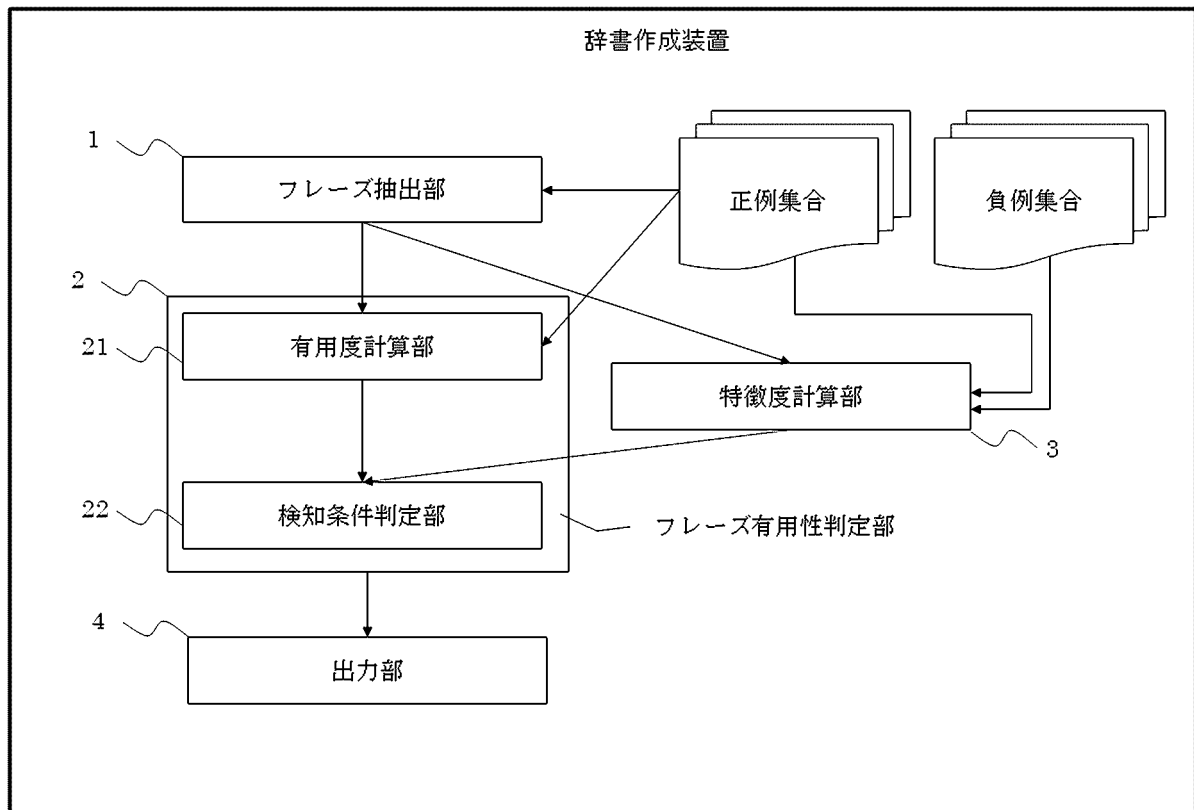
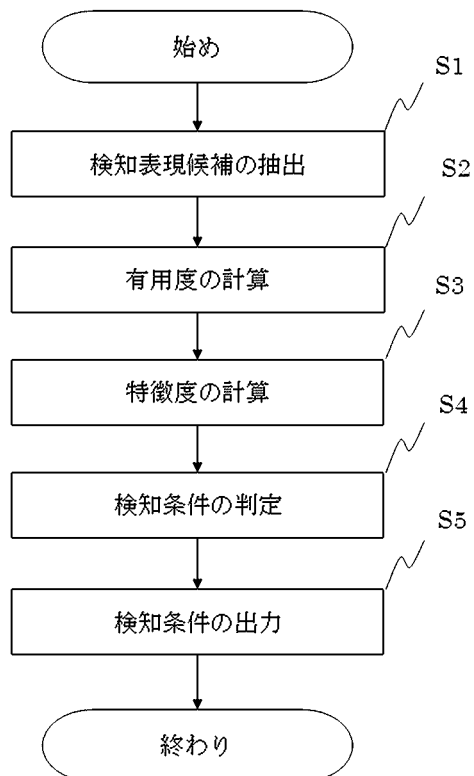


[図1]



[図2]



[図3]

正例集合	負例集合
<ul style="list-style-type: none"> ・メールを開いたらトロイの木馬に感染した。 ・あのメールはトロイの木馬らしいよ。 ・トロイの木馬に感染しないように注意！ ・*****～ 	<ul style="list-style-type: none"> ・メールを書いた。 ・風邪に感染？ ・*****～

[図4]

フレーズ	頻度（正例）	頻度（負例）	特徴度
トロイの木馬	3	0	3
トロイ	3	0	3
木馬	3	0	3
トロイの木馬に感染	2	0	2
木馬に感染	2	0	2
感染	2	1	1
メール	2	1	1
・・・			

※特徴度は、「頻度（正例）－頻度（負例）」で計算

[図5]

フレーズ	頻度(正例)	頻度(負例)	特徴度	長さ	有用度	スコア
トロイの木馬	3	0	3	2	6	18
トロイ	3	0	3	1	3	9
木馬	3	0	3	1	3	9
トロイの木馬に感染	2	0	2	3	6	12
木馬に感染	2	0	2	2	4	8
感染	2	1	1	1	2	2
メール	2	1	1	1	2	2
...						

※長さは、文節数で計算

※有用度＝フレーズ長さ×頻度（正例）

※スコアは、特徴度×有用度で計算 閾値 10以上

[図6]

フレーズ	頻度(正例)	頻度(負例)	特徴度	長さ	有用度	スコア
トロイの木馬	3	0	3	2	4.5	13.5
トロイ	3	0	3	1	1.5	4.5
木馬	3	0	3	1	1.5	4.5
トロイの木馬に感染	2	0	2	3	5	10
木馬に感染	2	0	2	2	3	6
感染	2	1	1	1	1	1
メール	2	1	1	1	1	1
...						

※長さは、文節数で計算

※有用度は、（長さ－0.5）×頻度（正例）で計算

※スコアは、特徴度×有用度で計算 閾値 10以上

[図7]

フレーズ	長さ	頻度(正例)	T	C	C-value	特徴度	スコア
トロイの木馬	2	3	2	1	2	3	6
トロイ	1	3	5	2	0.5	3	1.5
木馬	1	3	7	3	0.67	3	2
トロイの木馬に感染	3	2	0	0	6	2	12
木馬に感染	2	2	2	1	0	2	0
感染	1	2	4	2	0	1	0
メール	1	2	0	0	2	1	2
...							

C-valueの定義

$C\text{-value} = (\text{フレーズ長}) \times (\text{フレーズ頻度} - T/C)$ (C>0の場合)

$C\text{-value} = (\text{フレーズ長}) \times (\text{フレーズ頻度})$ (C=0の場合)

T: 着目するフレーズを包含するフレーズの出現頻度

C: 着目するフレーズを包含するフレーズの異なり数

スコアは、特徴度×有用度で計算 閾値5以上

[図8]

フレーズ	長さ	頻度(正例)	T	C	C-value	特徴度	スコア
トロイの木馬	2	3	2	1	1	3	3
トロイ	1	3	5	2	0	3	0
木馬	1	3	7	3	0	3	0
トロイの木馬に感染	3	2	0	0	4	2	8
木馬に感染	2	2	2	1	0	2	0
感染	1	2	4	2	0	1	0
メール	1	2	0	0	0	1	0
...							

C-valueの定義

$C\text{-value} = (\text{フレーズ長} - 1) \times (\text{フレーズ頻度} - T/C)$ (C>0の場合)

$C\text{-value} = (\text{フレーズ長} - 1) \times (\text{フレーズ頻度})$ (C=0の場合)

T: 着目するフレーズを包含するフレーズの出現頻度

C: 着目するフレーズを包含するフレーズの異なり数

[図9]

フレーズ	頻度 (正例)	長さ	有用度	頻度 (負例)	特徴度	スコア
トロイの木馬	3	2	4.5	0	3	13.5
トロイ	3	1	1.5	0	—	—
木馬	3	1	1.5	0	—	—
トロイの木馬に感染	2	3	5	0	2	10
木馬に感染	2	2	3	0	2	6
感染	2	1	1	1	—	—
メール	2	1	1	1	—	—
...						

※長さは、文節数で計算

※有用度は、 $(長さ - 0.5) \times 頻度(正例)$ で計算

※有用度 3 以上のみ特徴度計算→判定

※スコアは、特徴度 $\times$ 有用度で計算 閾値 10 以上