

ОРГАНІЗАЦІЯ ПЕРЕВІРКИ СТУДЕНТСЬКИХ І НАУКОВИХ РОБІТ НА ПЛАГІАТ

Олександр Меньяйленко¹, Олег Захожай²*ДЗ «Луганський національний університет ім. Тараса Шевченка»¹**ДЗ «Донбаський державний технічний університет»²***Анотація:**

Розглянуто наявні проблеми створення систем автоматизованої перевірки текстів на унікальність. На основі проведеного дослідження доведено наявність суттєвого впливу на достовірність результатів аналізу текстів на плагіат різноманітних методів штучного підвищення відсотку унікальності. Запропоновано варіант організації системи перевірки на плагіат, що дає змогу підвищити ступінь інваріантності результату аналізу завдяки використанню методів штучного підвищення унікальності текстів.

Ключові слова:

плагіат; системи перевірки унікальності текстів; методи штучного підвищення ступеня унікальності текстів.

Аннотация:

Меньяйленко Александр, Захожай Олег. Организация проверки студенческих и научных работ на плагиат.

Рассматриваются существующие проблемы создания систем автоматизированной проверки текстов на уникальность. На основе проведенных исследований доказано существенное влияние на достоверность результатов анализа текстов на плагиат различных методов искусственного повышения процента уникальности. Предложен вариант организации системы проверки на плагиат, которая позволяет увеличить степень инвариантности результата анализа благодаря использованию методов искусственного повышения уникальности текстов.

Ключевые слова:

плагиат; системы проверки уникальности текстов; методы искусственного повышения степени уникальности текстов.

Resume:

Menyaylenko Oleksandr, Zakhzhay Oleh. Organization of student and scientific works verification on plagiarism.

The article considers the existing problems of creating automated text verification systems for uniqueness. The conducted researches proved a significant influence of various methods of artificial increase of uniqueness on the reliability of the texts analysis results on plagiarism. The authors propose a variant of plagiarism verification system organization which allows increasing the degree of analysis result invariance when the methods of artificial increase of texts uniqueness are used.

Key words:

plagiarism; systems of text verification systems for uniqueness; methods of artificial increase of texts uniqueness.

Постановка проблеми. З розвитком інформаційних технологій і поширенням глобальних баз і сховищ даних з відкритим доступом дедалі ширшим стає намагання запозичувати чужі результати інтелектуальної праці. Останнім часом про плагіат досить часто можна почути в будь-яких сферах людської діяльності. На жаль, не стала винятком і галузь освіти та науки.

У галузі освіти плагіат (так званий «академічний плагіат») можна виявити під час перевірки будь-яких навчальних завдань – від контрольних, модульних робіт, рефератів, доповідей до курсових і дипломних робіт. Крім цього, академічний плагіат спостерігається також у процесі підготовки навчально-методичної літератури, наукових публікацій тощо.

Основними джерелами плагіату в навчальних роботах є відкриті Інтернет-ресурси з достатньо великими базами готових студентських робіт; іноді навіть окремі web-ресурси цілком присвячені певній тематиці.

Інша форма плагіату – виконання студентських навчальних завдань іншими особами. Широкий попит на такі послуги призвів до виникнення окремого напряму підприємницької діяльності, коли з'являються й успішно функціонують не тільки приватні виконавці, а й окремі фірми, що спеціалізуються на створенні розвиненого штату виконавців студентських робіт, а також на розміщенні

відповідної реклами й отриманні значної кількості замовлень. Іноді такі фірми об'єднують виконавців з різних держав, що виконують роботи з надзвичайно широкого спектра навчальних предметів і різними мовами.

Ще однією формою поширення плагіату є використання матеріалів студентських робіт, що виконувалися студентами старших курсів у попередні роки. У такому разі студенти старших курсів передають результати своєї інтелектуальної праці студентам молодших курсів. Ця форма плагіату набуває популярності тоді, коли викладачі з року в рік не змінюють тематику й варіанти завдань, що веде до поступового формування серед студентів бази даних готових робіт з визначеної дисципліни або групи дисциплін.

Тенденція до поширення й збільшення масштабів плагіату вказує на необхідність обов'язкового впровадження вищими навчальними закладами методів і засобів перевірки студентських робіт на плагіат і впровадження організаційних заходів щодо запобігання плагіату й підвищення якості та ефективності виконання студентами своїх навчальних завдань.

Формулювання цілей статті. Мета статті – розглянути наявні проблеми створення систем автоматизованої перевірки текстів на унікальність; довести суттєвий вплив різноманітних методів штучного підвищення відсотку унікальності на достовірність

результатів аналізу текстів на плагіат; запропонувати варіант організації системи перевірки на плагіат, що дасть змогу підвищити ступень інваріантності результату аналізу, завдяки використанню методів штучного підвищення унікальності текстів.

Виклад основного матеріалу дослідження. Для визначення масштабів поширення різних форм плагіату серед студентів 3-5 курсів вищого навчального закладу IV рівня акредитації було проведено анонімне анкетування, у якому пропонувалося відповісти на 3 запитання в анкеті, форма якої наведена в табл. 1. Умовами анкетування передбачалося, що респондент ставить позначку в полі «ТАК», якщо під час його навчання мав місце хоча б один факт, опис якого міститься в анкеті. Результати анкетування зведено в табл. 2.

Спираючись на дані, наведені в табл. 2, можна дійти висновку, що плагіат, на жаль, є достатньо поширеним способом виконання навчальних завдань серед студентства. При цьому слід зауважити, що суб'єкти

освітнього процесу здебільшого як джерела плагіату використовують не відкриті ресурси Інтернету, а роботи студентів-попередників, що вивчали дисципліну в семестрах, що передують поточному. Менш популярним є таке джерело плагіату, як замовлення робіт у сторонніх осіб, але й цей варіант має достатньо високий відсоток. Вочевидь, що співвідношення відсотків використання різних форм плагіату може бути неоднаковим для різних навчальних закладів, регіонів їхнього розташування, кількості студентів тощо. Так, наприклад, можна спрогнозувати найвищий відсоток виконання навчальних завдань іншими особами на платній або безоплатній основі в навчальних закладах, які розташовані у великих містах і мають порівняно більший контингент студентів. Це пов'язано з тим, що комерційні структури відповідного профілю акцентують свою діяльність безпосередньо там, де більша кількість суб'єктів освітнього процесу й, відповідно, прогнозується більш високий попит на послуги, що надаються.

Таблиця 1

Анкета анонімного опитування студентів

Запитання	ТАК	НІ
1. Чи користувалися Ви під час виконання своїх навчальних завдань готовими матеріалами, запозиченими з ресурсів Інтернету?		
2. Чи доручали Ви виконання своїх навчальних завдань іншим особам на платних або безоплатних умовах?		
3. Чи користувалися Ви під час виконання своїх навчальних завдань матеріалами, запозиченими в товаришів старших курсів?		

Таблиця 2

Результати анкетування студентів

Запитання	Відсоток
1. Відсоток респондентів, які під час виконання своїх навчальних завдань користувалися готовими матеріалами, запозиченими з ресурсів Інтернету	73,3%
2. Відсоток респондентів, що доручали виконання своїх навчальних завдань іншим особам на платних або безоплатних умовах	31,6%
3. Відсоток респондентів, що користувалися під час виконання своїх навчальних завдань матеріалами, запозиченими в студентів старших курсів	96,2%

З огляду на викладене вище, можна дійти висновку, що для запобігання плагіату серед суб'єктів освітнього процесу необхідно використовувати спеціальні системи перевірки на унікальність (оригінальність) текстів. Практика використання таких систем повинна стати нормою під час подання студентських і наукових робіт на перевірку або до захисту.

Робота систем перевірки текстів на плагіат базується на зіставленні наданих текстів з наявними в базі даних. У процесі роботи системи здійснюється попередній збір і фільтрація інформації про наявні продукти інтелектуальної праці [4, с. 323]. Після

попередньої обробки тексти індексуються й вносяться до бази даних. Надалі ця база даних використовуватиметься для порівняння з наданими для аналізу текстами. Кожний документ, завантажений для перевірки, ставиться в чергу на обробку. Після перевірки необхідного документа система формує звіт, у якому докладно наводить усі результати пошуку та рівень унікальності перевіреного матеріалу. Як правило, ті фрагменти, що не є унікальними, для наочності виділяються в тексті [6, с. 276], а також надається інформація про джерела запозичення окремих фрагментів неунікальних текстів.

На сьогодні є велике розмаїття програмних продуктів, які дають змогу перевірити текст на унікальність та ідентифікувати ознаки плагіату. Кожен з них має специфічні особливості, а також переваги й недоліки [4, с. 324; 3, 4, с. 165; 5, с. 55; 6, с. 79; 7, с. 251]. Аналіз таких систем дав змогу також виявити основні функції, що властиві багатьом з них, а саме:

- визначення відсотка унікальності (оригінальності) наведеного тексту;
- пошук фрагментів тексту, що були запозичені, а також ідентифікація джерел використаної інформації;
- маркування запозичених фрагментів тексту кольором для наочності, а також використання різного кольору для запозичених фрагментів з різних джерел;
- формування докладних звітів після перевірки тексту на унікальність.

Програмне забезпечення також, як правило, має змогу гнучкого налаштування алгоритмів та інструментарію пошуку подібності (наприклад, визначення кількості слів у шинглі, кількість вибірок з тексту тощо).

До додаткових функцій програмного забезпечення перевірки текстів на унікальність (оригінальність) належать:

- перевірка контенту різноманітних web-ресурсів на унікальність з формуванням докладного звіту, що містить зафіксовані джерела плагіату;
- робота зі списком проксі;
- пакетна перевірка наданих текстових файлів;
- можливість повторної перевірки текстів після видалення неунікальних фрагментів;
- збереження історії перевірок і ведення логу подій під час функціонування алгоритму.

Наведене розмаїття функціональних можливостей систем перевірки на плагіат дає змогу сформулювати основне завдання – розробку комплексного підходу до організації

перевірки студентських і наукових робіт у вищих навчальних закладах для підвищення якості освітніх послуг і підготовки висококваліфікованих фахівців. Таке завдання повинно однаково ефективно допомагати протидіяти плагіату всіх трьох форм, про які йшлося вище.

Виконання завдання. Для забезпечення виконання поставленого завдання були проаналізовані джерела плагіату. Відповідно до даних табл. 2, система перевірки текстів на унікальність повинна насамперед давати змогу ефективно протидіяти внутрішньому плагіату, коли студент запозичує результати інтелектуальної праці попередників, що вивчали відповідну дисципліну в попередніх семестрах (у процесі анкетування ця форма отримала найбільший відсоток використання). Крім цього, необхідно також забезпечити ефективну протидію іншим двом формам плагіату.

Використання найпоширеніших систем перевірки на плагіат, що базуються на зіставленні текстів, наданих для аналізу, з інформацією в мережі Інтернет, дає змогу ефективно визначити плагіат, джерело якого – загальнодоступні мережеві ресурси та частково той, що пов'язаний з виконанням роботи на замовлення в сторонніх осіб (зазвичай, у разі великих обсягів виконання робіт, виконавці можуть використовувати запозичену інформацію з мережі Інтернет або розміщувати виконані роботи на web-ресурсах, що дає змогу достовірно виявити плагіат через перевірку текстів по базі інтернет-ресурсів).

На рис. 1 наведено схему джерел виникнення плагіату в студентських і наукових роботах закладів освіти. Зауважимо, що академічний плагіат має місце під час розробок викладачами навчально-методичних, інструктивних матеріалів, навчальних посібників, підручників, методичних вказівок тощо.



Рис. 1. Джерела виникнення плагіату в студентських і наукових роботах

Зі схеми на рис. 1 видно, що в разі виконання робіт сторонніми особами (цей варіант також відповідає запозиченню готових робіт у студентів-попередників) текст може містити ознаки плагіату як з ресурсів Інтернету, так і з інших джерел, у зв'язку з тим, що сторонні виконавці можуть самостійно використовувати ресурси Інтернету або інші джерела, а також поповнювати бази джерел плагіату новими матеріалами (наприклад, розміщувати зроблені роботи у відкритих джерелах Інтернету). Також суб'єкт навчання може самостійно запозичити матеріали з Інтернету або інших джерел, що часто призводить до повторного плагіату. З іншого боку, якщо виконані роботи не розміщуються у відкритих джерелах інформації, то системи перевірки текстів на унікальність, що використовують для порівняння матеріали з Інтернету, не зможуть виявити такий плагіат. Отже, найбільш поширені системи перевірки унікальності текстів не зможуть забезпечити достовірність отримання даних про те, міститься плагіат в роботах чи ні.

Ураховуючи той факт, що найбільш поширеною формою плагіату є запозичення матеріалів у студентів, що раніше вивчали дисципліну й виконували подібні завдання, пропонується для перевірки текстів використовувати базу даних, що містить виконані студентами завдання. Тим самим

здійснюватиметься не тільки перевірка факту запозичення матеріалу в студентів, які раніше виконували такі ж завдання, а й відповідно до схеми на рис. 1, – непряма перевірка наявності матеріалів, що були використані з інших джерел і ресурсів Інтернету (як суб'єктом навчання особисто, так і особами, що виконували завдання за дорученням суб'єкта навчання). Крім цього, такий спосіб перевірки доцільно комбінувати з прямим аналізом матеріалів за відкритими джерелами в мережі Інтернет.

Отже, для забезпечення достовірної перевірки на наявність плагіату, база даних системи антиплагіату у вищому навчальному закладі повинна постійно поповнюватися навчальними роботами студентів, у яких не було виявлено плагіату. Крім того, до цієї бази даних необхідно вносити навчально-методичні, інструктивні матеріали (конспекти лекцій, підручники, навчальні посібники тощо), оскільки вони можуть бути джерелом як студентського, так і повторного плагіату (коли викладачі використовують запозичені тексти під час підготовки своїх навчально-методичних та інструктивних розробок, а студенти, зі свого боку, повторно запозичують ці тексти, виконуючи власні студентські роботи). З огляду на це, пропонується така система організації перевірки навчальних і наукових робіт на плагіат (див. рис. 2):

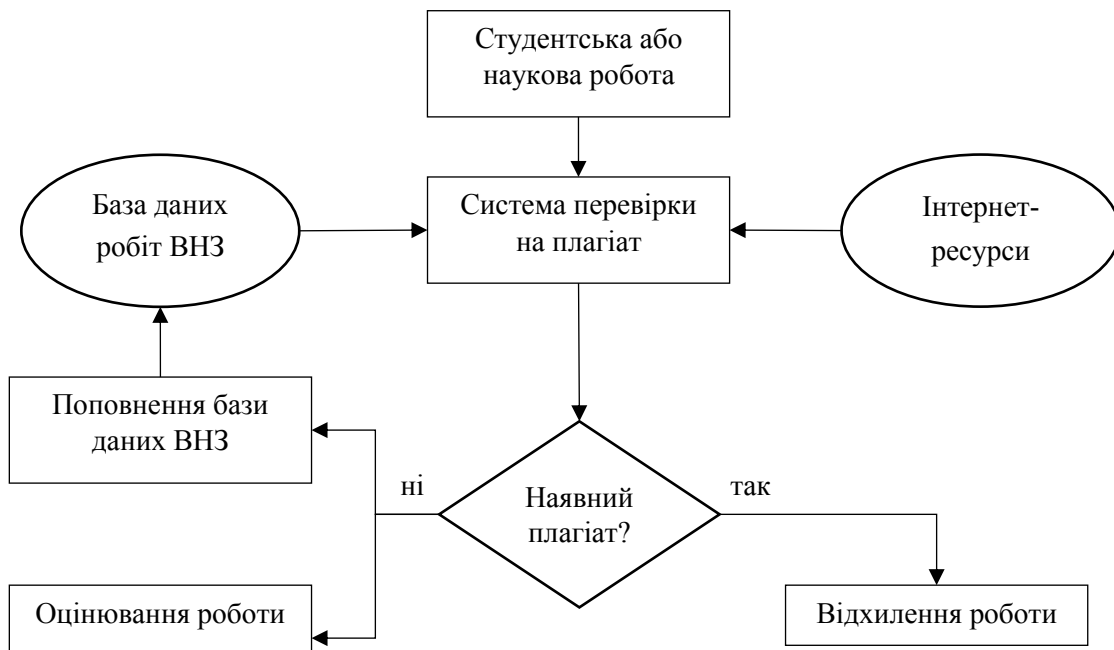


Рис. 2. Система організації перевірки студентських і наукових робіт на плагіат

У тому разі, коли за результатами перевірки тексту за базою даних робіт вищого навчального закладу (ВНЗ) й бази даних інтернет-ресурсів було виявлено плагіат – робота відхилялася від розгляду. Якщо плагіат не виявлено, то

роботу оцінюють і вносять до бази даних ВНЗ для використання надалі під час перевірки нових робіт. Для запобігання поданню робіт, що мають запозичені матеріали, які є результатами інтелектуальної праці студентів інших вищих

навчальних закладів, доцільно об'єднати бази даних різних ВНЗ для спільного використання під час аналізу. За таким же принципом пропонується проведення перевірки й подальше додавання до бази даних ВНЗ навчально-методичних та інструктивних розробок викладачів, що унеможливить виникнення повторного плагіату.

На особливу увагу заслуговує питання протидії штучному підвищенню унікальності текстів, що широко застосовується на практиці. Серед основних способів штучного підвищення унікальності слід назвати:

– перекладення запозиченого тексту іншою мовою;

– заміна символів, що мають однаковий вигляд у різних мовах (наприклад, заміна букв українського алфавіту: «а», «р», «Н», «о» та інших відповідними буквами латинського алфавіту).

Для виявлення впливу трансформації тексту на достовірність визначення відсотка унікальності проводився аналіз цілком

запозиченого тексту українською мовою та його перекладених варіантів. У табл. 3 наведено результати визначення відсотка унікальності українського тексту, який був цілком запозичений з мережі Інтернет, мовою оригіналу й у разі його автоматичного перекладу англійською та російською мовами. Для точності експерименту здійснювався онлайн переклад без подальшої корекції.

Дані табл. 3 свідчать, що переклад запозиченого тексту значно підвищує відсоток його унікальності. При цьому більший відсоток унікальності перекладеного англійського варіанта пояснюється тим, що англійський текст має більше розбіжностей з українським, ніж російський.

Для виключення цього негативного аспекту пропонується внесення перекладених варіантів текстів до бази даних системи перевірки на плагіат. Повторна перевірка цілком запозиченого тексту після вдосконалення бази даних дала результати, наведені в таблиці 4.

Таблиця 3

Відсоток унікальності (оригінальності) запозиченого українського тексту, перекладеного англійською та російською мовами

Варіант подання цілком запозиченого тексту	Відсоток унікальності
Текст українською мовою (оригінал)	0%
Текст, перекладений англійською мовою	65%
Текст, перекладений російською мовою	54%

Таблиця 4

Відсоток унікальності (оригінальності) запозиченого українського тексту після внесення до бази даних автоматично перекладених варіантів

Варіант подання цілком запозиченого тексту	Відсоток унікальності
Текст українською мовою (оригінал)	0%
Текст, перекладений англійською мовою	5%
Текст, перекладений російською мовою	3%

Презентований комплексний підхід до організації перевірки студентських і наукових робіт був апробований на базі науково-дослідної лабораторії кафедри інформаційних технологій та електронних систем ДЗ «Донбаський державний технічний університет». Для перевірки інформаційній системі були надані студентські роботи за спеціальностями 123 – «Комп'ютерна інженерія» та 171 – «Електроніка». Унаслідок експерименту було виявлено значне підвищення достовірності аналізу текстів на унікальність і стабільність цього показника для текстів різної спрямованості та форми подання. Основним недоліком такої системи є потреба в значно більшому обсязі пристроїв тривалого зберігання даних для розміщення багатоваріантних зразків текстів для порівняння.

Висновки. На основі вищенаведеного ми дійшли таких висновків.

1. Сучасний розвиток інформаційних технологій і сховищ даних з глобальним доступом стимулює розвиток і поширення методів і засобів запозичення чужих продуктів інтелектуальної праці.

2. Аналіз популярності форм плагіату у вищих навчальних закладах виявив досить високий відсоток студентів, які хоча б раз користувалися запозиченою інформацією. При цьому найбільш поширеною формою плагіату є отримання матеріалів від студентів, що раніше вивчали відповідну дисципліну й виконували подібні завдання.

3. На підставі проведених досліджень обґрунтовано варіант організації системи перевірки на плагіат для вищих навчальних закладів з використанням особистої бази даних

студентських і наукових робіт, яка постійно поповнюється новими матеріалами, що отримали великий відсоток унікальності. Для забезпечення більшого простору для пошуку плагіату доцільно об'єднувати бази даних вищих навчальних закладів для спільного використання під час аналізу.

4. З'ясовано, що трансформація тексту шляхом автоматичного перекладу іншими мовами веде до штучного підвищення відсотку унікальності. Для зменшення цієї помилки запропоновано доповнювати базу даних системи перевірки на плагіат автоматично перекладеними текстами без подальшого доопрацювання.

Список використаних джерел

1. Ліннік І. Програмне забезпечення для виявлення плагіату: практичний аспект [Електронний ресурс] / І. Ліннік // Науковий блог НаУ «Острозька Академія». – 2013. – Режим доступу: <http://naub.oa.edu.ua/2013/prohramne-zabezpechennya-dlya-vyavlennya-plahiatu-praktychnyj-aspekt>
2. Михайловський Ю. Б. Система Anti-Plagiarism як інструмент запобігання плагіату в навчальній та науковій діяльності / Ю. Б. Михайловський, Н. А. Длугунович // Вісник Хмельницького національного університету. Технічні науки. – 2013. – № 3. – С. 162–168.
3. Шарапова Е. В. Универсальная система проверки текстов на плагиат «Автор.net» / Е. В. Шарапова, Р. В. Шарапов // Информатика и её применения. – 2012. – № 3 (6). – С. 52–58.
4. Hariharan Sh. Automatic Plagiarism Detection Using Similarity Analysis [online] / Sh. Hariharan // The International Arab Journal of Information Technology. – 2012. – № 4 (9). – P. 322–326. – Available from: <http://www.ccis2k.org/iajit/PDF/vol.9,no.4/2796-4.pdf>
5. Kharat R. Semantically Detecting Plagiarism for Research Papers [online] / Kharat R., Chavan P. M., Jadhav V., Rakibe K. // International Journal of Engineering Research and Applications (IJERA). – 2013. – № 3 (3). – P. 77–80.
6. Singh R. Duplicity Detection System for Digital Documents [online] / R. Singh, C. Dutta // International Journal of Soft Computing and Engineering (IJSCE). – 2012. – № 5 (2). – P. 24–28. – Available from: http://www.iaeng.org/publication/IMECS2011/IMECS2011_pp272-277.pdf
7. Tschuggnall M. Detecting Plagiarism in Text Documents through Grammar-Analysis of Authors [online] / M. Tschuggnall, G. Specht // 15th GI-Symposium Database Systems for Business, Technology and Web, 11th March-15th March, 2013. – 2013. – P. 241–259. – Available from: <http://www.btw-2013.de/proceedings/Detecting%20Plagiarism%20in%20Text%20Documents%20through%20GrammarAnalysis%20of%20Authors.pdf>

Рецензент: Максимов О.С. – д.пед.н., професор

Відомості про авторів:

Меняйленко Олександр Сергійович
menyaylenko2@gmail.com

ДЗ «Луганський національний університет
ім. Тараса Шевченка»
пл. Гоголя, 1, м. Старобільськ, 92703, Україна

Захожай Олег Ігорович

zakhzhay_oleg@ukr.net
ДЗ «Донбаський державний технічний університет»
проспект Перемоги, 89, м. Лисичанськ,
93100, Україна

doi: <http://dx.doi.org/10.7905/nvmdpu.v0i18.1906>

Матеріал надійшов до редакції 11. 05. 2017 р.
Прийнято до друку 12. 06. 2017 р.

References

1. Linnik, I. (2013). *Software for plagiarism identification: practical aspect*. In : Scientific blog NaU "Ostrozka Akademiia". Retrieved from: <http://naub.oa.edu.ua/2013/prohramne-zabezpechennya-dlya-vyavlennya-plahiatu-praktychnyj-aspekt> [in Ukrainian]
2. Mykhailovsky, Yu. B., Dluhunovych, N. A. (2013). Anti-Plagiarism system as a plagiarism prevention tool in education and scientific activity. *Visnyk Khmelnyts'koho natsionalnoho universytetu*. Technical sciences, 3, 162–168. [in Ukrainian]
3. Sharapova, E. V., Sharapov, R. V. (2012). The universal system for plagiarism verification “Автор.net”. *Informatika i yeio primeneniia*, 3 (6), 52–58. [in Russian]
4. Hariharan, Sh. (2012). Automatic Plagiarism Detection Using Similarity Analysis. *The International Arab Journal of Information Technology*, 4 (9), 322–326. Retrieved from : <http://www.ccis2k.org/iajit/PDF/vol.9,no.4/2796-4.pdf> [in English]
5. Kharat, R., Chavan, P. M., Jadhav, V., Rakibe, K. (2013). Semantically Detecting Plagiarism for Research Papers. *International Journal of Engineering Research and Applications (IJERA)*, 3 (3), 077–080. [in English]
6. Singh, R., Dutta, C. (2012). Duplicity Detection System for Digital Documents. *International Journal of Soft Computing and Engineering (IJSCE)*, 5 (2), 24–28. Retrieved from http://www.iaeng.org/publication/IMECS2011/IMECS2011_pp272-277.pdf [in English]
7. Tschuggnall, M., Specht, M. (2013). Detecting Plagiarism in Text Documents through Grammar-Analysis of Authors. *15th GI-Symposium Database Systems for Business, Technology and Web, 11th March-15th March*, 241–259. Retrieved from <http://www.btw-2013.de/proceedings/Detecting%20Plagiarism%20in%20Text%20Documents%20through%20GrammarAnalysis%20of%20Authors.pdf> [in English]

Information about the authors:

Menyaylenko Oleksandr Serhiiovich
menyaylenko2@gmail.com

Luhansk Taras Shevchenko National University
1 Gogol square, Starobilsk, Ukraine

Zakhzhay Oleh Ihorovich

zakhzhay_oleg@ukr.net
Donbas State Technical University
89 Peremohy Ave., Lysychansk, Ukraine

doi: <http://dx.doi.org/10.7905/nvmdpu.v0i18.1906>

Received at the editorial office 11. 05. 2017.
Accepted for publishing 12. 06. 2017.