

Editorial: The challenge of Dissernet

Nick Rushby^{1,2}

Scientific Editor *Education & Self Development*

¹Kazan federal university, Kazan, Russia

²Conation Technologies LTD, United Kingdom

nick.rushby@conation-technologies.co.uk

ORCID: 0000-0002-2379-1402

The impact of Dissernet

In March 2016 the *E&SD* editorial team started a project to secure inclusion in the Scopus and WoS databases for *Education & Self Development*. This has developed into a programme which also intends to turn the Journal into an international publication for researchers worldwide wanting to access Russian educational research, and for Russian scholars wanting an accessible window on educational research in other countries (Rushby, 2016). Part of this project was to ensure that the Journal conformed to the highest editorial ethical standards and this meant that we were at least partially prepared for the impact of the Dissernet report which identified that a number of articles published in the Journal showed evidence of plagiarism and raised questions of editorial malpractice.

Although these problems occurred before the start of the transformation process, it became clear that urgent action was needed to repair the reputation of the Journal and ensure that processes were in place to ensure that such malpractice could not take place in the future. This editorial describes the action plan, developed by the Editorial Team to address the problem.

The E&SD response

The detailed ethical guidelines for the editorial team, for reviewers and authors – and for Kazan Federal University which publishes the Journal were in place before the end of November 2016 (see www.eandsdjournal.org/ethical-policy/). These were closely based on the guidelines set out by the Committee on Publication Ethics – COPE (2017). In addition to the general ethical policy, subsets for editorial board members, reviewers and authors were included in the notes for guidance sent to all of these actors.

An immediate action was to accept the resignation of two members of the editorial team who had been implicated in editorial misconduct. The more challenging task was to deal with those articles where there was evidence of plagiarism.

It was agreed that the eleven articles (all published in 2015) should be checked for plagiarism and retracted if necessary. Then, as quickly as practical, the other articles published between 2014 and 2016 would all be checked for plagiarism. As we will see, this is not a straightforward operation. It is important that instances of suspected plagiarism are handled carefully and are based on evidence rather than supposition.

Notes on plagiarism detection software

Plagiarism checkers compare the text of the article under consideration against a corpus of published papers, books, conference proceedings, theses, etc, looking for overlaps –

Note: A version of this Editorial with the title *The challenge of Dissernet: a case study*, was presented at The 6th International Scientific and Practical Conference, World-Class Scientific Publication – 2017: Publishing Ethics, Peer-Review and Content Preparation, Moscow in April 2017.

that is, sequences of words in the article that appear in previously published works. This means that their effectiveness is highly dependent on a comprehensive and up-to-date corpus. That corpus is continually growing but it is only as good as the material it contains: very recently published works or papers in the course of publication, may not appear.

iThenticate gives the user the option of including or omitting the bibliography and of including or excluding declared quotes. It is quite probable that two articles on a particular topic will refer to the same set of earlier published works. Where an author has quoted directly from another work, the quote should be enclosed in punctuation marks (" " in English and « » in Russian) and acknowledged. The use of short quotes does not constitute academic misconduct. It is also helpful to omit short word sequences – typically sequences of up to 10 words – since these are likely to appear in any piece of writing. The effects on the iThenticate similarity index can be significant. In one example (Article 7) the score when the bibliography and quotes were included was 21%: this was reduced to 11% when the bibliography and quotes were omitted. In this study, all of the iThenticate analyses were carried out with the omission of the bibliography, acknowledged quotes, and sequences of 10 words or less.

Which software?

Education & Self Development, like other journals publishing work in two or more languages needs to be able to carry out checks in different languages and therefore needs several software systems. To ensure consistency, there should also be parity in the thresholds applied to the different systems so that there is a consistency in what is – and is not – acceptable.

To this end, we examined three software systems – iThenticate, РУКОТекст and the Антиплагиат software.

iThenticate

Perhaps the most widely used plagiarism checking software for English language manuscripts is iThenticate. This claims to have the largest database of content including:

- 50 million journal articles, conference proceedings and books from over 800 leading scientific, technical and medical (STM) publishers,
- 110 million online and offline subscription content and research titles from 30 leading aggregators, databases and content providers, and
- 60 billion texts from the internet stretching back nearly a decade with 10 million web pages being indexed every day. (Turnitin, 2016)

iThenticate matches the texts of manuscripts against those texts in the same language in its database. The languages include: Chinese (simplified and traditional), Japanese, Thai, Korean, Catalan, Croatian, Czech, Danish, Dutch, Finnish, French, German, Hungarian, Italian, Norwegian (Bokmal, Nynorsk), Polish, Portuguese, Romanian, Serbian, Slovak, Slovenian, Spanish, Swedish, Arabic, Greek, Hebrew, Farsi, Russian, and Turkish. Crucially, it looks for matches within the same languages and so it will not recognise plagiarised passages that have been taken from a translation – unless that translation itself is available online and has been identified for inclusion in the database. Its effectiveness for detecting plagiarism in Russian texts is therefore determined by the number of Russian language sources in its database (see Turnitin, 2016).

РУКОТекст

Kazan Federal University has a licence to use РУКОТекст which uses an algorithm to search for semantic similarities in the content of the submitted text. The database contains over 150 million documents and is continually being increased.

РУКОНТекст identifies the percentage overlap (borrowings) from each of the sources that have been identified and calculates an overall figure which appears to be equivalent to the similarity index SI provided by iThenticate (see Rukontekst, 2017)

Антиплагиат

The Antiplagiат includes indexed online web pages, the collection of full texts and abstracts of dissertations from the Russian State Library, the Lexpro collection of texts of legal and regulatory documents and a collection of full text articles in the Elibrary.ru scientific electronic library.

The tool gives a report on the extent of the 'borrowing' – plagiarism with a ranked list of detected sources. There is a full text of document under scrutiny, which highlights plagiarised sections of text. The full report make it possible to examine and classify each plagiarised section (see Antiplagiат, 2017)

At the time of writing, KFU did not have independent access to Антиплагиат and was reliant on the reports provided by Dissernet.

The question

The need for parity raises the question of whether the results from the available systems are comparable for Russian language articles. Specifically:

Is there a correlation between the overlap scores calculated by iThenticate, РУКОНТекст and Антиплагиат for articles written in:

- a. Russian and
- b. English

It is up to the users of these tools to decide where the threshold scores should be set. The results of this study will enable us to make a informed decisions as to whether:

- i. Whether we can a single system to check **all** submissions and
- ii. Where the thresholds should be set to achieve parity

Methodology

Ten articles, five written in Russian and five written in English, published in *Education & Self Development* were selected at random from recent issues of the Journal. Using recent articles made it easier to obtain the MS Word version, rather than the pdf file used in publication.

The Russian manuscripts (titled Articles 1 through 5) were edited to omit the English language components (the English language title, abstract and keywords). Figures and tables were not omitted although these are not generally checked by the software.

Similarly, the English language articles (titled Articles 6 through 10) were edited to omit the Russian language components (the Russian language title, abstract and keywords).

All ten articles were then checked using iThenticate and РУКОНТекст. At the time, we did not have access to Антиплагиат. The resulting overlap scores are shown in Table 1. For all the iThenticate analyses, the bibliography, quotes and sources of less than 10 words were omitted.

Findings

The similarity index (SI) calculated by iThenticate and the percentage of Borrowings calculated by РУКОНТекст are shown in Table 1.

As might be expected, iThenticate yielded higher scores than РУКОНТекст for the English language articles while for the Russian language papers the results were reversed. This almost certainly reflects the content of the text databases used for comparison.

The ПУКОHTекст database is predominantly Russian while the Russian content of iThenticate is much smaller than the English content. This was confirmed by a detailed examination of the sources identified by each system. For a given article, iThenticate and ПУКОHTекст identify different sources.

Table 1. *Overlap scores for ten test articles*

Article number	Language	Length (words)	Overlap scores	
			iThenticate	ПУКОHTекст
1	Russian	2282	10%	30%
2		2775	12%	39%
3		4236	20%	39%
4		6473	8%	27%
5		6456	18%	39%
6	English	4141	1%	1%
7		7404	11%	0%
8		3762	5%	0%
9		3002	2%	3%
10		3566	2%	2%

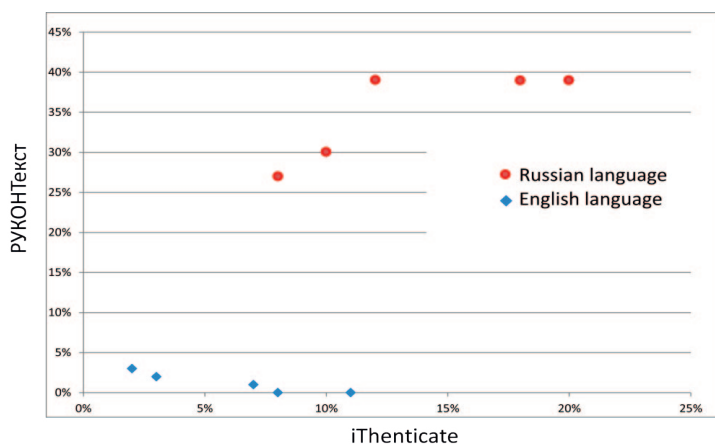


Fig. 1. *ПУКОHTекст and iThenticate scores for 10 example articles*

However, careful checking of the results for each test article identified other issues with ПУКОHTекст.

In one example, ПУКОHTекст identified a source (<http://studopedia.ru/>) that contributed over 38% of the test article. The publication dates of this source and the test articles were both 2015. However, there was compelling evidence that the test article had been written in early 2014. No precise publication date for the Studopedia article was given and there were no authorship details. It is at least possible that the Studopedia content was plagiarised from the test article without the author's consent, and that ПУКОHTекст then wrongly identified the plagiarist and victim.

In a second example of an article identified by Dissernet as having a high level of 'borrowing,' ПУКОHTекст failed to identify the source text and concluded that the article was free of plagiarism.

A similar problem was found in one example text using iThenticate, where the publication date was only given for one of the texts being compared. iThenticate made incorrect assumptions about the order of publication.

In the third example ПУКОHTекст identified a number of sources and gave the percentage 'borrowing' from each. The total of all these 'borrowings' was 292.6%! This suggests that there is a very high incidence of internal overlap between the various texts in the corpus used for comparison. In other words, it seems quite possible that a relatively short passage in the text under consideration might appear in many different corpus texts – possibly through unauthorised copying – and thus yield an inaccurately high overall score.

Support for this hypothesis comes from the detailed analysis of one of the Russian language articles. In this case, ПУКОHTекст identified 302 sources with a combined overlap that was far greater than 100%. Two of the significant overlaps were from the same source and it is unclear why ПУКОHTекст should list them separately.

When the same article was analysed using iThenticate, the software identified two sources contributing 3% (101 words) each. However, each of these sources was replicated up to 17 times in the database. While iThenticate recognises and can resolve this, it is possible that ПУКОHTекст does not.

It is not clear, in the light of this evidence that the results from ПУКОHTекст are reliable. Both iThenticate and Антиплагиат provide the facility for a line-by-line visual comparison of the source with the text being checked and enable the editor to confirm (or refute) the suggestion of plagiarism.

Although this study is drawing conclusions from a sample size that is too small to be statistically significant it does indicate that:

a. iThenticate works well for English language articles and less well for Russian language articles.

b. It is essential to carry out a visual check to ensure that the plagiarism software has not become confused.

The Dissernet articles

Dissernet identified eleven articles where they found evidence of 'borrowing.' We have also checked all of the English language articles published in the Journal since 2014 and all of those that are currently under consideration.

The Committee on Publication Ethics which is the internationally recognised authority on publication ethics, offers a series of guidelines and flowcharts for editors dealing with cases of academic misconduct such as plagiarism and redundant publication (COPE, 2017). *Education & Self Development* has followed the guidelines for dealing with plagiarism in published articles carefully (see figure 2).

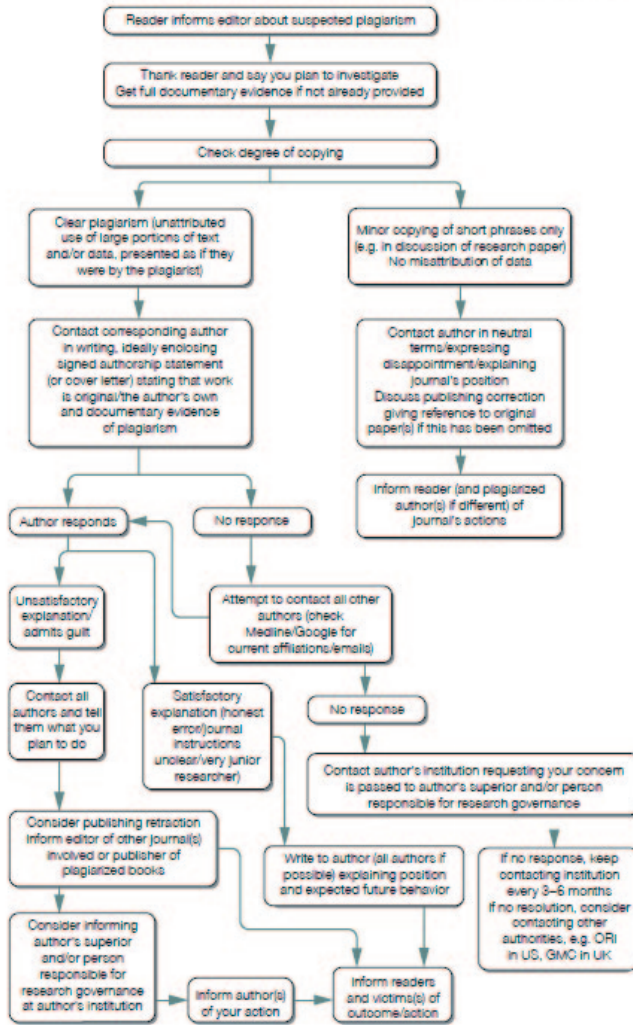
It is important to recognise that plagiarism checkers are not infallible. They can highlight articles where there is a strong suspicion of plagiarism or text re-cycling, but where there is a concern, the article should be desk-checked with the sources identified by the plagiarism checkers. This is time-consuming but necessary if editors are not to make unsubstantiated accusations of academic misconduct.

This detailed investigation confirmed that eight of the articles were heavily plagiarised. The authors were contacted and asked whether they could offer any explanation.

While some of these authors accepted the decision to retract, several challenged the decision on the grounds that, in Russia, text recycling is not considered to be plagiarism and is therefore acceptable. This cultural view of publication ethics is discussed by Rushby in the paper «Publication ethics – moral principles and cultural dissonance». While recognising that the ethics of publication vary between countries, he concludes that



What to do if you suspect plagiarism (b) Suspected plagiarism in a published manuscript



Note
The instructions to authors should include a definition of plagiarism and state the journal's policy on it

Developed for COPE by Liz Wager of Sideview (www.lizwager.com) © 2013 Committee on Publication Ethics First published 2006

A non-exclusive licence to reproduce these flowcharts may be applied for by writing to: cope_administrator@publicationethics.org

Fig. 2. COPE guidelines for suspected plagiarism in a published article (Reproduced under Creative Commons AttributionNonCommercial-NoDerivs license, <http://publicationethics.org/>)

«Academic publishing is becoming increasingly international and so, if the system is to work then authors and journals have to come to a common understanding of what is acceptable in academic publishing» (Rushby, 2017).

In seven cases, the evidence pointed towards self-plagiarism or text recycling where the publication dates of the suspect article and the source were very close. These appeared to be cases of simultaneous submission. The *Education & Self Development* Notes for Authors (<http://en.eandsjournal.org/for-contributors/>) now require authors to «Confirm that their contribution is original and that it has neither been published previously nor is currently being considered for publication elsewhere.» Authors must confirm this in their covering letter to the Editor-in-Chief when they make the submission. At the time, this was not a requirement and it is not clear that the source journals made this explicit in their notes for authors.

Thus, for these seven articles we decided to agree on the text of a 'statement of redundant publication' which draws attention to the fact that a similar version of the article was published elsewhere at around the same time.

The Journal will take similar action for any other articles that have been published and where there is verifiable evidence of plagiarism or redundant publication.

The outcomes

Of the eleven articles identified by Dissernet, we have retracted four where the visual check demonstrates unacceptable plagiarism. The overlaps in the remaining seven articles have arisen from simultaneous submission (redundant publication) where, as far as we can determine, neither journal required the authors to «Confirm that their contribution is original and that it has neither been published previously nor is currently being considered for publication elsewhere.» In these case we have issued a statement of redundant publication.

The notices appear in the Journal after this editorial.

Our investigations into the other articles published in the past three years (about 400 articles) are ongoing. We have checked the 18 articles written in English and have retracted one of them on the grounds of plagiarism.

We have also checked all of the English language articles under consideration and, as a result have declined to continue with one and asked another author for significant revisions to reduce the amount of unattributed borrowing to an acceptable level.

All the procedures have followed the appropriate COPE guidelines (COPE, 2017).

While the percentage of plagiarised articles may seem alarming it compares well with the experience of other journals using plagiarism checking tools. For example an analysis of plagiarism in articles submitted to the highly respected *British Journal of Educational Technology* carried out in 2013, found that over 11% of submissions returned an iThenticate score of over 40%!

under 30%	275	72.94%
30-39%	60	15.92%
40-49%	18	4.77%
over 50%	24	6.37%
Total	377	

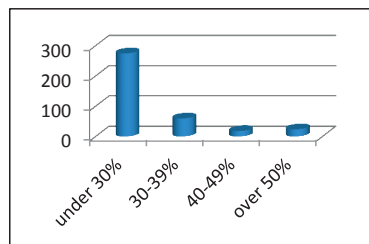


Fig. 3. Similarity index (SI) scores for submissions to BJET (2013)

Think of the editor

Dealing with academic misconduct is not a pleasant task. It takes a great deal of time (perhaps 3-4 hours for each article) and is profoundly depressing. Few editors enjoy having to tell authors that their articles must be retracted with all the consequences that this may have on their subsequent academic career. While we may have little sympathy for the serial plagiarist, the majority of cases involve authors who are inexperienced, unaware of publication ethics, under pressure to publish, or confused by journals failing to deal with submissions in a timely manner. Research supervisors need to do more to educate their students and to check their work **before** it is submitted to a journal.

Acknowledgements

The author thanks Laysan Kayumova for her assistance in checking the sample articles with РYКОHTекст, Olga Kirillova for her encouragement and guidance and Daria Khanolainen for translating the article into Russian.

References

- Antiplagiat* (2017) *About the system*. Accessed March 1, 2017 at: <https://www.antiplagiat.ru/Page/About>
- Apollodorus (1921). *Bibliotheca (The Library) (in Greek)*. With an English Translation by Sir James George Frazer, in 2 Volumes. Cambridge, MA; London: Harvard University Press; William Heinemann Ltd.
- COPE (2017) *Guidelines*. Committee on Publication Ethics. Accessed February 28, 2017 at: <http://publicationethics.org/resources/guidelines>.
- Rukontekst (2017) *Search for plagiarism and document analysis Rukontekst*. Accessed March 1, 2017 at: <https://text.rucont.ru/>
- Rushby, N.J. (2016) *Transforming a journal from an institutional focus to an international publication: a work in progress!* Paper at the Project 5-100 Seminar, Moscow, 11th November 2016.
- Rushby, N.J. (2017) *Publication ethics – moral principles and cultural dissonance*. *EON, Editorial Office News*. (June) pp4-9. <http://dx.doi.org/10.18243/eon/2017.10.5.3>
- Turnitin (2016) *Largest scholarly comparison database*. Oakland, CA. USA: Turnitin LLP. Accessed February 14, 2017: <http://www.ithenticate.com/content>