

# *Efficient Vocabulary Discovery in Late Antique Texts– Code Examples*

Andrew J. Hayes

May 22, 2025

The following are the code examples interactively demonstrated in the presentation. Each example follows the number that designates it.

(1)

```
lualatex -output-dir=build repaginate.tex
```

(2)

```
pdfgrep -e '[Ss]ch[ä]tz' HdE\ German.pdf -H --page-number=label
```

(3)

```
pdfgrep -e '[Ss]ch[ä]tz' HdE\ German.pdf -H -c
```

(4)

```
pdfgrep -e '[Ss]ch[ä]tz' HdE\ German.pdf HdV\ German.pdf -c
```

(5)

```
pdfgrep *.pdf -e 'G[aei]hen(n)?a' --page-number=label -H
```

(6)

```
pdfgrep -i -e '([Kk]ingdom)|(Königtum)|(\(Himmel\)[Rr]eich)' --page-number=label -H *.pdf
```

(7)

```
pdfgrep -P '(?![Zz]u )Ende(?! des [a-z])'  
--page-number=label -H -C 3  
--color=always *.pdf | grep -v -e '[Zz]u'
```

(8)

```
pdfgrep -P '(?![Zz]u )Ende(?! des [a-z])'  
--page-number=label -H  
--color=always *.pdf | grep -v -e '[Zz]u'
```

(9)

```
pdfgrep -e '[Ss]ch[ä]tz' HdE\ German.pdf  
-H --page-number=label > results.txt
```

(10)

```
pdfgrep *.pdf -e 'G[aei]hen(n)?a'
--page-number=label | awk -f grep_to_csv.awk > gehenna.csv
```

(11)

```
BEGIN { FS = ":"; print "text", ",", "citation" }
{
    cutoff = index($1, " ");
    # print "cutoff index is " cutoff # for debugging purposes
    print substr($1, 1, cutoff) ",", $2
}
```