

高亮 T_EX 和 L^AT_EX3 代码——使用 texhigh 宏包

霽月*

2025 年 8 月 22 日 v0.3.3

`texhigh` 宏包¹是专用来高亮 \TeX 文件的宏包。基于由 Rust 编写的命令行工具 `texhigh`²，处理 1.29MB 左右 (39,300 余行) 的 `expl3-code.tex` 只需 0.18s 左右 (操作系统为 Windows, CPU 为 i7-12700H)，处理速度约为 `minted` 宏包使用的 `pygmentize` (约 3.4s) 的 19 倍，`texhigh` 的增强模式也比它快 10 到 16 倍 (约 0.21s)。对于普通大小的 \TeX 代码，处理它们所需的时间相比于 \TeX 文件本身编译所需的时间，已经可以忽略不计。

`texhigh` 主要是在 \LaTeX 中为 `texhigh` 命令行工具提供交互接口。这要求在编译 \TeX 文件时启用 `--shell-escape`。

texhigh 除了可用于高亮 T_EX 文件，还支持计算文字的布局。基于此特性，texhigh 提供了输出颜文字的功能： $\varepsilon(\top\sim\top)3$ ，只需使用 `\kaomoji{颜文字}`。默认使用系统字体，也可自行设置 `\`($\geq\nabla\leq^*$)o`。

$\partial(\varphi^*\omega)\varphi^*\frac{\partial}{\partial t} = \partial(\omega) + L_{\varphi_*}\omega$

使用颜文字时可能会遇到字体问题，这时在字符间插入零宽词连接符 U+2060 或可解决。

使用 `\kaomoji*` 还支持把单行文字输出为图片:

% 这里的 fontsize 影响图片的大小, 从而影响清晰度

```
\kaomoji*[fontsize=\Huge]{🐼🐻🐱🐰🐉🐲🐮🐷🐽🐾🐾🐾} ←  
⇨ \includegraphics[height=12bp]{  
  
\kaomoji*[fontsize=50bp]{\Uchar"1F43C }{\includegraphics[height=25bp]{
```

代码 1



也可以自己封装一下这个命令：

*longaster@163.com

¹<https://github.com/Sophanatprime/texhigh>

²<https://github.com/Sophanatprime/texhigh-rs>

代码 2

```

\makeatletter
% fonts 键用于设置额外的字体。texhigh 会查找系统字体，一般无需另行设置
\NewDocumentCommand\inmoji{ D<>{\f@size\p@} ={\fontsize+} 0{} m }
  {\kaomoji*[fontsize={(\#1)*3},\#2]{\#3}
    {\includegraphics[height=\dimeval{\#1}]{}}}
\makeatother
\inmoji{🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒} \inmoji{^o^y}

```

🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒🕒 ^o^y

texhigh 提供 `\texhighverb`、`\texhighfile`、`\texhighinput` 这几个命令以及一个 texhigh 环境用于高亮 TeX 代码。

`\texhighverb` 用法和 `\verb` 类似，但没有带星号的版本，它不能作为其它命令的参数；
`\texhightext` 用于高亮文字，一般用于高亮已经处理过的文本，和 `\texhighverb` 相比，它可以作为命令的参数。`\texhighfile` 用于高亮一个文件，`\texhighinput` 则用于导入一个已经被处理过的文件。

texhigh 还有很强的可配置性。

为了实现处理 TeX 源码与输出结果的分离，texhigh 使用“类型”和“类别”来区分不同的记号。字符和控制序列是不同的“类型”，控制序列之间可以有不同的“类别”，例如是原语、 \LaTeX 函数等。类型不可改变，而“类别”可以自由修改。

每个类型都有一些命令用于更改它们的“类别”的显示效果，如，对于一个控制序列，可以使用 `\THSetClassCS` 改变显示效果。可以为它们设置前景色、背景色，甚至渐变色和底纹等等。实际上普通文字可以显示成什么效果，它们就可以做到同样的效果。具体修改方式可以参考文末 basic 样式的源码。

texhigh 利用 tikz 实现了渐变和底纹效果，同时也可直接集成到 tcolorbox 宏包中。只需要在加载 texhigh 之前加载这几个宏包。

代码 3

```

\usepackage{tikz}
\usepackage{tcolorbox}
\usepackage{texhigh}
\tcbset{listing engine=texhigh} % 使用这个即可切换至 texhigh
% 若使用 xeCJK，即在 XeLaTeX 中使用 ctex，最好设置
\SetKeys[texhigh/high]{
  font=\ttfamily\xecjksetup{CJKeclue={\hskip 0pt plus 0.08\baselineskip}}
}
% 这样可避免在显示代码时中英文之间出现不必要的空格。

```

识别行内数学公式：

`\texhighverb`!公式 $\int_a^b x^2 dx = \frac{1}{3} x^3 \Big|_a^b$ 。

代码 4

公式 $\int_a^b x^2 dx = \frac{1}{3} x^3 \Big|_a^b$ 。

渐变:

`\texhighverb`[style=tikz.gradient, use-ctab=latex3,
 \hookrightarrow config-file=config.cfg]|\sys_get_shell:nnNTF|

代码 5

`\sys_get_shell:nnNTF`

底纹:

`\makeatletter`
% #1: tikz options, #2: text
`\def\myshadetext#1#2{\texhigh@shadetext{#1}{\bfseries #2}}`
`\makeatother`
`{\LARGE`
% 在加载 texhigh 之前加载 tikz 宏包!
% 使用 grass.png 作为文字底纹, 依赖 tikz 的 fill.image 库, 会自动加载这个库。
`\texhighverb`[use-ctab=latex3, this-cs=`\myshadetext`{fill stretch
 \hookrightarrow image=grass.png}]
|\sys_get_shell:nnNTF|
}

代码 6

`\sys_get_shell:nnNTF`

中文命令识别 (TeX 原语带有下列线):

`\begin{texhigh}`[output=`\jobname.texhigh`, use-ctab=cjk]
`\def\好好好{中文 Good}`
`\好好好\relax`
`\end{texhigh}`

代码 7

`\def\好好好{中文 Good}`
`\好好好\relax`

类别码混合使用:

% 自动检测 `\makeatletter` 和 `\ExplSyntaxOn` 块,
% `\makeatletter` 和 `\ExplSyntaxOn` 必须在行首, 前面可以有空格
`\begin{texhigh}`[lexer-catcode={atletter, explon}]
`\def\foo@#1{[#1]} \foo@{F00} \@kernel`
`\makeatletter`

代码 8

```

\def\foo@:#1{[#1]} \foo@:#1{F00} \@kernel \scan_stop:
\ExplSyntaxOn
\cs_set:Npn \foo@: #1 { [#1] }
\foo@: {F00} \@kernel \scan_stop:
\ExplSyntaxOff
\@kernel \scan_stop:
\makeatother
\@kernel \scan_stop:
\end{texhigh}

```

```

\def\foo@#1{[#1]} \foo@{F00} \@kernel
\makeatletter
\def\foo@:#1{[#1]} \foo@:#1{F00} \@kernel \scan_stop:
\ExplSyntaxOn
\cs_set:Npn \foo@: #1 { [#1] }
\foo@: {F00} \@kernel \scan_stop:
\ExplSyntaxOff
\@kernel \scan_stop:
\makeatother
\@kernel \scan_stop:

```

也可手动调整类别码:

代码 9

```

\begin{texhigh}[
  lexer-catcode*={3}{9}{@=11, ?=11}, % [3, 9) 行, @ 和 ? 的类别码为 11
  % lexer-catcode*={3}{9}{ @?=11 }, % <- 可以合并
  lexer-catcode*={5}{7}{explon},      % [5, 7) 行, 启用 expl3 的类别码
]
\def\foo@#1{[#1]} \foo@{F00} \@kernel
\makeatletter
\def\foo@:#1{[#1]} \foo@:#1{F00} \@kerne? \scan_stop:
\ExplSyntaxOn
\cs_set:Npn \foo@: #1 { [#1] }
\foo@: {F00} \@kernel \scan_stop:
\ExplSyntaxOff
\@kernel \scan_stop:
\makeatother
\@kernel \scan_stop:
\end{texhigh}

```

```

\def\foo@#1{[#1]} \foo@{F00} \@kernel
\makeatletter
\def\foo@:#1{[#1]} \foo@:#1{F00} \@kerne? \scan_stop:
\ExplSyntaxOn

```

```

\cs_set:Npn \foo@: #1 { [#1] }
\foo@: {F00} \@kernel \scan_stop:
\ExplSyntaxOff
\@kernel \scan_stop:
\makeatother
\@kernel \scan_stop:

```

除了用 `use-ctab` 设置主类别码表以外，还可以添加额外的类别码表，优先选择靠后的类别码表：

```

\texhighverb[extra-catcode*=cjk, extra-catcode*=explon]
  {\cs_set:Npn \a_好的: #1 { aaa } \a_好的:}

% extra-catcode* 实际相当于设置 lexer-catcode* 的前两个值为 {0}{ }
% 即，从第 0 行开始，永不结束
\texhighverb[lexer-catcode*={0}{ }{cjk}]{\def\a好的#1{aaa} \a好的}

\cs_set:Npn \a_好的: #1 { aaa } \a_好的:
\def\a好的#1{aaa} \a好的

```

代码 10

还能进一步细化到列数：

```

% 1 相当于 [1, 0],
% [1, 9] 表示第 1 行第 9 个字符
% 这里就是从第一行的开始，直到第一行第 9 个字符（不含），空格的类别码为 0
% 设置类别码时，特殊字符必须转义，其它字符可转义也可不转义
\texhighverb[lexer-catcode*={1}{[1,9]}{\ =0}]{abd def #1{\space }}
%
%      ~~~转义      ~~~空格的类别码不再是 0

\texhighverb{abd def #1"\space"}

abd_def_#1{\space }
abd def #1"\space"

```

代码 11

`lexer-catcode*` 的前两个参数还支持更加复杂的模式，如纯文本正则表达式和 $\text{T}_{\text{E}}\text{X}$ 正则表达式。

例如检测 `\verb` 命令：

```

\texhighverb[lexer-catcode*={\c{verb}\*?\\}{\\}{str}]
  {a \verb|\macro in \verb| command \verb*|\macro in \verb| \macro}

a \verb|\macro in \verb| command \verb*|\macro in \verb| \macro

```

代码 12

纯文本正则表达式就是针对纯文本的正则表达式，日常见到的正则表达式都是这一类，tex-high 支持的纯文本正则表达式的完整语法见 <https://docs.rs/regex/latest/regex/#syntax>。

TeX 正则表达式是针对 TeX token 的正则表达式，L^ATeX 的 l3regex 就是这类，texhigh 支持的 TeX 正则表达式语法和 l3regex 基本一致，但暂不支持 \b \B \G \u 这几个转义序列，以及 \c 转义序列的否定形式（即暂不支持 [^\c{begin}\c{end}] 这类用法）。

代码 13

```
% \I 放在开头表示这是纯文本正则表达式，匹配源码
\begin{texhigh}[gobble=2, % 每行删除前两个字符
  lexer-catcode*=
    {\I\catcode\!=11[\s\]}
    {\I\endgroup\s}
    {\!=11},
]
\def\!mark#1{MARK1} \!mark
\begingroup
  \catcode\!=11
  \def\!mark#1{MARK2} \!mark
  \endgroup \!mark
\endgroup
\!mark
\end{texhigh}
```

```
\def\!mark#1{MARK1} \!mark
\begingroup
  \catcode\!=11
  \def\!mark#1{MARK2} \!mark
  \endgroup \!mark
\endgroup
\!mark
```

代码 14

```
% \T 放在开头表示这是 TeX 正则表达式，匹配 token
% \T 大多数时候可以省略，但当模式以数字或 [ 开头
% 时，\T 不可省略，否则被当作行号
\begin{texhigh}[gobble=auto, % 检测空格并删除
  lexer-catcode*=
    {\T\c{catcode}\!\c{!}\=11} % 注意字符转义
    {\c{endgroup}}
    {\!=11},
]
\def\!mark#1{MARK1} \!mark
\begingroup
  \catcode\!=11
  \def\!mark#1{MARK2} \!mark
  \endgroup \!mark
\endgroup
\!mark
\end{texhigh}
```

```
\def\!mark#1{MARK1} \!mark
\begingroup
  \catcode\!=11
  \def\!mark#1{MARK2} \!mark
  \endgroup \!mark
\endgroup
\!mark
```

lexer 也可混合使用正则表达式和行数：

代码 15

```
\begin{texhigh}[gobble=auto,
  lexer-catcode*={(\|cJ.|\\n)\c{ExplSyntaxOn}}{4}{\:\_ =11},
]
  \ExplSyntaxOn
    \cs_set:Npn \expl_off: { \ExplSyntaxOff } \expl_off:
  \ExplSyntaxOff
  \expl_off:
\end{texhigh}
```

```
\ExplSyntaxOn
  \cs_set:Npn \expl_off: { \ExplSyntaxOff } \expl_off:
\ExplSyntaxOff
\expl_off:
```

可用使用 `lines` 设置源文件需要保留的行数，`gobble=auto` 检测空格时只会检测保留下来的代码行：

代码 16

```
\begin{texhigh}[gobble=auto, % 检测空格并删除
  lines={2,8}, % 只保留 [2, 8) 行
  lexer-catcode*=
    {\T\c{catcode}\|c{!}\ =11} % 注意字符转义
    {\c{endgroup}}
    {\ !=11},
]
  \kill this line
\def\!mark#1{MARK1} \!mark
\begingroup
  \catcode`\ !=11
  \def\!mark#1{MARK2} \!mark
  \endgroup \!mark
\endgroup
\!mark
\end{texhigh}
```

```
\def\!mark#1{MARK1} \!mark
\begingroup
  \catcode`\ !=11
  \def\!mark#1{MARK2} \!mark
  \endgroup \!mark
\endgroup
```

代码 17

```
% \THSetCharReplacement{\ }{\textvisiblespace} % texhigh 已设置
\THSetCharReplacement{\$}{\S} % 把 $ 替换为 \S
\texhighverb[char-replacements={\ \$}, % 设置哪些字符要替换
  char-category*={symbol}{[\ \$]}{\mbox{\color{red}#1}}, % 修改颜色
] {\def\!a #1{$#1$} \!a {\ b}}
```

```
\def\!a #1{$#1$} \!a {\ b}
```

控制序列的名称里的字符也会被替换：

代码 18

```
\texhighverb[char-replacements={a=A, b=B, {=}{,}}]{abcd=$\abcd} % 一步到位
% 字符替换无需启用增强模式
\texhighverb[enhanced=false, char-replacements={a=A, b=B, {=}{,}}]
{abcd=$\abcd}
```

ABcd,\$\ABcd ABcd,\$\ABcd

char-category 也可以用来替换字符，但不会替换控制序列名称里的字符：

代码 19

```
\ExplSyntaxOn
\cs_new_protected:Npn \chartouni #1 { \fbox{ \int_to_Hex:n { `#1 } } }
\ExplSyntaxOff
\texhighverb[
% 这里使用正则表达式查找字符的类别，下面的正则表达式匹配是 Emoji 但不是
ASCII 的字符
char-category*={emoji}{[\p{Emoji}--\p{ASCII}]}{\chartouni{#1}\_}
] {Emoji: 🐼 🐼 🐼 🐼 🐼 🐼 🐼 🐼 🐼 🐼 🐼 🐼 🐼 🐼 }
```

Emoji: 1F400 1F403 1F405 1F407 1F409 1F40D 1F40E 1F410 1F412 1F413
1F415 1F416

可以使用 line-number 选项来为代码加上行号，first-line-number 可以设置起始行号。

代码 20

```
\begin{texhigh}[gobble=auto, line-number, first-line-number=42,
line-number-format={\color{gray}{0.5}\scriptsize\sffamily #1},
left-space=5mm, right-space=5mm,
]
\documentclass{article}
\usepackage[paper=a4,hmargin=2cm]{geometry} % 页面设置
\usepackage{fancyhdr} % 页眉
\begin{document}
Hello, \LaTeXe.
\end{document}
\end{texhigh}
```

42	\documentclass{article}	42
43	\usepackage[paper=a4,hmargin=2cm]{geometry} % 页面设置	43
44	\usepackage{fancyhdr} % 页眉	44
45	\begin{document}	45
46	Hello, \LaTeXe.	46
47	\end{document}	47

texhigh 支持与 listings 和 minted 类似的在高亮时让某些记号保持其原有作用的特性。

代码 21

```
\begin{texhigh}[texcl, escape-inside=||, escape-inside=\ES, gobble=auto]
% This \textbf{comment} line \emph{will be} escaped.
This \textbf{text} line \ES{\emph{won't} be} escaped.
This |\textbf{will}| be escaped.
\end{texhigh}
```

This **comment** line *will be* escaped.
 This \textbf{text} line *won't* be escaped.
 This **will** be escaped.

代码 22

```
\begin{texhigh}[gobble=auto, texcl, math-escape, % 允许输出行内公式
linenos, first-line-number=1, % texhigh 不会自动重设行号
left-space=5mm, line-number-pos=left, % 行号位置: left, right, both
line-number-space=2mm, % 修改行号与代码的间距
line-number-format={\small\color{red}[#1]}, % 修改样式, #1 参数为行号
the-line-number=\Alph{TeXHighLine}, % 修改 \theTeXHighLine
]
A expression $\displaystyle\int\sin x\mathrm{d}x=-\cos x+C$.
Another equation \(\sum\limits_{n=0}^{+\infty}\frac{1}{n^2}=\frac{\pi^2}{6}\).
% The text \emph{will} be printed, so is the $E^2=(pc)^2+(m_0c^2)^2$
% formula.
\end{texhigh}
```

[A] A expression $\int \sin x dx = -\cos x + C$.

[B] Another equation $\sum_{n=0}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

[C] The text *will* be printed, so is the $E^2 = (pc)^2 + (m_0c^2)^2$ formula.

代码 23

```
\begin{texhigh}[gobble=auto, line-number, line-number-pos=right,
right-space=5mm, comments-math-escape, % 只允许注释内的数学公式
]
A expression $\displaystyle\int\sin x\mathrm{d}x=-\cos x+C$
A expression % $\displaystyle\int\sin x\mathrm{d}x=-\cos x+C$
A expression % \(\displaystyle\int\sin x\mathrm{d}x=-\cos x+C\)
\end{texhigh}
```

A expression $\displaystyle\int\sin x\mathrm{d}x=-\cos x+C$ 4

A expression % $\int \sin x dx = -\cos x + C$ 5

A expression % $\int \sin x dx = -\cos x + C$ 6

事实上，它们有更通用的写法：

代码 24

```
\DeclareDocumentCommand\cs{0{\texttt}m}
  {#1{\textbackslash\detokenize{#2}}}%
\DeclareDocumentCommand\pkg{m}{\textsf{#1}}
\begin{texhigh}[gobble=auto,
% start 用于标记何时开始，为 TeX 正则表达式或纯文本正则表达式或纯文本，
% 使用正则表达式时一定要带着开始锚点 (^)!
% arguments 用于指示它获取的参数，像 ltcmd (xparse) 定义命令时使用的一样
% in-comments 用于表示它是否处于注释内，有三个可选值：must、never、any
range={cs}{escape, start=~\c{cs}, arguments=om, in-comments},
range={pkg}{escape, start=~\c{pkg}, arguments=m},
]
% The \cs{c_true_bool} is true in \pkg{l3bool}.
The \cs{c_false_bool} is false in \pkg{l3bool}.
\end{texhigh}

% The \c_true_bool is true in l3bool.
The \cs{c_false_bool} is false in l3bool.
```

`\THSetRange` 可以用来设置代码段 (range) 的格式。用法如下：

```
\THSetRange{<range id>}[<range settings>]{<begin code>}[<end code>]
```

当不给出 `<range settings>` 时，只会修改样式，而不会捕获代码段。

代码 25

```
\THSetRange{usepackage}
[start=~\c{usepackage}, arguments=omo]
{\THSetRange{argument.o}{\THcolor{blue}}%
\THSetRange{argument.m}{\THcolor{cyan}\bfseries}}
% \usepackage{hologo}
% remove-start 用于移除 start 的内容，这里 start 捕获 \AMS，
% 把 \AMS 移除，然后输出我们设置的内容，这里是 \hologo{AmS}
\THSetRange{logo}[escape, start=~\c{AMS}, remove-start]{\hologo{AmS}}

\begin{texhigh}[gobble=auto,]
\usepackage[width=210mm,height=297mm]{geometry}
\usepackage{amsmath,amsthm} % \AMS packages
\usepackage[xdvipdfmx]{color}[2025/06/01]
\end{texhigh}

\usepackage[width=210mm,height=297mm]{geometry}
\usepackage{amsmath,amsthm} % \AMS packages
\usepackage[xdvipdfmx]{color}[2025/06/01]
```

arguments 包括 `m o O d D r R s t v` 以及 `l u g G` 暂不支持 `e E c b`，并且 `! + = >` 也是无效的。此外，还支持一个特殊的符号 `^^J`，它用于捕获当前行剩下的内容，带一个参数，表

示是否要求大括号成对存在。

代码 26

```
\THSetRange{special}
% \THmN 是一个特殊的标记命令，用来表示 ^^J。
% ^^J 的参数 “1” 表示大括号必须成对存在。
% 由于 ^^J 捕获了当前行的剩余的内容，这也包括换行，这会导致
% 换行出现在 range 内部，本例中问题不大，但某些自定义的 range 就不一定了
% insert-ending 用来把定界符（这里就是换行符）“返还”出来
[start=^{\c{SP}}, arguments=1 m \THmN{1}, insert-ending]
{\THSetRange{argument.l}{\ensuremath{\lvert}}{\ensuremath{\rvert}}}%
\THSetRange{argument.m}{\ensuremath{\langle}}{\ensuremath{\rangle}}%
\THSetRange{argument.^~J}{\ensuremath{\lVert}}{\ensuremath{\rVert}}}]

\begin{texhigh}[gobble=auto]
  A \SP special {arguments} to the end.
  Another \SP fancy {parameter} of the line.
\end{texhigh}

A \SP| special |{arguments}|| to the end.||
Another \SP| fancy |{parameter}|| of the line.||
```

我们可以捕获各类括号和单词等：

代码 27

```
% 注意正则表达式要转义括号，start-is-arg 会把 start 也看做是参数的一部分
\THSetRange{bracket}[start={^\[}], arguments={r[]}], start-is-arg]
  {% 可进一步处理 [ 和 ]
    \THcolor[HTML]{007f00}}
\THSetRange{paren}[start={^\(}, arguments={r()}], start-is-arg]{ ←
↪ \THcolor[HTML]{7f007f}}
% \THSetRange{brace}[start={^{\{}}, arguments={m}], start-is-arg]{ ←
↪ \THcolor{purple}\bfseries}
\THSetRange{tikz-keywords}
  [start={\I^(tikzpicture|pic|controls|and)},]
  {\bfseries}
\makeatletter
% 定义一个样式，range-config 就是 \THSetRange 的第二个参数的键所属的路径
\texhighdefstyle[range-config]{tikz-keyword}
% 使用 insert-brace 会把所有参数放到一个大括号里
% 使用 u 参数捕获 #1，然后把它放到括号里
{escape, start={\L#1}, arguments=u{#1}, start-is-arg, insert-brace}
\THSetRange{[cycle]tikz-keywords}
  [tikz-keyword=cycle]
% 最终得到 \bfseries \texhigh@underline{cycle}
{\bfseries \texhigh@underline}
```

```

\THSetRange{[node]tikz-keywords}[tikz-keyword=node]{\bfseries ←
↪ \texhigh@underline}
\makeatother
\THSetClassCS[] {tikzcs}{\mbox{\THcolor{blue}\bfseries #1#2}}

\begin{texhigh}[gobble=auto,
  cs-category={tikzcs}{draw,fill,filldraw,path,node,coordinate}{\THPASS},
]
  \begin{tikzpicture}[scale=2]
    \draw[red] (0,0) -- +(1,1);
    \path[fill] (2,1) -- (2,2) -- (0,1) -- cycle;
    \draw (0,0) .. controls (1,1) and (2,1) .. (2,0) node {$A$};
  \end{tikzpicture}
\end{texhigh}

\begin{tikzpicture}[scale=2]
  \draw[red] (0,0) -- +(1,1);
  \path[fill] (2,1) -- (2,2) -- (0,1) -- cycle;
  \draw (0,0) .. controls (1,1) and (2,1) .. (2,0) node {$A$};
\end{tikzpicture}

```

我们可以自定义注释的显示方式，就是使用 `^^J` 参数：

代码 28

```

\THSetRange{comment-1}
  [start=\I^{\%}\s, arguments=\THmN{0}, insert-ending]
  {\THcolor{gray}{0.5}\THSetPlainStyle{cs,color}{}%
  [{\color{yellow!70!green}\rmfamily\bfseries \dotfill SC.}]
\THSetRange{comment-2}
  [start=\I^{\%e\_, escape, arguments=\THmN{1}, remove-start, ←
↪ use-argument, insert-ending]
  {\THcolor{red}\normalfont \textbf{SP: }}
\THSetRange{comment-3}
  [start=\I^{\%a, escape, arguments=m \THmN{1}, remove-start, insert-ending]
  {\def\grab#1{{\rmfamily\bfseries #1:\_}\ignorespaces}%
  \THcolor{purple}\sffamily \grab}
\THSetRange{comment-4}
  [start={\L//\THmS}, arguments=\THmN{0}, insert-ending]
  {\THcolor{green!40!black}\textrm{C style comment: }}

\begin{texhigh}[gobble=auto]
  This is normal % comment(1).
  This is %e Some comment 2.
  %a{comment 3} is the line.

```

```
// This is comment 4.
\end{texhigh}
```

```
This is normal % comment(1).....SC.
This is SP: Some comment 2.
comment 3: is the line.
C style comment: // This is comment 4.
```

由于 range 在捕获时不能嵌套捕获，使用自定义的注释会导致它内部无法捕获其它 range 了。一个解决办法是将此 range 设置为 escape，然后内部再使用 \texhighverb。

代码 29

```
\THSetRange{mycomment}
% \THmS 表示空格
[start={\L//\THmS}, escape, arguments=\THmN{1}, remove-start, ←
→ insert-brace, insert-ending]
{% \appendpercent “参数处理器” 把 \THmP (即 “%”) 放到参数的前面
\def\appendpercent#1{\edef\ProcessedArgument{\THmP\unexpanded{#1}}}%
\DeclareDocumentCommand{\mycommentverb}{>{\appendpercent} v}
{\texhighverb[remove-enabled-ranges={mycomment}]{#1}}%
% 如果是类别码为 14 且为是第一个 “%” 就换回 “//\THmS”
\def\firstComment#1{\if#1\THmP\relax //\THmS \def\firstComment{}} ←
→ \else #1\fi}%
\THSetClassCH[] {catcode.14}{\firstComment{#1}}%
\normalfont \mycommentverb}
\begin{texhigh}[gobble=auto, escape-inside=|]
The |\LaTeX| % comment can be |\emph{highlighted}|.
// also can be |\emph{highlighted}|, but isn't |\LaTeX|.
\end{texhigh}
```

```
The LATEX % comment can be highlighted.
// also can be highlighted, but isn't LATEX.
```

这个例子展示了如何检测并替换命令中的 @@ 符号。

代码 30

```
\ExplSyntaxOn
\tl_new:N \l__this_atat_tl
\cs_new:Npn \__this_atat_format:w #1 >
{
\tl_set:Nn \l__this_atat_tl {#1}
@@ = \l__this_atat_tl
}
\cs_new:Npn \__this_atat_cs_internal:nn #1#2
{
\group_begin:
```

```

\str_set:Nn \l_tmpa_str {#2}
\tl_replace_once:Nee \l_tmpa_str
{ \tl_to_str:n { _ @@ } } { _ _ \l__this_atat_tl }
\THcolor { purple } #1 \l_tmpa_str
\group_end:
}
\cs_new:Npn \__this_atat_cs_public:nn #1#2
{
\group_begin:
\str_set:Nn \l_tmpa_str {#2}
\tl_replace_once:Nee \l_tmpa_str
{ \tl_to_str:n { @@ } } { \l__this_atat_tl }
\THcolor { red } \bfseries #1 \l_tmpa_str
\group_end:
}
\THSetRange * {change-at} % 带 * 的不会自动加上 \begingroup \endgroup
[start={\I^<@@=}, escape, arguments=u{>}, remove-start]
{ \ensuremath{\langle} \__this_atat_format:w }
[ \ensuremath{\rangle} ]
\THSetClassCS[] {atat-i}{ \__this_atat_cs_internal:nn {#1} {#2} }
\THSetClassCS[] {atat-p}{ \__this_atat_cs_public:nn {#1} {#2} }
\ExplSyntaxOff

\begin{texhigh}[gobble=auto, use-ctab=latex3code,
cs-category*={atat-i}{_@@}{\THPASS}, % 注意要先匹配 _@@
cs-category*={atat-p}{@@}{\THPASS}, % 否则只会匹配到 @@
]
<@@=mycs>
\tl_new:N \l_@@_code_tl
\cs_set:Npn \@@_code: { \scan_stop: \l_@@_code_tl }
<@@=yourcs>
\tl_new:N \l_@@_code_tl
\cs_set:Npn \@@_code: { \scan_stop: \l_@@_code_tl }
\end{texhigh}

<@@=mycs>
\tl_new:N \l__mycs_code_tl
\cs_set:Npn \mycs_code: { \scan_stop: \l__mycs_code_tl }
<@@=yourcs>
\tl_new:N \l__yourcs_code_tl
\cs_set:Npn \yourcs_code: { \scan_stop: \l__yourcs_code_tl }

```

\THSetRange 的 start* 还支持嵌套，就是在 start* 中使用 start/start*。实际上，当外层的 start/start* 匹配成功后，就会立刻停止，如果它的 arguments 无法匹配，那么这个匹配

就是无效的，即使之后某些的 `start` 可以匹配成功，`texhigh` 也不会进行后续的匹配了。

但如果在 `start/start*` 中再使用 `start/start*` 和 `arguments`，假定内层的没有匹配成功或者无效，`texhigh` 会跳过这个，再继续寻找后面的 `start`。也就是说，如果 $\langle inner\ arguments \rangle$ 和 $\langle arguments \rangle$ 相同，那么匹配尽管不一定成功，但总是有效的。

代码 31

```
\THSetRange{sentence}
  [start*={ start='\I^\w', arguments='u{.}', start_is_arg=true }, ←
  → arguments={u{.}}, start-is-arg]
  {\THcolor{red}}
\THSetRange{half}
  [start*={ start='\I^\w', arguments='u{,}', start_is_arg=true }, ←
  → arguments={u{,}}, start-is-arg]
  {\THcolor{blue}}
\THSetRange{neither}
  [start*={ start='\I^\w', arguments='\THmN{f}' }, arguments={\THmN{0}}, ←
  → start-is-arg]
  {\THcolor{green!50!black}}[!]
```

`\texhighverb{a text without period}` `\texhighverb{a text without comma}`
`\texhighverb{a text with period.}` `\texhighverb{a text with comma,}`

a text without period! a text without comma!
 a text with period. a text with comma,

这个例子中 `start` 都是纯文本正则表达式 `\I^\w`，真正起作用的是它们使用的不同的参数，`texhigh` 首先检测是否匹配 `sentence` 的 `start` 和 `arguments`，然后检测 `half`，然后 `neither`。

再观察下面这个例子，没有内层的 `start`，由于最开始的 `sentence` 能够匹配，尽管其参数可能无效，`texhigh` 也不会再检测 `half` 和 `neither` 了。

代码 32

```
\THSetRange{sentence}
  [start=\I^\w, arguments={u{.}}, start-is-arg]
  {\THcolor{red}}
\THSetRange{half}
  [start=\I^\w, arguments={u{,}}, start-is-arg]
  {\THcolor{blue}}
\THSetRange{neither}
  [start=\I^\w, arguments={\THmN{0}}, start-is-arg]
  {\THcolor{green!50!black}}[!]
```

% 由于 arguments 匹配失败，texhigh 会报警告


```
\texhighverb{a text without period} \texhighverb{a text without comma}
\texhighverb{a text with period.} \texhighverb{a text with comma,}
```

```
a text without period a text without comma
a text with period. a text with comma,
```

不过要注意的是 `start*` 不能随意添加逗号, 且每个参数都不能少, `start*` 和 `arguments` 的参数单引号也不能少。 `start*` 有如下的几个情形, 必须严格按照给出的用法来写:

```
start*={ start='⟨start⟩', arguments='⟨arguments⟩', start_is_arg=⟨bool⟩ }
```

```
start*={ start={⟨start*⟩}, arguments='⟨arguments⟩', start_is_arg=⟨bool⟩ }
```

```
start*={ literal='⟨literal⟩' } 相当于 start={\L⟨literal⟩}
```

```
start*={ regex='⟨regex⟩' } 相当于 start={\I⟨regex⟩}
```

```
start*={ regtex='⟨regtex⟩' } 相当于 start={\T⟨regtex⟩}
```

`texhigh` 还能捕获某个 `range` 的内容, 就是用 `\THCollectRange`, 用法如下:

```
\THCollectRange{⟨do code⟩} \THrs{⟨range id⟩}... \THre{⟨range id⟩}
```

```
\THCollectRange{⟨do code⟩} \THes{⟨escape id⟩}... \THEe{⟨escape id⟩}
```

其中 `⟨do code⟩` 可以使用 3 个参数, 第一个为 `range/escape` 的 `id`, 第二个为 `range/escape` 的内容, 第三个为 `range/escape` 的原始内容。

代码 33

```
\THSetRange{keyword}
```

```
% 由于 texhigh 的实现方式所致, 如果 \b 断言出现在最开始的位置, 则无法生效,
% 需要使用 skip-if-pre 来跳过无效的匹配, 如果 skip-if-pre 或 skip-if-post
% 有一个匹配成功的话, 就会跳过该 range, texhigh 会继续找可能的匹配。
% skip-if-pre 检查 start 之前的内容, skip-if-post 检查 start 之后的内容,
% 它们都是纯文本正则表达式。
```

```
[start*={ regex='^\\b(and|or|not)\\b', in-comments=never, skip-if-pre= ←
→ \\w$ ]
```

```
{\THCollectRange{(styled: {\color{red}#3}, text: \detokenize{#2})%}
```

```
% 与 \THrs{keyword} 配对的 \THre{keyword} 会被移除, 需要补回来,
```

```
\THre{keyword}%
```

```
% 当代码执行到此处时, 原来的 \THrs{keyword} 已经被执行了, 需要补充回来
```

```
% 而 \THre{keyword} 还留在后边, 可以正确捕获, 成功后它也会被移除
```

```
\THrs{keyword}%
```

```
}
```

```
\texhighverb{and} \texhighverb{rand} \texhighverb{android}
```

```
\texhighverb{not} \texhighverb{cannot} \texhighverb{or} \texhighverb{nor}
```



```
(styled: and, text: and) rand android (styled: not, text: not) cannot (styled:
or, text: or) nor
```

有时我们不需要预定义的样式而只使用自己的样式，texhigh 提供了 `\THRemoveClass` 和 `\THRemoveClasses` 这两个命令。可用的类型为 `bp`, `cs`, `ch`, `rs`, `re`, `st`, `es`, `ee`, `pn`。

代码 34

```
\texhighverb{\makebox[\dimeval{5pt+5pt}]{\hbox{\LaTeX}}}{
\THRemoveClass{cs}{latex.document}% 删除控制序列类型中类别为
latex.document 的样式
\texhighverb{\makebox[\dimeval{5pt+5pt}]{\hbox{\LaTeX}}}{
\THRemoveClasses{cs}% 删除控制序列类型中除默认样式之外的所有样式，默认样式名
为“?”，可用 \THRemoveClass{cs}{?} 强制删除。
\texhighverb{\makebox[\dimeval{5pt+5pt}]{\hbox{\LaTeX}}}{
\makebox[\dimeval{5pt+5pt}]{\hbox{\LaTeX}}
\makebox[\dimeval{5pt+5pt}]{\hbox{\LaTeX}}
\makebox[\dimeval{5pt+5pt}]{\hbox{\LaTeX}}
```

texhigh 默认的定义以及部分命令的用法可参考 `texhigh.prelude.ths` 文件。

```
%%%---- File: texhigh.prelude.ths ----%%%
\ProcessKeyOptions[texhigh/prelude]

\THSetCharReplacement{\_}{\textvisiblespace}
% \THSetCharReplacement*{\_}{\ifincname\space\else\textvisiblespace\fi}
\THSetCharReplacement{\_}{\mbox{\THcolor{gray8}{\rightarrow}}}}

\texhighsetclassfallback{ch}{group.0}{group}
\texhighsetclassfallback{ch}{group.1}{group}
\texhighsetclassfallback{ch}{group.2}{group}
\texhighsetclassfallback{ch}{group.3}{group}
\texhighsetclassfallback{ch}{group.4}{group.1, group}
\texhighsetclassfallback{ch}{group.5}{group.2, group}
\texhighsetclassfallback{ch}{group.6}{group.3, group}
\texhighsetclassfallback{ch}{group.7}{group.1, group}
\texhighsetclassfallback{ch}{group.8}{group.2, group}
\texhighsetclassfallback{ch}{group.9}{group.3, group}
\texhighsetclassfallback{ch}{group.-1}{group.miss}
\texhighsetclassfallback{ch}{group.-2}{group.miss}
\texhighsetclassfallback{ch}{group.-3}{group.miss}
\texhighsetclassfallback{ch}{group.-4}{group.miss}
```

```

\texhighsetclassfallback{ch}{group.-5}{group.miss}
\texhighsetclassfallback{cs}{primitive.luatex}{primitive, tex, latex}
\texhighsetclassfallback{cs}{primitive.xetex}{primitive, tex, latex}
\texhighsetclassfallback{cs}{primitive.uptex}{primitive, tex, latex}
\texhighsetclassfallback{cs}{primitive.pdfTeX}{primitive, tex, latex}
\texhighsetclassfallback{cs}{primitive.etex}{primitive, tex, latex}
\texhighsetclassfallback{cs}{primitive.knuthTeX}{primitive, tex, latex}
\texhighsetclassfallback{cs}{primitive.widely}{primitive, tex, latex}
\texhighsetclassfallback{cs}{primitive.sometex}{primitive, tex, latex}
\texhighsetclassfallback{cs}{primitive.luametateX}{context}
\texhighsetclassfallback{cs}{plaintex}{tex}
\texhighsetclassfallback{cs}{latex.document}{latex}
\texhighsetclassfallback{cs}{latex.programming}{latex}
\texhighsetclassfallback{cs}{latex.internal}{internal, latex}
\texhighsetclassfallback{cs}{latex.public}{latex}
\texhighsetclassfallback{cs}{latex3.primitive}{latex3.function.kernel, ←
↪ latex3.function, latex3, primitive}
\texhighsetclassfallback{cs}{latex3.variable}{latex3, latex}
\texhighsetclassfallback{cs}{latex3.function}{latex3, latex}
\texhighsetclassfallback{cs}{latex3.variable.kernel}{latex3.kernel, latex3.variable, ←
↪ latex3}
\texhighsetclassfallback{cs}{latex3.function.kernel}{latex3.kernel, latex3.function, ←
↪ latex3}
\texhighsetclassfallback{cs}{latex3.variable.internal}{internal, latex3.variable, latex3}
\texhighsetclassfallback{cs}{latex3.variable.public}{latex3.variable, latex3}
\texhighsetclassfallback{cs}{latex3.function.internal}{internal, latex3.function, latex3}
\texhighsetclassfallback{cs}{latex3.function.public}{latex3.function, latex3}
\texhighsetclassfallback{cs}{math}{tex, latex}
\texhighsetclassfallback{cs}{symbol}{tex, latex}
\texhighsetclassfallback{rs}{math.inline}{math}
\texhighsetclassfallback{re}{math.inline}{math}
\texhighsetclassfallback{rs}{argument.m}{argument}
\texhighsetclassfallback{re}{argument.m}{argument}
\texhighsetclassfallback{rs}{argument.o}{argument.d, argument}
\texhighsetclassfallback{re}{argument.o}{argument.d, argument}
\texhighsetclassfallback{rs}{argument.O}{argument.d, argument}
\texhighsetclassfallback{re}{argument.O}{argument.d, argument}
\texhighsetclassfallback{rs}{argument.d}{argument}
\texhighsetclassfallback{re}{argument.d}{argument}
\texhighsetclassfallback{rs}{argument.D}{argument.d, argument}
\texhighsetclassfallback{re}{argument.D}{argument.d, argument}
\texhighsetclassfallback{rs}{argument.r}{argument.m, argument}
\texhighsetclassfallback{re}{argument.r}{argument.m, argument}
\texhighsetclassfallback{rs}{argument.R}{argument.m, argument}
\texhighsetclassfallback{re}{argument.R}{argument.m, argument}
\texhighsetclassfallback{rs}{argument.s}{argument.t, argument}

```

```

\texhighsetclassfallback{re}{argument.s}{argument.t, argument}
\texhighsetclassfallback{rs}{argument.t}{argument}
\texhighsetclassfallback{re}{argument.t}{argument}
\texhighsetclassfallback{rs}{argument.v}{argument}
\texhighsetclassfallback{re}{argument.v}{argument}
\texhighsetclassfallback{rs}{argument.l}{argument}
\texhighsetclassfallback{re}{argument.l}{argument}
\texhighsetclassfallback{rs}{argument.g}{argument.d, argument}
\texhighsetclassfallback{re}{argument.g}{argument.d, argument}
\texhighsetclassfallback{rs}{argument.G}{argument.d, argument}
\texhighsetclassfallback{re}{argument.G}{argument.d, argument}
\texhighsetclassfallback{rs}{argument.u}{argument}
\texhighsetclassfallback{re}{argument.u}{argument}
\texhighsetclassfallback{rs}{argument.U}{argument.u, argument}
\texhighsetclassfallback{re}{argument.U}{argument.u, argument}
\texhighsetclassfallback{rs}{argument.^^J}{argument.u, argument}
\texhighsetclassfallback{re}{argument.^^J}{argument.u, argument}

@ifpackageloaded{xcolor}{\RequirePackage{xcolor}}
\RequirePackage{ninecolors}

\THSaveStyle{plain}{\THSetPlainStyle{*}}
\THSaveStyle{plain0}{\THSetPlainStyle{bp,cs,ch,st,es,ee,pn,color}%
  \THSetClassRS{comment}{\begingroup\THColorStatus{1}\THcolor[gray]{0.5}}%
  \THSetClassRE{comment}{\endgroup}}

\THSaveStyle{basic}{%
  \THSetClassBP{?}
  {\ifhmode\discretionary
    {\hbox{\THcolor{gray8}\_ $\hookleftarrow$}}
    {\hbox{\THcolor{gray8}$\hookrightarrow\_}}
    {} \fi}%
  \THSetClassCS{texhigh}{\fboxsep\z@% for texhigh package itself
    \fcolorbox{yellow}{yellow}{\linespread{1}\bfseries\strut\THcolor{black}{#1#2}}}%
  ↪ #1=escape char, #2=cs name
  \THSetClassCS{latex3.primitive}{\mbox{\THcolor{red4}\bfseries#1#2}}%
  \THSetClassCS{latex3.kernel}{\mbox{\THcolor{red4}\bfseries#1#2}}%
  \THSetClassCS{latex3.variable}{\mbox{\THcolor{azure6}{#1#2}}}%
  \THSetClassCS{latex3.function}{\mbox{\THcolor{green5}{#1#2}}}%
  \THSetClassCS{internal}{\mbox{\THcolor{brown3}{#1#2}}}%
  \THSetClassCS{latex.document}{\mbox{\THcolor{magenta4}\bfseries#1#2}}%
  \THSetClassCS{latex.programming}{\mbox{\THcolor{yellow7}{#1#2}}}%
  \THSetClassCS{latex}{\mbox{\THcolor{yellow8}{#1#2}}}%
  \THSetClassCS{primitive}{\texhigh@underline{\THcolor{purple5}\bfseries#1#2}}%
  \THSetClassCS{?}{\mbox{\THcolor{magenta5}{#1#2}}}%
  \THSetClassCH{?}{#1}% char
  \THSetClassPN{?}{#1}% punctuation

```

```

\THSetClassCH{group}{\mbox{\THcolor{purple3}{#1}}}%
\THSetClassCH{group.1}{\mbox{\THcolor[HTML]{179FFF}{#1}}}%
\THSetClassCH{group.2}{\mbox{\THcolor[HTML]{DA6ED6}{#1}}}%
\THSetClassCH{group.3}{\mbox{\THcolor[HTML]{F8BA16}{#1}}}%
\THSetClassCH{group.miss}{\mbox{\THcolor{red}{#1}}}%
\THSetClassCH{digit}{\mbox{\THcolor{azure8}{#1}}}%
\THSetClassRS{comment}{\begingroup\THcolor[gray]{0.5}\THSetPlainStyle{cs,color}}%
\THSetClassRE{comment}{\endgroup}%
\THSetClassRS{parameter}{\begingroup\THcolor{magenta2}}%
\THSetClassRE{parameter}{\endgroup}%
\THSetClassRS{math}{\begingroup\THcolor{cyan7}}%
\THSetClassRE{math}{\endgroup}%
\THSetClassES{[]texcl}{\begingroup\@texhigh@reset@ctab\@texhigh@reset@font ←
→ \ignorespaces}%
\THSetClassEE{[]texcl}{\endgroup}%
}
\THUseSavedStyle{basic}

\long\def\texhigh@underline#1{\leavevmode\setbox\z@\hbox{#1}}%
\hbxt@wd\z@{\kern.05em
\vrule height-.25ex depth.4ex width\dimexpr\wd\z@-.1em\relax \kern.05em
\llap{\unhbox\z@}}

\if@texhighload@color
\relax
\fi
\ifpackageloaded{tikz}{\@texhighload@tikztrue}{}
\newbox\texhigh@picturebox
\if@texhighload@tikz
\RequirePackage{tikz}
\usetikzlibrary{shadings}
\usetikzlibrary{fill.image}
\protected\def\texhigh@shadetext#1#2{%
\setbox\texhigh@picturebox=\hbox{\texhigh@pdfliteral{7 Tr }#2}}%
\tikz[baseline=0,line width=0pt]\path\pgfextra{\rlap{\copy\texhigh@picturebox}}
[#1] (0,-\dp\texhigh@picturebox) rectangle (\wd\texhigh@picturebox,\ht ←
→ \texhigh@picturebox);}

\tikzset{texhigh/.is family,
texhigh/gradient primitive/.style={left color=blue,right color=cyan},
texhigh/gradient ?/.style={left color=red,right color=blue},
texhigh/gradient-style/.style={texhigh/gradient #1}}
\THSaveStyle{tikz.gradient}{%
\THSetClassCS{latex}{\texhigh@underline{\THcolor{purple}\bfseries#1#2}}
\THSetClassCS{primitive}
{\texhigh@shadetext{texhigh/gradient-style=primitive}{\bfseries #1#2}}%
\THSetClassCS{?}{\texhigh@shadetext{texhigh/gradient-style=?}{#1#2}}%
}

```

```

\fi

\@ifpackageloaded{tcolorbox}{
  \tcbuselibrary{listings@core}
  \def\tcb@texhigh@file#1#2{%
    {\edef\tcb@temp{\texhighfile[\unexpanded\expandafter{#1}]]}\tcb@temp{#2}}
  \def\tcb@texhigh@uselistlisting{\tcb@texhigh@file\kvtcb@texhighoptions ←
  ↪ \kvtcb@listingfile}
  \def\tcb@texhigh@usetemplisting{\tcb@texhigh@file\kvtcb@texhighoptions\kvtcb@tempfile}
  \def\tcb@texhigh@doc@usetemplisting{\tcb@texhigh@file\kvtcb@doctexhighoptions ←
  ↪ \kvtcb@tempfile}
  %
  \tcbset{
    texhigh options/.code=\edef\kvtcb@texhighoptions{\unexpanded{#1}},
    texhigh options=,
    texhigh options pre/.code=\edef\kvtcb@texhighoptions{\unexpanded{#1},\unexpanded ←
  ↪ \expandafter{\kvtcb@texhighoptions}},
    texhigh options app/.code=\edef\kvtcb@texhighoptions{\unexpanded\expandafter{ ←
  ↪ \kvtcb@texhighoptions,#1}},
    texhigh gobble/.style={texhigh options app={gobble=#1}},
    texhigh gobble/.default=auto,
    texhigh config file/.style={texhigh options app={config-file={#1}}},
    texhigh ctab file/.style={texhigh options app={ctab-file={#1}}},
    texhigh use ctab/.style={texhigh options app={use-ctab={#1}}},
    texhigh style/.style={texhigh options app={style={#1}}},
    texhigh detect catcodes/.style={texhigh options app={lexer-catcode={#1}}},
    listing engine/texhigh/.code={\let\tcbuselistlisting\tcb@texhigh@uselistlistinglisting
      \let\tcbusetemplisting\tcb@texhigh@usetemplisting
      \let\tcb@doc@usetemplisting\tcb@texhigh@doc@usetemplisting},
  }
}{}

```