

はさみこむ検索：行間を埋める検索システムの開発

はさんでポンッ

1. 背景

近年、さまざまな情報検索システムが提供されており、情報検索は生活に必要不可欠なものになりつつあり、ユーザが明確なターゲットをもつような検索においてはあまり問題がなくなってきた。一方で、検索の対象や意図が曖昧な場合の検索は非常に難しく、ユーザは求める情報にたどり着くことができない。例えば、AとBの間にある何らかのCを探すような検索は多くありえるのだが、現在の検索サービスでは実現されておらず、そのCを探すことができない。以下にそうしたAとBの間にある何らかのCを探すような検索の具体例を示す。

- 織田信長に関することと言えば、桶狭間の戦いがある、最後には本能寺で死んだことは知っているがその間には何があったのだろうか？
 - 親と飲みに行きたいが、近所の学生向けの居酒屋に連れて行くには忍びない。かといって研究室の近くにあるちょっとセレクトな居酒屋では高すぎる。2つの間くらいの居酒屋はないだろうか？
 - この本は簡単すぎるし、あの本は難しすぎる、2つの本の間くらいのものはないのか？
 - カレーを作る時、野菜を切ることと煮込むことは知っているけど、その間に何をやるのだろうか？
 - 結婚式の招待状がきた後、結婚式に参列するまでにどうすればいいのか？
- といったような具合である。つまり、2つの知っていることを使って間にあてはまるものを探すということである。このように2つのもの間にあるものを求めたい、といったシチュエーションはよくあるが、その発見を支援してくれるようなシステムは存在しない。

2. 目的

そこで本プロジェクトでは、ある2つの入力を与えると、その2つに挟まれたものを出力する、といったようなシステムを提案する。ここでいう「はさみこむ」とは、本手法として、AとBを与えてそこから挟みこむことにより対象を探す検索を実現するものである。そこで、Webからの知識発見を行う。

入力は2つのオブジェクトやイベント名などで、2つの入力間に存在する軸（入力によって変化する）で評価して出力するようなシステムの実装を行う。

このシステムでは、ユーザのわかっている状況やものを入力とすることにより、検索キーワードの洗練やウェブページの閲覧、横断的検索を行わずに与えられた2つのイベント、オブジェクトの間にあるものをあらゆる軸の観点からユーザは得ることができる。

3. 開発の内容

本プロジェクトでは、2つの入力を与えるとその間にあたるようなモノをランキングづけて出力するWebサービス、「はさんでポン」を実装した(図1)。結果をある観点で絞り込みたい場合は2つの入力に加え、「長さ」「高さ」「芸能人」「関東」などの絞り込むような語を入力することができる。



図1 検索結果画面

検索結果画面では、間に当てはまる候補がリストで表示される。上から順番に出現位置と出現頻度に基づいたランキングにより、候補が並んでいる。

出てきた結果がおもしろい！とユーザが感じた場合、間のモノリスト右側にある、「この結果はおもしろい！」ボタンを 押すことにより、評価がその検索キーワードに対して反映され、twitter の hasande_pon アカウントで自動的にその面白いと感じた検索キーワードが投稿される。

間のモノ候補の右側に表示されているのは、間のモノ候補と一緒によく記述されているモノで、タグクラウド形式で表示している。



図2 観点別結果表示

図2は観点別結果表示の画面である。これは、スニペット間でクラスタリングを行い、観点の分離を試みている。上図はうまくいった場合の例であり、ワカナゴとブリの間を求めた際に、関東での呼ばれ方の間(左上)、九州での呼ばれ方の間(左下)が分離出来ている。

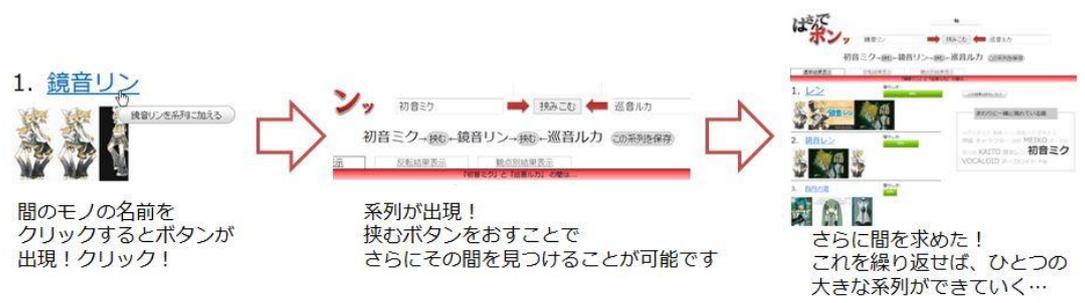


図3 どんどん挟む

最初に入力したキーワードとその結果から得た間のモノの間のモノをさらに求めることができる。これを繰り返すことにより、ひとつの大きな系列を作ることができる(図3)。作った系列は名前を付けて保存することができ、保存した系列はトップページで紹介される。そして、このトップページに表示された系列をさらに別の人が拡張していくことも可能である。

4. 従来技術（または機能）との相違

価格帯や経路といったものならば、何円から何円、どこからどこまでといったような価格や場所で挟む検索はすでにあるが、それらは決まった属性値や持っているデータベースにあるデータしか検索できない。提案システムでは、基準となる軸が動的に変化し、入力は具体例であるという点が新しい。値段や場所といった決まりきったものしか入力できない既存のシステムとは一線を画す。

5. 期待される効果

モノとモノの比較検索への応用が可能であると考えられ、また挟みこむ検索を再帰的に繰り返すことにより、系列を発見するといったことができ、ある種の系列発見を自動でできる可能性がある。系列がわかれば、系列の流れにそったページ推薦など応用が考えられる。

また、本システムでは、一般のリストとは異なり、どれが最も間らしいか、といったことがわかる。例えば、徳川将軍一覧のリストがあった場合、それを見ただけではどの将軍が有名なのか、といったことが一目ではわからないが、挟みこむ検索を用いれば、2つの将軍の間で最も有名なものが抽出できる。挟みこむ検索を再帰的に行うことで、重み付きのリストというものを求めることができ、新たなリストを作ることができると考えられる。

6. 普及（または活用）の見通し

Web上でサービスを公開 (<http://hasande.com/>) しており、現在ある程度のユーザを獲得している。ブラウザを問わず動作することを確認しているので、広報活動を進めていけば多くの人に利用してもらえると考えている。また、twitter アカウントをフォローしてもらうことにより、他の誰かのおもしろい検索例がタイムラインに流れてくるため、再活性を促せられる。

7. クリエータ名（所属）

旭 直人 （京都大学大学院情報学研究科 社会情報学専攻）

（関連 URL）

挟んでポン <http://hasande.com/>

twitter 上アカウント http://twitter.com/hasande_pon