

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-37420

(P2009-37420A)

(43) 公開日 平成21年2月19日(2009.2.19)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 17/30 (2006.01)	G06F 17/30 340A	5B075
G06F 13/00 (2006.01)	G06F 17/30 180Z	
	G06F 17/30 120B	
	G06F 13/00 540E	

審査請求 未請求 請求項の数 10 O L (全 18 頁)

(21) 出願番号 特願2007-201181 (P2007-201181)
 (22) 出願日 平成19年8月1日(2007.8.1)

(71) 出願人 500257300
 ヤフー株式会社
 東京都港区六本木六丁目10番1号
 (74) 代理人 100106002
 弁理士 正林 真之
 (72) 発明者 増山 毅司
 東京都港区六本木6丁目10番1号 ヤフー株式会社内
 (72) 発明者 佐藤 慎一
 東京都港区六本木6丁目10番1号 ヤフー株式会社内
 (72) 発明者 本野 秀樹
 東京都港区六本木6丁目10番1号 ヤフー株式会社内
 Fターム(参考) 5B075 PR08 QT10

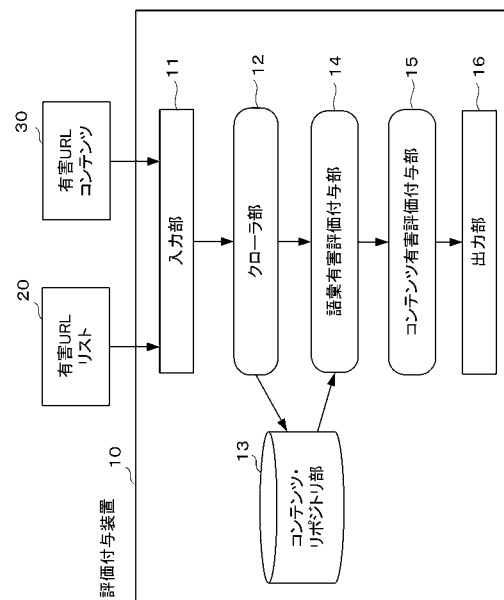
(54) 【発明の名称】 有害コンテンツの評価付与装置、プログラム及び方法

(57) 【要約】

【課題】従来手作業で行っていた、新しいキーワードやウェブサイトの抽出を自動で行い、より高精度に有害サイトのフィルタリングを提供すること。

【解決手段】予め定義された有害サイトのURLリストを入力とし、インターネット上の様々な有害サイトを自動的に探索(クローラ)するクローラ部を備える。そして、探索したURLのコンテンツのデータをコンテンツ・リポジトリ部(データベース)に蓄積、格納する。さらに、コンテンツ・リポジトリ部からコンテンツの文字データに含まれる語彙を抽出し、該語彙の所定のブラック度数と所定のグレー度数を求めて、語彙の有害度を計算する語彙有害評価付与部を備える。さらに、語彙有害評価付与部が計算した語彙の有害度数から前記コンテンツの有害度を計算するコンテンツ有害評価付与部を備えている。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

インターネット上のサイトの有害語彙及び有害コンテンツの評価付与を行う評価付与装置であって、

所定の有害 URL リストを取得し、該有害 URL リストに含まれる URL を随時アクセスして該 URL サイトのコンテンツを収集するクローラ部と、

前記コンテンツのデータを格納するコンテンツ・リポジトリ部と、

前記コンテンツ・リポジトリ部から前記コンテンツの文字データに含まれる語彙を抽出し、該語彙の所定のブラック度数と所定のグレー度数を求めて、語彙の有害度を計算する語彙有害評価付与部と、

前記語彙有害評価付与部が計算した前記語彙の有害度数から前記コンテンツの有害度を計算するコンテンツ有害評価付与部と、

を備えたことを特徴とする、評価付与装置。

10

【請求項 2】

前記所定のブラック度数である $AVG_Black(t)$ は、所定の選別されたブラック・ワードを基に、下記数 1 式によって計算することを特徴とする、請求項 1 に記載の評価付与装置。

【数 1】

$$AVG_Black(t) = \frac{\sum_{i=1}^m R(t, Black_i)}{m}$$

20

t : 単語またはフレーズ (形態素)

$Black_i$: i 番目のブラック・ワード

$R(X, Y)$: 単語またはフレーズ X と単語またはフレーズ Y のシン普森係数

m : ブラック・ワードの数

【請求項 3】

前記所定のグレー度数である $AVG_Gray(t)$ は、所定の選別されたグレー・ワードを基に、下記数 2 式によって計算することを特徴とする、請求項 1 または 2 に記載の評価付与装置。

30

【数 2】

$$AVG_Gray(t) = \frac{\sum_{j=1}^n R(t, Gray_j)}{n}$$

$Gray_j$: j 番目のグレー・ワード

n : グレー・ワードの数

【請求項 4】

前記語彙の有害度数である $SO(t)$ は、下記数 3 式によって計算することを特徴とする、請求項 1 乃至 3 に記載の評価付与装置。

40

【数 3】

$$SO(t) = \begin{cases} \frac{AVG_Black(t)}{\sqrt{\{AVG_Black(t)\}^2 + \{AVG_Gray(t)\}^2}} \cdot \frac{AVG_Black(t)}{AVG_Black(t)} & \text{if } AVG_Black(t) > \alpha, AVG_Gray(t) > \beta \\ 0 & \text{otherwise} \end{cases}$$

、 : 閾値

50

【請求項 5】

前記コンテンツの有害度数 $SO(c)$ は、下記数 4 式によって計算することを特徴とする、請求項 1 乃至 4 に記載の評価付与装置。

【数 4】

$$SO(c) = \gamma \cdot \frac{\sum_{k=1}^l SO(t_k)}{l} \quad \text{if } SO(t_k) > 0$$

c : コンテンツ

γ : 総単語またはフレーズに占める $SO(t) > 0$ の単語またはフレーズの比率

l : コンテンツ内の $SO(t) > 0$ の単語またはフレーズ数

【請求項 6】

前記ブラック・ワードは、複数のブラック・ワード候補の語彙に対して、操作員の目視判定結果と、該ブラック・ワード候補の語彙を含むサイトの URL をフィルタリングするツールの判定結果と、有害語辞書による判定結果を、それぞれ組み合わせることによって選別して、ブラック・ワードリストとして入力されることを特徴とする、請求項 2 に記載の評価付与装置。

【請求項 7】

前記グレー・ワードは、複数のグレー・ワード候補の語彙に対して、操作員のブラックでないという目視判定結果と、該グレー・ワード候補の語彙を含むサイトの URL をフィルタリングするツールのブラックでないという判定結果と、有害語辞書によるブラックでないという判定結果を、それぞれを組み合わせることによって選別して、グレー・ワードリストとして入力されることを特徴とする、請求項 3 に記載の評価付与装置。

【請求項 8】

ウェブページにおける検索結果のリストから有害と判定されたウェブサイトを下位に表示するか、または削除して前記検索結果の画面に表示する、請求項 1 乃至 7 に記載の評価付与装置。

【請求項 9】

インターネット上のサイトの有害語彙及び有害コンテンツの評価付与を行うためのコンピュータ・プログラムであって、

コンピュータに、

所定の有害 URL リストを取得し、該有害 URL リストに含まれる URL を随時アクセスして該 URL サイトのコンテンツを収集するステップと、

前記コンテンツのデータを格納するステップと、

前記格納されたデータから前記コンテンツの文字データに含まれる語彙を抽出し、該語彙の所定のブラック度数と所定のグレー度数を求めて、語彙の有害度を計算するステップと、

前記計算した前記語彙の有害度数から前記コンテンツの有害度を計算するステップと

、
を実行させることを特徴とする、コンピュータ・プログラム。

【請求項 10】

インターネット上のサイトの有害語彙及び有害コンテンツの評価付与を行うための方法であって、

所定の有害 URL リストを取得し、該有害 URL リストに含まれる URL を随時アクセスして該 URL サイトのコンテンツを収集するステップと、

前記コンテンツのデータを格納するステップと、

前記格納されたデータから前記コンテンツの文字データに含まれる語彙を抽出し、該語彙の所定のブラック度数と所定のグレー度数を求めて、語彙の有害度を計算するステップと、

10

20

30

40

50

前記計算した前記語彙の有害度数から前記コンテンツの有害度数を計算するステップと

、
を含むことを特徴とする、方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、インターネット上のサイトから有害語彙の検出と評価を行う評価付与装置、
処理プログラム、及び処理方法に関する。

【背景技術】

【0002】

インターネット上の情報を検索するためにサーチエンジンが利用されている。サーチエンジン
は一般的にウェブページのリンクを解析してページまたはサイトにスコア付けを行い、検索結果
のランキングを表示している。インターネット上にはジャンルを問わず膨大な数のウェブサ
イトが存在し、その中にはアダルトコンテンツや暴力的表現を含むものが多数存在する。
従来、検索エンジンは上記のとおり、リンク解析によって、検索結果を決定しているた
め、入力するキーワードによっては、アダルトコンテンツが検索結果の上位に表示され
ることも多かった。

【0003】

特に、子供がインターネット検索をする際にこれらのサイトが表示されないよう、ペア
レンタル・コントロールと呼ばれるフィルタリング・システムを導入することが増えている
。これらは一般的にキーワードや登録したURLに基づいてサイトをフィルタリングして、
検索結果に表示すべきでないページを決定している。そのため、新しいキーワードや
URLを常に発見し、更新しなければならない。従来、これらの更新作業は人手で行って
いたが、ウェブサイトの爆発的な増加に対応するのが難しく非効率であった。そのため、
例えば、特許文献1、2や非特許文献1のような技術が開示されている。

【0004】

特許文献1に記載の技術では、専門用語を抽出する対象ドキュメントが既にデータベー
スに保存されており、かつ専門用語と関連する分野のドキュメントであることを前提とし
ている。そのため、対象とするドキュメント数が限定されて、高精度で専門用語を抽出
することができた。しかし、対象をウェブサイト全体に広げた場合、ウェブドキュメントは
分野ごとに分類されていないという問題があり、専門用語を抽出する前に、まず対象と
するウェブドキュメントをウェブ上から収集する必要がある。また、ウェブサイトは日々更
新されるという特徴があり、さらに企業や官公庁だけでなく、個人の趣味・嗜好の基に作
成されるものも多く存在するため、学术论文などにくらべてノイズとなる情報がドキュ
メント中に多く含まれている可能性が高く、上記の技術とは別の視点が必要となる。

【0005】

特許文献2に記載の技術では、カテゴリごとに分類されたコンテンツ・データベース
を有し、検索要求に基づいて検索した結果のウェブページをコンテンツ・データベースと
比較し、コンテンツ・データベースに含まれるウェブサイトを検索結果から削除する方
法について記載されている。この技術では、カテゴリごとに分類されたコンテンツ・デー
タベースを作成する必要があるが、これは従来技術で述べたとおり、作成に非常に手間
がかかるという問題があるが、その問題の解決手段については開示されていない。

【0006】

非特許文献1には、自動車や映画のレビューを解析し、“excellent”と共起
する語が多い場合はポジティブなレビュー、“poor”と共起するフレーズが多い場合は
ネガティブなレビューが多いことを発見したことが開示されている。この方法では、相
互情報量を利用してレビューのP/N（ポジティブ/ネガティブ）を判定している。しか
し、本技術では、ウェブページの記載内容がポジティブなものかネガティブなものかを
判別することを目的としており、この技術をコンテンツの有害度の計算にあてはめよう
とすると、ブラック（有害）かホワイト（無害）かの判定になり、グレー（中間）の判定する

10

20

30

40

50

ことはできない。

【0007】

【特許文献1】特開2005-222263号公報

【特許文献2】特開2007-128119号公報

【非特許文献1】Peter D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of Association for Computational Linguistics (ACL'02), Philadelphia, Pennsylvania, USA, July 8-10, 2002, pp417-424, NRC 44946.

10

【発明の開示】

【発明が解決しようとする課題】

【0008】

本発明では、従来手作業で行っていた、新しいキーワードやウェブサイトの抽出を自動で行い、より高精度に有害サイトのフィルタリングを行うことを課題とする。また、それをペアレンタル・コントロールや検索結果のランク付けに応用することで、意図しないユーザにアダルトや暴力表現を含んだウェブページを表示しないことを課題とする。

【0009】

20

本発明の背景としては、検索結果のランキングにおいて、有害度の高いページの順位を下げて、そうでないページを積極的に上位に露出させたいという要求がある。しかしながら、一般的な検索エンジンは、リンク解析のみでコンテンツまで見ていないので、有害度が高いページも上位にランクされてしまうという問題点があった。ここで、有害度とは、子供に悪影響を及ぼす語彙を含んでいる度合いの意であり、例えば、アダルトサイト 有害度が高い、官公庁のトップページ 有害度が低い、といったものである。したがって、本発明の目的は、コンテンツ(ページ)に有害度を付与することである。ただし、本明細書では、「有害」または「有害度」の用語は、猥褻的、暴力的、中傷的な表現のみならず、ウェブページに掲載するには不適切な表現もしくはそれを測る度数を含むものとする。

【課題を解決するための手段】

30

【0010】

本発明では、上記課題に鑑み、以下のような解決手段を提供する。

【0011】

(1) インターネット上のサイトの有害語彙及び有害コンテンツの評価付与を行う評価付与装置であって、

所定の有害URLリストを取得し、該有害URLリストに含まれるURLを随時アクセスして該URLサイトのコンテンツを収集するクロウラ部と、

前記コンテンツのデータを格納するコンテンツ・リポジトリ部と、

前記コンテンツ・リポジトリ部から前記コンテンツの文字データに含まれる語彙を抽出し、該語彙の所定のブラック度数と所定のグレー度数を求めて、語彙の有害度を計算する語彙有害評価付与部と、

40

前記語彙有害評価付与部が計算した前記語彙の有害度数から前記コンテンツの有害度を計算するコンテンツ有害評価付与部と、

を備えたことを特徴とする、評価付与装置。

【0012】

本評価付与装置は、予め定義された有害サイトのURLリストを入力とし、インターネット上の様々な有害サイトを自動的に探索(クローラ)するクロウラ部を備える。そして、探索したURLのコンテンツのデータをコンテンツ・リポジトリ部(データベース)に蓄積、格納する。さらに、コンテンツ・リポジトリ部からコンテンツの文字データに含まれる語彙を抽出し、該語彙の所定のブラック度数と所定のグレー度数を求めて、語彙の有害

50

害度数を計算する語彙有害評価付与部を備える。さらに、語彙有害評価付与部が計算した語彙の有害度数から前記コンテンツの有害度数を計算するコンテンツ有害評価付与部を備えている。なお、評価付与（レイティング）とは、対象となる物事に対して、ある基準を適用して、等級分けをしたり、度合いを測ることをいう。

【0013】

このような構成を備えることによって、インターネット上の膨大なウェブサイトに含まれるデータから、その語彙（単語、またはフレーズ）を自動的に集めて、その語彙ごとに所定の「ブラック度数」と所定の「グレー度数」を計算して、その値を用いて語彙の有害度数（語彙有害評価付与度）を定量的に計算することができる。さらに、その語彙を含んだコンテンツの有害度（コンテンツ有害評価付与度）も定量的に計算することができる。

10

【0014】

この評価付与装置を作動させることによって、これまで目視で確認して分類していた有害用語や有害ウェブサイトを自動で収集することができる。さらに、上記で収集した用語やウェブサイト情報に基づいて、ペアレンタル・コントロールを行うことで、高精度にフィルタリングを行うことができる。

【0015】

また、本評価付与装置は、語彙有害評価付与部に下記の（2）～（5）に記載の計算式を用いて、計算する演算手段を構成として備えることで、ブラック度数やグレー度数から語彙 t の有害度数である $SO(t)$ （Semantic Orientationの略）を定量的に求めることができる。もちろん、計算式はこれに限定されるものでなく、他の公知の計算、例えば、非特許文献1に記載の計算方法を応用することも可能である。すなわち、この演算手段部分は他の計算式を用いたものと交換可能である。

20

【0016】

（2）前記所定のブラック度数である $AVG_Black(t)$ は、下記数1式によって計算することを特徴とする、（1）に記載の評価付与装置。

【数1】

$$AVG_Black(t) = \frac{\sum_{i=1}^m R(t, Black_i)}{m}$$

30

t : 単語またはフレーズ（形態素）

$Black_i$: i 番目のブラック・ワード

$R(X, Y)$: 単語またはフレーズ X と単語またはフレーズ Y のシン普森係数

m : ブラック・ワードの数

【0017】

（3）前記所定のグレー度数である $AVG_Gray(t)$ は、下記数2式によって計算することを特徴とする、（1）または（2）に記載の評価付与装置。

【数2】

$$AVG_Gray(t) = \frac{\sum_{j=1}^n R(t, Gray_j)}{n}$$

40

$Gray_j$: j 番目のグレー・ワード

n : グレー・ワードの数

【0018】

（4）前記語彙の有害度数である $SO(t)$ は、下記数3式によって計算することを特徴とする、（1）乃至（3）に記載の評価付与装置。

【数 3】

$$SO(t) = \begin{cases} \frac{AVG_Black(t)}{\sqrt{\{AVG_Black(t)^2\} + \{AVG_Gray(t)^2\}}} \cdot \frac{AVG_Black(t)}{\sqrt{\{AVG_Black(t)^2\} + \{AVG_Gray(t)^2\}}} & \text{if } AVG_Black(t) > \alpha, AVG_Gray(t) > \beta \\ 0 & \text{otherwise} \end{cases}$$

、 : 閾値

【0019】

10

さらに、本装置は、下記(5)に記載の計算式を用いた演算手段を備えることによって、コンテンツの有害度数 $SO(c)$ を定量的に求めることができる。この演算手段部分は他の計算式を用いたものと交換可能とする。

【0020】

(5) 前記コンテンツの有害度数 $SO(c)$ は、下記数4式によって計算することを特徴とする、(1)乃至(4)に記載の評価付与装置。

【数 4】

$$SO(c) = \gamma \cdot \frac{\sum_{k=1}^l SO(t_k)}{l} \quad \text{if } SO(t_k) > 0$$

20

c : コンテンツ

: 総単語またはフレーズに占める $SO(t) > 0$ の単語またはフレーズの比率

l : コンテンツ内の $SO(t) > 0$ の単語またはフレーズ数

【0021】

(6) 前記ブラック・ワードは、複数のブラック・ワード候補の語彙に対して、操作員の目視判定結果と、該ブラック・ワード候補の語彙を含むサイトのURLをフィルタリングするツールの判定結果と、有害語辞書による判定結果を、それぞれ組み合わせることによって選別して、ブラック・ワードリストとして入力されることを特徴とする、(2)に記載の評価付与装置。

30

【0022】

このような構成によれば、ブラック・ワード(単語だけでなくフレーズも含む)は、人間の目視判定結果、すなわち、例えば、5人中4人以上がブラックであると判定した結果を判定データとして取り込み、さらに、既存のURLフィルタリング・ツールを用いた判定データ、及び、既存の有害辞書を用いた判定データを取り込み、それらの判定データを組み合わせて、あるいは順次判定処理を行わせることによって絞り込みブラック・ワードの選別を行う。このように選別されたブラック・ワードに対して、本評価付与装置の有害度の評価付与を行うことで、さらに精度の向上が期待できる。

【0023】

40

以上の説明は、解決手段を評価付与装置として実現する態様で行ったが、下記(7)または(8)の態様のように、本発明は、上記の機能をコンピュータに実行させるコンピュータ・プログラム(ソフトウェア)、または、その処理方法として捉えることができる。

【0024】

(7) 前記グレー・ワードは、複数のグレー・ワード候補の語彙に対して、操作員のブラックでないという目視判定結果と、該グレー・ワード候補の語彙を含むサイトのURLをフィルタリングするツールのブラックでないという判定結果と、有害語辞書によるブラックでないという判定結果をそれぞれ組み合わせることによって選別して、グレー・ワードリストとして入力されることを特徴とする、(3)に記載の評価付与装置。

【0025】

50

このような構成によれば、(6)のブラック・ワードの選別と同様に、グレー・ワード(単語だけでなくフレーズも含む)についても同様な処理を行う。例えば、5人中1人以上がブラック・ワードと判定した語をグレー・ワード候補とし、次に、既存のURLフィルタリング・ツールや有害辞書によるフィルタリングがブラックでないとして判定した結果を加えていき、グレー・ワードの選別を絞り込んでいくことができる。このように選別されたグレー・ワードに対して、本評価付与装置の有害評価付与を行うことで、さらに精度の向上が期待できる。

【0026】

(8) ウェブページにおける検索結果のリストから有害と判定されたウェブサイトを下位に表示するか、または削除して前記検索結果の画面に表示する、(1)乃至(7)に記載の評価付与装置。

10

【0027】

以上の(1)~(8)の説明は、本発明の解決手段を、評価付与装置として実現する態様であるが、下記(9)または(10)の態様のように、本発明は、上記の機能をコンピュータに実行させるコンピュータ・プログラム(ソフトウェア)、または、その処理方法として捉えることができる。このような態様によって、(1)と同様な作用効果を得ることができる。

【0028】

(9) インターネット上のサイトの有害語彙及び有害コンテンツの評価付与を行うためのコンピュータ・プログラムであって、

20

コンピュータに、

所定の有害URLリストを取得し、該有害URLリストに含まれるURLを随時アクセスして該URLサイトのコンテンツを収集するステップと、

前記コンテンツのデータを格納するステップと、

前記格納されたデータから前記コンテンツの文字データに含まれる語彙を抽出し、該語彙の所定のブラック度数と所定のグレー度数を求めて、語彙の有害度を計算するステップと、

前記計算した前記語彙の有害度数から前記コンテンツの有害度を計算するステップと

、

を実行させることを特徴とする、コンピュータ・プログラム。

30

【0029】

(10) インターネット上のサイトの有害語彙及び有害コンテンツの評価付与を行うための方法であって、

所定の有害URLリストを取得し、該有害URLリストに含まれるURLを随時アクセスして該URLサイトのコンテンツを収集するステップと、

前記コンテンツのデータを格納するステップと、

前記格納されたデータから前記コンテンツの文字データに含まれる語彙を抽出し、該語彙の所定のブラック度数と所定のグレー度数を求めて、語彙の有害度を計算するステップと、

前記計算した前記語彙の有害度数から前記コンテンツの有害度を計算するステップと

、

を含むことを特徴とする、方法。

40

【発明の効果】

【0030】

本発明によれば、今まで目視で確認して分類していた有害用語や有害ウェブサイトを手動で収集することができる。さらに、上記で収集した用語やウェブサイト情報に基づいて、ペアレンタル・コントロールを行うことで、漏れなくフィルタリングを行うことができる。

【発明を実施するための最良の形態】

【0031】

50

以下、本発明の実施形態について図を参照しながら説明する。

【0032】

[評価付与装置の構成、処理フロー]

図1は、評価付与装置10の機能ブロックの概略を示す図である。本装置は、入力部11、クローラ部12、コンテンツ・リポジトリ部13、語彙有害評価付与部14、コンテンツ有害評価付与部15、出力部16を含んで構成される。

【0033】

有害URLリスト20は、それを予め所有するサイトやデータベースからのデータを、入力部11を介して本装置に入力されるものとする。クローラ部12は、この有害URLリスト20に基づいて、リストに含まれるURLの各サイトに順次アクセスし、そのサイトの内容(コンテンツ)を自動的に収集する。収集されたコンテンツのデータは、コンテンツ・リポジトリ部13に格納される。コンテンツ・リポジトリ部13は、一般にはデータベースであり、URL名、サイトアドレス、収集日時などに関連付けてコンテンツの内容が記録される。なお、クローラ方法自体は、公知の手法を用いてよい。

10

【0034】

語彙有害評価付与部14は、コンテンツ・リポジトリ部13からコンテンツの文字データを取り出し、形態素解析を行って、単語またはフレーズ(語彙)に分解する。ここで、形態素解析は公知の様々な方法を利用してよい。そして、得られた語彙それぞれに対して、有害評価付与処理を行って、語彙の有害評価付与(以後、これをSO(t)で表す)計算を行う。語彙の有害評価付与の方式の詳細は後述する。

20

【0035】

コンテンツ有害評価付与部15は、語彙有害評価付与部14によって、計算された語彙の有害評価付与を基に、コンテンツの有害評価付与(以後、これをSO(c)で表す)計算を行う。語彙の有害評価付与の詳細は後述する。計算された語彙の有害評価付与は出力部16を介して外部に出力され、URLの有害度数として、サイトのフィルタリングやURL検索結果の表示などに利用される。

【0036】

図2は、上記の装置が行うコンテンツの有害評価付与の処理フローをまとめた図である。すなわち、評価付与装置10は以下の処理を行う。

- (1) 外部サイトなどから有害URLリストをクローラ部が取り込む。
- (2) クローラ部はインターネット上にURLへのリクエストを発行する。
- (3) (有害)コンテンツを取得する。
- (4) (有害)コンテンツをコンテンツ・リポジトリに格納する。
- (5) 語彙の有害評価付与処理を行う。
- (6) コンテンツの有害評価付与処理を行う。

30

【0037】

[評価付与方式の詳細]

図3は、コンテンツの有害評価付与のイメージを示す図である。まず、クローラ部12によって収集された有害URLコンテンツ30は、形態素解析(品詞フィルタリング)処理を行って、形態素解析済み語彙リスト40を生成する。次に、語彙の有害評価付与計算では、予め定義されたブラック・ワードリスト50、グレー・ワードリスト60を用いる。ここで、グレー・ワードリストとは、ブラックかどうかの判定が難しい語を意味する。

40

【0038】

<有害評価付与の仮定>

本方式では以下の2つの仮定を前提としている。

(仮定1) ブラック・ワードが出現するコンテンツでは、異なるブラック・ワードも出現しやすい。

例えば、URL、http://www. . . com/dera/ . . .htmlのコンテンツは、以下のような内容を含んでいる。

【0039】

50

“ 「愛しのフ ラチオ」シリーズ第5弾！ 今回ももちろんデジタルモザイクだから、見たい部分がはっきり、くっきり、すっきりですよ！ フ ラチオ大好きな17人のギャルが自慢のフ テクでザー ン味くらべ。 ”

【0040】

これを、ある検索エンジンによる共起ヒット数を調べると下記のようになった。

「フ ラチオ」と「フ ラテク」の共起ヒット数 51,700件

「フ ラチオ」と「ザー メン」の共起ヒット数 513,000件

「フ ラテク」と「ザー メン」の共起ヒット数 54,300件

【0041】

(仮定2) グレー・ワードが出現するコンテンツでは、異なるグレー・ワードも出現しやすい。

例えば、URL、http://blogs. .co.jp/ .htmlのコンテンツは以下のような内容を含んでいる。

【0042】

“ . . .これR - 指定かかるシーンまであるwww、まあ原作がなにぶんエロゲーなのではないといえましょうがないのですが、じゃあ18禁とかちゃんと書いてくださいよ、いやもしそれでも買いますけども. . . ”

【0043】

これを、ある検索エンジンによる共起ヒット数を調べると、「エロゲー」と「18禁」の共起ヒット数 858,000件となった。したがって、アプローチとしては、コーパスから共起情報を獲得し、ブラック・ワードとグレー・ワードのどちらと共起しやすいかをモデル化する方式をとった。

【0044】

< 語彙の有害評価付与度 >

具体的には、語彙有害評価付与部14の演算手段は、下記数5~7式によって、語彙の有害評価付与度を求める。

【0045】

【数5】

$$AVG_Black(t) = \frac{\sum_{i=1}^m R(t, Black_i)}{m}$$

30

【0046】

【数6】

$$AVG_Gray(t) = \frac{\sum_{j=1}^n R(t, Gray_j)}{n}$$

【0047】

【数7】

40

$$SO(t) = \begin{cases} AVG_Black(t) \cdot \frac{AVG_Black(t)}{\sqrt{\{AVG_Black(t)^2\} + \{AVG_Gray(t)^2\}}} & \text{if } AVG_Black(t) > \alpha, AVG_Gray(t) > \beta \\ 0 & \text{otherwise} \end{cases}$$

【0048】

ここで、

50

t : 単語またはフレーズ (形態素)

、 : 閾値

Black_i : i 番目のブラック・ワード

R(X, Y) : 数 8 式で表される単語またはフレーズ X と単語またはフレーズ Y のシンブソン係数 (1 に近いほど関連性が高い)。

SO(t) : t の有害度 (0 ~ 1 の範囲)。1 に近いほどブラック度が高い。

【0049】

【数 8】

$$R(X, Y) = \begin{cases} \frac{|X \& Y|}{\min(|X|, |Y|)} & \text{if } |X| > 0, |Y| > 0 \\ 0 & \text{otherwise} \end{cases}$$

10

【0050】

図 4 は、語彙の有害評価付与方式のイメージを示す図である。

【0051】

ここで、SO(t) の特徴としては、

- ・ Black 軸に寄っているほどスコアが高い。(B による影響)
- ・ AVG_Black(t) が大きいほどスコアが高い。(A による影響)

が挙げられる。

20

【0052】

また、図中の B による影響としては、

- ・ Black 軸に寄っているほど、SO(t) は大きくなる。

例えば、(AVG_Black(t), AVG_Gray(t)) = (0.9, 0.3) の場合は、SO(t) は 0.8538 となり、同様に、(0.3, 0.9) の場合は、0.0948 となる。

- ・ AVG_Black(t) が大きいほど、SO(t) は大きくなる。

例えば、(0.9, 0.3) 0.8538、(0.5, 0.3) 0.4287 となる。

- ・ AVG_Gray(t) が小さいほど、SO(t) は大きくなる。

例えば、(0.9, 0.3) 0.8538、(0.9, 0.1) 0.8944 となる。

30

【0053】

また、図中の A による影響としては、

- ・ 角度 が同じ場合、AVG_Black(t) が大きいほど、SO(t) は大きくなる。

例えば、(1.0, 1.0) 0.7071、(0.5, 0.5) 0.3535 となる。

【0054】

<ブラック・ワードとグレー・ワード用のデータ収集方法>

40

データとしては、一般有害用語辞書 (例えば、アダルト、グラビア、出会い系) や有害検索キーワードを母集団とする。語彙の選別方法としては、

- ・ 形態素解析
- ・ 日本語の単語またはフレーズ (Alphabet 以外) の抽出
- ・ 「バスト」、「ソープ」など、ダイエットやファッションの用途で使われている場合

がほとんどである語の除去

以上の選別によって、今回収集できた語彙数は 441 語であった。

【0055】

<ブラック・ワードの選別方法>

まず、目視チェックによって、5人中4人以上がブラックと判定し、次に、URLフィ

50

ルタリング判定（対象語のWeb検索結果URLに対するチェック）やワードフィルタリング判定などのツールによるブラック判定を行った。その結果、今回収集できた語彙数は22語となった。このようにして、生成されたブラック・ワードリストは評価付与装置10の語彙有害評価付与部14の入力として用いられる。なお、上記の判定の順序は入れ替えて行ってもよい。

【0056】

ここで、URLフィルタリング判定（対象語のウェブ検索結果URLに対するチェック）は、次のステップにより実行される。

（1）有害URLを収集する。

（2）対象語のウェブ検索結果のURLをチェックし、（1）で収集したURLが含まれているかチェックする。

（3）（2）でURLが含まれていればブラックとする。

【0057】

また、ワードフィルタリング判定は、次のステップにより実行される。

（1）有害ワードを収集する。

（2）対象語が（1）に含まれているかをチェックする。

（3）（2）で含まれていればブラックとする。

【0058】

<グレー・ワードの選別方法>

まず、目視チェックとして、5人中1人以上がブラックと判定し、次に、URLフィルタリング判定（対象語のWeb検索結果URLに対するチェック）やワードフィルタリング判定などのツールによるブラックでない判定を行った。その結果、今回収集できた語彙数は293語となった。このようにして、生成されたグレー・ワードリストは評価付与装置10の語彙有害評価付与部14の入力として用いられる。URLフィルタリング判定、ワードフィルタリング判定については、前述のステップと同様である。なお、上記の判定の順序は入れ替えて行ってもよい。

【0059】

<コンテンツの有害評価付与方式>

具体的には、コンテンツ有害評価付与部15の演算手段は、コンテンツcの有害評価付与度SO(c)を下記の数9式で求める。

【0060】

【数9】

$$SO(c) = \gamma \cdot \frac{\sum_{k=1}^l SO(t_k)}{l} \quad \text{if } SO(t_k) > 0$$

t：単語またはフレーズ（形態素）

c：コンテンツ

γ：総単語またはフレーズに占めるSO(t) > 0の単語またはフレーズの比率、または0と1の任意の値で調整可能

l：コンテンツ内のSO(t) > 0の単語またはフレーズ数

SO(t)：tの有害度（0～1の範囲）。1に近いほどブラック性が高い。

SO(c)：cの有害度（0～1の範囲）。1に近いほどブラック性が高い。

【0061】

[具体的計算例]

図5、図6、図7は、実際のURLサンプルの評価付与計算例を示した図である。なお、以降の記載では、固有名詞、及び記載するのに差しさわりのある用語は、一部を伏せ字に置き換えている。

【0062】

10

20

30

40

50

ここでは、例として、URL、http://www. .com/dera/
.htmlのコンテンツとして以下の記述を調べた。

【0063】

“「愛しのフ　チーオ」シリーズ第5弾！　今回ももちろんデジタルモザイクだから、見たい部分のはっきり、くっきり、すっきりですよ！　フ　チーオ大好きな17人のギャルが自慢のフ　テクでザー　ン味くらべ。”

【0064】

図8、図9は、別のURLサンプルの評価付与計算例を示す図である。ここでは、URL、http://.co.jp/ /XXXXXXXX.htmlのコンテンツとして以下の記述を調べた。

10

【0065】

“...これR-指定かかるシーンまであるwww　まあ原作がなにぶんエロゲーなのではないといえましょうがないのですが　じゃあ18禁とかちゃんと書いてくださいよ、いやもしそれでも買いますけども...”

【0066】

以下に、実際のサンプル適用例を2つ示す。

【0067】

[サンプル適用例1(「ランジェリー」でブログ検索した場合)]

図10は、サンプル適用例1として、「ランジェリー」で通常のプログ検索した場合の検索結果の画面を示す図である。この検索結果で図中のかっこ内の数値が求められたコンテンツの有害度であり、コンテンツのタイトル欄の横に表示されている。

20

【0068】

[サンプル適用例2(「ランジェリー」でブログ検索した場合+有害フィルタ)]

図11は、サンプル適用例2として、「ランジェリー」で通常のプログ検索した結果に、更に有害フィルタをかけた場合の検索結果の画面を示す図である。ここで、有害フィルタの閾値、として0.02を用いているが、この値は、管理者または利用者が変更可能とする。この検索結果では、図中のかっこ内の数値が求められたコンテンツの有害度であるが、閾値以上の有害度を示すコンテンツを持つページは、検索結果から削除されて表示されている。このように、有害度を単に検索結果に表示するだけでなく、所定の閾値を超えたページを自動的に表示させないことで、子供などの利用者にも安心して検索エンジンを使用させることができる。

30

【0069】

[応用例]

なお、本発明は、以下のような応用(一部)も可能である。

【0070】

(1) 語彙の有害度を知りたい場合

・有害語彙の検知を行う。

文書分類やクラスタリングなどで、どの語が原因で有害と判定されたかの分析を行う。また、有害用語であるかどうか判断が難しい語彙の裏付けを行う。

【0071】

(2) コンテンツ(ページ)の有害度を知りたい場合

・検索問題への適用

有害度の高いページの検索ランキングを低くしたり、排除したりする。また、有害度の低いページを積極的に上位に露出させる。あるいは、リンク解析技術と組み合わせて検索結果のランキングを行う。

40

・文書分類・クラスタリング問題への適用

有害か、そうでないかというような2値分類とは異なる分類を行う。または、有害ページの検知を行う。

【0072】

[評価付与装置10のハードウェア構成]

50

図12は、本発明の好適な実施形態の一例に係る評価付与装置10のハードウェア構成を示す図である。

【0073】

評価付与装置10は、制御部130を構成するCPU131（マルチプロセッサ構成ではCPU132など複数のCPUが追加されてもよい）、バスライン105、通信I/F（I/F：インターフェイス）120、メインメモリ170、BIOS（Basic Input Output System）180、USBポート190、I/Oコントローラ160、キーボード及びマウスなどの入力装置110、並びに表示装置140を備える。

【0074】

I/Oコントローラ160には、テープドライブ151、ハードディスク153、光ディスクドライブ152、及び半導体メモリ154などの記憶装置150を接続することができる。

【0075】

BIOS180は、評価付与装置10の起動時にCPU131が実行するブートプログラムや、評価付与装置10のハードウェアに依存するプログラムなどを格納する。

【0076】

ハードディスク153は、評価付与装置10として機能するための各種プログラム及び本発明の機能を実行するプログラムを記憶する。

【0077】

光ディスクドライブ152としては、例えば、DVD-ROMドライブ、CD-ROMドライブ、DVD-RAMドライブ、CD-RAMドライブを使用することができる。この場合は各ドライブに対応した光ディスク1521を使用する。光ディスク1521から光ディスクドライブ152によりプログラムまたはデータを読み取り、I/Oコントローラ160を介してメインメモリ170またはハードディスク153に提供することもできる。また、同様にテープドライブ151に対応したテープメディア1511を主としてバックアップのために使用することもできる。

【0078】

評価付与装置10に提供されるプログラムは、ハードディスク153、光ディスク1521、またはメモリカードなどの記録媒体に格納されて提供される。このプログラムは、I/Oコントローラ160を介して、記録媒体から読み出され、または通信I/F120を介してダウンロードされることによって、評価付与装置10にインストールされ実行されてもよい。

【0079】

上述のプログラムは、内部または外部の記憶媒体に格納されてもよい。ここで、記憶媒体としては、ハードディスク153、光ディスク1521、またはメモリカードの他に、MDなどの光磁気記録媒体、テープメディア1511を用いることができる。また、専用通信回線やインターネットなどの通信回線に接続されたサーバシステムに設けたハードディスク153または光ディスクライブラリなどの記憶装置を記録媒体として使用し、通信ネットワークを介してプログラムを評価付与装置10に提供してもよい。

【0080】

ここで、表示装置140は、ユーザによるデータの入力を受け付ける画面を表示したり、評価付与装置10による演算処理結果の画面を表示したりするものであり、ブラウン管表示装置（CRT）、液晶表示装置（LCD）などのディスプレイ装置を含む。

【0081】

ここで、入力装置110は、操作員による入力の受け付けを行うものであり、キーボード及びマウスなどにより構成してよい。

【0082】

また、通信I/F120は、評価付与装置10を専用ネットワークまたは公共ネットワークを介して端末と接続できるようにするためのネットワーク・アダプタである。通信I

10

20

30

40

50

／ F 1 2 0 は、モデム、ケーブル・モデム及びイーサネット（登録商標）・アダプタを含んでよい。

【 0 0 8 3 】

以上の例は、評価付与装置 1 0 のハードウェア構成について主に説明したが、一般的なコンピュータに、プログラムをインストールして、そのコンピュータを、評価付与装置 1 0 として動作させ、上記で説明した機能を実現することもできる。

【 0 0 8 4 】

以上、本発明の実施形態について説明したが、本発明は上述した実施形態に限るものではない。また、本発明の実施形態に記載された効果は、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、本発明の実施例に記載されたものに限定されるものではない。

10

【 図面の簡単な説明 】

【 0 0 8 5 】

【 図 1 】 評価付与装置 1 0 の機能ブロックの概略を示す図である。

【 図 2 】 コンテンツの有害評価付与の処理フローを示す図である。

【 図 3 】 コンテンツの有害評価付与のイメージを示す図である。

【 図 4 】 語彙の有害評価付与方式のイメージを示す図である。

【 図 5 】 実際の URL サンプルの評価付与計算例（その 1）を示す図である。

【 図 6 】 実際の URL サンプルの評価付与計算例（その 2）を示す図である。

【 図 7 】 実際の URL サンプルの評価付与計算例（その 3）を示す図である。

20

【 図 8 】 別の URL サンプルの評価付与計算例（その 1）を示す図である。

【 図 9 】 別の URL サンプルの評価付与計算例（その 2）を示す図である。

【 図 1 0 】 サンプル適用例 1（「ランジェリー」でブログ検索した場合の検索結果を示す図である。

【 図 1 1 】 サンプル適用例 2（「ランジェリー」でブログ検索した場合 + 有害フィルタ（閾値 0 . 0 2）の検索結果を示す図である。

【 図 1 2 】 評価付与装置 1 0 のハードウェア構成を示す図である。

【 符号の説明 】

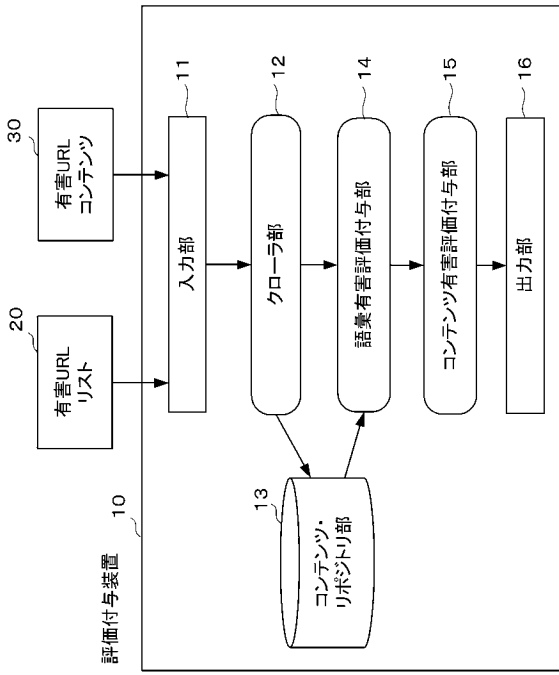
【 0 0 8 6 】

- 1 0 評価付与装置
- 1 1 入力部
- 1 2 クローラ部
- 1 3 コンテンツ・リポジトリ部
- 1 4 語彙有害評価付与部
- 1 5 コンテンツ有害評価付与部
- 1 6 出力部
- 2 0 有害 URL リスト
- 3 0 有害 URL コンテンツ
- 4 0 形態素解析済み語彙リスト
- 5 0 ブラック・ワードリスト
- 6 0 グレー・ワードリスト

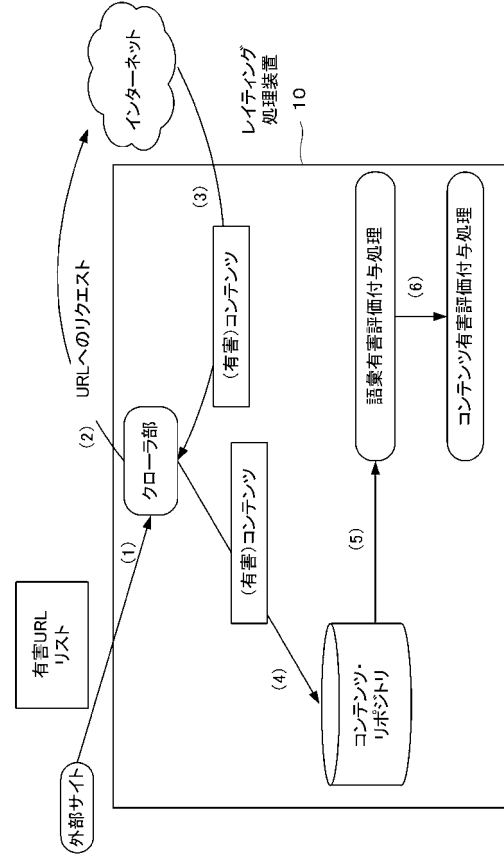
30

40

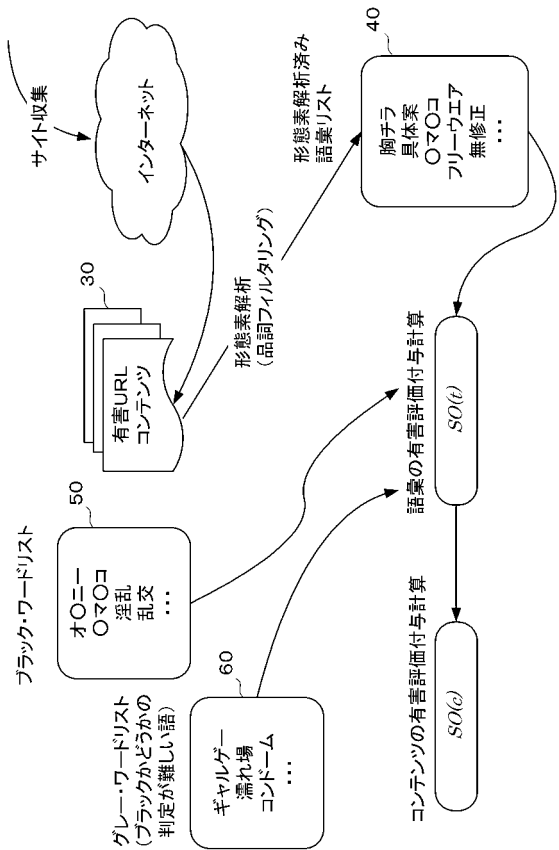
【 図 1 】



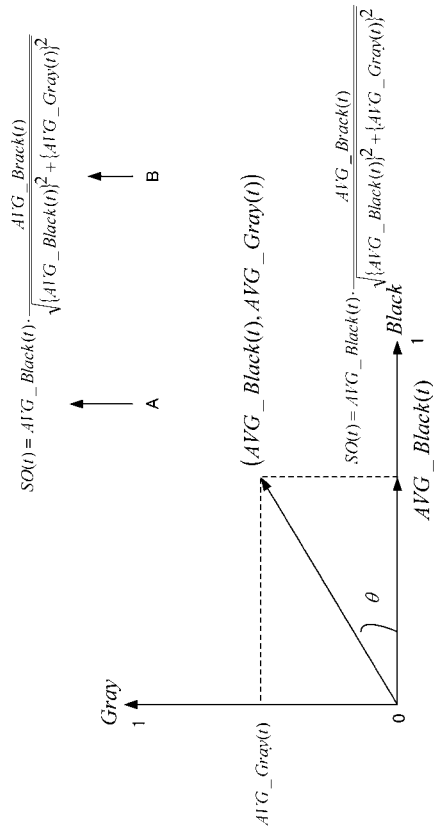
【 図 2 】



【 図 3 】



【 図 4 】



【 図 5 】

$$\begin{aligned}
SO(\text{フ〇〇チオ}) &= \text{AVG_Black}(\text{フ〇〇チオ}) \cdot \frac{\text{AVG_Black}(\text{フ〇〇チオ})}{\sqrt{\{\text{AVG_Black}(\text{フ〇〇チオ})\}^2 + \{\text{AVG_Gray}(\text{フ〇〇チオ})\}^2}} \\
&= 0.4733 \cdot \frac{0.4733}{\sqrt{(0.4733)^2 + (0.3232)^2}} \\
&= 0.3909 \\
\text{AVG_Black}(\text{フ〇〇チオ}) &= \frac{R(\text{フ〇〇チオ, オニ一}) + R(\text{フ〇〇チオ, オマ〇}) + R(\text{フ〇〇チオ, クン〇}) + \dots}{22} \\
&= \frac{528 + 248 + 230}{758 + 488 + 496 + \dots} \\
&= \frac{0.6966 + 0.5083 + 0.4637 + \dots}{22} \\
&= 0.4733 \\
\text{AVG_Gray}(\text{フ〇〇チオ}) &= \frac{R(\text{フ〇〇チオ, レイ〇}) + R(\text{フ〇〇チオ, レズ〇アン}) + R(\text{フ〇〇チオ, ロ〇ター}) + \dots}{277} \\
&= \frac{435 + 123 + 221}{758 + 388 + 758 + \dots} \\
&= \frac{0.5739 + 0.317 + 0.2916 + \dots}{277} \\
&= 0.3232
\end{aligned}$$

【 図 7 】

$$\begin{aligned}
SO(c) &= \gamma \cdot \frac{\sum_{k=1}^l SO(t_k)}{l} \\
&= \frac{3}{13} \cdot \frac{SO(\text{フ〇〇チオ}) + SO(\text{フ〇〇チク}) + SO(\text{フニ一〇ン})}{3} \\
&= \frac{3 \cdot 0.3909 + 0.4095 + 0.371}{13} \\
&= 0.0901
\end{aligned}$$

【 図 6 】

$$\begin{aligned}
SO(\text{フ〇〇チク}) &= \text{AVG_Black}(\text{フ〇〇チク}) \cdot \frac{\text{AVG_Black}(\text{フ〇〇チク})}{\sqrt{\{\text{AVG_Black}(\text{フ〇〇チク})\}^2 + \{\text{AVG_Gray}(\text{フ〇〇チク})\}^2}} \\
&= 0.4321 \cdot \frac{0.4321}{\sqrt{(0.4321)^2 + (0.1456)^2}} \\
&= 0.4095 \\
SO(\text{フニ一〇ン}) &= \text{AVG_Black}(\text{フニ一〇ン}) \cdot \frac{\text{AVG_Black}(\text{フニ一〇ン})}{\sqrt{\{\text{AVG_Black}(\text{フニ一〇ン})\}^2 + \{\text{AVG_Gray}(\text{フニ一〇ン})\}^2}} \\
&= 0.4295 \cdot \frac{0.4295}{\sqrt{(0.4295)^2 + (0.2506)^2}} \\
&= 0.371
\end{aligned}$$

【 図 8 】

$$\begin{aligned}
SO(\text{エロゲー}) &= \text{AVG_Black}(\text{エロゲー}) \cdot \frac{\text{AVG_Black}(\text{エロゲー})}{\sqrt{\{\text{AVG_Black}(\text{エロゲー})\}^2 + \{\text{AVG_Gray}(\text{エロゲー})\}^2}} \\
&= 0.1133 \cdot \frac{0.1133}{\sqrt{(0.1133)^2 + (0.1234)^2}} \\
&= 0.0766 \\
\text{AVG_Black}(\text{エロゲー}) &= \frac{R(\text{エロゲー, オニ一}) + \dots}{21} = 0.1133 \\
\text{AVG_Gray}(\text{エロゲー}) &= \frac{R(\text{エロゲー, ロ〇ター}) + \dots}{253} = 0.1234 \\
SO(18禁) &= \text{AVG_Black}(18禁) \cdot \frac{\text{AVG_Black}(18禁)}{\sqrt{\{\text{AVG_Black}(18禁)\}^2 + \{\text{AVG_Gray}(18禁)\}^2}} \\
&= 0.3214 \cdot \frac{0.3214}{\sqrt{(0.3214)^2 + (0.3213)^2}} \\
&= 0.2273 \\
SO(\text{フ〇ン}) &= 0 \\
SO(\text{本}) &= 0
\end{aligned}$$

SO(デジタル) = 0

