

ウェブ検索ディレクトリの自動 構築とその改良 -鳥式改-

鳥澤健太郎、他10名

情報通信研究機構MASTERプロジェクト言語基盤グループ

北陸先端科学技術大学院大学
奈良先端大学院大学

発表者：伊藤輝将

はじめに

- ・様々なトピックに関する問題回避、あるいは行動に関する未知のアイデアについて情報を求めるため検索エンジンを利用する

↓しかし...

「意外」なキーワードを入力する必要がある
(こうしたキーワードはユーザの「意識に昇っていない」)

↓

システム側から提示する必要がある

鳥式改

- ・キーワードの想起を支援するための検索ディレクトリ
- ・ユーザが最初に入力したキーワードに対して、関連語を意外なものまで提示し、検索に利用できる

鳥式改の特徴(1)

- ①現在180万語という大量のトピックをカバーしている
 - Web文書に自然言語処理を適用することで自動生成されている

- ②いくつかの意味的カテゴリに属する関連語のみを提示する
 - 価値ある情報を効率よく検索できるようにするため

鳥式改の特徴(2)

③トピックの上位概念の名称を自動的に獲得し、
大量に保持している

-広範な関連語を提示するため

例:

トピック「東京大学」に対する「大学」

④単に関連語を提示するだけでなく、関連相互の関係性も提示できる

鳥式改の特徴(3)

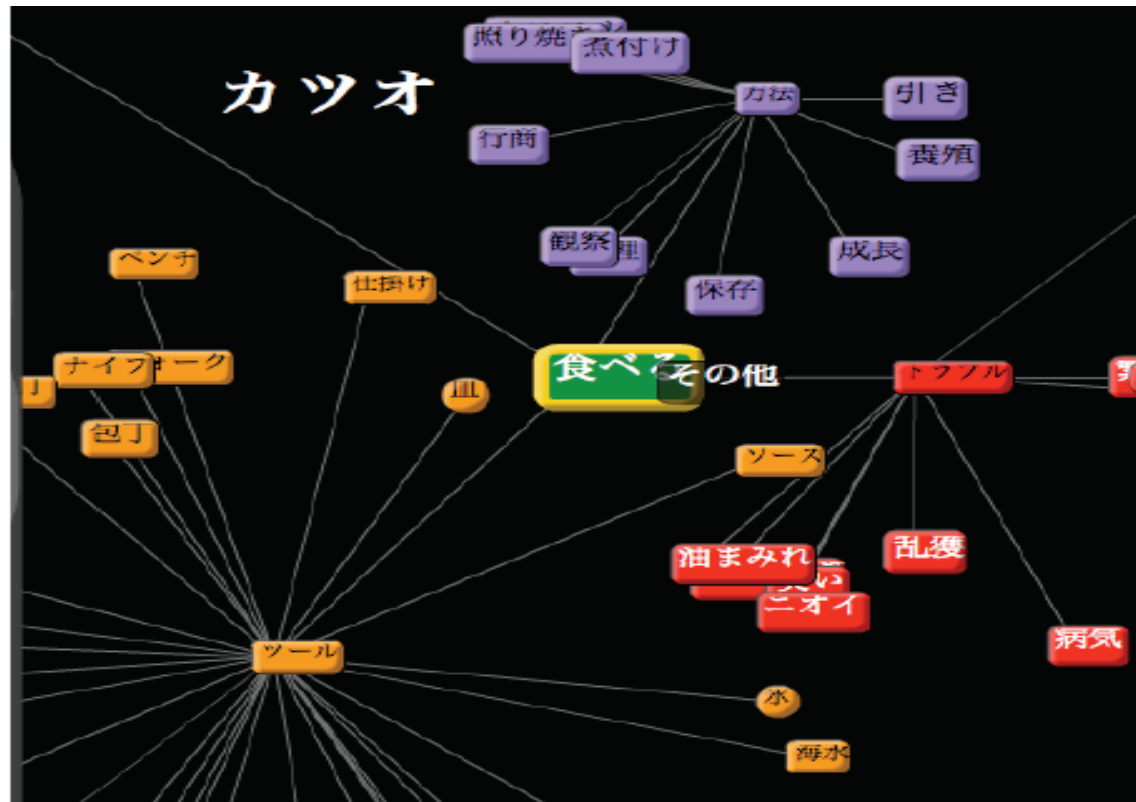


図-2:「カツオ」の「食べる」文脈での関連語

類似表現の関係とアナロジー(1)

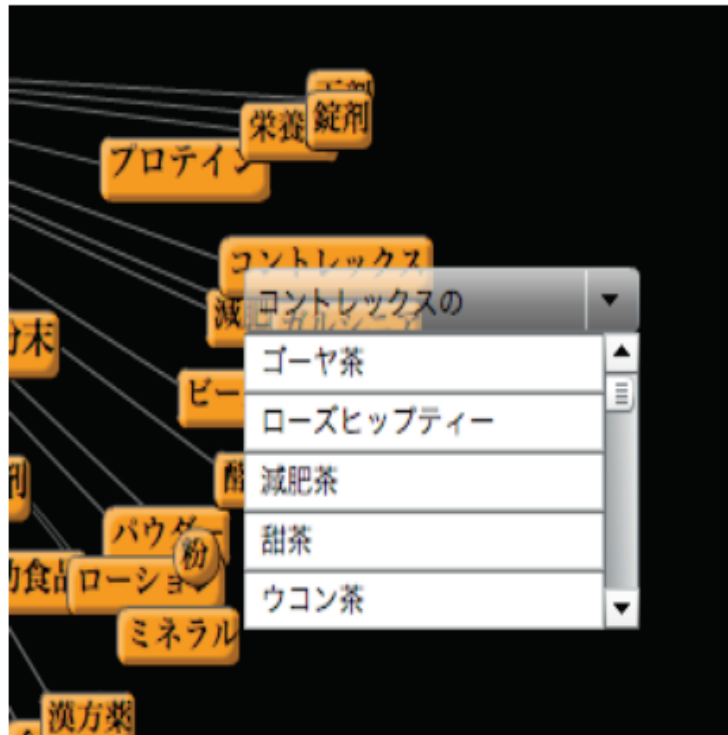


図-3: ダイエットのツールとしての「コントロールクス」の類義語を提示した例



図-4: 青酸カリの類似表現を鳥式改で提示した例

類似表現の関係とアナロジー(2)



図-5:ガスコンロ利用時の具体的方法とその類似表現

因果関係

- ・関連語に関係して、自動獲得した因果関係を表示することも可能



図-6:「うつ病」の原因の提示



図-7:「中毒」が引き起こす結果としての「死亡」

今後の展開

- より多様な意味的関係をWebより抽出して、鳥式改を巨大なネットワークとし、対話システムなどのバックエンドとして活用して、ユーザのおかれている文脈を考慮して適切なアドバイスをするといったシステムを開発することを考えている