付録 A. ８文書の形態素解析結果から生成したターム行列（一部）

	A	B	C	D	E	F	G	H	I	J
1		1_GCしても残り	2_誤ったAPI情報	3_見よう見まね	4_機器のエラー	5_割り込み処理	6_日付の扱い変	7_日付の扱いか	8_単位によって化ける数値.txt	
2	GC	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
3	MVVM	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
4	No	0.08	0.00	0.00	0.11	0.10	0.00	0.00	0.00	
5	Thread	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
6	いる	0.14	0.13	0.12	0.09	0.05	0.10	0.08	0.05	
7	が	0.09	0.11	0.22	0.08	0.15	0.16	0.12	0.16	
8	から	0.08	0.10	0.00	0.15	0.00	0.00	0.00	0.00	
9	こと	0.06	0.09	0.05	0.00	0.00	0.03	0.00	0.06	
10	する	0.28	0.17	0.16	0.25	0.20	0.18	0.19	0.16	
11	て	0.19	0.14	0.16	0.14	0.05	0.15	0.11	0.11	
12	で	0.07	0.04	0.00	0.00	0.09	0.17	0.14	0.18	
13	と	0.09	0.04	0.06	0.05	0.11	0.07	0.12	0.04	
14	ない	0.09	0.07	0.00	0.00	0.00	0.05	0.05	0.09	
15	に	0.07	0.13	0.15	0.08	0.13	0.09	0.12	0.09	
16	の	0.16	0.15	0.28	0.20	0.21	0.20	0.25	0.20	
17	は	0.12	0.15	0.09	0.08	0.04	0.11	0.10	0.11	
18	プラット	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
19	まで	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
20	メモリ	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
21	も	0.16	0.00	0.00	0.00	0.08	0.00	0.00	0.00	
22	モデル	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
23	リーク	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
24	れる	0.09	0.10	0.09	0.05	0.05	0.03	0.05	0.05	
25	安定	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
26	異常	0.13	0.00	0.00	0.22	0.22	0.00	0.00	0.00	
27	印字	0.10	0.00	0.00	0.14	0.00	0.00	0.00	0.00	
28	画面	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
29	解放	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

付録B. 正規化と重み付けについて

各文書の長さ（＝文章量・単語数）は互いに異なり，その正規化と重み付けの調整を行う必要がある．

（例：10 万語の文章 a にターム A が出現した回数が 3 回． 100 語の文章 b にターム A が 3 回出現した場合，結果は同じターム A の 3 回出現であっても，a，b それぞれに対する重要度は異なる）

それぞれ一般的な正規化(Normalization)，局所的重みを索引語頻度(TF:Term Frequency)，大域的重みを文書頻度逆数（IDF:Inverse Document Frequency）として影響調整を行う．TF(tf)と IDF(idf)二つの指標に基づいて計算される．

$$\mathrm{tfidf}_{ij} = \mathrm{tf}_{ij} \cdot \mathrm{idf}_i \qquad \mathrm{tf}_{ij} = \frac{n_{ij}}{\Sigma_k n_{kj}} \qquad \mathrm{idf}_i = log \frac{|D|}{|\{d:d \ni t_i\}|}$$

$n_{i,j}$ は単語 t_i の文書 d_j における出現回数，$\sum_k n_{k,j}$ は文書 d_j におけるすべての単語の出現回数の和， $|D|$ は総文書数， $|\{d:d \ni t_i\}|$ は単語 t_i を含む文書数である． そのため，idf は一種の一般語フィルタとして働き， 多くの文書に出現する語（一般的な語）は重要度が下がり，特定の文書にしか出現しない単語の重要度を上げる役割を果たす．

TF-IDF 法(TF=Term Frequency=単語の出現頻度）と IDF（Inverse Document Frequency=逆文書頻度）,，“ウィキペディア日本語版,” 13 9 2016. https://ja.wikipedia.org/wiki/Tf-idf.）

付録 C. 用語解説

アブストラクションシート（ABS）:
バグ票同様，欠陥を個票形式で記録したものであり，インシデント情報（発生事象，実害，原因）及び欠陥情報（欠陥モデルの各要素（誘発因子，過失因子，プログラム欠陥，増幅因子，表出現象））の両方が記録されている．

クラスタリング:
異なる性質のものが混ざりあっている集団（対象）の中から互いに似たものを集めて集落（クラスター）を作り，対象を分類する方法の総称．客観的な基準（本稿ではユークリッド距離）に従って科学的に分類ができるため，現在，様々な場面で利用されており，例えばマーケティングリサーチにおいては，ポジショニング確認を目的としたブランドの分類などにも用いられている．（http://www.macromill.com/landing/words/b003.html）

ワードクラウド（Word Cloud）:
文章中で出現頻度が高い単語を複数選び出し，その頻度に応じた大きさで図示する手法．ウェブページやブログなどに頻出する単語を自動的に並べることなどを指す．文字の大きさだけでなく，色，字体，向きに変化をつけることで，文章の内容をひと目で印象づけることができる．（goo辞書）