A Tool for Analysing Higher-Order Feature Interactions in Preprocessor Annotations in C and C++ Projects

David Korsman
Radboud University Nijmegen
Nijmegen, The Netherlands
david.korsman@ru.nl

CDN (**Diego**) Damasceno
Radboud University Nijmegen
Nijmegen, The Netherlands
https://damascenodiego.github.io/
d.damasceno@cs.ru.nl

Daniel Strüber
Chalmers | University of Gothenburg (SE)
Radboud University Nijmegen (NL)
https://www.danielstrueber.de/
danstru@chalmers.se

@damascenodiego

26th ACM International Systems and Software Product Line Conference (SPLC'22) – Sept. 12–16, 2022

Radboud University

SPLC'22

Pre-processors in C and C++

```
1 #ifdef FEAT_BIGINT
2   #define SIZE 64
3 #else
4   #define SIZE 32
5 #endif
6
7 ... allocate(SIZE) ...;
```

Alternative macro definitions

```
1 #ifdef FEAT_WINDOWS
2   #include <windows.h>
3 #else
4   #include <unix.h>
5 #endif
6
7 ... fopen(...) ...;
```

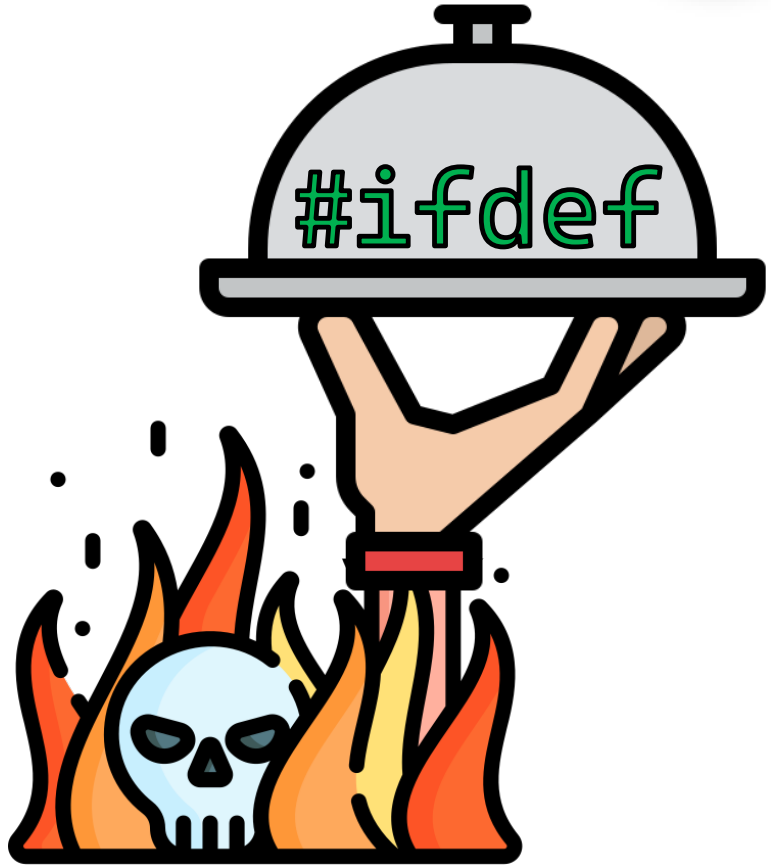
Alternative Includes

```
1 #ifdef FEAT_SELINUX
2   #define FEAT_LINUX 1
3   #undef  FEAT_WINDOWS
4 #endif
5
6 #ifdef FEAT_WINDOWS
7 ...
```

Conditional feature definitions

```
1 #ifdef FEAT_RAND
2 int rand() { ... }
3 #else
4 #define rand(...) 0
5 #endif
6
7 int i = 3 + rand();
```

Alternative function definitions



The #ifdef Hell

Our Contribution: `pdparser`

http://github.com/dkorsman/pdparser

A Python-based tool for automated reasoning of structural feature interactions in preprocessor directives of programs written in C and C++

Functionalities

1. Fully relies on standard python libraries (i.e., RegEx , argparse)
2. Provides a CLI to analyze source code files in a C/C++ project (incl. batch mode)
3. Exports results in JSON format (incl. variability information, features in multiple projects)
4. Project showcase and help menu for getting started

Showcase experiment: Vim and Emacs

Showcase experiment (Number of features)

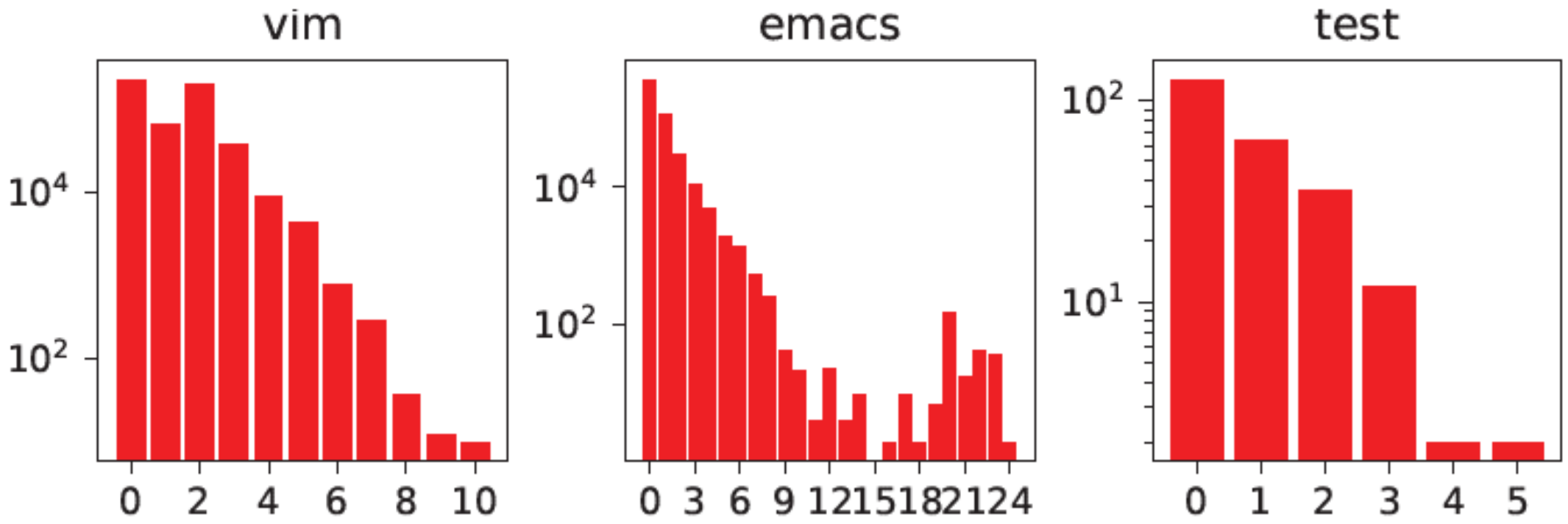


Figure 1: Number of features involved

Showcase experiment (LOC at each nesting level)

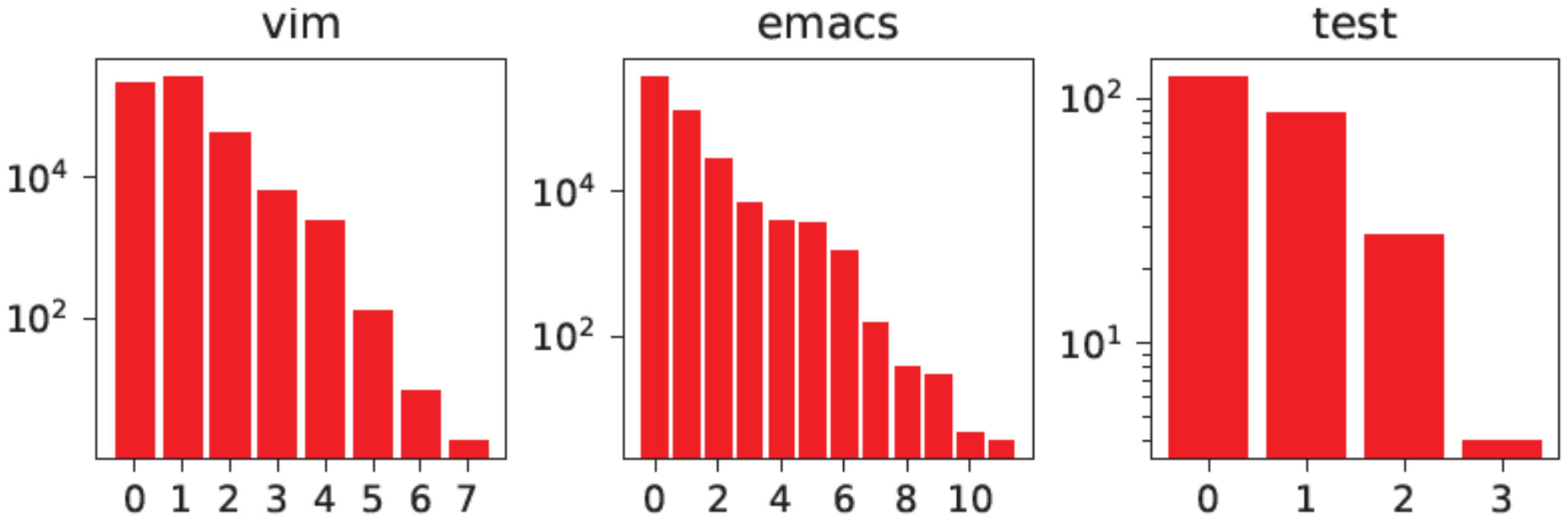


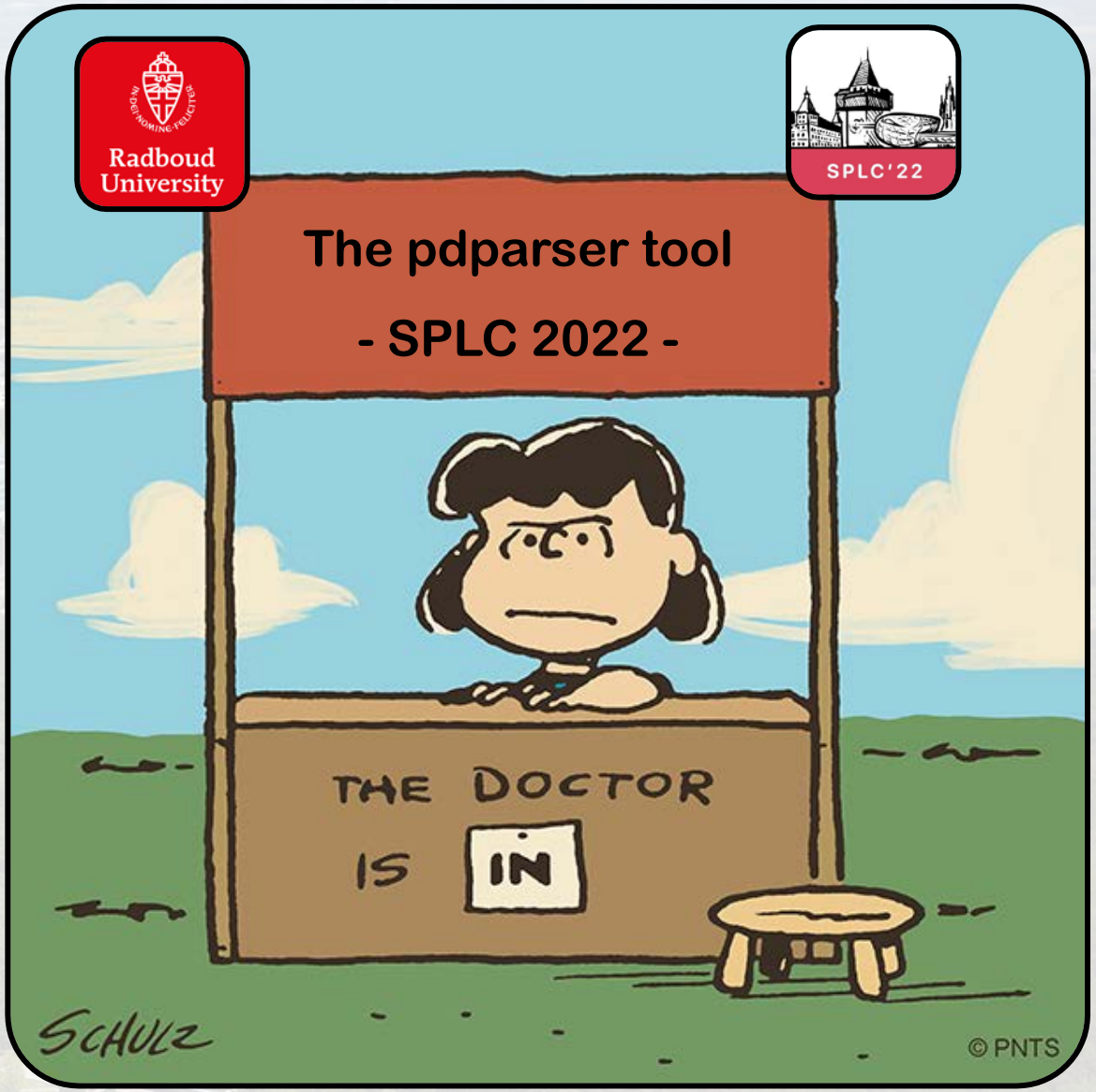
Figure 2: Number of lines at each nesting level

Come and visit our stand!

David Korsman
david.korsman@ru.nl

CDN (**Diego**) Damasceno
d.damasceno@cs.ru.nl

Daniel Strüber
danstru@chalmers.se



pdparser vs. KernelHeaven

Code extractor name	Input files	Implem. Lang.	Preproc. Block	Annot. AST	Quant. Report
Undertaker [20]	*.c, *.h, *.s	C++	■		
Code Block [13]	*.c, *.h, *.s	Java	■		
srcML [7]	*.c, *.h	Java		■	
TypeChef [12]	*.c	Java		■	
pdparser [5]	*.c, *.h, *.cpp, *.hpp, *.cxx, *.m, *.cc	Python	■		■

Table 1: Code Extractors in KernelHeaven [13] vs. pdparser

pdparser @ GitHub

http://github.com/dkorsman/pdparser

dkorsman / pdparser Public

Code Issues Pull requests Actions Projects Wiki Security Insights

master 1 branch 0 tags Go to file Add file Code

damascenodiego Update README.md d92e892 on 18 Jun 18 commits

test	Add test project	6 months ago
README.md	Update README.md	2 months ago
count-project-features.py	Add count_features to project feature counting	8 months ago
download-all-experiments.sh	Add script to download all experiments	2 months ago
pdparser.py	Add source code line counters and unique feature filter	6 months ago
repair-all-projects.sh	Fix some parser errors in projects	10 months ago
run-all-projects.sh	Add some line-based statistics in addition to block-based statistics	9 months ago
source_parser.py	Add source code line counters and unique feature filter	6 months ago
util.py	Add file/folder blacklist feature, unify print system	9 months ago

pdparser @ CLI

http://github.com/dkorsman/pdparser

```
usage: pdparser.py [-h] [-o OUTPUT] [-q [CAT [CAT ...]]] [-d] [-f] [-B] [-jr]
                   [-jp] [-u UNIQUE_FEATURE_FILTER]
                   source
  -o, --output          the source folder to operate on
  -q, --quiet           silence a category of messages
  -d, --hide-code       hide code from output files except
                        preprocessor directives and annotations
  -f, --features        try to get all involved features in defined(X)
  -B, --no-blacklist    ignore pdparser-blacklist.lst file
                        (to get all files again)
  -jr, --json-result    store a json file with general results in
                        the output folder
  -jp, --json-pinpoint  store json files with all locations of nesting levels
                        and feature interactions
  -u UNIQUE_FEATURE_FILTER, --unique-feature-filter UNIQUE_FEATURE_FILTER
                        the count-project-features.json to use for filtering
                        UNIQUE features - pass only if you want to filter
```

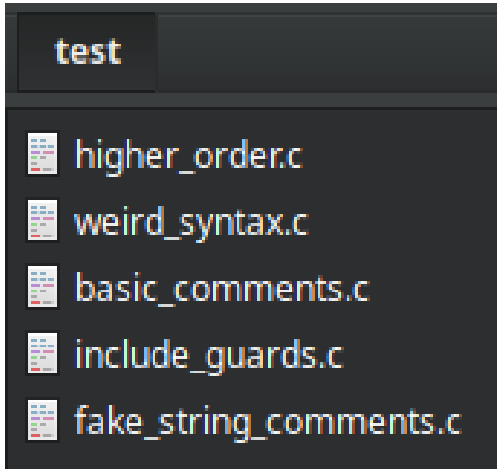
pdparser @ Youtube (Demo video)

A Tool for Analysing Higher-Order Feature Interactions in Preprocessor Annotations in C and C++ Projects

Showcase experiment

Vim and Emacs

Showcase experiment (Test project)




```
const char *s = "/*";

#ifdef GOOD_8
The comment start is part of a string
#endif

const char *s2 = "*/";

fn("asdf", /* "asdf" ...

#ifdef BAD_2

The comment start is not actually part of a string

*/ "asdf");
```

```
fn("This string does\" NOT end here! /* NOT a \"comment");

#ifdef GOOD_11
An escaped quote does not end the string
#endif
```

```
/* We start a comment here
"and we suddenly end it here */ "this is a string" // "

#ifdef GOOD_12
That string wasn't real was it?
#endif
```

Showcase experiment (Number of features)

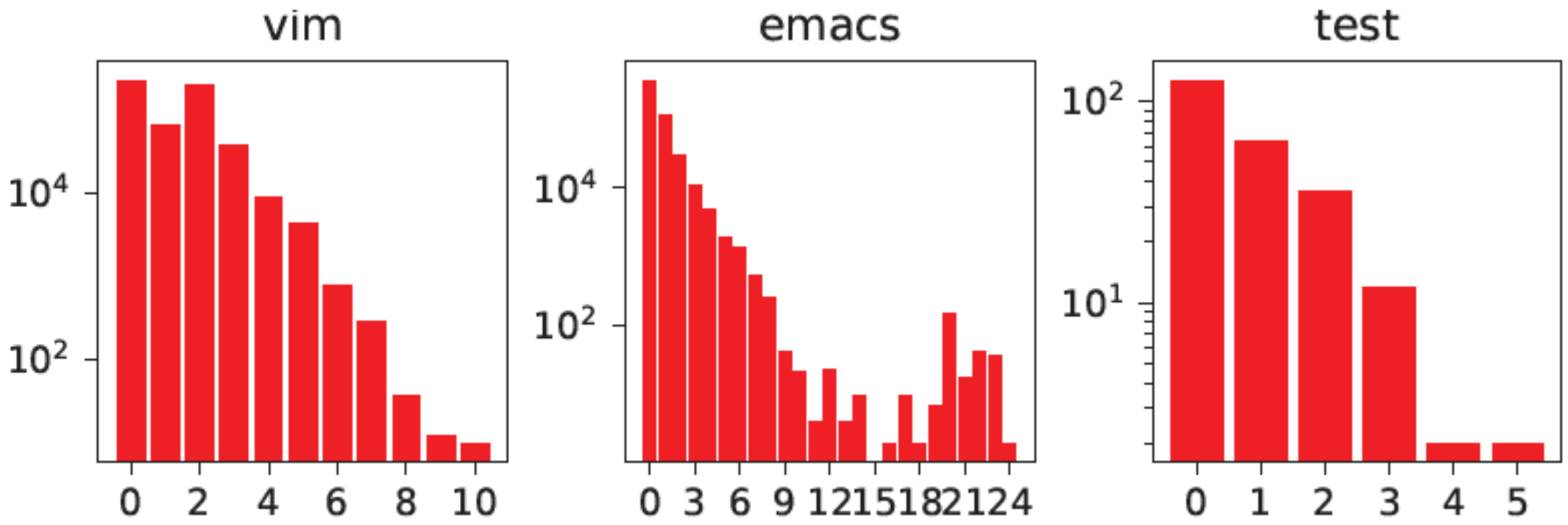


Figure 1: Number of features involved

Showcase experiment (LOC at each nesting level)

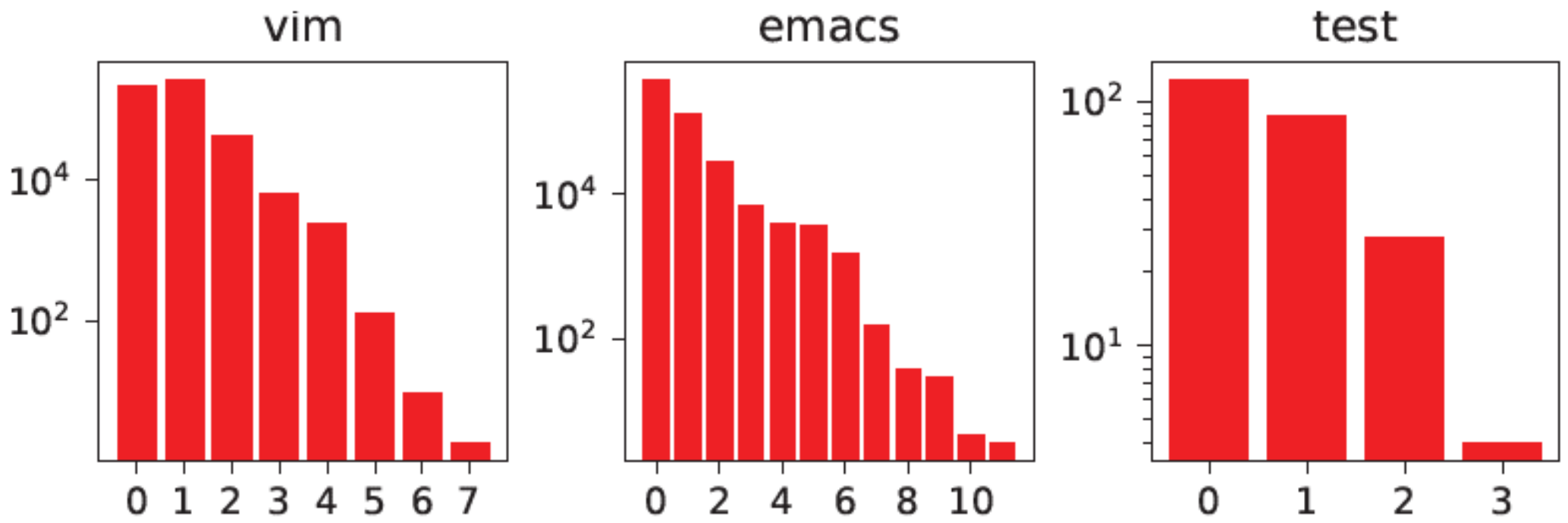


Figure 2: Number of lines at each nesting level

Showcase experiment (Results)

- Annotated source files

```
#ifdef HAVE_SCHED_GETSCHEDULER
// [pdparser] L1, only if (defined(HAVE_SCHED_GETSCHEDULER))
// [pdparser] Features involved: {'HAVE_SCHED_GETSCHEDULER'}
        #ifdef HAVE_GETPRIORITY
// [pdparser] L2, only if (defined(HAVE_SCHED_GETSCHEDULER)) && (defined(HAVE_GETPRIORITY))
// [pdparser] Features involved: {'HAVE_SCHED_GETSCHEDULER', 'HAVE_GETPRIORITY'}
        #else
// [pdparser] L2, only if (defined(HAVE_SCHED_GETSCHEDULER)) && !(defined(HAVE_GETPRIORITY))
// [pdparser] Features involved: {'HAVE_SCHED_GETSCHEDULER', 'HAVE_GETPRIORITY'}
        #endif
// [pdparser] L1, only if (defined(HAVE_SCHED_GETSCHEDULER))
// [pdparser] Features involved: {'HAVE_SCHED_GETSCHEDULER'}
#else
// [pdparser] L1, only if !(defined(HAVE_SCHED_GETSCHEDULER))
// [pdparser] Features involved: {'HAVE_SCHED_GETSCHEDULER'}
#ifdef HAVE_GETPRIORITY
// [pdparser] L2, only if !(defined(HAVE_SCHED_GETSCHEDULER)) && (defined(HAVE_GETPRIORITY))
// [pdparser] Features involved: {'HAVE_SCHED_GETSCHEDULER', 'HAVE_GETPRIORITY'}
#endif
// [pdparser] L1, only if !(defined(HAVE_SCHED_GETSCHEDULER))
// [pdparser] Features involved: {'HAVE_SCHED_GETSCHEDULER'}
#endif
// [pdparser] L0, always
// [pdparser] Features involved: set()
```

- Results in JSON format


```
{
 "n_source_files": 469,
 "n_code_lines": 270642,
 "n_total_lines": 412288,
 "n_ignored_files": 1483,
 "n_parser_errors": 0,
 "n_possible_guards": 134,
 "n_features": 777,
 "features": [
  "SQLITE_DISABLE_PAGECACHE_OVERFLOW_STATS",
  "HEALTH_LISTEN_BACKLOG",
  "HAVE_MONGOC",
  ...
 ],
 "time_taken": 3.8627485579345375
 ...
}
```

Showcase experiment (#else without #if at Emacs)

```
/* comment */ #define ANSIC
 #define ANSIC
 #else
typedef void (proc) ();
```

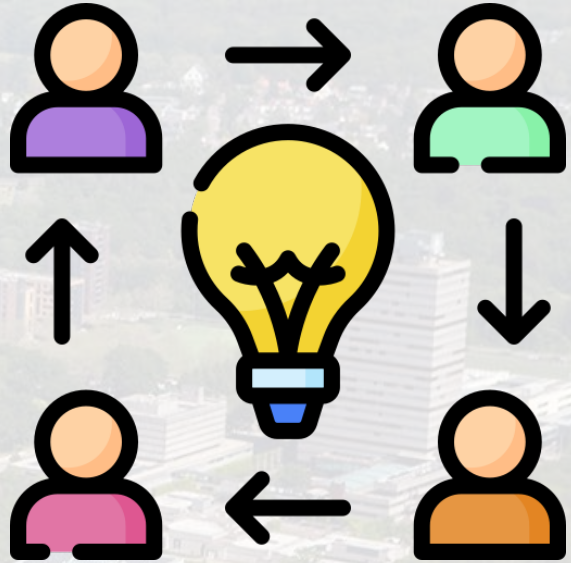
https://github.com/emacs-mirror/emacs/blob/98365c7b1e1e1d3d5f7185f2d4a2baa1c65b4540/test/manual/etags/c-src/h.h#L86

Thank you!

David Korsman
david.korsman@ru.nl

CDN (**Diego**) Damasceno
d.damasceno@cs.ru.nl

Daniel Strüber
danstru@chalmers.se



Questions?