Interesting Patterns in R Code

CodeInspector — Monthly-Meeting

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- Assignments
- Loops
- Controlflow
- Data-Access
- Meta

Top imported packages

Package Code		*User Code*	
4.69 % 1 112	`stats`	5.81 % 1 212	`ggplot2`
3.94 % 934	`utils`	5.65 % 1 179	`dplyr`
3.58 % 848	`dplyr`	3.70 % 772	`tidyverse`
3.43 % 812	`rlang`	2.99 % 623	`lme4`
3.38 % 802	`testthat`	2.18 % 455	`car`
2.84 % 674	`cli`	1.96 % 408	`plyr`

Ways of import

	::	:::	lib.	req.	req.Ns.	var.	load.	*apply	attach.
Package	66.61 % 15 788	2.69 % 637	22.04 % 5 224	3.84 % 911	3.74 % 887	0.70 % 165	0.28 % 66	0.09 % 21	0.01 % 2
User	9.00 % 1 876	0.20 % 42	83.20 % 17 349	7.26 % 1 514	0.12 % 24	0.03 % 6	0.00 % 0	0.20 % 42	0.00 % 0

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- Assignments
- Loops
- Controlflow
- Data-Access
- Meta

	Comments	
	Package Code	*User Code*
Roxygen/Total	492 510/884 950	1 486/311 344
	Used Imports	
`@import`	305	8
`@importFrom`	5 587	6
`@importClassesFrom`	6	0
`@importMethodsFrom`	8	0
`@useDynLib`	605	0
	Used Exports	
`@export`	33 186	13
`@exportClass`	27	0
`@exportMethod`	81	0
`@exportS3Method`	14	0
`@exportPattern`	1	0

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- Assignments
- Loops
- Controlflow
- Data-Access
- Meta

	Package Code	*User Code*
`Total`	131 953	12 122
`Lambdas`	0.02 % 22	0.00 % 0
`Assigned`	72.88 % 96 161	56.27 % 6 821
`Direct-Call`	0.07 % 91	0.00 % 0
`Nested`	19.34 % 25 521	18.48 % 2 240
`Recursive`	0.80 % 1 058	0.08 % 10

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- **Function Calls**
- Values
- Assignments
- Loops
- Controlflow
- Data-Access
- Meta

	Package Code		*User Code*
1 956 887	*all*	980 964	*all*
6.83 % 133 650	`c`	11.22 % 110 020	`c`
2.93 % 57 398	`list`	2.02 % 19 849	`length`
2.90 % 56 842	`length`	1.93 % 18 975	`library`
2.58 % 50 534	`expect_equal`	1.92 % 18 797	`summary`
2.16 % 42 279	`is.null`	1.84 % 18 009	`aes`
1.90 % 37 238	`test_that`	1.66 % 16 303	`list`
1.72 % 33 688	`return`	1.34 % 13 162	`element_text`

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- Assignments
- Loops
- Controlflow
- Data-Access
- Meta

	Package Code	*User Code*
Numbers	1 094 489	844 344
Imaginary	0.13 % 1 436	0.00 % 37
Integers	9.76 % 106 843	0.04 % 346
FloatHex	0.00 % 12	0.00 % 0
Logical	164 249	64 421
Special	94 590	20 586
Strings	852 103	488 999

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- **Assignments**
- Loops
- Controlflow
- Data-Access
- Meta

	Package Code	*User Code*
Operator	772 263	396 999
←	94.55 % 730 160	81.43 % 323 286
↞	0.55 % 4 211	0.11 % 431
=	4.80 % 37 048	18.08 % 71 769
→	0.02 % 145	0.25 % 1 004
↠	0.00 % 1	0.00 % 0
Others	0.09 % 698	0.13 % 509
Special	82 800	86 171
Nested	6 479	1 680

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- Assignments
- Loops
- Controlflow
- Data-Access
- Meta

	Package Code	*User Code*
`for`	19 298	13 679
`while`	1 357	389
`repeat`	233	50
`break`	1 112	205
`next`	932	282

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- Assignments
- Loops
- Controlflow
- Data-Access
- Meta

	Package Code	*User Code*
if-then-else	216 722	13 773
nested	70.15 % 152 022	70.51 % 9 711
constant	0.35 % 750	0.50 % 69
variable	11.46 % 24 844	5.25 % 723
switch-case	3 421	46

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- Assignments
- Loops
- Controlflow
- **Data-Access**
- Meta

	Package Code	*User Code*
[	186 893	188 888
empty	0.67 % 1 257	0.11 % 204
constant	23.37 % 43 684	20.96 % 39 587
variable	34.14 % 63 805	22.75 % 42 970
[[	62 904	17 248
empty	0.02 % 10	0.00 % 0
constant	48.50 % 30 508	49.87 % 8 601
variable	47.10 % 29 625	46.45 % 8 011
chained	87 134	90 362
by name	271 577	283 081
by slot	19 980	3 198

Feature Extraction

- Packages
- Roxygen2
- Function Definitions
- Function Calls
- Values
- Assignments
- Loops
- Controlflow
- Data-Access
- Meta

	Package Code	*User Code*
Count	25 691	4 230
File Length		
min	1	1
max	29 211	7 100
avg	145	348
median	70	184
Line Length		
min	0	0
max	116 068	33 341
avg	32	44
median	28	36

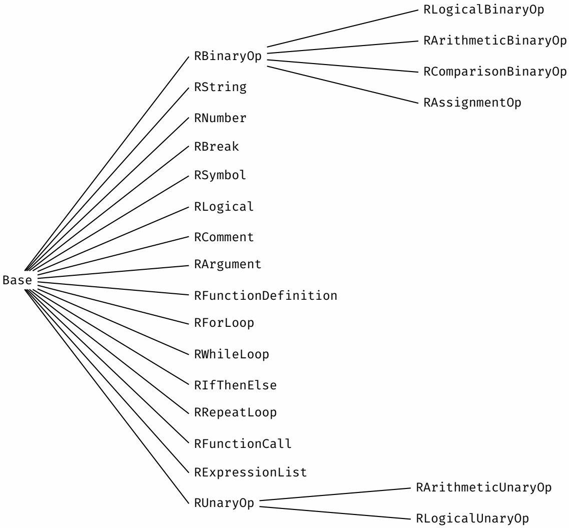
Quad Generation

- We can generate Quads from the normalized AST
- Limited to files that can be parsed

```
<https://uni-ulm.de/r-ast/filename/0> <https://uni-ulm.de/r-ast/type> "exprlist" <filename> .
<https://uni-ulm.de/r-ast/filename/0> <https://uni-ulm.de/r-ast/children-0> <https://uni-ulm.de/r-ast/filename/1> <filename> .
<https://uni-ulm.de/r-ast/filename/1> <https://uni-ulm.de/r-ast/location> <https://uni-ulm.de/r-ast/filename/2> <filename> .
<https://uni-ulm.de/r-ast/filename/2> <https://uni-ulm.de/r-ast/start> <https://uni-ulm.de/r-ast/filename/3> <filename> .
<https://uni-ulm.de/r-ast/filename/3> <https://uni-ulm.de/r-ast/line> "1"^^<http://www.w3.org/2001/XMLSchema#integer> <filename> .
<https://uni-ulm.de/r-ast/filename/3> <https://uni-ulm.de/r-ast/column> "1"^^<http://www.w3.org/2001/XMLSchema#integer> <filename> .
<https://uni-ulm.de/r-ast/filename/2> <https://uni-ulm.de/r-ast/end> <https://uni-ulm.de/r-ast/filename/4> <filename> .
<https://uni-ulm.de/r-ast/filename/4> <https://uni-ulm.de/r-ast/line> "1"^^<http://www.w3.org/2001/XMLSchema#integer> <filename> .
<https://uni-ulm.de/r-ast/filename/4> <https://uni-ulm.de/r-ast/column> "1"^^<http://www.w3.org/2001/XMLSchema#integer> <filename> .
<https://uni-ulm.de/r-ast/filename/1> <https://uni-ulm.de/r-ast/lexeme> "1" <filename> .
<https://uni-ulm.de/r-ast/filename/1> <https://uni-ulm.de/r-ast/type> "NUM_CONST" <filename> .
```

- Question: Which domain should we use?

Ontology



- Base
  - RBinaryOp
    - RLogicalBinaryOp
    - RArithmeticBinaryOp
    - RComparisonBinaryOp
    - RAssignmentOp
  - RString
  - RNumber
  - RBreak
  - RSymbol
  - RLogical
  - RComment
  - RArgument
  - RFunctionDefinition
  - RForLoop
  - RWhileLoop
  - RIfThenElse
  - RRepeatLoop
  - RFunctionCall
  - RExpressionList
  - RUnaryOp
    - RArithmeticUnaryOp
    - RLogicalUnaryOp

https://github.com/Code-Inspect/ontology/issues/1

Open Questions

- What other features are of interest? (if any)
- Which domain should we use for the quads?
- What parts of the data-flow-information are important?