社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－DBS－143（48）
2007／7／3

検索目的に基づくスニペットの動的再生成による ウェブ検索結果の個人適応化

高見 真也† 田中 克己†

† 京都大学大学院 情報学研究科 社会情報学専攻
〒606-8501 京都市左京区吉田本町
E-mail: †{shie,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし ウェブ検索エンジンは，ウェブページを発見するためだけではなく，知識やサービスにアクセスするための道具としても使われるようになってきている．そのため，ユーザが入力した検索語に応じて，広告コンテンツやサービスへの誘導リンク等も検索結果に表示されるようになった．しかし，検索語だけでますます多様化するユーザの検索目的を把握することは難しく，検索語で表示内容，表示順序が一意に決定される検索結果では，求める情報までの経路が最適化されているとはいえない．我々はユーザの検索目的に適した検索結果を提供するために，検索結果として示すべき概要文（スニペット）を二種類の軸で分類した．そして，最適化された検索結果を実現するためのスニペットの動的再生成を紹介する．我々は，このように検索目的に適した検索結果表示を行うことを SPO（Search Purpose Optimization）と呼んでいる．

キーワード スニペット，検索結果の個人適応化，ウェブ情報検索

Customizing Web Search Results by Dynamic Re-Generation of Web-Snippet based on Search Purpose

Shinya TAKAMI† and Katsumi TANAKA†

† Department of Social Informatics, Graduate School of Informatics, Kyoto University
Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
E-mail: †{shie,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract Web search engines are used as a tool not only to find web pages but also to access some knowledge and services. Therefore, the links to advertising contents and services etc. came to be displayed in the search results according to the search query that the user had input. However, it is difficult for such systems to know user's search purpose because it is more and more diversified. The route to target information is not necessarily optimized in the search results when the search query defines both the ranking and the content in the search results. We classified the outline (Web-Snippet) in the search result by two criteria to offer a suitable search result for user's search purpose. Then we introduce dynamic re-generation of the Web-Snippet to achieve the optimization of search results. We call such display of search result suitable for the search purpose SPO (Search Purpose Optimization).

Key words Web-Snippet, Customization of Search Results, Web Information Retrieval

1. はじめに

インターネットユーザの増加に伴い，一般のユーザが容易にコンテンツを制作できることで，膨大な情報がウェブ空間に溢れるようになった．本来，ウェブページを探すために利用されていたウェブ検索エンジンは，ウェブページの爆発的な増加により，登録制ディレクトリ検索型からロボット自動収集制キーワード検索型へと移行してきた．また，ウェブページを構成する素材が高度化したことで，ユーザの検索目的も多様化し，ウェブページそのものではなく，ウェブページを構成するテキストの一部や画像などを探す目的でも使われるようになってきている．マーケティング理論の分野でも，消費者の行動プロセスを 5 つのステージに分類した AIDMA (Attention → Interest → Desire → Memory → Action) モデルから，近年ではインターネットを意識し，購買前に検索を行い，購買後にブログなどの CGM (Consumer Generated Media) で情報を共有する

AISAS (Attention → Interest → Search → Action → Share) モデルが提唱されている．このように，ウェブ情報検索はインターネットユーザにとって大変重要な利用目的の一つとなっている．

ウェブ上で何らかの情報を探す場合，我々は通常ウェブ検索エンジンに検索語の組み合わせをクエリとして入力し，返された検索結果のうちごく限られた上位のものだけを対象に，目的とする情報が含まれていそうなウェブページを探す作業を繰り返し行っている．一般にデータベースへの問い合わせをクエリと呼ぶが，本論文では情報検索の場合に限定し，ウェブ検索エンジンに与える検索語の組み合わせをクエリと呼ぶことにする．多くのウェブ検索エンジンは，検索結果として，タイトル，URL および概要文（以下，スニペットと呼ぶ）を含むウェブページのリストを返す．そのようなシステムにおいて，クエリによく適合するウェブページが検索結果の上位にランクされることはもちろん重要であるが，たとえまったく同じクエリが入力されたとしても，その目的により，システムが返すべきスニペットは同じであるとは限らない．そこで，我々は検索目的に応じた検索結果の個人適応化を実現することで，ウェブ情報検索を支援できるのではないかと考えている．

2. 検索結果の個人適応化

2.1 クエリ拡張とクラスタリング

ウェブ情報検索に関する研究分野では，HITS [1] や PageRank [2] といった優れたランキングアルゴリズムがいくつか提案されている．それらは，ハイパーリンクの構造解析による客観的な評価基準をもとに，あるクエリを含む数千，数万のウェブページ群から多くの人々が求めるものを上位にランクする手法としては，十分価値のある結果を提供している．しかし，多くの場合，検索の目的はウェブページの URL リストを取得することではなく，あるウェブページ上に存在する何らかの情報を見つけることにある．そのため，ウェブ検索エンジンが返す結果の上位に含まれるウェブページ群が目的にそぐわない場合，目的のウェブページがより上位にランクされるように，クエリを再考し再検索が行われることが多い．そこで，クエリに追加または削減すべき単語の提案などを行うことで，ユーザの意図に適した検索結果を提供しようとする研究が行われている．

また，再検索は行わず，検索結果上位 k 件を対象にして，クラスタリングやリランキングを行うことで，ウェブ情報検索の支援を行おうとする研究が注目されている [3] [4] [5]．検索結果のクラスタリングは，対象とするものがウェブページかスニペットかで二種類に分類することができる．ウェブページを対象としたクラスタリングの場合，各ウェブページ毎に特徴ベクトルを生成し，その類似度を評価する方法などが用いられる．しかし，近年のウェブページは，複数のブロックに種類の違うコンテンツが配置されていることも多く，またページの単位で話題が区切られているとは限らない．そのため，ウェブページに複数の話題が存在すると類似度が低くなってしまう可能性がある．また，スニペットを対象としたクラスタリングの場合，特徴を評価するには情報量が少なすぎるといった問題や，スニペットがウェブページのどこから抽出されたものであるかによって，精度が左右されるという問題がある [6] [7]．

スニペットの各要素がウェブページのどこから抽出された断片であるかは大変重要な情報である．なぜなら，スニペットの各要素に含まれる検索語のウェブページ内での位置が，その意味や重要性に深く関係しているからである．ほとんどのスニペットは検索語を少なからず含むが，スニペットとしては抽出されていなくとも，他にそれら検索語を含む断片が対象のウェブページには存在している可能性がある．さらに，複数のウェブページに類似した断片が存在したとしても，それらがスニペットとして抽出されなければ，クラスタリング時に類似しているとは見なされない．つまり，検索結果のクラスタリングを行う際には，ウェブページの場合は対象とする範囲，スニペットの場合はその生成手法に注意する必要がある．

2.2 スニペットの改良

株式会社アイレップ SEM 総合研究所らの「インターネットユーザの検索行動調査」[8] によると，ウェブ検索エンジンの利用者が検索結果の中から実際にウェブページを確認するかどうかを決定する判断材料として，クリックする場合はタイトル，スニペットの順に，クリックしない場合はスニペット，タイトルの順に内容を確認する傾向にあることが報告されている．米国における同様の調査では，その順序が反対になっているが，それは大きな問題ではなく，ウェブページの特性を実際に開いて確認する前に判断する材料として，スニペットが重要な役割を果たしているということが示されていることに注目したい．このように，検索結果におけるスニペットはウェブページの特性を推測する際に重要な情報と見なされており，我々はその役割をよりよく果たすためにスニペットを改良することに着目した．

現行のウェブ検索エンジンにより生成されるスニペットの多くは，ウェブページから断片的に抽出された検索語を含むテキストにより構成されるのであって，必ずしも意味的に抽出されているわけではない．つまり，スニペットは検索語の組み合わせにより動的に生成されるため，概要文として見た場合，ウェブページの全体を包括する内容ではなく，ほんの限定された一部の内容だけを示している可能性がある [9]．また，断片的に抽出されたテキストをウェブページ内での出現順に単純結合しただけのスニペットは，意味的なつながりを持たず一貫性に欠ける概要文となることが多い．

現存するウェブページの多くは，文字情報だけではなく，画像を含むマルチメディアコンテンツを含んでいる．HTML や XML の構造は，ときに文脈における重要性や意味に影響を与える場合がある．例えば，ウェブページ内での意味や重要性は，その単語がタイトル部分に存在するか，本文に使用されているかによって違うため，その特性を利用して詳細度の違う 2 つの単語の関係を抽出しようとする研究もある [10]．一方で，HTML などの構造化テキストであるウェブページから生成されるにも関わらず，スニペットは文字情報だけからなる．そのため，スニペットは人間が読む事によってのみ理解され得るコンテンツである．

このように，現行のスニペットはウェブページの特性を定量

的には表現しておらず，クエリが決定されると一意に決定されるため，ユーザにとってはクエリ依存で静的な概要文である．また，人間がそれらを読むことでしか理解出来ない．しかし，各検索語がウェブページのどこに存在しているかなどの情報を獲得する事はそれほど難しいわけではない．また，クエリに依存しない要約との違いを評価することで，著者の意図とユーザの目的の適合度を示すことができる．そのような情報は，ウェブページの特性を定量的に評価しており，我々がウェブページの全容を推測する際に役立つ．そのため，ウェブページに関する視覚化された定量的評価をスニペットに付加したり，クエリが同じ場合でもユーザの目的に適したスニペットを動的に提供することで，ユーザがウェブページの特性を推測する作業を支援することができると我々は考えている．

3. スニペット生成手法の分類

本来，図書館における検索の目的は本を探すことであり，様々な動機があるにせよ検索対象は本であった．検索を行うための情報も本の内容すべてが対象になっているわけではなく，著者名や内容の一部などのメタデータを利用している．しかし，ウェブページはその内容すべてに容易にアクセスでき，情報提供以上の価値を発信するようになったことで，ウェブ検索エンジンはウェブページを探すためだけの道具ではなくなってきている．ほとんどの場合，その検索対象はウェブページであるが，多くの場合そのウェブページ上に存在する情報の一部やサービスを利用するために検索が行われている．検索目的は，知識の獲得とサービスの利用に大別されるが，我々は前者の目的のウェブ情報検索を支援することに注目した．

知識の獲得が検索目的の場合，目的の知識を含むウェブページを提示できれば最適な検索結果となる．しかし，目的の知識を含むかどうかをシステムが判断することは極めて難しい．そこで，ウェブ検索エンジンは，目的の知識を含むかどうかをユーザの判断に委ねるために，タイトルやスニペットといった判断材料を検索結果として提供している．

我々はユーザの検索目的に適した検索結果を提供するために，スニペットの生成手法を二種類の軸で分類した．一つ目の軸は，スニペットを生成する際に考慮するウェブページの数である．単体独立型は一つのウェブページから得られる情報だけで生成するタイプで，全体依存型は検索結果などに含まれる他の複数のウェブページ集合から得られる情報を考慮して生成するタイプである（図1）．

もう一つの軸は，ウェブページからスニペットを生成する際に，内容の包括性を考慮するかどうかである．断片集約型は検索語を含むといった何らかの基準で抽出された断片をまとめてスニペットにするタイプで，包括要約型は検索語などには依存しない全体の内容を包括した要約としてのスニペットを生成するタイプである．このように，単体独立型または全体依存型，断片集約型または包括要約型を基準に分類した場合，スニペットの生成手法は図2のI型からIV型に整理することができる．

I型は単体独立型かつ断片集約型で，既存のウェブ検索エンジンの多くがこの手法を採用している．検索対象についての知識が少ない場合，手がかりとして入力された検索語を含む周辺情報を提供すべきであり，このような場合はI型のスニペットが適している．例えば，検索語として「京都」と「湯豆腐」が入力された場合，ユーザの検索目的は湯豆腐が食べられるお店を探すことかもしれないし，湯豆腐に使われる豆腐のことが知りたいのかもしれない．そこで，I型の生成手法を採用することで，「京都で湯豆腐を食べるなら〜がおすすめ」「京都の〜では湯豆腐に嵐山の〜というお店の豆腐が」といった内容のスニペットを提供することができる．

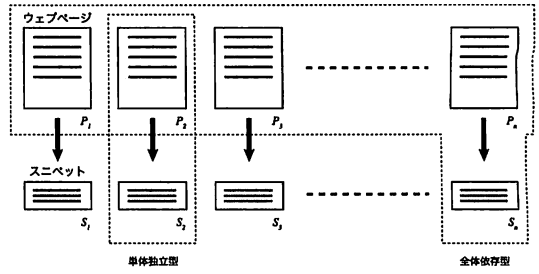

図1 単体独立型と全体依存型

Fig. 1 Single-independent Type and All-considered Type

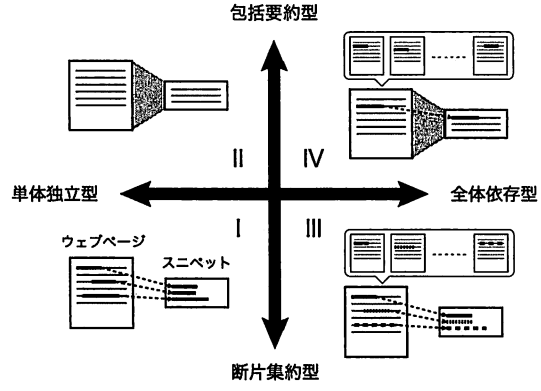

図2 スニペット生成手法の分類

Fig. 2 Classification of Generation Method of Web-Snippet

II型は単体独立型かつ包括要約型である．検索対象についての知識がいくらかある場合，特定の種類のウェブページを探している場合が多いため，その場合はウェブページの種類を推測しやすい包括的な情報を提供すべきであり，II型のスニペットが適している．例えば，検索語として「京都」「湯豆腐」「食べる」が入力された場合，ユーザの検索目的はおそらく湯豆腐料理店を探すことだと思われる．このとき，「京都の湯豆腐料理店一覧」といった内容がスニペットとして含まれている場合，いくつかのお店が紹介されているウェブページであると容易に推測出来る．このようなテキストは，必ずしも検索語を含むとは限らない．

ウェブ情報検索の目的が，検索結果に含まれる複数のウェブページを比較したり横断的に確認することである場合，それぞれのウェブページを独立に評価するのではなく，他との関係も考慮する必要があり，III型またはIV型のスニペットが適して

いる．III 型は全体依存型かつ断片集約型であり，他のウェブページには存在しない排他性の高い断片をスニペットとして抽出する場合，検索結果全体を見れば特徴的な情報を把握することができる．例えば，検索語として「京都」と「湯豆腐」が入力された場合，「京都で〜が食べられるのは当店だけ」といった内容を含むスニペットを提供することができる．また，全体依存型かつ包括要約型である IV 型のスニペットでは，他のウェブページと相関性の高い部分が示されることで，所在地や営業時間の比較などを行うことができる．

4. 検索目的に応じたスニペット生成

4.1 単体独立型

現行のスニペットの多くは，クエリがユーザの目的を表していると考えられることから，ウェブページから抽出された検索語を含む断片の組み合わせで構成されている．スニペットを生成する際には，キーワード抽出の手法を適用するために，各文を一つの単位として取り扱い，クエリに適合した重要文を抽出することが基本的なアプローチとされている [11]．このようなクエリ適合度によって評価された重要文抽出によるアプローチは，クエリ依存型抽出手法によるスニペット生成と見なすことができ，図 2 における I 型に該当する．

ユーザの目的を表すクエリ適合度を評価基準とした重要文抽出と比較して，クエリに依存しない要約は，著者の意図を反映するものである．このようなクエリ独立型要約手法により生成されたスニペットは，図 2 における包括性を重視した II 型となる．コンピュータによるクエリ独立な要約生成に関する研究は，古くから行われており，様々な手法が提案されている [12] [13] [14]．要約をその機能をもとに分類すると，以下に示す二種類のタイプに分けることができる [15]．

- 指示的（Indicative）
- 報知的（Informative）

指示的な要約とは，原文が読むべきものかどうか，自分の関心に合うかどうかなど，目的への適合性を判断するために原文を参照する前の段階で利用されるものである．また，報知的な要約とは，原文の代わりとして利用されるものである．そのため，検索結果におけるスニペットは指示的な要約としての役割を担っているが，クエリ依存による手法では指示的，クエリ独立による手法では報知的な側面が強いスニペットが生成されると考えることができる．コンピュータによる自動要約の多くは，様々な観点から文の重要度に重み付けを行い，ランキング上位の重要文を選択し，その出現順に並べることで要約が生成される．抽象化または言い換えによる読みやすさの向上や文短縮によるアプローチなども試みられているが，本研究では，クエリ依存型抽出手法との共存も考慮し，重要文抽出によるアプローチを要約生成手法として採用することにする．重要度評価に用いられる情報には様々なものがあるが，奥村らはそれらを以下の六種類に分類している [15]．

- テキスト中の単語の重要度
- テキスト中あるいは段落中での文の位置情報
- テキストのタイトル等の情報
- テキスト中の手がかり表現
- テキスト中の文あるいは単語間のつながりの情報
- テキスト中の文間の関係を解析したテキスト構造

H. P. Edmundson の実験によると，これらは位置情報，手がかり表現，単語の重要度の順で効果的であると報告されており，またそれらを組み合わせることで精度が向上されるとの結果が示されている [16]．ただし，その後の様々な検証により，対象となるテキストの種類によって，効果的な手法にばらつきが出ることも知られている．

従来型要約手法では，テキストの内容をもとに要約は静的に決定できるという考え方で実現されてきたが，それは近年提案されているクエリによる重み付けを加えた場合でも基本的には変わっていない．しかし，対象となるテキストとクエリが同じ場合でも，検索目的の多様性から，指示的な要約としての効果をより高める為には，何らかの基準で一意に決定されるものではなく，ユーザの意図によって動的に変化する要約の方がウェブ情報検索には有効であると考えられる．

4.2 全体依存型

図 2 で示したスニペットの分類において，比較や分類といった検索対象を集合的に評価することが目的の場合は，III 型や IV 型のスニペットを生成すべきである．このような全体依存型のスニペットも，重要文抽出手法により生成することができる．断片集合型の場合，検索結果に含まれる上位 k 件のウェブページを対象にして，$TF \times iDF$ 値 [17] の大きな単語に重みを与えることで，スニペットとして生成される内容がウェブページごとに独自性をもつことになる．また，包括要約型の場合は，要約を生成する手法に加えて，$TF \times DF$ 値の大きな単語等に重みを与えることで，他のウェブページとの関係を考慮したスニペットを生成することができる．

5. スニペットの動的再生成

5.1 クエリ依存とクエリ独立

我々はクエリ依存型抽出手法とクエリ独立型要約手法を統合し，ユーザの意図により再生成可能な重要文抽出によるスニペットの生成手法を提案する．スニペット生成のために各文の重要度を評価する特徴ベクトルは以下のように生成する．

（1） ウェブページに存在する各文の重要度を計算し，特徴ベクトル v_t を生成（クエリ独立型要約）

（2） 検索語を含む文に重み付けを与え，特徴ベクトル v_q を生成（クエリ依存型抽出）

（3） 特徴ベクトル v_t および v_q を，α および $1-\alpha$ の比率で統合（式 1）

$$v_{tq} = \alpha \times v_t + (1-\alpha) \times v_q \quad (1)$$

まず，クエリ独立な要約を生成するための特徴ベクトル（v_t）を生成する．文の重要度を評価する手法については，上で紹介した重要度評価を利用する．このプロセスは，クエリに関係なく生成可能なため，事前にシステム側で用意しておくことができる．次に，クエリが決定した後，クエリ依存な抽出を行うために，検索語を含む文に重み付けを与えた特徴ベクトル（v_q）

を計算する．このとき，出現数の多い検索語や複数の検索語を含む文の重要度を高くする．最後に，二つの特徴ベクトル（v_t, v_q）を統合し，総合的な特徴ベクトルを生成する．

スニペットは，統合された特徴ベクトル（v_{tq}）における重要度の高い文を規定数選択し，出現順に配置することで生成する．このとき，α を変動させることにより，各文の総合的な重要度が変化するため，生成されるスニペットが変化することになる．我々はこのように生成された改良型スニペットを「Rich-Snippet」と呼んでいる．

また，クエリ独立な要約を生成するための特徴ベクトル（v_t）とクエリ依存な抽出を行うための特徴ベクトル（v_q）の類似度を計算することで，ウェブページに対するクエリの位置付けを評価することができる．本研究では，それをウェブページの主題を表すトピックに対するクエリの適合度と捉え，特徴ベクトル間のコサイン類似度をトピック-クエリ適合度（R_{tq}）として定義する（式 2）．

$$R_{tq} = \frac{v_t \cdot v_q}{||v_t|| \cdot ||v_q||} \qquad ||x|| = \sqrt{\sum_{i=1}^{n} x_i^2} \qquad (2)$$

トピック-クエリ適合度は，ウェブページとクエリが決定されると一意に定義される評価値で，クエリで表現されたユーザの目的とウェブページ著者の意図とのずれを定量的に評価したものである．トピック-クエリ適合度が大きい場合は，クエリが含まれる断片がウェブページの主題に近いことを意味し，小さい場合は，複数の話題がウェブページに存在するか全体の内容としてはクエリを含む断片はあまり重要でない可能性が高いことを意味する．この評価値をもとに検索結果のリランキングを行うことができる．

我々が提案した手法により，システムは動的に再生成が可能なスニペットをユーザに提供することができる．二つの特徴ベクトルを統合する際の α 値を大きくすることで，より包括要約型，つまり，クエリ独立な要約としての側面が強いスニペットを生成することができる．逆に，α 値を小さくすることで，より断片集約型，つまり，クエリに依存したスニペットを生成することができる．そのため，クエリが決定された場合にクエリ依存型抽出手法により一意に生成される現行のスニペットと比べ，Rich-Snippet はユーザによる可変かつ動的なスニペットであるといえる．この特徴は，同じクエリが入力された場合でも，ユーザの意図する結果が同じであるとは限らないという問題を解決する可能性をもつ．

5.2 Private View の実装と今後の課題

我々はクエリ依存またはクエリ独立なスニペットを動的に再生成させることができるウェブ検索インタフェースである Private View を実装した．ユーザは断片集約型または包括要約型の度合いをスライドバーにより変更することで，入力したクエリを変更することなくスニペットの内容を変化させることができる．図 fig:pview はスライドバーを包括要約型の度合いを強める方向へ移動させた場合のスニペットの変化を示している．

現在のプロトタイプでは，トピック-クエリ適合度は表示されていないが，検索目的に適したスニペットに動的に変更できることで，従来の静的なスニペットに比べ，ウェブページの全容をより正確に把握することができる．また，トピック-クエリ適合度が色の違いで視覚化されることにより，ウェブページの主題がクエリで表現される検索目的に適しているかどうかをテキストを読むことなくすばやく把握することができるようになる．このように，ウェブページに対する定量的な評価を視覚化することで，ユーザが目的のウェブページを見つける効率を改善させることができる．今後は，全体依存型スニペットの再生成を実現し，検索結果の変化と検索目的の適合度について評価を行う予定である．

6. 関連研究

ウェブ検索エンジンの普及に伴い，検索結果の個人適応化に着目した研究がいくつか行われている．Yahoo! Research は，「Yahoo! Mindset」[18] と呼ばれるユーザの検索意図や検索目的に適した検索結果を表示するウェブ検索インタフェースを提供している．彼らのシステムでは，ウェブページの種類を commercial（商品購買）と non-commercial（商品情報）とに分類しており，ユーザが購買目的または情報収集目的の度合いを選択することで検索結果をリランキングすることができる．このシステムもまた，検索結果の個人適応化を実現しているが，スニペットの機能は拡張されていない．Paolo Ferragina らは，スニペットの内容をもとに検索結果のクラスタリングを行おうとしている [6] [7]．しかし，既存のウェブ検索エンジンにより提供される現行のスニペットを扱うために，いくつか精度上の問題が報告されている．また，検索結果を類似度やコミュニティベースのスニペット・インデックスを利用して個人適応化しようとする研究もある [19] [20]．これらは検索結果を分類するには有効な手法であるが，スニペットの再生成は考慮されていない．

7. おわりに

ウェブ検索エンジンを利用したウェブ情報検索は図書検索と違い，検索対象はウェブページの枠を超えたウェブ空間に存在するすべての情報である．そのため，検索対象は複雑化し，検索目的は多様化してきている．本論文では，検索目的に適した検索結果を示すためにスニペットの生成手法を二種類の軸で分類し，スニペットの動的再生成を行う手法を提案した．我々は，このように検索目的に適した検索結果表示を行うことを SPO (Search Purpose Optimization) と呼んでいる．

謝辞

本研究の一部は，文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」，異メディア・アーカイブの横断的検索・統合ソフトウェア開発（研究代表者：田中克己）ならびに，文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」，計画研究「情報爆発時代に対応するコンテンツ融合と操

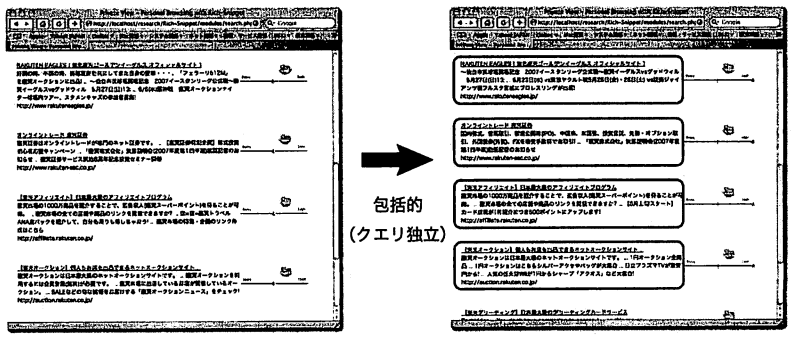

図 3 Private View: スニペットの動的再生成

Fig. 3 Private View: Dynamic Re-generation of Web-Snippet

作環境融合に関する研究」(研究代表者：田中克己，A01-00-02，課題番号 18049041) および計画研究「情報爆発に対応する新 IT 基盤研究支援プラットフォームの構築」(研究代表者：安達淳，Y00-01，課題番号：18049073) によるものです．ここに記して謝意を表すものとします．

文　献

[1] J. M. Kleinberg: "Authoritative sources in a hyperlinked environment", J. ACM, **46**, 5, pp. 604–632 (1999).

[2] S. Brin and L. Page: "The anatomy of a large-scale hypertextual web search engine", Proceedings of the seventh international conference on World Wide Web 7, Amsterdam, The Netherlands, The Netherlands, Elsevier Science Publishers B. V., pp. 107–117 (1998).

[3] M. A. Hearst and J. O. Pedersen: "Reexamining the cluster hypothesis: scatter/gather on retrieval results", Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR-1996), New York, NY, USA, ACM Press, pp. 76–84 (1996).

[4] Y. Wang and M. Kitsuregawa: "Evaluating contents-link coupled web page clustering for web search results", Proceedings of the eleventh international conference on Information and knowledge management (CIKM-2002), New York, NY, USA, ACM Press, pp. 499–506 (2002).

[5] E. J. Glover, K. Tsioutsiouliklis, S. Lawrence, D. M. Pennock and G. W. Flake: "Using web structure for classifying and describing web pages", Proceedings of the 11th international conference on World Wide Web (WWW-2002), New York, NY, USA, ACM Press, pp. 562–569 (2002).

[6] P. Ferragina and A. Gulli: "A personalized search engine based on web-snippet hierarchical clustering", Special interest tracks and posters of the 14th international conference on World Wide Web (WWW-2005), New York, NY, USA, ACM Press, pp. 801–810 (2005).

[7] F. Geraci, M. Pellegrini, P. Pisati and F. Sebastiani: "A scalable algorithm for high-quality clustering of web snippets", Proceedings of the 2006 ACM symposium on Applied computing (SAC-2006), New York, NY, USA, ACM Press, pp. 1058–1062 (2006).

[8] 株式会社アイレップ SEM 総合研究所，株式会社クロス・マーケティング："インターネットユーザの検索行動調査", Technical report (2006). Available as http://www.semirep.jp/info/20060626.pdf.

[9] E. Amitay and C. Paris: "Automatically summarising web sites: is there a way around it?", Proceedings of the ninth international conference on Information and knowledge management (CIKM-2000), New York, NY, USA, ACM Press, pp. 173–179 (2000).

[10] S. Oyama and K. Tanaka: "Query modification by discovering topics from web page structures", Proceedings of the 6th Asia-Pacific Web Conference (APWeb-2004), Vol. 3007 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 553–564 (2004).

[11] Y. Hu, G. Xin, R. Song, G. Hu, S. Shi, Y. Cao and H. Li: "Title extraction from bodies of html documents and its application to web page retrieval", Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR-2005), New York, NY, USA, ACM Press, pp. 250–257 (2005).

[12] H. P. Luhn: "The automatic creation of literature abstracts", IBM Journal of Research and Development, **2**, 2, pp. 159–165 (1958).

[13] C. D. Paice: "The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases", Proceedings of the 3rd annual ACM conference on Research and development in information retrieval (SIGIR-1980), Kent, UK, UK, Butterworth & Co., pp. 172–191 (1981).

[14] G. Salton: "Automatic Text Processing – The Transformation, Analysis, and Retrieval of Information by Computer", Addison–Wesley (1989).

[15] 奥村学，難波英嗣："知の科学：テキスト自動要約"，オーム社 (2005).

[16] H. P. Edmundson: "New methods in automatic extracting", J. ACM, **16**, 2, pp. 264–285 (1969).

[17] G. Salton and C. Buckley: "Term weighting approaches in automatic text retrieval", Technical report, Ithaca, NY, USA (1987).

[18] Yahoo! Research: "Yahoo! Mindset", http://mindset.research.yahoo.com/.

[19] M. Dontcheva, S. M. Drucker, G. Wade, D. Salesin and M. F. Cohen: "Summarizing personal web browsing sessions", Proceedings of the 19th annual ACM symposium on User interface software and technology (UIST-2006), New York, NY, USA, ACM Press, pp. 115–124 (2006).

[20] O. Boydell and B. Smyth: "Community-based snippet-indexes for pseudo-anonymous personalization in web search", Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR-2006), New York, NY, USA, ACM Press, pp. 617–618 (2006).