4. 検索連動型広告を用いたクエリ拡張に基づ くタスクを含むウェブページの発見

本稿では,3 章で述べたタスク検索を実現するための第 一歩として,タスクを含むウェブページを網羅的に発見す るための手法を提案する.1 章でも述べたとおり,通常の ウェブ検索を行っただけでは,ある目的を達成するための タスクを網羅的に収集することは難しい.そこで,本稿で は,検索エンジンの検索連動型広告に着目し,そこからタ スクに関連した動詞を用いて入力クエリを拡張すること で,元の入力クエリから得られるウェブページには含まれ ていないようなタスクを含んだウェブページを発見する.

4.1 検索連動型広告を用いたクエリ拡張 前章で説明したように,タスクとは,目的を達成するた

めに必要な,ユーザが取るべき行動と考えることができ る.Yamamoto らや田麥らは,そうしたタスクは検索エン ジンの検索連動型広告と非常に関連が深いことを指摘して いる.現在,多くの商用検索エンジンでは,ユーザが入力 したクエリに対して,広告主によって登録された広告のう ち関連度の高いものが,検索結果ページ上に表示されるよ うになっている.この検索連動型広告とタスク検索の関係 に関して,たとえば,Yamamoto らは,タスク検索は一般 的に達成することが困難であるため,そのようなドメイン に多くの広告が登録されており,1 つ 1 つの広告は特定の 行動やトランザクションに関連していると述べている [1]. また,田麥らは,あるサービスを検索する場合,その多く は検索連動型広告に出現しており,検索連動型広告を利用 することでサービスに関するウェブを効率的に検索できる ことを示している [8].

そこで,本稿ではクエリ拡張を行う際の拡張語を選択す るための情報源として,検索連動型広告を利用する.検索 連動型広告を利用することで,タスクに関連し,かつ通常 のウェブ検索結果からでは得られないようなタスクに関連 した語句がクエリ拡張語として得られることが期待される.

4.2 手法の流れ 提案手法の概要を図 4 に示す.本手法は以下の流れで

ウェブページを取得する. ( 1 ) タスク(目的となる状態)を表したクエリ q を受け

取る. (2) q0 でウェブ検索を行い,n 件の検索連動型広告

{a1, . . . , an} を取得する. ( 3 ) {a1 . . . an } のタイトルおよびスニペットに対して形態

検索連動型 広告

動詞・サ変名詞 の抽出

拡張 拡張 拡張 拡張 拡張 クエリ クエリ クエリ クエリ クエリ

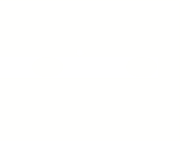
ウェブ

結果ページを 統合

図 4 提案手法の概要 Fig. 4 Overview of our method.

素解析を行い,出現する動詞およびサ変接続名詞を抽 出し出現回数を計算する(得られた単語集合を W ,単 語 w ∈ W の出現回数を tfw とする).

(4) tfw の高い単語上位 m 件({w1′ ...wm′ } とする)を取 得する.



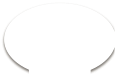
   

広広告告タタイイトルト・ルス・ニスペニッペトッ 広告タイトル・スニペッ

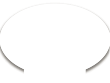


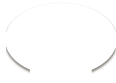
ト ト







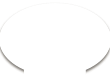




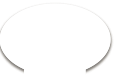


















































Web Web Web Web Web 検索 拡張 拡張 拡張 拡張 拡張

ページ ページ ページ クエリ クエリ クエリ

ページ ページ クエリ クエリ

(5)

ウェブページ集合を D = φ とし, (4)で得られたそ れぞれの単語 wi(1 ≤ i ≤ m)について,|D| = k と なるまで以下のプロセスを繰り返す. (a) 元のクエリに得られた単語を追加したクエリ

qi =q∧wi を作成する. ( b ) 得られたクエリ qi でウェブ検索を行い,上位 k′

件のウェブページ {d(i) . . . d(i) } を取得する. 1 k′

( c ) 得られたウェブページを d(i) から順に D に追加 1

(D ← D ∪ {dk′ })する.このとき,|D| = k とな 1

Web 拡張

ページ拡張 クエリ

拡張 クエリ

クエリ

ればステップ(6)へ移行する. ( 6 ) 得られたウェブページ集合 D を返す.

上記の手法を適用することで,元のクエリ q で得られる ウェブページ集合より多くのタスクを含んだ,つまり得ら れるタスクという観点で再現率の高いウェブページ集合を 集めることができると考えられる.