



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

A.K. Mustafina, Zh.B. Kalpeeva, A.K. Mazhenov and A.A. Sabyrzhan  
*Kazakh National Agrarian University, Almaty, Kazakhstan*

**Key words:** The checking system for uniqueness, plagiarism, anti-plagiarism software information uniqueness, w-shingling, translit, sinonimayzer

### Corresponding Author:

A.K. Mustafina  
*Kazakh National Agrarian University, Almaty, Kazakhstan*

Page No.: 20-25

Volume: 15, Issue 1, 2020

ISSN: 1816-9503

International Journal of Soft Computing

Copy Right: Medwell Publications

**Abstract:** This study describes the process of creating the University system of check texts for plagiarism. We compared and analyzed other systems recognize plagiarism in texts. Researchers 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 “Antiplagiat Kaz NAU” to check text documents.

## 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<sup>[1, 2]</sup>.

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 researcher<sup>[3]</sup>.

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

## MATERIALS AND METHODS

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 (Fig. 1).

The most effective way to represent documents as sets for the purpose of identifying lexically similar documents is to construct from he document the set of

Table 1: The results of checking 2 graduation works

Researchers	Language of document (%)	Percentage of the borrowing		
		Antiplagiat. RU (%)	Strikeplagiarism.com (%)	Antiplagiat KazNAU (%)
Zh. Sakenova	Kazakh	8.36	9.2	18
A. Azhirov	Russia	37.07	38.5	23

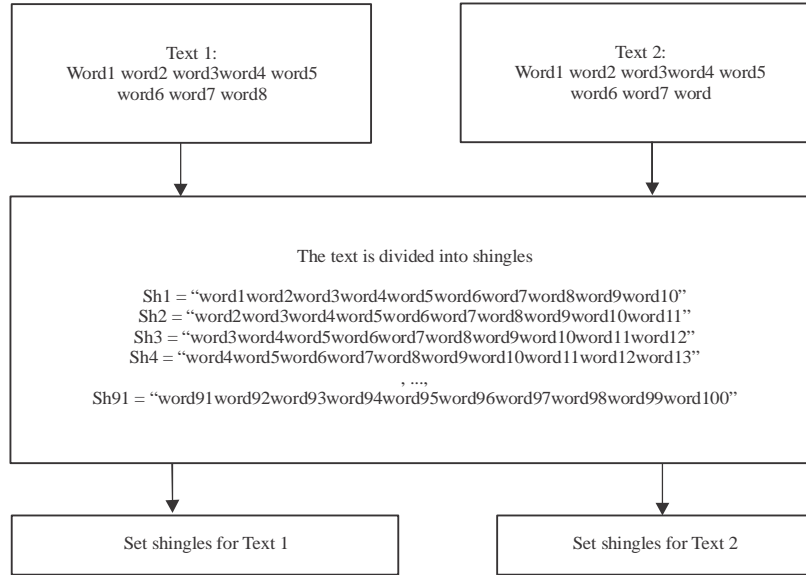


Fig. 1: Splitting into shingles

short strings that appear with in 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 shingling’s intersection and union or:

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

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 uniqueness of the system lies in the search algorithms that were developed by the researchers 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 Plagiatius, Strike plagiarism.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 1).

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.

## RESULTS AND DISCUSSION

The researchers 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



Fig. 2: Main page of the system “AntiPlagiat kazNAU”

#	Автор	Группа	Название работы	ФИО руководителя	Дата проверки	Уникальность	Plagiat	Document	Certification
1	Бердиев Нуркен Маликулы	ПА-405К	Алматы облысы жағдайында ралс дауаланып өкөндөпөтө минералдық тыңайтқыштардың тиімділігі	Уленбаева Гулнур	2017-05-31 17.11.18	94%	6%	Document	Certification
2	Расилова Айгүль Аумбеловна	ТППЖ-409 Р	Использование селекционно-генетических параметров в селекции депрессивных овец с полуробутой шерстью	Ким Галина	2017-05-31 17.29.16	67%	33%	Document	Certification
3	Телебергенов Аян Бекенулы	О3-405К	Алматы облысы Үйгентас орман және жануарлар дүниесі қорғау мемлекеттік мекемесінде қолданылатын омыртқасыз жолдары	Байаташов Муктар	2017-05-31 17.26.28	92%	8%	Document	Certification
4	Мусан Ержан Айбоқтылы	О3-405К	Жамбыл облысы, Мойынқум орман және жануарлар дүниесі қорғау мемлекеттік мекемесіндегі қарақобан қорғау және қолдау шаралары	Керімбеке Сатыбалды	2017-05-31 17.22.58	98%	2%	Document	Certification
5	Махатов Асманбай Сатыбалды	ТППЖ-407К	«Onara Ranch» агрофирмасындағы етті бағытты ұрушылардың ет өнімдері	Аманжолов Кедірбай	2017-05-31 17.22.35	86%	14%	Document	Certification
6	Абдуллин Рахима Бақытқызы	ТППЖ-409 Р	Мясная продуктивность молодянка депрессивной породы овец, разводимых в условиях племязоа «Имади» Алтайской области	Адьянова Шолпан	2017-05-31 17.14.06	66%	34%	Document	Certification
7	Иван Жайыл Райсаулы	О3-405К	Биоциклонное разнообразие культурных животных в Азиатском государственном прерорном заказнике	Керімбеке Сатыбалды	2017-05-31 17.23.30	66%	34%	Document	Certification
8	Жақыпбек Назерке Қамалқызы	МР-408К	Оңтүстік Қазақстан облысы, Мақтарал ауданындағы «Абай» ауылының сумен қамтамасыз ету жобасы	Жапарұлова Ермекулы	2017-05-31 16.36.39	74%	26%	Document	Certification
9	Ерлы Бірысты Жанжолқызы	МР-408К	Оңтүстік Қазақстан облысы, Салыарал ауданы «Достық» шаруа қожалығының жергілікті суару	Жапарұлова Ермекулы	2017-05-31 16.36.38	64%	36%	Document	Certification
10	Жақыпбек Назерке Қамалқызы	МР-408К	Оңтүстік Қазақстан облысы, Мақтарал ауданындағы «Абай» ауылының сумен қамтамасыз ету жобасы	Жапарұлова Ермекулы	2017-05-31 16.36.37	64%	36%	Document	Certification
11	Тоқсанбай Аяулы Аюрбеқызы	АТТ-417К	Алматы облысындағы «Ақпелек» шаруа қожалығында сары сару және сүтті ағышты ендуді механизациялауы	Ушакирова Анар	2017-05-31 16.52.17	86%	14%	Document	Certification

Fig. 3: The results of inspections of diploma works of students of KazNAU

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.

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 pre-loaded text Fig. 4 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.

Наберите ФИО автора \* Author of Work

---

Название работы \* Name of Work

---

Аннотация \* Annotation

---

Наберите ФИО руководителя \* Supervisor's name

---

Тематическая рубрика

---

Тип работы

---

Выберите язык

---

Choose File No file chosen

Fig. 4: From to pass validation

Автор: Шарип Алтынай

Руководитель: Сахариянов Ақылбай

Тема: Дикроцелиозбен ауырған қой етінің санитариялық сапасын сезімдік және биохимиялық көрсеткіштері бойынша анықтау

Проверка завершена, оригинальность теста: 84%

Save

КОММЕРЦИАЛЬК ЕМЕС АКЦИОНЕРЛІК ҚОҒАМ ҚАЗАҚ ҰЛТТЫҚ АГРАРЛЫҚ УНИВЕРСИТЕТІ «Ветеринария» факультеті 5В120200 «Ветеринарлық санитария» мамандығы «Ветеринариялық санитариялық саралтау және гигиена» кафедрасы Дипломдық жұмысты (жобаны) орындау ТАПСЫРМАСЫ Студент Шарип Алтынай Жұмыс (жоба) тақырыбы «Дикроцелиозбен ауырған қой етінің санитариялық сапасын сезімдік және биохимиялық көрсеткіштері бойынша анықтау» Университет бойынша 2016 жылдың 26 қазанындағы 510 К бұйрығымен бекітілген Дайын жұмысты (жобаны) тапсыру мерзімі 2017 ж. 2 мамыр Жұмыстың (жобаның) бастапқы деректері 1 Ауру қоздырғышының морфологиясы мен биологиясы 2 Эпизоотологиялық көрсеткіштері 3 Аурудың дамуы 4 Клиникалық беллері және балау 5 Мясo и мясные продукты.Методы отбора проб. АНЫҚТАМАЛАР Дипломдық жұмыста төмендегі анықтамалар мен терминдер пайдаланылды: ДИКРОЦЕЛИОЗ ауыл шаруашылығы малдарының кеннен тараған гелминтозды ауруларының бір түрі ДДСҰ дүниежүзілік денсаулық сақтау ұйымы FRUTICICOLA FRUTICUM құрлық ұлуының Fruticicola туыстастықтарына жататын ұлулар FORMICA RUFIBARBIS Formica туыстастығына жататын құмырсқа ГЕЛЬМИНТОКАПРОЛОГИЯЛЫҚ ЗЕРТТЕУ мал қиын бірнеше рет жуу арқылы зерттеу ТРЕМАТОДАЛАР жалпақ құрттар, бұлар қоздыратын ауруларды трематодоздар деп атайды ДЕФИНИТИВТИ ИЕСІ бұл жыныстық жетілген паразиттер өмір сүретін мал организмі АГОНИЯ малдың өлер алдында жаңталасып жатқан кезі МИРАЦИДИИ дикроцелий жұмыртқасында дамыған бірінші балаң құрттық форма АНАЛЫҚ СПОРОЦИСТА дикроцелийдің екінші балаң құрттық кезеңі ЖЕТКІН СПОРОЦИСТЕР дикроцелийдің үшінші балаң құрттық кезеңі ЦЕРКАРИИ дикроцелийдің төртінші балаң құрттық кезеңі МЕТАЦЕРКАРИИ дикроцелийдің бесінші балаң құрттық кезеңі. БЕЛПЛЕУЛЕР МЕН «Агроуниверситет» баспасы, 2003. 460 бет. 6 Шуклин Н.Ф., Қырықбайұлы С.и др. Экспертиза доброкачественности и радиационной безопасности продуктов их стандартизация и сертификация. Алматы, «CREDOS», 1 2 3 том, 2011, 426

Fig. 5: Verification process the text document for plagiarism

The system supports authorization of users. 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. Types of system users are divided into several categories.

Administrator B 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 (Fig. 5).

Автор: Телеугали Еркем Ерланұлы - Уникальность:96%

Название работы: „Көксу, ЖШС-де қант қызылшасын тасымалдау жұмыстарын ұйымдастыруды жетілдіру

Кафедра:”Машиноиспользование”

#	плагиат: 4%	Название работы	Кафедра
1	0.75%	Алматы қаласындағы “Қамқор” ЖШС Жөніп автомобильге ТҚ және алдыңғы асықтың топсасын шығару қондырғысын жобалау	”Машиноиспользование”
2	0.5%	Одёрлі аймағында Ssang Yong машинасының тосқауы құрылысын жобалау және жетілдіру	”Механика и конструирование сельскохозяйственной техники”
3	0.25%	Алматы облысы, Үшарал қаласында көлік жөндеу станциясында жөніп автомобильдерге техникалық қызмет көрсетуді жетілдіру	”Машиноиспользование”
4	0.25%	Совершенствование организации дорожного движения на улично-дорожной сети мкр. Шанырақ г.Алматы	”Машиноиспользование”
5	0.25%	Қазақстанның оңтүстік-шығыс аймағында рапстың себу мерзімінің өнімділігіне әсері	”Агронмия”
6	0.25%	«Star Motors» ЖШС автомобильдер техникалық қызмет көрсету және клапандарды түзету қондырғысын жобалау	”Машиноиспользование”
7	0.25%	Алматы-Тараз автомагистралінде келкілерге жөндеу және шеберханасын жобалау	”Машиноиспользование”
8	0.25%	Логистический подход в деятельности транспортной компании «АВТ» г. Алматы	”Машиноиспользование”
9	0.25%	ШҚО облысы, Ұршақ ауданында «Өділет» ТҚ орталығында жүк автомобильдеріне техникалық қызмет көрсетуді жетілдіру	”Машиноиспользование”
10	0.25%	«Агрошколадинг» акционерлік қоғамының кәсіпорнында тапсырыс беруші базасын тандау ұрақын басқаруды жетілдіру»	”Машиноиспользование”
11	0.25%	Жамбыл ауданының Ұзынағаш ауылының Жамбыл көшесіндегі жол қозғалысын ұйымдастыруды жақсарту	”Машиноиспользование”
12	0.25%	Есел айырысу операцияларының есебі мен аудиті	”Энергосбережения и автоматика”
13	0.25%	Исследование проблемы загрязнения атмосферного воздуха от автотранспортных средств в условиях г. Алматы	”Аграрная техника и технология”

Fig. 6: The results of originality of borrowing the list of detected sources

Student has access to their documents uploaded to the system. After filling in all required fields and upload the text document passes the validation process (Fig. 5) then the user displays the percentage of originality of borrowing, the list of detected sources (Fig. 6).

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
- The user can access the personal account
- Formation of own base of internal sources of proven work
- Classification of documents by type (theses, dissertations, course works, essays, etc.)
- 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

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<sup>[4]</sup>.

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

The system is designed as internet service<sup>[5]</sup>, 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.

## **CONCLUSION**

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 research 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**

01. Sharapova, E.V., 2012. Universal system for checking texts for plagiarism. *Inf. Sci. Appl.*, 6: 52-58.
02. Sliva, A.V., V.N. Fokina, A.V. Abramova and M.E. Shirokova, 2015. Methods of improving the software tools to detect plagiarism. *Remote Virtual Learn.*, 7: 92-99.
03. Kalpeyeva, Z.B. and A.K. Mustafina, 2013. IT-infrastructure of university based on cloud computing. *Intl. J. Comput. Sci.*, 10: 176-179.
04. Mustafina, A.K. and Z.H.B. Kalpeyeva, 2012. Cloud-based solutions. *Bull. KazNTU. Almaty*, 4: 202-209.
05. Zagorulka, U.A., N.V. Salomatina, A.S. Seriy, A. Sidorova and B.K. Shestakov, 2013. Identifying Fuzzy Duplicates with Automatic Creation of Thematic Collections of Documents. Nauka Publisher, Novosibirsk, Russia,.