

УДК 004.42

ББК 32.97

Е. В. Шарапова

Аспирант

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ СИСТЕМЫ «АНТИПЛАГИАТ» ДЛЯ ОБНАРУЖЕНИЯ ЗАИМСТВОВАНИЙ

В работе проводится исследование возможностей системы «Антиплагиат» как средства обнаружения заимствований. Анализируются особенности работы системы, возможности обнаружения подходов к сокрытию заимствований. Исследования показали, что система «Антиплагиат» может использоваться для поверхностного анализа небольших текстов и не позволяет обнаруживать измененные формы заимствований. Система не обрабатывает множество источников сети Интернет, что существенно снижает эффективность ее работы.

Ключевые слова: антиплагиат, плагиат, заимствование, обнаружение плагиата, обнаружение заимствований, текст.

E. V. Sharapova

Postgraduate student

RESEARCH OF OPPORTUNITIES OF THE SYSTEM «ANTI-PLAGIARISM» FOR THE DETECTION OF BORROWING

The paper deals with the study of possibilities of system «Anti-plagiarism» detection of borrowing. Analyses the peculiarities of the operation of the system, the probability of finding approaches to the concealment of borrowing. Research has shown that the «Anti-plagiarism» can be used for surface analysis of small texts and does not allow detecting modified forms of borrowing. The system does not handle many sources of the Internet, which significantly reduces the effectiveness of its work.

Key words: anti-plagiarism, plagiarism, borrowing, the detection of plagiarism, detection borrowings, text.

Развитие компьютерных технологий и сети Интернет привело к появлению больших объемов информации, доступной всем желающим. Множество каталогов, библиотек и архивов содержит в себе массу информации, воспользоваться которой может любой человек. Это дало возможность использовать всю доступную информацию в различных отчетных работах, статьях, книгах и диссертациях, в том числе без каких либо изменений и указания первоисточников. Множество громких скандалов последних лет показало, что проблема обнаружения фактов заимствований (называемых часто плагиатом) стоит достаточно остро.

Наиболее известным решением в области обнаружения заимствований является система «Антиплагиат», функционирующая с 2005 года [1]. Система применяется множеством учебных заведений для обнаружения кусков

текстов, заимствованных из других источников.

Система «Антиплагиат» позволяет осуществлять поиск по большому количеству коллекций рефератов, контрольных работ, учебников и загруженных пользователями документов, хранящихся в собственной базе системы [2].

Несмотря на большую известность, способности системы эффективно обнаруживать факты заимствований остаются под вопросом.

Для оценки возможностей системы были составлены тестовые выборки, состоящие из текстов, взятых с различных сайтов сети Интернет. В результате анализа тестов в системе «Антиплагиат» было установлено, что система неплохо находит тексты, скопированные из рефератов и учебников. Тексты их энциклопедий и Википедии находятся не всегда. Информация с региональных, тематических и новостных сайтов практически не обнаруживается системой. А ведь большое число заимствование осуществляется именно из таких источников.

Такое поведение объясняется достаточно просто – система «Антиплагиат» осуществляет поиск только по собственной базе, формируемой определенным образом. Это позволяет существенно ускорить скорость поиска заимствований. Но при этом значительно снижает качество проверки.

Еще одним недостатком системы является ограничение размера проверяемого текста 3000 или 5000 символами (доступно после регистрации). Это объясняется желанием ускорить время работы алгоритмов поиска [3], которые существенно замедляются при увеличении размеров текста.

В связи с тем, что в учебных заведениях работы часто проверяют на наличие плагиата, все чаще стали применяться подходы к сокрытию фактов заимствования. Существует большое количество подобных подходов [4]:

1. Корректировка родов, чисел и времен входящих в текст слов. Часто замена ограничивается изменением слов «выполнил» на «выполнила» или «выполнили», «я» на «мы» и т.д.

2. Незначительное изменение заимствованного текста. Например, изменение по одному слову в предложении.

3. Сокращение заимствованного текста путем удаления слов, предложений, абзацев, рисунков, формул и т.д.

4. Перестановка частей текста, абзацев и предложений местами.

5. Обход систем проверки на плагиат путем замены русских букв на аналогичные по написанию английские буквы и т.д.

6. Замена знаков препинания: «.» на «,» и обратно, « » на «.» и т.д.

7. Замена пробелов на невидимые буквы (написанные, например, белым цветом).

8. Осуществление ручной или автоматической синонимизации текста и т.д.

Для исследования возможностей системы «Антиплагиат» был составлен ряд тестов. Для этого были взяты тексты, полностью обнаруживаемые системой как плагиат (100% соответствий) и последовательно модифицированы с использованием каждого из вышеописанных подходов [5]. Все тесты имели приблизительно одинаковый размер в 2000 символов и содержали в среднем по 400 слов. Было составлено по 10 тестов каждого вида.

В результате анализа тестов системой «Антиплагиат» было установлено:

1. Система «Антиплагиат» не смогла обнаружить корректировку родов, чисел и времен слов.

2. Незначительное изменение заимствованного текста обнаруживается не всегда (зависит от того, как близко располагаются изменяемые слова друг от друга).

3. Сокращение заимствованного текста путем удаления слов, предложений, абзацев система «Антиплагиат» способна обнаружить.

4. Перестановка частей текста, абзацев и предложений местами также обнаруживается системой.

5. Обход систем проверки на плагиат путем замены русских букв на аналогичные по написанию английские буквы обнаруживается в текущей версии системы.

6. Замена знаков препинания не влияет на работу системы (видимо, знаки препинания не учитываются).

7. Замена пробелов на невидимые буквы может быть обнаружена визуально при вставке текста в окно проверки. Сама система «Антиплагиат» обнаружить подобные замены не может.

8. Синонимизация текста системой не обнаруживается.

Таким образом, исследования показали, что система «Антиплагиат» не позволяет обнаруживать многие подходы к сокрытию фактов заимствований текстов. Наряду с отсутствием поддержки поиска по сети Интернет, это существенно снижает эффективность проверки текстов и в значительной мере снижает точность в обнаружении заимствований. По этой причине, несмотря на широкую известность, систему «Антиплагиат» следует применять только для поверхностного анализа текстов. Для детального

анализа текстов на предмет заимствований следует использовать другие системы [6] проверки текстов.

ЛИТЕРАТУРА

1. Антиплагиат [Электронный ресурс]. – Режим доступа: <http://www.antiplagiat.ru/> (дата обращения: 14.06.2013)
2. Шарапов Р.В., Шарапова Е.В. Система проверки текстов на заимствования из других источников // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: XIII Всероссийская научная конференция «RCDL'2011». Воронеж, 19-22 октября 2011 г.: труды конференции – Воронеж: Издательско-полиграфический центр Воронежского государственного университета, 2011, С. 233-238.
3. Зеленков Ю.Г., Сегалович И.В. Сравнительный анализ методов определения нечетких дубликатов для WEB-документов // Труды 9-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL'2007: сб. работ участников конкурса – Переславль-Залесский, 2007. – Т. 1. – С. 166-174.
4. Шарапов Р.В. Анализ подходов к обнаружению заимствованных текстов // Журнал «Современные наукоемкие технологии» – М: Российская академия естествознания, 2011 г. № 3, С. 47-49.
5. Шарапов Р.В., Шарапова Е.В. Исследование плагиата в работах студентов // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог 2012» (Бекасово, 30 мая – 3 июня 2012 г). Вып. 11 (18). Том 1 – М: Изд-во РГГУ, 2012 – с. 578-586.
6. Шарапов Р.В., Шарапова Е.В. Универсальная система проверки текстов на плагиат "Автор.NET" // Информатика и ее применения, 2012, Том 6, Выпуск 3, С. 52-58.

REFERENCES

1. *Antiplagiat*. Available at: <http://www.antiplagiat.ru/> (accessed 14 June 2013).
2. Sharapov R.V., Sharapova E.V. *Sistema proverki tekstov na zaimstvovaniia iz drugikh istochnikov* [System of verification texts on borrowing from other sources]. *Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kollektsii: XIII Vserossiiskaia nauchnaia konferentsiia «RCDL'2011»* [Electronic libraries: perspective methods and technologies, electronic collections of the XIII all-Russian scientific conference RCDL'2011]. Voronezh: Publishing center of the Voronezh state university, 2011, pp.233-238.
3. Zelenkov Iu.G. Segalovich I.V. *Sravnitel'nyi analiz metodov opredeleniia nechetkikh dublikatov dlia WEB-dokumentov* [Comparative analysis of methods of determination of fuzzy duplicate of a WEB document]. *Trudy 9-oi Vserossiiskoi nauchnoi konferentsii «Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kollektsii» RCDL'2007* [Proceedings of 9-th all-Russian scientific conference «digital libraries: advanced methods and technologies, digital collections» RCDL'2007]. Pereslavl-Zalessky, 2007, V.1. pp. 166-174.
4. Sharapov R.V. Analysis of approaches to the detection of borrowed texts. *Sovremennye*

- naukiemkie tekhnologii – Modern high technologies*, 2011, no.3, pp.47-49 (in Russian).
5. Sharapov R.V., Sharapova E.V. *Issledovanie plagiata v rabotakh studentov* [Research of plagiarism in the work of students]. *Komp'yuternaia lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoi mezhdunarodnoi konferentsii «Dialog 2012»* [Computational linguistics and intellectual technologies: materials of the annual international conference "Dialogue-2012"]. Moscow, 2012, pp. 578-586.
 6. Sharapov R.V., Sharapova E.V. Universal system of verification texts on plagiarism "Автор.NET". *Informatika i ee primeneniia – Informatics and applications*, 2012, Vol.6, no.3, pp. 52-58 (in Russian).
- .

Информация об авторе

Шарапова Екатерина Викторовна (Российская Федерация, г. Муром) – Аспирант. Муромский институт Владимирского государственного университета. E-mail: sharapovamivlgu@gmail.com

Information about the author

Sharapova Ekaterina Viktorovna (Russian Federation, Murom) – Postgraduate student. Murom institute of Vladimir state university. E-mail: sharapovamivlgu@gmail.com