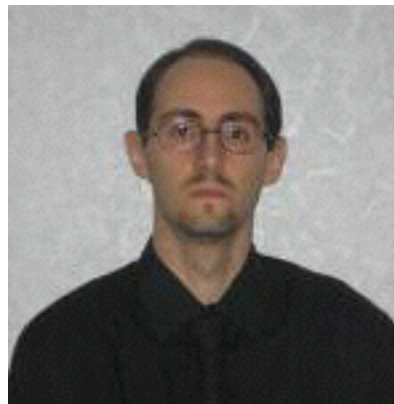


Code siblings: technical and legal implications of copying code between applications



Daniel M. German
University of Victoria
Canada



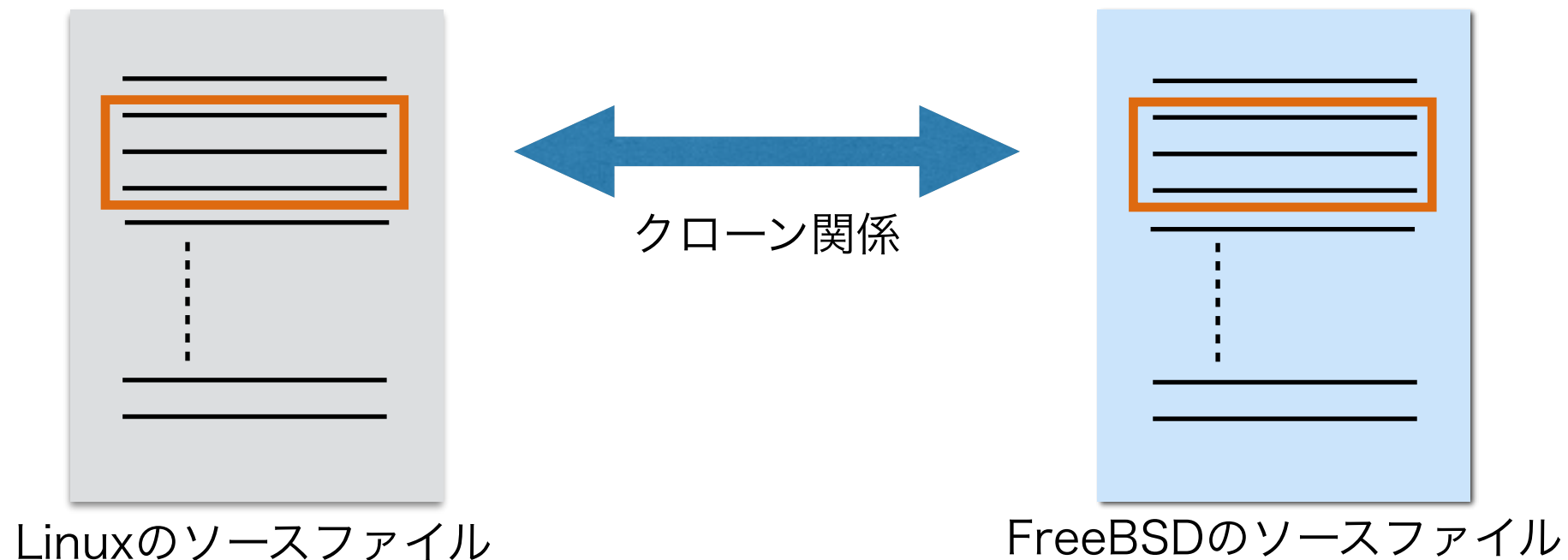
Massimiliano Di Penta
University of Sannio
Italy



Yann-Gaël Guéhéneuc
École Polytechnique
de Montréal
Canada

Code Siblingとは

- ✧ 異なるソフトウェア間でクローン関係にあるコード片



- ✧ Code Siblingが作成される理由
 - 他のソフトウェアの機能を，そのままコピーして取り入れる
 - 効率的な保守を行うため，コンポーネントを共通化している

背景

✧ Code Siblingの作成

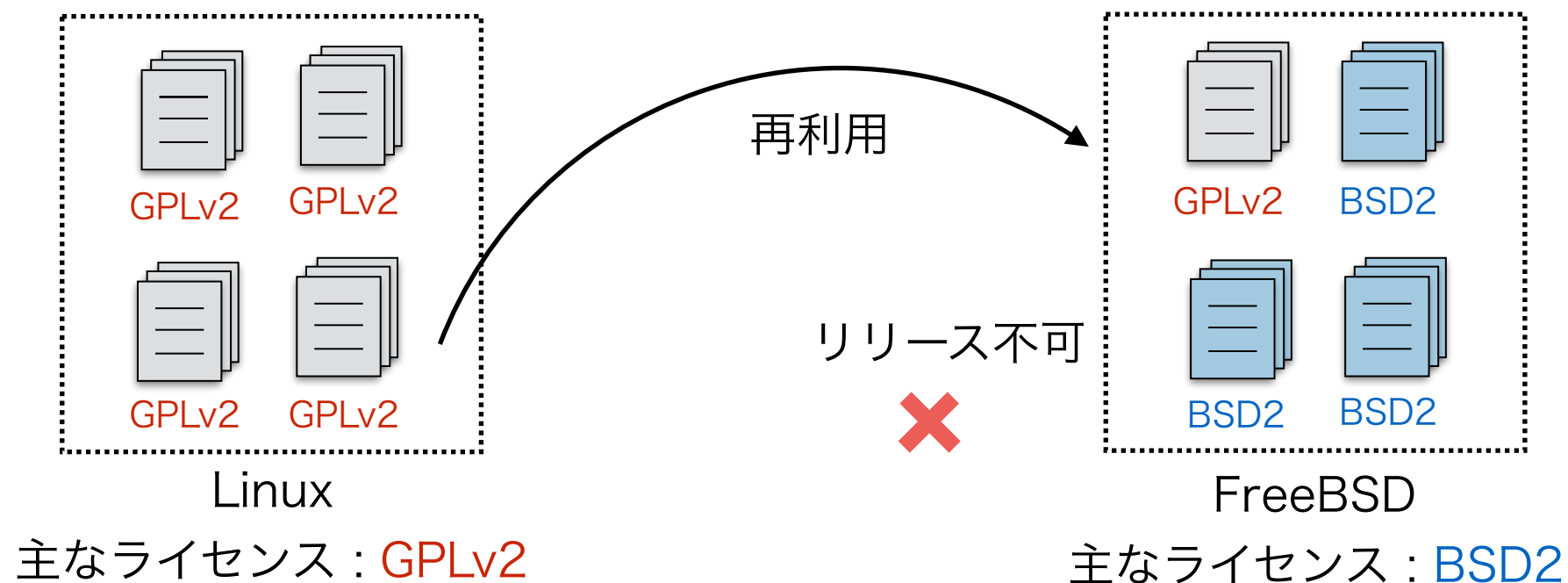
- Code Siblingは、他のソフトウェアのソースコードを再利用することによって作成される
- Code Siblingは、特に、ドライバ等のコンポーネントが多く存在するOSなどで多く見られる

✧ Code Siblingの作成は、法的問題をクリアした場合のみ許される

- ライセンス不整合問題

ライセンス不整合とは

- ❖ 複数のライセンスでリリースされているソフトウェアにおいて、ライセンスを遵守できていない状態



- リリースすると、GPLv2のコピーレフト条項に違反する
 - ★ GPLのソフトウェアの2次著作物は、GPLでなければならない

制約の強いライセンスから制約の弱いライセンスへの
再利用は不整合を起こしやすい

問題

- ✧ ライセンスの不整合問題が、Code Siblingの作成に影響を与えている可能性がある
 - Code Siblingの作成を妨げているライセンスがあるか？
 - ★ 再利用元や再利用先のライセンスによっては、Code Siblingを作成することができない
 - Code Siblingが広がっていきやすいライセンスがあるか？
 - ★ Code Siblingが制約の弱いライセンスだと、不整合問題が起こりにくい

ソフトウェアを再利用して開発を行う場合、ソフトウェアの**ライセンスの決定**は重要なプロセスとなる

目的

- ✧ ライセンスが、Code Siblingの作成に与える影響を明らかにする
- ✧ Research Question
 - RQ1 : Unix系OSには、どれくらいのライセンスが含まれているか
 - RQ2 : Unix系OS間で、どれくらいのCode Siblingが存在するか
 - **RQ3 : Code Siblingは、どのようなライセンスが使われているか、また、ライセンスが異なっているのはなぜか**

データセット

- * Linux , FreeBSD, OpenBSDの最新スナップショット
 - 選んだ理由：最も古いファイル（1993~1995年）が同年代であるため、検出できないCode Siblingが少ない

Table 1. Characteristics of the three kernels.

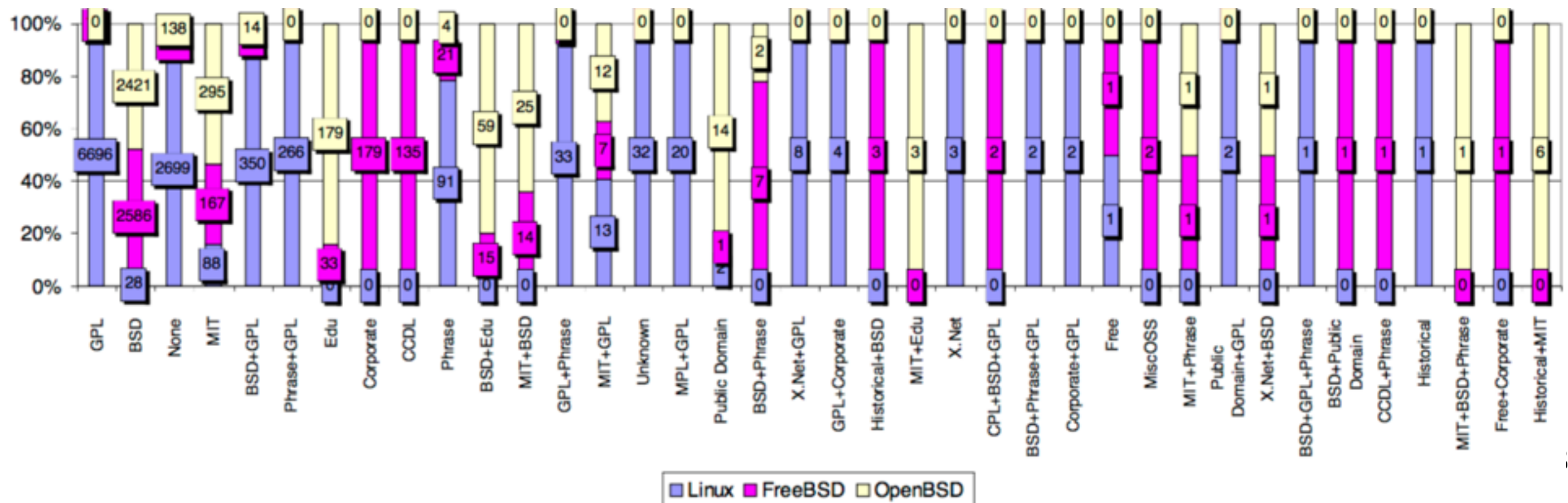
	Linux	FreeBSD	OpenBSD
Snapshot/release dates	Jan, 18 2009	Jan, 23 2009	Jan, 27 2009
# of .c files	10,343	13,699	3,175
# of .h files	9,422	7,855	3,468
KLOC	7,262	8,456	2,301

RQ1の分析方法と結果

＊ RQ1：Unix系OSには，どれくらいのライセンスが含まれているか

- 分析方法：スナップショットのライセンスを，ライセンス特定ツールFossologyで特定する

- 実験結果



RQ2の分析方法

✧ RQ2 : Unix系OS間で、どれくらいのCode Siblingが存在するか

- 分析方法

1.3つのスナップショットを1つにして、CCFinderXを使ってコードクローンを検出

2.検出されたクローンのうち、クローンペアがそれぞれ異なるシステムに存在するクローンをCode Siblingとする

3.クローンペアが複数ファイルに渡って存在しているものを省く

RQ2の結果

- ❖ RQ2 : Unix系OS間で、どれくらいのCode Siblingが存在するか

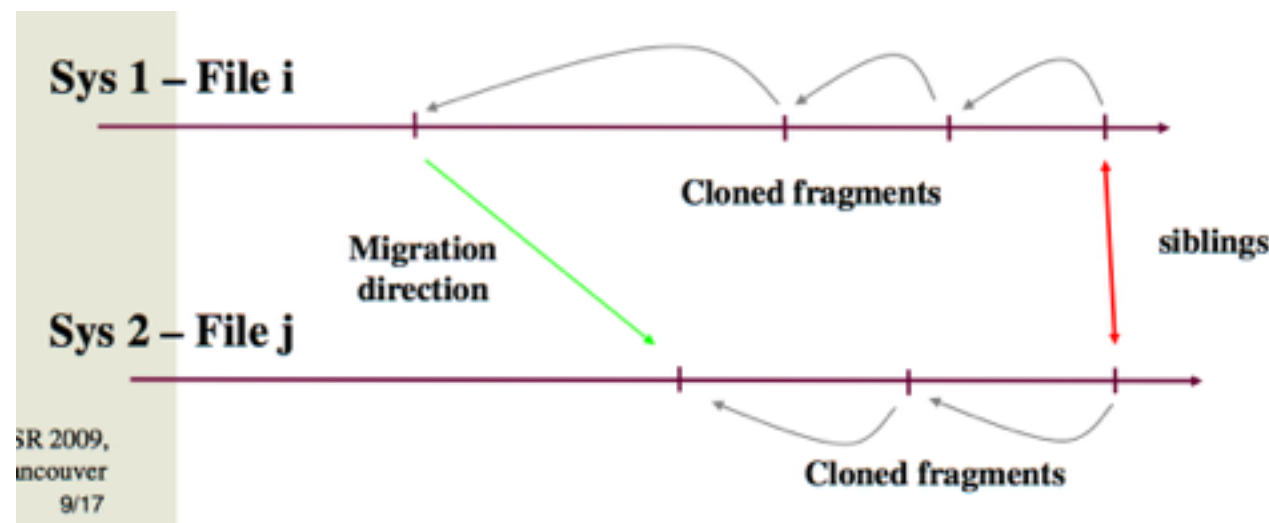
Table 2. Number of files and file pairs with clones between Linux and the BSDs.

CCFinder Outputs		
Metrics	FreeBSD vs. Linux	OpenBSD vs. Linux
# of Clone pairs	2208	749
# Files Linux	296	100
# Files BSD	289	94
# File Pairs	412	161
# File Pairs (same name)	189	47

Filtering Results		
Metrics	FreeBSD vs. Linux	OpenBSD vs. Linux
# Files Linux	224	63
# Files BSD	226	61
# File Pairs	238	73
# File Pairs (same name)	185	45

RQ3の分析方法

- ✧ RQ3 : Code Siblingは, どのようなライセンスが使われているか, また, ライセンスが異なっているのはなぜか
 - 分析方法 : Code Siblingが作成されたリビジョンを特定する
 - ★ Code Siblingが以前のリビジョンでも存在したかどうか, Linux側とBSD側でClone検出しながら遡っていく



RQ3の結果(1/2)

✧ RQ3 : Code Siblingは, どのようなライセンスを含んでいるか

✧ 表4から以下のことが分かった

- ✧ FreeBSD→Linuxが多い
- ✧ FreeBSDのCode Siblingのほとんどは, 2001年以前に作成されている
- ✧ ほとんどのCode Siblingのライセンスは, 異なっている
- ✧ 特に, Linux→FreeBSDでは, ほとんどが制約の弱いライセンスに変更されている

Table 4. Linux vs. FreeBSD siblings with license change.

Siblings introduced first in FreeBSD and then in Linux				
License FreeBSD	License Linux	# of Files	Before 2002/01/01	After 2002/01/02
BSD	GPL	8	1	7
BSD	MIT	2	–	2
BSD	None	2	–	2
Corporate	BSD+GPL	89	79	10
GPL	None	1	1	–
Phrase	BSD+GPL	1	1	–
X.Net+BSD	MIT	1	–	1
TOTAL		104	82	22

Siblings introduced first in Linux and then in FreeBSD				
License Linux	License FreeBSD	# of Files	Before 2002/01/01	After 2002/01/02
BSD+GPL	Corporate	8	–	8
GPL	BSD	17	–	17
GPL	BSD+GPL	1	–	1
GPL	CPL+BSD+GPL	1	–	1
MIT	BSD	1	–	1
MIT+GPL	None	2	–	2
None	BSD	1	–	1
Phrase+GPL	MIT	2	–	2
TOTAL		33	–	33

RQ3の結果(2/2)

- ✧ RQ3 : Code Siblingのライセンスが異なっているのはなぜか
- ✧ SCSI AIC 7xxx Driver (1994年にLinuxで誕生)
 - 1.1995年, OpenBSDがLinuxのコードを導入する
 - 2.NetBSDがOpenBSDをフォーク
 - 3.1996年, FreeBSDがNetBSDのドライバを導入

2つのBSD系OSを経由して, FreeBSDにコードが混入
Linuxにコードを追加した開発者とFreeBSDに追加した開発者は同じ人

Code Siblingの例

```
/* FreeBSD */
struct ahd_pci_identity *
ahd_find_pci_device(aic_dev_softc_t pci)
{
    ...
    ... ahd_pci_identity ...
    ...
    ... aic_pci_read_config ...
    ... aic_pci_read_config ...
    ... aic_pci_read_config ...
    ... aic_pci_read_config ...
    ... ahd_compose_id ...
    /* If we are configured to attach to HostRAID
    * controllers, mask out the IROC/HostRAID bit
    * in the */
    if (ahd_attach_to_HostRAID_controllers)
        full_id &= ID_ALL_IROC_MASK;
    [rest of function identical except
    for naming conventions]
}
```

```
/* Linux */
const struct ahc_pci_identity *
ahc_find_pci_device(ahc_dev_softc_t pci)
{
    ...
    const ... ahc_pci_identity ...
    ...
    ... ahc_pci_read_config ...
    ... ahc_pci_read_config ...
    ... ahc_pci_read_config ...
    ... ahc_pci_read_config ...
    ... ahc_compose_id ...
    /* If the second function is not hooked up, ignore it.
    * Unfortunately, not all MB vendors implement the
    * subdevice ID as per the Adaptec spec, so do our best
    * to sanity check it prior to accepting the subdevice
    * ID as valid. */
    if (ahc_get_pci_function(pci) > 0
        && ahc_9005_subdevinfo_valid(vendor, device, subvendor,
        subdevice) && SUBID_9005_MFUNCENB(subdevice) == 0)
        return (NULL);
    [rest of function identical except
    for naming conventions]
}
```

Figure 2. Excerpt from code siblings in the `aic7xxx` drivers of Linux and FreeBSD. (For the sake of brevity identical code has been replaced with “...”. Note how both functions use different naming conventions and one section of the function, the `if` statement, is very different in both.)

未解決な問題

- ✧ なぜ、Linuxの方が歴史がありコミュニティも大きいのに、FreeBSD➡LinuxのCode Siblingの混入が多いのか
 - 2001年までは、FreeBSDの勢いがあったかもしれない
 - Open BSDがリリースされたこともあり、2009年からFreeBSDのコミュニティが縮小している
- ✧ その他
 - FreeBSDにGPLが含まれているのはなぜか
 - Code Siblingを管理する方法は難しい

まとめ

- ✧ 複数のライセンスでリリースされているUnix系OS間のCode Siblingについて分析した

- わかったこと

- ★ 多くのCode Siblingを確認することができた

- ・ 特に, Free BSD→Linuxへの混入が多い

- ★ Code Siblingのライセンスは, 各プロジェクトで異なっていた

- ✧ 今後の課題

- ライセンス不整合チェックツール

- Code Siblingやライセンス変更が, 開発者やライセンスに影響を与える要因であるかどうかを調査していく