

WWW ハイパーリンクの種類を考慮したコンテンツ評価の可能性¹

3 P-3

千種 健太郎* 西埜 覚** 苗村 憲司*

*慶應義塾大学 政策・メディア研究科

{chigu | naemura}@mag.keio.ac.jp

**通信・放送機構 横浜コンテンツリサーチセンター

1. はじめに

WWW ではハイパーリンク（以下、リンク）を用いることによって、様々なページやコンテンツ間を行き交うことができる。ページ間のリンクには様々な形態がある。リンク先としては HTML ページや画像、動画などのコンテンツがある。またリンクの見せ方としては現在見ているページを入れ替えて表示するもの、フレーム構造を持ち一部だけを入れ替えるもの、全く新しいウィンドウを開きそこに表示させるものなどが挙げられる。実際にこういった仕組みは、どの程度使われているのかを調べた。

また、ひとつのページからどのくらいのリンクが張られているのか、そこには何かの傾向があるのかなど、リンクについての実態を調べた。その傾向を元に、リンクで結ばれたページ同士に関連性については、コンテンツを評価する際に利用可能かどうかを検証する。

ここでいう「リンク」は、HTML ファイル上で "a href" で示されていて主に WWW 上ページ自体を指し、"img src" などで呼び出されている画像等は数に入れていない。

2. リンクの数

サーチエンジン「こねっと goo」²で、「月食」「徳川家康」「ユーゴスラビア」「源氏物語」という異なる 4 つのジャンルのキーワードで検索をした。その結果の上位 20 サイトについて、4 階層リンクをたどったページを対象とした。こねっと goo 及び上記のようなキーワードを用いたのは、ここでコンテンツ評価の対象としているものが、初等・中等教育に役立つサイトを想定しているためである。そのために初等・中等教育で使われそうなサーチエンジン、キーワードを用いた。

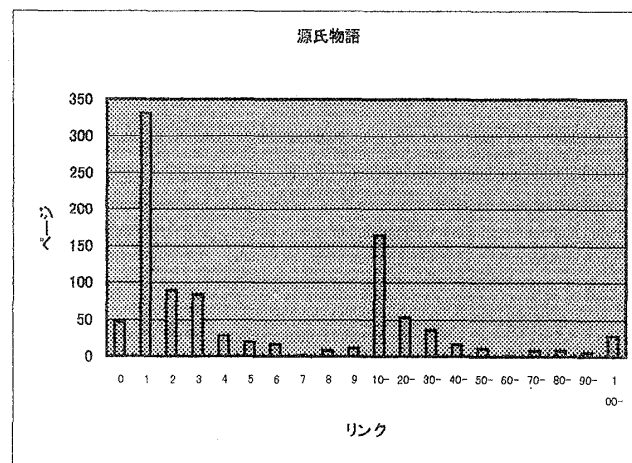
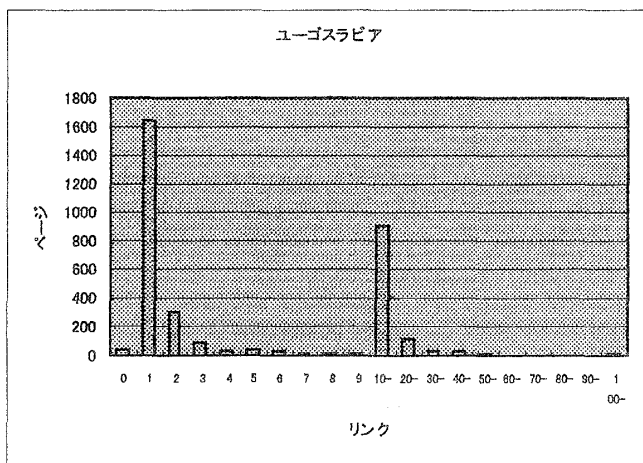
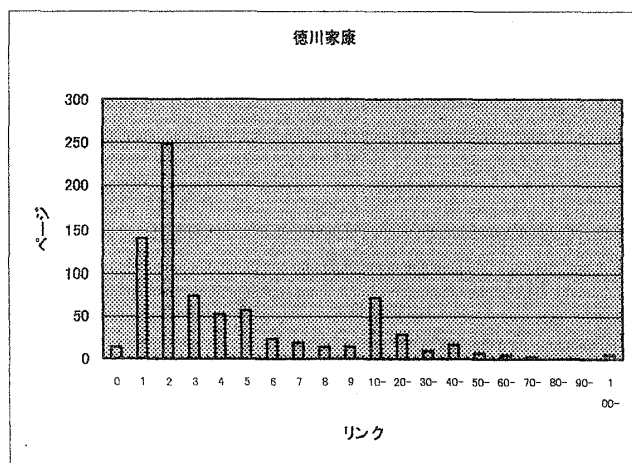
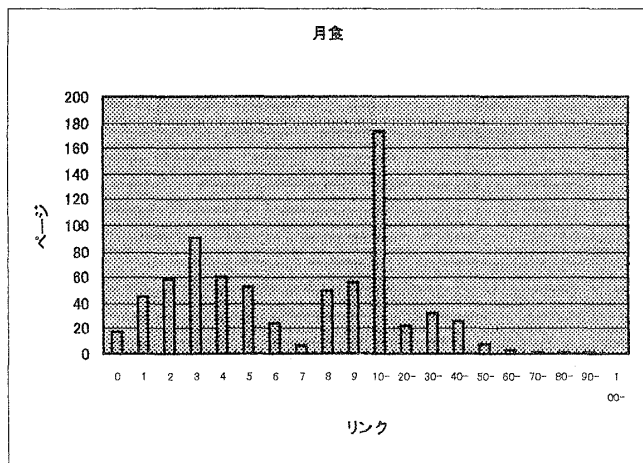
ここで得られたページについて、それぞれのページがどれだけのリンクを持っているか (<a href> タグの数) を調べた。その結果が次ページのグラフである。キーワードから得られたページの中のリンク数を調べ、横軸の数字はリンクの数を表し、縦軸の数字はそれだけリンクがあったページ数を表している。縦軸に "10-" とあるのはリンクの数が 10 以上 20 未満であったことを表す。リンクの数は 1 ~ 10 までは 1 刻みで、そこから 100 までは 10 刻み、残りは 100 以上となっている。これらを見た結果、キーワードによってばらつきはあるものの、リンク数は 2 前後のものが多いということがわかった。また、もうひとつの注目点として、リンク数が 10-19 のところがどのキーワードについても周辺よりも大きくなっている。これは、個人のちょっとしたリンク集を作る場合にこのくらいのリンク数になることが多いことが推測される。

¹ Types of WWW hyperlinks and their applicability to content rating. Kentarou Chigusa*, Satoru Nishino**, Kenji Naemura*

*Graduate School of Media and Governance Keio University

**Telecommunications Advancement Organization of Japan

² <http://www.goo.wnn.or.jp/>



以下は対象としたキーワードと、リンクの平均値の表である。

キーワード	ページ数	1ページ当たりのリンク数
月食	726	10.44
徳川家康	798	7.82
ユーゴスラビア	3302	7.52
源氏物語	969	15.17

3. コンテンツ評価への応用

任意のページ A が、教育に有益であると判断された場合、そこからリンクをされているページ B も同じように有益である可能性が高いと考えられる。この原理を用いると全てのページを人間が見る前に、暫定的な値としてページを評価することができる。現時点で求められているのはあるページからのリンクの数である。それに加えページ A からページ B にリンクが張られていて、かつ、ページ B からページ A にリンクが張られている、相互リンクの状態を調べる必要がある。そうした上で、全部のリンクから戻っていくリンクを引き、純粋に「次」のページを評価することが可能となる。また、有益な任意のページから出発した場合にどの程度の深さまで有益であると言えるかを検証する。

このように、ハイパーリンクの性質や数を調べることにより、将来的にコンテンツの半自動評価にやくだてることができる。

参考文献

[1] 西埜、満澤、苗村：“WWW 上の有害な情報を効率的に検出する一手順”，電子情報通信学会 1999 年総合大会論文集, 情報・システム 1, P.278(1999)