



文書クラスタリングの技法

F. 文書構造の視覚化のための技法

1. VIBE

2. GUIへのSOMの応用



VIBEとは？

- 文書集合を視覚的に提示するためのシステム
- 文書クラスタリングが主目的ではない
- 入力した主題概念に応じて結果的に文書がクラスタ化される



VIBEの動作(1/2)

- 利用者が3つの概念に関心をもったとする。
- 3つの概念をそれぞれ
“document retrieval”
“scientific visualization”
“virtual reality”
とする。
これらをそれぞれPOIと呼ぶ。



VIBEの動作(2/2)

- 利用者は各POIの位置をコンピュータの画面上に指定する。
- ある文書をこの画面上に座標として配置する場合、POIとの類似度によって座標が計算される。



VIBEのイメージ

document retrieval



document scientific



document retrieval scientific visualization

virtual reality



scientific visualization



virtual reality



GUIへのSOMへの応用

- SOMとは文書構造を視覚化するためのツールである。複数の属性からなる入力データを教師なしで自動的に分類できる。
- 特徴は長方形の出力領域に、主題ごとに長方形の区域を設定できる。
- Javaアプレットなどを使って各区域を色分けすれば「文書の地図」を表示することが可能。

SOMのイメージ

