# About Creating of the University System of Checking Texts for Plagiarism

Akkyz Mustafina, Zhuldyz Kalpeeva, and Aray Sabyrzhan International IT University, Almaty, Kazakhstan Email: {amustafina, zh.kalpeyeva}@iitu.kz, araisabyrzhan@gmail.com

> Aibek Mazhenov Satpayev University, Almaty, Kazakhstan Email: aibek.mazhenov@gmail.com

Abstract—This article describes the process of creating the University system of check texts for plagiarism. We compared and analyzed other systems recognize plagiarism in texts. Authors considered different ways to "bypass", common methods and system disorders. The results of the study complement the existing theoretical understanding in the field of software development for the detection of borrowing text. Totals open prospects for further study with the aim of improving of scientific-pedagogical method ensuring process of development of an information system "AntiplagiatKazNAU" to check text documents.

Index Terms—the checking system for uniqueness, plagiarism, anti-plagiarism software, information uniqueness, w-shingling, translit, sinonimayzer

## I. INTRODUCTION

With the development of information technologies and the Internet appeared a huge opportunity in getting various information, including teaching character, useful for self- education and broaden horizons. But along with this there was also the problem of drawing the results of others labor without giving reference to the source. In the network appeared the entire sites that offer users to download essays, term papers and dissertations. In such situation in educational institutions need to implement an effective means of combating plagiarism. One such tool is the creation of computer system check text for plagiarism [1], [2].

Plagiarism—intentionally committed by a natural person illegal use or possession of protected results of other people's creative work, which is accompanied by communication to the other persons false information about himself as about the actual author [3].

The purpose of system is to improve the quality of preparation of final qualifying works and scientific works of students and their motivation for academic integrity. Managers of qualification works check presentation, text document, disclosing of theme and originality of the text.

The corresponding check stimulates students to a more serious attitude in the preparation of final qualifying

Manuscript received November 20, 2018; revised February 2, 2019.

work and supervisors for a more careful study of the works.

#### II. METHODS OF SOLVING

Methodological basis of research made a systematic approach, the analysis of the primary data, expert interview, monitoring of media, specialized databases and monitoring official statistics and special industry publications.

We begin by phrasing the problem of similarity as one of finding sets with a relatively large intersection. We show how the problem of finding textually similar documents can be turned into such a set problem by the technique known as "shingling."

The most effective way to represent documents as sets, for the purpose of identifying lexically similar documents is to construct from the document the set of short strings that appear within it. If we do so, then documents that share pieces as short as sentences or even phrases will have many common elements in their sets, even if those sentences appear in different orders in the two documents.

W-shingling algorithm designed to search for copies and duplicates the text under consideration in the web document. Tool to detect plagiarism. The w denotes the number of tokens in each shingle in the set. For a given shingle size, the degree to which two documents *A* and *B* resemble each other can be expressed as the ratio of the magnitudes of their shinglings' intersection and union, or

$$r(A,B) = \frac{|S(A) \cap S(B)|}{|S(A) \cup S(B)|} \tag{1}$$

where |A| is the size of set A. The resemblance is a number in the range [0, 1], where 1 indicates that two documents are identical. This definition is identical with the Jaccard coefficient describing similarity and diversity of sample sets.

The stages, which is a text subjected to comparison:

- The canonization of the text;
- Splitting into shingles (Fig. 1);
- Hashes of shingles;
- A random sample of 84 values of the control amounts:
- Comparing, determination of result.

© 2019 J. Adv. Inf. Technol. doi: 10.12720/jait.10.1.24-28

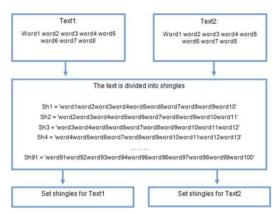


Figure 1. Splitting into shingles

The uniqueness of the system lies in the search algorithms that were developed by the authors independently, and in the formation of their own collections according to the profile of the University. There are other alternative systems that can recognize plagiarism in texts, such as Plagiarism.ru, Text.ru, Plagiarism.NET, ETXT, Antiplagiat, Advego Plagiatus, Strikeplagiarism.com.

The following is a comparative analysis of the test the two theses in these online services and in the system "Antiplagiat KazNAU" that are used to check text for plagiarism (Table I).

TABLE I. THE RESULTS OF CHECKING 2 GRADUATION WORKS

Authors		Zhansaya Sakenova	Arsen Azhirov	
Language of d	ocument	kazakh	russia	
% of the borrowing	antiplagiat.ru	8.36%	37.07%	
	strikeplagiarism .com	9.2%	18%	
	plagiat.kaznau.kz	38.5%	23%	

The results of the comparison showed that the system "Antiplagiat" and "Strikeplagiarism.com" have an approximately identical level of borrowing text searches across multiple collections of sources, and describe in detail a large number of sources, which is a plus of these systems.

#### III. RESULTS

The authors developed a self-Plagiarism system for the University of KazNAU, allows you to check text documents for borrowings and to determine the total (in percentage) the originality of the text (Fig. 2). This system can be successfully used by students of the Kazakh national agrarian University to check for

plagiarism of final qualification works, course papers, essays and other text documents. Multi-level in nature and multifunctional in purpose the system is intended to students, undergraduates, doctoral students, students, teachers of higher educational institutions, colleges, as well as a wide range of users wishing to check the material for borrowings without reference to the source.



Figure 2. Main page of the system "Antiplagiat KazNAU"

In world practice, the use of such systems is not new. But existing systems are paid or shareware. Besides using a diverse base of work for comparison. Some free check a document only on open sources from the Internet, and access to private databases is paid. An important issue in the functioning of such systems is the formation base of operations. The system allows part or full check of the document the subject of borrowing texts from public Internet sources and from their own base, which is partially closed to outside users. In General, the service is free, anyone student or teacher can check the document on the borrowing of material, but to produce a report on the borrowing of the text you want to log into the system.

The main goal is verification of the originality and recognition information about whether the copied text is from an existing source database, or has a high percentage of uniqueness.

The result is a validation report document (Fig. 3). The report shows those portions of the document that were found in the source collection System, and specifies in percentage the ratio of borrowed and original text.

After researching loaded text the system gives a percent of its originality, borrowing, and leads the list of the detected sources. In order to see which fragment is taken from a source requires authorization.

Types of system users are divided into several categories:

- Administrator a specialist who maintains the website in good technical condition, change the access levels of verification for students and teachers;
- Teacher-head of Department, the results of inspections of qualifying works of students of this Department;
- Lecturer-head of the qualification, available for all inspection reports of a student who passed a text document for checking;
- Student has access to their documents uploaded to the System.

The system supports users authorization. In authorizing the System, the user enters his personal

account, where he receives a set of services corresponding to their user rights. In personal account user can view the statistics checked for plagiarism works, as well as to report on the percentage of borrowing. The teacher through a personal account of the available reports on the work of students who indicated it as a supervisor.

The recommended threshold value of uniqueness is on average 75% of the original text of the total work,

however this value can be set by the administration of the University itself.

The level of access for system users depending on the position of teachers, staff. For students of KazNAU access level is the same.

After filling in all required fields and upload the text document passes the validation process (Fig. 4), then the user displays the percentage of originality, of borrowing, the list of detected sources (Fig. 5).

#	Author	Group name	Name of project	Supervisor's name	checked date	Unique	Plagiat	Document	Certification
1	Калдыбаев Азамат Талгатович	MTMO-201.	Сұйық әртекті жүйелерді ажырату үрдісін зерттеу	Оспанов Абдыманап	2018-06- 21 10:48:39	96%	4%	Document	Certification
2	Калтаева Айдын Муратовна	MTMO-201.	Күнбағыс майын бөлу кезіндегі сығымдау үрдісін зерттеу	Оспанов Абдыманап	2018-06- 20 16:27:16	89%	11%	Document	Certification
3	Райсова Акнур Кайратовна	МТПРП-201.	Балалар тағамы технологиясында экструзионды- модифицирленген ұнды пайдалану	Тимурбекова Айгуль	2018-06- 20 16:23:14	84%	16%	Document	Certification
4	Умбетов Ернар Ержанович	ЮП-102 (зем.)	АЛАШ ҚАЙРАТКЕРЛЕРІНІҢ КӨЗҚАРАСТАРЫНДАҒЫ ҚАЗАҚСТАНДАҒЫ ЖЕР САЯСАТЫНЫҢ ҚҰҚЫҚТЫҚ МӘСЕЛЕЛЕРІ	Куандыков Бахытжан	2018-06- 20 14:49:50	87%	13%	Document	Certification
5	Нұрсеит Абай Мырзағалиұлы	ЮП-102 (зем.)	ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ЖЕР РЕФОРМАЛАРЫНЫҢ ӨЗЕКТІ МӘСЕЛЕЛЕРІ: ТАРИХИ-ҚҰҚЫҚТЫҚ АСПЕКТІЛЕРІ ЖӘНЕ ДАМУ БАГЫТТАРЫ	Тайторина Бинур	2018-06- 20 14:39:22	96%	4%	Document	Certification
6	Досмуханбетов Дархан Балтабекович	ЮП-101 (аграр.)	ШАРУА (ФЕРМЕР) ҚОЖАЛЫҒЫ ҚЫЗМЕТІН ҚҰҚЫҚТЫҚ РЕТТЕУ	Аюпова Зауре	2018-06- 20 14:32:43	57%	43%	Document	Certification
7	Төлеу Абылай Әділбекұлы	ЮП-102 (зем.)	ҚАЗАҚ ӘДЕТ ҚҰҚЫҒЫНДАҒЫ МАЛ ДАУЫ	Куандыков Бахытжан	2018-06- 19 15:18:22	88%	12%	Document	Certification
8	Кемельбек Еламан Қабылбайұлы	ЮП-102 (зем.)	АГРАРНО-ПРАВОВАЯ ПОЛИТИКА КАЗАХСТАНА ВО ВЗГЛЯДАХ ПЕРВОГО ПРЕЗИДЕНТА РЕСПУБЛИКИ КАЗАХСТАН	Кунхожаева Гульнара	2018-06- 19 15:08:48	88%	12%	Document	Certification
9	Жаңбырбаев Бейбарс Ералыұлы	ЮП-101 (аграр.)	ҚАЗАҚ ӘДЕТ ҚҰҚЫҒЫНДАҒЫ ЖЕР ДАҰЫ	Куандыков Бахытжан	2018-06- 19 15:02:17	93%	7%	Document	Certification
10	Кондыбаева Алтынай Канатовна	МЮП-202	КОНСТИТУЦИЯЛЫҚ-ҚҰҚЫҚТЫҚ ҚАГИДАЛАР	Тайгамитов Газиз	2018-06- 19 11:59:16	48%	52%	Document	Certification

Figure 3. The results of inspections of diploma works of students of KazNAU



Figure 4. Verification process the text document for plagiarism

The basic functionality of the System:

 Search for loanwords in the text documents in the format .docx.

- Ability to work with the System through the website.
- Create inspection reports with the release of the borrowed fragments of text without reference to the sources and an indication of the list of loan sources.
- Classification of documents by type (theses, dissertations, course works, essays, etc.).
- The user can access the personal account.
- Formation of own base of internal sources of proven work.
- Ability to check work on open bases on the Internet.
- The ability to connect to paid databases of other universities.
- Detection in scanned document such of its parts as contents, bibliography and quotations are enclosed in quotation marks, and the exclusion of such texts from plagiarism.

#### Author: Төлеубек Дархан Төлеубек - Уникальность:93%

Name of project: Алматы облысы, Балқаш ауданының әкімшілік қызметін жетілдіру

Cafedras:Кафедра "Менеджмент и организация агробизнеса"

#	Плагиат: 7%	Name of project	Cafedras
1	1.13%	Қызылорда облысы, Шиелі орман және жануарлар дүниесін қорғау мемлекеттік мекемесінде тұрақты орман питомнигін ұйымдастыру	"Лесные ресурсы и охотоведение"
2	1.13%	Мемлекеттік бюджеттің табыс белігінің қалыптасуындағы салық органдарының ролі (Алматы облысы, Талғар қаласының салық комитетінің мысалдары негізінде)	Кафедра "Менеджмент и организация агробизнеса"
3	0.68%	Инновациялық жобаларды қалыптастыру принциптері және бағалау әдістері («Бизнес Сервис GROUP» ЖШС - нің мәліметтері бойынша)	Кафедра "Менеджмент и организация агробизнеса"
4	0.45%	Қазақстан Республикасында фискалды салық жүйесінің қалыптасуы	Кафедра "Менеджмент и организация агробизнеса"
5	0.45%	Қысқа мерзімді міндеттемелердің есебі мен аудиті	Кафедра "Менеджмент и организация агробизнеса"
6	0.45%	Талғар қаласы бойынша Кіріс басқармасының салық төлеушілер есебінің ақпараттық жүйесін құру	"Математика және физика"
7	0.45%	Энергия үнемдеуіш желдету жүйесін әзірлеп қой телдету қорасын электрмен жабдықтау	Энергосбережение и автоматика
8	0.45%	Бюджетпен есеп айырысу операцияларының есебі мен аудиті	Кафедра "Менеджмент и организация агробизнеса"
9	0.45%	Екінші деңгейлі банктердегі қаржылық менеджметті жетілдіру ("Халық банк" АҚ мысалында)	Кафедра "Менеджмент и организация агробизнеса"
10	0.23%	Бюджетпен есеп айырысу операцияларының есебі мен аудиті	Кафедра "Менеджмент и организация агробизнеса"
11	0.23%	Әлеуметтік-экономикалық жағдайда ауыл шаруашылығы өндірісін құрылымдық басқаруды жақсарту жолдары (Алматы облысы, Балқаш ауданы жобасында)	Кафедра "Менеджмент и организация агробизнеса"
12	0.23%	Ауыл экономикасын тиімді басқарудың басым бағыттары (Алматы облысы Балқаш ауданы Бақбақты ауылы мысалында)	Кафедра "Менеджмент и организация агробизнеса"
13	0.23%	Роз-бенгал антигенін дайындауды игеру және оны практикада қолдану	"Акушерство, хирургия и биотехнология воспроизводства животных"
14	0.23%	«Ауылшаруашылық саласын басқару тиімділігі (Алматы облысы, Балқаш ауданының мөліметтері негізінде)»	"Экология"
15	0.23%	Алматы облысы, Балқаш ауданы ауыл шаруашылығын басқару жетілдіру жолдары	Кафедра "Менеджмент и организация агробизнеса"

Figure 5. The result of originality, of borrowing, the list of detected sources

It should be noted that with the development of systems for the verification of texts for plagiarism began to appear a variety of methods and ways to "bypass". Any verification system is not perfect, has its weaknesses, there is always a way to cheat the system [4].

Qualitative analysis of the works authenticated through our system, have revealed a number of violations in order to "bypass" algorithm search plagiarism. Among the most common "tricks" include:

- Full or partial replacement of the letters on the various characters of the alphabet;
- The merging of multiple words and sentences by removing spaces between words;
- Full or partial replacement of a text into another text, such as excerpts from literature and textbooks;
- Dilution sentences introductory words;
- Find the topic in English and translate it;
- The use of sinonimayzer words are identical;
- Violation of the integrity of the structure of the work;
- Check a simple piece of work.

Algorithms of systems provide methods of protection from "above" aimed at avoiding detection of plagiarism. In order to ensure that the correct document loaded into the system, after loading the system recognizes the number of characters in the text and the file size. This information will be included in the reference and the

teacher can ascertain by comparing the data with the original document.

The system is designed as Internet service [5], [6], all the calculations occur on the technical equipment of the Kazakh national agrarian University, therefore, students and teachers can operate the system from any computer connected to the Internet.

### IV. CONCLUSIONS

Control system of texts can only perform the function of intermediary between the expert and the collection of documents, where there is a search of borrowing and the check is a time consuming intellectual process. The decision about the originality of a thesis remains with the supervisor, after a deep meaningful analysis of the work, which is impossible without a careful reading of the text.

The use of the University system Antiplagiat is a factor influencing the improvement of quality of preparation of final qualification works. The corresponding validation test stimulates students to a more serious relationship in the preparation of final qualifying work, and supervisors - for a more careful study of the theses.

# REFERENCES

 E. V. Sharapova and P. V. Sharapov, "Universal system for checking texts for plagiarism "Author.net"," *Information Science* and Application, vol. 6, no. 3, pp. 52-58, 2012.

- [2] A. V. Sliva, V. N. Fokina, A. V. Abramova, and M. E. Shirokova, "Methods of improving the software tools to detect plagiarism," *Remote and Virtual Learning*, no. 7, pp. 92-99, 2015.
- [3] A. K. Mustafina and Z. B. Kalpeyeva, "IT-infrastructure of university based on cloud computing," *International Journal of Computer Science Issues*, vol. 10, no. 5, pp. 176-179, Sep. 2013.
- [4] A. K. Mustafina and Z. B. Kalpeyeva, "Cloud-based solutions," *Bulletin of KazNTU*, pp. 202-209, 2012.
  [5] U. A. Zagorulka, N. V. Salomatina, A. S. Seriy, A. Sidorova, and
- [5] U. A. Zagorulka, N. V. Salomatina, A. S. Seriy, A. Sidorova, and B. K. Shestakov, "Identifying fuzzy duplicates with automatic creation of thematic collections of documents, web-based publications," Novosibirsk, 2013.
- [6] Internet service "Antiplagiat" [Official. site]. [Online]. Available: https://www.antiplagiat.ru



**Akkyz Mustafina** was born in Almaty, Kazakhstan on March 7, 1963. She graduated from Al-Farabi Kazakh National University in Almaty, Kazakhstan in June 1981.

After receiving her education, in 1983, Akkyz started to work in Ministry of Automation and Control Soviet Union Systems, Scientific Production Association as a system integrator. During her work she participated, then managed of creating software industry at ICS

for Soviet Union non-ferrous metallurgies. In 1991, she was assigned as a teacher of programming languages and database management in Academy of Transport and Communications, as the same time successfully managed creating staff workplaces for railway companies. At 1997, she was engaged in the development of a subsystem for the coordination of the educational process of the Corporate Information System. After her diligent work, the task of forming curricula for specialties, calculating the academic load of departments, etc were introduced to Satpayev University. Since 2004, she was director of Information Technology Department. Simultaneously, she managed the work on the development of structured cable network and the creation of an "E-university" of Satpayev University. Under her leadership was developed educational WAP-portal of the University, hardware and software portal "News Stand", as well as remoted multimedia complex for teaching technical specialists by program "Creation of information systems for the development of the state language and other languages of the peoples of Kazakhstan" for the Committee on Languages of the Ministry of Culture and Information of the Republic of Kazakhstan. She was on the authors who regulated credit system of education and testrating system of the University. From February 2016 to August 2017, she worked as director of Institute Information Technology and associated professor of the of Automation and Information technology Department at the Kazakh National Agrarian University, Almaty, Kazakhstan. During her work time she created new official university website, Antiplagiarism system, web-services as "Attendance", "Event manager", University repository and provided the effective functioning of AIS "Platonus". Since September 2017, she worked as a vice-rector at the University of Foreign Languages and Business Careers. Under her leadership, the university successfully passed the institutional and specialized accreditation of 10 educational programs, and she coordinated the work on the creation of an official website and automation of the educational process of the university. Currently, she is working at International IT University.

Since 2009, Akkyz Mustafina is the Candidate of Technical Science in the field system analysis and data processing management at Satpayev University, Almaty, Kazakhstan. She is author of more than 70 scientific works and 2 teaching aids, also she has 15 copyright works.

She was one of the commission in the expert group of the NAAR of Ministry of Education and Science of the Republic of Kazakhstan. Akkyz was a chairman of the MAC (Main Attestation Commission) and the Republican competition of scientific projects schoolchildren "Daryn", participated in the projects of Tempus. Scientific interests of Akkyz Mustafina are related to the modeling of technological and business process of automation, the design of information systems, computer network and machine learning, the development and implementation of software for automated systems of various composition and purpose, also works with data. She is married, has two children.



**Zhuldyz Kalpeeva** was born on April 14, 1980. She graduated from Kyrgyz National University bachelor degree in June 2001 and master degree in June 2003. She received her doctor degree at Satpayev University in June 2014.

Since 2006 she was assigned head of "WEB and Media of Information Technologies Department" at the Satpayev University, Almaty, Kazakhstan. Her previous position

was a head of the "Automation and IT Department" in the KazNAU.Currently Zhuldyz is Assistant Professor of the "Computer Engineering and Telecommunications Department" at the International IT University, Almaty, Kazakhstan.

Zhuldyz Kalpeeva is Doctor (Ph.D.) and author of more than 40 scientific papers, 4 of international publications indexed by Scopus:

- 1. Uskenbayeva, R.K., Kalpeyeva, J.B., Kassymova, A.B. Organization of computational processes in distributed cloud environments // 54th Annual Conference of the Society of Instrument and Control Engineers of Japan, SICE 2015.
- 2. Uskenbayeva, R.K., Kuandykov, A.A., Cho, Y.I., Kalpeyeva, Zh.B. Tasks scheduling and resource allocation in distributed cloud environments // International Conference on Control, Automation and Systems, 2014.
- 3. Uskenbayeva, R.K., Kuandykov, A.A., Im, C.Y., Kalpeyeva, Zh.B., Kozhamzharova, D.K. Organization of computing processes in the large heterogeneous distributed systems // 44th International Symposium on Robotics, ISR 2013.

Her main attention is directed to managing IT department of the university, design and setting of tasks. Field of Interest: cloud computing, business processes, distributed systems. She is married, has two children.



Aray Sabyrzhan was born in Almaty, Kazakhstan on October 3, 1995. She graduated from International IT University in Almaty, Kazakhstan in June 2017. Her major of bachelor degree is CS and Software Engineering.

During her studying years she worked as a full-stack software developer in KazNAU. After receiving her education, she continued her direct as software developer in Choco-

family holding company. Currently she's working at startup company Amplify Live with project LEA. She engage with big data as users responds and web development.

Aray is co-author of 4 scientific papers which connected with WEB-services, Data Science, Data visualization and Logistics systems. Also she is interested in physical activities like: snowboarding, skiing, swimming and camping.