



## Text Analyzing Algorithm for Speech Synthesizer of Uzbek Language

**Bakhtiyor Akmuradov<sup>1</sup>, Utkir Khamdamov<sup>2</sup> and Djамshid Sultanov<sup>3</sup>**

<sup>1</sup>Lecturer, Department of Hardware and Software Provision of Management Systems in Telecommunications Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan.

<sup>2</sup>Professor, Department of Hardware and Software Provision of Management Systems in Telecommunications Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Tashkent, Uzbekistan.

<sup>3</sup>Associate Professor, Department of Hardware and Software Provision of Management Systems in Telecommunications, Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Tashkent, Uzbekistan.

(Corresponding author: Djамshid Sultanov)

(Received 22 September 2020, Revised 28 December 2020, Accepted 27 January 2021)

(Published by Research Trend, Website: [www.researchtrend.net](http://www.researchtrend.net))

**ABSTRACT:** Simplifying the use of modern systems of storage, transmission and processing of information has become one of the most pressing issues today. In this regard, the production of technologies that convert text control into speech signals is one of the important directions. This article describes an algorithm for analyzing incoming text in synthesis systems that convert textual information in Uzbek into speech signals. Electronic text editing tools and ways to express Uzbek text electronically are also considered.

**Keywords:** Electronic text, speech, speech synthesis, text analysis, algorithm.

### I. INTRODUCTION

All mechanical and electronic devices developed to date are aimed at simplifying and facilitating lifestyles by automating the performance of any task in human life and increasing work efficiency. The results of this effort can be seen in the achievements of modern electronic technology in the 21st century, from the primitive weapons that humanity developed to defend itself in the age of creation.

It is no longer possible to imagine our lives as modern advanced and popular electronic systems, personal computers, mobile services, television, radio and other electronic storage and transmission techniques. Such technologies are able to easily perform the functions associated with the processing of large and small amounts of information, from the simple tasks of our daily lives to the most complex tasks.

Improving the management and control of such modern systems, simplifying the interface for the exchange of information between man and the system has become one of the most pressing issues today.

Typically, modern systems use special input-output devices. Life can be made easier by organizing them with the help of sound signals, making it easier for people with disabilities to use modern technology, and expressing simple and complex information in the form of speech signals.

To this end, a number of measures are being taken to recognize textual information and convert it into speech signals or, conversely, to convert speech signals into textual information. As a result of the application of the results of the study, speech synthesis systems have been developed and are being implemented in many languages.

In particular, the number and quality of systems that synchronize incoming signal electronic text information is

increasing day by day. How to determine the results of the production of such systems is to link them closely with the literary norms of that language.

An electronic text is a text produced by any electronic medium that combines the features of oral and written speech. Much research has also been done on the study and analysis of the properties of electronic texts. Including Yu. V. Balakina's "Electronic text: a fundamentally new type of text?" The research paper provides a detailed description of the characteristics of electronic text, intertextual relations, multimedia capabilities, completeness, format, status, functions, language features, state of creation and other features [1].

R.V. Mesheryakov's research work "The structure of the system of synthesis and speech recognition" studied the algorithmic problems in the optimal use of computer system resources and the creation of information processing programs in the speech synthesis of various systems [2].

Unlike computers, texts are very easy for people to read and understand. On the other hand, if text analysis needs to be analyzed using software, it is a complex task that requires a combination of computer linguistics and statistical methods. In this case, the use of modern systems of intellectual analysis of texts is required. A.M. Situlsky and A.B. Ivannikov's research on "Intellectual Analysis of Text" discusses intellectual analysis in detail [3].

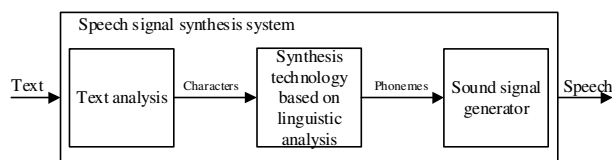
In general, the algorithms of any speech synthesis system depend directly on the form and properties of the incoming information. When the incoming information is in the form of text, the system's methods of analysis and synthesis and linguistic algorithms need to be developed based on the characteristics of the text. In particular, the development of a system that converts electronic text into Uzbek into speech signals will solve many of the

problems of all those who communicate in our language in keeping records and using modern technology.

## II. A SYSTEM FOR SYNTHESIZING SPEECH SIGNALS FROM TEXT

Speech synthesis is the process of organizing the artificial development of human speech. A computer system used for this purpose is called a speech synthesizer and can be implemented in software or hardware products. A system for synthesizing speech signals from text data converts the normal text of a language into speech signals.

The quality of a speech synthesizer is measured by its resemblance to a human voice and its ability to understand clearly. The technology of synthesizing speech signals from textual information can help people with visual impairments or reading difficulties in daily life and learning. It also simplifies the use of modern information and communication technologies. To date, many speech synthesizers have been developed, which are functionally composed of three main parts. (Fig.1).



**Fig. 1.** A general structure of speech signal synthesis system.

The first part converts the text of the system into written electronic equivalents of words, containing various elements of text, words, symbols, numbers, abbreviations, and so on. This process is often referred to as text normalization, processing, or lexical analysis. Also, based on the analysis of the text, a plan for the organization of the next work process of the system is developed. The most important process is to identify all the elements in the text and make them ready for synthesis. This process may require reference to almost all grammatical and spelling rules of a given language. For this reason, in most speech synthesizers, most of the synthesis time is spent on text analysis.

The next section deals with the process of assigning phonetic transcriptions for each word and dividing the text into prosodic units, such as sentences, sentences, and paragraphs. The process of attaching phonetic symbols to words is called conversion from text to phoneme or from grapheme to phoneme. Phonetic transcription and the interconnection of information form the symbolic lexical meaning described in the preceding section.

In the last section, sound signals are generated based on phonemes sorted according to the text elements. Generated sound signals are output in the form of speech signals through selected play elements for the system [4].

In general, speech signals synthesized by such systems differ from natural speech in a number of features. Today, using the advances of science and technology, research is being done to bridge the gap between synthesized artificial speech and natural speech.

The process of synthesizing speech signals from text is usually referred to as the electronic form of the text. PCs, mobile devices, and other electronic devices can now be used to express text electronically.

## III. CODING SYSTEMS OF ELECTRONIC TEXT

With the development of computers, coding became necessary to create an electronic form of all the information that humans handled.

Binary coding is one of the most common ways of presenting information. In computers, robots, and digitally controlled machines, all information related to electronic devices is encoded in binary. This situation is explained by the fact that electronic devices operate in the binary number system.

From a computer perspective, text is made up of individual characters. Characters include not only uppercase and lowercase letters, but also numbers, punctuation, special characters such as "=", "(", "&", etc., and even spaces between words.

Texts are entered into computer memory using the keyboard. Buttons are written with familiar letters, numbers, punctuation and other symbols. They access random access memory via binary code. This means that each character is represented by an 8-bit binary.

A table in which all the characters of the computer alphabet are assigned serial numbers is called a coding table. Different coding tables are used for different computers.

ASCII (American Standard Code for Information Interchange) is a table of American standard codes for data exchange. The ASCII code table consists of 3 main parts, which can be seen in Table 1 below.

**Table 1: The structure of the ASCII coding table.**

Serial number in decimal notation	Binary code	Feature
0 - 31	00000000 - 00011111	These are called control signs, and their function is to display or print text, give an audible signal, assign text, control the process, and so on.
32 - 127	00100000 - 01111111	The standard part of the table includes lowercase and uppercase letters of the Latin alphabet, decimal numbers, punctuation marks, all types of parentheses, commercial and other symbols.
128 - 255	10000000 - 11111111	As an alternative part of the table, the second half of an ASCII code table, called a code page, may have different options, each with its own number.

There are different options for the second alternative part of this coding system. Therefore, the ASCII code usually refers to the main part, which is the first 0 to 127.

The number of characters used as text elements is not limited to 255. This can be explained by the large number of types and numbers of characters in the existing languages of the world. The ASCII standard does not have the capacity to express them in electronic text. The Unicode standard is currently used to address this shortcoming.

Unicode is a universal encoding standard used to represent characters that are not part of ASCII, and includes characters that are used in writing in almost every language in the world. Currently, this standard is used as the main standard on the Internet.

The standard was proposed in 1991 by the non-profit organization Unicode Consortium. The application of this standard allows to encode a large number of characters from different writing systems: in documents encoded in accordance with the Unicode standard, Chinese characters, mathematical symbols, letters of the Greek alphabet, Latin and Cyrillic letters, musical notes and other symbols can be encoded.

The standard consists of two main parts: the Universal character set (UCS) and the Unicode transformation format (UTF). The universal character set includes characters that correspond to Unicode, and each character is represented as a non-negative integer, usually written in hexadecimal with the prefix U+.

#### IV. TEXT STRUCTURE

When it comes to text files, we need to consider the structure of files with the extensions txt, doc, rtf, docx. Doc or docx text files are commonly used in office and other document work. Depending on the nature of the text elements in them, their structure can be expressed as follows.

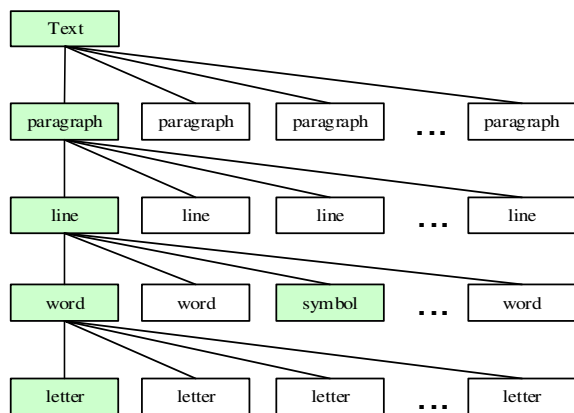


Fig. 2. Text structure.

As you can see in the picture above, text files are an organizational structure made up of hierarchical links of elements from simple characters to pages.

The smallest element of this structure is the letters, one or more of which combine to form a word with an independent meaning.

Structurally, a line is made up of one or more word and character combinations. A set of one or more lines forms a paragraph. One or more paragraphs are referred to as text. Such files are called text files because the main component of the file under study is text.

In general, text files are electronic structures with a structure of one or more pages. Note that a text file may

contain a single page and may not contain any text elements. This does not technically change the properties of the text file.

#### V. TEXT ANALYSIS

As a result of the development and popularity of modern information and communication technologies, the use of electronic forms of written speech is growing. At present, electronic document management systems have been introduced in almost all areas. As a result of the proliferation of personal computers, not only documents of governmental and non-governmental organizations, but also personal drafts of any person are stored electronically. In such electronic texts, information is represented on the basis of a sequence of different characters.

In general, the following three groups of symbols are most commonly used in such electronic text:

1. Letters
2. Symbols
3. Numbers

Each of the symbols in this group has its own characteristics, some represent a sound, some a value, some a condition according to a standard, and so on. Taking into account all the signs and circumstances that can be found in the majority of texts in the Uzbek language, its structure can be summarized as follows.

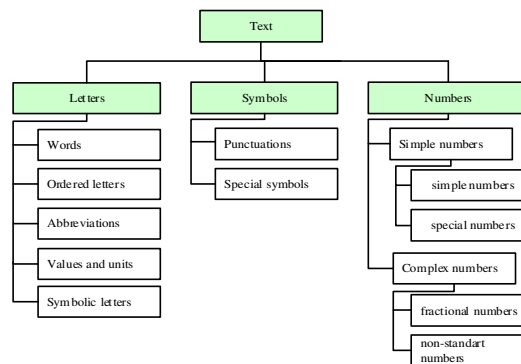


Fig. 3. Hierarchy of text elements.

As you can see from the picture, text is usually a sequence of characters that express a meaning. The bulk of such characters are letters.

#### VI. TEXT ELEMENTS REPRESENTED BY LETTERS IN UZBEK LANGUAGE

A letter is a symbol or set of characters that represents a sound. In the text, such elements can be found in different elements, arranged in a sequence. To them

1. Words
2. Ordered letters
3. Abbreviations
4. Values and units
5. Symbolic letters can be an example.

The literary features of the target language and the structure of the alphabet used play an important role in the expression of words. On September 2, 1993, the Law of the Republic of Uzbekistan "On the introduction of the Uzbek alphabet based on the Latin script" was promulgated in our country, and Latin graphics began to be used in practice [5]. Since the amendment of this law on May 6, 1995, the current Latin alphabet has been

officially in force. The following table shows the current Uzbek alphabet based on the Latin script [6]. As can be seen from the table, 24 of the 29 letters are represented by a single character. Of these, 2 are a combination of letters and characters and 3 are a

combination of letters. From this it can be concluded that 29 sounds can be expressed in writing in Uzbek.

The corresponding codes in Unicode for the electronic representation of these characters can be seen in Table 3 below.

**Table 2: Latin alphabet of the Uzbek language.**

Types \ №	1	2	3	4	5	6	7	8	9	10	11	12
Letters	Aa	Bb	Dd	Ee	Ff	Gg	Hh	li	Jj	Kk	Ll	Mm
	Nn	Oo	Pp	Qq	Rr	Ss	Tt	Uu	Vv	Xx	Yy	Zz
Letter + Symbol	O'o'	G'g'										
Letter + Letter	Shsh	Chch	ng									
Symbol	'											

The words in the table are based on a certain sequence of uppercase and lowercase letters. Words make up sentences and sentences make up paragraphs.

In some cases, the letter symbols can be found in the form of a sequence that has no meaning. Examples of such cases are the alphabetical order of the characters, only the ordered sequence of vowels, the ordered sequence of consonants, or the order of consonant and non-consonant sounds, and the sequence of letters selected by some other algorithm [7]. In such cases, the letters are usually pronounced differently than in the case of ordinary words. That is, the alphabet is based on pronunciation. The following table shows the alphabetical pronunciation of letters.

Abbreviations are another of the things that are represented by a sequence of letters in which the meaning is defined in the text. Abbreviations consist of an extended sequence to ensure that they can be used or have a large number of digits. Most conditional abbreviations are written in capital letters.

- AKT - Axborot KommunikatsiyaTexnologiyalari (Information Communication Technologies);
- AQSH - Amerika Qo'shmaShtatlari (United States of America);
- BMT - BirlashganMillatlarTashkiloti (United Nations);
- DXX - DavlatXavfsizlikXizmati (State Security Service);
- GDR - GermaniyaDemokratik Respublikasi (German Democratic Republic).

There are many such examples. It should be noted that these abbreviations are formed by shortening the sequence of words in the Uzbek language. However, there are some types of abbreviations that have been borrowed from other languages and their extended form may be lost in the Uzbek language.

- ISO - XalqaroStandartlashtirishTashkiloti (International Standard Organization);
- NATO - ShimoliyAtlantikaShartnomasiTashkiloti (North Atlantic Treaty Organization);
- NBU - O'zbekistonMilliyBanki (National Bank of Uzbekistan);
- TCP - UzatishniBosqarishProtokoli (Transmission Control Protocol);
- UDP - FoydalanuvchiDatagrammasiProtokoli (User Datagram Protocol).

**Table 3: Unicode representation of letters of the Uzbek language in Latin script.**

NO.	Uppercase letters	Punctuation code(Unicode-dec)	№	Lowercase letters	Punctuation code(Unicode-dec)
1	A	65	1	a	97
2	B	66	2	b	98
3	D	68	3	d	100
4	E	69	4	e	101
5	F	70	5	f	102
6	G	71	6	g	103
7	H	72	7	h	104
8	I	73	8	i	105
9	J	74	9	j	106
10	K	75	10	k	107
11	L	76	11	l	108
12	M	77	12	m	109
13	N	78	13	n	110
14	O	79	14	o	111
15	P	80	15	p	112
16	Q	81	16	q	113
17	R	82	17	r	114
18	S	83	18	s	115
19	T	84	19	t	116
20	U	85	20	u	117
21	V	86	21	v	118
22	X	88	22	x	120
23	Y	89	23	y	121
24	Z	90	24	z	122
25	O'	79+699	25	o'	111+699
26	G'	71+699	26	g'	103+699
27	Sh	83+104	27	sh	115+104
28	Ch	67+104	28	ch	99+104
			29	ng	110+103

The most widely used abbreviations can be used to solve such inaccuracies in text analysis or speech signal synthesis systems. However, this is not a complete solution to the problem, as it is not possible to summarize abbreviations in all languages and in all fields.

**Table 4: Pronunciation of letters of the Latin alphabet in Uzbek language.**

№	1	2	3	4	5	6	7	8	9	10	11	12
Types												
Letters	Aa	Bb	Dd	Ee	Ff	Gg	Hh	li	Jj	Kk	Ll	Mm
Pronunciation	a	be	de	e	ef	ge	he	i	je	ke	el	em
Letters	Nn	Oo	Pp	Qq	Rr	Ss	Tt	Uu	Vv	Xx	Yy	Zz
Pronunciation	en	o	pe	qe	er	es	te	u	ve	xe	ye	ze
Letters	O'o'	G'g'	Shsh	Chch	ng							
Pronunciation	o'	g'e	she	che	nge	apostrophe						

Also, abbreviations that have different meanings may look the same when abbreviated. Therefore, it is advisable to organize the reading of abbreviations that do not exist on the basis of the formed abbreviations.

The main part of the structure of most scientific and technical texts consists of symbols representing sizes and their units. Although they are denoted by one or more letters, they represent a whole and independent meaning internationally. When reading the symbols representing the sizes, it is necessary to say the full name of the size. Here are a few examples of them.

- U-Kuchlanish (Voltage)
- P-Bosim (Pressure)
- F-Kuch (Power)
- F -Chastota (Frequency)
- Q-Elektrziyad (Electric charge)

One solution to this problem is to place elements of such quantities and their units in the database of systems that analyze text and synthesize them into speech signals so that physical quantities can be distinguished from ordinary words when they occur in the text.

In the text, several units of measurement can be given in words or their symbols. The synthesizer must be able to recognize any characters in the text and distinguish them from ordinary words. To do this, you need to create a database of standard units. Below are some examples of standard units.

- Mkm - MIKROMETR
- g - GRAMM
- l - LITR
- V - VOLT
- A - AMPER
- W - WATT

A series of letters or letter combinations representing such units can be continued as desired. In order to develop a relatively perfect system for analyzing text and synthesizing it into sound signals, it will be necessary to create an electronic database of units of measurement based on world standards.

Symbol recognition is one of the most difficult problems in text analysis. The main emphasis of the issue is that the author can mark such characters with any letter, depending on the content of the text.

However, for systems that do not provide for content analysis of the text, it is sufficient to identify the symbols used in the world standard. It should be noted that only conditional symbols are used here. Here are a few examples of such characters.

For example:

-Names of chemical elements in the Mendeleev periodic table;

-Determination of magnetic poles (S, N) or type of semiconductor (p, n);

-Symbols representing the functions of computer input-output devices (Tab, Alt, Ctrl, PgUp, PgDn, F1, F2..., etc.).

In all cases where the symbols represented by the letters of the alphabet are observed, the properties of the letters are accepted unchanged. When abbreviations, units of measurement, and letters and combinations of symbols are observed, it is advisable to read the full standard names.

## VII. SYMBOLS IN THE UZBEK TEXT

Text is usually a series of words and phrases that describe information related to a field. A number of symbols are used to sort and semantically separate these sentences. Such characters can be conditionally divided into two groups according to their function:

1. Punctuations
2. Special symbols

Punctuation is an important graphic tool for accurate, expressive, and logical expression of written speech in a particular language, as well as for showing the logical grammatical relationships between parts of written speech. Punctuation marks belong to the central, main character system, which differs in some respects from the additional, auxiliary character system. The use of punctuation in writing has its own system. This system - the amount of punctuation, the order of placement and the set of principles of application - creates punctuation. They are also important in expressing a variety of intellectual relationships and psychological states that cannot be expressed by other means of writing or linguistic units, and make it easier to understand written speech. The main function of punctuation is to show the semantic division of speech, as well as to help determine its syntactic structure and intonation.

The number of punctuation marks in modern Uzbek language is 10, including dots, question marks, exclamation marks, commas, parentheses, hyphens, semicolons, colon and quotation marks. The method and order of use of punctuation is determined by the principles of punctuation logical-grammatical, methodological and differentiating [8].

The corresponding codes in the standard Unicode system for expressing such punctuation marks in electronic text can be seen in Table 5 below.

**Table 5: Punctuation is represented by Unicode system.**

No	Punctuation	Punctuation code (Unicode - dec)	Pronunciation (Uzbek language)	Pronunciation (English language)
1.	!	33	Undovbelgisi	Exclamation mark
2.	"	34	Qo'shtirnoq	Quotes
3.	(	40	Ochiqqavs	Open parentheses
4.	)	41	Yopiq qavs	Close parentheses
5.	,	44	Vergul	Comma
6.	-	45	Tire	Dash
7.	.	46	Nuqta	Dot
8.	:	58	Ikki nuqta	Colon
9.	;	59	Nuqtali vergul	Semicolon
10.	?	63	So'roq belgisi	Question-mark

**Table 6: Special marks are represented by Unicode system.**

No	Special marks	Special mark code (Unicode - dec)	Pronunciation (Uzbek language)	Pronunciation (English language)
1.	#	35	Panjara	Hash
2.	\$	36	AQSH dollari	US dollar
3.	%	37	Foiz	Percent
4.	&	38	Ampersand	Ampersand
5.	*	42	Yulduzcha	Asterisks
6.	/	47	Drop	Slash
7.	@	64	Kuchukcha	At
8.	[	91	Ochiqqavsbosma	Open bracket
9.		92	Vertikalchiziq	Vertical bar
10.	]	93	Yopiqqavsbosma	Close bracket
11.	^	94	Daraja	Caret
12.	°	176	Gradus (daraja)	Degree

In addition to punctuation, there are other types of characters in the text. Such symbols are special characters, each of which has a specific meaning or function. In some cases, such symbols may have different meanings depending on the content of the text. The following table shows the most common special characters and their corresponding codes in Unicode for electronic representation.

The series of marks in the table 6 can be continued as desired. Such symbols can have different forms and meanings depending on the field and direction, as well as the diversity of languages and alphabets. Depending on the content of the text and the features of the written language, such special marks can be used as desired.

### VIII. NUMBERS IN UZBEK LANGUAGE

Plain text is followed by letters and punctuation, and followed by numbers. Numbers are one of the basic concepts in mathematics, a mathematical tool used to count things and determine quantities [9]. The concept of number, in its simplest form, originated in primitive society due to the need to count things, and has evolved over the course of human activity. Initially, the concept of whole positive numbers, and then a series of infinite

natural numbers (1, 2, 3, 4, 5 ...) emerged. The quantitative and semantic properties of numbers are studied in arithmetic subjects.

Most of the official and unofficial documents in our country use Arabic numerals in the decimal system [10]. Arabic numerals are ten mathematical symbols consisting of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, which can be used to represent any number, large or small. The following table 3 shows the electronic representation of Arabic numerals in the Unicode system.

**Table 7: Numbers are represented by Unicode system.**

No	Numbers	Number code (Unicode - dec)	Reading (Uzbek)	Reading (English)
1.	0	48	Nol	Zero
2.	1	49	Bir	One
3.	2	50	Ikki	Two
4.	3	51	Uch	Three
5.	4	52	To'rt	Four
6.	5	53	Besh	Five
7.	6	54	Olti	Six
8.	7	55	Yetti	Seven
9.	8	56	Sakkiz	Eight
10.	9	57	To'qqiz	Nine

You can create any number from the combination of numbers given in the table 3. These numbers are divided into two groups according to their nature.

1. Simple numbers
2. Complex numbers

Simple numbers can be divided into two groups. The first is simple numbers, such numbers are formed on the basis of a sequence of numbers and are read in Uzbek according to the rules of reading numbers, units, tens, hundreds, thousands or thousands, groups of millions, billions, etc. The numbers in the second group are special numbers, which are represented as a series of numbers. For example, standard serial numbers, phone numbers, credit or visa card numbers, barcodes, and more. Numbers of this type are usually compared to their standard forms, and the corresponding number is read separately for each index.

One of the biggest problems with using numbers in writing and speaking is the representation of complex numbers. The reason it is called a complex number is that it has different properties, but other symbols can be used in addition to them. It is useful to divide them into two groups when studying and analyzing complex numbers.

1. Fractional numbers
2. Non-standart numbers

A fraction is the result of measuring a quantity, that is, comparing that quantity with another quantity. A fraction is a number in mathematics that consists of one or more parts. A fraction is a ratio of two integers:  $n/m$  or  $n/m$ . Here  $m$  is called the denominator of the fraction, and  $n$  is called the image [11, 12]. If the denominator of a fraction is a number of tens, then such fractions are called decimal fractions. There are separate spelling and reading rules for decimal places in Uzbek.

Non-standard numbers include all numbers in languages and norms other than the Arabic numerals that are currently used as standard. Such numbers may vary depending on the system and language used. Roman numerals are an example of the most common non-standard numbers in our written speech. The rules for writing Roman numerals and generating multi-digit numbers are very different from Arabic numerals. Numbers are used not only in arithmetic expressions and mathematical formulas, but also in expressing the exact number of dates, times, coordinates, or things, sorting text, grouping them by meaning, and so on. In any case, the value of the number is accepted unchanged.

### IX. ALGORITHM FOR ANALYSING ELECTRONIC TEXT

The following algorithm can be used to analyze a text file consisting of the above text elements and make it ready for integration into speech fragments. This algorithm is recommended for the analysis of electronic texts in Uzbek. Given the grouping of text elements in the above order, the analysis algorithm can be expressed as follows.

As you can see from the block diagram, it is necessary to determine how many pages the incoming text is initially. From the next step, the existing text elements on each page are analyzed. Preliminary processing is performed to form appropriate speech fragments of all elements.

In order to save time in the analysis process, it is advisable to start by identifying the group of elements that have the least database and the least memory space. According to the proposed algorithm, the text analysis is performed on the basis of the following sequence:

- At the initial stage, the existing numbers on the current page are identified and edited.
- In the second step, all elements of the existing character group on the page you are viewing are edited.
- In the next step, all the letter characters present on the current page are analyzed. The types of elements represented by letters are identified and their preparation for the process of forming sound fragments is carried out. Given that the above information can be used to represent many types of text elements using letters, this is the most complex stage.

Based on these steps, each page of the text file is edited. As a result of editing, statistics are generated for each group of elements identified at each stage. This information can look something like this:

**Table 8: Table of text analysis statistics.**

Pages	Number of letters	Number of symbols	Number of digits
1.	x	y	z
...			

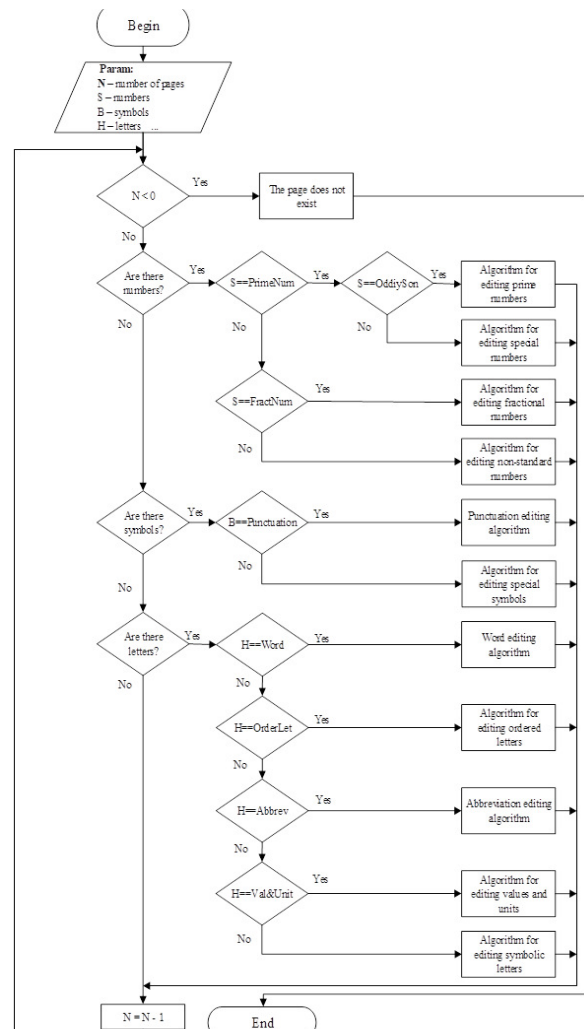
Based on the results of the analysis, it will be possible to organize the operation of the system. Each type of group of elements has its own complexity of editing. It

also takes some time to sound and present each element.

### X. CONCLUSION

The developed algorithm can be used to analyze the Uzbek text and make the elements sound. The organization of the identification of elements in the text by the code in the Unicode system allows the system to represent all the characters of the world standard.

In order to avoid interruptions and delays, the synthesis system distributes memory and other resources proportionally to the x: y: z state based on the results of the analysis. For example, if we assume that the main part of the text consists of numbers and formulas, the system directs the main resources to sound the elements of the text in this group. Or in the case of complex numbers, the system is required to solve the problem of synthesizing this unknown number before the process of sounding other characters is completed, and so on. This feature of synthesis systems is especially important in cases where real-time operation is required.



**Fig. 4.** Block diagram of the algorithm for analyzing electronic text in the Uzbek language.

## REFERENCES

- [1]. Балакина, Ю. (2016). Электронный текст: принципиально новый тип текста. *Вестник Волгоградского государственного университета*, Т.15. No. 3, 17-27.
- [2]. Мещеряков, Р. (2009). Структура систем синтеза и распознавания речи. *Известия Томского политехнического университета. Инжиниринг георесурсов*, №5, 121-126.
- [3]. Цитульский, А., Иванников А., РоголИ. (2020). Интеллектуальный анализ текста. *Stud Net*. № 6, 476-483.
- [4]. Tabet, Y., Boughazi M., Affifi S., (2015). A Tutorial on Speech Synthesis Models. *Procedia Computer Science*, 73, 48-55.
- [5]. O'zbekiston Respublikasi Qonuni (1993). "Lotinyozuviga asoslangano'zbekalifbosini joriyetishto'g'ri sida". 931-XII-son.
- [6]. O'zbekiston Respublikasi Qonuni (1995). "Lotinyozuviga asoslangano'zbekalifbosini joriyetishto'g'ri sidagi qonunigao'zgartirishlari kiritish haqida". 71-I-son.
- [7]. Otamirzayeva, S., Yusupova M. (2004). O'zbektili, fonetika, grafika, or foepiyavaorfoografiya. *O'AJBNT Markazi*.
- [8]. Maxmudov, N., Madvaliyev A., Maxkamov N. (2015). O'zbektilipunktuasiasining asosiy qoidalari. *O'zbekiston*.
- [9]. Меннингер, К. (2011). История цифр. Числа, символы, слова. М.: ЗАО Центрполиграф, 543.
- [10]. Boltayev, B., Mahkamov M., Azamatov A., Rahmonqulova S. (2017). Informatikava axborot texnologiyalari. *O'zbekiston milliy ensiklopediyasi Davlat milliy nashriyoti*.
- [11]. Wu, H. (2011). The mis-education of mathematics teachers. *Notices of the AMS*, 58(3), 372-384.
- [12]. Kalyanova, L. M. (2019). The project method in teaching special disciplines to students of technical universities. *International Journal on Emerging Technologies*, 10(3), 23-26.

**How to cite this article:** Akmuradov, B., Khamdamov, U. and Sultanov, D. (2021). Text Analyzing Algorithm for Speech Synthesizer of Uzbek Language. *International Journal of Emerging Technologies*, 12(1): 108–115.