

サイト固有のカテゴリ構造を考慮したサジェストインタフェースの提案

木村久美子*1, 中清志都加*2, 堀 雅洋*1

*1: 関西大学大学院総合情報学研究科

*2: 関西大学総合情報学部

1

キーワードサジェスト機能

- キーワード入力段階で、キーワードと合致する複数の候補をユーザに提示
 - 不特定多数ユーザの入力履歴, キーワードの共起関係に基づく



2

キーワード検索における付加情報


- キーワード候補だけを拠り所とした検索 [White 2007]
 - ユーザを誤った結果へ誘導する危険性を指摘
 - 例: 「世界初の女性宇宙飛行士」を調べるタスク
 - 誤って到達した結果「初のアメリカ人女性宇宙飛行士」
 - 検索結果またはスニペット等を伴う結果リストを一度も閲覧していない初期状態では判断が難しい



- サジェストインタフェースで、ページ内容を特徴付ける情報を付加することも有用

3

汎用検索エンジンによるサイト内検索

- 探索範囲をURLによって制限
 - サイト内検索ではより具体的なキーワードが入力される傾向がある [Chau 2005]
- 
- サイト内の絞り込みに有用な情報を提示できるとは限らない
 - サイト内で取り扱う情報や、構造を考慮すべき

4

サジェストインタフェースの例(1)

単語共起関係に基づく階層構造化[Joho 2003]

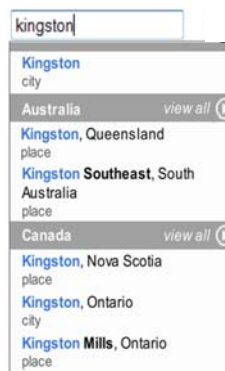
storm (119)	tropical storm (103)
tropical (86)	wind (45)
rain (53)	hurricane (34)
water (52)	disaster (32)
island (45)	injury (19)
united (44)	atlantic (17)
damage (42)	gulf (17)
weather (36)	weather service (12)
coast (34)	louisiana (8)
mile (34)	meteorologist (7)
agriculture (34)	torrential (7)
writer (32)	wire storm (7)
lo (30)	martinique (6)
pacific (29)	tropical depression (6)
temperature (28)	

- 階層構造によって探索効率が向上
 - より多様で、正確な文書を見つけることが可能
- 精度と再現率の向上は見られず

5

サジェストインタフェースの例(2)

語彙の汎化関係に基づくグループ化[Amin 2009]



- 地理情報のグループ化
 - ユーザにとって馴染みのある分類体系を構築可能
- 汎用の意味分類 (WordNet) によるグループ化
 - 個別の検索状況において効果的な検索支援に至らず
 - ユーザにとって馴染みがなく、検索状況に適した分類体系でなかった

6

研究の目的

- サイト固有のカテゴリ構造を考慮したサジェストインタフェースの実現
 - サイト構造に従った絞り込み検索を行う

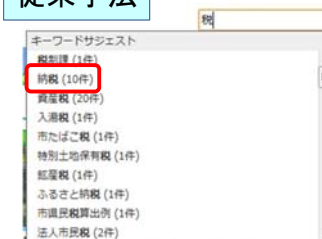


- 動的なサイトナビゲーション支援による閲覧効率の向上

7

従来手法との比較

従来手法



提案手法



- 提案手法では、自動的に該当カテゴリでの絞り込み検索を行う
 - ヒットするページ数の低減による閲覧時間の短縮

8

サジェストインタフェースのデモ



サジェストインタフェースのデモ



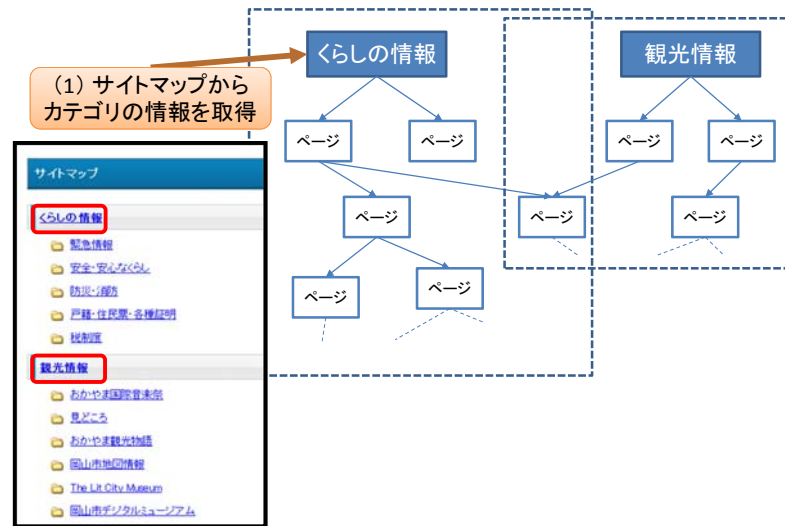
解決方法

- サジェスト候補のグループ化
 - Webサイト内のページ・キーワードをサイト構造に従ってグループ化
 - サイトマップを使用
- サジェスト候補に付加情報を表示
 - ページタイトル文字列

サイトマップ

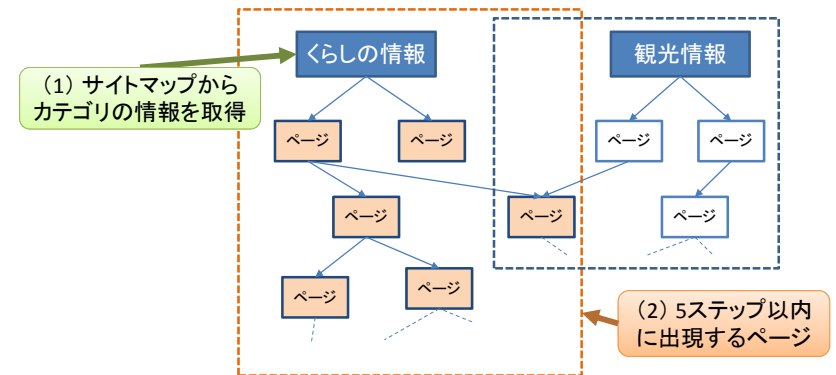
- Webページで取り扱う情報
 - サイト独自に構造化
 - 上位構造はサイトマップとして多くのサイトで提供されている
- ↓
- サイトマップの分類体系を利用し、サジェスト候補をグループ化

サイトマップ情報に基づく ページのグループ化 (1/3)



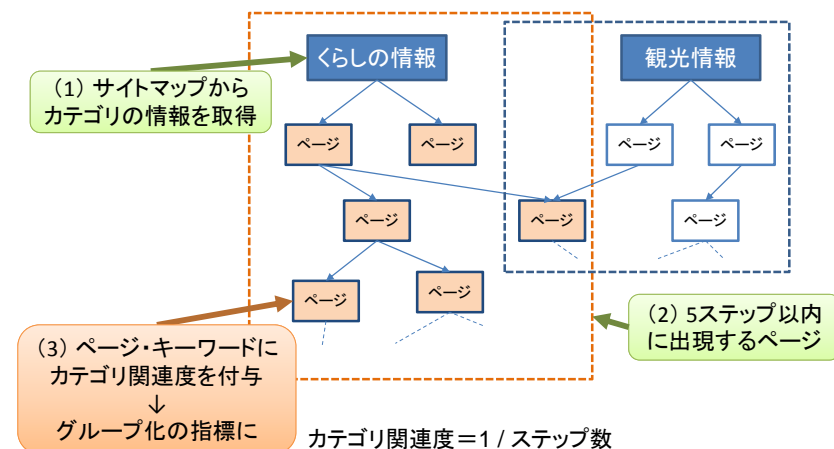
13

サイトマップ情報に基づく ページのグループ化 (2/3)



14

サイトマップ情報に基づく ページのグループ化 (3/3)



15

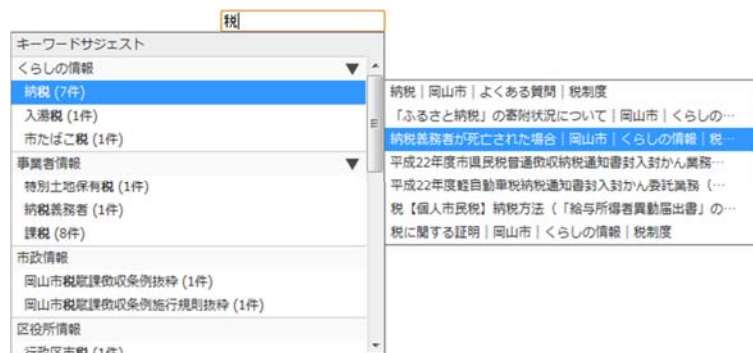
キーワードの抽出

- 前提となる文字列の取得
 - ページ内容に関わる文字列を用いる
 - リンクラベル, ページタイトル
 - ページ固有でない情報は除外
 - ヘッダ, フッタ, グローバルナビゲーションなど, 複数ページに繰り返し出現する文字列を除外
- 形態素解析によってキーワードを抽出
 - 連続して出現した名詞は結合
 - 漢字1文字, かな2文字以下のキーワードは除外

16

付加情報: タイトルリスト

- キーワードに合致するページ群のタイトル



17

評価実験

- 目的
 - サジェストキーワードのグループ化表示による閲覧効率向上の検証
- 実験協力者
 - 関西大学総合情報学部生16名
- 対象サイト: 2サイト
 - 各カテゴリのトップから5ステップまでクロール
 - 自治体サイト(岡山市) (6549ページ)
 - 大学サイト(国士舘大学) (3169ページ)

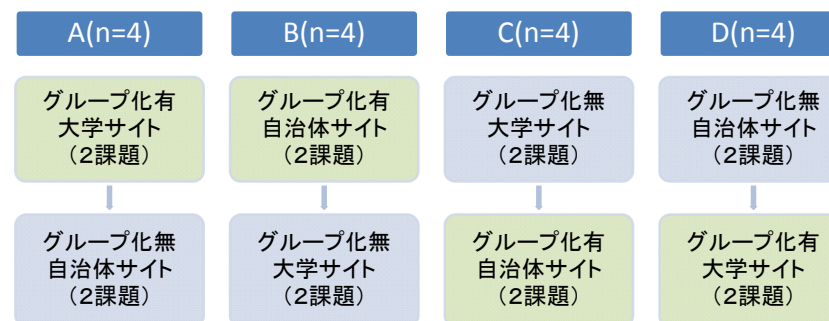
18

実験課題

- 探索課題
 - 各課題サイトについて2問ずつ出題
- 課題例:
 - 自治体サイト
 - 「図書館に行きたいと考えていますが、どこにあるのかわかりません。どこに図書館があるかが分かるページを探してください。」(正解:2ページ)
 - 大学サイト
 - 「武道に関する研究を行っている施設があると聞きました。その研究施設について書かれているページを探してください。」(正解:1ページ)

19

実験手順



- 2要因(グループ化有無×サイト種別)協力者内計画
 - 課題の出題順序はカウンターバランスをとって相殺

20

実験結果

- 30秒毎の課題達成率
 - 2要因分散分析: 有意差, 交互作用なし
 - 90秒時点で16名中14名が正解した(頭打ち状態)
- 事後アンケート
 - サジェスト機能の使いやすさ
 - 16名中13名が使いやすいと回答
 - グループ化有のサジェスト機能の使いやすさ
 - 16名中11名が使いやすいと回答


21

平均精度

- 精度: 検索クエリによってヒットした文書群のうち, 正解文書がどれだけ含まれているか
- 2要因分散分析
 - グループ化の有無: 有意差あり ($p < 0.05$)
 - グループ化あり(0.81) \gg なし(0.66)
 - サイト種別: 有意差あり ($p < 0.05$)
 - 自治体サイト(0.84) \gg 大学サイト(0.63)
 - 交互作用なし

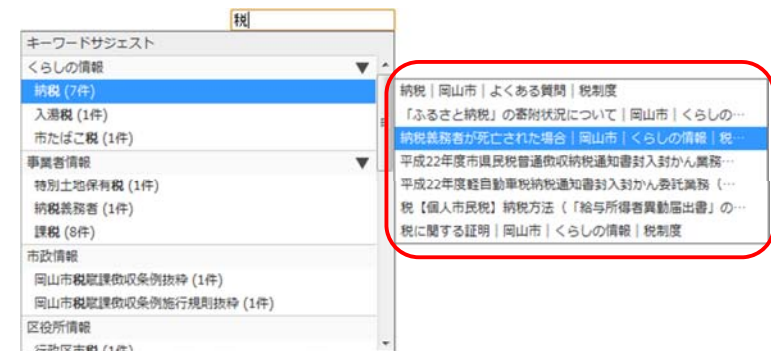
22

精度についての考察

- キーワード候補リストのグループ化表示によって検索精度が向上
 - ユーザが適切なカテゴリを選択できている
 - 課題達成率の向上に貢献していない
- 
- 付加情報の表示が達成率に影響している可能性

23

付加情報: タイトルリスト



24

付加情報に関する結果(1)

- タイトルリストの利用率
 - 全検索回数の36% (38/105)
 - ほとんどの協力者が両方の遷移方法を併用
- キーワード入力完了からキーワード候補orタイトルリストのクリックまでに要した平均操作時間
 - タイトルリスト ≧ キーワード候補リスト
(21.6秒) (12.0秒)

有意差あり (p<0.01) [Wilcoxonの順位和検定]

25

付加情報に関する結果(2)

- 課題達成率
 - キーワード候補をクリック→検索結果ページ
 - 55.2% (38/67)
 - タイトルリスト
 - 71.1% (27/38)

26

付加情報についての考察

- タイトルリストを利用した場合
 - サジェスト内容をよく吟味している
 - 課題の達成率にも貢献している可能性
- アンケート
 - 「関連する別ページを参照・比較できる」



- タイトルリスト内の情報によって、グループ化の有無にかかわらず課題達成率が底上げされている可能性
- 付加情報の有無を要因とした評価実験による検証

27

サジェスト機能についての考察

- 通常のサイトナビゲーションと異なり、深さによらず1回の選択で目的ページに到達可能
 - ただし、クロール範囲でのみ
- 本手法では、指定ステップ以降のページは検索不可
 - クロール遷移数を7ステップに設定した場合
 - 自治体サイト: 11228ページ (= 6549 + 4679)
 - 大学サイト: 3898ページ (= 3169 + 729)
 - クロールステップ数を増やすことで対処可能
 - ナビゲーションリンクを併用することで探索可能
 - 目的の情報の近くまでユーザを誘導

28

今後の課題

- 付加情報の有無を要因とした評価実験
- 現状のサイトナビゲーション(+サイト内検索)との比較実験
 - どの程度の深さにあるページまでサジェストの対象とするか
- よりサイト内探索に適したグループ化インタフェースの検討
- 広いユーザ層での評価

参考文献

- [Chau 2005]Chau, M., Fang, X., and Sheng, O. R. L.: Analysis of the query logs of a Web site search engine. Journal of the American Society for Information Science and Technology, Vol. 56, No. 13, pp. 1363-1376, (2005)
- [White 2007]White, R. W. and Marchionini, G.: Examining the effectiveness of real-time query expansion. Information Processing and Management, Vol. 43, No. 3, pp. 685-704, (2007)
- [Joho 2003]Joho, H., Coverson, C., Sanderson, M., and Beaulieu, M.: Hierarchical presentation of expansion terms. In: Nyberg, K., Heys, H.M. (eds.) SAC 2002. LNCS, vol. 2595, pp. 645-649. Springer, Heidelberg (2003)
- [Amin 2009]Amin, A. K., Hildebrand, M., van Ossenbruggen, J. R., Evers, V., and Hardman, L.: Organizing Suggestions in Autocompletion Interfaces . European Colloquium on IR Research - ECIR , pp. 521-529, (2009)