LuaCensor: A package for redacting sensitive information

Elijah Z Granet*

23 February 2022 Version 1.1.0

Contents

1	Ver	y Quick Guide	1			
	1.1	Purpose	1			
	1.2	Usage	1			
	1.3	The warning option	2			
	1.4	Requirements	2			
	1.5	Limitations	2			
	1.6	Demonstration	3			
2	More detailed information					
	2.1	Purpose	4			
	2.2	The censoring mechanism	4			
	2.3	The accessibility feature	4			
	2.4	Bugs and development	5			
	2.5	Licensing	5			
	2.6	Some useful advice	5			
3	Implementation 6					
	3.1	Dependencies	6			
	3.2	Fonts	7			
	3.3	Pesky enviroments	8			
	3.4	The cnsr command	9			
	3.5	Accessibility features	10			
	3.6	The Lua magic	11			
	3.7	The warning option	15			

^{*}e-mail: ezg21@cantab.ac.uk

4	Version History					
	4.1	1.1.0	17			
	4.2	1.0.0	17			

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 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.

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

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.

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

1.6 Demonstration

%In the preamble: \usepackage{luacensor}
\begin{quote}
\footnotesize \cnsr{Whereas recognition} of the \cnsrtrue\cnsr{inherent}
\upper dignity and of the equal and inalienable rights of all members of the

→ duman family} is the foundation of freedom, justice and peace in the → world,

Whereas recognition of the is 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, 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.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{1.1.0}
   \ProvidesPackage{luacensor-dev}
   [2022/02/22 \luacensorversionnumber\
   Redact sensitive information using Lua]
   % !TeX program = lualatex
   % !TeX encoding = utf8
    % This work may be distributed and/or modified under the
   % 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!
    \RequirePackage{xcolor}
    \RequirePackage{graphicx}
```

3.2 Fonts

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

```
86
   %The little spaces let justification happen
87
   % x chosen as an arbitrary average width
   93
   }{
   $$$$$$$$$$$$$$$$$$$$$$$$$$$$
   % This option works perfectly
   %fine, it's just less pretty
   %but a good fallback because
100
   % 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
105
   %the below declarations are to prevent warnings about shapes not being
   → available
   ItalicFont={Source Sans Pro Black},BoldItalicFont={Source Sans Pro
107
   → 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
   \newcommand{\onething}{\cnsrfnt = }
113
   \newcommand{\twothings}{\cnsrfnt = = }
114
   \newcommand{\donothing}{ }
115
116
```

3.3 Pesky enviroments

```
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
      \verb|\expandafter| provide command| csname env@\#1@process| endcsname {} % \\
      \verb|\difundefined{#1}{}| \\ \verb|\RenewEnviron{#1}{}| \\ \\
135
136
    \makeatother
137
    \label{lem:lemmand} $$\operatorname{\down}[1]_{\%}$
138
    \ifcnsr{}\else%
139
140
    \newenvironment*{hidden}{\begin{@empty}
    }{\end{@empty}}
142
    \voidenvironment{hidden}
143
144
    145
    146
    147
```

3.4 The cnsr command

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

```
\voidenvironment{tabular*}
164
                \voidenvironment{}
165
                % I don't know how many
                % 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
170
                 → marks floating above censored letters.
                171
                172
                173
                \mbox{renewcommand}(\\)[1]{}
                \rcmall \renewcommand \{\'\}[1]\{\}
175
                \mbox{renewcommand} \^{[1]{}}
176
                \mbox{renewcommand} \" \[ 1 \] \ \
177
                \mathbf{H}[1]
                179
                \rchin{l} \rch
180
                \rcmall \renewcommand \\ \k\}[1]\{\}
                 \mbox{renewcommand}\{\l\}[1]\{\}
                 \mbox{renewcommand} \= [1]{}
183
                \renewcommand{\b}[1]{}
184
                \renewcommand{\.}[1]{\}
                \mbox{renewcommand} \d [1] {}
186
                \rcmall \renewcommand \\r\[1]\{\}
187
                \renewcommand{\u}[1]{}
                \renewcommand{\v}[1]{}
189
                \mbox{renewcommand}\{\t\}[1]\{\}
190
                \mbox{renewcommand} \o\} [1] {}
                \mbox{renewcommand} \{i\}[1]\{\}
192
                193
```

3.5 Accessibility features

```
%but more importantly
204
   % it means screen readers don't subject
205
  \%. their users to the meaningless reading out of unicode black squares 50

→ times in a row!

   210
   \BeginAccSupp{method=plain,ActualText={TEXT REDACTED}}%
   \rndstring{#1}
   \EndAccSupp{}%
213
    \else%
215
   216
   217
   % if the conditional is off
   % the command does absolutely nothing
219
   220
   %3%3%3%3%3%3%3%3%
223
   224
```

3.6 The Lua magic

```
235
236 --fulsome thanks to TeX.SE users Henri Menke and David Carlisle, without whom

→ none of this would be possible
237 local function rndstring()
```

```
local toks = token.scan_toks(s)
239
             local on = true
240
     for n, t in ipairs(toks) do
241
         if t.csname == "begin" or t.csname == "end" then
242
             on = false
243
      -- The below is necessary as TeX primitives can break the code otherwise
      → because they do not use brackets
245
246
        if not(on) and t.cmdname == "right_brace" then
247
248
           -- This prevents needless errors about gibberish up commands
     end
250
      if on and t.csname == "&" then
251
       local letter = token.create'donothing'
252
      toks[n] = letter
253
254
255
257
      elseif on and t.csname == "%" then
258
       local letter = token.create'donothing'
259
      toks[n] = letter
261
     elseif on and t.csname == "$" then
262
       local letter = token.create'donothing'
      toks[n] = letter
265
     elseif on and t.csname == "#" then
266
       local letter = token.create'donothing'
267
      toks[n] = letter
268
      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
281
      elseif on and t.csname == "~" then
282
       local letter = token.create'donothing'
      toks[n] = letter
284
```

```
285
     elseif on and t.csname == "^" then
286
      local letter = token.create'donothing'
287
288
      toks[n] = letter
     elseif on and t.cmdname == "letter" then
289
      -- The below is the randomness part of this, which I admit is fairly
      → arbitrary, but will more often artificially shorten strings than lengthen
      \hookrightarrow them, as testing found if lengthening was too frequent, it led to really

→ unsightly long strings.

                 local f = math.random (1,20)
291
                  if f == 1 then
292
                     local letter = token.create'donothing'
      toks[n] = letter
295
                          elseif f == 2 then
296
                     local letter = token.create'donothing'
297
      toks[n] = letter
298
      elseif f == 3 then
299
        local letter = token.create'donothing' toks[n] = letter
      elseif f == 4 then
                     local letter = token.create'twothings'
302
       toks[n] = letter
        elseif f == 5 then
        local letter = token.create'donothing' toks[n] = letter
305
306
        else
                     local letter = token.create'onething'
308
      toks[n] = letter
309
                          end
                          elseif
311
                  on and t.cmdname == "spacer" then
312
                  local f = math.random (1,20)
                          if f == 2 then
314
                     local letter = token.create'donothing'
315
      toks[n] = letter
316
                                            elseif f == 3 then
317
318
                     local letter = token.create'donothing' toks[n] = letter
319
      elseif f == 4 then
      local letter = token.create'donothing'
321
       toks[n] = letter
322
                                              elseif f == 5 then
323
324
325
                     local letter = token.create'twothings' toks[n] = letter
326
      elseif f == 6 then
327
328
```

```
329
                     local letter = token.create'donothing' toks[n] = letter
330
                        elseif f == 7 then
331
        local letter = token.create'donothing' toks[n] = letter
332
334
        else
335
                     local letter = token.create'onething'
336
      toks[n] = letter
337
338
                           end
339
                           elseif
341
                  on and t.cmdname == "other_char" then
342
                  local f = math.random (1,20)
343
                          if f == 2 then
344
                     local letter = token.create'donothing'
345
      toks[n] = letter
346
                                             elseif f == 3 then
347
348
                     local letter = token.create'donothing' toks[n] = letter
349
      elseif f == 4 then
350
      local letter = token.create'donothing'
351
       toks[n] = letter
352
                                               elseif f == 5 then
353
354
355
                     local letter = token.create'twothings' toks[n] = letter
356
      elseif f == 6 then
357
358
359
                     local letter = token.create'donothing' toks[n] = letter
                        elseif f == 7 then
        local letter = token.create'donothing' toks[n] = letter
362
363
364
365
                     local letter = token.create'onething'
366
      toks[n] = letter
368
369
                           end
370
         end
371
      --Drop the token in and move on
372
             token.put_next(toks)
373
374
     local lft = lua.get_functions_table()
375
```

3.7 The warning option

```
382
   % WARNING FUN YAY
   384
   385
   386
   % Definitely this whole section
   %is there to be user modified, because
   % depending on language, jurisdiction
   %type of document etc, everyone will need
391
   %a specific warning style. So the important
   % part of the code here is the
   % conditional and global [warning]
   % option, because that's the magic value addedd
   397
   398
   % fonts for the warning:
   %I chose default LaTeX fonts
   % here to be changed as users wish
   405
   \newfontface\wrnstncl{QT Military}
407
   \verb|\newcommand{\warnword}{WARNING}|
   \newfontface\smbl{Deja Vu Sans Bold}
   \newcommand{\danger}{\smbl △\normalfont}
   \newcommand{\warnformat}{\sffamily\bfseries \color{red}}
411
   \verb|\newcommand{\textwarn}{%}|
412
```

```
This document is {\underline{NOT}}} redacted. It contains private and
413
     _{\rightarrow} confidential personal data, and may {\underline{NOT}}} be distributed,
        published, or shown to those without the right to view such information.
     _{\,\,\hookrightarrow\,\,} The publication of the information in this document may constitute a
     _{\rightarrow} contempt of court, punishable by a term of imprisonment.%
     }%
     \newcommand{\textsafe}{This document has been altered to remove sensitive
415
        personal data. It is cleared for publication and dissemination.
416
     \definecolor{darkgreen}{rgb}{0.0, 0.2, 0.13}
417
     \definecolor{darkspringgreen}{rgb}{0.09, 0.45, 0.27}
418
            \definecolor{forestgreen}{rgb}{0.13, 0.55, 0.13}
     \newcommand{\dquad}{\danger\danger\danger\danger}
     \newcommand{\dangersign}[1]{\scalebox{2}{\huge\danger}}
421
     \newcommand{\dangerblock}{\scalebox{2}%
422
     {\huge\danger\quad%
423
     \danger\quad\danger}}
424
     \newcommand{\warnblock}{{\Large\wrnstncl%
425
     \warnword\quad\warnword%
     \quad\warnword}}
427
     \newcommand{\tworules}{\hrule width \hsize height .7pt\vskip2pt\hrule width
428
     → \hsize height .7pt}
     \newcommand{\allwarning}{%
429
     \dangerblock\\\warnblock\\%
430
     \normalfont\smallskip\warnformat\textwarn%
431
432
     \newcommand{\confwarning}{%
433
     434
     435
     % The warning option
436
     437
     %%%%%%%%%%%%%%%%%%%%
438
     \ifwarning
439
     \ifcnsr
440
     441
     % a note saying document is redacted
443
     444
     %%%%%%%%%%%%%%%%%%%%
     \begin{center}
446
     \color{forestgreen}
447
     \tworules\vskip5pt
448
     \normalsize\normalfont\sffamily\bfseries\textsafe
449
     \vskip5pt\tworules
450
     \end{center}
451
     \else
     453
```

```
454
   % The WARNING for un-redacted docs
455
   457
   \begin{center}\color{red}\tworules\vskip 5pt\allwarning
    \vskip5pt\tworules
    \end{center}
   \fi
461
   \else\fi
462
463
   \AtBeginDocument{\confwarning}\clearpage
464
```

4 Version History

4.1 1.1.0

22 February 2022: Package creation

4.2 1.0.0

18 February 2022: Package creation