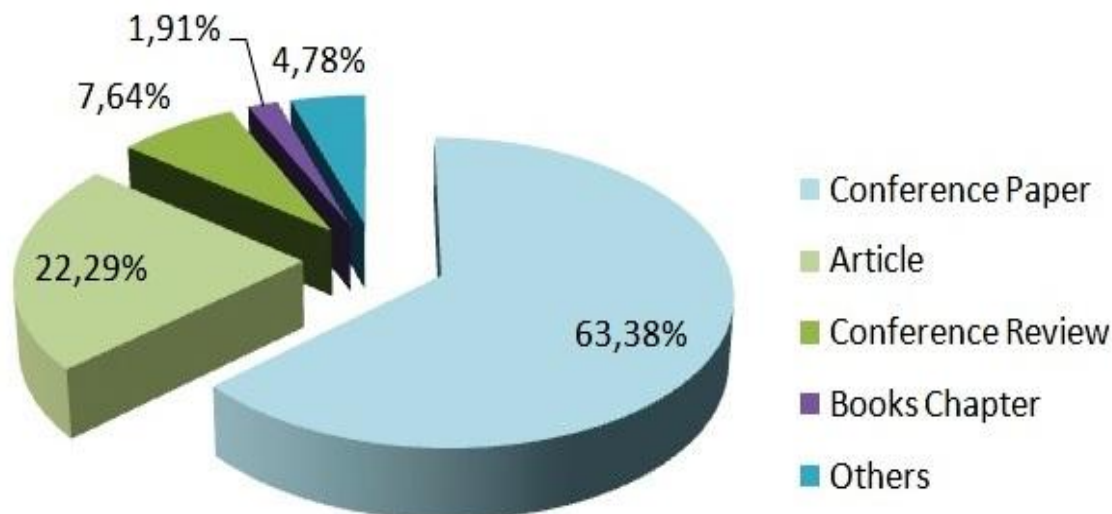

Figure 4. Publications by type

In terms of dates, the first published document is entitled "The catacomb project: building a user-centered portal the conversational way", under the responsibility of [Ginsburg 2002]. In the article, which received a quote, the author comments on the need to unify business applications such as database and documents in general, to promote centralized access to this information. As a solution, a conversation portal is proposed, the Catacumb project, directed to the end user, which extends the functionality of Chatbot ALICE. From the point of view of the author, this solution promotes scalability, extensibility and coordination between end users and developers.

In relation to the most cited articles, [Turney and Littman 2003], entitled "Measuring Praise and Criticism: Inference of Semantic Orientation from Association", appears first with 748 citations. In the document, the authors discuss an automated system with the objective of measuring the semantic orientation (positive or negative) of words of a text. According to the authors, this orientation is associated with the expressions used in the sentences, indicating a character of criticism or praise. This system would be applied to the classification of texts, the tracking of opinions in online dialogues, analysis of responses to surveys and in chatbots. In the case of chatbots, the suggested system could help in choosing the most appropriate response type for the moment.

3. Conclusion

This document demonstrates, based on the information contained in the Scopus database, a portion of the research involving the theme Chatbots.

An initial feature concerns the little published material (to the total were 314 documents) when compared to other subjects. This condition is credited to the fact that the first publication occurred in the year 2002, being considered in this way a recent theme. However, a significant increase in the number of publications has been detected since 2016, possibly driven by the increase in investments in the area of Artificial Intelligence by organizations, by the competitive and increasingly globalized market, as well as by the increase in performance of the hardware of the computers in general, thus allowing the execution of more robust applications. In this context, it is suggested to extend the research in other international databases, such as the IEEE or even the Capes Portal, afterwards making a comparison with the data collected in this research.

In relation to the content of the analyzed articles, one observes its adherence to academic environments and researches, usually oriented to educational contexts, in order to help students in their daily activities. However, recent products such as Amazon's Echo, Microsoft's Cortana, Apple's Siri or even Google Now, strengthen the advancement of research in the area aimed at the general public. These products, as indicated by their manufacturers, make intensive use of neural networks, considered an emerging concept in the field of Artificial Intelligence. Here is a potential subject to be explored in future works.

Considering the quantity and diversity of titles, we can note the significant presence of interdisciplinarity, ratifying that Artificial Intelligence contemplates several distinct areas, such as health, education, computing, linguistics, psychology, among others. Thus, the relevance of the theme is emphasized, with a vast horizon to be explored.

References

Artero, A. (2009). *Inteligência Artificial: teoria e prática*. 1st ed. São Paulo: Livraria da Física.

Coppin, B. (2015). *Inteligência Artificial*. 1st ed. Rio de Janeiro: Ltc.

Franco, C. (2014). *Inteligência Artificial*. 1st ed. Londrina: Educacional.

Freire, P. (2013). *Aumente a qualidade e a quantidade de suas publicações científicas: manual para elaboração de projetos e artigos científicos*. 1st ed. Curitiba: Crv.

Fryer, L. and Carpenter, R. (2006). Bots as language learning tools. *Language Learning and Technology*. **10** (3) 8-14.

Ginsburg, M. (2002). The catacomb project: building a user-centered portal the conversational way. *Fourth International Workshop on Web Information and Data Management*. 84-87, ACM Press.

Huang, J., Zhou, M. and Yang, D. (2007). Extracting Chatbot Knowledge from Online Discussion Forums. In: *20TH International Joint Conference on Artificial Intelligence, Iijcai 2007,* 423 - 428.

Lima, L. A. (2014). *Estudo de implementação de um robô de conversação em curso de língua estrangeira em ambiente virtual: um caso de estabilização do Sistema Adaptativo Complexo*: Doctoral Thesis. Belo Horizonte: Universidade Federal de Minas Gerais.

Russell, S. and Norvig, P. (2013). *Inteligência Artificial*. 3rd ed. Rio de Janeiro: Elsevier.

Turney, P. and Littman, M. (2003). Measuring praise and criticism: Inference of Semantic Orientation from Association. *Acm Transactions On Information Systems*. **21** (4) 315-346, Association for Computing Machinery (ACM).