

LuaCensor: A package for redacting sensitive information

Elijah Z Granet*

23 February 2022

Version 1.1.0

Contents

1	Very Quick Guide	2
1.1	Purpose	2
1.2	Usage	2
1.3	The warning option	2
1.4	Requirements	3
1.5	Limitations	4
1.6	Demonstration	4
2	More detailed information	5
2.1	Purpose	5
2.2	The censoring mechanism	5
2.3	The accessibility feature	5
2.4	Bugs and development	6
2.5	Licensing	6
2.6	Some useful advice	6
3	Implementation	7
3.1	Dependencies	7
3.2	fonts	8
3.3	Removing pesky environments	9
3.4	The ‘cnsr’ command	10
3.5	the Lua magic	12
3.6	The warning ’ option	15

*e-mail: ezg21@cantab.ac.uk

4	Version History	18
4.1	1.1.0	18
4.2	1.0.0	18

1 Very Quick Guide

1.1 Purpose

This package redacts sensitive information using Lua, and adds accessibility support.

1.2 Usage

The package is called with:

```
\usepackage{luacensor}
```

Sensitive information is enclosed within the command:

```
\cnsr{John Smith}
```

When the outputted document is intended for authorised readers who are supposed to see sensitive information, no further action is needed. When the outputted document is for general audiences, who are *not* supposed to see sensitive information, add the following line to the preamble:

```
\cnsrtrue
```

This activates the censoring globally.

1.3 The **warning** option

For many documents, the presence or absence of redaction in the document will be obvious by black bars in the first page. However, for other documents, particularly long reports or, for example, a court judgment with a title page, it is useful to give an indication. This can be done by calling the package with:

```
\usepackage[warning]{luacensor}
```

This prints automatically prints a warning at the top of the page indicating either that the text is redacted or unredacted. The default ‘safe’ and ‘danger’ warnings, below, are there mostly as a place holder, because it is anticipated that the precise wording of the warning will vary by jurisdiction.

The default danger warning:



This document is NOT redacted. It contains private and confidential personal data, and may NOT be distributed, published, or shown to those without the right to view such information. The publication of the information in this document may constitute a contempt of court, punishable by a term of imprisonment.

The default safe warning:

This document has been altered to remove sensitive personal data. It is cleared for publication and dissemination.

As I noted above, it is unlikely these default options will be suitable for most situations, and for this reason it is easy to change. To alter the text used in the ‘safe’ option, simply use:

```
\renewcommand{\textsafe}{This is the text in safe mode}
```

To change the warning heading option for the ‘danger’ text, renew the command `\warnword` to change the word ‘warning’, renew the command `\dangerblock` to alter or remove the warning triangles. The other commands for more custom changing of the warning are all in the source code and easily altered for even more custom work.

```
\renewcommand{\textwarn}{This is the text in danger mode}
```

To alter the warning triangles,

1.4 Requirements

This package will **only** work in Lua^LTeX. The package works out of the box with a standard TeX distribution, but ideally, I would strongly suggest installing the free (*libre & gratis*) ‘Redacted’ font from Google Fonts, which gives a quite aesthetically pleasing black bar effect.¹

¹Available at: <https://fonts.google.com/specimen/Redacted>; a version is also available at the GitHub repo of this project: <https://github.com/ezgranet/luacensor>

2 More detailed information

2.1 Purpose

This package is a relatively lightweight and aesthetically pleasing censorship solution which includes accessibility features to allow screen readers to be aware that content has been redacted.

2.2 The censoring mechanism

The package uses Luas's `toks` filter to replace all `UTF8` characters with a single glyph (`•` in the case of `Redacted`, and a Unicode black rectangle in the fallback `TEX` default font `Source Sans Pro`). In both font options, these combine visually into a single line (though this can be deconstructed in a text editor).

However, while changing all characters into a single character is effective in *most* cases, this alone would not be sufficiently secure for reliable usage. This is because knowing the length of a censored name could be combined with other information in, for example, a Family Court judgment, to allow for what lawyers call 'jigsaw identification' (*eg*, where there is only one person with an eight letter surname who meets the other details given in the judgment).

Therefore, the package adds an extra layer of security by randomly changing the length of strings during the censorship phase; censored strings can thus be either longer or shorter by a few characters. This means that while the area of the blacked out content will be *approximately* similar to the length of the uncensored string (which means wireframing more or less works), it cannot be used to reverse engineer information about the censored content.

2.3 The accessibility feature

One concern about document redaction is ensuring that visually impaired readers of your document, who use screen reading software to listen to your text, may encounter problems with censored content. If the screen reading software skips over the censored text altogether, it will be a very confusing jump for the visually impaired user. If the screen reading software reads the replacement characters, it will be very annoying for the visually impaired reader to hear, in a censored paragraph, the same character being read out *ad nauseam* (*eg*, 'Asterisk, asterisk, asterisk...').

To overcome this limitation, the package uses the `accsup` package to add an 'actual text' feature which will lead screen readers (and utilities like `pdf to text`) to replace the string of replacement characters with the two words 'TEXT REDACTED'. This also will be encountered by naïve users who try to outdo the package by copying and pasting the black blocks from Adobe™ Acrobat or Reader. (However, because other PDF readers, like Apple's Preview, do not implement accessibility features, this is **not** an additional security feature and is not on its own sufficient to work for redaction; if it were otherwise, the rest of the package would be unnecessary)

2.4 Bugs and development

All bugs, feature requests, or other technical points should be submitted to the package's official Github page.²

2.5 Licensing

The software is free and open-source software licensed under the Latex Public Project Licence, version 1.3c.³

2.6 Some useful advice

This package is really good at some things, but if you find it breaks down on censoring complex L^AT_EX code, the existing `ENSOR` package on CTAN is excellent (albeit less good with accessibility), and works with non-Lua versions of T_EX. Incidentally, you can use both this package and `ENSOR` in the same file without trouble; this (*not* a penchant for annoying tech-speak) is why the main command in this package is `ENSR` without vowels.

²<https://github.com/ezgranet/luacensor>

³<https://www.latex-project.org/lppl/>

3 Implementation

```
7 %luacensor.sty
8 %luacensor.sty
9 \def\luacensorversionnumber{1.1.0}
10 \ProvidesPackage{luacensor}
11 [2022/02/22 \luacensorversionnumber\
12 Redact sensitive information using Lua]
13 % !TeX program = lualatex
14 % !TeX encoding = utf8
15 % This work may be distributed and/or modified under the
16 % conditions of the LaTeX Project Public License, either version 1.3
17 % of this license or (at your option) any later version.
18 % The latest version of this license is in
19 % http://www.latex-project.org/lppl.txt
20 % and version 1.3 or later is part of all distributions of LaTeX
21 % version 2005/12/01 or later.
22 %
23 % This work has the LPPL maintenance status `maintained'.
24 %
25 % The Current Maintainer of this work is Elijah Z Granet
26 %%%%%%%%%%%
27 %%%%%%%%%%%
28 % option (we'll come back
29 % to this later
30 %%%%%%%%%%%
31 %%%%%%%%%%%
32 \newif\ifwarning
33 \warningfalse
34 \DeclareOption{warning}{\warningtrue}
35 \ProcessOptions*
```

3.1 Dependencies

```
36 %%%%%%%%%%%
37 %%%%%%%%%%%
38 % DEPENDENCIES
39 %%%%%%%%%%%
40 %%%%%%%%%%%
41 \RequirePackage{luacode}
42 \RequirePackage{environ}% http://ctan.org/pkg/environ
43
44 \RequirePackage{verbatim}
45 % ^ for the censoring
```

```

46 \RequirePackage{accsupp}
47 %^for accessibility
48 \RequirePackage{fontspec}
49 %^for black lines
50 %in theory, you could do
51 %a lighter version of this
52 %package with just asterisks
53 %or '[TEXT-REDACTED]'
54 %And perhaps that would be better for
55 %the environment with printing
56 %BUT I MADE MY CHOICE!
57 \RequirePackage{xcolor}
58 \RequirePackage{graphicx}

```

3.2 fonts

```

59 %%%%%%%%%%%
60 %%%%%%%%%%%
61 % FONTS
62 %%%%%%%%%%%
63 %%%%%%%%%%%
64 %%%%%%%%%%%
65 %%%%%%%%%%%
66 % redacted is prettier and free to download
67 %%%%%%%%%%%
68 %%%%%%%%%%%
69 % Strongly recommended
70 %%%%%%%%%%%
71 %%%%%%%%%%%
72 %%%%%%%%%%%
73 %%%%%%%%%%%
74 \IfFontExistsTF{Redacted}{%
75 \newfontface\cnsrfnt[%%%%%%%%%
76 %the scale is arbitrary, but kind of works
77 %Scale=1.1,
78 %%the below declarations are to prevent warnings about shapes not being
79 %↪ available
80 %WordSpace=0,
81 ItalicFont={Redacted},%
82 BoldItalicFont={Redacted},%
83 BoldFont={Redacted},%
84 SmallCapsFont={Redacted}]{Redacted}
85 \newcommand{\onething}{\cnsrfnt\cdot}
86 \newcommand{\twothings}{\cnsrfnt\bullet}

```

```

86 \newcommand{\donothing}{\cnsrfnt\ }
87 %%%%%%%%%
88 %%%%%%%%%
89 %The little spaces let justification happen
90 %%%%%%%%%
91 %%%%%%%%%
92 %%%%%%%%%
93 %%%%%%%%%
94 % • chosen as an arbitrary average width
95 %%%%%%%%%
96 %%%%%%%%%
97 }{
98 %%%%%%%%%
99 %%%%%%%%%
100 % This option works perfectly
101 %fine, it's just less pretty
102 %%but a good fallback because
103 % Source Sans is in TeX dists by default
104 %%%%%%%%%
105 %%%%%%%%%
106 \newfontface\cnsrfnt[Scale=1.01,%To allow for separate use of source sans in
↪ text
107 WordSpace=0,%To make it all one black line
108 %the below declarations are to prevent warnings about shapes not being
↪ available
109 ItalicFont={Source Sans Pro Black},BoldItalicFont={Source Sans Pro
↪ Black},BoldFont={Source Sans Pro Black},SmallCapsFont={Source Sans Pro
↪ Black}}{Source Sans Pro Black}
110 %%%%%%%%%
111 %%%%%%%%%
112 % Bit of unicode magic below to make the black line effect
113 %%%%%%%%%
114 %%%%%%%%%
115 \newcommand{\onething}{\cnsrfnt - }
116 \newcommand{\twothings}{\cnsrfnt - - }
117 \newcommand{\donothing}{ }
118 }

```

3.3 Removing pesky environments

```

119 %%%%%%%%%
120 %%%%%%%%%
121 % A neat fallback for disappearing things...
122 %%%%%%%%%

```

```

123 %%%%%%%%%%%
124 %%%%%%%%%%%
125 %%%%%%%%%%%
126 % FULL CREDIT
127 % and FULSOME THANKS
128 % TO TEX.SE USER
129 % Werner for the code below
130 %%%%%%%%%%%
131 %%%%%%%%%%%
132 \makeatletter
133 \newcommand{\voidenvironment}[1]{%
134   \expandafter\providecommand\csname env@#1@save@env\endcsname{%
135   \expandafter\providecommand\csname env@#1@process\endcsname{%
136   \ifundefined{#1}{\RenewEnviron{#1}}}%
137 }
138 \makeatother
139 \newcommand{\hddn}[1]{%
140 \ifcnsr{}\else%
141 #1\fi}
142 \newenvironment*{hidden}{\begin{@empty}
143 }{\end{@empty}}
144 \voidenvironment{hidden}
145
146 %%%%%%%%%%%
147 %%%%%%%%%%%

```

3.4 The ‘cnsr’ command

```

148 %%%%%%%%%%%
149 %%%%%%%%%%%
150 %%%%%%%%%%%
151 %%%%%%%%%%%
152 % the CENSOR COMMAND
153 %%%%%%%%%%%
154
155 \newif\ifcnsr\cnsrfalse
156 \newcommand{\cnsr}[1]{%
157 \ifcnsr{%
158 \voidenvironment{equation*}%
159 \voidenvironment{equation}%
160 \voidenvironment{table}%
161 \voidenvironment{table*}%
162 \voidenvironment{tabular}%
163 \voidenvironment{tabular*}%

```

```

164 \voidenvironment{}%
165 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
166 % I don't know how many
167 % people use TEX native accent commands
168 % in LuaTEX given that using Unicode is more
169 %people's style. But just in case, because these can lead to stray accent
   ↪ marks floating above censored letters.
170 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
171 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
172 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
173 \renewcommand{\`}[1]{}%
174 \renewcommand{\'}[1]{}%
175 \renewcommand{\^}[1]{}%
176 \renewcommand{\"}[1]{}%
177 \renewcommand{\H}[1]{}%
178 \renewcommand{\~}[1]{}%
179 \renewcommand{\c}[1]{}%
180 \renewcommand{\k}[1]{}%
181 \renewcommand{\l}[1]{}%
182 \renewcommand{\=}[1]{}%
183 \renewcommand{\b}[1]{}%
184 \renewcommand{\.}[1]{}%
185 \renewcommand{\d}[1]{}%
186 \renewcommand{\r}[1]{}%
187 \renewcommand{\u}[1]{}%
188 \renewcommand{\v}[1]{}%
189 \renewcommand{\t}[1]{}%
190 \renewcommand{\o}[1]{}%
191 \renewcommand{\i}[1]{}%
192 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%d
193 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
194 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
195 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
196 % here we have the accsupp magic
197 % this operates by replacing the 'x's
198 % or unicode black squares as the case
199 % may be with an alt text
200 % this serves a dual purpose of both making
201 %pdftotext not break with huge strings of meaningless characters
202 %but more importantly
203 % it means screen readers don't subject
204 % their users to the meaningless reading out of unicode black squares 50
   ↪ times in a row!
205 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
206 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
207 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
208 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

209 \BeginAccSupp{method=plain,ActualText={TEXT REDACTED}}%
210 \rndstring{#1}%
211 \EndAccSupp{}}%
212 \else%
213 %%%%%%%%%%%
214 %%%%%%%%%%%
215 % if the conditional is off
216 % the command does absolutely nothing
217 %%%%%%%%%%%
218 %%%%%%%%%%%
219 #1%
220 \fi}
221 %%%%%%%%%%%
222 %%%%%%%%%%%

```

3.5 the Lua magic

```

223 % %%%%%%%%%%%
224 %%%%%%%%%%%
225 % The LUA MAGIC PART
226 %%%%%%%%%%%
227 %%%%%%%%%%%
228 %%%%%%%%%%%
229 %%%%%%%%%%%
230 \begin{luacode}

```

```

231 --fulsome thanks to TeX.SE users Henri Menke and David Carlisle, without whom
232 ↪ none of this would be possible
233 local function rndstring()
234     local toks = token.scan_toks(s)
235     local on = true
236     for n, t in ipairs(toks) do
237         if t.csname == "begin" or t.csname == "end" then
238             on = false
239         -- The below is necessary as TeX primitives can break the code otherwise
240         ↪ because they do not use brackets
241         end
242     if not(on) and t.cmdname == "right_brace" then
243         on = true
244         -- This prevents needless errors about gibberish up commands

```

```

244 end
245 if on and t.csname == "&" then
246     local letter = token.create'donothing'
247     toks[n] = letter
248
249 elseif on and t.csname == "%" then
250     local letter = token.create'donothing'
251     toks[n] = letter
252
253 elseif on and t.csname == "$" then
254     local letter = token.create'donothing'
255     toks[n] = letter
256
257 elseif on and t.csname == "#" then
258     local letter = token.create'donothing'
259     toks[n] = letter
260
261 elseif on and t.csname == "_" then
262     local letter = token.create'donothing'
263     toks[n] = letter
264
265 elseif on and t.csname == "{" then
266     local letter = token.create'donothing'
267     toks[n] = letter
268
269 elseif on and t.csname == "}" then
270     local letter = token.create'donothing'
271     toks[n] = letter
272
273 elseif on and t.csname == "~" then
274     local letter = token.create'donothing'
275     toks[n] = letter
276
277 elseif on and t.csname == "^" then
278     local letter = token.create'donothing'
279     toks[n] = letter
280 elseif on and t.cmdname == "letter" then
281     -- The below is the randomness part of this, which I admit is fairly
282     ↪ arbitrary, but will more often artificially shorten strings than lengthen
283     ↪ them, as testing found if lengthening was too frequent, it led to really
284     ↪ unsightly long strings.
285     local f = math.random (1,20)
286     if f == 1 then
287         local letter = token.create'donothing'
288         toks[n] = letter
289
290     elseif f == 2 then

```

```

288         local letter = token.create'donothing'
289     toks[n] = letter
290     elseif f == 3 then
291         local letter = token.create'donothing' toks[n] = letter
292     elseif f == 4 then
293         local letter = token.create'twothings'
294     toks[n] = letter
295     elseif f == 5 then
296         local letter = token.create'donothing' toks[n] = letter
297
298     else
299         local letter = token.create'onething'
300     toks[n] = letter
301         end
302     elseif
303         on and t.cmdname == "spacer" then
304         local f = math.random (1,20)
305         if f == 2 then
306         local letter = token.create'donothing'
307     toks[n] = letter
308         elseif f == 3 then
309
310         local letter = token.create'donothing' toks[n] = letter
311     elseif f == 4 then
312     local letter = token.create'donothing'
313     toks[n] = letter
314         elseif f == 5 then
315
316
317         local letter = token.create'twothings' toks[n] = letter
318     elseif f == 6 then
319
320
321         local letter = token.create'donothing' toks[n] = letter
322         elseif f == 7 then
323     local letter = token.create'donothing' toks[n] = letter
324
325
326     else
327         local letter = token.create'onething'
328     toks[n] = letter
329
330         end
331     elseif
332         on and t.cmdname == "other_char" then
333     local f = math.random (1,20)
334

```

```

335         if f == 2 then
336             local letter = token.create'donothing'
337             toks[n] = letter
338                 elseif f == 3 then
339
340                     local letter = token.create'donothing'   toks[n] = letter
341             elseif f == 4 then
342                 local letter = token.create'donothing'
343                 toks[n] = letter
344                 elseif f == 5 then
345
346
347                     local letter = token.create'twothings'   toks[n] = letter
348             elseif f == 6 then
349
350
351                 local letter = token.create'donothing'   toks[n] = letter
352                 elseif f == 7 then
353                 local letter = token.create'donothing'   toks[n] = letter
354
355
356             else
357                 local letter = token.create'onething'
358                 toks[n] = letter
359
360                 end
361             end
362         end
363         --Drop the token in and move on
364         token.put_next(toks)
365     end
366     local lft = lua.get_functions_table()
367     --make a global command
368     lft[#lft + 1] = rndstring
369     token.set_lua("rndstring", #lft, "global")

```

```

370 \end{luacode}

```

3.6 The **'warning'** option

```

371 %%%%%%%%%%
372 %%%%%%%%%%
373 % WARNING FUN YAY
374 %%%%%%%%%%
375 %%%%%%%%%%
376 %%%%%%%%%%
377 %%%%%%%%%%
378 % Definitely this whole section
379 %is there to be user modified, because
380 % depending on language, jurisdiction
381 %type of document etc, everyone will need
382 %a specific warning style. So the important
383 % part of the code here is the
384 % conditional and global [warning]
385 % option, because that's the magic value added
386 %%%%%%%%%%
387 %%%%%%%%%%
388 %%%%%%%%%%
389 %%%%%%%%%%
390 %%%%%%%%%%
391 %%%%%%%%%%
392 % fonts for the warning:
393 %I chose default LaTeX fonts
394 % here to be changed as users wish
395 %%%%%%%%%%
396 %%%%%%%%%%
397 \newfontface\wrnstncl{QT Military}
398 \newcommand{\warnword}{WARNING}
399 \newfontface\smb1{Deja Vu Sans Bold}
400 \newcommand{\danger}{\smb1 \triangle\normalfont}
401 \newcommand{\warnformat}{\sffamily\bfseries \color{red}}
402 \newcommand{\textwarn}{This document is {\underline{NOT}} redacted. It
↳ contains private and confidential personal data, and may {\underline{NOT}}
↳ be distributed, published, or shown to those without the right to view
↳ such information. The publication of the information in this document may
↳ constitute a contempt of court, punishable by a term of imprisonment.}
403 \newcommand{\textsafe}{This document has been altered to remove sensitive
↳ personal data. It is cleared for publication and dissemination.
404 }
405 \definecolor{darkgreen}{rgb}{0.0, 0.2, 0.13}
406 \definecolor{darkspringgreen}{rgb}{0.09, 0.45, 0.27}
407 \definecolor{forestgreen}{rgb}{0.13, 0.55, 0.13}
408 \newcommand{\dquad}{\danger\danger\danger\danger}
409 \newcommand{\dangersign}[1]{\scalebox{2}{\huge\danger}}
410 \newcommand{\dangerblock}{\scalebox{2}{\huge\danger\quad\danger\quad\danger}}
411 \newcommand{\warnblock}{\Large\wrnstncl\warnword\quad\warnword\quad\warnword}
412 \newcommand{\tworules}{\hrule width \hsize height .7pt\vskip2pt\hrule width
↳ \hsize height .7pt}

```

```

413 \newcommand{\allwarning}{\dangerblock\\\warnblock\\\normalfont\smallskip\warnformat\textwarn
↵ }
414 \newcommand{\confwarning}{%
415 %%%%%%%%%%%
416 %%%%%%%%%%%
417 % The warning option
418 %%%%%%%%%%%
419 %%%%%%%%%%%
420 \ifwarning
421 \ifcnsr
422 %%%%%%%%%%%
423 %%%%%%%%%%%
424 % a note saying document is redacted
425 %%%%%%%%%%%
426 %%%%%%%%%%%
427 \begin{center}
428 \color{forestgreen}
429 \tworules\vskip5pt
430 \normalsize\normalfont\sffamily\bfseries\textsafe
431 \vskip5pt\tworules
432 \end{center}
433 \else
434 %%%%%%%%%%%
435 %%%%%%%%%%%
436 % The WARNING for un-redacted docs
437 %%%%%%%%%%%
438 %%%%%%%%%%%
439 \begin{center}\color{red}\tworules\vskip 5pt\allwarning
440 \vskip5pt\tworules%
441 \end{center}%
442 \fi%
443 \else\fi}
444 %%%%%%%%%%%
445 %%%%%%%%%%%
446 % Allow \maketitle
447 % on same page
448 % yay
449 %
450 %%%%%%%%%%%
451 %%%%%%%%%%%
452 \ifwarning\let\oldmaketitle\maketitle\renewcommand{\maketitle}{\let\newpage\relax\maketitle}\else\fi
453 %%%%%%%%%%%
454 %%%%%%%%%%%
455 % print the warning at the start of the document
456 %%%%%%%%%%%
457 %%%%%%%%%%%
458 \AtBeginDocument{\confwarning}

```

4 Version History

4.1 1.1.0

22 February 2022: Added the `\warning` option and fixed a few errors in the code resulting from \TeX primitives causing issues.

4.2 1.0.0

18 February 2022: Package creation