コンピュータサイエンス

平成21年度　戦略的創造研究推進事業（さきがけ）採択研究

「圧縮データ索引に基づく巨大文書集合からの関連性マイニング」

知能情報工学科　准教授　坂本比呂志

巨大文書集合に内在する関連情報 → キーワード検索 → 圧縮データからの関連性マイニング

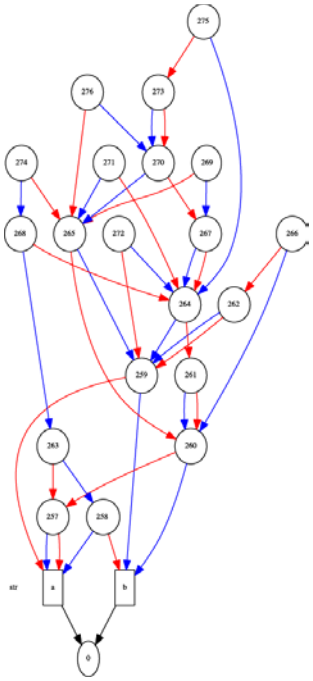
問題点

■**時間的・空間的困難**　巨大データへの索引付け

■**意味的困難**　キーワード抽出を超える検索の枠組み

情報を圧縮して検索

（テキストデータを数列に置き換え、情報をうまく圧縮）



社会への応用

冗長な部分を削ぎ落とすことで**重要情報を特定**。
埋もれた知識の発見を目指す。

大規模文書分類

- 特許データ
- 新聞記事
- 科学技術論文の関連性に基づくクラスタリング

超高速検索

- 遺伝子データ
- プログラムソース
- 特許データからの類似箇所検索

将来の展望

使われている言語を問わないため、
画像や動画、音声などのデータにも
応用も可能に。