4K-01

# 意外性のある検索クエリの推薦方法の提案

鈴木　永史郎†　　杉本　徹‡

芝浦工業大学大学院　理工学研究科†　　芝浦工業大学　工学部‡

## １．研究背景と目的

Web 検索を用いて事実や出来事について調べるとき，システムによって推薦された検索クエリを用いることがある．しかし，一般的によく知られている情報が得られるクエリが推薦されることもあり新たな知識を獲得できないことがある．これに対し，推薦システムの研究では，未知の情報を提示するために意外性を示す Serendipity を指標とした研究がなされている．例として加藤らの研究[1]がある．加藤らは五感に関連したオノマトペを特徴量とした飲食店の推薦を行い，オノマトペを使用しない場合の推薦と比較して高い意外性の評価値が得られ，意外性のある飲食店の推薦が可能であることを示した．一方で，検索クエリを推薦アイテムとした研究はなされていない．

本研究では，Serendipity に着目し，思いつきづらく，予想できない情報が得られる検索クエリを意外性のある検索クエリとしてユーザに推薦する方法の提案を行う．本研究では，ユーザが調べたい情報を検索語，検索語と共に入力し検索範囲を限定する語を検索範囲限定語と呼ぶ．また，推薦する検索クエリは検索語と検索範囲限定語をスペース区切りの形で結合したものとする．

## ２．検索サジェストの収集

推薦する検索クエリは Google サジェストを用いて収集し，Google 検索への入力には，検索語のみ，および，たとえば「検索語　あ」という形で五十音，濁点，半濁点，拗音，アルファベットなどを検索語に対しスペース区切りで加えた 137 個のパターンを用いた．

## ３．意外性のある検索クエリの調査

推薦する検索クエリは，意外性があるだけでなくユーザにとって有用なクエリである必要がある．そこで，クエリの意外性および有用性をアンケートによって調査しその結果を用いてクエリに対する意外性の定義を行った．

Proposal of a Recommendation Method of Serendipitous Queries

†Eishiro Suzuki Graduate School of Engineering and Science,Shibaura Institute of Technology

‡Toru Sugimoto College of Engineering,Shibaura Institute of Technology

### ３．１．　意外性に関するアンケート調査

アンケートでは「夏目漱石」に関する検索クエリ 132 個について 9 名の被験者に夏目漱石と検索範囲限定語の組み合わせが意外であると感じるかについて 1～5 の 5 段階で評価してもらった．評価値の平均が 3.0 を上回った検索範囲限定語は 38 個あった．結果の一部を表 1 に示す．

表 1. 検索語に対する検索範囲限定語の意外性の評価値の平均

| 検索語 | 検索範囲限定語 | 評価値の平均 |
|---|---|---|
| 夏目漱石 | スコットランド | 4.7 |
| 夏目漱石 | 野球 | 4.1 |
| 夏目漱石 | 長男 | 1.8 |
| 夏目漱石 | 小説 | 1.0 |

この結果から，検索語に対して検索範囲限定語を思いつくことが難しい場合に意外性があると考えられる．

### ３．２．　有用性に関するアンケート調査

意外性に関するアンケートの評価値の平均が 3.0 を上回ったクエリについて，6 名の被験者に各検索クエリによる Google の検索結果のスニペットを見せ，その情報を知ることができて嬉しいと感じるかについて 5 段階で評価してもらった．評価値が 3.0 未満のクエリは 23 個あった．結果の一部を表 2 に示す．

表 2. 検索クエリに対する有用性の評価値の平均

| 検索クエリ | 評価値の平均 |
|---|---|
| 夏目漱石　スコットランド | 4.7 |
| 夏目漱石　野球 | 3.6 |
| 夏目漱石　癖 | 2.2 |
| 夏目漱石 キリスト教 | 2.0 |

この結果から，評価値が 3.0 未満のクエリは得られる検索結果を容易に推測でき，ユーザにとって思いがけない情報が得られないため評価が低くなっていると考えられる．そこで検索クエリから得られる検索結果が思いがけない情報であるとき，有用性があるとする．

### ３．３．　意外性のある検索クエリの定義

アンケート調査の結果を踏まえて，本研究では検索クエリに対する意外性を「思いつきづらさ」と「得られる情報の予想しづらさ」の 2 つの指標によって定義する．具体的には，2 つの指標の値を算出しその積によって意外性を判断する．

## ４．意外性のある検索クエリの算出

### ４．１．　思いつきづらさの算出

思いつきづらさは検索語と検索範囲限定語の関連性を算出し，その値の低さによって判断する．2 つの語の関連性の算出には，Wikipedia のカテ

ゴリリンクを利用する．カテゴリリンクは，Wikipedia における記事とカテゴリとの所属関係を表現し，カテゴリもまた別のカテゴリとの所属関係を持つことでグラフ構造をなしている．

カテゴリリンクを用いた単語間の関連性の算出式は伊藤らの研究[2]に基づく．検索語 s および検索範囲限定語 t をそれぞれ記事のタイトルとするページが存在する場合，カテゴリリンクを辿ることで記事 s から t に到達できる経路の総数を l，各経路の長さを$p_k$(1≦k≦l)とするとき，単語の関連の強さのスコア pf は以下の式で表される．

$$pf(s,t)=\sum_{k=1}^{l}\frac{1}{p_k} \quad (1)$$

(1)式は，ある記事から辿ることができるカテゴリの総数が多い場合，経路が多くなり値が大きくなる．そこで，2 つの記事 s,t においてそれぞれの所属カテゴリの総数 cf(s),cf(t)のうち小さい方の値の逆数を pf にかけることで対処する．よって，関連性のスコア pficf は以下の式で表される．

$$picf(s,t)=pf(s,t)\cdot icf(s,t) \quad (2)$$

$$icf(s,t)=\frac{1}{min(cf(s),cf(t))} \quad (3)$$

### ４．２．　得られる情報の予想しづらさの算出

得られる情報の予想しづらさは，検索語と同じ範疇(人名，組織名など)に属する検索語から収集した検索クエリの集合を用いて，検索範囲限定語の出現頻度の少なさによって判断する．つまり，得られる情報の予想しづらさを，s と同じ範疇の検索語のうちその検索語から収集した検索クエリに t が含まれるものの個数 stf(s,t)によって判断する．

### ４．３．　推薦する検索クエリの算出と提示方法

4.1，4.2 節で述べた「思いつきづらさ」「得られる情報の予想しづらさ」の値の積を求め，その値の低さによって検索クエリの意外性を表す．Google サジェストから収集した検索クエリ<s, t>に対して以下の式でスコアを算出し，その値が小さい順に提示を行う．

$$score(s,t)=picf(s,t)\cdot stf(s,t) \quad (4)$$

### ５．評価実験

実験に使用する検索語は人名 4 語および組織名 3 語である．被験者は本学の学生 14 名である．評価は推薦された検索クエリと検索結果を見て，意外であると感じる度合いを 5 段階で回答してもらうことで行う．使用する検索クエリの数は 1 つの検索語につき 9 語または 10 語である．また，ベースラインとして Google サジェストから「検索語」と「検索語+スペース」によって得られた検索クエリを用いた．各検索語における評価値の平均を図 1 に示す．

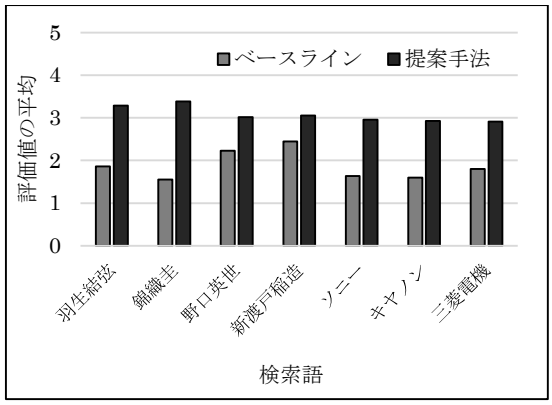


図 1．意外性の評価値の平均の比較結果

全ての検索語において提案手法はベースラインと比べ高い評価値を示した．また，2 つの手法について t 検定を行ったところ 1%水準で帰無仮説が棄却された．つまり，提案手法とベースラインは有意に差があることが認められた．

ここで，提案手法の評価値の平均は約 3 であり，平均で見た場合意外性のある検索クエリを推薦できていないように見える．これは，クエリによっては評価値として 1 および 2 が与えられるものも含まれているため平均としては低くなったためである．そこで，評価値の平均が 4.0 以上のクエリを調査すると，「キヤノン　ウィリアムズ」や「羽生結弦　ゲーム」などがあり，高い評価値のクエリが存在していることがわかる．また，提案手法により提示された検索クエリに対して被験者が回答として与えた全 476 個の評価値のうちの 41.8%が 4 以上であった．このことから，提案手法は一定数の意外性のある検索クエリを提示できていると考える．

### ６．まとめと今後の展望

本研究では，検索クエリに対する意外性を「思いつきづらさ」と「得られる情報の予想しづらさ」を指標として用いて定義し，意外性のある検索クエリの推薦手法を提案した．被験者実験により，ベースラインと比べて意外性のある推薦ができることが示された．今後は，より多くの検索語に対応できるようにするため，検索クエリの収集方法や Wikipedia 以外のシソーラスの活用について検討する必要がある．

### 参考文献

[1] 加藤亜由美他，“五感と関連するオノマトペを用いた意外性の高い飲食店推薦”，人工知能学会論文誌 30(1)，pp.216-228(2015)

[2] 伊藤雅弘他，“Wikipedia からの連想シソーラス構築プロジェクト”，第 20 回セマンティックウェブとオントロジー研究会 Wikipedia ワークショップ(2009)