

cregit: identifying contributors of source code



Daniel M German
University of Victoria
dmg@uvic.ca

Kate Stewart
Linux Foundation

Bram Adams
Polytechnique of Montreal

**Intellectual
property**

**Open
Source**

**Mining
Software
Repositories**



Linux



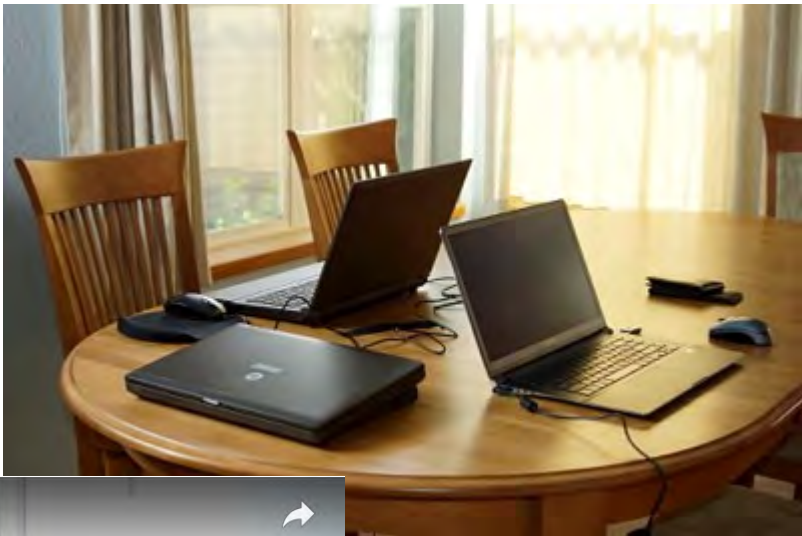




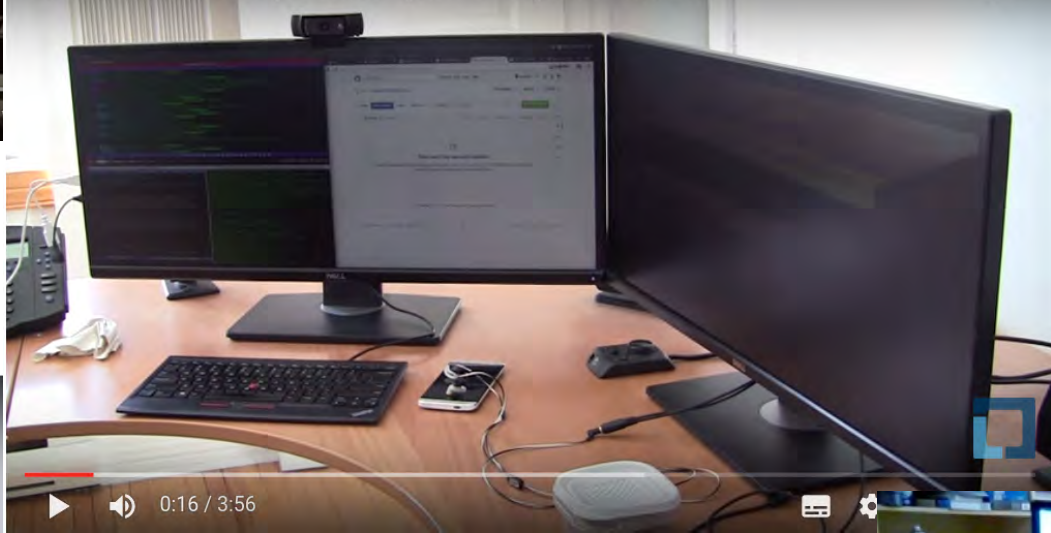




Linux Kernel Developer Workspaces: John Linville, Red Hat



Linux Kernel Developer Work Spaces: Stephane Gruber



Linux Kernel Developer Work Spaces: Tejun Heo, Red Hat



Linux is not only software...
it is also software engineering

Git promotes Linux practices:

Git promotes Linux practices:

- code reviews

Git promotes Linux practices:

- code reviews
- pull requests

Git promotes Linux practices:

- code reviews
- pull requests
- isolation

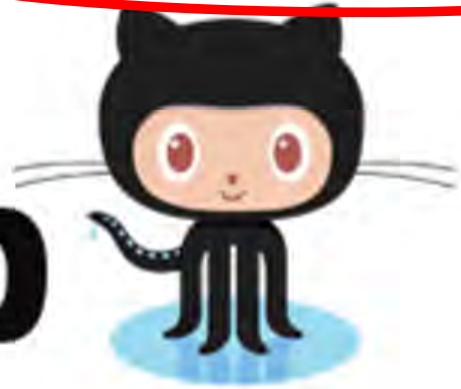
Git promotes Linux practices:

- code reviews
- pull requests
- isolation

and github popularizes them

GitHub for non-technical people

GitHub



contributions welcome

ROBERT MCMILLAN BUSINESS 09.02.12 6:40 AM

FROM
COLLABORATIVE
CODING TO
WEDDING
INVITATIONS:
GITHUB IS GOING
MAINSTREAM

GitHub for the rest of us



Credit: Flickr/Nick Quaranto

Git made it possible for programmers to coordinate distributed work across teams -- now GitHub makes it possible for everyone else

credit: increasing provenance information



Cincinnati Library
image in the public domain

diff-able

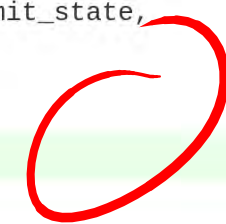
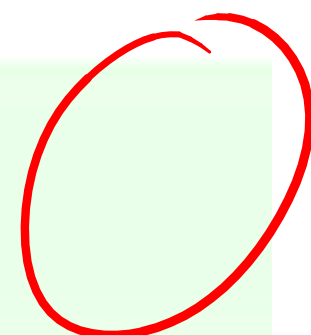
new

old

```
82
83 static struct alloc_state commit_state;
84
85 void *alloc_commit_node(void)
86 {
87 - static int commit_count;
88   struct commit *c = alloc_node(&commit_state,
89     sizeof(struct commit));
90 - c->index = commit_count++;
91   return c;
92 }
93
```

```
82
83 static struct alloc_state commit_state;
84
85 +unsigned int alloc_commit_index(void)
86 +{
87 +   static unsigned int count;
88 +   return count++;
89 +}
90 +
91 void *alloc_commit_node(void)
92 {
93   struct commit *c = alloc_node(&commit_state,
94     sizeof(struct commit));
95 + c->index = alloc_commit_index();
96   return c;
97 }
98
```

diff-able













**IF YOU MUST SQUAWK,
BLAME THEM - AND
WATCH OUT FOR ME!**



REPRODUCTION OF MORRIS PLAN BANKERS ASSOCIATION METER POSTAGE EMBLEM

Image in the public domain

 mmu-notifiers: core Andrea Arcangeli authored on 28 Jul 2008	cddb8a5	29	<code>#include <linux/mmu_notifier.h></code>
 Linux-2.6.12-rc2 Linus Torvalds authored on 16 Apr 2005	1da177e	30	<code>#include <linux/fs.h></code>
 mm: per-thread vma caching Davidlohr Bueso authored on 7 Apr 2014	615d6e8	31	<code>#include <linux/mm.h></code>
 [PATCH] namespaces: add nspr... Serge E. Hallyn authored on 2 Oct 2006	ab51601	32	<code>#include <linux/vmacache.h></code>
 [PATCH] move capable() to capa... Randy.Dunlap authored on 11 Jan 2006	c59ede7	33	<code>#include <linux/nsproxy.h></code>
 Linux-2.6.12-rc2 Linus Torvalds authored on 16 Apr 2005	1da177e	34	<code>#include <linux/capability.h></code>
 Task Control Groups: add fork()/... Paul Menage authored on 19 Oct 2007	b4f48b6	35	<code>#include <linux/cpu.h></code>
 Linux-2.6.12-rc2 Linus Torvalds authored on 16 Apr 2005	1da177e	36	<code>#include <linux/cgroup.h></code>
 hugetlb: reserve huge pages for... gormanm authored on 24 Jul 2008	a1e7877	37	<code>#include <linux/security.h></code>
 seccomp: add system call filteri...	e2cfabd	38	<code>#include <linux/hugetlb.h></code>
		39	<code>#include <linux/seccomp.h></code>

Line

“History is written by the winners”

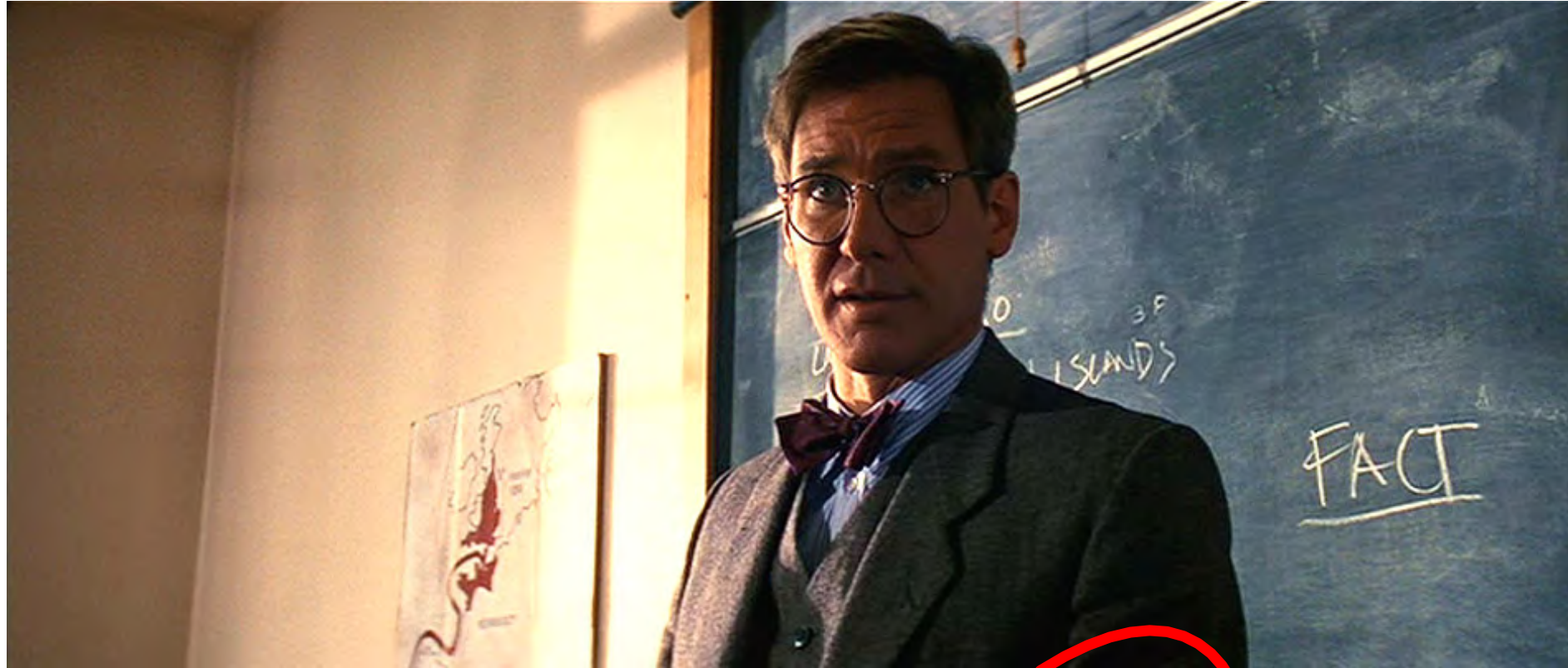


By Y. Karsh. Image in the public domain in Canada. Copyrighted in the US

7.6 Git Tools - Rewriting History

Rewriting History

Many times, when working with Git, you may want to revise your commit history for some reason. One of the great things about Git is that it allows you to make decisions at the last possible moment. You can decide what files go into which commits right before you commit with the staging area, you can decide that you didn't mean to be working on something yet with the stash command, and you can rewrite commits that already happened so they look like they happened in a different way. This can involve changing the order of the commits, changing messages or modifying files in a commit, squashing together or splitting apart commits, or removing commits entirely – all before you share your work with others.



“Archeology is the **search for fact, not truth**. If it's truth you're interested in, Dr. Tyree's Philosophy class is right down the hall.”

-- Indiana Jones

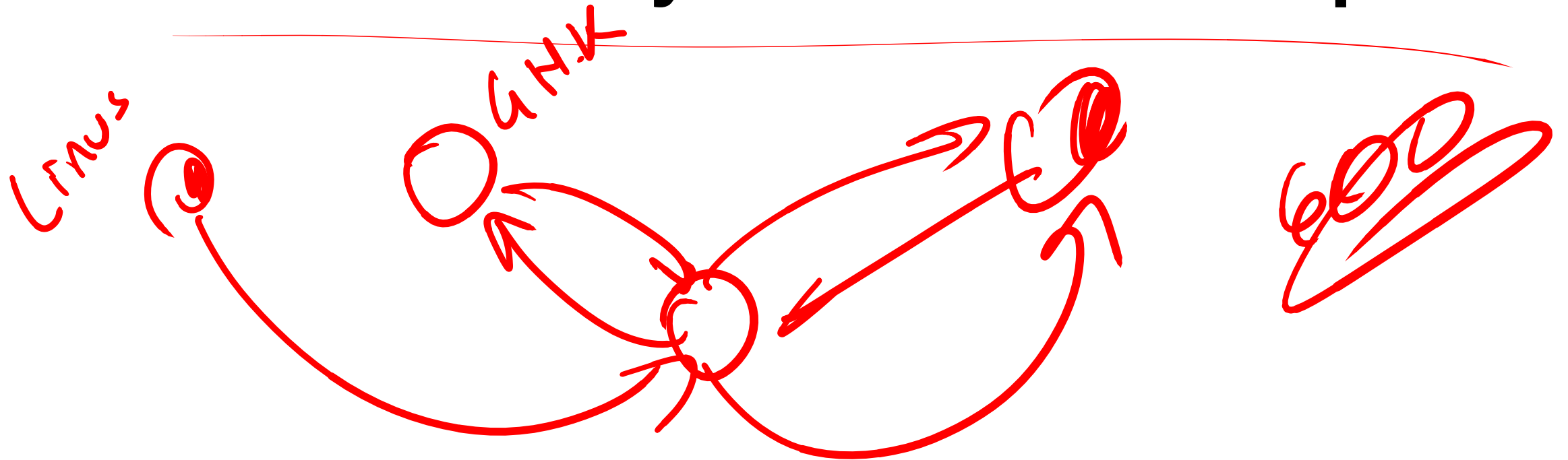
Image Copyright Walt Disney Company

The history in git is likely to be incomplete

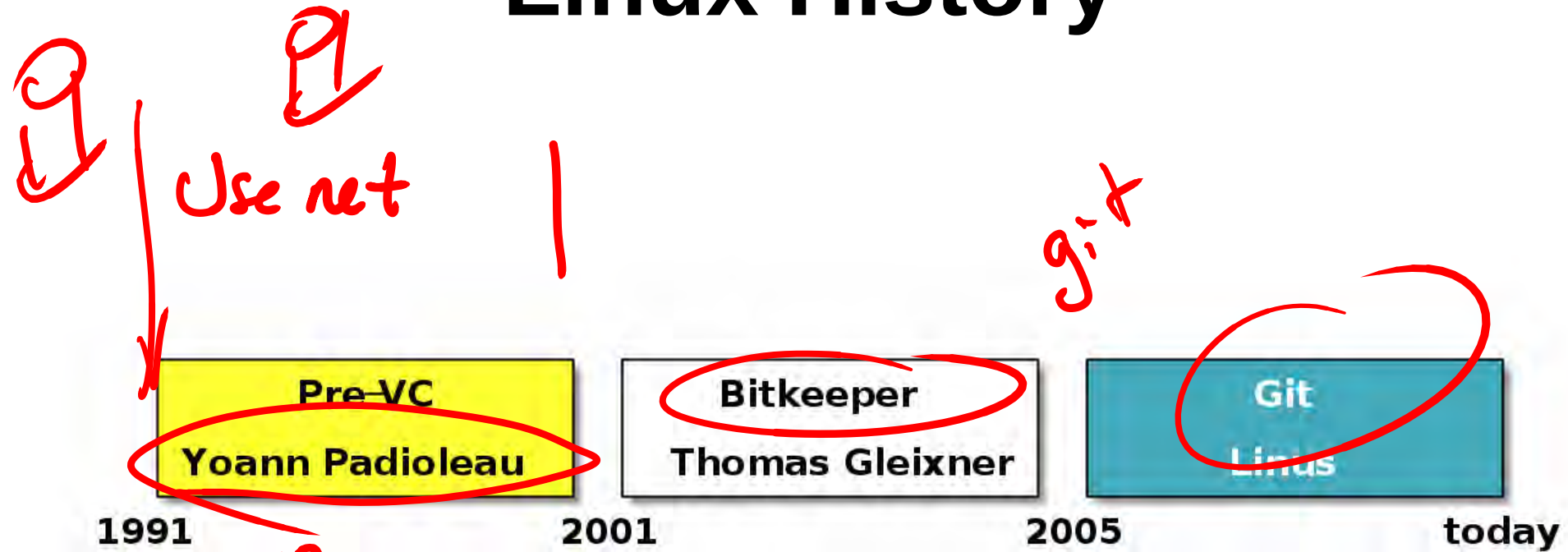


Yet, what can we do with it?

Problem #1: History is broken into repos



Linux History



- Many CVS repositories used before git
- Some code today not under git (eg. quilt)
- Lots of historical information spread in many mailing lists

Problem #2: git tracks changes per line

Warning

- The author of a commit (git parlance) is not necessarily the author of the code
 - Code imported from another source
 - Refactorings
 - Moving code
 - Code derived from other code
- contributors*

```
107
108 void alloc_report(void)
109 {
110     REPORT(blob, struct blob);
111     REPORT(tree, struct tree);
112     REPORT(commit, struct commit);
113     REPORT(tag, struct tag);
114     REPORT(object, union any_object);
115 }
```

blame

when -

Commit

 Add specialized object allocator Linus Torvalds authored on 19 Jun 2006	855419f	107	
 alloc: include any-object allocati... peff authored on 10 Jun 2014	c335d74	110	<code>REPORT(blob, struct blob);</code>
 alloc.c: remove the alloc_raw_c... Ramsay Jones authored on 13 Jul 2014	225ea22	112	<u><code>REPORT(commit, struct commit);</code></u>
 alloc: include any-object allocati... peff authored on 10 Jun 2014	c335d74	113	<code>REPORT(tag, struct tag);</code>
 Add specialized object allocator Linus Torvalds authored on 19 Jun 2006	855419f	115	<code>}</code>

git



Add specialized object allocator

855419f

Linus Torvalds authored on 19 Jun 2006

107

108 void alloc_report(void)

109 {



alloc: include any-object allocati...

c335d74

peff authored on 10 Jun 2014

110

REPORT(blob, struct blob);

111

REPORT(tree, struct tree);



alloc.c: remove the alloc_raw_c...

225ea22

Ramsay Jones authored on 13 Jul 2014

112

REPORT(commit, struct commit);



alloc: include any-object allocati...

c335d74

peff authored on 10 Jun 2014

113

REPORT(tag, struct tag);

114

REPORT(object, union any_object);



Add specialized object allocator

855419f

Linus Torvalds authored on 19 Jun 2006

115

}

Problem: git tracks changes per line

but developers write code (tokens)

```
void alloc_report(void)
{
    REPORT(blob, struct blob);
    REPORT(tree, struct tree);
    REPORT(commit, struct commit);
    REPORT(tag, struct tag);
    REPORT(object, union any_object);
}
```











tokenization →

```
name|void
name|alloc_report
parameter_list|(
name|void
parameter_list|)
block|{
name|REPORT
argument_list|(
name|blob
argument_list|,
expr|struct
name|blob
argument_list|)
expr_stmt|;
name|REPORT
argument_list|(
name|tree
argument_list|,
expr|struct
name|tree
argument_list|)
expr_stmt|;
name|REPORT
argument_list|(
name|commit
argument_list|,
expr|struct
name|commit
argument_list|)
```

version
long dependent.






 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	107	
	108	<code>void alloc_report(void)</code>
	109	<code>{</code>
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	110	<code>REPORT(blob, struct blob);</code>
	111	<code>REPORT(tree, struct tree);</code>
 alloc.c: remove the alloc_raw_c... 225ea22 Ramsay Jones authored on 13 Jul 2014	112	<code>REPORT(commit, struct commit);</code>
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	113	<code>REPORT(tag, struct tag);</code>
	114	<code>REPORT(object, union any_object);</code>
 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	115	<code>}</code>






peff

 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	504	<code>name void</code>	 alloc.c: remove the alloc_raw_c... 297856b Ramsay Jones authored on 13 Jul 2014	528	<code>name commit</code>
	505	<code>name alloc_report</code>			
	506	<code>parameter_list {</code>	 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	529	<code>argument_list ,</code>
	507	<code>name void</code>		530	<code>expr struct</code>
	508	<code>parameter_list }</code>		531	<code>name commit</code>
	509	<code>block {</code>	 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	532	<code>argument_list)</code>
	510	<code>name REPORT</code>		533	<code>expr_stmt ;</code>
	511	<code>argument_list </code>		534	<code>name REPORT</code>
	512	<code>name blob</code>	 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	535	<code>argument_list {</code>
	513	<code>argument_list ,</code>		536	<code>name tag</code>
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	514	<code>expr struct</code>		537	<code>argument_list ,</code>
	515	<code>name blob</code>		538	<code>expr struct</code>
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	516	<code>argument_list)</code>		539	<code>name tag</code>
	517	<code>expr_stmt ;</code>		540	<code>argument_list)</code>
	518	<code>name REPORT</code>		541	<code>expr_stmt ;</code>
	519	<code>argument_list {</code>		542	<code>name REPORT</code>
	520	<code>name tree</code>		543	<code>argument_list {</code>
	521	<code>argument_list ,</code>		544	<code>name object</code>
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	522	<code>expr struct</code>		545	<code>argument_list ,</code>
	523	<code>name tree</code>		546	<code>expr union</code>
	524	<code>argument_list)</code>		547	<code>name any_object</code>
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	525	<code>expr_stmt ;</code>	 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	548	<code>argument_list)</code>
	526	<code>name REPORT</code>		549	<code>expr_stmt ;</code>
	527	<code>argument_list {</code>		550	<code>block }</code>
	528	<code>name commit</code>			

to kern



 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	107	
	108	void alloc_report(void)
	109	{
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	110	REPORT(blob, struct blob);
	111	REPORT(tree, struct tree);
 alloc.c: remove the alloc_raw_c... 225ea22 Ramsay Jones authored on 13 Jul 2014	112	REPORT(commit, struct commit);
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	113	REPORT(tag, struct tag);
	114	REPORT(object, union any_object);
 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	115	}

 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	504	name void
	505	name alloc_report
	506	parameter_list (
	507	name void
	508	parameter_list)
	509	block {
	510	name REPORT
	511	argument_list (
	512	name blob
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	513	argument_list ,
	514	expr struct
	515	name blob
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	516	argument_list)
	517	expr_stmt ;
	518	name REPORT
	519	argument_list (
	520	name tree
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	521	argument_list ,
	522	expr struct
	523	name tree
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	524	argument_list)
	525	expr_stmt ;
	526	name REPORT
	527	argument_list (
	528	name commit

 alloc.c: remove the alloc_raw_c... 297856b Ramsay Jones authored on 13 Jul 2014	528	name commit
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	529	argument_list ,
	530	expr struct
	531	name commit
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	532	argument_list)
	533	expr_stmt ;
	534	name REPORT
	535	argument_list (
	536	name tag
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	537	argument_list ,
	538	expr struct
	539	name tag
	540	argument_list)
	541	expr_stmt ;
	542	name REPORT
	543	argument_list (
	544	name object
	545	argument_list ,
	546	expr union
	547	name any_object
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	548	argument_list)
	549	expr_stmt ;
	550	block }

 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	107 108 void alloc_report(void) 109 {
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	110 REPORT(blob, struct blob); 111 REPORT(tree, struct tree);
 alloc.c: remove the alloc_raw_c... 225ea22 Ramsay Jones authored on 13 Jul 2014	112 REPORT(commit, struct commit);
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	113 REPORT(tag, struct tag); 114 REPORT(object, union any_object);
 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	115 }

to kern






 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	504 name void 505 name alloc_report 506 parameter_list (507 name void 508 parameter_list) 509 block { 510 name REPORT 511 argument_list (512 name blob 513 argument_list , 514 expr struct 515 name blob 516 argument_list) 517 expr_stmt ; 518 name REPORT 519 argument_list (520 name tree 521 argument_list , 522 expr struct 523 name tree 524 argument_list) 525 expr_stmt ;
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	526 name REPORT 527 argument_list (528 name commit

 alloc.c: remove the alloc_raw_c... 297856b Ramsay Jones authored on 13 Jul 2014	528 name commit
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	529 argument_list , 530 expr struct 531 name commit 532 argument_list)
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	533 expr_stmt ; 534 name REPORT 535 argument_list (536 name tag 537 argument_list , 538 expr struct 539 name tag 540 argument_list) 541 expr_stmt ; 542 name REPORT 543 argument_list (544 name object 545 argument_list , 546 expr union 547 name any_object 548 argument_list)
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	549 expr_stmt ; 550 block }
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	

Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	107	void alloc_report(void)
alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	110	REPORT(blob, struct blob);
alloc.c: remove the alloc_raw_c... 225ea22 Ramsay Jones authored on 13 Jul 2014	111	REPORT(tree, struct tree);
alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	112	REPORT(commit, struct commit);
alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	113	REPORT(tag, struct tag);
alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	114	REPORT(object, union any_object);
Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	115	}






Commit





Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	504	name void	alloc.c: remove the alloc_raw_c... 297856b Ramsay Jones authored on 13 Jul 2014	528	name commit
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	505	name alloc_report	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	529	argument_list ,
Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	506	parameter_list (Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	530	expr struct
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	507	name void	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	531	name commit
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	508	parameter_list)	Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	532	argument_list)
Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	509	block {	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	533	expr_stmt ;
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	510	name REPORT	Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	534	name REPORT
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	511	argument_list (alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	535	argument_list (
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	512	name blob	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	536	name tag
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	513	argument_list ,	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	537	argument_list ,
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	514	expr struct	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	538	expr struct
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	515	name blob	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	539	name tag
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	516	argument_list)	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	540	argument_list)
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	517	expr_stmt ;	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	541	expr_stmt ;
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	518	name REPORT	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	542	name REPORT
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	519	argument_list (alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	543	argument_list (
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	520	name tree	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	544	name object
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	521	argument_list ,	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	545	argument_list ,
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	522	expr struct	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	546	expr union
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	523	name tree	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	547	name any_object
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	524	argument_list)	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	548	argument_list)
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	525	expr_stmt ;	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	549	expr_stmt ;
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	526	name REPORT	alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	550	block }
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	527	argument_list (
alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	528	name commit			

 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	107	
	108	void alloc_report(void)
	109	{
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	110	REPORT(blob, struct blob);
	111	REPORT(tree, struct tree);
 alloc.c: remove the alloc_raw_c... 225ea22 Ramsay Jones authored on 13 Jul 2014	112	REPORT(commit, struct commit);
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	113	REPORT(tag, struct tag);
	114	REPORT(object, union any_object);
 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	115	}

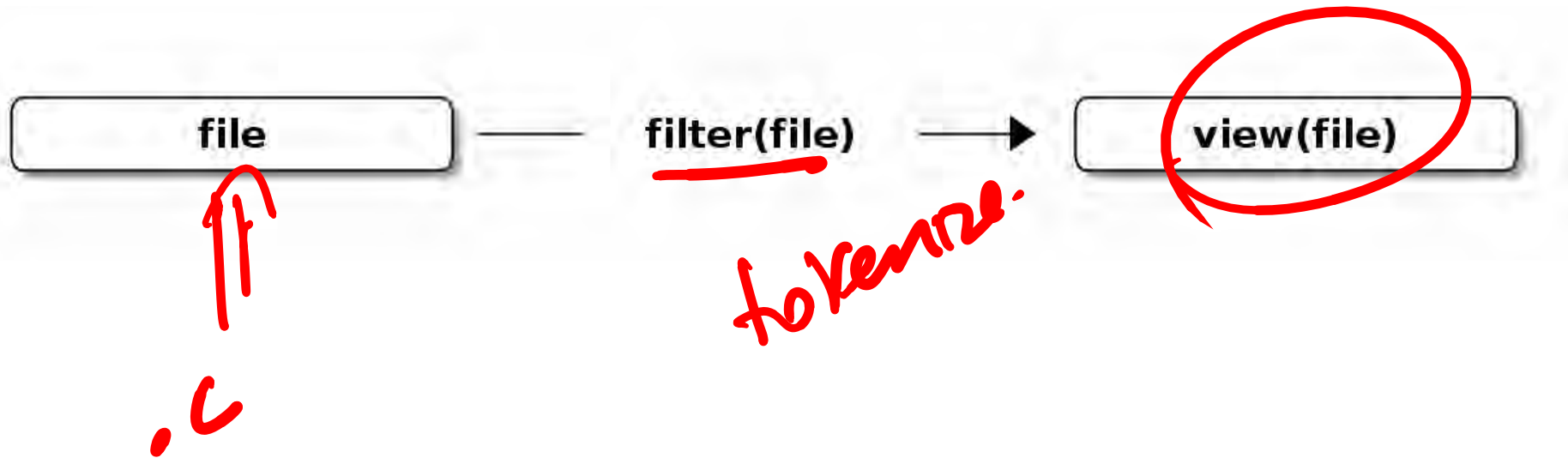
wrong attr per line.

All these sections have wrong attribution

 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	504	name void
	505	name alloc_report
	506	parameter_list (507 name void
	508	parameter_list)
	509	block {
	510	name REPORT
	511	argument_list (512 name blob
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	513	argument_list ,
	514	expr struct
	515	name blob
	516	argument_list)
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	517	expr_stmt ;
	518	name REPORT
	519	argument_list (520 name tree
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	521	argument_list ,
	522	expr struct
	523	name tree
	524	argument_list)
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	525	expr_stmt ;
	526	name REPORT
	527	argument_list (528

 alloc.c: remove the alloc_raw_c... 297856b Ramsay Jones authored on 13 Jul 2014	528	name commit
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	529	argument_list ,
	530	expr struct
	531	name commit
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	532	argument_list)
	533	expr_stmt ;
	534	name REPORT
	535	argument_list (536 name tag
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	537	argument_list ,
	538	expr struct
	539	name tag
	540	argument_list)
	541	expr_stmt ;
	542	name REPORT
	543	argument_list (544 name object
	545	argument_list ,
	546	expr union
	547	name any_object
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	548	argument_list)
	549	expr_stmt ;
	550	block }

Evolutionary Views of VC Repos



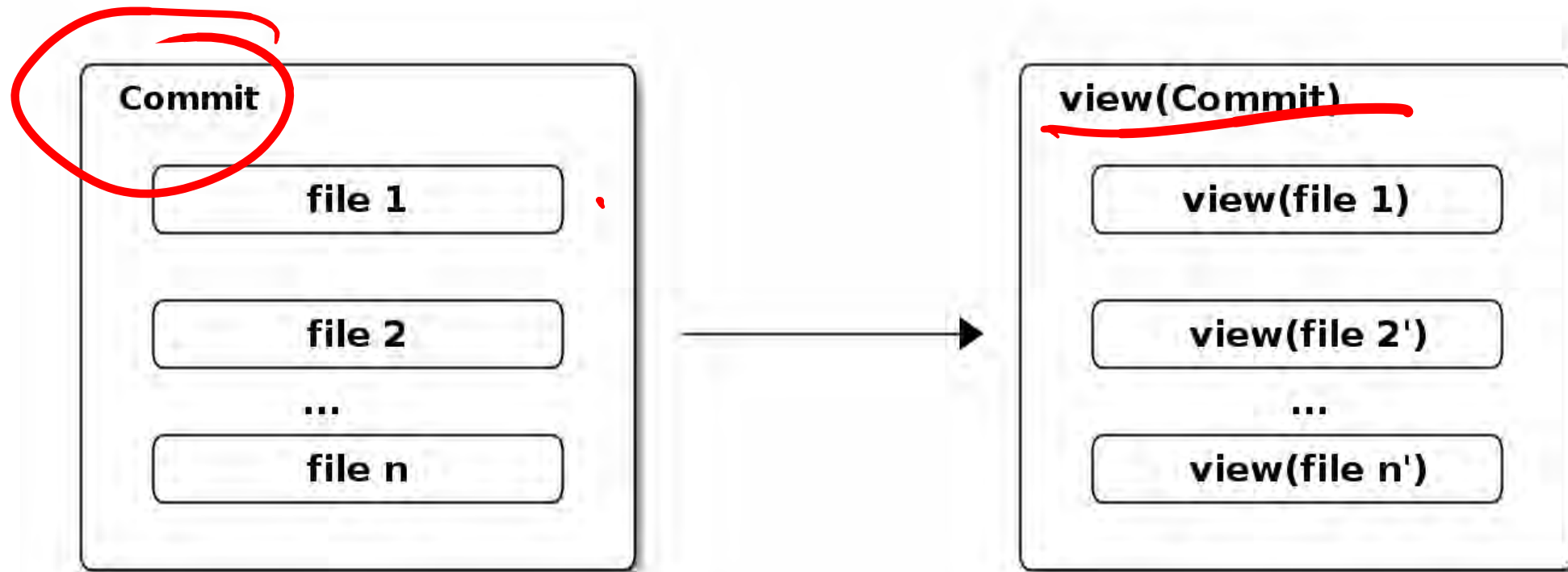
file

filter(file)

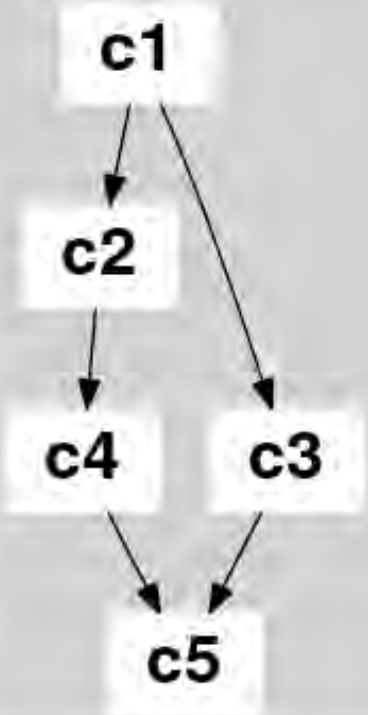
view(file)

```
void alloc_report(void)
{
    REPORT(blob, struct blob);
    REPORT(tree, struct tree);
    REPORT(commit, struct commit);
    REPORT(tag, struct tag);
    REPORT(object, union any_object);
}
```

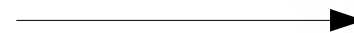
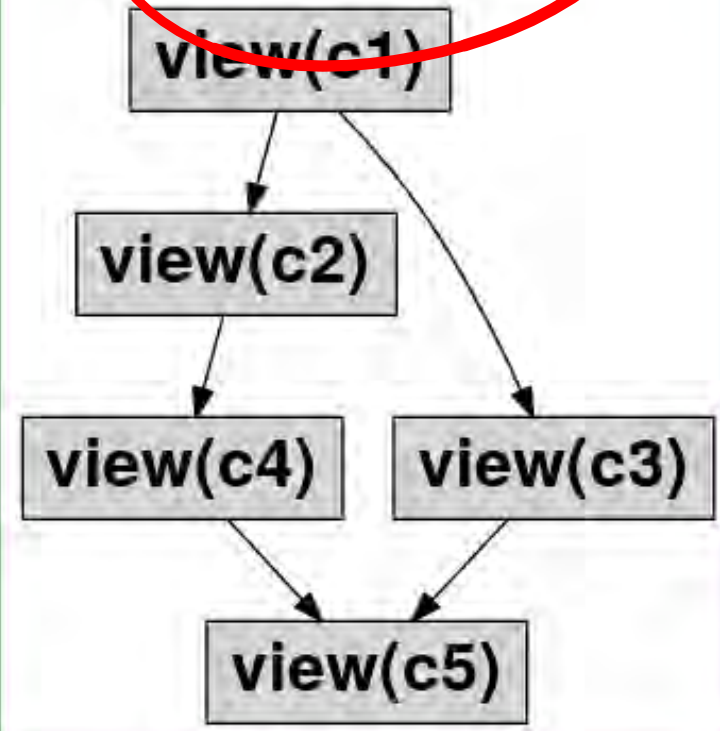
```
name|void
name|alloc_report
parameter_list|(
name|void
parameter_list|)
block|{
name|REPORT
argument_list|(
name|blob
argument_list|,
expr|struct
name|blob
argument_list|)
expr_stmt|;
name|REPORT
argument_list|(
name|tree
argument_list|,
expr|struct
name|tree
argument_list|)
```



Original Repo



View Repo



History for **git** / **alloc.c**

Commits on Jul 13, 2014

alloc: factor out commit index ...
peff committed with gitster on 13 Jul 2014

move setting of object->type to alloc_* functions ...
peff committed with gitster on 13 Jul 2014

alloc: write out allocator definitions ...
peff committed with gitster on 13 Jul 2014

alloc.c: remove the alloc_raw_commit_node() function ...
Ramsay Jones committed with gitster on 13 Jul 2014

Commits on Jun 12, 2014

commit: push commit_index update into alloc_commit_n ...
peff committed with gitster on 10 Jun 2014

alloc: include any-object allocations in alloc_report ...
peff committed with gitster on 10 Jun 2014

Commits on Oct 16, 2013

History for **git-token** / **alloc.c**

our user

Commits on Jul 23, 2016

alloc: factor out commit index ...
peff committed with dmgerman on 13 Jul 2014

move setting of object->type to alloc_* functions ...
peff committed with dmgerman on 13 Jul 2014

alloc: write out allocator definitions ...
peff committed with dmgerman on 13 Jul 2014












alloc.c: remove the alloc_raw_commit_node() function ...
Ramsay Jones committed with dmgerman on 13 Jul 2014

commit: push commit_index update into alloc_commit_node ...
peff committed with dmgerman on 10 Jun 2014

alloc: include any-object allocations in alloc_report ...
peff committed with dmgerman on 10 Jun 2014

 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	107 108 void alloc_report(void) 109 {
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	110 REPORT(blob, struct blob); 111 REPORT(tree, struct tree);
 alloc.c: remove the alloc_raw_c... 225ea22 Ramsay Jones authored on 13 Jul 2014	112 REPORT(commit, struct commit);
 alloc: include any-object allocati... c335d74 peff authored on 10 Jun 2014	113 REPORT(tag, struct tag); 114 REPORT(object, union any_object);
 Add specialized object allocator 855419f Linus Torvalds authored on 19 Jun 2006	115 }

Per Line

 Linus Torvalds authored on 19 Jun 2006	505 name alloc_report 506 parameter_list (507 name void 508 parameter_list) 509 block { 510 name REPORT 511 argument_list (512 name blob 513 argument_list , 514 expr struct 515 name blob 516 argument_list) 517 expr_stmt ; 518 name REPORT 519 argument_list (520 name tree 521 argument_list , 522 expr struct 523 name tree 524 argument_list) 525 expr_stmt ; 526 name REPORT 527 argument_list (528 name commit	 alloc.c: remove the alloc_raw_c... 297856b Ramsay Jones authored on 13 Jul 2014	528 name commit
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014		 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	529 argument_list , 530 expr struct 531 name commit
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006		 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	532 argument_list) 533 expr_stmt ; 534 name REPORT 535 argument_list (536 name tag
 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014		 alloc: include any-object allocati... 9224247 peff authored on 10 Jun 2014	537 argument_list , 538 expr struct 539 name tag 540 argument_list) 541 expr_stmt ; 542 name REPORT 543 argument_list (544 name object 545 argument_list , 546 expr union 547 name any_object
 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006		 Add specialized object allocator a0345b4 Linus Torvalds authored on 19 Jun 2006	548 argument_list) 549 expr_stmt ; 550 block }
 alloc.c: remove the alloc_raw_c... 297856b			

Per Token

commit: push commit_index update into alloc_commit_node

[Browse files](#)

Whenever we create a commit object via `lookup_commit`, we give it a unique index to be used with the `commit-slab` API. The theory is that any "struct commit" we create would follow this code path, so any such struct would get an index. However, callers could use `alloc_commit_node()` directly (and get multiple commits with index 0).

Let's push the indexing into `alloc_commit_node` so that it's hard for callers to get it wrong.

Signed-off-by: Jeff King <peff@peff.net>

Signed-off-by: Junio C Hamano <gitster@pobox.com>

🔗 master 📁 v2.10.0-rc1 ... v2.0.2



peff committed with gitster on 10 Jun 2014

1 parent [c335d74](#) commit [969eba6341a5af8ac52c67e26462548ed05e23e3](#)

📄 Showing 2 changed files with 10 additions and 4 deletions.

[Unified](#)[Split](#)

12 ■■■■ alloc.c

[View](#)

@@ -47,10 +47,18 @@ union any_object {

```
47
48 DEFINE_ALLOCATOR(blob, struct blob)
49 DEFINE_ALLOCATOR(tree, struct tree)
50 -DEFINE_ALLOCATOR(commit, struct commit)
51 DEFINE_ALLOCATOR(tag, struct tag)
52 DEFINE_ALLOCATOR(object, union any_object)
```

```
-47
-48 DEFINE_ALLOCATOR(blob, struct blob)
-49 DEFINE_ALLOCATOR(tree, struct tree)
50 +DEFINE_ALLOCATOR(raw_commit, struct commit)
51 DEFINE_ALLOCATOR(tag, struct tag)
52 DEFINE_ALLOCATOR(object, union any_object)
```

commit: push commit_index update into alloc_commit_node

[Browse files](#)

Whenever we create a commit object via `lookup_commit`, we give it a unique index to be used with the `commit-slab` API. The theory is that any "struct commit" we create would follow this code path, so any such struct would get an index. However, callers could use `alloc_commit_node()` directly (and get multiple commits with index 0).

Let's push the indexing into `alloc_commit_node` so that it's hard for callers to get it wrong.

Signed-off-by: Jeff King <peff@peff.net>

Signed-off-by: Junio C Hamano <gitster@pobox.com>

cregit: 969eba6341a5af8ac52c67e26462548ed05e23e3 <http://github.com/git/git/commit/969eba6341a5af8ac52c67e26462548ed05e23e3>

🔗 master



peff committed with dmgerman on 10 Jun 2014

1 parent [9224247](#)

commit [f9bf93c4b8683a42e6863e7faf1223c8dfc28478](#)

📄 Showing 2 changed files with 35 additions and 16 deletions.

Unified

Split

37 ■■■■ alloc.c

View

✳ @@ -104,7 +104,7 @@ end_macro

104 begin_macro

105 name|DEFINE_ALLOCATOR

106 argument_list|(
~~107 -argument|commit~~

108 argument_list|,

109 argument|struct commit

104 begin_macro

105 name|DEFINE_ALLOCATOR

106 argument_list|(
107 +argument|raw_commit

108 argument_list|

109 argument|struct commit


Up to 4.7

Language	SLOCs				Tokens	
	files	blank	comment	code	#Comments	code
C	23,040	2,223,857	2,046,386	11,331,924	879,186	69,563,180
C/C++ Header	17,506	426,656	729,310	2,797,340	461,940	12,379,194
	40,546	2,650,513	2,775,696	14,129,264	1,341,126	81,942,374

Persons in blame:


Line: 12,005

Token: 12,087

 **vsprintf.c: use noinline_for_stack** [cf3b429](#)
JoePerches authored on 24 May 2010

 **Linux-2.6.12-rc2** [1da177e](#)
Linus Torvalds authored on 16 Apr 2005


 **vsprintf: give it some care to ple...** [7b9186f](#)
andre-rosa authored on 14 Dec 2009

 **Linux-2.6.12-rc2** [1da177e](#)
Linus Torvalds authored on 16 Apr 2005

 **lib/vsprintf.c: replace while with ...** [43e5b66](#)
Villemoes authored on 12 Feb 2015

 **Linux-2.6.12-rc2** [1da177e](#)
Linus Torvalds authored on 16 Apr 2005

 **lib/vsprintf.c: replace while with ...** [43e5b66](#)
Villemoes authored on 12 Feb 2015

 **vsprintf: give it some care to ple...** [7b9186f](#)
andre-rosa authored on 14 Dec 2009

 **Linux-2.6.12-rc2** [1da177e](#)
Linus Torvalds authored on 16 Apr 2005

```
121 static noinline_for_stack
122 int skip_atoi(const char **s)
123 {
124     int i = 0;
125
126     do {
127         i = i*10 + *((*s)++) - '0';
128     } while (isdigit(**s));
129
130     return i;
131 }
132
```

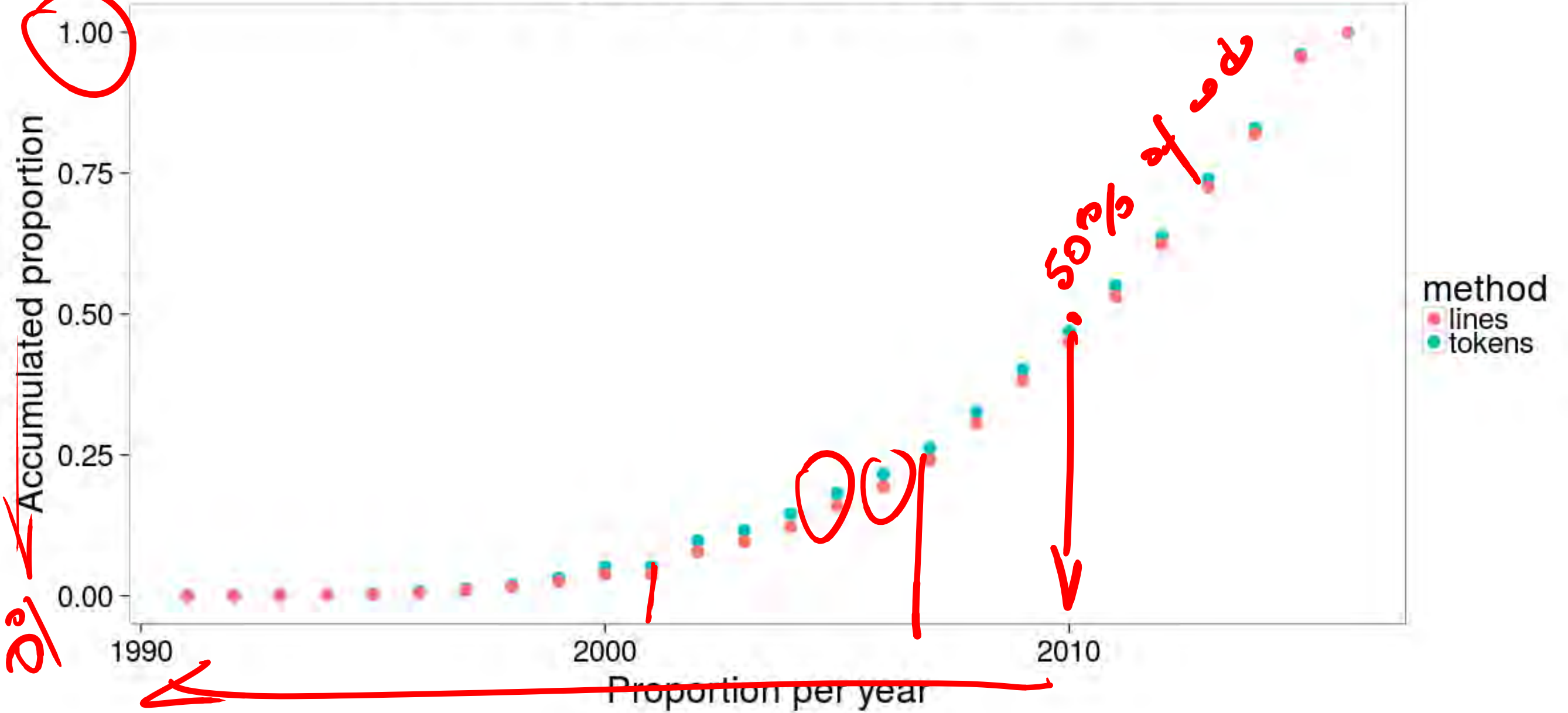
Switch to demo

Aggregated information

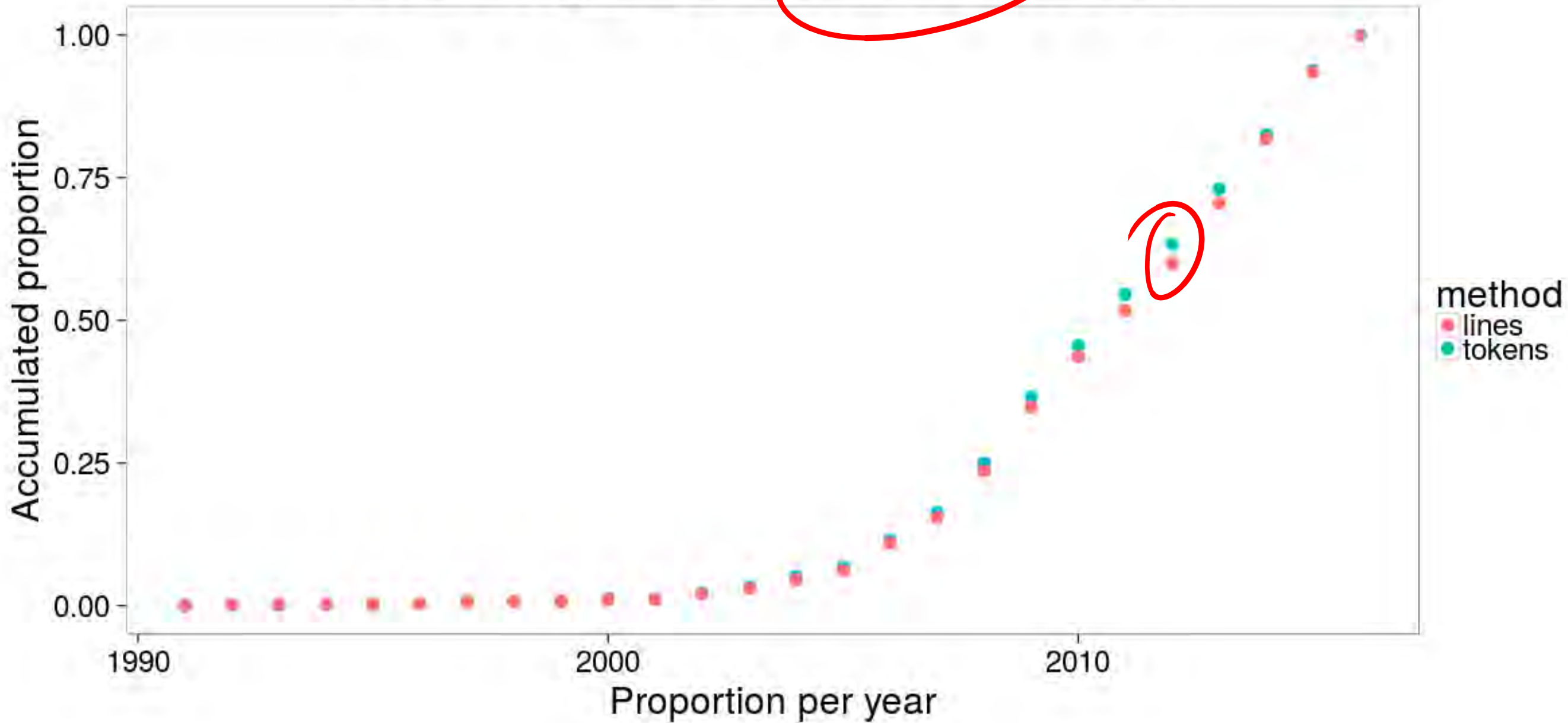
Comparing token vs Line

- Does it make a difference to count contribution
 - Per line, vs
 - Per Token
- Eg:
 - Linux Kernel Report

Year of origin of code in Linux as of v4.7



Year of origin of code in kernel/ as of v4.7



Linux

Token

index	personid	ctokens	prop	accum
1	pre-git	4286384	5.15	5.15
2	alex deucher	2532472	3.04	8.19
3	larry finger	1424969	1.71	9.9
4	linus torvalds	1077539	1.29	11.19
5	jaroslav kysela	983822	1.18	12.37
6	peng tao	731861	0.88	13.25
7	andrew morton	728759	0.88	14.13
8	mark brown	648497	0.78	14.91
9	david howells	624218	0.75	15.66
10	hans verkuil	601867	0.72	16.38
11	ben skeggs	590972	0.71	17.09
12	david s. miller	541773	0.65	17.74
13	arend van spriel	501988	0.6	18.34
14	david daney	491898	0.59	18.93
15	alan cox	469412	0.56	19.5
16	laurent pinchart	461247	0.55	20.05
17	christoph hellwig	427959	0.51	20.56
18	mauro carvalho chehab	404112	0.49	21.05
19	johannes berg	401906	0.48	21.53
20	takashi iwai	367443	0.44	21.97

Line

index	personid	clines	prop	accum
1	pre-git	745226	3.81	3.81
2	alex deucher	609927	3.12	6.93
3	larry finger	295550	1.51	8.44
4	linus torvalds	247626	1.27	9.7
5	jaroslav kysela	185692	0.95	10.65
6	david howells	181493	0.93	11.58
7	mark brown	156540	0.8	12.38
8	peng tao	149764	0.77	13.14
9	andrew morton	135377	0.69	13.83
10	joe perches	125900	0.64	14.48
11	ben skeggs	123504	0.63	15.11
12	david daney	123350	0.63	15.74
13	hans verkuil	123131	0.63	16.37
14	mauro carvalho chehab	119616	0.61	16.98
15	david s. miller	118700	0.61	17.59
16	alan cox	106285	0.54	18.13
17	johannes berg	105760	0.54	18.67
18	christoph hellwig	103596	0.53	19.2
19	laurent pinchart	99141	0.51	19.71
20	takashi iwai	96905	0.5	20.2

kernel/

Token

Line

index	personid	clines	prop	accum	index	personid	ctokens	prop	accum
1	steven rostedt	27274	10.38	10.38	1	steven rostedt	94860	9.81	9.81
2	peter zijlstra	19704	7.5	17.88	2	peter zijlstra	80844	8.36	18.18
3	thomas gleixner	14139	5.38	23.26	3	paul e. mckenney	47737	4.94	23.12
4	paul e. mckenney	13353	5.08	28.34	4	thomas gleixner	43543	4.51	27.62
5	tejun heo	11482	4.37	32.71	5	tejun heo	33074	3.42	31.05
6	ingo molnar	9522	3.62	36.33	6	jason wessel	32133	3.32	34.37
7	jason wessel	8216	3.13	39.46	7	ingo molnar	29380	3.04	37.41
8	frederic weisbecker	6961	2.65	42.1	8	frederic weisbecker	28123	2.91	40.32
9	rafael j. wysocki	6264	2.38	44.49	9	alexei starovoitov	27944	2.89	43.21
10	alexei starovoitov	6231	2.37	46.86	10	tom zanussi	25010	2.59	45.8
11	tom zanussi	6144	2.34	49.2	11	rafael j. wysocki	21292	2.2	48.0
12	oleg nesterov	5314	2.02	51.22	12	oleg nesterov	18696	1.93	49.94
13	eric w. biederman	3987	1.52	52.74	13	eric w. biederman	17823	1.84	51.78
14	rusty russell	3558	1.35	54.09	14	rusty russell	14845	1.54	53.32
15	masami hiramatsu	3166	1.2	55.29	15	al viro	14210	1.47	54.79
16	al viro	3000	1.14	56.44	16	masami hiramatsu	13423	1.39	56.18
17	pre-git	2737	1.04	57.48	17	pre-git	11244	1.16	57.34
18	john stultz	2576	0.98	58.46	18	srikar dronamraju	10757	1.11	58.45
19	dave young	2546	0.97	59.43	19	eric paris	10311	1.07	59.52
20	jiri olsa	2489	0.95	60.37	20	dave young	10197	1.06	60.57

Many small changes

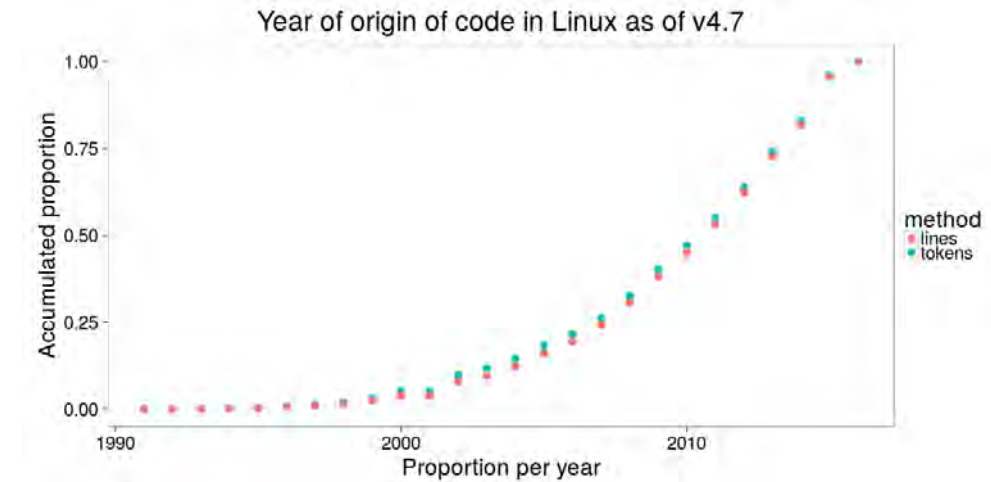
- **9.5** % of non-merge commits added **3 or less** tokens and removed **3 or less** tokens

Warning

- cregit does not tell you where the code comes from
 - Nor who authored it
- It only gives you one more layer of information
- True provenance might never be feasible (in practice)

cregit: what does it tell us

- On the large
 - Token and Line are equivalent



- On the small
 - Provide a fine grained view of the evolution of the code

lib/ctype.c at v4.7

```
/*  
 * linux/lib/ctype.c  
 *  
 * Copyright (C) 1991, 1992 Linus Torvalds  
 */
```

```
#include <linux/ctype.h>  
#include <linux/compiler.h>  
#include <linux/export.h>
```

```
const unsigned char _ctype[] = {  
_C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, /* 0-7 */  
_C, _C|_S, _C|_S, _C|_S, _C|_S, _C|_S, _C|_S, _C, _C, _C, _C, /* 8-15 */  
_C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, /* 16-23 */  
_C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, _C, /* 24-31 */  
_S|_SP, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, /* 32-39 */  
_P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, /* 40-47 */  
_D, _D, _D, _D, _D, _D, _D, _D, _D, _D, _D, _D, _D, _D, _D, _D, /* 48-55 */  
_D, _D, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, /* 56-63 */  
_P, _U|_X, _U|_X, _U|_X, _U|_X, _U|_X, _U|_X, _U|_X, _U|_X, _U, /* 64-71 */  
_U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, /* 72-79 */  
_U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, _U, /* 80-87 */  
_U, _U, _U, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, _P, /* 88-95 */  
_P, _L|_X, _L|_X, _L|_X, _L|_X, _L|_X, _L|_X, _L|_X, _L|_X, _L, /* 96-103 */
```

```
494 pre-git          1991-09-17  
2 pre-git          1991-11-11  
1 pre-git          1992-05-22  
1 pre-git          1992-08-01  
98 pre-git         1996-07-02  
12 Arnaldo Carvalho de Melo 2003-10-07  
1 André Goddard Rosa 2009-12-14  
6 Paul Gortmaker  2011-11-16
```

- pre-git 585
- arnaldo carvalho de melo 7
- paul gortmaker 4
- andre goddard rosa 1

What is next? Who reviewed the code

- We are currently adding the review information in mailings lists to the kernel
 - We have excellent heuristics to **match** email reviews to a given commit

What is next: cregit future

- Enhance the commit/change information
 - Origin analysis: where did the code come from?
 - Who truly created/authored the code?
 - Is the new code a derivation of other code?

Code is a sequence of changes

- Each change (the commit)
 - Has one or many copyright-authors
 - It might be based on previously written code
 - Moved code
 - Copied code
 - Improved code
 - Its history might be recorded somewhere else

Code is a sequence of changes

- **The *current*** version is the composition of commits from beginning of (project) time
- A commit is a “**function**” that **changes** the **previous code** into the **new code**
- This **function** might have **multiple inputs**
 - Previous version of the code
 - External code
 - This code might be created in a different repository and **might already have a rich history**
 - Knowledge from the developer (the creative part)
 - Generated code

What is next: cregit future

- Create new **views** of history
 - As long as it is diffable, cregit can track its evolution
- Use cregit as a **tool** to inspect **history**
 - Example: (switch to github)

cregit: identifying contributors of source code

Daniel M German
University of Victoria
dmg@uvic.ca

Kate Stewart
Linux Foundation

Bram Adams
Polytechnique of Montreal

cregit: identifying contributors of source code

Daniel M German
University of Victoria
dmg@uvic.ca

Kate Stewart
Linux Foundation

Bram Adams
Polytechnique of Montreal

The screenshot displays a code analysis tool interface. On the left, a list of commits is shown, each with a contributor's profile picture, name, and a commit ID. The commit 'alloc.c: remove the alloc_raw_c...' by Ramsay Jones is highlighted in yellow. On the right, a code snippet for 'void alloc_report(void)' is shown. The code contains several 'REPORT' calls. A yellow box highlights the 'REPORT(commit, struct commit);' line, with arrows pointing to the commit entry for Ramsay Jones in the commit list. Other contributors like Linus Torvalds and peff are also listed in the commit list and have their names highlighted in the code snippet.

```
void alloc_report(void)
{
    REPORT(blob, struct blob);
    REPORT(tree, struct tree);
    REPORT(commit, struct commit);
    REPORT(tag, struct tag);
    REPORT(object, union any_object);
}
```

cregit: identifying contributors of source code

Daniel M German
University of Victoria
dmg@uvic.ca

Kate Stewart
Linux Foundation

Bram Adams
Polytechnique of Montreal

The screenshot shows a diff of the file 'alloc.c'. The left side lists commits with their authors and dates. The right side shows the code changes. A commit by Ramsay Jones (225e422) is highlighted in yellow. The code shows a function 'void alloc_report(void)' with several 'REPORT' calls. Arrows point from the commit list to the corresponding lines in the code diff.

```

    sprint_backtrace(sym, value);
else if (*fmt != 'f' && *fmt != 's')
    sprint_symbol(sym, value);
else
    sprint_symbol_no_offset(sym, value);

return string(buf, end, sym, spec);
#else
return special_hex_number(buf, end, value, sizeof(void *));
#endif
}
    
```

Contributors

	Person	Tokens	Prop
linus torvalds	linus torvalds	65	44.83%
joe perches	joe perches	39	26.90%
frederic weisbecker	frederic weisbecker	16	11.03%
namhyung kim	namhyung kim	13	8.97%

cregit: contributors to the Linux kernel

Home

Release 4.8

Release 4.7

Cregit version 1.0-rc1

cregit: identifying contributors of source code

Daniel M German
University of Victoria
dmg@uvic.ca

Kate Stewart
Linux Foundation

Bram Adams
Polytechnique of Montreal

The image shows a commit history on the left and a code diff on the right. The commit history lists several commits, including 'Add specialized object allocator' by Linus Torvalds and 'alloc.c: remove the alloc_raw_c...' by Ramsay Jones. The code diff shows the implementation of the 'alloc_report' function, with lines 109-115. The diff highlights changes in the 'REPORT' macro and the 'alloc_report' function body. Arrows point from the commit history to the corresponding lines in the code diff.

Code is a sequence of changes

- **The current** version is the composition of commits from beginning of (project) time
- A commit is a “**function**” that **changes** the **previous code** into the **new code**
- This function might have multiple inputs
 - Previous version of the code
 - External code
 - This code might be created in a different repository and **might already have a rich history**
 - Knowledge from the developer (the creative part)
 - Generated code

The image shows the cregit website interface. The top navigation bar includes 'cregit: contributors to the Linux kernel', 'Home', 'Release 4.8', 'Release 4.7', and 'Cregit version 1.0-rc1'. A sidebar on the left contains a hamburger menu icon.

```

    sprint_backtrace(sym, value);
else if (*fmt != 'f' && *fmt != 's')
    sprint_symbol(sym, value);
else
    sprint_symbol_no_offset(sym, value);

return string(buf, end, sym, spec);
#else
return special_hex_number(buf, end, value, sizeof(void *));
#endif
}
    
```

Contributors

	Person	Tokens	Prop
linus torvalds	linus torvalds	65	44.83
joe perches	joe perches	39	26.90
frederic weisbecker	frederic weisbecker	16	11.03
namhyung kim	namhyung kim	13	8.97