

An Exploratory Study of the Evolution of Software Licensing

Massimiliano Di Penta

Daniel M. German

Yann-Gael Gueheneuc, Giuliano Antoniol

背景

- FOSSは様々なライセンスの下でユーザーに利用可能となっている
- しかし法的な変更などで、ライセンス記述をアップデートする必要がある
- このアップデートは後にシステムを再利用する際に、問題が起きる可能性がある

目的

- ライセンスアップデートの分析
 - 開発者がどのようにしてライセンス記述を更新するのか
 - いつ、どのように変化するのを知りたい

RQ

1. どの程度の頻度でライセンス記述は変更されるのか？
2. どの程度の範囲でライセンス記述は変更されているのか？
3. 著作年の変更はライセンス記述にどのような影響を与えているか？

アプローチ

- ライセンスアップデートを自動的に追跡する手法を提案し、六つの異なるFOSSライセンスのアップデートを調査
- 対象FOSS
 - ArgoUML, Eclipse-JDT, FreeBSD , OpenBSD kernels, Mozilla, Samba.

対象FOSS

Table 3: Main characteristics of the six systems.

Characteristics	ArgoUML	Eclipse-JDT	FreeBSD	OpenBSD	Mozilla	Samba
Languages	Java	Java	C	C	C/C++	C
Release ranges	0.10-0.20	1.0-3.0	2.0-7.1	2.0-4.4	M3-1.7.13	1.9-3.0
#of source files ranges	777-1,421	578-3,274	895-6,729	3,359-6,483	4,845-12,436	299-860
KLOC ranges	129-280	79-697	325-3,292	994-2,242	1,827-4,104	156-332
CVS/SVN start dates	2000-09-14	2001-05-02	1993-06-12	1995-10-18	1998-03-28	1996-05-04
CVS/SVN end dates	2005-12-30	2006-11-07	2009-02-16	2009-02-07	2008-01-11	2004-04-03
Analyzed file revisions	32,582	128,611	195,077	110,430	468,747	29,018
# of committers	40	51	383	212	681	35

六つのシステムを選ぶ理由

- 長いリビジョンを持っているFOSSということ
- ライセンス変更の関連性と範囲の証明のために異なったシステムを調査している
- 大きなシステムなので、それによって様々なライセンス変化を観察できる

分析ステップ1

- ライセンス記述抽出
 - 一般的に最初にコメント文として記述されているのがライセンス記述なので最初のコメント文を抽出する
- ライセンス変化確認
 - Information Retrieval Vector Space Modelsを用いてタグづけを行い、コサイン類似性を確認して変化を確認する

分析ステップ2

- ライセンス分類

- Fossologyというライセンス分類ツールを使用する、分類は最初のリビジョンから開始
- 二つ続いてコサイン類似性が0.99未満の場合に改めてライセンス分類を行う

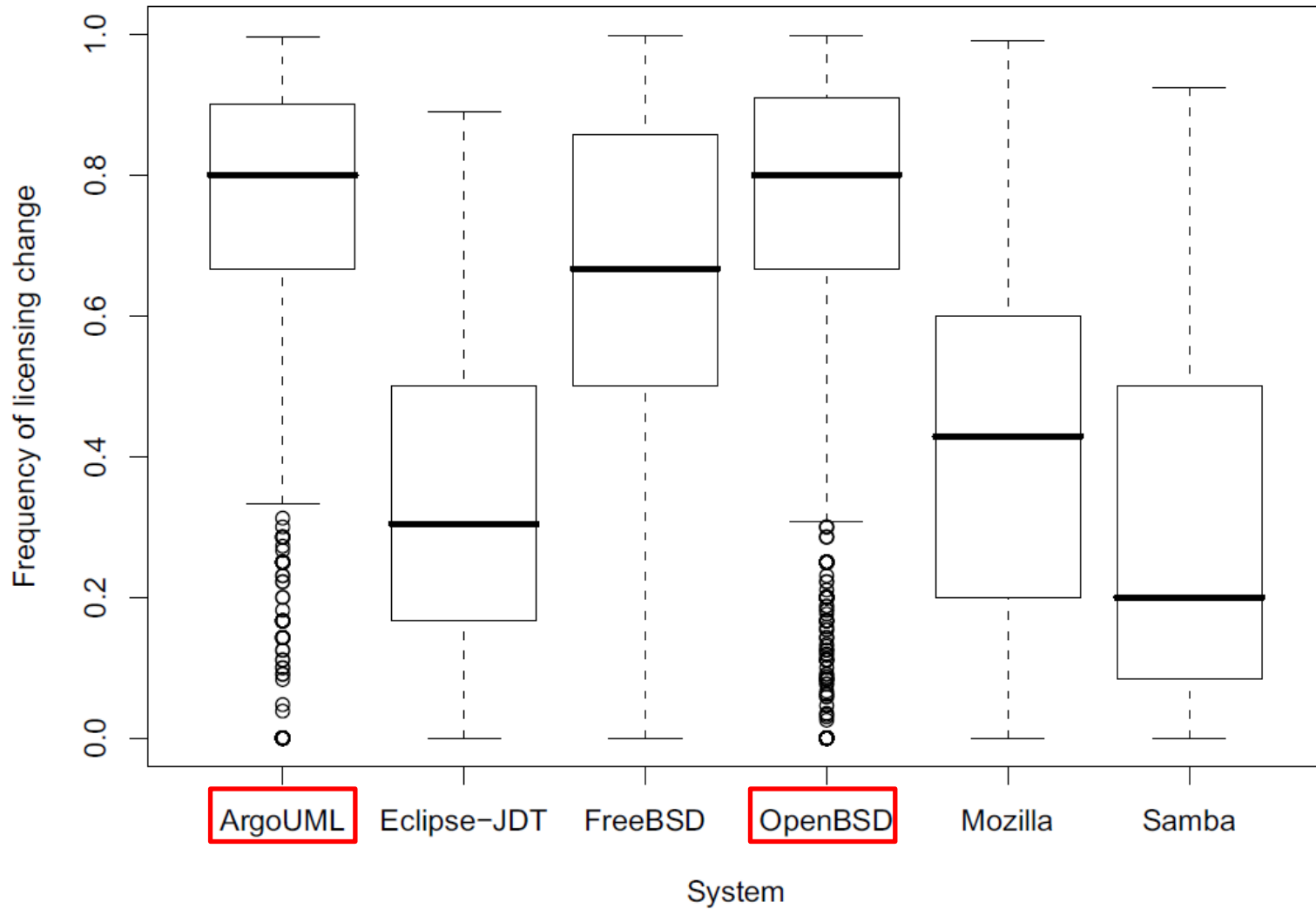
分析ステップ3

- 著作権年の変化の確認
 - 2桁または4桁の数列を抽出し、今回の年の範囲である1990~2009とあっているか確認する

RQ1

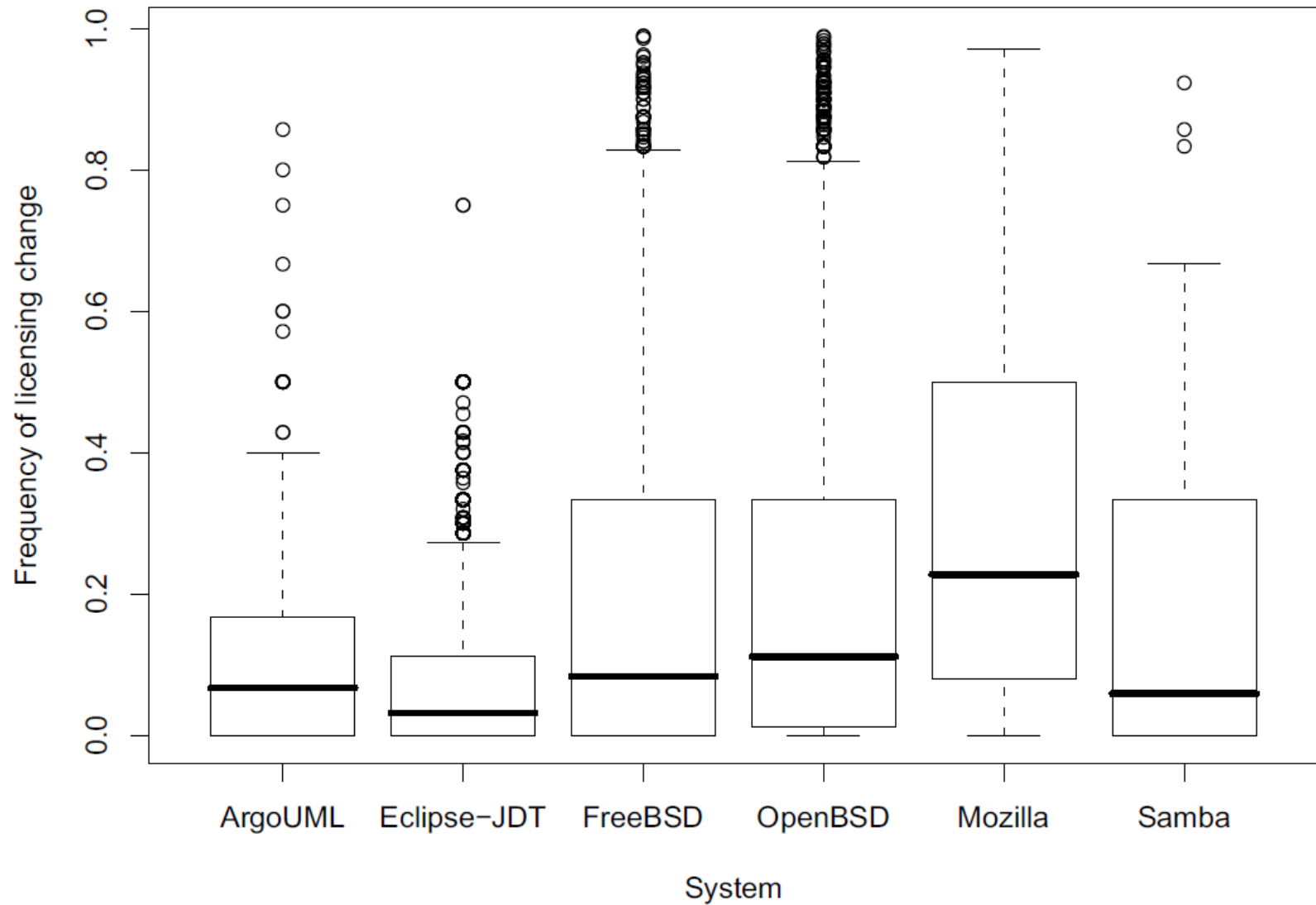
- どの程度の頻度でライセンス記述は変更されるのか？

RQ1 : 結果



(a) All changes

RQ1 : 結果



(b) Cosine < 0.99

RQ1 : 補足

- 図aは全ての変更の頻度の図です
- 図bは一つ前のリビジョンの類似性が0.99以下の場合の変化のみを考慮した場合のこと（つまり大きな変化の頻度です）

RQ1：結果

- FOSS毎によって変化傾向は異なっている
- OpenBSD,ArgoUMLが他のライセンスよりも大きな変化傾向がある
- 大きな変化はあまり起きない

RQ2

- どの程度の範囲でライセンス記述は変更されるのか？

RQ2 : 結果

Mozilla	
NPL → 'NPL v1.1'-style+GPL v2+LGPL v2.1	2914
NPL → 'Dual MPL GPL'-style+MPL	1274
'Dual MPL GPL'-style+MPL → NPL	1194
GPL v2+MPL → 'Dual MPL GPL'-style+MPL	942
MPL → 'MPL v1.1'-style+Dual MPL GPL	908
NPL → 'MPL v1.1'-style+Dual MPL GPL	543
NPL → GPL v2+MPL	375
MPL → 'Dual MPL GPL'-style+'MPL v1.1'-style	361
NPL → GPL v2+LGPL v2.1+NPL	149
GPL+NPL → 'MPL v1.1'-style+Dual MPL GPL	148
GPL+NPL → 'Dual MPL GPL'-style+MPL	144
Others	1736

RQ2 : 結果

- ArgoUML
 - 特に変化なし
- EclipseJDT
 - 最初はライセンスを備えておらずCPL-v0.5からCPL-v1.0に更新されて、最後にEPL-v1.0になった
- FreeBSD
 - 様々なライセンスが使われているがとくにBSDがよく利用されている、FreeBSD自体はBSD-clauses2の条件の元に利用できる

RQ2 : 結果

- OpenBSD
 - FreeBSDの派生なのでFreeBSDと似ている違いとしてはFreeBSDよりもBSD-clauses3の傾向がある
- Mozilla
 - NPLからGPL-v2、MPL-v1.0の二つに分離していった、バグによって一度GPLからNPLに戻るという自体が起きていたがすぐに修正された

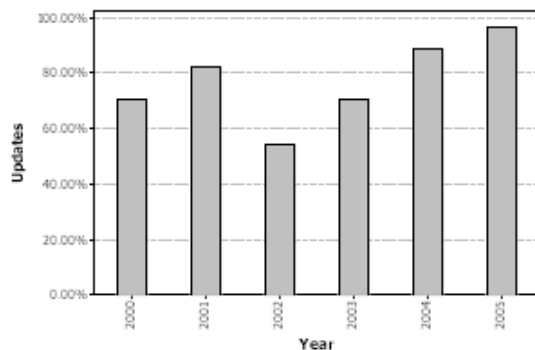
RQ2 : 結果

- Samba
 - 特に変化はなくGPL-v2のもとで利用されていた

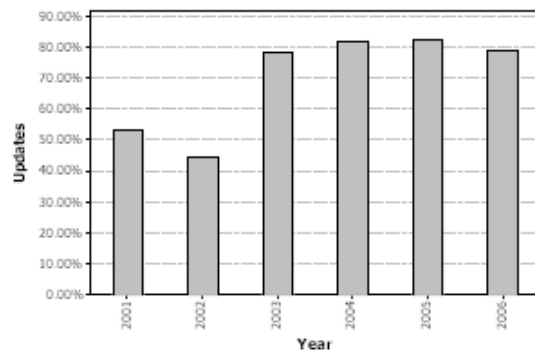
RQ3

- 著作年の変更はライセンス記述にどのような影響を与えているか？

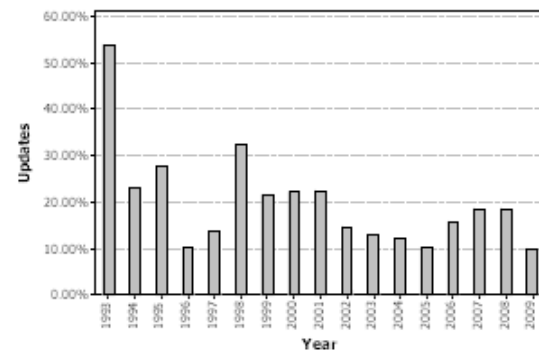
RQ3 : 結果



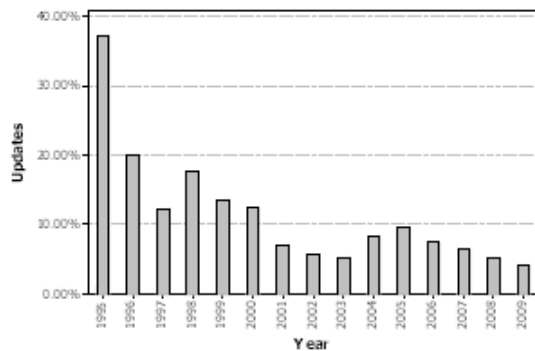
(a) ArgoUML



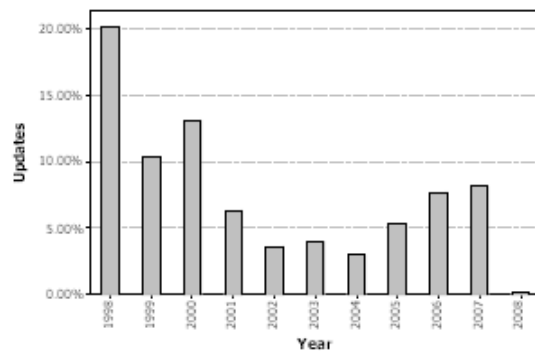
(b) Eclipse-JDT



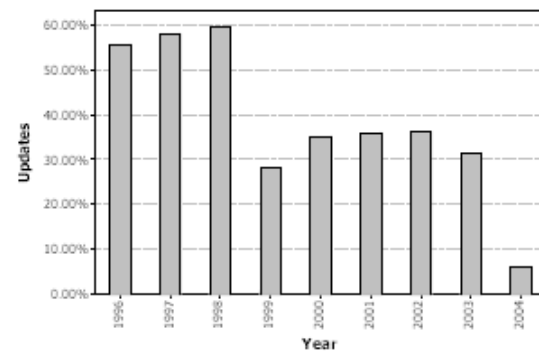
(c) FreeBSD



(d) OpenBSD



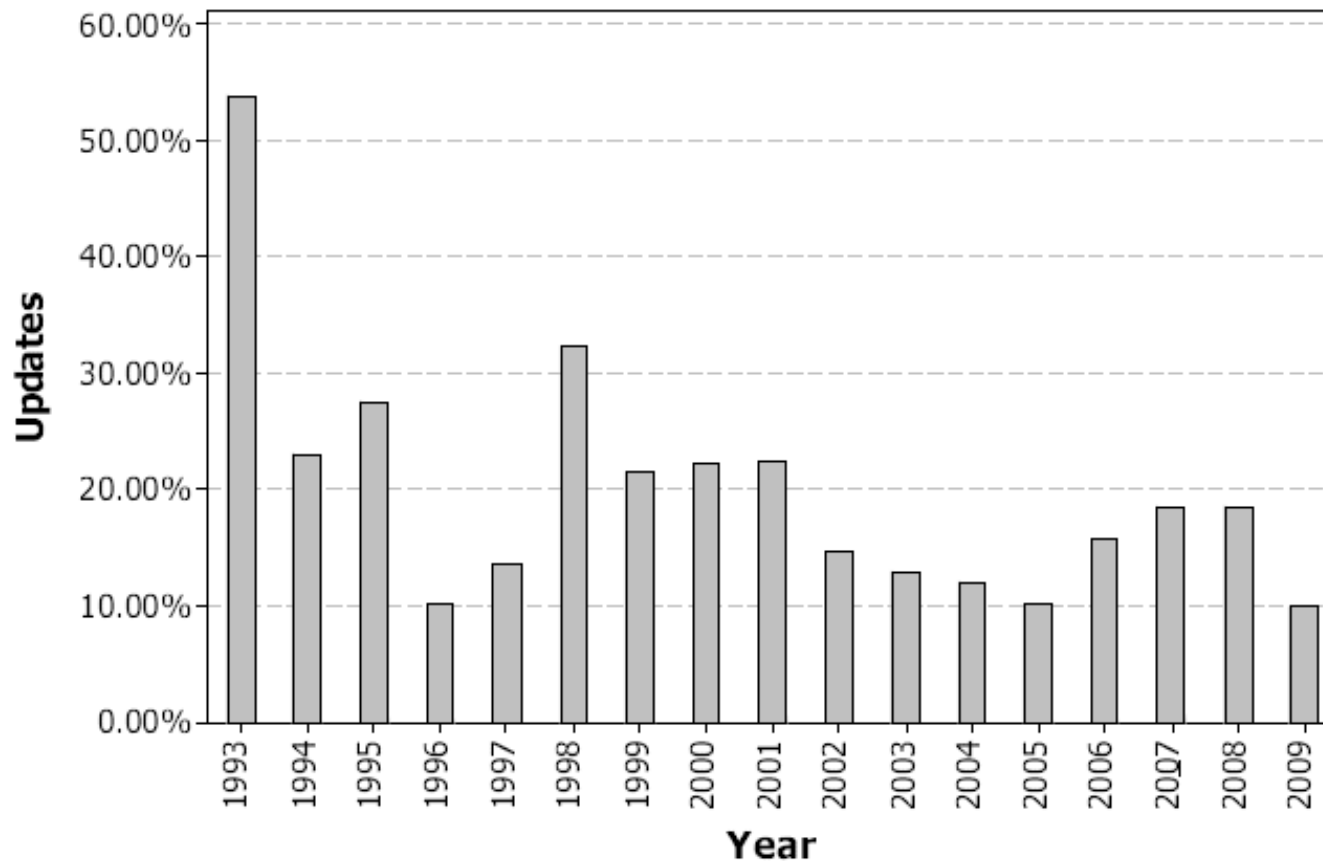
(e) Mozilla



(f) Samba

Figure 2: Percentage of files modified during a year for which the copyright year was updated.

RQ3 : 結果



(c) FreeBSD

RQ3 : 結果

Table 6: Relationship between file changes and copyright year updates.

Systems	Changes to files with outdated copyright year			Changes to files with outdated copyright year			p-values	Eff. sizes
	Mean	Median	σ	Mean	Median	σ		
ArgoUML	3.3	2.0	3.7	4.9	3.0	6.0	< 0.01	0.3
Eclipse-JDT	3.8	2.0	6.1	6.0	3.0	8.4	< 0.01	0.3
FreeBSD	4.3	2.0	6.5	5.4	3.0	7.5	< 0.01	0.1
OpenBSD	2.6	1.0	4.9	4.7	2.0	8.3	< 0.01	0.5
Mozilla	4.7	2.0	10.2	4.7	2.0	10.2	< 0.01	0.0
Samba	5.8	2.0	12.4	12.6	7.0	17.1	< 0.01	0.4

RQ3 : 結果

- 表6の図の意味が正しく読み取れてません
- 表6を説明している文書は論文の8枚目のtable6～の英文です
- Mann-Whitneyの両側検定のp値を用いている

RQ3 : 結果

- ライセンス記述の著作年が更新されたファイルは更新されていないファイルに比べて多くの頻度で変更が行われている

まとめ

- ライセンスアップデートについて6つのFOSSを対象に調査を行った
- ライセンスの変化頻度はFOSSによって異なっている
- ライセンスの著作権年が更新されたファイルは変化が起きやすい