

## WWW における有用性の高いページの特定手法について

野末道子 鉄道総合技術研究所	安形輝 亜細亜大学	石田栄美 国立情報学研究所
久野高志 作新学院大学	池内淳 大東文化大学	上田修一 慶應義塾大学

〔要旨〕 増加する Web ページから必要な情報を探し出す上で、情報源としての有用性が重要な要素となってきた。有用性は一般にそれぞれの閲覧者のニーズによって異なり、主観によって左右されると考えられるが、複数被験者による判定実験の結果、有用性についてはほぼ共通の基準を持っていることが示された。さらに有用性判断を行った 3000 ページの実験環境において、Web ページに出現したワードを用いた有用ページの自動判定システムを試作した。判定ワード群法とサポートベクターマシン法による判定評価においては、サポートベクターマシンの評価結果が高い正解率を示した。

キーワード

WWW、有用性、サポートベクトルマシン

### Evaluation of the Method to Detect Useful Web Pages

NOZUE Michiko	Railway Technical Research Institute (michiko@rtri.or.jp)
AGATA Teru	Asia University
ISHIDA Emi	National Institute for Informatics
KUNO Takashi	Sakushin Gakuin University
IKEUCHI Atushi	Daito Bunka University
UEDA Shuichi	Keio University

#### Abstract

For retrieving required information from increasing Web pages, usefulness is becoming an important element for the source of information. Although it is thought that usefulness generally changes with the needs of each user and is influenced by subjectivity, it is shown that we have an almost common standard about usefulness as a result of judgment experiment by nine judges.

In this study, we developed a usefulness judging system by using the words which appears on web pages in a 3000-page test environment. In the evaluation of the judgements by the method of vector space model and that of the support vector machine, latter shows higher degree of effectiveness.

## 1 Web ページと有用性

WWW は、出現してからわずか 10 年たらずのメディアであるが、社会や生活の中に浸透している。30 億ページを超えるとされる Web ページは、全体量が増えるにつれて、情報源としての重要性をまし、インターネットの常時接続が通常となるここにより、印刷体を超える情報入手手段となっている。

初期においては、Web ページは、全体に質が低いという批判がなされてきたが、量的な増大とともに質の高い Web ページの絶対量が増加したため、こうした批判は無意味となっている。一方では、Web ページの評価に対して関心が集まりはじめています。

本稿では、2 章においては著者らの先行研究における Web ページ評価の枠組みと主観的な判断である「有用性」に関する評価実験について紹介する。3 章以降では、複数被験者による Web ページの有用性判定実験について示し、さらにその自動的な判定手法を提案する。判定実験対象としたのは日本語で表現される Web ページである。

現在、サーチエンジンにおいては、有用性に関して、被リンク数に基づく自動判定方法検索結果の自動が定着しているが、本研究においてはサイト全体ではなく、個々の Web ページに含まれる情報のみを利用した。

## 2 Web ページ評価の枠組み

ここで言う Web ページの「有用性」とは、その閲覧者、利用者にとって役に立つかどうかであるとともに、誰にとっても役に立たないページ、いわゆる「くず」ページではないという両面を意味する。こうした有用性は、一般にそれぞれの閲覧者、利用者のニーズによって異なり、主観によって左右されると考えられるが、ここでは、多くの人々が有用性について共通の基準を持っていると仮定する。現に、ページランクなどの手法によって提示されるサーチエンジンの検索結果について、利用者の大多数は満足している。

### 2.1 伝統的な評価基準とその問題点

1999 年に Alexander らは、伝統的な評価基準として(1)権威、(2)正確さ、(3)客観性、(4)カレ

ントであること、(5)収録範囲、があげられてきたが、これらはそのままでは Web ページには適用しがたいと主張している<sup>1)</sup>。例えば、「権威」はその主題について適切な著者によって書かれているかによって判断する。印刷書籍の場合は、著者の経歴や業績によって判断可能であるのに対し、Web ページの場合は著者名が記載されていないことも多く、記載されていたとしても、著者紹介等の情報源に辿りつくのは必ずしも容易でない。

そこで、これまでの Web ページの評価基準に含まれる評価項目と Web ページのアクセスを増やすために必要とされている項目とを洗い出し、これらをまとめて、Web ページの評価の枠組みを作成した。さらにそこから独自の評価項目を導き、被験者に Web ページを見せ、それぞれの評価項目の重要度をみることにした。

### 2.2 評価の視点と評価規準

Web ページの評価基準に関する Web ページは数多く存在しているが、その中で Smith は、インターネット上の情報源の評価基準に関するホームページを開設し、この中には、個人や機関が作成した 58 種の基準がリンクされている<sup>2)</sup>。そして Smith は、これらをまとめて表 1 のような評価項目を作っている<sup>3)</sup>。

表 1 Smith による評価項目

分類	項目
範囲 (scope)	主題の広がり(Breadth), 詳しさ(Depth), カバーする年代(Time), 形式(Format)
中味 (Content)	正確さ(Accuracy), 権威(Authority), 新しさ(Currency), 類似のものがない(Uniqueness), リンク(Links), 文章の質(Quality of writing)
デザイン(Graphic and multimedia design)	
目的(purpose)	対象利用者(Audience)
評価の仕組み(reviewing)	
利用者支援(Workability)	使いやすさ(User friendliness), コンピュータの環境(Required computing environment), 探索機能(Searching), 一覧機能(Browsability and organisation), 対話機能(Interactivity), アクセス性能(Connectivity)
支払う費用(Cost)	

次に、12種の国内のWebページ作成者向けのガイドに記載されている要素を分析した。これらは、主としてアクセスを増やすための観点から考えられたものであって、情報源となるページを想定したものではない。しかし、両者には共通する評価項目が多くみられた。

これらのWebページを情報源として評価する基準と、アクセスを増やすための項目とを検討し、Webページの評価について作成者、利用者、物理的アクセス状況の三つの視点から整理した<sup>4)</sup>。

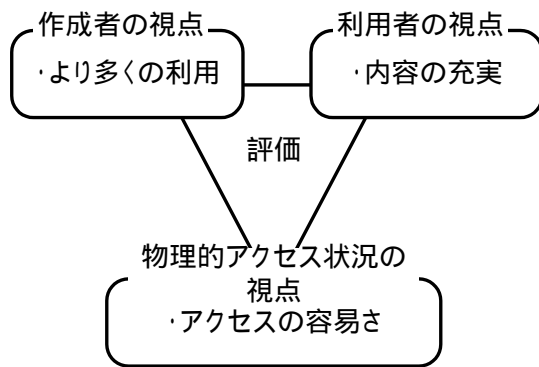


図1 Webページ評価の視点

a) 作成者の視点

当該Webページに対するアクセスを増加させるための評価項目である。これはWebページの構築に関わるもので、テーマの独自性、明確さなどが含まれる。

b) 利用者の視点

利用者がWebページを閲覧する際に何らかの情報を得るのに関わる項目である。内容の充実が大きく関与し、正確さ、速報性などが含まれる。

c) 物理的アクセスの視点

当該Webページにアクセスする際の快適さに関わる評価項目である。ページのデザインに関連し、その量や利用者のコンピュータ利用環境への配慮などが含まれる。

これら三つの視点のうち、作成者と利用者の視点に即して、15項目の独自の評価基準を選定した(表2の評価項目)。

表2 Webページ評価の枠組みと評価項目

Smithの項目	アクセスの増加	評価項目	視点
使いやすさ	操作性	量が豊富である	作成者の視点
一覧機能		見やすくするための工夫がある	
「デザイン」	ページ構成 タイトル 統一性	ページのデザインがよい ページタイトルが適切である ページ内のテーマが統一されている	
リンク	リンク リンク集	他のページへのリンクが多い	
対象利用者			利用者の視点
類似のものなし	独自性		
	コンセプト	テーマがわかりやすい テーマが明確である 詳しい内容である	
権威		作者に専門的な知識がある 信頼できる作者である	
正確さ		内容が正確である	
新しさ		最新の内容である 定期的に更新されている	物理的アクセスの視点
文章の質		正しい日本語でかかれている	
アクセス性能	軽快さ 画像、音声等少 ページが小さい		
探索機能	データベース		その他
対話機能	掲示板等		
「評価の仕組み」, 「支払う費用」			
	多言語、ソフト ウェア		

2.3 評価項目と有用性との関係

次に、被験者にwebページを見せ、前項に示した評価項目による評価実験を行った。この調査の概要は次の通りである。

a) 調査方法の概要

被験者に対象ページごとに「よい情報源」であるかどうかを判定させ、さらに各評価項目について5段階(5:強く思う、4:そう思う、3:どちらでもない、2:そう思わない、1:全く思わない)で判定させる。

b) 評価項目

最初に被験者を学生26名として予備調査を行ったが、その結果に基づき、同じ回答パターンを示した「テーマがわかりやすい」と「テーマが明確である」をまとめ「テーマが明確でわかりやすい」とした。

c) 対象ページの選定

調査の対象としたページ集合は、以下のような手順で収集した。

Yahoo! Japan から約22万のURLを取得  
ロボットにより2レベル(と3レベルの1部)までの約500万のURLを取得

無作為な5000URLを抽出し、画像等まで含めてダウンロード

自動タイプ判定システムにより「標準とさ

れた」ページ群から 1000URL を無作為抽出

なお 最後の段階でページのタイプ判定を行っているが、これは「リンク集」、「掲示板」などを排除するためである。実際には、先行研究 5)において高い精度(76.9%)で「標準」ページを判定できたタイプ判定システムを使った。

d) 被調査者

被調査者は、社会人、主婦、学生各 3名の計 9名である。各被調査者は 500 ページを判定したが、計 1000 ページのうち 500 ページは 6 名、残りの 500 ページは 3 名で判定した。

e) 調査結果

項目「このページはよい情報源である」と各評価項目との相関は表 3 のようになった。

表 3 「よい情報源」と評価項目間の相関

評価項目	相関係数
このページはよい情報源である。	1.0000
テーマがわかりやすく、明確である。	0.5798
信頼できる作者である。	0.4975
内容が正確である。	0.4969
詳しい内容である。	0.4878
見やすくするための工夫がある。	0.4594
ページ内のテーマが統一されている。	0.4308
量が豊富である。	0.4110
作者に専門的な知識がある。	0.3846
ページタイトルが適切である。	0.3790
ページのデザインがよい。	0.3520
正しい日本語でかかれている。	0.3290
最新の内容である。	0.3109
定期的に更新されている。	0.2829
他のページへのリンクが多い。	0.2398

まず、被調査者の属性によって評価の違いの有無をみたが、性別、年齢、インターネット利用歴による差はとりたててみられなかった。

各評価項目を最短距離法でクラスタ分析結果した結果を図 2 に示した。

2.4 各評価項目と定量的指標の相関

対象となった各 Web ページと文字数、タグ数、リンク数、各タグの出現数などと定量的指標の相関を分析した。

項目「このページはよい情報源である」と定量的指標の間には、直接的な相関は見られなかった。

一方、「ページのデザインがよい」と画像数 (0.26) 「他ページへのリンクが多い」とリンク数 (0.40) 「量が豊富である」と文字数 (0.25) などの間には、ある程度の相関が見られた (括弧内は相関係数)。

以上から、被調査者が判断の基準とした上位の項目は、いずれも当該 Web ページから判断できず、外部の情報をもちいるしかなく、また、有用性を定量的特徴から判断することも困難であることが判明した。

そこで、さらに判定実験の規模を拡大し、新たな実験環境を構築するとともに、有用な Web ページと有用でない Web ページ上に出現する語に着目した有用性の自動判定を試みた。

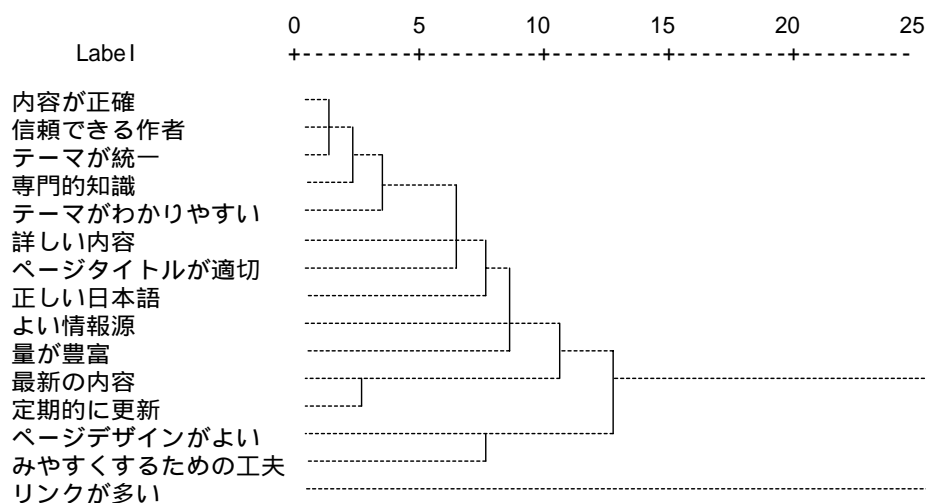


図 2 評価項目間のクラスタ分析結果

### 3 有用性自動判定実験

#### 3.1 実験用データの作成について

以下では、Web ページの有用性自動判定のための実験用データ作成方法について概説する。まず、任意の意味をもたない文字と数字をキーワード群とし、五つの検索エンジンを用いて検索を行い、約一万件の URL 集合を収集した後に、それらの URL 集合から無作為に 3,000 件の URL を抽出して、画像を含めた全ての情報をダウンロードした。

ここで、ロボットによって直接的にページを収集するのではなく検索エンジンを用いたのは、実験用ページ集合を Web ユーザが日常的にアクセスする可能性のあるページ群に模するためであり、その結果、収集された 3,000 ページの平均バイト数は 10,000.1 バイト、平均文字数は 3,468.8 字となった。

次に、収集されたページ群に対して、それぞれ有用性の判定を行った。判定を行ったのは、20代から 30代までの男女(9名)であり、いずれも、日常的にインターネットを利用し、Web ページをブラウズしていることを条件とした。ここでは、判定者は三名ずつ 3 グループに分けられ、各自が 1,000 ページずつを担当した。従って、1 ページにつき三名が判定を行っていることとなる。

また、本来、主観的な有用性の概念について、判定者間で一定の共通イメージを持ってもらうために、質問文では、Yahoo! (<http://www.yahoo.co.jp/>) に登録できるか否かをその判定基準として設定した。具体的には、「そのページを見て、有用かどうかを判断してください。あなたが Yahoo の査定者になったと仮定した場合に、そのページを登録するかどうかを判断して下さい。」と尋

ね、「採用する」、「採用したいが問題あり」、「どちらとも言えない」、「残念だが採用しない」、「採用しない」の五段階評価を行ってもらった。

なお、判定の際、判定者間でのコミュニケーションはなく、すべて各人の判断によっている。また、個人個人による内的な判定基準が大きく変化する可能性を考慮して、作業期間は最長でも一ヶ月までとした。

判定者間での判定結果の一致度は、述べ 3,000 ページ中、「三者一致」が 521 ページ (17.4%)、「二者一致」が 1876 ページ (62.5%)、「三者不一致」が 603 ページ (20.1%) であった。二者一致までで全体の約 80% を占めていることから、充分高い値が得られたと言えるだろう。

次に、判定結果の分布を表 5 に示す。ここで、「判定値」は回答者による判定結果そのものであり、「平均値」は回答者による判定結果の分散を考慮し、ページごとに、三人の判定値を 1~5 までの数値に置き換えた後に(「採用する」=1~「採用しない」=5) その平均値を算出し、小数点以下を四捨五入したものである。また、今回の有用性自動判定実験は、後者の「平均値」に基づいて行っている。

表 5 の「判定値」を見ると、前述の通り、検索エンジンを用いて収集したページ群であることから、全体で「採用する」が半数を超えるなど、Web の実態と比して、有用性の高いページの比率が多くなっていると考えられる。「平均値」については、全体で最も多いのは「採用したいが問題あり」(43.3%) であり、「採用しない」(2.1%) が最も少なくなった。

表 4 有用性判定結果の分布

	判定値		平均値	
	件数	割合	件数	割合
採用する(1)	4,824	53.6%	816	27.2%
採用したいが問題あり(2)	1,066	11.8%	1,298	43.3%
どちらとも言えない(3)	906	10.1%	417	13.9%
残念だが採用しない(4)	529	5.9%	405	13.5%
採用しない(5)	1,610	17.9%	64	2.1%
(空欄)	65	0.7%		
合計	9,000	100%	3,000	100%

\*1 判定値は回答者による判定結果。1 ページ当り 3 人が判定しているため、述べ 9,000 ページとなっている。

\*2 平均値は、3 人の判定結果の平均値 (小数点以下四捨五入)

### 3.2 有用性の自動判定手法

Web ページの有用性の判定は、判定ワード群による判定手法とサポートベクトルマシン（以下 SVM とする）の二つから行い、結果を比較している。それぞれの処理は図 3 に示す手順で行った。

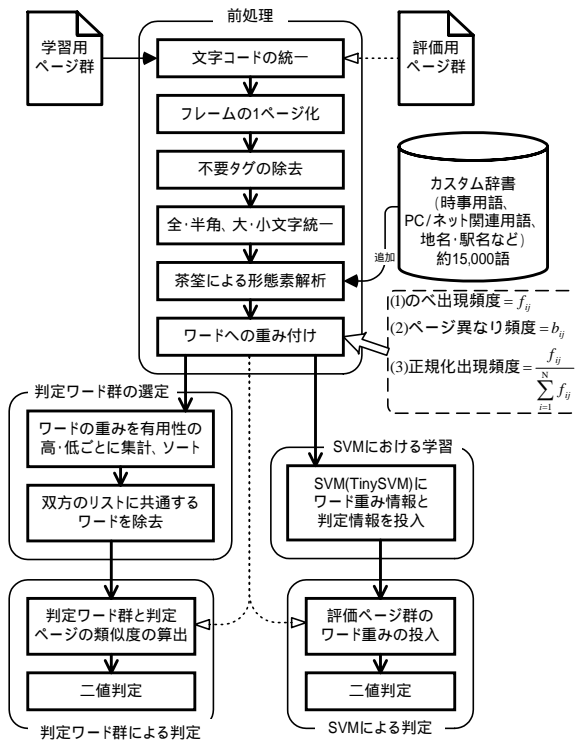


図 3 有用性判定の手順

判定ワード群、SVM による判定を行う上での基本的な仮定としては、有用性が高いページ群と低いページ群では、使われる語彙の出現傾向が異なっていることである。つまり、この傾向の差異を利用し、有用性判定を行うのである。

本実験では有用性平均値が 3.5 未満を有用ページ、3.5 以上を不用ページとして 2 値分類判定を行った。以下の文章中における有用、不用ページという言葉はこの 2 値分類に基づいて記述しているものである。

#### 3.2.1 前処理による出現ワード取得と重み付け

今回のテスト集合から、各ページの出現ワードを取得するために、以下の前処理を行った。

#### 文字コードの統一

混在する文字コードを、EUC に統一した。

#### フレームの 1 ページ化

フレームタグによって複数のファイルから構成されているページを 1 ファイル化した。

#### 不要タグの除去

表示書式を記述するタグ情報部分を除去した。

#### 全・半角、大・小文字の統一

カナ、英、数文字における、半角文字の全角化、英小文字を大文字化して統一した。

#### 茶釜による形態素解析

Web ページでは時事用語や固有名詞が多く出現する。ここでは追加辞書として、時事用語、PC/ネット関連用語、地名・駅名などから構成される約 15,000 語を登録した。解析により取得した形態素集合から、記号文字を除いたものをすべて、各ページの出現ワードとした。

#### ワードへの重み付け

重みの算出の方法としてはいくつか考えられるが、ここでは(1)ページ異なり出現頻度、(2)延べ出現頻度、(3)正規化出現頻度の三つの手法を検討した。延べ出現頻度とは、ページ群全体を  $D$  としたとき、各ページ  $d_i$  ( $i = 1, 2, 3, \dots, N$ ) における語  $t_j$  ( $j = 1, 2, 3, \dots, M$ ) の出現頻度  $f_{ij}$  である。ページ異なり頻度  $b_{ij}$  はページ中にあるワードが出現するかを示しているため、0 か 1 の二値となる。正規化出現頻度は、のべ出現頻度

$f_{ij}$  をページ  $d_i$  中の出現頻度の総計  $\sum_{j=1}^M f_{ij}$  で正規化したものである。

この手順により取得した各ページの出現ワード数（のべ）と有用性平均値との関係を表 6 に示す。表 6 出現ワード数と有用性の結合度数分布

	有用性判定(平均値)				合計	
	4-5 有用度 が低	3-4 有用度 がやや 低	2-3 有用度 がやや 高	1-2 有用度 が高		
出現 ワ ド 数	0	44	42	142	11	239
	1-100	45	129	107	125	406
	101-500	48	197	317	434	996
	501-1000	9	65	177	286	537
	1001-2000	9	62	146	217	434
	2001以上	18	70	142	158	388
	合計	173	565	1031	1231	3000

フレーム処理、タグ除去の自動処理の不備により、ワードを取得できなかったページが若干含まれているものの、ワードが全く取得できなかったページが239ページ、全体集合のうち約8%となり、これらは今回の判定手法においては判定不能ページとなった。

### 3.2.2 判定ワード群による判定

判定手順としては、学習フェーズとして、学習用ページ群における有用、不用ページ群別に各ワード毎の重みを集計、各ワード毎の重みの高い順にソートし、50位単位で500位までの各ワード集合を生成、双方のリストに共通する語を除去し、正(有用)、負(不用)の重みを持つ判定ワード群とする、判定フェーズとして、判定ワード群と判定ページの類似度を算出、類似度値が正負のいずれかにより、二値判定を行う、というものになる。

特定の学習ページと判定ページによる判定結果への影響をなくすため、交差検定を行いこの平均値をとった。表7は、語の出現頻度の算出手法別の判定結果を示したものである。

表7 各ワード重み算出手法別判定結果

	異なり	延べ	正規化	平均
正解率	57.3%	65.9%	60.3%	61.2%
不正解率	23.5%	17.0%	20.9%	20.5%
同点率	5.5%	3.9%	5.6%	5.0%
0ヒット率	5.7%	5.2%	5.2%	5.4%
判定不能率	8.0%	8.0%	8.0%	8.0%

ここで正解率は有用、不要を正しく判定、不正解率は判定誤り、同点率は有用、不要語の出現確率が同点、0ヒット率はそれぞれのワード群に共通ワードがない、判定不能率は出現ワード数が0であったページの割合である。

学習・判定セットによる判定ワード違いはみられたものの、いずれのセットにおいても、延べの出現頻度で順位付けしたワードリストを用いて作成した判定ワード群が、他の2手法による結果の成績を上回った。この理由として、延べの出現頻度から生成されたワード群において、有用、不用ワードリスト間での重複が少なかったために、判定ワードの数が多かったことも一因と考えられる。しかし、判定ワードが少なくてもかなりよい判定結果となったものもあり、一概にワード数の多さを有効性に結びつけることはできない。

判定ワード群の特徴としては、有用ページと比較して、不用ページにおける名詞、動詞などの内容語以外が非常に多く見られた。この結果から、判定精度を高めているのは、必ずしも有用ページ判定ワード群上位となる「情報」「利用」等のワードのみならず、「ね」「じゃ」「いい」「もう」などの不要ワード群の上位となるワードであると考えられる。

なお、判定ワード群の選択に際して、名詞、動詞、形容詞等の内容語による判定と、助詞、助動詞、接頭詞、感嘆詞等を含む出現語全てによる判定のいずれが良いか予備実験を行った。この結果、内容語のみの抽出を行った場合には、人間が判断すると一見「有用そうな」ワードが選択されるものの、判定実験における全出現ワードとの判定精度比較において、2~20%正解率が劣ることがわかった。そこで、本実験では、判定ワード群、SVMともに全ワードを選択した。

次に、多すぎる語で判定することは処理上、非効率であることから、判定ワード群をどの規模で生成するのが効率的であるかを検討した。この結果を示したのが図4~6のグラフである。

いずれの手法においても、150~200位でほぼ正解率は横ばいとなり、この近辺でのワード選択が最も効率的という結果になった。

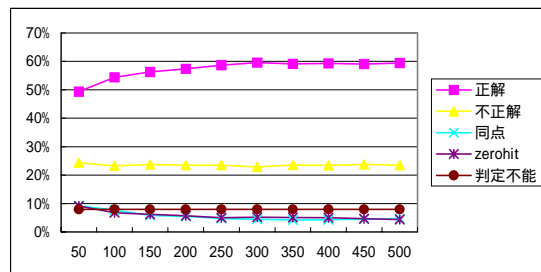


図4 判定ワード群生成規模別判定結果の推移 (出現頻度算出方法は異なりによる)

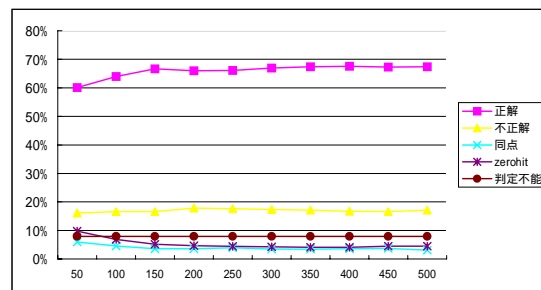


図5 判定ワード群生成規模別判定結果の推移 (出現頻度算出方法はのべによる)



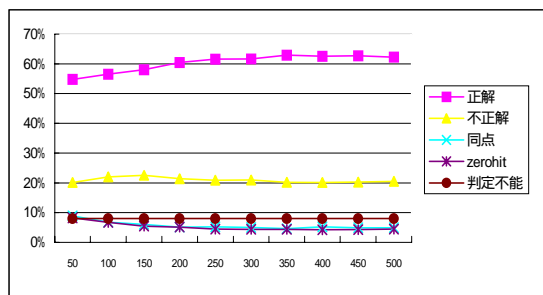


図6 判定ワード群生成規模別判定結果の推移 (出現頻度算出方法は正規化によるもの)

### 3.2.3 SVM による判定

判定ワードによる判定においても、ある程度 実際的な判断が可能であると考えられるが、この結果では、ランダムな判定をした場合に比べても、良好な結果であるとは言いがたい。

そこで、文書分類や自然言語処理分野において高い精度を示している SVM を利用して、判定を行った。判定ツールは工藤氏により提供されている TinySVM<sup>1)</sup> を使用している。多くのカーネル関数が提案されているが、ここで用いたのは以下の Polynomial Kernel 関数である。

$$K(x, y) = (x \cdot y + 1)^d$$

ベクトル要素としては各ワードを、各々の要素に対し、のべ、異なり、正規化のワード重みを与えた。1つもワードの出現しない判定不能データは外し、2761 ページを今回の実験対象とした。

判定は判定ワード群と同様に交差検定を行い、平均値を取った。判定結果を表8に表す。

表8 SVM による有用性判定結果

	異なり	延べ	正規化
再現率	93.98%	89.25%	90.08%
判定精度	90.81%	92.29%	92.07%
正解率	85.87%	83.44%	88.08%

この表における再現率は、有用ページがどれだけ再現できたかという割合で90%前後である。

また、判定精度は、有用と判定されたものの中で有用ページだったものの割合で、いずれも90%台となった。

最後の正解率は、有用、不用の判定が正しくできた割合であり、この結果はいずれも85%前後であるが、若干正規化のものが他の手法を上回った。実際には不用ページの割合は14.5%であることから、全て有用と判定した場合に比べて正規化の結果は若干これを上回っていることになる。

## 4 おわりに

本論文では、Web ページに対する有用性という観点について検討を行い、判定実験を通して、利用者間で共通性のある有用性判断基準が存在することを示した。またこれらの有用性判断の際に、表示ワードにより有用性の判定が行われているという仮説を立て、自動判定の可能性を2つの手法により検証した。

判定ワード群による手法の特徴としては、学習フェーズにおいてあらかじめ有効な判定ワード群を生成しておくことができ、有用性の判定を比較的簡単に行うことができ、応用範囲も広い点である。また、判定ワード群の個別要素であるワードについて、より詳細な分析を行うことも可能である。

一方、SVM による判定結果により、判定精度と有用ページの再現率としては、かなりのレベルで達成可能であるが、判定正解率をさらに上げるための工夫が必要である。

今回の自動判定実験では、人手による判定実験の際に得ている Yahoo のカテゴリ情報や、フォームやリンク集といったページページタイプ情報を利用していない。今後は、これらの情報も含め、さらに判定精度を高めることが可能かどうかを検討中である。

1) <http://d.aist-nara.ac.jp/~taku-ku/software/TinySVM/>

### 【引用文献】

- Alexander, J. E. Tate A.T. Web wisdom : how to evaluate and create information quality on the Web. Mahwah, Lawrence Erlbaum Associates, 1999.156p.
- Smith, Alastair. Evaluation of information sources <<http://www.vuw.ac.nz/~agsmith/evaln/evaln.htm>>[2002-07-07]
- Smith, Alastair. VUW Department of Library and Information Studies, New Zealand. Criteria for evaluation of Internet Information Resources. <<http://www.vuw.ac.nz/~agsmith/evaln/index.htm>>[2002-07-07]
- 上田修一;久野高志; 安形輝;石田栄美. Web ページ評価の視点と基準. 三田図書館・情報学会研究大会発表論文集 2000 年度. p.33-36(2000)
- 久野高志; 安形輝; 石田栄美; 上田修一. Web ページのタイプ判定法. 2000 年度日本図書館情報学会春季研究大会発表要綱. p.55-58(2000)