

探索アクティビティの共有と視覚化 に基づく協調型情報探索

†神戸大学大学院自然科学研究科 ‡京都大学大学院情報学研究科

上田 正明† 中島 伸介‡ 角谷 和俊‡ 田中 克己‡

本研究の動機(1/2)

- グループにとって価値のある情報を探す

(例: グループ旅行のプラン選定 etc.)

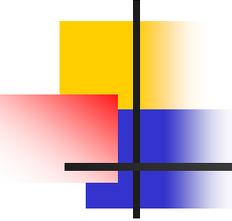


- グループの代表者が情報を探索する場合

- 全員の意見を反映させるのは困難
- 1人に負担がかかる
- 効率が良いとはいえない



では、グループ全員で情報を探索する場合はどうだろうか？



本研究の動機 (2/2)

- **グループ全員で情報を探索する場合**

- 各々の探索活動を明示的に表現できない
- お互いの探索活動の関係を把握できない

➡ **効率的な協調探索を行えない**

したがって、本研究では、

各ユーザの**探索アクティビティ**を取得・保存し、
共有することで、効果的な協調作業を実現する...

「協調型Web情報探索システム」を提案する

探索アクティビティとは

どのような**特徴**のWebページから

どれほどの**数**を

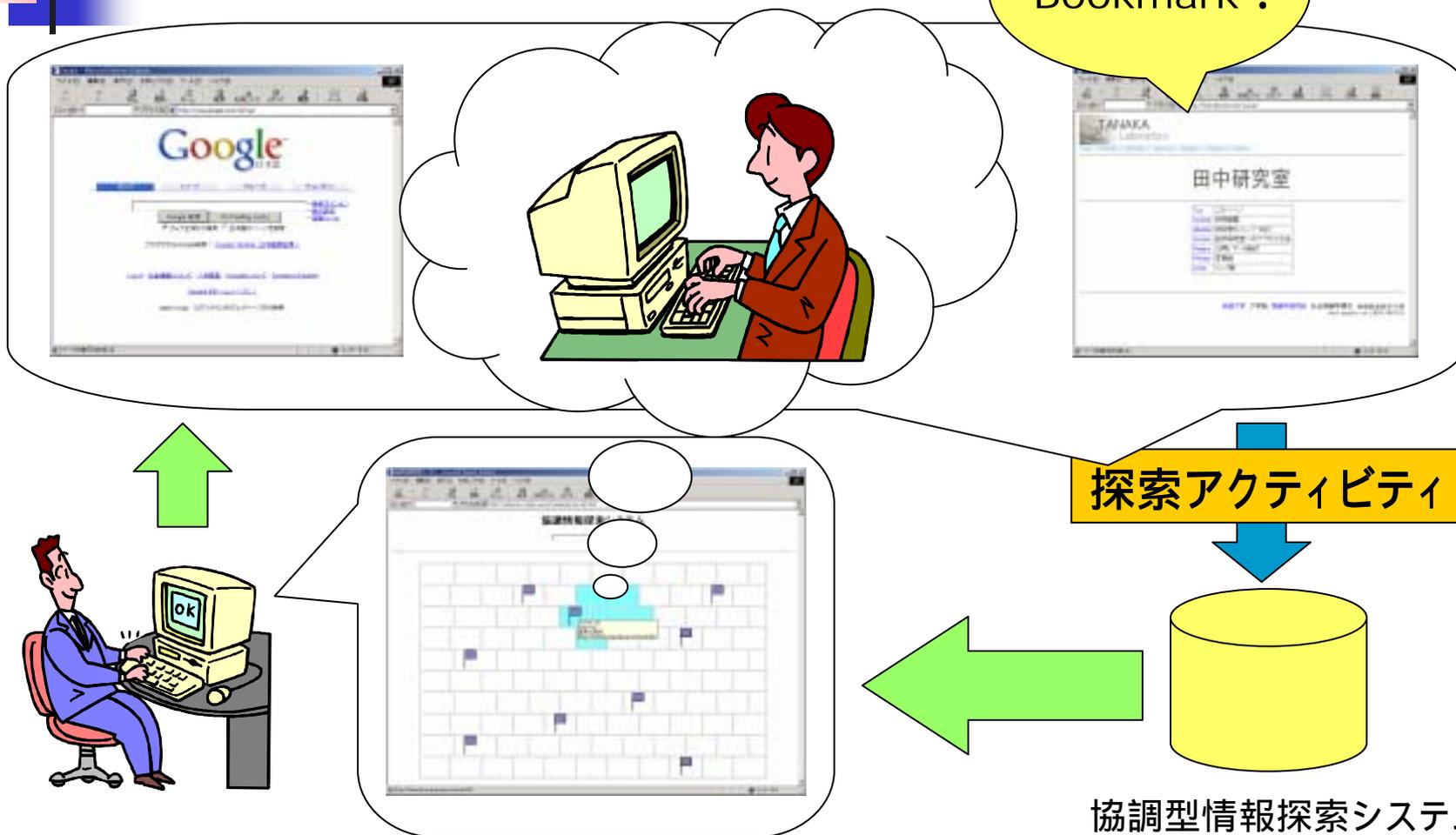
探索した結果**どれを選んだのか**



個々の探索活動の全体のことをいう

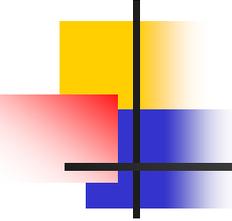
想定している探索手順

Bookmark !



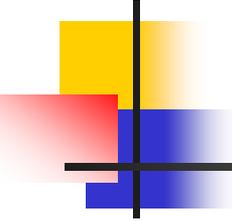
探索アクティビティ

協調型情報探索システム



アプローチ

- どのように探索アクティビティを保存するか？
 - ➡ コンテキストブックマークの利用
- どうすれば探索アクティビティを直観的に把握できるか？
 - ➡ 探索アクティビティの視覚化
- どうすれば他のメンバーのアクティビティを見られるか？
 - ➡ 探索アクティビティの共有
- どうすれば他のメンバーと意見交換ができるか？
 - ➡ 探索アクティビティの合成
 - ➡ 探索アクティビティの問合せ利用



コンテキストブックマーク

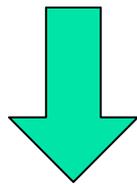
- 中島らが提案（第63回情報学会全国大会，DBWeb2001）
- 「ブックマークしたページ」と「それまでに閲覧したページ」
特徴の違いから，ユーザの意図を抽出
- 保持情報
 - 一般的なWebブックマークと同様の情報（URL，タイトル）
 - 閲覧履歴の各ページのURL
 - 代表的キーワード：閲覧履歴の各ページ内のtf・idf値の高い単語
 - ランク値：閲覧範囲と閲覧内容の類似度から算出

これにより探索アクティビティを保存

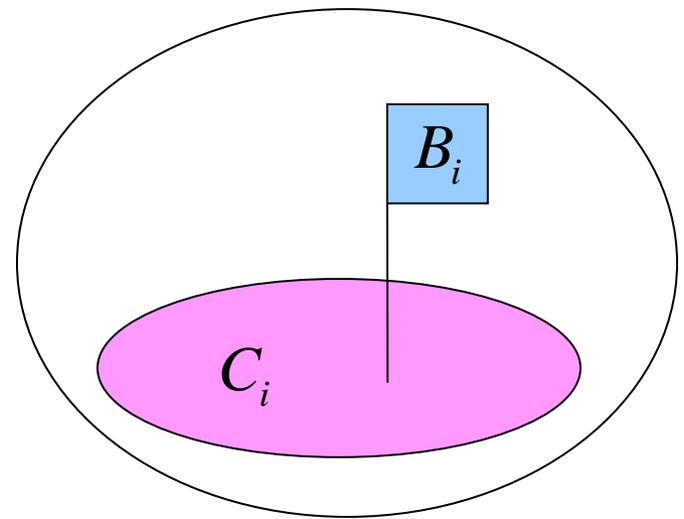
コンテキストブックマーク

- 探索アクティビティのイメージ
 - 探索アクティビティ A_i

コンテキスト集合 C_i



ブックマーク集合 B_i



探索アクティビティの視覚化と共有

■ 視覚化にあたって...

■ 各Webページの特徴ベクトル

- ページ内のテキストを取得
- 「茶筌」を用いて形態素解析
- 名詞1つを1次元とし、その出現頻度を値とする特徴ベクトルを生成

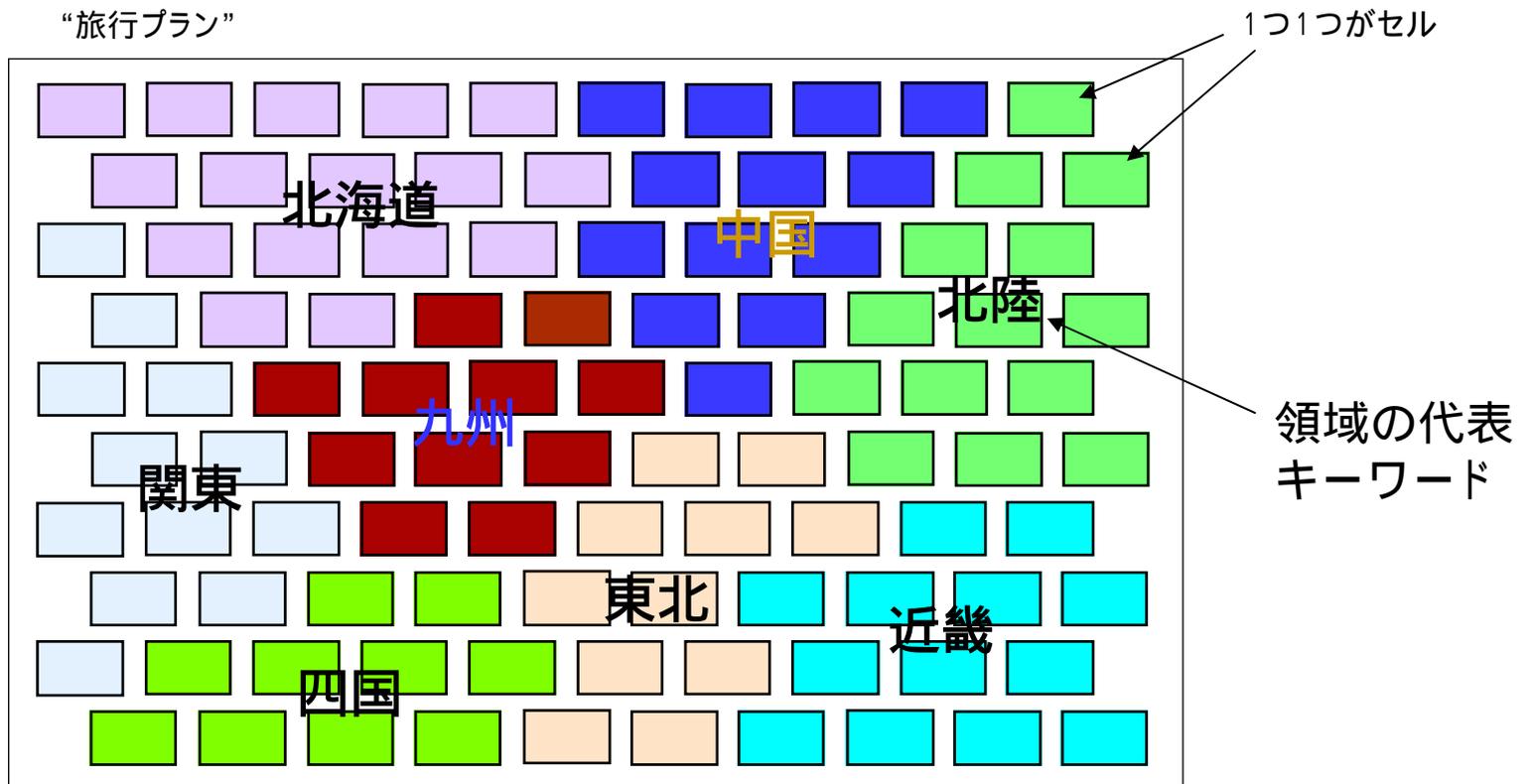


■ SOM (Self-Organizing Maps: 自己組織化マップ)

- ニューラルネットワークの1種
- データを2次元平面に配置された入れ物(セル)に割り振る
- 類似した特徴を持つデータが近くなるように割り振る
- 1つのセルには複数のデータが割り振られることもある

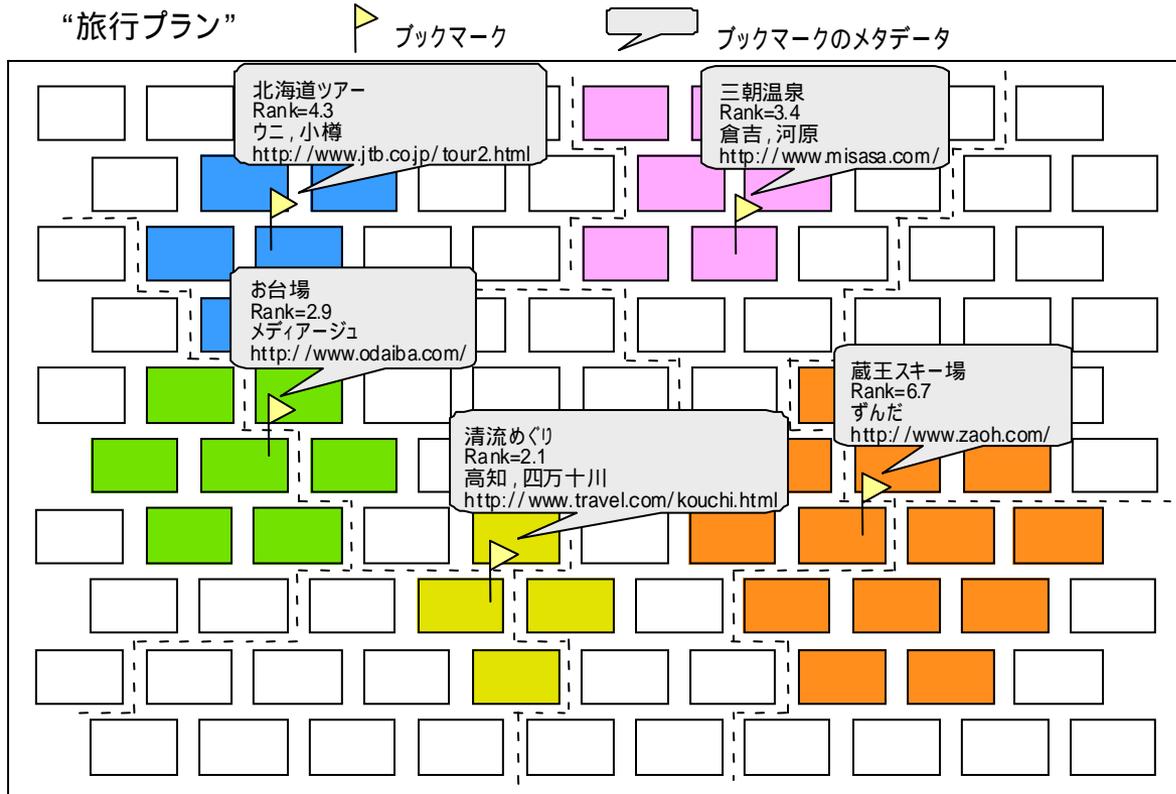
探索アクティビティの視覚化と共有

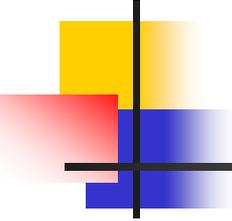
- SOMで検索結果を視覚化したイメージ



探索アクティビティの視覚化と共有

- 「自分が探索を行った範囲」を視覚的に理解可能



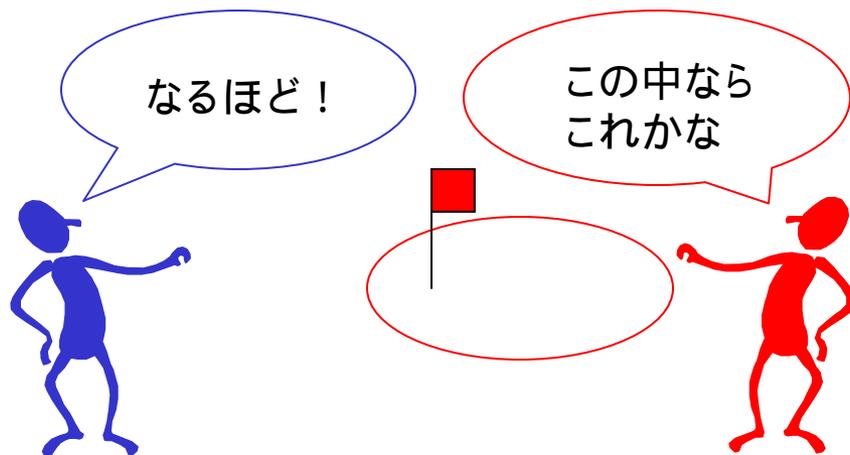


探索アクティビティの視覚化と共有

- 視覚化されたコンテキストブックマークを共有することで...
 - 他人の探索範囲も視覚的に理解可能
 - 相手との探索内容の関係も理解可能
 - 互いの得意・不得意分野などに応じた分担が可能

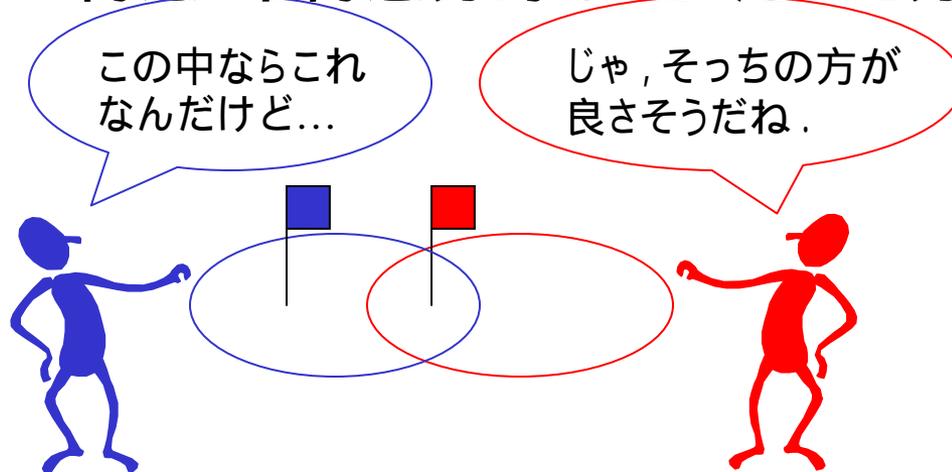
探索アクティビティの視覚化と共有

- 視覚化されたコンテキストブックマークを共有することで...
 - 他人の探索範囲も視覚的に理解可能
 - 相手との探索アクティビティの関係も理解可能
 - 互いの得意・不得意分野などに応じた分担が可能



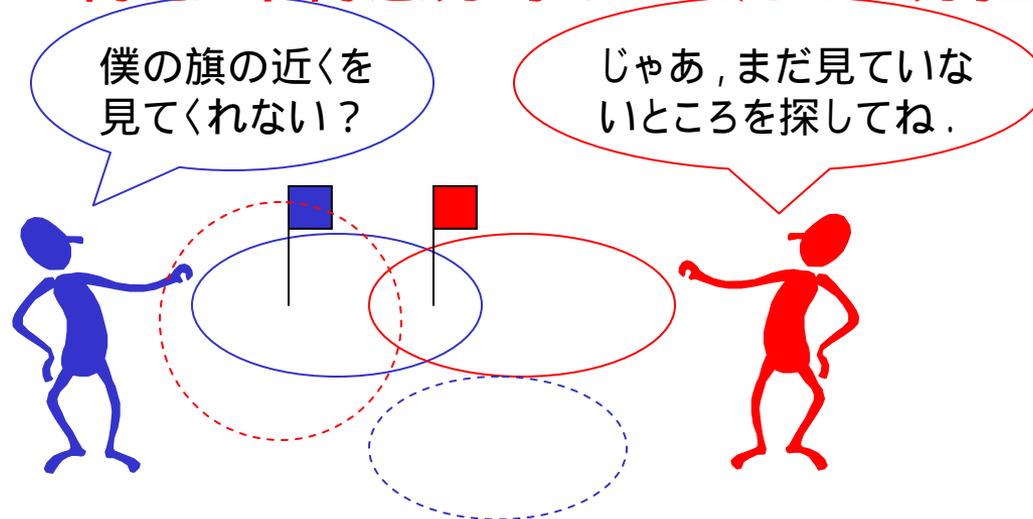
探索アクティビティの視覚化と共有

- 視覚化されたコンテキストブックマークを共有することで...
 - 他人の探索範囲も視覚的に理解可能
 - 相手との探索アクティビティの関係も理解可能
 - 互いの得意・不得意分野などに応じた分担が可能



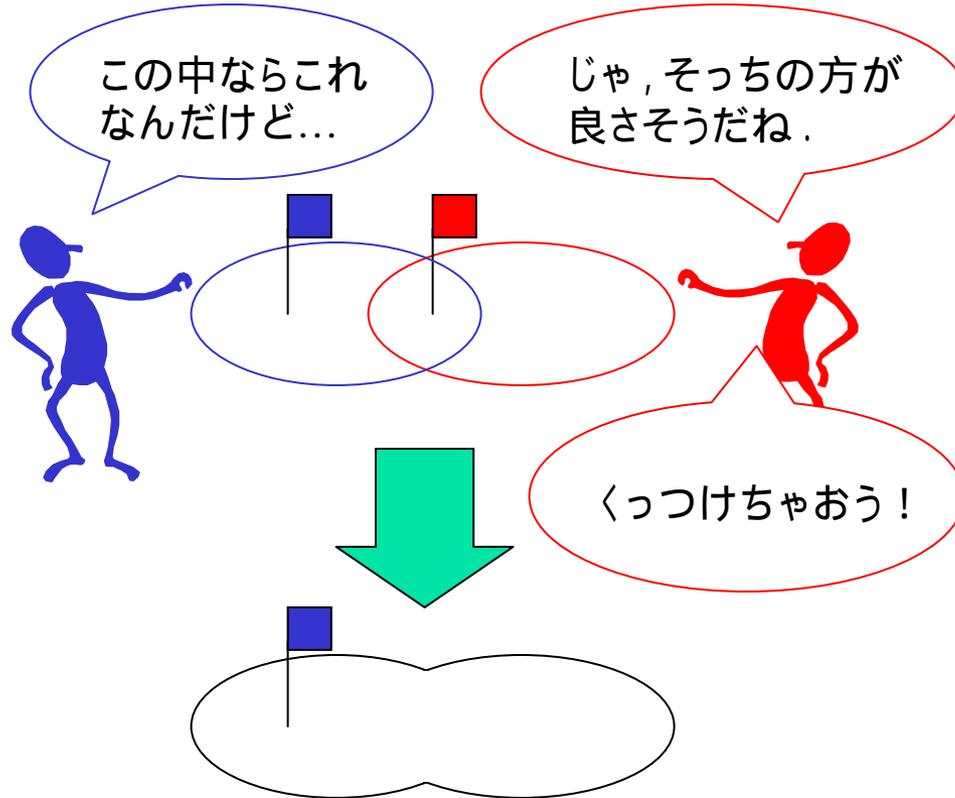
探索アクティビティの視覚化と共有

- 視覚化されたコンテキストブックマークを共有することで...
 - 他人の探索範囲も視覚的に理解可能
 - 相手との探索アクティビティの関係も理解可能
 - 互いの得意・不得意分野などに応じた分担が可能

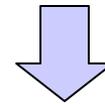


探索アクティビティの合成

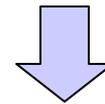
- 他のユーザと合意した探索アクティビティを生成



繰り返し行う



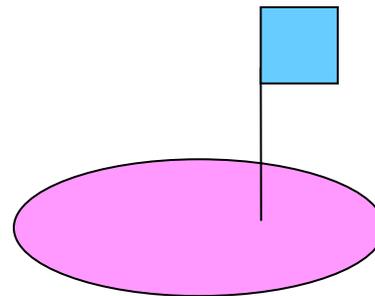
多くの探索アクティビティを1つにまとめる



重複する探索アクティビティの整理ができる!

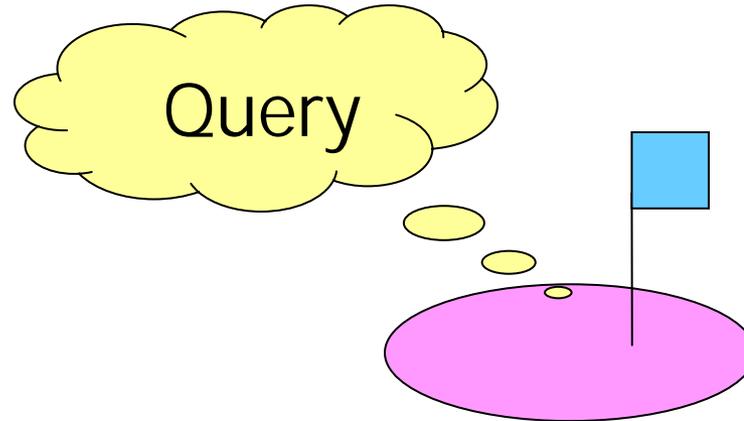
探索アクティビティの問合せ利用

- 「コンテキストブックマーク」 = 「ユーザの意図」
抽象化することで質問を生成
他の領域に適用
領域内でユーザの意図に最も合致したページを得る



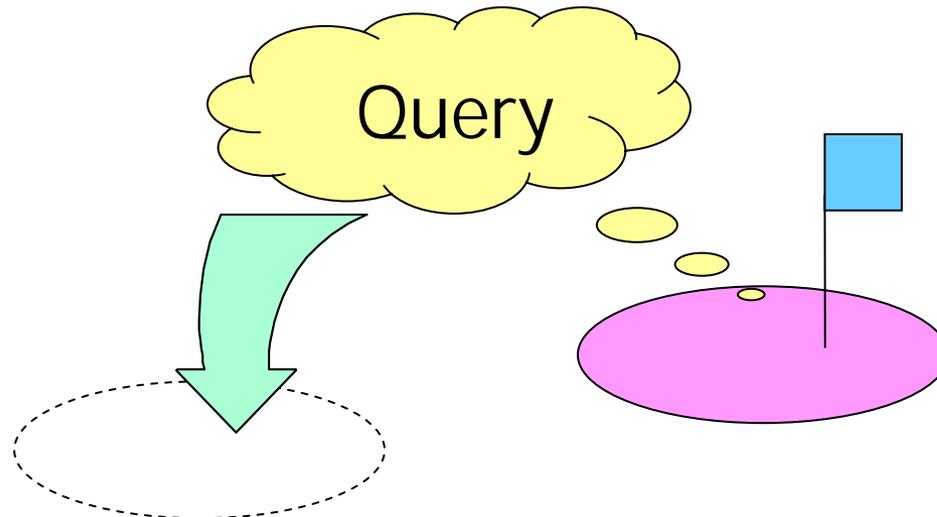
探索アクティビティの問合せ利用

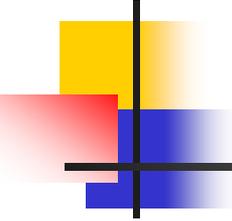
- 「コンテキストブックマーク」 = 「ユーザの意図」
抽象化することで質問を生成
他の領域に適用
領域内でユーザの意図に最も合致したページを得る



探索アクティビティの問合せ利用

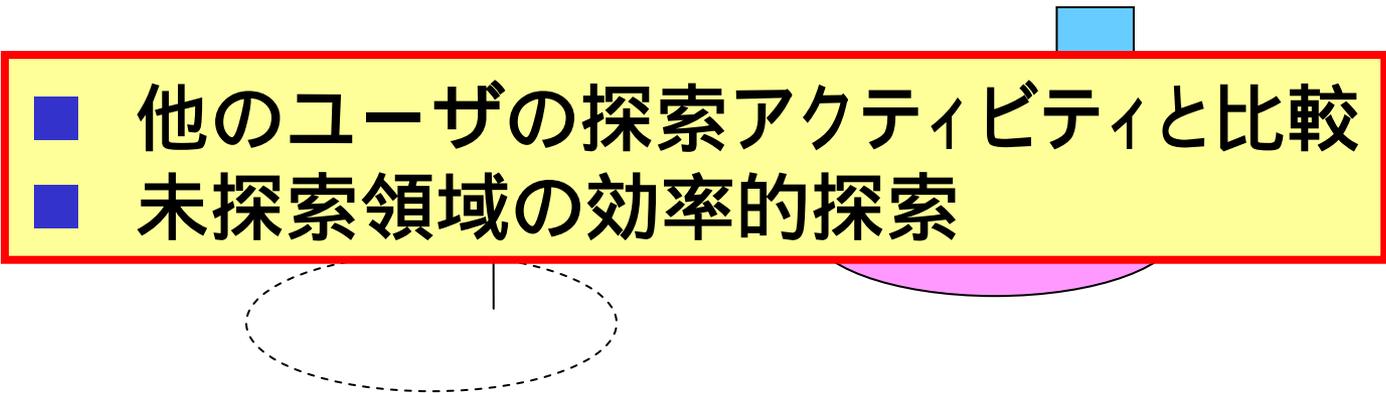
- 「コンテキストブックマーク」 = 「ユーザの意図」
抽象化することで質問を生成
他の領域に適用
領域内でユーザの意図に最も合致したページを得る

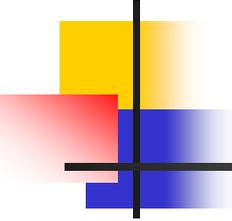




探索アクティビティの問合せ利用

- 「コンテキストブックマーク」 = 「ユーザの意図」
抽象化することで質問を生成
他の領域に適用
領域内でユーザの意図に最も合致したページを得る

- 
- 他のユーザの探索アクティビティと比較
 - 未探索領域の効率的探索



探索アクティビティの問合せ利用

■ 抽象化の手法

- ブックマークされたページの特徴キーワードの出現頻度が最大
- それまでに閲覧されたページの特徴キーワードの出現頻度が 0
- 上記のいずれにも該当しない単語の出現頻度は不問

■ 抽象化の手法

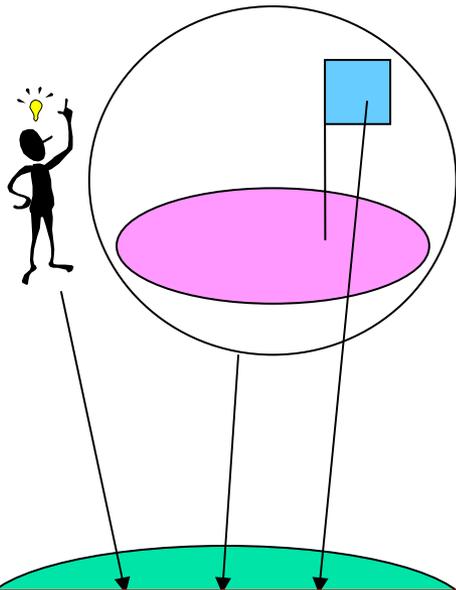
- 探索アクティビティ A_i から生成した質問 $q_i(A_i)$

$$p_{q_i}(A_i) = b_i + \alpha \sum_j (b_i - c_{ij})$$

b_i	ブックマークされたページの特徴ベクトル
c_{ij}	それまでに閲覧されたページの特徴ベクトル
α	定数

探索アクティビティの問合せ利用

実験：未探索領域への適用



・ユーザによる評価

すべてのページをユーザの意図に合っているかどうか ×で評価(理想の評価)

・ブックマークされたページで評価

コンテキストに関係なく, そのページの特徴そのものに類似したページ順にランク付け

・コンテキストブックマークによる評価

コンテキストから得た 必要な単語と不要な

が, にどれだけ近い結果を出すかで, ユーザの意図がくみ取れているかを実験

探索アクティビティの問合せ利用

実験結果

実際に検索をしたわけではないので、
再現率と適合率に近い基準である
2つの正解率で評価

正解率	0.63	0.76
正解率	0.60	0.66

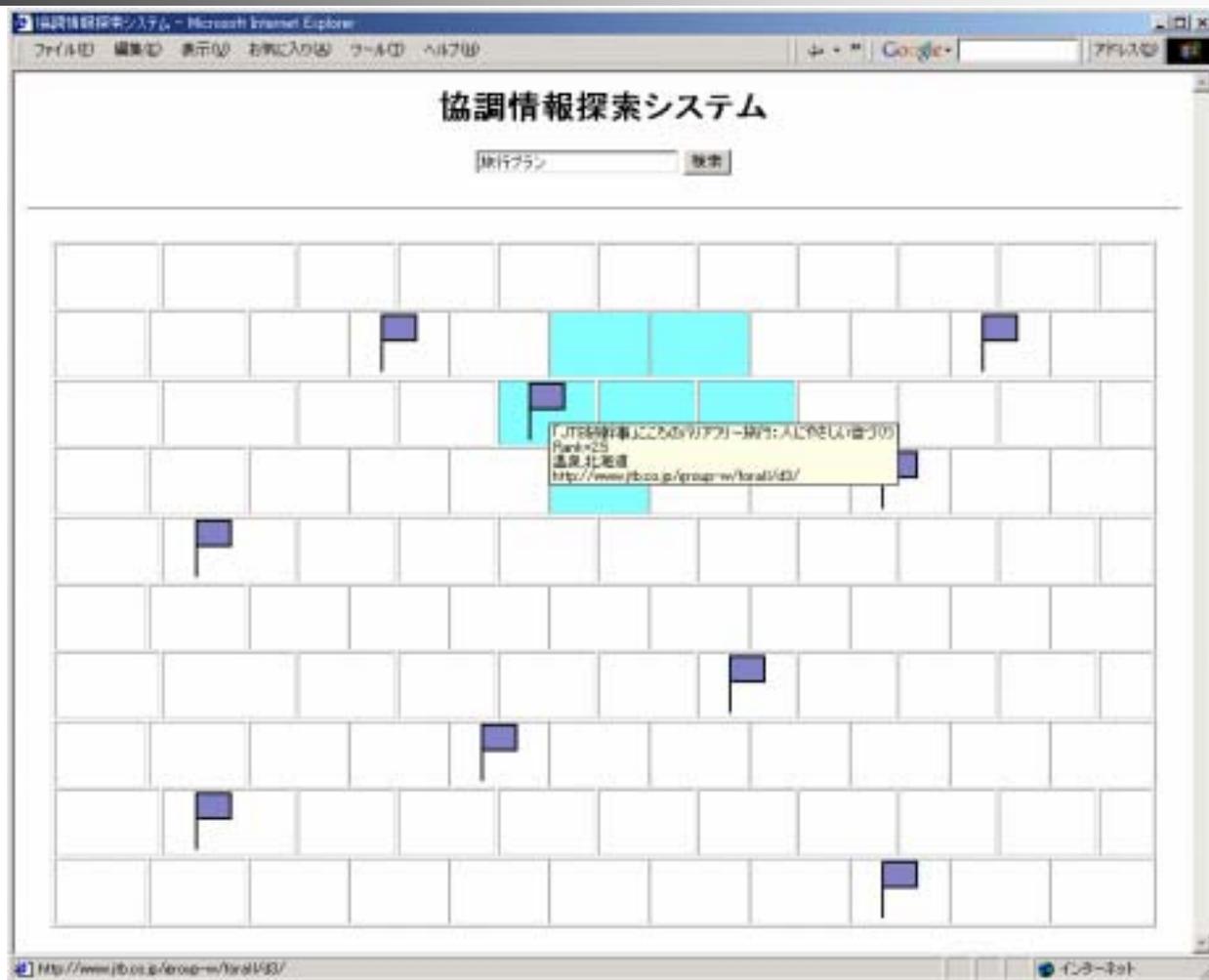
再現率 実際に検索された情報 必要な情報

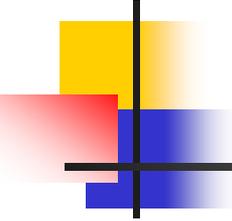
正解率 実際に検索された情報 必要な情報

a: ユーザーが としてページの数 b: としてページが全て出てきた順位

ブックマークされたページの特徴だけでも、ある程度の精度はあるが、コンテキストの内容を反映させることで、精度が上がる事が確認された。

プロトタイプ





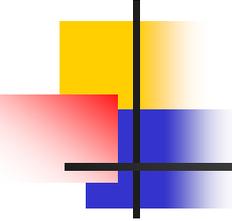
まとめ

■ 本研究の提案

- コンテキストブックマークによる探索アクティビティの保存
- SOMを用いた探索アクティビティの視覚化と共有
- 探索アクティビティの合成・問い合わせ利用による
ユーザ間協調の支援

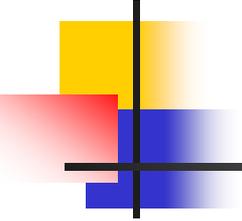
■ 今後の予定

- プロトタイプの実装の完成 および システムの評価



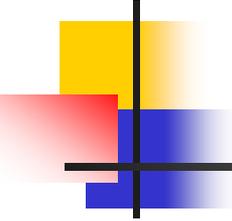
検討事項

- 視覚化・マッピング手法の検討
 - 今回はSOMを用いたが, SOM以外の手法でも応用可能かどうか
- 情報共有空間に対する操作の検討
 - 他のユーザの探索結果に対する注釈付け
 - 探索結果への評価の与え方
- システムの評価方法の検討



以上で発表を終わります。
ありがとうございました。





言葉の違い

- 閲覧履歴
 - 過去にずっと見てきたもの
- 探索アクティビティ
 - 探索するために見てきた履歴
- コンテキストブックマーク
 - 探索アクティビティ + 解析したユーザの意図

実験内容

■ コンテキストブックマークによる評価

基準となるベクトルの生成法

必要な単語

$$\begin{array}{l} \text{ブックマーク} \\ \text{されたページ} \\ \text{での出現頻度} \end{array} + \left[\begin{array}{l} \text{ブックマーク} \\ \text{されたページ} \\ \text{での出現頻度} \end{array} - \begin{array}{l} \text{コンテキスト内} \\ \text{のページでの} \\ \text{平均出現頻度} \end{array} \right] \times \text{定数}$$

不要な単語

$$\begin{array}{l} \text{ブックマーク} \\ \text{されたページ} \\ \text{での出現頻度} \end{array} - \left[\begin{array}{l} \text{ブックマーク} \\ \text{されたページ} \\ \text{での出現頻度} \end{array} - \begin{array}{l} \text{コンテキスト内} \\ \text{のページでの} \\ \text{平均出現頻度} \end{array} \right] \times \text{定数}$$

その他の単語

ブックマークされたページでの出現頻度

正解率

$$\text{再現率} = \frac{\text{実際に検索された情報}}{\text{必要な情報}}$$



$$\text{正解率} = \frac{\text{上位}a\text{位にはいった数}}{a}$$

a: ユーザが としたページの数

$$\text{適合率} = \frac{\text{必要な情報}}{\text{実際に検索された情報}}$$



$$\text{正解率} = \frac{a}{b}$$

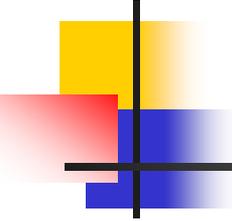
b: としたページが全て出てきた順位

未探索ページ: 10個 の数(a): 4個 の場合

順位 1 2 3 4 5 6 7 8 9 10

× × × ×

正解率 = 3/4 正解率 = 4/6



研究の違い

■ 上田(この発表)

- グループによる協調した情報探索を想定
- コンテキストブックマークで探索アクティビティを保存・提示

■ 黒田(直前の発表)

- 個人による情報探索を想定
- コンテキストブックマークを検索・再利用