

情報検索のための学習成長型対話システムの開発

- ユーザの嗜好を反映した情報検索の実現 -

1 背景

近年、ウェブ上の情報は爆発的に増加しているが、googleをはじめとする検索エンジンによって、ある単語(検索クエリ)が含まれるウェブページのみを高速に検索することが可能となっている。しかしながら、情報の増加に伴い、特定性の低い単語を検索クエリにした場合に数億ページの検索結果が得られることも少なくなく、それらを全て閲覧することは不可能である。そこで、より特定性を高めるために検索クエリを増やし、より適合性の高い情報のみに絞り込む試みが有効だが、そのためには適切な検索クエリを構築する技術が要求される。しかし、ユーザの検索技術は検索経験に依存することが報告されており[1]、検索に不慣れなユーザは不適切な検索クエリを過剰に設定し、適合性の高い検索結果が得られないという事態が起こりうる。

それに対し、検索エンジンは例えば以下の工夫を施しており、それぞれの利点と欠点を示す。

パーソナライズド検索 ユーザのウェブ閲覧履歴を元に、ユーザの嗜好に合わせて検索結果を並べ替える

利点 例えば計算機に関心のあるユーザに対しては、それに関連した記事を優先的に上位に位置づけることで、適合性の高い情報を見つけやすくする

欠点 閲覧履歴を収集するのに時間がかかる。並べ替えた結果のみ(または並べ替える前の結果とともに)を提示するため、ユーザにとって、嗜好が反映されたかどうかを判定するのは容易ではない

クエリ拡張 入力したクエリに対して、他のユーザがよく併記している別のクエリを追加クエリの候補として提示する

利点 検索に不慣れなユーザに対して、有用なクエリを提示できる

欠点 多くの他ユーザが共通して用いたクエリを利用するため、ユーザの検索要求に適合したクエリであるとは限らない

このような背景において、検索結果からユーザに必要な情報のみを絞り込む、あるいはシステムが能動的に推薦するといった技術が要求されている。

2 目的

本プロジェクトの目的は、ユーザの検索要求に対し、ユーザの嗜好をとらえた上での確かな情報のみを提示することである。そのために対話システムというインターフェースを採用する。対話の中で得られるユーザ固有の情報は計り知れず、それを検索エンジンに利用することで、ユーザの嗜好に合わせた情報検索を実現する。しかし、多くの対話履歴を得ていくには時間がかかるため、本プロジェクトではユーザの記述したウェブサイトなどからユーザ固有の情報を得ることで代用する。

将来的にはユーザとの対話履歴を学習することで検索エンジンとしての性能が上がり、それが対話を活性化させるという正のスパイラルを構築していく。

3 開発の内容

本システムは以下の順に動作し、システム発話の生成を行う。

1. ユーザの検索要求 (発話文) から検索クエリ (単語集合) を自動構築
ユーザはクエリとして単語ではなく、文を入力する。本機能はその文を解析し、検索クエリとなる単語を自動的に構築する。この際、特定性の高い単語、すなわちクエリとする難易度の高い単語は入力されないことを想定している (Figure 1).
2. 検索クエリ拡張
1で構築したクエリは特定性の低い単語をクエリとしているため、以下の2種類の方針でクエリを拡張する。
 - (a) ユーザ非依存のクエリ拡張
1のクエリのウェブ検索結果中で共起頻度の高い単語を拡張クエリとする。ユーザの嗜好に依存しない手法であり、従来のクエリ拡張に相当する。
 - (b) ユーザの嗜好に合わせたクエリ拡張
ユーザとの対話履歴や、ユーザの記述したウェブサイト、収集したブックマークなど、ユーザ固有の情報を利用して、1のクエリに対して、ユーザの嗜好を反映した単語をクエリとして拡張する。手法の詳細は [2] を参照されたい。
3. 複数の検索エンジンでウェブ検索を行う
2までで構築されたクエリに対して、様々な検索エンジンを用いて検索結果を得る。利用する検索エンジンは以下の通りである。
 - google
 - google news
 - yahoo
 - 楽天市場
 - Youtube
4. 検索結果の文書からシステム発話の生成
複数の検索結果の文書中の文章に対して、ユーザ発話および拡張されたクエリと最も類似した一文を選択し、システム発話として出力する。

これらの機能を統合した対話システムを、ウェブブラウザ上で専用のインターフェース (Figure 1) およびウェブサービスの twitter 上 (Figure 2) で利用することができる。



Figure 1: Ruby on Rails による専用インタフェース



Figure 2: twitter 上での本システム

4 開発成果の特徴

4.1 検索エンジンとの比較

開発成果の特徴を、検索エンジンにウェブ検索と商品検索のどちらを利用するかに大きく分け、従来の検索エンジンとの違いを述べる。

4.1.1 ウェブ検索

従来のウェブ検索では、例えば“プリウス”で検索した場合、オフィシャルサイトやニュース、Wikipediaといった、クエリに関する代表的な情報が出力される (Figure 3)。それに対



Figure 3: Google で“プリウス”を検索した結果

して、本システムはウェブ検索結果における検索順位をユーザ発話に基づいて再計算するため、個人のブログ中の一文がシステム発話として出力される場合が多い。オフィシャルの情報は広く知られているものが多いため、本システムの方がユーザとの対話を盛り上げられる可能性があり、かつユニークな情報を手に入れられる可能性が高い。

クエリ拡張を行った場合、オフィシャルサイト以外の様々なウェブサイトが検索されるため、ユーザにとって予想外の発話が得られる可能性が高くなる (Figure 4)。このように、従来のウェブ検索と比較すると、本システムはユーザの関心を引きやすい結果を提示できると考えられる。

4.1.2 商品検索の場合

ウェブ検索において商品検索を利用すると、ユーザの嗜好を利用するかどうかによって、結果に大きな違いが生じる。まず、楽天市場の日本酒カテゴリに対して、クエリ拡張を全く行

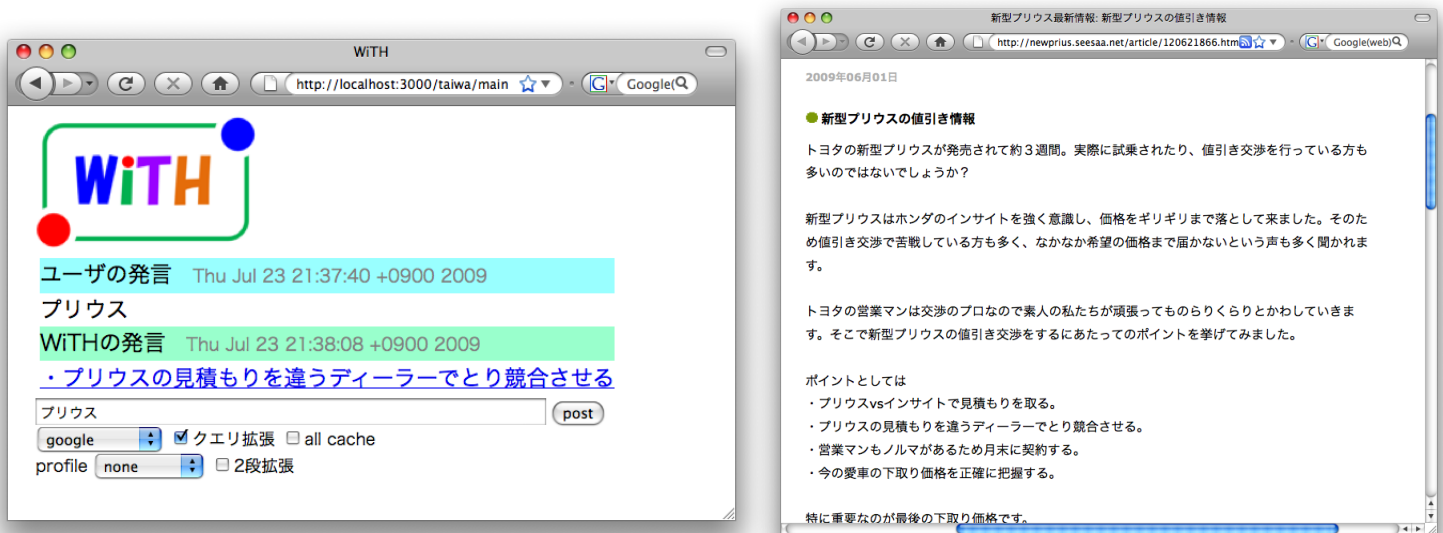


Figure 4: 本システムでウェブ検索を行った結果と、引用元のウェブサイト

わない場合の結果を Figure 5 に示す. 次に, 勝屋久 PM の嗜好に合わせてクエリ拡張を行



Figure 5: クエリ拡張を全く行わない場合の結果と、その引用元のウェブサイト

い, 拡張されたクエリに対して, 関連する単語をさらに拡張した場合の結果を Figure 6 に示す. ここでは, フランスで制作され逆輸入された無添加純米原酒の日本酒が提示された. 実際に勝屋久 PM に確認をしたところ, 日本酒では“無添加純米酒”が好みであるということであり, この日本酒は嗜好に合う可能性が高く, 大変興味深いということであった.

最後に, 他ユーザの嗜好を利用した場合に, 結果がどう変化するのかを示す. 日本酒の好きな人のブログとして, “おたくさは日本酒大好き!”¹というサイトを利用して, ユーザの嗜好データを構築し, 同じように対話を行った. その結果を Figure 7 に示す. 著者に確認を取ることは行っていないが, 少なくとも勝屋久 PM の場合と大きく嗜好の異なる日本酒が提示されていることは明らかである.

¹<http://ameblo.jp/nipponnosakeumai/>

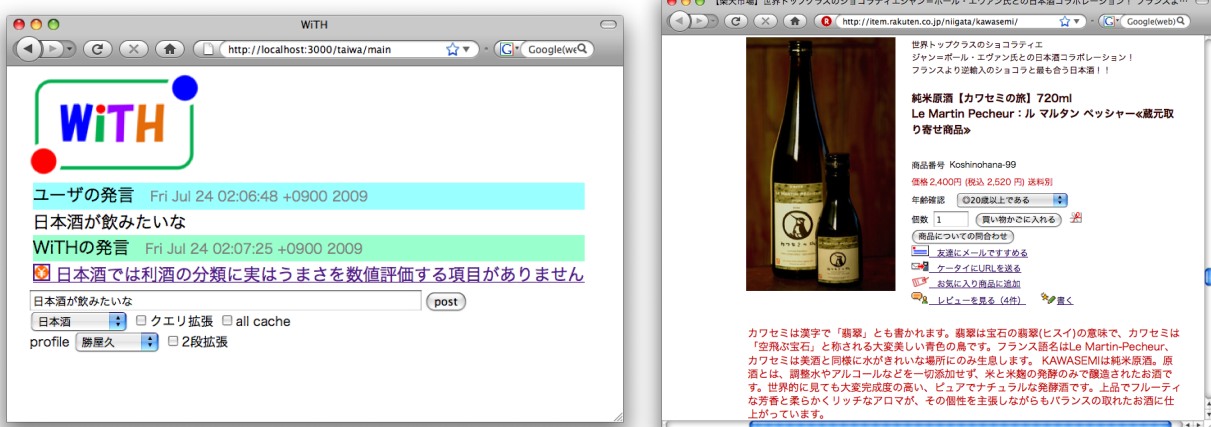


Figure 6: 勝屋久 PM の嗜好を利用して検索を行った結果と、その引用元のウェブサイト

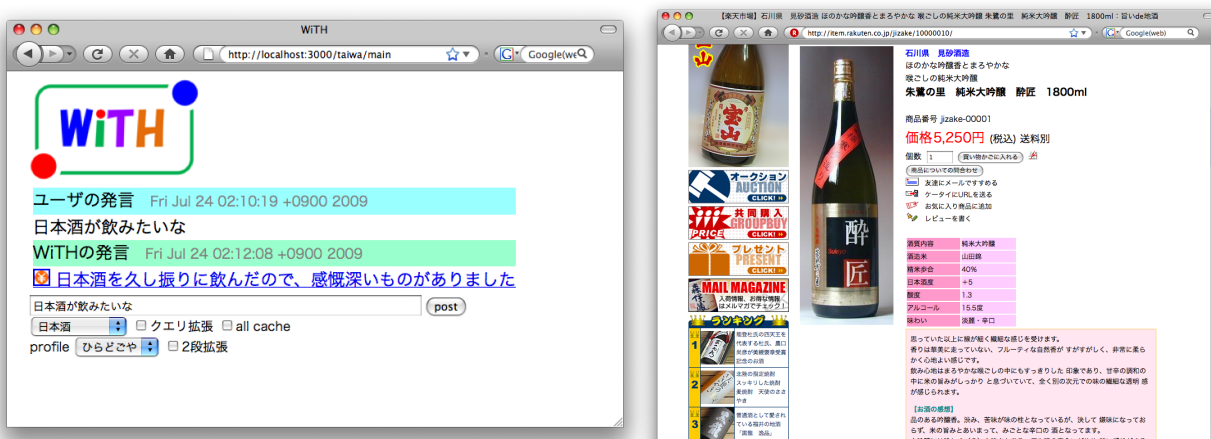


Figure 7: “おたくさは日本酒大好き！”の著者の嗜好を利用して検索を行った結果と、その引用元のウェブサイト

5 期待される効果

本システムは twitter 上で利用できるため、“どんな話題に対しても気の利いた返事を返すことができる”という点が大きな特徴である。ユーザは必ずつぶやき返してくれる follower として、本システムを楽しむことが出来る。また、つぶやきの履歴からユーザの嗜好を推定することが出来るので、嗜好の似た別のユーザを紹介することも可能である。mixi をはじめとする SNS や、twitter といった、従来の電話やメールといった通信手段の枠を超えた通信が活発に行われている今日において、そのユーザたちを繋げていくシステムとして本システムの発展性は高いと考えられる。

また、学術面においても対話システムを実用段階まで引き上げる試みの一つとして、様々な知見が得られる可能性が高い。

6 普及の見通し

利用する検索エンジンの性質や、利用している要素技術の性能に影響されるなど、システムの頑健性にはまだ多くの改善点があげられる。しかし、検索エンジンの次のステップとしてパーソナライズは必須であると考えられ、googleでは部分的ではあるがそれが実現されつつある。本システムは対話システムであると同時に、パーソナライズド検索エンジンとしての性質も持っており、両要素の結合的発展を促せる可能性を持っていると考えられる。

twitterのようなサービスは今後も発展が続くことが予想され、その中でユーザ同士をどう繋げていくのかは大きな課題である。本システムはそのアプローチの1つとして開発を継続していく。

7 開発者名

水野淳太 奈良先端科学技術大学院大学

村田祐一 テルモ株式会社

References

- [1] 中島悠, 土方嘉徳, 西田正吾. 検索経験と領域知識の www 情報検索行動に与える影響. 第2004 巻, pp. 25-32. 社団法人情報処理学会, 20040521.
- [2] 水野淳太, 村田祐一, 勝屋久. ユーザの嗜好を反映したクエリ拡張を用いた情報検索・推薦システムの開発. 楽天研究開発シンポジウム 2009, 2009.