



現リストのすべてである。そのため、具体的には次の検索式の演算を行う。

$W_p \text{ NEAR } (W_{S1} \text{ OR } W_{S2} \text{ OR } \dots \text{ OR } W_{SN})$

$W_p$ : 検索語  $\{W_s\}$  評価表現リスト

### 3. 3 優先順位の設定

優先順位の設定は、まず近接演算処理で抽出した評判情報に対して正規表現で記述された表1のような優先度判定ルールを適用する。

ID	優先度判定ルール
1	$\_KEY\_*(は が も)\*_REP\_$ (検索語と評価表現が「は」「が」「も」を挟んで存在する)
2	$\_KEY\_*(。 . ? !)\*_REP\_ REP\_*(。 . ? !)\*_KEY\_$ (検索語と評価表現が別の文にある)

表1 優先度判定ルールの例 (KEY\_は検索語、\_REP\_は評価表現を示す)

次に、総ルール数がnのときn次元の配列を用意し、n個のルールを満たすか満たさないかを1/0の値でセットする。例えば、総ルール数が4つで評判情報がID1とID3のルールを満たす場合は(1,0,1,0)である。各配列のボタンにはスコアを対応付けてあり、このスコアが評判情報の優先度を示す。なお、各配列パタンのスコアはあらかじめ用意したサンプル評判情報に対して人間が付けたスコアを配列ボタン毎に平均値を取ることにより決定した。

### 3. 4 評判情報の分類

評判情報の分類は、肯定・否定別の分類とサイト別の分類がある。肯定・否定別の分類は、評価表現の近傍に否定表現が出現したか否かで判定する。否定表現とは、例えば「ない」「ません」「なかった」「不(接頭語)」のような、評価表現の意味を打ち消す役割をもつ表現である。サイト別の分類は、評判情報を含むHTMLページのURLを用いて、あらかじめ決められたサイト種に分類する。サイト種は、評判情報が良く出現するサイトの種類であって、例えば「ニュースサイト」「掲示板」「ショッピングサイト」が該当する。

### 4. 試作システム

図2に3節の検索方式に従い作成した評判情報検索システムの検索語「モバイルギア」商品カテゴリ「コンピュータ」に対する検索結果を示す。用意した評価

表現の数(コンピュータ関連)は100程度である。また、近接演算の距離は、検索語を基準として前方40Byte、後方80Byteとした。各ページの評判情報は優先順に並んでおり、そのスコアは文脈的に評判情報らしい程高いスコアを与え最高5点としている。また、分類結果として肯定的又は否定的かを示すアイコン、サイト別のアイコンをそれぞれ表示している。

この試作システムの適合率(指定した商品名に関する評判が検索される精度)を測定したところ、上位10件で90%、上位20件で75%であった(16個の検索語の平均値)。なお、単純に検索語を含むページ数は16個の検索語の合計で9390、その内評判情報を含むページに限定することで12%に絞り込まれた。



図2 検索結果画面

### 5. おわりに

本稿では、評判情報検索の機能及びそれを実現するための検索方式について述べた。本検索方式は、オンラインショッピングにおける商品購入支援や、インターネット上の情報モニタリングなどへの応用が考えられる。今後の課題は検索の高速化、評判の重要性に基いた優先度判定等である。

#### 参考文献

[1] 福島俊一 石黒義英 喜田弘司 山田洋志 松田勝志, "目的特化型サーチエンジンの開発", 情報処理学会第61回全国大会予稿集(3), p137-138, 2000

[2] Doorenbos, R., Etzioni, O., and Weld, D. "A scalable comparison-shopping agent for the World-Wide Web.", In Proceedings of the First International Conference on Autonomous Agents Agents'97(Marina del Rey, Calif., Feb. 5-8) ACM Press, N.Y., 1997, pp.39-48.