How to cheat the page limit

Citation for published version:
Duivesteijn, W, Hess, S & Du, X 2020, 'How to cheat the page limit', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, pp. e1361. https://doi.org/10.1002/widm.1361

Digital Object Identifier (DOI):
10.1002/widm.1361

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery

Received: 7 November 2019 | Revised: 16 January 2020 | Accepted: 17 January 2020

DOI: 10.1002/widm.1361

PERSPECTIVE

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

WILEY

How to cheat the page limit

Wouter Duivesteijn[1,2] | Sibylle Hess[1,2] | Xin Du[1,2]

[1]Proceedings Chair of ECMLPKDD 2019, Würzburg, Germany

[2]Data Mining Group, Technische Universiteit Eindhoven, Eindhoven, the Netherlands

Correspondence
Wouter Duivesteijn, Proceedings Chair of ECMLPKDD 2019, Würzburg, Germany.
Email: w.duivesteijn@tue.nl

Abstract

Every conference imposing a limit on the length of submissions must deal with the problem of page limit cheating: authors tweaking the parameters of the game such that they can squeeze more content into their paper. We claim that this problem is endemic, although we lack the data to formally prove this. Instead, this paper provides a far from exhaustive summary of ways to cheat the page limit, a case study involving the papers accepted for the Research and Applied Data Science tracks at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD) 2019, and a discussion of ways for program chairs to tackle this problem. Of the 130 accepted papers in these two ECMLPKDD 2019 tracks, 68 satisfied the page limit; 62 (47.7%) turned out to spill over the page limit, by up to as much as 50%. To misappropriate a phrase from Darrell Huff's "How to Lie with Statistics," we intend for this paper not to be a manual for swindlers; instead, nefarious paper authors already know these tricks, and honest program chairs must learn them in self-defense.

This article is categorized under:
Commercial, Legal, and Ethical Issues > Fairness in Data Mining

KEYWORDS

conference organization, fairness, LaTeX, reviewing process, TeXnical Desk Reject Phase

1 | HOW TO CHEAT THE PAGE LIMIT: AUTHORS' VIEWPOINT

So you are a scientist, you have a good idea, and hence you are preparing a paper to submit to a big international conference. For instance, you are a data miner, and you have set your sights on ECMLPKDD 2019 (European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2019). In many scientific disciplines, this compels you to prepare your manuscript in LaTeX, and the conference will have outlined a set of rules along which you are supposed to do so. For instance, ECMLPKDD publishes their accepted papers in Springer's *Lecture Notes in Computer Science*, and directs you for more information to the corresponding web page (Springer, 2019b). The conference website informs you that within this format, the maximum length of papers is 16 pages, including references. Armed with this knowledge, you download a copy of the Guidelines for Proceedings Authors (Springer, 2019a), and you commence writing!

Suppose that, once you emerge from your burst of creative writing and once the dust settles, you run into a problem: the paper you produced is actually 18 pages long instead of the allowed 16. This is a bummer. At this point, the appropriate path to take, the high road, is to reformulate your thoughts and arguments in a more concise manner. Remove sentences, reformulate arguments such that they take up less space, consider whether this train of thought is really necessary for the overall story line. There are two problems with this solution. On the one hand, it is a lot of work. On the other hand, if you try to convey your thoughts while being more economic with the number of words you use, you run the risk of making your paper more difficult to read. This would put you at a disadvantage, compared to those authors who use nefarious tricks to squeeze their 18-page paper into 16 pages. Fear not! You too can learn these tricks (summarized in Table 1)!

1.1 | Load alternative fonts

Springer prescribes the use of LaTeX's standard Computer Modern font (Springer, 2019a, section 2.3). This font eats loads of space. If one were to load the Times font, with \usepackage{times}, this will easily eat a page out of your paper.

For advanced users, one might want to load the packages helvet and courier along with times, so that you have the plausible deniability that you are really only doing this for aesthetic reasons. Also, if you fear that the times package is too obvious, you might instead want to use either \usepackage{newtxtext,newtxmath} or \usepackage{stix}.

A drawback of this method is that loading alternative fonts is really obvious: a reviewer might actually catch you doing this, and a small subset of those reviewers might elect to reject your paper over it. This risk may be too much for you to take.

1.2 | Reduce font size

Springer prescribes that all plain text is written in 10 pt format, except the abstract and captions (both of which may be written in 9 pt format) and section headers (which are bigger) (Springer, 2019a, abstract and section 2.1). To get around these limitations, many font size altering LaTeX commands are at your disposal: \small, \smaller, \tiny, \footnotesize, \scriptsize, and \fontsize{7pt}{7.5pt} are some examples. If you put one of those in a mathematical equation, small gains are made; if you put a \scriptsize inside a table, you can gain significant space!

For advanced users, one might want to employ a \scriptsize inside the \lstset settings of the listings package. This is used to control the appearance of code in your paper, and the font size change is not that obvious even to the most careful of proceedings chairs. Alternatively, one could pass the font size changing commands as options to the caption package (where \small would be allowed, but none of the others), or use them to change the appearance of algorithm environments by changing \SetAlFnt or \SetCommentSty.

TABLE 1 Condensed summary of main offending commands

Fonts	Vertical spaces	Figures, tables, margins
\usepackage{times}	\vspace{−1cm}	\textfloatsep
\usepackage{helvet}	\vskip −1cm	\floatsep
\usepackage{courier}	\\[−3mm]	\intextsep
\usepackage{newtxtext}	\abovedisplayskip	\usepackage{wrapfig}
\usepackage{newtxmath}	\belowdisplayskip	\resizebox
\usepackage{stix}	\algomargin	\scalebox
\small	\bibsep	\adjustbox
\smaller	\bibpreamble	\arraystretch
\tiny	\itemsep	\tabcolsep
\footnotesize	\linespread{0.9}	\usepackage[margin=1in]{geometry}
\scriptsize	\baselinestretch	\usepackage{fullpage}
\fontsize{7pt}{7.5pt}	\enlargethispage{1cm}	\usepackage{a4wide}

Note: Tricks for advanced users, working only in specific contexts, are left out of this table; their explanation requires plain text paragraphs, so they are represented best in nontabular form.

1.3 | Reduce vertical spacing…

Springer does not allow authors to insert or remove vertical spacing (Springer, 2019a, section 4.1). However, if you are not too aggressive with it, it is unlikely that reviewers will catch you doing so. Lots of whitespace is left around section headers in the standard Springer style, which you might want to reduce by \vspace{−1cm} or somesuch. If you are really old school, you might want to use the plan TeX command \vskip −1cm. If you need to reduce the effect of a single newline, a \\[−3mm] might just give you the subtle edge you need without being too obvious.

1.3.1 | …around equations

If you feel that the mathematical equations get too much whitespace, you can set some flexible spaces to narrower amounts. With \abovedisplayskip you can control the amount of whitespace before an equation and \belowdisplayskip controls the whitespace after an equation.

1.3.2 | …around algorithms

To reduce whitespace around algorithms, set \algomargin to a smaller value.

1.3.3 | …in and around the bibliography

To reduce whitespace between entries of the bibliography, one can set \bibsep to a smaller length. Additionally, for the true hackers, the command \bibpreamble is meant to print some text at the start of the bibliography; you can abuse it to provide some additional negative space commands.

1.3.4 | …between items

If one uses itemize, enumerate, or description environments, one can reduce the spacing between individual items with the \itemsep length.

1.3.5 | …throughout the document

For advanced users, you might not want to keep typing negative vspaces everywhere in the document. There is a neat trick that allows you to reduce vertical spacing between all lines in one single command: just set \linespread{0.9}. This does change the physical appearance of the document quite a bit, so it is a bit risky. Renewing the \baselinestretch command to a lower factor has a similar effect.

1.4 | Reduce spacing around figures

Many academic papers contain figures, not entirely dissimilar from Figure 1. Springer guidelines leave authors relatively free in designing the contents of figures. However, the standard style does prescribe the spacing of these figures, and it eats quite a bit of space. We could tackle this with negative \vspaces, but subtler ways are available. You can control the distance between floats (figures, tables, algorithms, …) on the top (or bottom) of the page and the text by setting the length \textfloatsep. If several floats follow one another, you can control the distance between them with \floatsep. If floats are inserted inside the page text (using the h placement option), the distance to the surrounding text is controlled with \intextsep.

If you have a figure as narrow as Figure 1, you might be annoyed by the waste of horizontal space. If you have multiple figures, of course, you can combine them into one using the relevant LaTeX packages, but this document only has

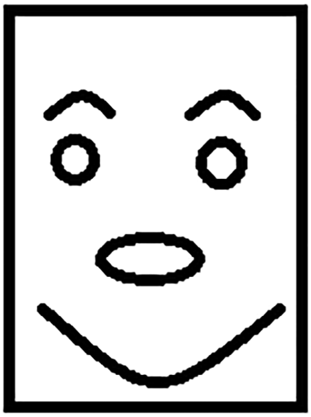

FIGURE 1 This portrait is fictitious. No identification with actual persons (living or deceased), especially any ECMLPKDD 2019 chairs, is intended or should be inferred

a single version. The standard LaTeX solution for this is to employ the wrapfig package. This package enables the figure to be on either the left or the right side of the page, while the normal text flows on alongside. The Springer guidelines are unclear on whether or not this is acceptable. One could see the use of wrapfig as a temporary edit to the page margin, which surely is not allowed, but this is a matter of opinion. If you use wrapfig, and use the vanilla wrapfigure environment, you will end up with comically large vertical spacing surrounding your figure, and you might be tempted to counter those with negative vspaces. While the Springer guidelines forbid inserting or removing vertical spacing, this particular spacing is inserted by the wrapfig package, which makes removing the spacing you just inserted a knotty argument to disentangle. The question whether wrapfig is compatible with the Springer guidelines is unclear at the time of writing; it would be prudent for Springer to provide clarity in this matter.

1.5 | Reduce table space

You can reduce the spacing *around* tables in the same way as the spacing around figures, as discussed in the previous section. However, there are additional cheats involving tables, which would not be considered cheats when they would involve figures. The core point is that tables are supposed to display plain text (typically, numbers), for which Springer prescribes a standard font size (10 pt). Any reduction of this font size, be it implicit or explicit, violates the Springer guidelines.

If you are a data miner, chances are that you want to display the dominance of your method over the competition. The more competitors you beat on the more datasets, the better! Hence, many data mining papers have at least one quite large result table. If your table is not entirely square, you have the choice of mode. A table in portrait mode is more likely to fit within the margins, but it eats a lot of vertical space. A table in landscape mode is more economical in space, but it is more likely to walk outsize of a page margin. If this happens, you can reduce the horizontal size of your table with \resizebox, \scalebox, or \adjustbox. Collateral benefit is that this often also reduces the vertical size of your table, which gives you more lines of text to write your arguments. The implicit reduction of table content font size violates the Springer guidelines.[1]

If you wish to reduce the row height of tables, you can set \arraystretch to a lower factor. Since this inserts or removes vertical spacing, this violates Springer's guidelines. If you wish to reduce the space between table columns, you can reduce the setting of \tabcolsep. Since this does not directly affect vertical spacing, its compliance with Springer's guidelines is unclear; proceed with caution.

For advanced users, you can get around all these hacks by creating your table in an external piece of software (Excel, or Paint if you want to be really creative), turning it into a figure file (png, pdf, jpg, ...), and then importing that file as a figure. This makes it prohibitively time-consuming for the proceedings chairs to recreate your table as a standard LaTeX table, while also making it impossible for the proceedings chairs to check the font size in your pictured table.

1.6 | Reduce margins

The Springer standard style leaves huge white margins around the edges of the paper. If you want to make more economical use of the pages, you might want to reduce those margins. If you want to be really brazen about this, use

\usepackage[margin=1in]{geometry}. Slightly less nuclear options are \usepackage{fullpage} and \usepackage{a4wide}. Only use this if preceding cheats in this section do not deliver the desired results! Changing the margins is really visible in the final papers, and if a reviewer catches it, this is sufficient grounds to desk reject your paper.

For advanced users, if you only need to steal the bottom margin on a single page, you could employ \enlargethispage{1cm}. A similar effect can be achieved by collecting multiple figures that together would not fit within the margins of a single page, and put them into a single float environment.

2 | CASE STUDY: THE 130 PAPERS ACCEPTED AT THE RESEARCH AND APPLIED DATA SCIENCE TRACKS OF ECMLPKDD 2019

We have no hard data across multiple conferences or venues to say anything substantial about how widespread this phenomenon is. However, in our role as Proceedings Chairs at ECMLPKDD 2019 (European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2019), we have acquired the material to analyze the extent of space cheating in a single edition of a single conference. At this conference, 130 papers were accepted for publication in the Research and Applied Data Science tracks. All authors of these 130 papers were required to send us their LaTeX sources, so that we could prepare the pre-proceedings versions of their paper, including the Springer-style headers and such. With those LaTeX sources, we did the following.

First, we compiled the sources as the authors delivered them to us. This gives us the page length of the papers as the authors intended, on our system. Subsequently, we removed all commands listed as space cheating in the preceding sections of this paper. This process gives us the page length of the papers, in the form compliant with Springer's guidelines (Springer, 2019a).

2.1 | Results

With the sole exception of \enlargethispage, all forms of space cheating mentioned so far appeared in at least one of the 130 papers.

Paper length histograms are given in Table 2, both in the form as originally submitted by the authors, as in the reformatted form. In Table 3, we summarize for both versions how many papers complied with the limit.

The first thing to notice is that even in the originally submitted version, 10 of the 130 papers do not comply with the page limit (cf. Table 3). This may be an artifact of people using old versions of the LNCS LaTeX class and the associated bibliography style. It may also result from small discrepancies in LaTeX distributions across machines and operating systems. These 10 may deserve the benefit of the doubt.

The second thing to notice is that with all space cheating removed, 62 out of 130 papers are over the page limit (cf. Table 3). A whopping 47.7% of papers are overlong. There is a huge variety in the degree to which papers are overlong. The vast majority (49 out of 62) of these papers roll over onto page 17 (cf. Table 2) and that is it. Some of

TABLE 2 Histogram of paper lengths in pages (partially) used, across the 130 papers accepted for the Research and Applied Data Science tracks at ECMLPKDD 2019

Number of pages	12	13	14	15	16	17	18	19	20	21	22	23	24
Submitted version	1	0	2	5	112	10	0	0	0	0	0	0	0
Reformatted version	1	0	1	5	61	49	8	3	1	0	0	0	1

Note: Notice that the page limit is 16 pages. The second row contains the histogram of the papers compiled as they were originally submitted; the third row contains the histogram of the papers recompiled after all space cheating commands were removed.

TABLE 3 Condensed form of Table 2, retaining only the information whether papers were over or under the limit

Complied with page limit?	Yes	No
Submitted version	120	10
Reformatted version	68	62

these papers have only a single reference spill over on page 17, making the space cheating rather benign, while other make use of the entire 17th page. It is clear, though, that this form of space cheating is at another level than the paper that manages to squeeze a 24-page paper (when complying with the guidelines) into a 15-page submission (when employing tricks).

Finally, we would like to remark that of the papers remaining on the correct side of the page limit in their reformatted version, not all are innocent either. Only 31 of the 130 papers did not contain any of the mentioned space cheats. Two additional papers had a single negative \vspace; all others had more serious space cheats. However, for 35 papers, the space cheating did not result in a change in the number of pages the paper used, which is why we do end up with 68 papers complying with the page limit in reformatted form.

3 | HOW TO TACKLE THE PROBLEM

In this digital day and age, you might wonder why we even bother with page limits to begin with. Would not it be a better idea to just let authors write as much as they want, and relegate this discussion to the dustbin of history along with the concept of an actual physical proceedings book? In smaller research fields, this may be a very reasonable solution to the situation sketched so far in this paper. The problem[2] in data mining is that the field is currently going through such a boom, that reviewers are severely overstretched. Collectively, we can barely handle the reviewing load as it is, and if we would not keep papers that are to be reviewed relatively short, we would move the reviewing load one order of magnitude even further away from sustainable. Hence, paper lengths must remain upper bounded.

If reviewer load were the only consideration, one could imagine replacing the page limit with another proxy for paper length, such as the word count. If one were to limit the number of words used in a paper, LaTeX space cheating commands would become irrelevant, so the form of cheating discussed in this paper will be resolved. The benefit is clear, but there are two drawbacks to this solution. On the one hand, a page limit per paper provides a reasonable estimate of an upper bounded of the total number of pages in the complete conference proceedings. For instance, ECMLPKDD publishes its proceedings with Springer. With the current number of accepted papers, and if those papers adhere to the stated page limit, the ECMLPKDD organization knows for a fact that the proceedings will fit within the upper page limit of a three-volume Springer proceedings. If the page limit is replaced by a word limit, the number of pages can vary wildly from paper to paper, and the total number of pages in the conference proceedings may spill over into a fourth volume. ECMLPKDD is charged per volume, so this forms a financial risk for the conference organization. On the other hand, a word count is not straightforward. We must make very clear to the authors exactly how the word count of a paper is computed. Do mathematical formulas count? If so, how? Do the numbers of numbered section headings count? If in-text citations count, how many words are they? When citing a single paper as "[1]," or as "(Huff & Geis, 1954)," do these citation styles have different word counts? When citing multiple papers in a single command, how many words is "[1,2,3]"? Do captions of tables and figures count? Do words in axis labels in graphs count? How about numbers on those axes? For all these questions, reasonable answers can be defined. It is entirely possible for conference organizers to formulate answers to these questions and communicate them to paper authors. However, it is very likely that whatever reasonable answers we choose, the resulting word count will not match the inbuilt word counts delivered by software such as Adobe Acrobat or Microsoft Word. These mismatches will inevitably frustrate paper authors. Since these two problems seem prohibitive for word count to be incorporated in conference organization, we do not see the page limit replaced any time soon.

As long as we still have a page limit, you might wonder if this is really such a big deal. Do we really need to be so strict with people skirting around the margins? The thing is, as we outlined in the second paragraph of the Introduction, when writing an overlong paper, the high road would be to be more economic with your phrasing. When reducing the overlong paper to the page limit, when authors take the high road, they will cut corners in the number of words, sentences, arguments used to fortify their case. Less fortification for the paper's arguments reduces the probability that the paper is accepted. This is not a problem on its own, *as long as the playing field is level*. If other overlong papers get to use tricks to squeeze under the page limit, and if those tricks are not punished, they do not suffer this reduced probability. Hence, if we do not counter page limit cheating tricks, we put authors taking the high road at a disadvantage, which is not a signal we as scientific community should want to broadcast.

Therefore, we say: when a page limit is in place, authors should be forced to comply with it following the accompanying guidelines as strictly as humanly possible.

3.1 | Benefits and limitations of automation in this process

The contents of this paper are the direct result of lots of manual work on behalf of the proceedings chairs. For a conference of the magnitude of ECMLPKDD, this is still doable; if one were to replicate this analysis for a conference of the magnitude of AAAI 2019, where 1,150 papers were accepted, it is a ridiculous amount of work. Beyond that, not all proceedings chairs may feel strongly enough about this topic to invest the time necessary for this analysis. However, the job does seem like a good candidate for outsourcing to automation.

Springer's *Data Mining and Knowledge Discovery* journal, for instance, has an online system for paper submission enforcing style guide matters. When submitting a paper to the journal, you must upload your LaTeX sources to their website, and they will compile a submission PDF for you from your sources. This gives them a lot of control on what happens within those source files. Could not we use a similar system to solve our problem?

3.1.1 | System overload

The problem is that the situation for journals and conferences are not identical. For relatively small scientific communities, the answer may very well be positive. However, as mentioned before, we are data miners, and data mining is booming. So booming, in fact, that the sheer volume of paper submissions managed to break the Microsoft Conference Management Toolkit around the paper submission deadline of NeurIPS 2019 (Peng, 2019). The simple confluence of thousands of papers attracted by a submission deadline manages to break our electronic systems already; if we were to add to this process an automated LaTeX compilation from source files, this would inflate this problem to an unacceptable degree. Hence, we cannot make automated source compilation from scratch part of the paper submission process.

We can, however, demand that authors upload a zip file with their source material to the existing systems instead of their self-compiled PDF, and analyze the contents of that zip directly after the paper submission deadline.

3.1.2 | The problem with automated desk rejects

Another problem is that we must keep the human in the loop. Several of the commands we highlight as space cheats can be used for both nefarious and perfectly legal means. For instance, \resizeboxing a figure is perfectly acceptable, while \resizeboxing a table illegally reduces the font size. People can reduce \arraystretch to cheat space, but they can also increase \arraystretch simply to make their table look a bit better. We cannot simply automatically detect some commands and reject the paper using them without seeing the papers. Instead, the best we can do is run an automated check on whether any of the possibly offending commands appear in the LaTeX sources, and if so, flag up the paper for manual inspection by the proceedings chairs.

In essence, any technical solution that we would suggest for this problem would run into the same benefits and drawbacks as any antivirus program: it aims to detect malicious code. Both for detecting virus code and LaTeX space cheating commands, technological solutions can only make manipulation more difficult; it cannot reasonably be expected to completely eradicate the problem.

3.2 | Our proposal: The TeXnical Desk Reject Phase

We propose the following change in the conference paper submission and review procedure. After the paper submission deadline, plan a week for the TeXnical Desk Reject Phase. It works as follows:

1. Paper bidding starts as usual, and papers are assigned to reviewers as usual.
2. For the paper submission deadline, authors submit a zip file with their LaTeX sources, along with their self-compiled PDF. The author-submitted PDFs are sent to reviewers.
3. Directly after the paper submission deadline, the proceedings chairs run the submitted zip files through an automated system, to be designed. The only functionality is that the system checks whether any of the commands listed in this paper appear in the submitted LaTeX sources. If at least one of the commands appears, the paper gets flagged.

4. In the week after this automated check, the proceedings chairs recompile all the flagged papers, removing all commands that are deemed offending (keeping in mind the subtleties around \resizebox and \arraystretch as pointed out in the previous section). If the recompiled version complies with the page limit, the flag is removed. If the recompiled version violates the page limit, the paper is TeXnically desk rejected.
5. The program chairs inform the authors of TeXnically desk rejected papers of the decision.
6. The program chairs inform the reviewers having TeXnically desk rejected papers in their batch of the decision, pointing out that their review for this paper is no longer required.

A precondition for this procedure to work smoothly is that we must clearly communicate to all authors that we plan to include this phase, and under which conditions their papers will be TeXnically desk rejected. The extensive list of examples of what we consider page cheating should be available to authors, so that no one can complain that they did not know what is or is not illegal. As an exemplar of such information distribution, the AAAI 2020 Author Kit contains a PDF guide (AAAI Press Staff, 2020), whose Tables 1 and 2 list LaTeX commands and packages, respectively, that must not be used.

A potential banana skin in the implementation of the TeXnical Desk Reject Phase, is that reviewers may already have started reviewing papers in their batch while steps 3 and 4 are still ongoing. Hence, desk rejects could be issued after the reviewers have already begun their work. One would want to prevent that this causes frustration among the reviewers, which could reduce their willingness to review for this venue in the future. Two policies are therefore of key importance when implementing this phase. On the one hand, reviewers must be explicitly informed of this way of working. On the other hand, steps 3 and 4 must take as small an amount of time as possible. If reviewers can be ascertained that TeXnical desk rejects will be sent out in at most a week, this imposes an upper bounded on their potential for frustration. In this sense, a little expectation management can go a long way.

4 | CONCLUSIONS

LaTeX offers far too many ways to change the appearance of a paper. As a consequence, it is far too easy for authors to squeeze in more material than the page limit would allow if one were to play the game fairly. We outline several such ways to cheat the page limit, and illustrate its effect on the accepted papers in the Research and Applied Data Science track at ECMLPKDD 2019. As we have seen in Table 3, 47.7% of the papers violate the page limit when recompiled according to the Springer guidelines, with the biggest offender adding a whopping 50% extra pages (cf. Table 2). To tackle the problem, we propose to include a TeXnical Desk Reject Phase into the process, directly after the paper submission deadline.

In some instances, deciding whether LaTeX code violates the guidelines can be in the eye of the beholder. This article explores examples of \arraystretch and \resizebox that can be either in compliance with or in violation of the guidelines, depending on the context in which they are used. Since lines can be blurry, we do not claim that the TeXnical Desk Reject Phase will completely eradicate the problem of cheating the page limit. However, we are confident that its introduction will reduce the fraction of papers violating the page limit from 47.7% toward a substantially lower percentage. Part of this effect will be due to raised awareness of the problem: we do not think that 47.7% of paper authors use space cheating commands because of malicious intent; instead, many of these authors will use such commands without realizing that their use constitutes cheating. In this sense, the publication of this paper also sends a signal to the community.

While this paper is largely written from the viewpoint of data miners and a specific data mining conference, we think that its lessons can be generalized beyond this scope to any venue where submissions have to comply with a page limit. We think that the automated system required to support proceedings chairs in completing the new phase will not be too difficult to build, yet it ought to substantially reduce the workload involved in performing this task. When done well, the TeXnical Desk Reject Phase will result in more fairness in the conference submission evaluation procedure, by being harsher toward papers not playing by the rules.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Wouter Duivesteijn: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; supervision; validation; visualization; and writing-original draft, review and editing. **Sibylle Hess**: Conceptualization; investigation; methodology; project administration; supervision; and writing-review and editing. **Xin Du**: Conceptualization; investigation; methodology; project administration; supervision; and writing-review and editing.

ORCID

Wouter Duivesteijn https://orcid.org/0000-0003-0412-8864

ENDNOTES

[1] Notice that these three kinds of boxes are compatible with the Springer guidelines when used to resize figures. This is often even necessary to allow a figure to fit on a page. Conversely, thusly resizing tables or algorithms runs afoul of font size specifications, and hence violates the guidelines.

[2] We call it a problem because it poses a problem for the particular situation sketched in this paper. Outside of this particular context, experiencing a boom in your research field can of course be quite beneficial.

RELATED WIREs ARTICLES

Internet of things and data mining: From applications to techniques and systems
On the accuracy of linear regression routines in some data mining packages
Evolutionary computation for training set selection

REFERENCES

AAAI Press Staff. (2020). *AAAI press formatting instructions for authors using LaTeX—A guide*. Retrieved from https://www.aaai.org/Publications/Templates/AuthorKit20.zip

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. (2019). *ECMLPKDD*. Retrieved from http://www.ecmlpkdd2019.org/

Huff, D., & Geis, I. (1954). *How to lie with statistics*. London, England: Victor Gollancz.

Peng, T. (2019). *Paper submissions break neurips 2019 paper submission system*. Retrieved from https://medium.com/syncedreview/paper-submissions-break-neurips-2019-paper-submission-system-884a60e32a82

Springer. (2019a). *Guidelines for proceedings authors*. Retrieved from ftp://ftp.springernature.com/cs-proceeding/svproc/guidelines/Springer_Guidelines_for_Authors_of_Proceedings_CS.pdf

Springer. (2019b). *Information for authors of springer computer science proceedings*. Retrieved from https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines

How to cite this article: Duivesteijn W, Hess S, Du X. How to cheat the page limit. *WIREs Data Mining Knowl Discov*. 2020;10:e1361. https://doi.org/10.1002/widm.1361