



#### 4. システム構成

以下の3個のパネル(マイニング処理と可視化 IF の組)によって構成する。Rを利用して単語ベクトル, ポジショニングマップを作成する SOM 処理モジュールパネル, ポジショニングマップを表示しユーザに単語集合を作成してもらいURLベクトルを作る MAP 表示モジュールパネル, Wekaを利用して決定木を作成し推薦単語を表示する決定木処理モジュールパネル。

処理の詳細を図2のシステム構成図の記号に合わせて以下に説明する。

- HTML ページ内の javascript によりスニペットを取得し, 統合環境の入力形式のテキストファイルを作成し, 保存する。
- 統合環境の制御処理部にテキストを入力することによって, 統合環境が ChaSen による形態素解析を行い, text オブジェクトを生成する。
- SOM 処理モジュールによって, text オブジェクトを利用して単語ベクトルと URL ベクトルを作成する。
- SOM 処理モジュールにより, R を実行してポジショニングマップを作成する。
- MAP 表示モジュールにより, ポジショニングマップを表示する。SOM 処理モジュールと Display モジュールにより, ポジショニングマップの各領域に割り当てられた単語の詳細を表示する。(図3参照)
- 決定木処理モジュールにより, ユーザがクラス分けを行い, URL の特徴ベクトルにクラス属性を追加する。
- 決定木処理モジュールにより, Weka を実行して決定木を作成する。
- 決定木処理モジュールにより, 決定木のデータを取得し, 結果を整理してユーザに提示する。

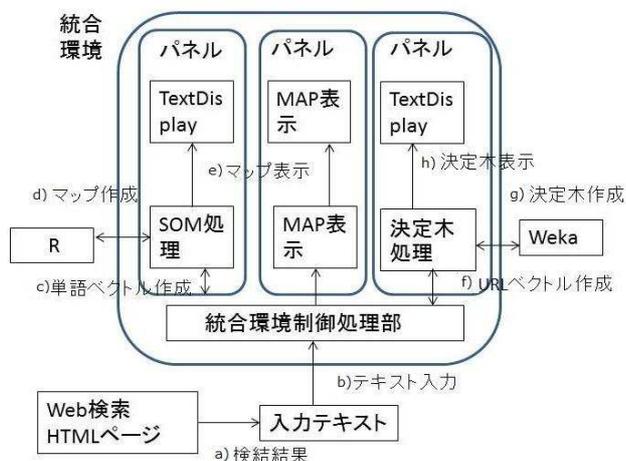


図2 システム構成図

#### 5. 実験結果

Web 検索支援システムの有用性を調べるために, Web 検索支援システムを使用した場合と, Google の予測変換・検索候補を利用した場合, またランキングサイト「Rerank.jp」[山本 2008]を利用した場合の, 目的のサイトにたどり着くまでの所要時間の比較実験を行った。ランキングとは, ユーザの興味に合わせて検索結果のサイト順序を再ランキングして表示するというものである。

「量子力学」から検索を開始して, 最終的に「統一場理論」について書かれているサイトにたどり着けるかどうかと, その過程についてまとめる。本検索支援システムでは(1)「現象, ファイル」

が推薦語として示され, (2)「量子力学, 現象」で検索し「数学, 研究所, 巨視的」が推薦語として示され, (3)「量子力学, 巨視的」で検索し「関数, 量子, 支配, 効果, 学問」が推薦語として示され, (4)「量子力学, 量子, 支配」で検索し「統一場理論」にたどりついた。かかった時間は14分である。Google 検索では, 予測変換や検索候補として出された「入門, とは, 発展, 応用, 理論」などを利用して, 各検索ワードに対して 10 ページずつ閲覧していった。その結果, システムを使用したときの時間を越えてもサイトを発見することができなかった。Rerank.jp では, 提示された「理解, 粒子, 世界の謎, 無限」などを上位に持っていく処理をしたが, キーワードがほぼ更新されず同じサイトばかり表示されるため, 途中で検索を断念した。

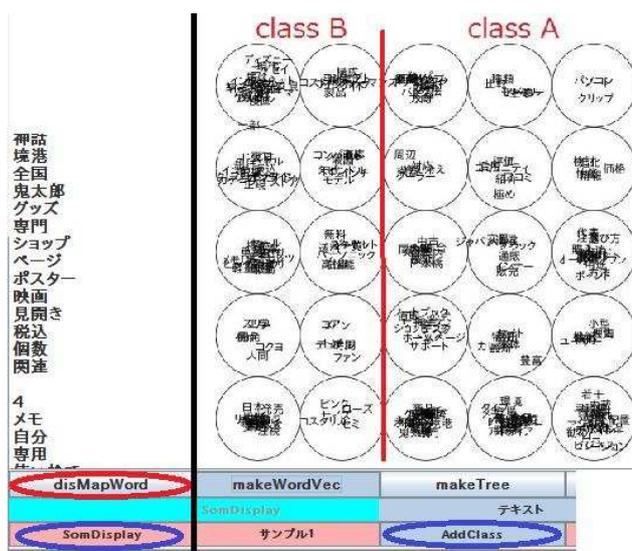


図3 ポジショニングマップの表示画面

#### 6. おわりに

検索結果のスニペットに存在する単語を自己組織化マップによるマップを提示し, ユーザに興味のある単語集合を選択してもらうことによって, 再検索に有効な単語を推薦するシステムを提案した。実験結果より, 検索対象の知識が乏しいユーザが, より専門的な情報にたどり着けることを確認できた。多くの人が興味をもつ情報は協調フィルタリングによって効率的に得ることができるが, 少数の者しか興味を持たない情報を検索するには本システムが有用であると考えられる。

#### 参考文献

[砂山 2011] 砂山 渡, 高間 康史, ダヌシカボレガラ, 西原 陽子, 徳永 秀和, 串間 宗夫, 松下 光範, : Total Environment for Text Data Mining, 人工知能学会論文誌, Vol.26, No.4,2011.

[豊田 2008] 豊田秀樹: データマイニング入門, 東京図書 2008.

[IAN 1999] IAN H.WITTEN,EIBE FRANK: Data Mining, MORGAN KAUFMANN PUBLISHERS ,1999.

[コホネン 1996] T.コホネン: 自己組織化マップ, シュプリンガー・フェアラーク東京, 1996.

[キンラン 1995] J.R.キンラン: AIによるデータ解析,トッパン,1995.

[山本 2008] 山本岳洋, 中村聡史, 田中克己: Rerank-By-Example: 編集操作の意図伝播によるウェブ検索結果のランキング, 情報処理学会論文誌 データベース Vol. 49, No. SIG 7,2008.