

平成 21 年度 情報工学コース卒業研究報告要旨

渡邊 研究室	氏 名	松 岡 研 二
卒業研究題目	適合・不適合指示情報に基づいたウェブ検索結果のリランキング	

近年，ブログや SNS の普及により，Web ページの数が爆発的に増大している．それに伴い，膨大な数の Web ページから目的のページを探し出すことができる検索エンジンの需要が高まっている．最も利用されているキーワード型の検索エンジンは，ユーザの入力したクエリと呼ばれる検索語を基に，Web ページを抽出しユーザに提示する．しかし，複数の意味を持つ語が存在するため，検索結果に不要なページが含まれることがある．このとき，ユーザは検索結果から目的のページを探す必要がある．また目的のページが見つからないとき，クエリを修正して再検索する必要がある．

本論文では，適切な検索結果が得られない場合における目的のページの抽出を目的とする．検索結果を閲覧するとき，それぞれのページがユーザにとって必要かどうかの評価が無意識になされている．そこで，本論文ではページに対する適合・不適合の評価を基に検索意図を推測し，それを検索結果に反映することで検索結果を改善する手法を提案する．本手法の全体像は図のようになっている．適合ページと不適合ページの特徴の差異が，適合ページと不適合ページを分ける基準となると捉え，差異に基づき評価基準を推定する．検索結果のリランキングでは，推定した評価基準において，適合ページに類似し，不適合ページに類似しないページを抽出することで，検索結果の閲覧の支援を行う．修正したクエリの再検索では，適合ページの評価基準であり，かつ検索結果数の少ない語を元のクエリに追加することでクエリを修正し，ユーザに提示することで再検索の支援を行う．

本手法の有効性を確認するため，リランキングとクエリの修正に関する評価実験を行った．リランキングの実験では，対抗手法として BM25 と Rocchio アルゴリズムを組み合わせた手法を用い，24 個の検索タスクにおいて，評価回数ごとの適合率と再現率を算出した．その結果，対抗手法と比較して評価回数 1 回から 5 回の平均再現率が 15.8% 改善され，評価基準を用いない場合と比較して 5.3% 改善された．また，クエリの修正の実験では，評価回数 4 回の情報を基にクエリを修正し，クエリ候補の上位 5 語により再検索を行った．その結果，再検索前の検索結果と比較して，上位 20 件の適合率が平均で 33.0% 改善された．本実験により，評価基準を用いたリランキングとクエリの修正の有効性が確認できた．

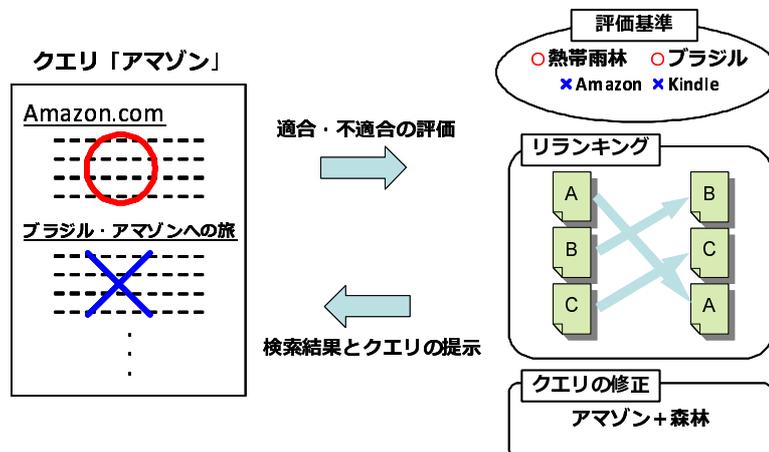


図: 本手法の全体像