

西埜 覚* 西塚 一樹* 苗村 憲司**

通信・放送機構 横浜コンテンツリサーチセンター* 慶應義塾大学環境情報学部**

1. はじめに

インターネットで公開される WWW ページを受信側で表示を制御する方法として、W3C(WWW コンソーシアム)は PICS(Platform for Internet Content Selection)ルール[1]を勧告している。このルールに基づくレイティングシステムとして RSACi(Recreational Software Advisory Council on the Internet)が知られている。RSACi は 4 カテゴリ・5 レベルのレイティングを行っており、コンテンツ提供者が HTML 構文に RSACi のレベルを記述すれば、一般に用いられているブラウザでフィルタリングの実現が可能である。しかし、最近 RSACi の課題(セルフレイティングの客観性、受信者の利用目的に無関係なレイティングの固定性等)が明らかになってきた。これらの課題の解決には、Yale 大バルキン教授らが提案する「3 層ケーキモデル(以下、三層モデル)」[2]の実現が期待されている。このモデルは、「コンテンツ提供者によるページを特徴づける基本語句(30~60 個程度の用語)の指定(第一層)。第三者による複数のレイティングテンプレートの提供(第二層)。レッドリスト、グリーンリストによるページ表示の制御(第三層)」によってレイティング、フィルタリングを実現する。このモデルの特徴は、個別のページにはレイティング情報が付けられておらず、第一層で指定された基本語句を第二層で提供されるテンプレートで評価してレイティングを可能にしたことである。多様なテンプレートを用意することで利用目的にあったフィルタリングが可能になる。これによって、RSACi が抱えるいくつかの課題を解決することが期待されている。三層モデルは、レイティング、フィルタリングシステムの国際標準化を目指している ICRA(Internet Content Rating Association)[3]が RSACi に代わるシステムとして注目し、その適用の検討を進め、近い将来の採用を目指している。

2. 三層モデル適用への公開ページの課題

三層モデルでは、RSACi の META タグ PICS ラベルに代わる新しい META タグがページの HTML 構文に記述され、そこにコンテンツ提供者がページの特徴を示す基本語句を選んで埋め込む。基本語句を埋め込むためのページ作成ツールが提供され、このツールを用いることで簡単に語句の埋め込みが可能になる。しかし、ツールを用いずに作成したページや意図的に基本語句の埋め込みを行わないページも多数公開されることが予想される。このような場合、レイティングテンプレートに与える基本語句が存在せず、レイティングが決定出来ずフィルタリングシステムが働かないという問題が発生する。

WWW ページのテキスト処理を行ない、有害な情報(=学校の授業を妨げる情報)を効率的に検出する研究開発を行ってきた[4]。この有害な情報の検出に適用した方式を基本語句が埋め込まれていないページに適用して、基本語句を決定・生成し、それをテンプレートで評価して PICS ルールに基づくページのアクセス制御を実行する方式を提案する。

3. ページ(URL)と基本語句を連携する DB

ページの基本語句を決定し、URL との関連付けを行なうデータベース(DB)を作成する。

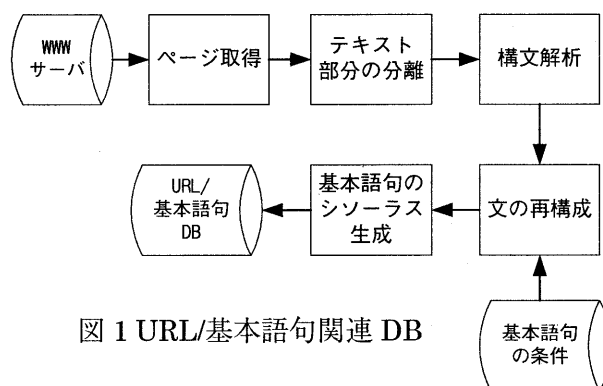


図 1 URL/基本語句関連 DB

¹ A proposal of approaching to the method for the international rating standard of contents on WWW, Satoru NISHINO, Kaduki NISHIDUKA, Telecommunications Advancement Organization of Japan, and Kenji NAEMURA, Keio University

- ①収集した個別のページについて、テキスト記述の部分で構文解析を実行する。
- ②ページの特徴を抽出するために、構文解析で得られた結果を単語、単語の組み合わせや短い文章に再構成する。
- ③この再構成の時、単語等の出現頻度や、テキストが記述されていたHTML構文(例えば、Title文、Keywords 定義等)要素に基本語句の条件を与え、ページを特徴づける基本語句を決定する。
- ④ページの特徴づけの適用範囲を拡大するために基本語句と関連語のシソーラスを生成する。
- ⑤個別のページにテキストの記述がない(例えば、サムネイル画像やバナー画像でリンク集を構成するページ等)場合は、リンク先のページにテキスト記述が存在すればリンク先ページで決定した基本語句を適用する。
- ⑥META タグに基本語句が埋め込まれている場合は、その語句を適用する。
- ⑦上述した方法で基本語句を決定し、ページのURLをキーとするDB(URL/基本語句DBという)を作成する。

4. テンプレートの適用と PICS ラベル生成

一般に用いられているパソコン用ブラウザには PICS ルールに基づくフィルタリング機能が用意されている。これに PICS ラベルを与えれば、受信者が事前に設定したレイティングレベルでフィルタリング機能が働き、ページの表示、表示ブロックが可能である。前述した URL/基本語句DBには、PICS ラベルが存在しないので、基本語句をテンプレートで評価して PICS ラベルを生成して与えなければならない。

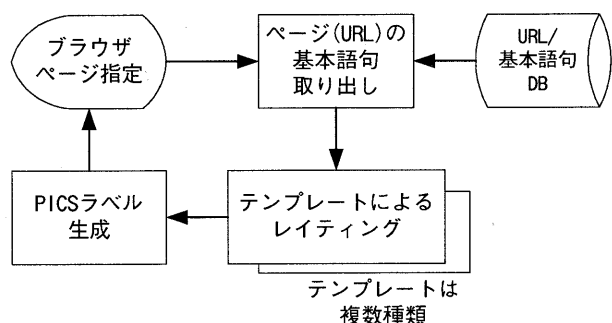


図2 PICS ラベルの生成

- ①META タグに基本語句が埋め込まれていないページをブラウザ利用者が参照しようとしたとき、ページのURLをキーにして、URL/基本語句

- DBから該当するページの基本語句を取り出す。
- ②取り出した基本語句を利用者の目的に合ったレイティングテンプレート(テンプレートは複数種類が用意される)で評価し、レイティングレベルを決定し、PICS ラベルを生成する。このとき、基本語句の取り出しとテンプレートによるレベル決定は一体化した処理、または分離処理のどちらでも構わない。
- ③生成した PICS ラベルをブラウザに渡す。ブラウザは PICS ラベルに含まれるレイティングレベルでページの表示、表示ブロックを制御する。
- ④なお、参照したいページの META タグに基本語句を埋め込まれていれば、その基本語句がレイティングテンプレートによって直接処理される。

5. まとめ

本提案は三層モデル、ICRA の動向等から得た情報を元に、その適用の検討途中をまとめたものである。

三層モデルの実現は、コンテンツ提供者には表現の自由が確保される。また、利用目的にあった多様なテンプレートを準備することで、ページ参照者には目的にあった情報の選択・参照が可能になる。これによって、インターネットの健全な発展に寄与する期待が大きい。

一方、三層モデルに関して、ICRA の検討は継続されており、未だ決定していない事項も多い。また、RSACi 等のように既に長年利用されてきたレイティング、フィルタリングシステムとの継続性や互換性を保証する等、解決すべき課題が残っている。

ページのテキスト記述から基本語句を決定することは、既に試みた有害な情報の候補を検出する方式を適用出来る部分も多い。本提案方式の可能性、有効性について、プログラムを作成して試みたいと考えている。

参考文献

- [1] <http://www.w3.org/PICS/>
- [2] http://webserver.law.yale.edu/infosociety/filtering_report.html
- [3] <http://www.icra.org/>
- [4] 西埜、他：“WWW の教育用格付けの効率化技術に関する検討”, 情処学会 99 年後期全国大会, 3X-04(1999)