

Web 検索結果におけるスニペット表示の一手法

目本信太郎[†]

[†] 岡山県立大学情報工学部 〒719-1197 岡山県総社市窪木 111

E-mail: †meki@cse.oka-pu.ac.jp

あらまし 多くの場合、検索エンジンは検索結果として URL のほかに 3 行程度のテキスト (以下スニペット [3] と呼ぶ) を返してくることが多い。必要とする情報が日付や場所、人の名前などの単純なデータの場合、ユーザはこれらの情報をスニペットを読むだけで得られることも多い。しかしながら、キーワードがお互いに離れて出現する場合には、スニペットは短いいくつかのテキストの断片の集合となるので、ユーザは必要な情報をスニペットから得られないことも多い。そのような場合、検索エンジンは必ずしも 3 行のスニペットを返すことにこだわるべきではなく、必要に応じて 3 行より長いスニペットを返すべきである。しかし、そうすると検索結果が長大になり、検索結果の一覧性が損なわれる。本稿では、3 行より長いスニペットを、小さな別ウインドウに表示することによって、検索結果の一覧性を損なうことなく、より豊富な情報をスニペットから得られるようにする手法を提案する。

キーワード 検索エンジン, 検索結果, スニペット

Shintaro MEKI[†]

[†] Faculty of Computer Science and System Engineering, Okayama Prefectural University

111 Kuboki, Soja, Okayama, 719-1197 Japan

E-mail: †meki@cse.oka-pu.ac.jp

1. ま え が き

インターネットは現在の情報化社会の情報インフラとなった。インターネットでは情報は Web サーバ上に存在することが多い。しかしながら、インターネットには多くの Web サーバが存在し、それぞれの Web サーバも多くの Web ページを持っている。ユーザはどの Web サーバが知りたい情報を持っていて、どの Web ページにそれが書かれているか、すなわち位置情報 (URL) を知っていなければいけない。さもなければユーザはインターネット上の情報を利用することができない。

それゆえ多くのユーザは検索エンジンを使って URL を検索する。検索エンジンはあらかじめ作ってある索引を使って URL を検索し、多くの場合 URL のほかに約 3 行のテキスト (スニペットと呼ばれる) を URL とともに返す。ユーザはこれらのスニペットに目を通して、返された URL の中からいくつかを選択し、どの Web ページを実際に閲覧するかを決定するのである。そして、ユーザは実際にそれらの Web ページを閲覧することによって必要とする情報を得るのである。

しかしながら、必要とする情報が日付や場所、人の名前などといった単純なデータの場合、スニペットがそれらの情報を含んでいることが多く、それゆえ返されたスニペットを読むだけで必要とする情報を得ることができる。

図 1 を参照して頂きたい。これらの検索結果は、数年前まだ Bill Gates がマイクロソフトのチェアマンをしている時、Google [2] を使って、” マイクロソフト” と” チェアマン” をキーワードとして検索を行った時の、検索結果の上位 5 件である。この場合ユーザは実際の Web ページを閲覧する必要はない。なぜならば、スニペットを読むだけでマイクロソフトのチェアマンが Bill Gates であることを理解することができるからである。

Web ページの中には非常に長大で重いものも存在し、そのような Web ページを閲覧するには長い時間が掛かる。それゆえスニペットを読むだけで必要とする情報を得ることができるのであれば理想的である。

このアイデアはキーワードが Web ページ上でお互いに接近して現れるときだけうまく働き、キーワードがお互いに離れて現れるときにはあまりうまく働かない。なぜならば、そのような場合にはスニペットは多くの短いテキストの断片になり、スニペットから有効な情報を得られなくなるからである。

本稿では、3 行より長いスニペットを、検索結果とは別の小さなウインドウに表示し、検索結果の一覧性を損なうことなく、より豊富な情報を検索結果のスニペットから得られるようにする手法を提案する。

本稿は以下のように構成されている。次の章では基本的なア

(1) 【楽天市場】 マイクロソフト Halo3 リミテッドエディション【Xbox360 用 ...

マイクロソフト Halo3 リミテッドエディション【Xbox360 用】【税込】 9UE-00010 ヘイロー 3 リミテッド [9UE00010 ヘイロー 3 リミテッド]【デジタル 0702】. 商品番号, 4988648505889-52-3916. メーカー希望小売価格, 7140 円 (税込). 販売価格, 6380 円 (税込) 送料 ...

item.rakuten.co.jp/jism/4988648505889-52-3916-n/

%E6%98%AF%E9%80%99%E7%89%87%E5%97%8E~%E5%A6%82%E6%9E%9C%E...

(2) eetimes.jp マイクロソフト社、「WinHEC」で「Windows Home Server」の ...

米 Microsoft (マイクロソフト) 社は 2007 年 5 月 15 日 (米国時間)、米国のカリフォルニア州ロサンゼルスで開催されたハード ... Microsoft 社のチェアマンである Bill Gates 氏は WinHEC の基調講演の中で、「Windows Vista は、パソコン周辺機器を大きく変えつつ ...

www.eetimes.jp/contents/200705/18865_1_20070520163200.cfm

(3) マイクロソフトカップ 2006 - MSN スポーツ

朝日新聞社は 55 年間、全国社会人大会を主催者の一つとして応援させていただき、3 年前からマイクロソフトカップを一生懸命応援させていただいています。新聞読者の一人でも多くにこの大会の意義深さを 中締めあいさつ ■ トップリーグ 真下昇チェアマン ...

inews.sports.jp.msn.com/rugby/mscup2006/pre/vip01.html

(4) マイクロソフトカップ 2006 - MSN スポーツ

本件に関し、ジャパンラグビー トップリーグチェアマン 真下 昇 (ましも・のぼる) のコメントを以下の通り、お知らせいたします。 ... このページ上に表示される記事、あるいはリンク先の記事内容は MSN およびマイクロソフトの見解を反映するものでは ...

inews.sports.jp.msn.com/rugby/mscup2006/news/20060102-16.html

(5) 米マイクロソフト、Xbox を公開 — ネット — マイコミジャーナル

米ラスベガスで開催されている 2001 International CES で、米マイクロソフトのチェアマン&チーフソフトウェアアーキテクトのビル・ゲイツ氏は基調講演を行い、その講演において新しいゲームマシン「Xbox」を公開した。この Xbox は Intel 製の 733MHz の CPU、 ...

journal.mycom.co.jp/news/2001/01/07/12.html

図 1 簡単な質問の検索結果

アイデアについて述べる。第3章では実験結果と考察について述べる。最後の章は本稿の結論である。

2. 基本的なアイデア

前の章では、ユーザは実際の Web ページからではなく、スニペットから必要な情報を得られることがあるということを書いた。しかしあまりうまくいかないこともある。

この章では、われわれは検索結果を表示する方法を少し変更することを提案する。

多くの検索エンジンは、URL と一緒に高々3行程度のスニペットを検索結果として返す。そこで、ユーザが3行のスニペットではよくわからないと感じた場合のために、検索エンジンが3行より長いスニペットを返すようにするのである。

キーワードがお互いに接近して Web ページ上に現れるときは、スニペットは一連の文章となるので、検索エンジンが3行だけのスニペットを返すのは適切である。しかしながら、もし検索エンジンが3行より長いスニペットを返すのならば、スニペットが必要な情報を含んでいる可能性はより高くなるはずである。

すなわち、スニペットは必ずしもすべてのキーワードを含んでいるわけではないが、スニペットが長くなるのにしたがって、スニペットが必要な情報を含んでいる確率は高くなる。それゆえ、長いスニペットは必要な情報を含んでいることが期待できるから、より望ましいと考えられる。

なお、本稿では検索結果の一つの表示法を提案しているのであって、新しい検索アルゴリズムを提案しているのではない。だから、検索エンジンは自分で新たに作成してもよいし、既存の検索エンジンの検索結果を利用してもよい。

3. 実験結果

普通ならばプロトタイプシステムを作成して実験を行うところであるが、まだ実装は行っていないので、手作業で実験を行い、どのようなイメージのシステムになるのかを伝えることにする。

図2は、Google を使ってキーワードを「タウリン」と「効能」として、通常の検索を行った結果である。各検索結果には約3行のテキストのスニペットが含まれているが、それらは短すぎるのでそれらの意味するところを理解するのは困難である。

そこで、第1位の検索結果のスニペットをリンクテキストとしておき、それをクリックしたとき、3行より長いスニペットが、小さな別ウィンドウに表示されるように、検索結果に手を加えた。図3はそのスニペットをクリックして、長いスニペットが小さなウィンドウに表示されているところの画面のダンプである。

JavaScript を使うと、新しく開くウィンドウのサイズを指定することもできるらしいが、この実験では Google の検索結果に手を加えただけなので、図2の Web ページは HTML で記述されており、新しく開くウィンドウのサイズは指定できない。そこで手作業でウィンドウをリサイズしたものが図3である。

この図を見ると、検索エンジンの検索結果の一覧性が損なわ

れることなく、より長いスニペットが表示されており、実際の Web ページを参照することなく、知りたいことがわかるようになっていくことが理解できる。

なお、本稿ではスニペット表示の一手法を提案しているのであって、小さなウィンドウに表示されているスニペットは、適当に選んだに過ぎず、これが最適なスニペットであると主張しているわけではない。

4. 結 論

本稿では、我々は長いスニペットを小さな別ウィンドウに表示することを提案し、手作業による実験結果を示した。

実験結果から、長いスニペットはその意味するところを理解することができ、ユーザは検索結果の Web ページを順次閲覧する時間を省くことができることがわかる。

我々はたった一つの実験結果を示しただけだが、提案した手法はスニペットを長くして別の小さなウィンドウに表示することであり、それゆえスニペットの情報は損なわれることはない。スニペットの持つ情報はおそらく豊かになるであろうし、少なくなることはまったくない。したがってどんな場合にもユーザは普通の検索結果よりも多くの情報を、提案した手法による実験結果から得ることができると考えられる。

以前筆者らは、モバイル環境において、検索結果のスニペットを3行より長くする、という手法を提案した[4]。その手法には、検索結果が長大になり、一覧性が損なわれてしまうという欠点があったが、本稿で提案した手法には、そのような欠点は存在しない。

また、デスクトップ環境にも、モバイル環境にも適用可能である。

今回われわれが行った実験では、小さなウィンドウに表示されるスニペットは適当に選んだだけなので、どのようなスニペットが最適なのかを明らかにすることは、今後の検討課題である。

文 献

- [1] Adobe: <http://www.adobe.com/>.
- [2] Google: <http://www.google.com/>.
- [3] 高見真也, 田中克己: ウェブページに対する定量的評価の視覚化による情報検索支援, DBWeb2006, pp.67-73, 情報処理学会, 2006.
- [4] 目本信太郎: モバイル環境における Web 検索結果の柔軟な表示法, DBS-144, pp.115-121, 情報処理学会, 2008.
- [5] 興石純子, 佐藤哲司: モバイル環境に適した検索結果の提示法に関する研究, 日本データベース学会論文誌, Vol.8, No.1, 2009.



図 2 検索結果

