

リンクに基づいた明示的 Web ページ評価法

平林 真実^{†1,†2} 大月 一弘^{†3} 清光 英成^{†3}
森下 淳也^{†3} 北村 新三^{†2} 絹川 達也^{†4}

本論文では、リンクに対する意図を反映した Web ページの点数化方法を提案する。我々は、Web ページ作成者がリンクを設定する際の意図に注目し、リンクに対し推薦度を導入している。CLEVER や Google などではすべてのリンクは推薦として想定されていた。推薦度を導入することで、リンクに対してページ作成者の意図を明確に指定できるようになる。また、我々の方法では、Web ページの内容やリンク構造などのデータからは導出できない外部評価を評価点付けに利用している。本提案方法では、まず Web 利用者が外部評価に基づきページの初期評価を行い、さらに推薦度を付けられるように拡張されたリンク構造により、各ページの評価点が求まる。同時に、本方法によりページ間の評価伝播とそのアルゴリズムを示している。

Semantic Web Page Scoring Based on Intention of Hyperlinks

MASAMI HIRABAYASHI,^{†1,†2} KAZUHIRO OHTSUKI,^{†3}
HIDENARI KIYOMITU,^{†3} JUNYA MORISHITA,^{†3} SHINZO KITAMURA^{†2}
and TATSUTA KINUGAWA^{†4}

We propose a method for scoring a web page to which is reflected an intention of the corresponding hyperlinks. In this paper, we focus to the semantics of hyperlinks such that a web page author intends as a recommendation. Conventional web search engines, CLEVER and Google, recognize the meanings of all hyperlinks as the recommendations. In our Recommendation Linkage introduces a rating of recommendation that is specified by the author explicitly. Also, our scoring method use external factors that are not derived from web page contents and/or hyperlinks. Our proposing method is as follows. First, a user judges the scores of pages based on the external factors. Then, the score of a linked (destination) page is calculated by using some established recommendation linkages. Moreover, we show the propagation mechanism and algorithm of a base-page score via recommendation linkage.

1. はじめに

Web におけるページ評価方法としては、検索技術を中心に多くの方法が研究、検討され、また利用されている¹⁾。キーワード検索や構文解析を基本とする内容に基づく方法から、Google²⁾における PageRank^{3),4)}や CLEVER における Kleinberg らによる方法⁵⁾、Clan Graph⁶⁾などのリンク構造解析を基本とするものが注目されている。

利用者が Web ページを見てページを判断する際には、誰がページを作成したのか、どのような立場で書いているのか、どのような目的で書かれているのかなど Web ページ(集合)自体からは導出できない Web とは独立した外部評価を行っている。つまり、利用者が Web を見るときに暗黙的に行っている外部評価は、Web を見る目的や環境に応じて恣意的な判断の下に行われていると考えられる。

そこで、このようなページに対する外部評価に基づいて、リンク構造によりすべてのページに評価点を付け、ページ評価の判断へ利用する方法を提案する。本提案方法では、Kleinberg⁵⁾も指摘したように従来は推薦の意味でのみとらえられることの多かったリンクに対して、推薦度を付与できるように拡張した。リンクに推薦度を付けることで、リンク構造による評価に役立つだけでなく、ページ作成者のリンク設定に関する意図を明示的に表明でき、ある種の意味付けが可

†1 岐阜県立国際情報科学芸術アカデミー

International Academy of Media Arts and Sciences

†2 神戸大学大学院自然科学研究科

Graduate School of Natural Science, Kobe University

†3 神戸大学国際文化学部

Department of Cross Cultural Studies, Kobe University

†4 神戸大学大学院総合人間科学研究科

Graduate School of Cultural Studies and Human Science, Kobe University

能となる。

本論文では、外部評価と推薦度付きのリンク構造から、ページ間の相対評価を用いて行うページの評価の考え方について示す。さらにページ評価点として評価最大値を採用することでページ評価を行う手法を提案する。最後に、本手法に基づき実験を行い、その妥当性を示し、本ページ評価点の利用法についても議論を行う。

2. ページ評価の考え方

我々の提案するページ評価方法では、Web に含まれる情報からは導出できない Web に対する外部評価と、リンク先ページへの推薦度を明示的に表明できるように拡張されたリンクを持つリンク構造よりページ評価を行う。

2.1 ページ評価の要素

ページに対する外部評価

Web ページの外部評価とは、ページ(集合)内コンテンツやリンク構造などとは独立で、Web からは導出できないデータを用いた判断あるいは評価である。

我々は普段、URL などからページを運営している組織やページ作成者を推測し、暗黙のうちにページ評価に利用している。このとき、URL から運営組織やページ作成者が推測できるが、その組織や作成者に対する評価自体は、本来 Web ページのコンテンツやリンクから得られるものではない。つまり、新聞社のニュース記事や有名な企業の広報ページならば高い評価を持って読むことができるが、個人ページ内の情報に対して最初から高い評価を持ってページを見ることはない。このような評価は、Web とは別の外部情報として新聞社や企業、個人を評価しているのであり、Web 閲覧者の持つ知識に基づいたものと考えられる。

また、評価を行うとき、Web データから得られた情報から Web 閲覧者がどのような評価を行うかも、評価者や評価基準によって様々であり、目的や状況に応じた視点に基づいて任意に決定されるものである。

本方式では、外部評価として任意の恣意的視点から決定されるページの評価を利用する。

推薦度付きリンク

推薦度付きリンクは、リンク作成者が推薦という視点からリンク先ページに対して行っている評価を推薦度として指定できるようにリンクを拡張したものである。

リンクには様々な意図があるが、従来それを表明する方法がなかった。ただしリンクの意図は、推薦だけではなく単純な参照や「戻る」などの構造を示すため

のリンクなど様々であり、推薦度だけですべてを表すことはできない。また、自分の作成したページに対するリンクと他人の作成したページに対するリンクでは、リンクの持つ意味が異なっている。

しかしながら、従来はリンクするかしないかの選択肢しかなく、リンク構造解析を利用したページ評価では推薦の意味のみでしかとらえられていなかった。推薦度を導入することで、ページ作成者がリンク先ページに対してどの程度好意的に評価しているか、あるいは構造だけを示し推薦を含まないリンクなのかを推薦という視点から明示的に表明できるようになる。

2.2 リンクによるページ評価の概念

本方式では、ページによるページの評価という考えによりリンクによる評価を行う。つまり、リンクを介したページ間の相対評価という考え方をを用いている。

ここで、リンクにより構成される Web 構造を、ページをノード集合 N 、リンクを枝集合 E とする有向グラフ G とし、以下のような表記を行う。

$$G = (N, E)$$

$$N = n_i \mid \text{ページ集合中のページ}$$

$$E = e_{ij} \mid \text{ページ集合中の } n_i \text{ から } n_j \text{ へのリンク}$$

$$q_i : n_i \text{ の外部評価点}$$

$$w_{ij} : \text{リンク } e_{ij} \text{ に付与された推薦度}$$

$$(\text{ただし, } 0 \leq w_{ij} \leq 1)$$

w_{ij} の値については、

- 自ページの評価点以上を他ページの評価点として与えることは不自然であること、
- 推薦しないことを負の推薦度で表現することも考えられるが、本論文では負の評価点を考慮しないこと、

を意図している。

図 1 にページ n_a を起点とし、ページ n_i に至る経路 x 上の推薦度を用いたページ評価の例を示す。図 1 (a) は初期状態である。ページ n_a の外部評価点 q_a とページ n_a からページ n_b へのリンクに付与された推薦度 w_{ab} 、ページ n_b からページ n_i のリンクに付与された推薦度 w_{bi} が与えられている。また、ページの外部評価点 q_b 、 q_i が 0 点となっているのは、あらかじめ与えられていないため便宜上 0 点としている。

図 1 (b) は、リンクによる評価が行われた結果である。ページ n_a を起点として、 q_a を 100 点と評価すると、ページ n_b の評価点は、推薦度が 0.8 であるため 100×0.8 つまり 80 点となる。さらに、ページ n_i の評価点はページ n_i への推薦度が 0.5 であるため 80×0.5 つまり 40 点となる。

ここで、

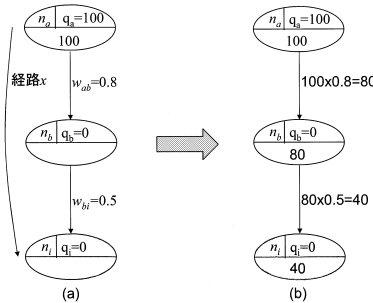


図1 リンクによるページ評価
Fig. 1 Page evaluation by links.

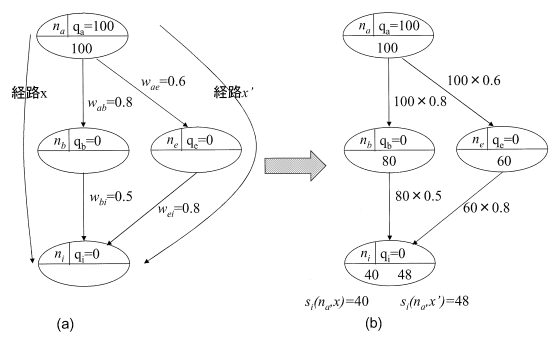


図3 経路が2つある場合の評価例
Fig. 3 Evaluation with multiple paths.

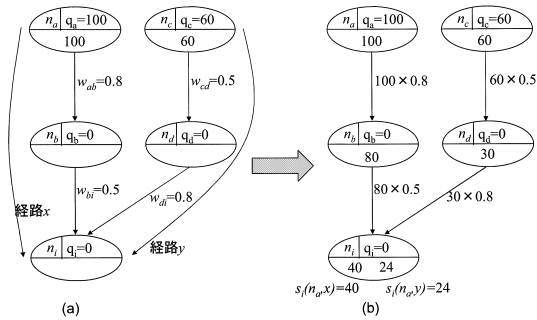


図2 複数リンクによる評価例
Fig. 2 Evaluation with multiple start-pages.

$s_j(n_j, x)$: ページ n_i からページ n_j へのある経路 x による評価点
とし、各ページに外部評価点が与えられているとすると、

$$s_i(n_a, x) = q_a \cdot w_{ab} \cdot w_{bt} = 100 \times 0.8 \times 0.5 = 40 \quad (1)$$

となる。

図2では、ページ n_i に対して複数の被リンクが存在する場合を考える。この図ではページ n_i に対して、 $n_a \rightarrow n_b \rightarrow n_i$ を通る経路 x と $n_c \rightarrow n_d \rightarrow n_i$ を通る経路 y が存在する。

ページ n_a を起点とした場合は、図1と同様にページ n_i の評価点は40となる、しかしページ n_c を起点とした場合には、ページ n_i の評価点は24となる。つまり起点となるページの違いによって、ページ n_i の評価点は異なることになる。

すなわち、ページ n_i の評価点は一意には決まらず、

$$\begin{aligned} s_i(n_a, x) &= q_a \cdot w_{ab} \cdot w_{bt} = 100 \times 0.8 \times 0.5 = 40 \\ s_i(n_c, y) &= q_c \cdot w_{cd} \cdot w_{dt} = 60 \times 0.5 \times 0.8 = 24 \end{aligned} \quad (2)$$

となる。

さらに図3では、ページ n_a からページ n_i に対し

て2つの経路が存在する。この場合には、経路上のリンク推薦度の違いによって、ページ n_i の評価点が異なることになる。

すなわち、経路 $n_a \rightarrow n_e \rightarrow n_i$ を x' すると

$$s_i(n_a, x') = q_a \cdot w_{ae} \cdot w_{ei} = 100 \times 0.6 \times 0.8 = 48 \quad (3)$$

となる。

以上のように、あるページの評価点は、任意のページを起点としてどのような経路をたどって評価されたかによって決まる。すなわち、あるページに注目したとき、その評価点は起点となるページと経由するリンクにより異なることになり、あるページを起点とした、そのページの評価として非常に多くの評価点が存在する。

我々がWebページを見てリンクをたどるときを考えると、リンク先ページがどのようなページにリンクされているかは既知ではなく、自分のたどっているリンクによってのみ判断していると考え、1つのページに起点や経路によって複数の評価が存在するのは妥当といえる。多数の評価点からページ評価点を決定する方法としては、評価の個数や偏差を計算するなど、いくつかの方法が考えられる。

我々は、評価点の決定について、最大評価点の採用を方針とした。すなわち、ページ n_i に対する評価点を ps_i とし、集合 Z_i をあるページ n_i への評価点の集合とすると、

定義： ページ n_i の評価点は

$$ps_i = \max(q_i, \{z | z \in Z_i\}) \quad (4)$$

ただし、 $\{z | z \in Z_i\}$ は Z_i に属するすべての要素を表す

と表される。

評価点として最大値を採用することは、少なくともそれだけの評価がなされているという評価のポテンシャルを示していると考えられる。起点とな

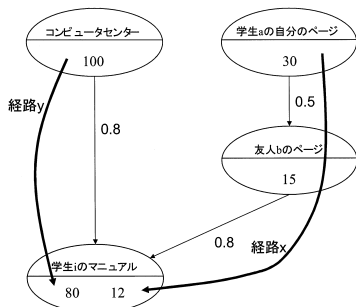


図 4 評価点の採用の判断
Fig. 4 Judgement in score while browsing.

るページからの何らかの積極的な評価が、リンクにより伝播しつつ、あるページの評価点が決定されることから、少なくともある程度の推薦という評価を行ったページ群が存在している。つまり、ある視点からの評価において、良いページである可能性を持っていると考えられる。

たとえば、図 4 に示すように、大学において学生 a があるシステムの使い方を探しているとき、経路 x の場合、自分のページを起点として友人 b のページのお勧めリンクを経由し学生 i のページに到達し、評価点は 12 となる。ところが、学生 i 書いたマニュアルが良くできていて、システムを管理しているコンピュータセンターからお勧めリンクが張ってあるとする。このとき、経路 y が存在し、学生 i のページの評価点は 80 となる。

この例では、学生 i のマニュアルページに対する評価点は、コンピュータセンターのお勧めを信頼して 80 点とするのが妥当と思われる。少なくともコンピュータセンターは、学生 i のページを良いページだと判断している。経路 x を経由して学生 i のページにたどり着いた学生 a は、そのページを 12 点相当にしか判断できない。このとき、そのページに 80 点が付けられていることやコンピュータセンターからお勧めリンクが張られていることを知れば、ページの判断において利用できる。

このように、評価点として最大評価点を採用することは、なんらかの良い評価がされているポテンシャルを示すものとしてページ評価における要素として有効であると考えられる。

2.3 ページ評価の伝播

本方式の特徴の 1 つとして、評価点として最大評価点を採用することで、各ページがどのページを起点とするリンクによって評価が伝播したかが一意に決定されることがある。各ページがどのページの外部評価を

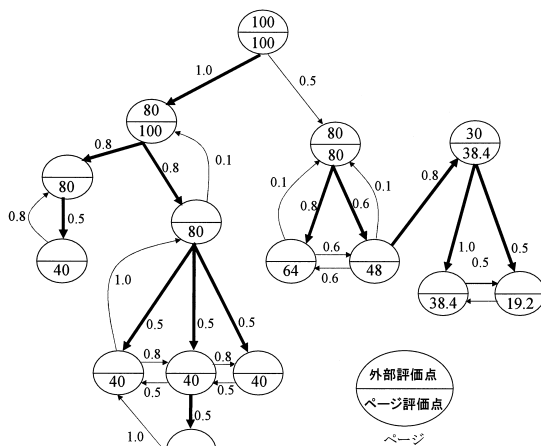


図 5 評価の伝播木
Fig. 5 Propagation tree of evaluation.

起点として評価が伝播したか決まることで、Web の持つグラフ構造から、評価伝播の木構造を抽出することができる。Web 構造に存在しているサイクル構造も、推薦度が $0 \leq w_{ij} \leq 1$ であるため、サイクルしても評価点は減少するだけで評価の伝播がサイクルすることはない。

図 5 にリンクによる評価点の決定と、その評価の伝播の例を示す。

図 5 において外部評価点がないページは、外部評価がされていないか 0 点であるかどちらかである。

本来の Web のグラフ構造から、評価点に関するリンクだけを抜き出し（太線）、それ以外のリンクを消す（灰色）ことにより、どのページを起点として評価が伝播したかが木構造となって現れている。この木構造により、あるページがどのページを起点として推薦されているか、誰が推薦しているのかが明確に示される。

このような Web におけるページ間の関係、あるいはページ作者間の関係は、評価点の基準となる外部評価の視点によって様々な解釈が可能である。

また、自分の作成したページに対して評価の伝播を確認することで、Web 読者の視点から、どのページを推薦したことになっているか、誰に推薦してもらっているかを確認できる。Web における自ページの評価関係の確認は、ページの自己管理にも有効と考えられる。

3. ページ評価点計算方法

前章で示したページ評価モデルにより、Web に対して適用を行うための計算方法を示す。

3.1 評価点の導出

V_j をノード n_j に到達可能なすべてのノードの集合, X_j を V_j に属するノードを基点としノード n_j に到達するすべての経路の集合とする. ノード n_i への評価点の集合 Z_i を求めるには, X_j に属するすべての経路に対する評価点 (すなわち, 起点となるノード $n_i \in V_j$ の外部評価 q_i に対して経路上にあるすべてのリンク e_{ab} 上の推薦度 w_{ab} を積算した値) の中から最大値となる評価点を見つけなければならない. 経路集合 X_j には, サイクルが含まれていることもあり, その要素数は相当数にのぼる可能性があり, 式 (4) を直接計算することは困難である. そこで本章では, 式 (4) を変形しページ n_j の評価点 ps_j を求めるためのアルゴリズムを導出する.

今, ノード n_a がノード n_j に対するリンクを持つ, つまり e_{aj} が存在するものとする. X_j の要素のうち, 経路に枝 e_{aj} を含む経路の集合を Y_j^a とする.

Y_j^a のうちノード n_j に対して最も高い評価を行う (起点ノード, 経路) の対は, n_a に対して最も高い評価を行う対と同じものである. なぜなら, 各対の n_j に対する評価は, 各対の n_a に対する評価に w_{aj} を積算したものであるからである. ここで, この対には, n_a 自身を起点とする対も含まれる. すなわち,

$$\max(s_j(A, x) | x \in Y_j^a) = ps_a \cdot w_{aj} \quad (5)$$

(ただし, A は V_a に属するすべてのノード)

と表すことができる. ここで, N'_j をノード n_j に対して直接リンクしているノードの集合とすると, 式 (4) は, 式 (5) を用いることにより, 隣接ノードの評価点 ps_a とそのノードからの推薦度 w_{aj} のみの関係式で表すことができ,

$$ps_j = \max(q_j, \{ps_a \cdot w_{aj} | n_a \in N'_j\}) \quad (6)$$

となる. この式を解けば, 各ノードの評価点 ps_i を求めることができる.

ここで, すべての w_{ij} において, $0 \leq w_{ij} \leq 1$ であったことに注目すれば, 次の性質が成り立つ.

性質 1 すべてのノードの中で最大の外部評価点 q を持つノードを n_i とすれば, $ps_i = q_i$ となる.

性質 2 すべてのノードに対する評価点 ps_i のうち, $1 \sim k$ 番目までに大きい値を持つノードの ps_i が求まったものとする. このとき, $k+1$ 番目に大きな値を持つノードは, すでに求まっている $1 \sim k$ 番目のノードのいずれかから直接リンクされているか, あるいは $1 \sim k$ 番目以外のノードの中で最大の外部評価点 q_j を持つノードである.

3.2 アルゴリズム

以上から, 各ページの評価点計算のためのアルゴリ

ズムを求めた. 本アルゴリズムでは, 各ページに対し外部評価を利用してページの初期値を設定し, まず最大値を持つノードを決定し, さらに次に大きい値を持つノードを決定するという繰返しにより, 高い評価点を持つノードから順番に評価点が決定していく.

また, このアルゴリズムによりすべての評価点は, ページ数を n , リンク数を m としたとき, $O(n \log(n+m))$ で計算可能となる. 以下にアルゴリズムを示す.

すべての ps_j は以下のアルゴリズムによって得ることができる. ここでは, pc_j を評価の現在値を入れるもの, $label_j$ はノード n_j の現在の状態, $cnode$ は現在注目するノードをそれぞれ示すものであるとする.

[手順 1]: [初期化]

すべてのノードに対して, $pc_j = q_j$, $label_j = 0$ とする. ここで, $i = 0$, $cnode = 0$ とし, 手順 2 に進む.

[手順 2]: [最大の評価点を持つページへの注釈]

$label$ の値が 1 でないすべてのノード n_k に対して, 最大の pc_k を持つノード n_k を探す [同値の場合, 任意のノードを選択する].

$cnode = k$, $ps_k = pc_k$, $label_k = 1$, $i = i + 1$ と置く.

$i > n$ ならば手順 4 に進み, それ以外なら手順 3 に進む.

[手順 3]: [注目ノードからの評価点を計算する]

$cnode$ からのリンクを持ち, $label$ の値が 1 でないすべてのノードに対して, $pc_j = \max(pc_j, pc_{cnode} \times w_{cnode_j})$ とする. 手順 2 に進む.

[手順 4]: 終了 [すべての ps_k の値が計算された]

4. 提案方法の適用

提案方式により求まるページ評価は, 外部評価として恣意的に決定される値に影響される. つまり, 誰がどのように決めた外部評価を利用したのかによって, 各ページの評価点が異なり, 同時に評価点の持つ意味も異なってくる. 今回は, 神戸大学国際文化学部の Web サイトを対象とし, 実験用 WWW サーバに神戸大学の国際文化学部のミラーを作成した. 本サイトの総ページ数は 3,053, リンク数は 7,734 である.

4.1 外部評価点の作成

大学のサイトでは, 組織としての公的な情報から, 教員の公的/私的な情報, さらに学生が作ったページまで, 様々な作成者によるページが混在する. Web 閲覧者が大学サイト内のページを評価する手段の 1 つとして, ページを作成した組織や部署, 個人などの地位や立場などを判別して, それに基づいて評価を行う方

表 1 外部評価の種別と値

Table 1 Classification of external valuation and values.

種別	内容	外部評価点
faculty	学部当局	100
division	学科	60
relevant	関連研究	20
committee	学部内委員会	80
teacher	教員	50
student	学生	30
admin	サイト管理部署	70
project	学部内プロジェクト	80

法がある。

たとえば、ある大学のサイトにおいて学内のコンピュータシステムについてのマニュアルを、

- システムを管理するコンピュータセンターが
- 教官が授業での使用に
- 学生が個人で

それぞれ作成したとする。

これらのマニュアルのうち、まずコンピュータセンターのマニュアルを読むのがよいだらうと判断すると思われる。どのマニュアルが最も分かりやすいのか、使いやすいのかは分からないが、マニュアルについて最も責任を持つべき部署（あるいは個人）はコンピュータセンターであり、最も公式な内容だと考えるからである。

そこで、サイト内に対する外部評価の基準として、公式度を採用し、表 1 のような種別と評価点を設定した。表 1 においては、学部自身が管理するページを最高点（100 点）としている。つまり、各ページは、そのページ内容を管理する組織・あるいは個人（以下では、これをページ作成者と呼ぶ）に応じて外部評価点が付与される。

実際の Web のファイルシステムにおいては、たいいていの場合、各ページ作成者の管理するページは、ディレクトリ単位で管理されていることが多く、作成者名、作成者の種別、作成者の管理するページ群のトップページの URL（通常は index.html など）を指定するテーブルを準備しておけば、すべての Web ページに対して外部評価点を付与することができる。また、各作成者が新規にページを作成したり、ページを削除したりした場合にも、自動的に外部評価点が付与される。

4.2 リンクによる推薦度の指定

リンクによる推薦度を表すために、表 2 に示すようなリンク属性を用意し、これを各ページのリンク（a href）タグ）に埋めこんだ。ここでは、推薦度に応じて 6 種類の属性を用いた。本来同一の値になるようなページへのリンク（1 つの論文を章ごとに別ページにした

表 2 リンク属性一覧

Table 2 Link attributes and recommendation rates.

属性名	説明	推薦度
equivalent	章ごとに分割された文章など意味的に同一なページへのリンク	1.0
official	ページ自己評価において公式なページへのリンク	0.95
personal	ページ自己評価において非公式なページへのリンク	0.4
endorse	他作成者ページに対する推薦を意味するリンク	0.8
introduce	他作成者ページに対する単なる紹介を意味するリンク	0.3
ignore	「戻る」などのナビゲーションだけのリンク	0.0

場合など)には、equivalent を「戻る」「back」などの推薦とは直接関係のないリンクには、ignore を付与した。また、一般的なリンクに関しては、同一作成者内のリンクか他の作成者へのリンクを区別し、それぞれ推薦度の高さに応じて、official と personal, endorse と introduce, の各 2 種類の属性を準備した。実験における属性付け作業においては、「back」や「戻る」などと書かれたリンクは ignore とする、「同一作成者内のリンクのデフォルトは official とする」「他作成者へのリンクのデフォルトは endorse とする」などといった自動化処理を図ることにより作業を軽減した。

4.3 実験結果

実験では、本提案方式、各ページ作成者の外部評価のみによる評価、PageRank⁴⁾の 3 つ方法により各ページの評価点を求めた。PageRank による評価点は、本提案方法の合計点と同じになるように正規化を行っている。なお、random surfer model の値は 0.15 として計算した。

表 3 に、いくつかの代表的なページを示す。この表では各ページについて、提案方法による評価点、PageRank による評価点、ページの外部評価点、提案方法の評価点の決定される際の評価点の伝播を示している。表中の (1) ~ (4) は「林」で検索を行った場合に検出されるページである。本サイト内には、教員である林が作成したページ（表中の (1) と (2)）、学生の林が作成したページ（同 (3)）、ならびに、学部委員会が作成した教員紹介の中に林教授のページ（同 (4)）が存在する。(1) は林教授のトップページであり、講座のページからのみリンクされているが、リンク属性は ignore（自分の属する組織からの単純なリンクの評価は外部評価点に折込み済みのため無視することにしたため）となっており、外部評価点である 50 点が提案方式の評価点となっている。また、(3)、(4) も同様にそれぞれ

表 3 ページ評価の例
Table 3 Scores of characteristic pages.

ページ内容	提案方法	Page Rank	外部評価点 (q_i)	評価伝播
(1) 林教授のトップページ	50	0	50	50
(2) 林教授のプロフィールページ	64	4	50	80→64
(3) 林という学生のページ	30	29.2	30	30
(4) 教員情報 林教授	80	600	80	80
(5) 学生 A の作成したレポートのページ	60.8	7.2	30	80 → 64 → 60.8
(6) 学生 B の作成したページ	30	85.5	30	30

れの外部評価点そのまま評価点となっている。(2)は、林教授の業績の自己紹介ページである。このページは、(4)のページから「最新の詳しい業績」としてリンクが張られている。つまり、このページは、学務委員会(評価点80)が公式度の高いページとして推薦してはいるものの、コンテンツ自体を管理しているわけではない。このため、提案方法による評価点は、(4) > (2) > (1) > (3) という順になっているが、公式度という観点からは妥当であると思える。

これに対し、PageRankの結果は、作成者に対する評価とは関係なく、一部のページで突出した点を持つページが存在している。これらのページは、多くの被リンクを持つものであり、PageRankの持つ特性を反映したものといえる。なお、ページ(1)のPageRankの値は正確には0点ではないが、被リンクが1つしかなく、小数点以下3桁で丸めたため0点となっている。

ページ(5)は学生Aが作成したページであり、あるプロジェクトのレポートが書かれている。このページは学生Aの個人ページ内に置かれているが、プロジェクトの公式レポートの1つであり、プロジェクトページからendorseリンクが張られている。プロジェクトのページからのendorseリンクにより学生Aのプロジェクトに関するページが評価されることで、プロジェクトの外部評価は80点が影響し60.8点となる。このようにリンクによる高い外部評価を持つページからの評価伝播が発生している。ただし、学生Aのページのうち高い評価点を得ているのは、プロジェクトに関するページだけであり、それ以外のページでは学生への外部評価である30点となっている。これは、リンクによる評価伝播では「戻る」などのリンクはignore属性として判断されるため、評価伝播は発生しないことによる。一般に同一作成者の作成したペー

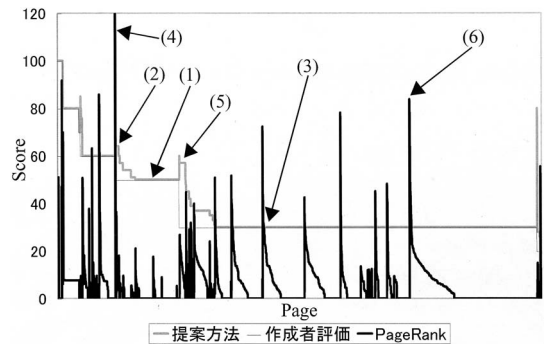


図 6 各方式の評価点

Fig. 6 Result of page evaluation for each methods.

ジでも「戻る」などのリンクを取り除くと関連性の深いページへのリンクしか張っていないことが多く、関連性の低いページにまで評価点が上がってしまう可能性は低いと考えられる。

ページ(6)は学生Bの作成したページであり、PageRankにおける特徴的な評価を示している。このページは多数のページからのリンクされているためPageRankにおいて高い評価点を与えられている。このリンクの多くは学生B自身のページからのリンクが多くを占めているがPageRankではこのようなページの評価点も上がってしまう。提案方法では外部評価である30点のみが与えられている。

図6に各評価方法による評価点の全体分布と、表3の各ページの位置を示す。この図では、ページ作成者への外部評価点、提案方法による評価点、ページ番号の優先度によりページを並べ替え、 x 軸としている。

提案方法による評価点は、ほぼ作成者評価と一致しているが、一部のページで作成者評価を上回っている。評価点が上がっているページは、高い外部評価を持つ作成者により作成されたページからリンクされたページであり、我々の想定する公式度として妥当性を持つと考えられる。

本提案方法では、基本的には作成者評価に準じた評価点を持ちつつ、ページ(2)やページ(5)のように評価の高い作成者からリンクをされることで個人ページ内のページでありながら高い評価を得ることができる。

上の実験では、各ページの評価点は、最低でもその作成者の外部評価点が付与された。しかし、同じ教員が作成したページにおいても、プロフィールや研究業績、趣味のページとでは、公式さが異なるはずである。

そこで、作成者に対する外部評価点を各ページ作成者の管理ページ群のトップページ(インデックスページ)に対してのみ与え、その他のページに対する外部評価点を0点として実験を行った。すなわち、

表 4 自己管理付き評価の例
Table 4 Scores of pages with self-management.

ページ内容	自己管理付き評価	一律外部評価	外部評価点 (q_i)	作成者評価
(1) 林教授のトップページ	50	50	50	50
(2) 林教授のプロフィールページ	60.8	64	0	50
(3) 林という学生のページ	27	30	0	30
(4) 教員情報 林教授	76	80	0	80
(5) 学生 A の作成したレポートのページ	57.8	60.8	0	30
(6) 学生 B の作成したページ	30	30	30	30
(7) 学部関連組織の紹介	30	100	0	100
(8) 学生の趣味のページ	12	30	0	30
(9) 終了したシンポジウムのページ	0	100	0	100

$$q_i = \begin{cases} \text{作成者への外 (トップページ)} \\ \text{部評価点} & (7) \\ 0 & (\text{他ページ}) \end{cases}$$

とした。

外部評価点を各ページ作成者のトップページに対してのみ与えることで、ページ作成者が自己作成ページを推薦度付きリンクで自己評価を行う。つまり、外部評価は作成者に対して与えられる作成者評価として考えることができ、各ページ作成者がその評価の範囲内で自分の作成したページを評価することができる。これは、作成者による自己作成ページの自己管理と考えられる。

表 4 にこの実験でのいくつかの代表的なページを示す。表中の自己管理付き評価がこの実験での評価点である。一律外部評価の結果は表 3 における提案方法と同じ方法であり、作成者のページに対し一律に外部評価点を与えている。また、作成者評価とは作成者に対する外部評価点 (その作成者のトップページの外部評価点) を示すものであり、表 3 の外部評価と同じ値となっている。

ページ (1) ~ (6) は表 3 のページ (1) ~ (6) と同じページである。ページ (2), (3), (4), (5) においては、 q_i は 0 点であるが、トップページからのリンクのより official 属性の評価がされ評価点が付けられている。

ページ (7) は学部の関連組織の紹介ページであるため、作成者は学部当局であり作成者評価は 100 点である。しかし、紹介としての introduce 属性のリンクが

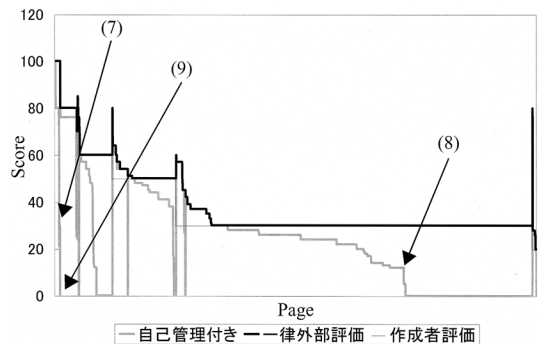


図 7 自己管理付き評価点

Fig. 7 Result of scores with self-management.

貼られているために評価は 30 点となっている。同様にページ (8) では、趣味のページに対して学生自身が personal 属性のリンクを貼ることで評価点を低く抑えることができる。ページ (9) は、学部主催のシンポジウムの案内が書かれたページである。シンポジウムの開催に先立って学部当局の管理領域内に作成され、トップページからリンクがされていた。その後シンポジウム終了とともにトップページからのリンクが削除されている。学部トップページからのリンクが削除されたため、いくつかの被リンクは残っているものの、提案方法による評価点は 0 点となっている。

図 7 にこの方法による各評価方法の実験結果を示す。この結果でも提案方法は作成者評価におおむね準じた結果となっているが、高い作成者評価を持つが低い評価を持つページや、0 点となるページが現れている。しかし図 6 とは異なり、高い外部評価を持つ作成者のページであっても点数の低いものや 0 点となっているページが存在する。これらのページはリンク切れや作成者によるページ自己管理の結果として現れたものである。

この結果から、外部評価点を式 (7) に従い、作成者トップページにのみ与え、その他のページを 0 点とした場合でも、公式度の高いページは高い評価点を得られ、リンク切れなどで有用性が低くなったページの評価点が下がっている。全体としては外部評価に従いつつ、ページの公式度をより反映した評価点となっていると考えられる。

図 8 では、太い線が評価の伝播木を表す。図には、プロジェクトトップページ、学生 A トップページ、学生 B トップページの異なる 3 つのルートを持つ木、伝播木 1 ~ 3 が存在している。また灰色の領域は各作成者の管理領域を示す。学生 B の作成したすべてのページは伝播木 2 に属し、学生 B 自身の管理領域内に収まっている。一方、学生 A のレポートのページは、学

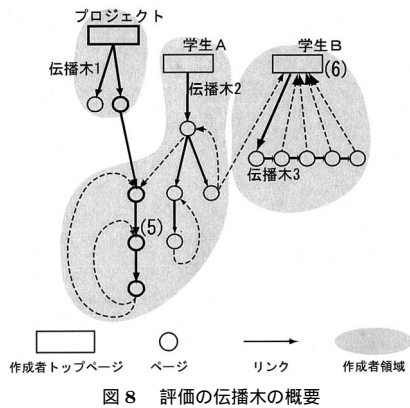


図8 評価の伝播木の概要

Fig. 8 A case of propagation tree in this experiment.

生 A の管理領域にありながら、プロジェクトをルートとする伝播木 1 に含まれている。つまり、学生 A のレポートは、公式さという視点からはプロジェクトのページであると見なすことができる。一方、学生 A の管理領域では、トップページをルートとする伝播木 2 に属するページと、伝播木 1 に属するページが存在している。つまり、同じ学生 A の作成しているページであるが、プロジェクトに属するページと学生 A に属するページかが伝播木によって区別することができる。このように評価の伝播木によって、外部評価点の高いどのような人が、そのページを推薦しているかを知ることで、ページに対する判断に利用できる。

提案方法では、高い外部評価を持つページからのリンクの存在に大きな影響を受ける。この性質を利用することでリンク有無のチェックにも利用することができる。作成者評価点と提案方法による点を比較することで、本来あるべきリンクのリンク切れや不要になったページの発見に役立つ。つまり、ページ作成者によるページの自己管理だけではなく、サイトとしてのページ管理にも有効であると考えられる。

以上のように提案方法による評価は、サイトとしての公式さの判断や作成者によるページ自己管理において有効な材料となると考えられる。

5. 議 論

本提案方式の利用の可能性と展望について議論を行う。

Web を閲覧する際に利用する場合、ページ評価の伝播木を可視化し、利用者へ提示することが考えられる。伝播木を参照しながらページを閲覧することにより、読んでいるページが誰によって推薦されているかを確認することができ、より安心して Web 閲覧を行うことができる。評価の伝播木を可視化することでサ

イトの構造やページ作成者を推測することができ、閲覧者の情報判断の手段として利用可能である。

組織のサイトに対しての適用については、大学において実験により示したように情報の信頼性の評価への利用が考えられる^{7),8)}。

大学のサイトでは、組織としての公的な情報から、教員の公的/私的な情報、さらに学生が作ったページまで、多様なページが混在している。このような組織では、同じサイト内にある情報であっても、その情報に対する組織としての責任や情報に対する信頼性が異なっている。提案方法を使うことで、ページに対する信頼性やページの自己管理への応用が期待できる。大学と同様に、NPO などの情報が混在している組織においても、サイト内の情報の管理に利用できると考えられる。

一方、企業などのサイトでは、現在ほとんどのサイトが公的な情報だけが書かれており、公式さの低い個人的な情報は書かれていないことが多い。公式度を外部評価とした提案方式は、このようなサイトに対しては有効ではない。しかし、提案方式では外部評価とリンク属性を利用することにより、サイトにおいて公式度の低いページ作成者を指定したり、リンク属性により公式度を低くしたりできる。つまり、ページ評価点が高いということを利用することで、組織として公式でないことを表明できる。本評価点を組織における公式度と適用することで、従来排除されがちだった自由度の高いページを企業サイト内に作成できるかもしれない。このように公式度としての評価点は、公式で有益な情報の判断だけではなく、自由度の高い柔らかい情報に対する判断にも応用可能である。

提案方法では、ある恣意的な視点から決定される外部評価を基準としたページ評価点が得られる。ただし、外部評価は恣意的な視点において決定できるが、リンクによる評価は推薦度を利用したため推薦という視点のみで評価されている。したがって、外部評価と推薦の組合せにより、評価点は外部評価の恣意性をそのまま反映したものではない。しかし、外部評価決定の視点として、何をを用いるかでまったく異なる意味を持つ評価点を得ることができる。

個人的な興味や趣味に基づいて外部評価を行うことで自分に適したページを発見するための利用や、検索エンジンの結果と組み合わせることで、興味ある話題についての関連ページ群を木構造として、引き出すこともできる可能性がある。また、格付け機関のような組織により外部評価が行われた場合には、Web 情報に対する信頼性の評価や Web ページの格付けにも利

用できると考えられる。

6. ま と め

本論文では、外部評価と推薦度を記述できるように拡張したリンクを用いた、新しいページ評価手法の提案を行った。本方法では、Web データから導出できない外部評価と、ページ作成者の意向を反映するリンク推薦度を導入することで、新しい評価点を導入している。また、本手法をサイト内に適用することで、ページ評価に有効な妥当な結果が得られた。

参 考 文 献

- 1) 佐藤光弘, 梶浦正浩, 多田智之, 池野篤司, 森口修: WWW における情報検索技術の動向, 電子情報通信学会誌, Vol.82, No.12, pp.1237-1242 (1999).
- 2) <http://www.google.com>.
- 3) Page, L., Brin, S., Motwani, R. and Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web (1998). <http://www-db.stanford.edu/backrub/pageranksub.ps>
- 4) Haveliwala, H.: Efficient Computation of PageRank, Stanford Technical Report (1999).
- 5) Kleinberg, J.M.: Authoritative Sources in a Hyperlinked Environment, *J. ACM*, Vol.46, No.5, pp.604-632. (1999).
- 6) Terveen, L. and Hill, W.: Finding and visualizing inter-site clan graphs, *Proc. CHI '98*, pp.448-455, ACM Press (1998).
- 7) 絹川達也, 平林真実, 清光英成, 大月一弘, 北村新三: 作成者責任の概念を用いた Web ページ評価法のサイトへの適用実験, 情報処理学会研究報告, Vol.2001, No.70, pp.281-288 (2001).
- 8) 平林真実, 堤 智也, 絹川達也, 大月一弘, 北村新三: 大規模サイトにおける WWW ページ評価手法の検討, 情報学シンポジウム, pp.75-82 (2001).

(平成 14 年 6 月 28 日受付)

(平成 14 年 10 月 5 日採録)

(担当編集委員 江口 浩二)



平林 真実 (正会員)

1964 年生。1987 年筑波大学第三学群情報学類卒業。1989 年同大学院理工学研究科修了。同年 (株) リコー入社。慶應義塾大学環境情報研究所研究員を経て、現在岐阜県立国際情報科学芸術アカデミー助教授。2000 年神戸大学大学院自然科学研究科社会人入学。Web を中心としたコミュニケーションシステム、表現システムに興味を持つ。ACM, 日本バーチャルリアリティ学会, ヒューマンインターフェイス学会各会員。



大月 一弘 (正会員)

1958 年生。1981 年京都大学工学部数理工学卒業。1986 年同大学院工学研究科博士課程単位取得退学。同年神戸大学教養部助手。現在、同大学国際文化学部教授。工学博士。情報通信システム, 柔構造データベース, 電子図書館システム, 教育工学等の研究に従事。電子情報通信学会, IEEE 会員。



清光 英成 (正会員)

1994 年図書館情報大学図書館情報学部卒業。1996 年同大学院図書館情報学研究科図書館情報学専攻修士課程修了。1998 年奈良先端科学技術大学院大学情報科学研究科博士後期課程中退。同年神戸大学経済学部助手。2001 年同大学国際文化学部講師, 2002 年同大学国際文化学部助教授, 現在に至る。データベースシステム研究, Web 情報システム研究, Web データの個別化・環境適応研究に従事。IEEE Computer Society, ACM 各会員。



森下 淳也(正会員)

1956年生．1979年神戸大学理学部卒業．1984年同大学大学院自然科学研究科博士課程修了．同年神戸大学総合情報処理センター助手．1998年姫路獨協大学外国語学部助教授．1997年神戸大学国際文化学部助教授．2001年同大学国際文化学部教授．現在に至る．学術博士．オブジェクト指向データベースの研究等に従事．米国物理学会，AAAS等の会員．



北村 新三(正会員)

1962年神戸大学工学部計測工学科卒業．東レ(株)勤務．1966年神戸大学大学院工学研究科(修士課程)修了．大阪大学工学部助手(電気工学科)．神戸大学助教授(計測工学科)を経て，1985年教授(現在，情報知能工学科)．1995年2月大学院自然科学研究科長，1997年2月工学部長併任(2000年3月まで)．2002年4月副学長．1998年度システム制御情報学会会長，ロボット制御，医療情報処理，ニューラルネットワーク，自律分散システム，創発システム等の研究に従事．



絹川 達也

2001年神戸大学国際文化学部コミュニケーション学科卒業．現在同大学大学院総合人間科学研究科博士前期課程在籍．Web-DB関連に興味を持つ．