LuaCensor: A package for redacting sensitive information

Elijah Z Granet*

7 April 2024 Version 1.1.1

Contents

1	Ver	y 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		10			
	3.5	the Lua magic	12			
	3.6	The 'warning' option	15			

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

4	Version History				
	4.1	1.1.1	18		
	4.2	1.1.0	18		
	4.3	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.







This document is <u>NOT</u> redacted. It contains private and confidential personal data, and may <u>NOT</u> 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 LualAT_EX. The package works out of the box with a standard T_EX distribution, but ideally, I would strongly suggest installing the free (*libre* & *gratis*) 'Redacted' font from Google Fonts, which gives a quite æsthetically pleasing black bar effect.¹

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

1.5 Limitations

The package is completely effective at censoring text formatted with normal LATEX commands, by which I mean that it is impossible for someone to ascertain the original text (or even its precise length) from the outputted PDF when the \cnsrtrue option has been activated.

However, math mode is used, it will censor numbers, but not operators or TEX (as opposed to Unicode) operators. This is probably fine for most instances, but unacceptable where security is of the highest priority, and I would not really recommend using this package to censor highly secret formulæ; the censor package in your TEX distribution will do a better job of that. The package is set to completely disappear the output (as opposed to black bar over) of the math, align, equation, tabular, and a few other environments, as disappearing these environments proved more secure than the piecemeal blacking out I saw. It is probable that there are packages and macros that will break the cnsr macro, and therefore, care should be taken to always examine output before public distribution.

Users should also be aware that many TEX primitives confuse the package, in particular \vskip, \hskip etc that take arguments outside curly braces cause difficulties; the best solution is simply to either use the LATEX alternatives (eg, \hspace) or enclose the primitives in the \hddn command which simply disappears them.

1.6 Demonstration

%In the preamble: \usepackage{luacensor} \
\text{begin}{quote} \cnsr{\Whereas recognition} of the \cnsrtrue\cnsr{\inherent} \
\dignarrow dignity and of the equal and inalienable rights of all members of the \(
\rightarrow human family\} is the foundation of freedom, justice and peace in the \(
\rightarrow world,\)
\cnsr{\Whereas disregard and contempt for human rights have resulted in \(
\rightarrow barbarous acts which have outraged the conscience of mankind, and the \(
\rightarrow advent of a world in which human beings shall enjoy freedom of speech \(
\rightarrow and belief and freedom from fear and want has been proclaimed as the \(
\rightarrow highest aspiration of the common people,\} \(
\text{end}{\quote}\)
\text{Whereas recognition of the of the common people,} \(
\text{the foundation of freedom, justice and peace in the world,} \)

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, 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 pdftotext) to replace the string of replacement characters with the two words 'TEXT RE-DACTED'. 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-sour ce 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 LATEX code, the existing censor package on CTAN is excellent (albeit less good with accessibility), and works with non-Lua versions of TEX. Incidentally, you can use both this package and censor in the same file without trouble; this (not a penchant for annoying tech-speak) is why the main command in this package is cnsr without vowels.

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

³https://www.latex-project.org/lppl/

3 Implementation

```
%luacensor.sty
    \def\luacensorversionnumber{1.1.1}
   \ProvidesPackage{luacensor}
   [2024/04/06 \luacensorversionnumber\
   Redact sensitive information using Lua]
   % !TeX program = lualatex
   % !TeX encoding = utf8
    % This work may be distributed and/or modified under the
14
   % conditions of the LaTeX Project Public License, either version 1.3
   % of this license or (at your option) any later version.
   % The latest version of this license is in
   % http://www.latex-project.org/lppl.txt
   % and version 1.3 or later is part of all distributions of LaTeX
    % version 2005/12/01 or later.
   % This work has the LPPL maintenance status `maintained'.
22
   % The Current Maintainer of this work is Elijah Z Granet
   $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
   % option (we'll come back
    % to this later
   \newif\ifwarning
   \warningfalse
    \DeclareOption{warning}{\warningtrue}
    \ProcessOptions*
```

3.1 Dependencies

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

3.2 fonts

```
59
  % FONTS
60
  % redacted is prettier and free to download
  66
  % Strongly recommended
   72
   \IfFontExistsTF{Redacted}{%
73
  \newfontface\cnsrfnt[%%%%%
74
  %the scale is arbitrary, but kind of works
75
   %Scale=1.1,
   %the below declarations are to prevent warnings about shapes not being
   → available
   %WordSpace=0,
78
  ItalicFont={Redacted},%
   BoldItalicFont={Redacted},%
   BoldFont={Redacted},%
   SmallCapsFont={Redacted}]{Redacted}
   \newcommand{\onething}{\cnsrfnt\ • }
   \newcommand{\twothings}{\cnsrfnt\ • •}
84
   \newcommand{\donothing}{\cnsrfnt\ }
```

```
86
   87
   %The little spaces let justification happen
   % • chosen as an arbitrary average width
93
   $$$$$$$$$$$$$$$$$$$$$$$$$$$
   % This option works perfectly
   %fine, it's just less pretty
100
   %but a good fallback because
   % Source Sans is in TeX dists by default
   %%%%%%%%%%%%%%%%%%%
   \newfontface\cnsrfnt[Scale=1.01,%To allow for separate use of source sans in
   WordSpace=0,%To make it all one black line
106
   %the below declarations are to prevent warnings about shapes not being
   ItalicFont={Source Sans Pro Black},BoldItalicFont={Source Sans Pro
108
   → Black},BoldFont={Source Sans Pro Black},SmallCapsFont={Source Sans Pro
    → Black}]{Source Sans Pro Black}
   109
   % Bit of unicode magic below to make the black line effect
111
   112
   %%%%%%%%%%%%%%%%%%%%
113
   \newcommand{\onething}{\cnsrfnt = }
   \newcommand{\twothings}{\cnsrfnt = = }
115
   \newcommand{\donothing}{ }
116
   }
```

3.3 Removing pesky environments

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

3.4 The 'cnsr' command

```
147
    999999999999999999999999999999
148
    149
    150
    % the CENSOR COMMAND
    152
    \newif\ifcnsr\cnsrfalse
    \newcommand{\cnsr}[1]{%
155
    \ifcnsr{%
156
    \voidenvironment{equation*}%
    \voidenvironment{equation}%
158
    \voidenvironment{table}%
159
    \voidenvironment{table*}%
    \voidenvironment{tabular}%
    \voidenvironment{tabular*}%
162
    \voidenvironment{}%
163
```

```
164
    % I don't know how many
165
   % people use TEX native accent commands
    % in LuaTEX given that using Unicode is more
    %people's style. But just in case, because these can lead to stray accent

→ marks floating above censored letters.

    169
    170
    \renewcommand{\'}[1]{\}%
172
    \rcmal{'}[1]{}%
173
    \mbox{renewcommand} \^{[1]{}}
    \mbox{renewcommand} \" \[1] \ \
175
    \mathbf{H}[1]
176
    \rcmall \renewcommand \{\ \ \ \ \ \ \ \ \ \} [1] \{\}%
177
    \mathbf{c}_{1}{}
    \mathbf{k}[1]
179
    \mbox{renewcommand}\{\l][1]{}
180
    \mbox{renewcommand} \=\[1]_{}%
    \mathbf{b}[1]
    183
    \mbox{renewcommand} \d [1]{}
184
    \renewcommand{\r}[1]{}
    \mbox{renewcommand}\{\u\}[1]{}%
186
    \renewcommand{\v}[1]{}%
187
    \mathbf{t}[1]
    \rcmall \rcmall renewcommand {\o}[1]{}%
189
    \renewcommand{\i}[1]{}%
190
    192
    193
    % here we have the accsupp magic
195
    % this operates by replacing the 'x's
196
    % or unicode black squares as the case
   % may be with an alt text
    % this serves a dual purpose of both making
199
    %pdftotext not break with huge strings of meaningless characters
200
    %but more importantly
    % it means screen readers don't subject
   %. their users to the meaningless reading out of unicode black squares 50
203
    204
    205
    206
    \BeginAccSupp{method=plain,ActualText={TEXT REDACTED}}%
```

```
\rndstring{#1}%
209
  \EndAccSupp{}}%
210
  \else%
  212
  % if the conditional is off
  % the command does absolutely nothing
  216
  217
  #1%
219
  %3%3%3%3%3%3%3%3%
220
```

3.5 the Lua magic

```
--fulsome thanks to TeX.SE users Henri Menke and David Carlisle, without whom

→ none of this would be possible

     local function rndstring()
231
         local toks = token.scan_toks(s)
232
             local on = true
233
     for n, t in ipairs(toks) do
         if t.csname == "begin" or t.csname == "end" then
235
236
     -- The below is necessary as TeX primitives can break the code otherwise
237
      → because they do not use brackets
238
        if not(on) and t.cmdname == "right_brace" then
          on = true
241
           -- This prevents needless errors about gibberish up commands
242
     end
243
```

```
if on and t.csname == "&" then
244
       local letter = token.create'donothing'
245
       toks[n] = letter
246
247
     elseif on and t.csname == "%" then
248
       local letter = token.create'donothing'
       toks[n] = letter
250
251
     elseif on and t.csname == "$" then
252
       local letter = token.create'donothing'
253
      toks[n] = letter
254
     elseif on and t.csname == "#" then
       local letter = token.create'donothing'
257
      toks[n] = letter
258
259
      elseif on and t.csname == "_" then
260
       local letter = token.create'donothing'
261
      toks[n] = letter
263
      elseif on and t.csname == "{" then
264
       local letter = token.create'donothing'
265
      toks[n] = letter
267
     elseif on and t.csname == "}" then
268
       local letter = token.create'donothing'
       toks[n] = letter
270
271
      elseif on and t.csname == "~" then
       local letter = token.create'donothing'
273
      toks[n] = letter
274
     elseif on and t.csname == "^" then
276
       local letter = token.create'donothing'
277
      toks[n] = letter
278
      elseif on and t.cmdname == "letter" then
279
      -- The below is the randomness part of this, which I admit is fairly
280
      \hookrightarrow arbitrary, but will more often artificially shorten strings than lengthen
      \leftrightarrow them, as testing found if lengthening was too frequent, it led to really

→ unsightly long strings.

                  local f = math.random (1,20)
281
                  if f == 1 then
282
                     local letter = token.create'donothing'
283
       toks[n] = letter
284
                           elseif f == 2 then
                     local letter = token.create'donothing'
287
```

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

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

```
369 \end{luacode}
```

3.6 The 'warning' option

```
% WARNING FUN YAY
372
    373
    374
    375
    % Definitely this whole section
377
    %is there to be user modified, because
378
    % depending on language, jurisdiction
379
    %type of document etc, everyone will need
380
    %a specific warning style. So the important
    % part of the code here is the
382
    % conditional and global [warning]
    % option, because that's the magic value added
    385
    386
    388
    % fonts for the warning:
391
    %I chose default LaTeX fonts
392
    % here to be changed as users wish
    394
    395
    \newfontface\wrnstncl{QT Military}
396
    \newcommand{\warnword}{WARNING}
397
    \newfontface\smbl{DejaVuSans-Bold}
398
    \newcommand{\danger}{\smbl △\normalfont}
399
    \newcommand{\warnformat}{\sffamily\bfseries \color{red}}
    \mbox{\newcommand{\textwarn}{This document is {\underline{NOT}}} \ \ reducted. It}
401
       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.}
    \newcommand{\textsafe}{This document has been altered to remove sensitive
402
        personal data. It is cleared for publication and dissemination.
403
    \definecolor{darkgreen}{rgb}{0.0, 0.2, 0.13}
404
    \definecolor{darkspringgreen}{rgb}{0.09, 0.45, 0.27}
           \definecolor{forestgreen}{rgb}{0.13, 0.55, 0.13}
406
    \newcommand{\dquad}{\danger\danger\danger\danger}
407
408
    \newcommand{\dangersign}[1]{\scalebox{2}{\huge\danger}}
    \newcommand{\dangerblock}{\scalebox{2}{\huge\danger\quad\danger\quad\danger}}
409
    \newcommand{\warnblock}{{\Large\wrnstncl\warnword\quad\warnword\quad\warnword\}}
410
    \newcommand{\tworules}{\hrule width \hsize height .7pt\vskip2pt\hrule width
411
     \rightarrow \hsize height .7pt}
    \newcommand{\allwarning}{\dangerblock\\\warnblock\\\normalfont\smallskip\warnformat\textwarn
412
    → }
```

```
\newcommand{\confwarning}{%
413
            414
            415
            % The warning option
416
            \ifwarning
419
            \ifcnsr
420
            % a note saying document is redacted
423
            $$$$$$$$$$$$$$$$$$$$$$$$$$$
            425
            \begin{center}
426
            \color{forestgreen}
427
            \tworules\vskip5pt
            \normalsize\normalfont\sffamily\bfseries\textsafe
429
            \vskip5pt\tworules
430
            \ensuremath{\setminus} end\{center\}
431
            \else
432
            433
            % The WARNING for un-redacted docs
435
            436
            437
            \begin{center}\color{red}\tworules\vskip 5pt\allwarning
438
            \vskip5pt\tworules%
439
            \end{center}%
440
            \fi%
441
            \else\fi}
442
            443
            %%%%%%%%%%%%%%%%%%%%
444
            % Allow \maketitle
445
            % on same page
446
            % yay
447
           449
            450
            \label{thm:lemand} $$ \left( \operatorname{lot}_{{\left( \operatorname{lot}_{\operatorname{maketitle}}} \right)} \right) $$ if warning (\end{thm:lemand (\end{thm:lemand})} $$ if warning (\end{thm:lemand}) $$ if warnin
            452
            453
            % print the warning at the start of the document
            455
            456
            \AtBeginDocument{\confwarning}
457
```

4 Version History

4.1 1.1.1

 $\,$ 6 April 2024: Fixed an issue with OTF for dejavusans not loading by using the postscript name.

4.2 1.1.0

22 February 2022: Added the `warning' option and fixed a few errors in the code resulting from $T_{\hbox{\footnotesize E}}X$ primitives causing issues.

4.3 1.0.0

18 February 2022: Package creation