AN OVERVIEW OF A STATE OF THE ART ON DEVELOPING SOFT COMPUTING-BASED LANGUAGE EDUCATION AND RESEARCH SYSTEMS: A SURVEY OF ENGINEERING ENGLISH STUDENTS IN UZBEKISTAN

Oybek Eshbayev *
Department of English language,
Tashkent State University of
Economics
eshyev1995@gmail.com

Abdulxalim Maxmudov
"Digital Pedagogy" section, Uzbek
State Scientific-Research Institute of
Pedagogical Sciences named after T.N.
Qori Niyazi
abdu-hm@yandex.ru

Ravshan Rozikov
Department of English language,
Tashkent State University of
Economics
ravshan.rr.79@mail.ru

ABSTRACT

This manuscript presents a state of –the –art overview on developing soft computing-based intelligent systems relevant to language education and research to inspire technology-savvy teacher-researchers to create their interactive tools. In this regard, two important fields (Natural Language Processing and Data Analysis) of computer science have been considered as the main directions of that linguistic computation because of their amazing software applications like automatic writing tutors, text summarizers, and so-called corpus-based apps. The paper also outlines some ethical considerations of system usage so that system designers should devise relevant design features in the future thereby teachers overcoming some classroom challenges. Unlike previous methods of studies on computer-assisted language teaching and research, our work devises a multidisciplinary method to reach the state-of-art. Through this method, we summarize the relevant literature source from two perspectives, linguistic and computational ones. The rationale of this method by the survey of engineering English students confirms that fuzzy logic-based Natural Language Processing and Data Analysis approaches can be employed effectively in language education and research. Finally, based on the multidisciplinary literature review method, the manuscript produces an overview of state-of-the-art performance as the outcome.

CCS CONCEPTS

• **NLP**; • **DA**; • **Stratification**; • **data-driven learning**; • **ESP**; • **State–of-the-art**; • **Natural language processing**; • **Data analysis**; • **Intelligent-systems**; • **Corpus-based-pedagogy**;

1 INTRODUCTION

Soft computing inspired by the human mind uses approximate calculations to find solutions to real-world problems. It has been frequently applied to a broad range of disciplines, accompanying various soft computing techniques (fuzzy logic, machine learning, and artificial neural networks, just to name a few) [14, 15]. From those disciplines, natural language processing and data analysis are of interest to the focus of this manuscript. With Natural Language Processing (NLP), it has become possible to make computers read text, hear speech, interpret it and identify important parts while data analysis enables us to manage extremely large volumes of data to extract insights that can inform the decision-making process [1–5, 8, 11–13]. By incorporating some soft computing techniques into them, we can handle even challenging real-life tasks. For example, fuzzy logic allows us to handle partial truth even though there is less information about both problem and solution, which is not likely to be done with traditional logic. Additionally, neural network systems as part of soft computing can learn useful presentations without human factor intervention. In language education and research contexts, intelligent systems consistent with NLP and DA enhance the quality of teaching and research [15]. Because students are motivated to learn the discourse of their discipline and develop their research writing competence when automated tutoring systems are employed. Fuel to ignite the inspiration of this article comes from the fact that nowadays, language education and research are becoming increasingly important [16, 17]. And to provide the state of art of creating intelligent systems conforming to NLP and DA can be the appropriate response to this situation. This is also beneficial for human resource development, which

is the main driving force of any economy since language-related consciousness comes first.

Inserting educationally sound and informative insights into the algorithms of those intelligent systems can meet the urgent needs of language learners. To explain the state of the art, the paper also uses the context of engineering English students in Uzbekistan whose learning habits are in the need of digital support. The method proposed in this paper also provides the outline of general guidelines on how to create the state of the art systems suitable for language education and research.

The main contributions of this manuscript are expected to be the followings:

Ethical practices of language education and research when using such automated systems.

Guidelines for system designers to consider when creating such automated systems conforming to soft computing techniques.

2 LITERATURE REVIEW AND THEORETICAL DEVELOPMENT

2.1 Linguistic perspective

The need for technology in education and research has gained increasing popularity as a method to develop students' research writing competencies and discipline-specific language skills. Motivated by this need, many educational researchers and system designers have developed automated tutoring systems based on language corpus datasets to assist the academic community. One typical example [1, 2] is the research-writing tutor that provides instantaneous feedback on students' research articles or drafts, developed by IOWA state university faculty (Elena Cotos et al.). This automated evaluation tool relies on a machine-learning algorithm to analyze the individual sections of inserted research articles across 30 different disciplines [2]. The list of such automated systems continues with criteria, my access, writing power, and the like. However, the main drawback of this system is the lack of a teacher interface to alter progress reports according to individual specifications or needs and technical functions to enable teacher-made comments. Further research is needed for a new feature design based on classroom observations and teachers' input as shown in [Cotos 2016].

[17] investigated the academic writing challenges faced by postgraduate engineering English students and identified that those challenges are primarily at sentence levels but most of the students are concerned with the discourse level challenges. Some strategies do not satisfy the program's expectations. The study of [17] has significant pedagogical implications including the need of providing proper writing models and writing interventions for those students.

Meanwhile, it is important to note that NLP systems are seemed to mirror three basic approaches of corpus-based pedagogy: top-down, bottom-up, and mixed approaches [7]. To be clear, register analysis refers to a bottom-up analysis of lexico-grammatical features of target discourse while the top-down approach deals with the analysis of text genres. As for the mixed one, the best links between these two worlds of language analysis are considered for designing.

The recent pedagogical concept "data-driven learning" has reached increasing popularity among language educators and researchers. Based on language corpora as the core of this concept, they conducted computer-aided researches and have come to pedagogical decisions about the ethical practices of using intelligent corpus-based systems. [8] suggests that automated writing systems are a good motivation for students to engage in the continuous process of revising and editing. Teachers should also check randomly student-writing samples and give suggestions for them so that they can clarify incorrect machine messages. Additionally, to increase the student confidence in the intelligent systems, a positive attitude towards the programs and patient demonstration of system functions for students are recommended for teachers. Finally, teachers should also encourage students to put quality first over quantity or machine scoring and not blindly trust machine scoring since the social nature of writing has to be preserved.

Another main contribution to the ethical practices of using such intelligent systems was made by [9]. She built her argument for the integrated model of lexical and corpus-based approaches. With the example of engineering English, she presented the finding of her computer-assisted research that both sub-technical and academic English should be paid more attention by English Specific Purposes (ESP) teachers whose job is to teach discipline-specific English for the students of that discipline. She also illustrated amazing classroom activities for enhancing students' lexical acquisition.

The next important contribution [10] is a three-part methodology for choosing specialized vocabulary that is more suitable to teach data-driven learning. Domain experts can rely on this approach to choose words more confidently and make connections between specialized vocabulary and their functions. This method also supports the system designers to reanalyze their language data sets based on domain expert judgments.

2.2 Computational perspective

One of the main accomplishments to support language education and research has been the development of soft computing-based data summarization systems that are one of the most challenging tasks of NLP [3, 4]. These systems can reduce effectively the workload of the tedious data-to-knowledge translation process, thereby avoiding irrelevance and ambiguity of reading the document as well as mental torture of comprehending or summarizing the document and many others. One novel direction in this kind of summarization tool is made by [Smits et al. 2018], which is a soft computing-based approach and a visualization of linguistic summary of data (so-called term cloud). By employing soft computing techniques (especially fuzzy logic), a term cloud that includes linguistic terms can be formed to represent effectively the datasets. Most importantly, to achieve a more easy and speedy analysis of data, direct summarization of a dataset with conjunctions of terms has been recommended [3].

Additionally, [Tayal et al. 2017] developed an automatic text summarizer by using their title identification and semantic representation mechanisms. This was mainly done by employing Subject Verb Object rules and tag-based training that were the core parts of their soft computing approach [4]. The authors of this system concluded that in the future, to clean ill-structured sentences (e.g., punctuation and spelling errors) or the whole text from different sources, improvement of the system's robustness is in the need of

AN OVERVIEW OF A STATE OF THE ART ON DEVELOPING SOFT COMPUTING-BASED
LANGUAGE EDUCATION AND RESEARCH SYSTEMS: A SURVEY OF ENGINEERING ENGLISH
STUDENTS IN UZBEKISTAN

ICFNDS 2021, December 15, 16, 2021, Dubai, United Arab Emirates

further research. They also stressed that summarization is not only the job of a common summarization tool but also the matter of specialization.

[5] proposed a framework to tackle part of speech tagging and sentiment classification tasks, which are the core two techniques of the natural language processing domain. In this general framework, neural network and (probabilistic) soft logic have been combined to improve the performance of the NLP model. This framework helped to solve the uninterpretability problem of classifying texts and sentiment analysis.

The letter of [6] is important for this study since it maintains one recent general concept (stratification)for natural language processing and big data analysis. This letter highlighted the main prospects of these two domains: NLP and Big Data. The application of stratification in NLP and DA can result in user access to more predictive choices of language and efficient data management. Stratification uses distinctive features so-called incremental enlargement, which is not included in soft computing techniques.

In the direction of data exploration and knowledge discovery, a novel data mining algorithm has been recommended by [12]. This algorithm based on fuzzy functional dependencies and linguistic interpretations supports domain experts (linguists included) to make decisions during data analysis. This allows us to express useful knowledge from big data by linguistic terms, which is made possible by fuzzy modifiers and quantifiers.

Since the vast majority of research in soft computing uses DA and NLP methods, we related the important applications of linguistics (such as corpus-based approaches) to soft logic related research focus of our DA & NLP literature review, which results in reducing massive datasets and extracting core information about the state-of-the-art systems for language education and research. The next section of this manuscript is structured in the following way. First, we begin by giving a brief overview of key articles from a computing perspective and note that those have been rarely applied to existing intelligent systems of language and research. Next, we summarize the important applications of linguistics concerning the computing field. From this background, we conduct a quantitative analysis (surveying) of the applications of soft logic-based intelligent systems to language education and research to validate the effectiveness of the proposed state-of-the-art performance.

3 METHODOLOGY

The systematic approach used in this study to summarize the state of the art was intended to be used by automated system designers and tech-savvy English for Specific Purposes instructors to align methods and algorithms with the needs of the classroom and target discourse. To set ethical practices of using such systems, and guidelines for system designers to design wisely, this study decided to focus on a systematic literature review based on NLP and DA domains since a majority of existing automated programs are built upon those programming logics. A pool of more than 10 key articles has been established and analyzed from two perspectives (linguistics and computing). Domain experts were also involved in this literature review process so that the importance of methods and algorithms can be evaluated carefully to respond current needs of the students and changing patterns of the discourse community.

As the result, with the framework of [5], two main tasks of natural language processing, part of speech tagging and sentiment analysis based on soft logic are considered to be more elaborate to improve the performance of NLP systems. The general concept of stratification in [6] allows users of NLP systems more predictive choices of language. Soft computing-based text summarization methods [3, 4] also support the idea of creating more elaborate summarization systems for language research and education. A novel data-mining algorithm that has been recommended by [12] is selected for the decision-making process of data analysis.

From the point of linguistic perspective, three basic approaches of corpus linguistics (bottom-up, top-down, and mixed) are considered quality keepers for the process of automated system designs. This is also the fact that all of the domains have their language discourse communities that communicate with each other within certain specialized transfers and genres. The three-part methodology proposed by [10] and one of [9] are included as modern approaches for selecting vocabulary into language datasets.

To verify the importance of this state-of-the-art design approach, the context of engineering English students and researchers in Uzbekistan was surveyed concerning testing the pilot automated systems that incorporated target methods mentioned above. Within 3 months, the students were surveyed with questionnaires online and interviewed about the effectiveness of methods and digital tools. Existing research writing tutor by the IOWA State University team was also proposed since it mirrors three approaches of corpus-based linguistics and 70 % of participants of this survey (total number = 100 people) showed positive feedback about this software. Below (see Figure 1) is the text summarizer system designed based on the soft logic approach that was intentionally created using python programming language thereby showing the effectiveness of selected methods.

Chemical engineering students and researchers checked the performance of this software inserting domain-specific texts and gave more positive comments (80 % of total participants). The reduction rate of this innovative software was equal to 56% as can be seen below (Figure 2). Red and green colors indicate the locations of the main ideas in different parts of the text.

From what has been outlined above, we can see that there are enough effective soft logic-based methods that computer scientists have created so far for DA and NLP. They can also enhance the quality of intelligent systems used for language education and research if used appropriately. However, the main missing element in such intelligent systems is the realization of linguistic practices (classroom activities, pedagogical methods, and teacher intervention). From an above-mentioned background of linguistics, the system designer can safely create needs-responsive features of their systems to help teachers handle some challenging tasks of language education and research. So, that is all about the state-of-the-art performance.

4 RESULTS

The initial literature review produced a state-of-the-art system design that was significantly useful for the practitioners by summarizing the pool of more than 10 key articles from two different domains linguistics and soft computing. Students and researchers
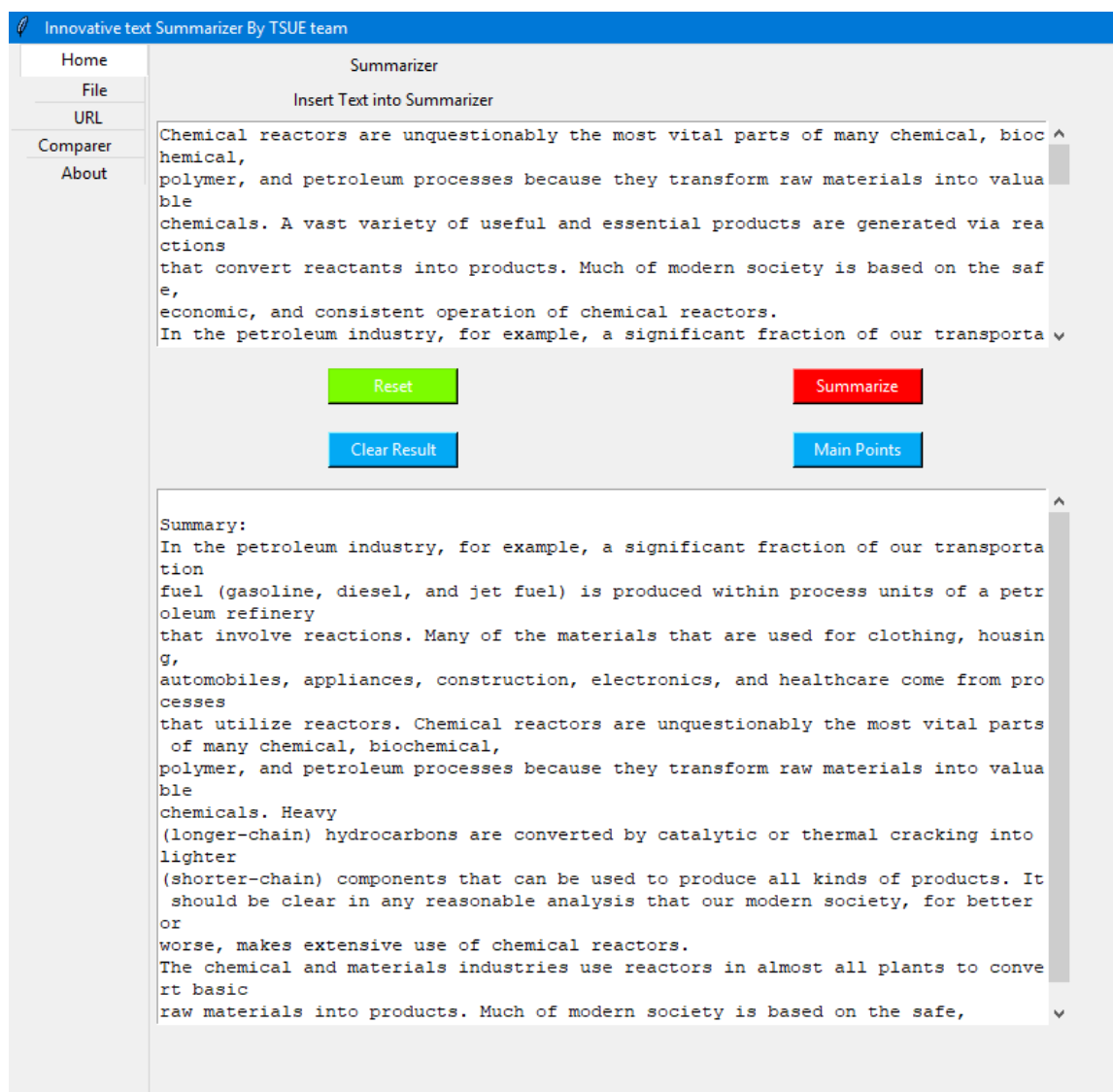
Figure 1: Innovative text summarizer including soft logic.

could investigate the samples of state-of-the-art systems in practice and left mostly positive comments on the survey (70% and 80%).

The following ethical practices of language education are identified:

● Conforming to the standards of corpus-based pedagogy to build language datasets and interfaces.

● Integration of lexical approach with corpus linguistics to enrich students working lexicon

● Take the consideration of students corpora into the account of a system design plan.

● Building term clouds to provide concise representations of textual data based on a soft computing approach.

● Balanced attention for academic and sub-technical vocabulary in English specific purposes research and system design.

Below is also the outline of resulting guidelines for system designers:

● Using soft logic-based approaches for POS tagging and sentiment analysis to improve the performance of NLP systems.

● Using stratification techniques in NLP and DA as a more advanced way of creating automated systems.

● Employing soft logic techniques into text summarization software design and automated tutoring systems.

● Creating automatic scoring systems for a spontaneous speech by using NLP techniques (Part of Speech tagging, vector space model, language model).

5 DISCUSSION AND CONCLUSION

The government's goal in issuing sovereign international bonds on the international market is not only to provide the economy with

AN OVERVIEW OF A STATE OF THE ART ON DEVELOPING SOFT COMPUTING-BASED
LANGUAGE EDUCATION AND RESEARCH SYSTEMS: A SURVEY OF ENGINEERING ENGLISH
STUDENTS IN UZBEKISTAN

ICFNDS 2021, December 15, 16, 2021, Dubai, United Arab Emirates

Original Text

Chemical reactors are unquestionably the most vital parts of many chemical, biochemical, polymer, and petroleum processes because they transform raw materials into valuable chemicals. A vast variety of useful and essential products are generated via reactions that convert reactants into products. Much of modern society is based on the safe, economic, and consistent operation of chemical reactors.

In the petroleum industry, for example, a significant fraction of our transportation fuel (gasoline, diesel, and jet fuel) is produced within process units of a petroleum refinery that involve reactions. Reforming reactions are used to convert cyclical saturated naphthenes into aromatics, which have higher octane numbers. Light C4 hydrocarbons are alkylated to form high-octane C8 material for blending into gasoline. Heavy (longer-chain) hydrocarbons are converted by catalytic or thermal cracking into lighter (shorter-chain) components that can be used to produce all kinds of products. The unsaturated olefins that are used in many polymerization processes (ethylene and propylene) are generated in these reactors. The polluting sulfur components in many petroleum products are removed by reacting them with hydrogen.

The chemical and materials industries use reactors in almost all plants to convert basic raw materials into products. Many of the materials that are used for clothing, housing, automobiles, appliances, construction, electronics, and healthcare come from processes that utilize reactors. Reactors are important even in the food and beverage industries, where farm products are processed. The production of ammonia fertilizer to grow our food uses chemical reactors that consume hydrogen and nitrogen. The pesticides and herbicides we use on crop fields and orchards aid in the advances of modern agriculture. Some of the drugs that form the basis of modern medicine are produced by fermentation reactors. It should be clear in any reasonable analysis that our modern society, for better or worse, makes extensive use of chemical reactors.

Text Summary

In the petroleum industry, for example, a significant fraction of our transportation fuel (gasoline, diesel, and jet fuel) is produced within process units of a petroleum refinery that involve reactions. Many of the materials that are used for clothing, housing, automobiles, appliances, construction, electronics, and healthcare come from processes that utilize reactors. Chemical reactors are unquestionably the most vital parts of many chemical, biochemical, polymer, and petroleum processes because they transform raw materials into valuable chemicals. Heavy (longer-chain) hydrocarbons are converted by catalytic or thermal cracking into lighter (shorter-chain) components that can be used to produce all kinds of products. It should be clear in any reasonable analysis that our modern society, for better or worse, makes extensive use of chemical reactors. The chemical and materials industries use reactors in almost all plants to convert basic raw materials into products. Much of modern society is based on the safe, economic, and consistent operation of chemical reactors.

Figure 2: One sample of summarizing chemical engineering text using the pilot software. Here, 302 worded text (on the left) 156 worded summary text (on the right)

investment funds, but also to place sovereign international bonds in international financial markets. It was concluded that banks will be able to place their bonds in the global financial markets and on this basis to create a basis for independent financing of investment projects.

It is inevitable for system designers and technology-savvy teacher-researchers to encounter problems that require the selection of appropriate methods and modern techniques to be solved while creating needs-responsive intelligent systems.

In this regard, the state of the art of designing such automated systems provides several advantages over natural language processing and data analysis domains from linguistic and computational perspectives. From computational points of view, for the natural language process, the problem of interpretability on part of speech tagging and sentiment analysis is well solved by employing the framework based on soft logic and neural network to improve the performance of NLP systems. Employing the recent concept of stratification on NLP gives users more predictive options to determine the required level of accuracy in their written or spoken performance. In addition, one of the most challenging tasks of NLP is advanced-level automated text summarization, which can be effectively solved by using a soft computing approach consisting of SVO rules and tag-based training.

Regarding the data analysis process on automated system design, a novel descriptive algorithm based on fuzzy functional dependencies and linguistic interpretations supports the data mining process to discover useful knowledge. The paper [3] of soft computing-based big data summarization proposes term clouds to represent data from linguistic summaries instead of using the visualization of statistical measures. The incremental enlargement from the stratification process holds the good potential to manage large-sized data

sets. Miguel Badaracco also proposed a remarkable fuzzy linguistic algorithm for Intelligent Tutoring System based on competencies [15].

From the perspective of linguistics, it is wise that the system designers would devise novel features to overcome some major classroom challenges like teacher commenting mechanisms, more reliable scoring systems. Additionally, adhering to the standards of corpus-based pedagogy such as transfer analysis and genre analysis is safer for the system designer and digitally literate teachers. Olga Mudraya's integrated model of the lexical and corpus-based approach and significant pedagogical implications of [17] should be taken into consideration when constructing datasets and user interfaces since it recommends a novel pedagogical implication of using academic and sub-technical vocabulary, learning activities, writing interventions.

Authentic application of the state-of-the-art design method of intelligent systems based on NLP and DA is experienced in the real-life engineering English context where the formation of datasets and performing the important functionalities of the pilot automated systems have been analyzed. The findings show the usefulness of the manuscript method.

Finally, it should be stressed that it is the qualifications and mastery of system designers and tech-savvy educators that determine the best ways to create needs-responsive technologies well suited to the classroom conditions while different methods and algorithms still try to approach the faster and convenient ways of designing automated systems from different perspectives (computing and linguistics). That is why future researches should be conducted to update constantly the proposed state-of-the-art design method.

ACKNOWLEDGMENTS

REFERENCES

[1] E. Cotos. 2016. Computer-assisted research writing in the disciplines. IOWA State University Digital repository, English publications. Adaptive educational technologies for literacy instruction, 225 – 242. https://lib.dr.iastate.edu/engl_pubs/238/

[2] Ramaswamy Nandhini. 2012. Online tutor for research writing. Graduate Theses and Dissertations. 12743. https://lib.dr.iastate.edu/etd/12743

[3] Grégory Smits , OlivierPivert , Ronald R.Yager , and PierreNerzic. 2018. A soft computing approach to big data summarization. Fuzzy Sets and Systems, 348 (2018) 4–20. https://doi.org/10.1016/j.fss.2018.02.017

[4] Madhuri A.Tayal, Mukesh M. Raghuwanshi, and Latesh G. Malik .2017. ATSSC: Development of an approach based on soft computing for text summarization. Computer Speech and Language, 41 (January 2017) 214–235. http://dx.doi.org/10.1016/j.csl.2016.07.002

[5] Mourad Gridach. 2020. A framework based on (probabilistic) soft logic and neural network for NLP. Applied Soft Computing, Volume 93, (August 2020), 106232. https://doi.org/10.1016/j.asoc.2020.106232

[6] Mehdi Rajabi Asadabadi , Morteza Saberi, and Elizabeth Chang .2018. Letter: The concept of stratification and future applications. Applied Soft Computing, 66 (2018) 292–296. https://doi.org/10.1016/j.asoc.2018.02.035

[7] Elena Cotos.2017. Language for Specific Purposes and Corpus-based Pedagogy. The Handbook of Technology and Second Language Teaching and Learning, (June 2017) 248-264.https://doi.org/10.1002/9781118914069.ch17

[8] Pei-ling Wang.2013.Can Automated Writing Evaluation Programs Help Students Improve Their English Writing?. International Journal of Applied Linguistics & English Literature, Vol. 2 No. 1 (January 2013) 6-12. http://dx.doi.org/10.7575/ijalel.v.2n.1p.6

[9] Olga Mudraya.2006.Engineering English: A lexical frequency instructional model. English for Specific Purposes, Volume 25, Issue 2,(2006) 235-256. https://doi.org/10.1016/j.esp.2005.05.002

[10] Philippa Otto. 2021. Choosing specialized vocabulary to teach with data-driven learning: An example from civil engineering. English for Specific Purposes,61(5)(January 2021)32-46. DOI:10.1016/j.esp.2020.08.003

[11] Suma Bhat and Su-Youn Yoon.2015. Automatic assessment of syntactic complexity for spontaneous speech scoring. Speech Communication 67 (2015) 42–57. http://dx.doi.org/10.1016/j.specom.2014.09.005

[12] Miljan Vučetić , Miroslav Hudec and Boško Božilović .2020. Fuzzy functional dependencies and linguistic interpretations are employed in knowledge discovery tasks from relational databases. Engineering Applications of Artificial Intelligence 88 (2020) 103395. https://doi.org/10.1016/j.engappai.2019.103395

[13] Dr. D.K. Sreekantha.2019.Big Data Analytics using Soft Computing Techniques: A Study Communication and Computing Systems, research gate,(July 2019) DOI:10.1201/9780429444272-67

[14] Alain Jérôme, Fougères and EgonOstrosi. 2021.Fuzzy engineering design semantics elaboration and application. Soft Computing Letters, (November 2021) 100025. https://doi.org/10.1016/j.socl.2021.100025

[15] Miguel Badaracco and Luis Martínez. 2013. A fuzzy linguistic algorithm for adaptive test in Intelligent Tutoring System based on competences. Expert Systems with Applications, Volume 40, Issue 8 (June 2013) 3073-3086. https://doi.org/10.1016/j.eswa.2012.12.023

[16] Ken Hyland and Feng (Kevin) Jiang .2021. Delivering relevance: The emergence of ESP as a discipline. English for Specific Purposes 64 (2021) 13–25. https://doi.org/10.1016/j.esp.2021.06.002

[17] Linda H.F. Lin and Bruce Morrison.2021. Challenges in academic writing: Perspectives of Engineering faculty and L2 postgraduate research students. English for Specific Purposes, Volume 63,(July 2021) 59-70 https://doi.org/10.1016/j.esp.2021.03.004