

WSA : Web 検索代数に基づく情報検索支援システムの実装

片山聡一郎[†] 遠山 元道^{††}

[†] 慶應義塾大学大学院 理工学研究科 開放環境科学専攻 〒 223-8522 横浜市港北区日吉 3-14-1

^{††} 慶應義塾大学 理工学部 情報工学科 〒 223-8522 横浜市港北区日吉 3-14-1

E-mail: [†]katasou@db.ics.keio.ac.jp, ^{††}toyama@ics.keio.ac.jp

あらまし Web は非常に巨大かつ多様性に富んだ情報メディアであり、ここから自分が欲しい情報を取り出すためには、その情報要求によってはいくつもの Web ページをネットサーフィンする必要があったり、複数の検索結果を統合する必要があったりと複雑な検索を行わなければ欲しい情報が得られない事がある。この様な検索における検索の効率化やユーザの検索における負担軽減のために、ユーザがインターネットを利用して情報要求を満たす行為を定性的に表現する数学的基盤として、Web Surfing Algebra(WSA) を提案した。本論文では WSA に基づく情報検索支援システムを提案する。そしてこのシステムを利用した検索例を示す事により、従来の検索手法では答えを得る事が難しかった情報要求の答えをこれまでよりも容易に得られる事を実証し、WSA 及び WSA に基づく情報検索支援システムの有効性を示す。

キーワード 情報検索, Web, 検索支援システム

WSA : Implementation of Information Retrieval Support System based on Web Surfing Algebra

Soichiro KATAYAMA[†] and Motomichi TOYAMA^{††}

[†] School for Open and Environmental Systems, Graduate School of Science and Technology,
Keio University

Hiyoshi3-14-1, Kouhoku-ku, Yokohama-shi, 223-8522 Japan

^{††} Department of Information and Computer Science, Faculty of Science and Technology,
Keio University

Hiyoshi3-14-1, Kouhoku-ku, Yokohama-shi, 223-8522 Japan

E-mail: [†]katasou@db.ics.keio.ac.jp, ^{††}toyama@ics.keio.ac.jp

Abstract Web is a huge and variety information carrier. So many users have to do a complicated search that is do netsurf of many Web pages for gaining information which I want. We proposed Web Surfing Algebra(WSA) which is mathematical representation of action of collecting information with internet, so that users can do netsurf easily. We have proposed supporting system of information retrieval based on WSA in this paper. We show with actual example that WSA and this system are effectivity.

Key words Information Retrieval, Web, retrieval support system

1. はじめに

Web は非常に巨大かつ多様性に富んだ情報メディアであり、ここから自分が欲しい情報を取り出すために google [1] など Web 検索エンジンが用いられている。またユーザの検索要求が多様化複雑化するのに応じて検索結果の質を向上させるための研究も盛んに行われている [3] [2].

Web 検索の流れは、ユーザが知りたい情報があった場合に、

ユーザはまずはその情報が載っている Web ページをイメージし、そのページが検索されると思われる検索クエリを検索エンジンに与える。その結果検索エンジンはクエリの条件を満たす Web ページをあらかじめ収集してある Web ページ集合から選び出し、ユーザに提供する。ユーザはそこから知りたい情報を探し出し自分の要求を満たす事ができるという流れである。つまりユーザは検索クエリを通じて自分の情報要求を検索エンジンに伝えている。

例えば「沖縄の駅弁について調べたい」とユーザが思った場合、ユーザは検索エンジンに「沖縄 AND 駅弁」というクエリを与える。この場合検索エンジンから返ってきた結果から、壺川駅の駅弁の情報を得る事ができる。

この様に多くの場合、ユーザの情報要求に含まれている単語を適当に組み合わせる事によって作られる検索クエリを1回検索エンジンに与える事により自分の情報要求を満たす事ができる。

しかしながらユーザの要求は様々であり、例えば宜野湾の名物を新宿で体験したいので、「宜野湾の名物を体験できる新宿の店を調べたい」といった事をユーザが思った場合、この情報要求を従来の検索エンジンによって満たす事は難しい。

なぜならば、この場合先程の例の様に、ユーザの情報要求に含まれる単語を組み合わせると「宜野湾 AND 新宿」や「宜野湾 AND 名物 AND 新宿」といった検索クエリを検索エンジンに与えても、最終的に知りたい新宿のお店の情報は検索結果にほとんど現れないからである。この原因としては、例え新宿で宜野湾の名物を出しているお店があったとしても、その店の紹介文の中に「宜野湾」や「名物」といった単語が含まれているとは限らないからである。

従って、この場合従来の検索エンジンを用いてこの情報要求を満たすためには、まず「宜野湾 AND 名物」で検索を行い一旦宜野湾の名物を調べ、その後改めてその名物を扱っている新宿のお店について検索を行う必要がある。この方法により知りたい店の情報は得られるものの、この方法を行うためには、宜野湾の名物を調べた結果をなんらかの形で保存しておき、新宿の店を調べる際にその結果を入力し直さなければならない。これはユーザにとって負担が大きい。

ユーザの情報要求に含まれている単語を適当に組み合わせる事によって作られる検索クエリを検索エンジンに与えるだけでは調べたい内容の答えを得られない場合、何度も検索エンジンに検索クエリを与え、Web ページをネットサーフィンする必要があり、また1回1回の検索で得られた情報を次の検索に利用するという事が必要である。このような検索においてユーザの負担を軽減し、検索を効率よく行う事ができる様にする事はWeb 検索を従来よりも便利にする事につながると言える。

1回検索クエリを検索エンジンに与えるだけでは知りたい情報を得る事ができない、複数のWeb ページをネットサーフィンする必要がある検索を支援するシステムを構築するためには、複数回行われる検索の手順及び検索結果の履歴を管理し扱う事ができる必要がある。そこでユーザのネットサーフィン行為に対応した代数であるWeb Surfing Algebra(WSA)を提案した[6]。WSAを利用する事により、ユーザのWeb における情報探索行為を代数として扱う事が可能となる。そして、複雑な検索に対する検索支援システムについて議論する事が出来るようになり、その様にして構築された検索支援システムにより、従来よりユーザの検索負担を減らす事ができたり、検索を効率よく行う事ができたり、検索を共同作業で行う事ができたりすると考えている。最終的には、ユーザが知りたい情報がある時、情報を得るための手順をナビゲートし、ユーザはその指示に従

うだけで効率よく、素早く知りたい情報を得る事ができるシステムを構築する事につながる事ができるのではないかと考えている。

本論文ではWSAに基づく情報検索支援システムを提案し、それに従って実装を行なった。その結果従来の検索システムでは答えを得る事が困難であった情報要求を従来よりも簡単に答えが得られる様になった事を示す。

2章でWeb Surfing Algebraの定義を行い、3章でWSAに基づく情報検索支援システムを提案する。4章では3章で述べた情報検索支援システムの実装方法について述べる。5章では実装した情報検索支援システムの実行例を示す。6章では従来手法との比較および情報探索行動とWSA式との対応の観点から本システムの考察を行う。そして7章でまとめを述べる。

2. Web Surfing Algebra(WSA)

本章では、Web Surfing Algebraの定義及び、そこで使用される関数及び演算子について述べる。

2.1 Web Surfing Algebraの定義

Web Surfing Algebraを以下の様に定める。

Web Surfing Algebra(WSA)とはユーザのネットサーフィン行為に対応した代数である。ユーザのネットサーフィン行為及び、ネットサーフィン中システムで行われる処理について1つ1つを以下に述べる関数、演算子に対応させた代数である。WSAが対象とするのはユーザのWeb上の情報・データに対する情報探索行動である。言い換えると、ある情報を知りたいといった情報要求やあるデータを集めたいといった情報要求に対して、Webを駆使する事によりその情報要求を満たすプロセスである。

2.2 Web Surfing Algebraにおける用語の定義

WSAにおいて扱われる用語を以下に紹介、説明する。

- word: 単語

名詞、動詞、形容詞など、意味を持った文字列。

- words: 単語集合

0個以上の単語を要素とする集合で、その順番も区別をする順序付きの集合と定める。

- query: クエリ

検索エンジンにおける検索クエリを指すものとする。単語、”AND”、”OR”、”NOT”、”(”、”)”からなる文字列。単語はクエリに含まれる。

- queries: クエリ集合

0個以上のqueryを要素とする集合で、その順番も区別をする順序付きの集合と定める。単語集合はクエリ集合に含まれる。

- result: 検索結果

検索エンジンに対して検索クエリを与える事によって得られる検索結果。Webページのタイトル、URL、ページの要約からなる組と定める。

- results: 検索結果集合

0個以上のresultを要素とする集合で、その順番も区別をする順序付きの集合と定める。

- URLs: URL 集合

0 個以上の URL を要素とする集合で、その順番も区別をする順序付きの集合と定める。

- attribute : 属性

名詞、形容詞など特徴を示す文字列。開いている Web ページから探し出さなければならない単語や文章がどのような内容か説明する。

- WSA 式

単語及び以下に紹介する関数、演算子からなる式。各関数、演算子には入力、出力が定められており、それらが間違いなく対応しているものを WSA 式とする。WSA 式はユーザのネットサーフィンにおける情報に対する要求及びそのためのプロセスを WSA により対応付けしたものである。WSA 式の処理は一番内側にある関数、演算子の処理から行っていき、一番外側の関数または演算子が処理し終わった段階でユーザの目標が達成された事となる。なお式を処理していく途中で、出力が 0 件であるなど十分な情報が得られない場合には適宜式の処理を遡る事となる。

2.3 WSA で使用される関数及び演算子

WSA で使用される関数及び演算子は情報探索行動における行為及びその目的と対応付けされている。例えば、ユーザが検索エンジンに検索クエリを与えその結果を受け取るという行為が WSA では Q 関数として対応付けられている。WSA で使用される関数及び演算子の名称、対応する行為、入力、出力について関数及び演算子の種類ごとに紹介する。

なお演算子と関数の違いはユーザが選択する行為が含まれるか含まれないかにより区別し、ユーザが選択する行為が含まれているものを演算子、含まれていないものを関数と呼ぶ事にする。

2.3.1 ユーザ及びシステムの Web 検索行為に対応した関数、演算子

- Q(queries) : Q 関数

query を検索エンジンで検索し、検索エンジンの結果 (Web ページのタイトル, URL, ページの要約の集合) を得る。入力が複数の query からなる場合 1 つずつ処理していく。また、入力されたクエリ集合の要素数が 0 の場合には、要素数 0 の検索結果集合を出力とする。

入力 … クエリ集合 (queries)

出力 … Web ページのタイトル, URL, ページの要約を組とする検索結果集合 (results)

- $\sigma(\text{results})$: σ 演算子

入力された検索エンジンの結果からユーザは、その後の WSA 式の処理を進めていく上で要求されていると判断した URL を 0 個以上選択する。

入力 … Web ページのタイトル, URL, ページの要約を組とする検索結果の集合 (results)

出力 … ユーザが選択した URL 集合 (URLs)

- $\pi_{\{\text{attribute}\}}(\text{URLs})$: π 演算子

入力された URL の Web ページを開きユーザに提示する。ユーザは指定された属性を持つ (条件を満たす) と判断した単語を選択し登録する。

入力 … URL の集合 (URLs)

出力 … 指定された属性 (attribute) を持つとユーザが選択、登録した単語集合 (words)

検索において通常ユーザは結果の上位数件しか利用しない。そこで集合の全要素の上位 n 件のみを出力とする場合、関数、演算子に添え字 n を付ける事によって表現する。但し集合の要素数が n 件に満たない場合は集合の要素全てとする。

例えば $Q_{10}(\text{query1})$ は query1 を検索エンジンに検索クエリとして与え、返ってきた検索結果の上位 10 件、言い換えると $Q(\text{query1})$ によって出力される検索結果集合の上位 10 件を出力とするという事である。

2.3.2 単語の集合に対して操作する関数及び演算子

- dis (queries) : disjunction

入力されたクエリの集合間を "OR" で結合する。

入力 … クエリ集合 (queries)

出力 … 入力された集合の要素を集合の要素の順番に "OR" でつなぎ、両端を括弧で囲む事によって得られるクエリ集合 (queries), 得られる集合の要素数は 1 となる

例 : $\text{dis}(\{\text{query1}, \text{query2}\}) = (\text{query1 OR query2})$

- con (queries) : conjunction

入力されたクエリの集合間を "AND" で結合する。

入力 … クエリ集合 (queries)

出力 … 入力された集合の要素を集合の要素の順番に "AND" でつなぎ、両端を括弧で囲む事によって得られるクエリ集合 (queries), 得られる集合の要素数は 1 となる

- words \cup words : union (和集合)

2つの単語集合の和集合をとり、新たな単語集合を得る

入力 … 2つの単語集合 (words)

出力 … 2つの単語集合の和集合をとり、その中から重複するものを取り除いた結果得られる単語集合 (words)

- words \cap words : intersection (積集合)

2つの単語集合の積集合をとり、新たな単語集合を得る。

入力 … 2つの単語集合 (words)

出力 … 2つの単語集合の積集合をとる事により得られる単語集合 (words), 2つの集合で共通する要素が存在しないならば空集合を返す

- words \sqcup words : 選択 union

ユーザは 2つの単語集合を見て、それらの中から出力に含めたくない要素を選択。それらを除き残った要素の和集合をとり新たな単語集合を得る。

入力 … 2つの単語集合 (words)

出力 … 2つの単語集合の単語の中からユーザが選択された要素を取り除き、残った単語の和集合をとり、その中から重複する要素を取り除いた結果得られる単語集合 (words)

- words \sqcap words : 選択 intersection

ユーザが 2つの単語集合を見比べ、それらの中から全く同じまたはほぼ同じ要素を選択。それらを要素とする新たな単語集合を得る。

入力 … 2つの単語集合 (words)

出力 … ユーザが 2つの単語集合を見比べ、ユーザが選択した

単語集合 (words)

処理内容はユーザが選択した単語集合を結果として出力するという意味では□と同じであるが、ユーザに対する要求 (指示) が異なるため別の演算子として用意した。

- asc (words) : asc 関数

入力された単語の集合を昇順に並び替える。

入力 … 単語の集合 (words)

出力 … 入力された単語の集合を昇順に並び替える事によって得られる単語の集合 (words)

- dec (words) : dec 関数

入力された単語の集合を降順に並び替える。

入力 … 単語の集合 (words)

出力 … 入力された単語の集合を降順に並び替える事によって得られる単語の集合 (words)

- queries AND queries : AND

入力された2つのクエリ集合の各々から要素を1つずつ取り出し、それらを”AND”でつなぎ合わせて新たな検索クエリを作り出す。クエリ集合が複数の要素からなる場合には、全ての組み合わせについて処理を行い、それらを全てが出力するクエリ集合の要素となる。

入力 … 2つのクエリ集合 (queries)

出力 … 入力された2つのクエリ集合の各々から1つずつ要素を取り出し、その2つの要素を文字列”AND”で結合し、両端を括弧で囲む事によって query を作成する事を、入力された2つの集合の要素の全ての組み合わせについて行う事によって作られるクエリ集合 (queries)

例 : query1 , query2 AND query3 = (query1 AND query3) , (query2 AND query3)

- queries OR queries : OR

入力された2つのクエリ集合の各々から要素を1つずつ取り出し、それらを”OR”でつなぎ合わせて新たな検索クエリを作り出す。クエリ集合が複数の要素からなる場合には、全ての組み合わせについて処理を行い、それらを全てが出力するクエリ集合の要素となる。

入力 … 2つのクエリ集合 (queries)

出力 … 入力された2つのクエリ集合の各々から1つずつ要素を取り出し、その2つの要素を文字列”OR”で結合し、両端を括弧で囲む事によって query を作成する事を、入力された2つの集合の要素の全ての組み合わせについて行う事によって作られるクエリ集合 (queries)

- queries NOT queries : NOT

入力された2つのクエリ集合の各々から要素を1つずつ取り出し、それらを”NOT”でつなぎ合わせて新たな検索クエリを作り出す。クエリ集合が複数の要素からなる場合には、全ての組み合わせについて処理を行い、それらを全てが出力するクエリ集合の要素となる。

入力 … 2つのクエリ集合 (queries)

出力 … 入力された2つのクエリ集合の各々から1つずつ要素を取り出し、その2つの要素を文字列”NOT”で結合し、両端を括弧で囲む事によって query を作成する事を、入力された2

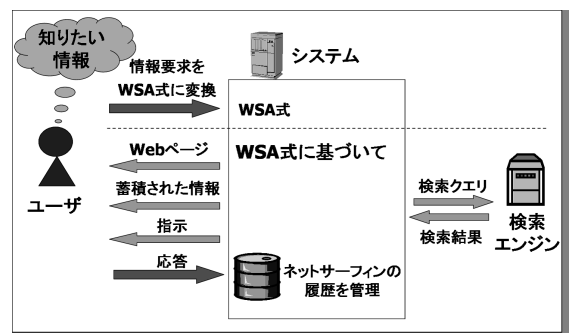


図1 WSAに基づく情報検索支援システムの構成

つの集合の要素の全ての組み合わせについて行う事によって作られるクエリ集合 (queries)

なおこれらの関数、演算子の性質は文献[6]で述べている様に各関数、演算子毎に異なる性質を持つ。

2.4 WSA 式の例

これまで挙げてきた関数、演算子を利用する事により、ユーザが Web 上から必要な情報を得る行為が WSA 式によりどの様に表現されるか例を示す。

例。「ノートパソコンの機種名を知りたい」という場合、WSA 式で表すと以下の様になる。

$\pi_{\text{機種名}}(\sigma(Q(\text{ノートパソコン})))$

式の処理は原則的に括弧の内側から処理されていくため、この式を処理する場合には、まず始めに Q 関数を処理し「ノートパソコン」を検索エンジンにかける。次に σ 演算子を処理するので、その検索結果をユーザに提示し、その中から URL を選択してもらう。最後に π 演算子を処理する。選択された URL のページを開き、ユーザに機種名をピックアップしてもらう事となる。なおこの場合、もしも選択された URL のページから機種名が見つからない場合には式の処理を遡る事になる。

3. Web Surfing Algebraに基づく情報検索支援システムの構成

WSA に基づく情報検索支援システムの構成について説明する。本システムの構成を図1に示す。

システムの流れは以下の通りである。

1. ユーザは何か知りたい情報があつた場合、その情報要求を WSA 式に変換し WSA 式をシステムに与える。
2. システムはこの WSA 式に基づいて処理を行なう。WSA 式を内側から処理する。
3. WSA 式に従って、例えば検索エンジンにクエリに与え、その結果をユーザに提供する。その際システムはユーザに対して次に何をしたいかの指示を適宜同時に提供する。
4. ユーザはシステムから提示された結果及び指示に対して応答する。
5. 応答結果はシステムに履歴として管理される。
6. またユーザの応答に応じてシステムは、新たなクエリを検索エンジンに与えその結果をユーザに提供したり、WSA 式で次に処理する関数、演算子の処理を行うのに十分な量の履歴が集まった段階でその情報を整理した上でユーザに提供したりす



図 2 Q(query) の実行画面

る。この時も情報と共に適宜指示を一緒に提供する。

7. それに対してユーザは再び応答する。

8. この様なユーザとシステムのやり取りを繰り返し、WSA 式の処理が全て終わり、ユーザの情報要求が満たされた段階で終了する。

従って、ユーザは始めに WSA 式をシステムに与えた後は、システムからの指示に従って応答する事により情報要求を満たす事ができる。これは従来の Web 検索においてユーザが自分で次に何をすべきか考えたり、結果を保存したりしなければならなかった事と比較すると、ユーザは簡単に効率よく欲しい情報を得る事が可能となる。

4. 実 装

現段階では各関数、演算子の実装が終わっており、それらをつなぎ合わせる事により、あらかじめ定められている WSA 式に対応した検索支援システムを構築する事が可能である。

一方、現段階では途中で WSA 式を変更する事や式の流れを指定する事はできず、式の処理を遡る方法も単純に前の関数、演算子に戻る様になっておりこの点に関しては今後さらに実装を進めていきたいと考えている。

各関数、演算子は PHP5.0.5 及び HTML により記述された Web ページにより実装されている。データベースエンジンは SQLite を利用した。

以下、各関数、演算子毎に見ていく。

- Q(queries) 図 2 の様な入力フォームが表示され、関数に入力された query が入力フォームに入る (複数の query が入力されている場合には 1 つずつ処理する)。“Search!”をクリックする事により、その query が検索エンジンに与えられ、検索結果が関数の出力としてデータベースに保存される。これがこの関数における出力である。保存される内容はページの URL、ページのタイトル、ページに関連するテキストサマリーである。“完了”をクリックする事により次の関数、演算子のページに飛ぶ。Yahoo!検索 Web サービスのウェブ検索 Web サービスを利用した [7]。リクエストパラメータを設定する事が可能だが、1 回の検索に対する返却結果の数は 50 とした。従って、WSA における Q_{50} (queries) に対応する。

- σ (results)

入力された検索結果を図 3 の様に表示する。“OK”が押されるとユーザがチェックボックスを選択した検索結果の URL が上にあるものから順番に別のデータベースに格納される。これがこの演算子の出力となる。全てが格納し終わったところで次の関数、演算子のページに移る。

- $\pi_{attribute}$ (URLs)

入力された属性名が図 4 の様に表示される。入力された URL の集合が同図の様に表示される。ここで URL を選択する事に

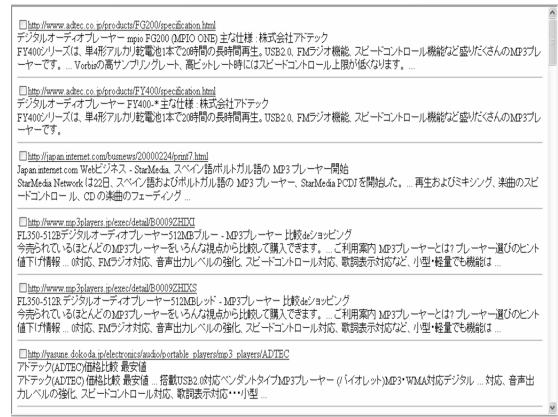


図 3 σ (results) の実行画面



図 4 $\pi_{attribute}$ (URLs) の実行画面 (1)



図 5 $\pi_{attribute}$ (URLs) の実行画面 (2)

よりそのページを別窓で開く。また“メモを開く”をクリックする事により別窓に図 5 の様に入力フォームが表示される。この入力フォームを介して登録された単語は別のデータベースに格納される。これがこの演算子の出力となる。“完了”をクリックする事により次の関数、演算子のページに移る。

- dis (queries) / con (queries)

入力された集合から要素を 1 つ取り出し、空の文字列に格納する。まだ集合に要素が残っている場合には、新たに要素を 1 つ取り出し、先程の文字列の後ろに OR(AND) を連結しその後ろに新たな要素を連結する。これを集合の要素が無くなるまで繰り返す。要素が無くなったらその文字列を別のデータベースに格納する。これがこの関数の出力となる。格納後自動的に次のページに移る。

- words \cup words / words \cap words

2つのデータベースを介して2つの単語集合が入力される。それぞれの集合の要素を別々の配列に一旦格納する。その後その2つの配列に関して和集合(積集合)をとる事によって新たな配列を得る。それを別のデータベースに要素を1つ1つ順番に格納する。これがこの関数の出力となる。格納が終わったら自動的に次のページに移る。

- words \sqcup words / words \sqcap words

2つのデータベースを介して2つの単語集合が入力される。そ

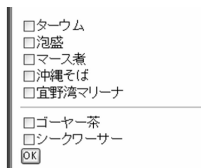


図 6 words □ words の実行画面

れらを図 6 の様に表示する。OK が押されるとユーザがチェックボックスを選択している単語に関して上にあるものから順番に別のデータベースに格納される。これがこの演算子の出力となる。全てが格納し終わったところで次の関数、演算子のページに移る。

- asc (words) /dec (words)

入力された単語集合の要素を 1 つずつ取り出し配列に格納する。その後その配列の要素を昇順 (降順) に並べ改めて別のデータベースに配列の要素を 1 つずつ格納していく。これがこの関数の出力である。格納が終わったら自動的に次のページに移る。

5. 実行例

本章では実装した WSA の関数、演算子を利用する事により情報要求の答えを得る事を実際に行った結果を示す。

5.1 実行例

- 実行例 1

「スピードコントロール機能のついている MP3 プレーヤーの値段を調べたい」という情報要求を WSA 式で表すと

$\sigma(Q(\pi_{50} \text{機種名}(\sigma(Q(\text{スピードコントロール AND MP3 プレーヤー})))) \text{ AND (値段 OR 価格)})$

となる。式の処理の流れ及び各々の入力出力とデータベースの関係を図示したものが図 7 である。なお、名前に tmp と付いているデータベースは処理し直す毎に初期化されるデータベースである。また AND ノードの部分では単純に入力された 2 つの文字列を AND で連結した結果を次に送っている。

最終的な結果として得られた URL のページから「FG200-512」「FL350-512B」「iFP-799」などの機種についての値段の情報を得られた。

この情報要求に対して「スピードコントロール AND MP3 プレーヤー AND 値段」として検索した場合には、本システムを利用した場合と比較して情報を得られた機種の種類が少なかった。その原因としては「スピードコントロール」という単語は製品の機能紹介のページに存在する単語であるが、機能紹介のページに必ずしもその商品の値段が載っているとは限らないため、「スピードコントロール」「値段」2 つの単語が存在するページは限られてしまうからと考えられる。

- 実行例 2

「宜野湾の名物を体験できる新宿の店を調べたい」という情報要求を WSA 式に変換する際、様々な式が考えられるがその一例を示すと

$\sigma(Q(\text{con}(\pi_{50} \text{名物}(\sigma(Q(\text{宜野湾 AND 名物})))) \text{ AND 新宿}))$

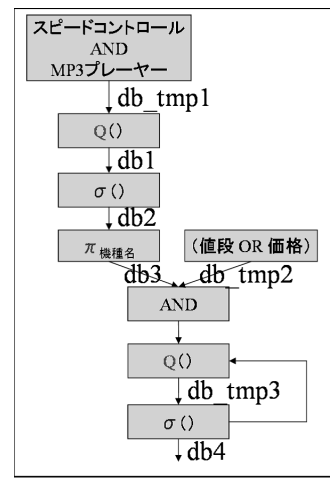


図 7 実行例 1 の式の処理の流れ

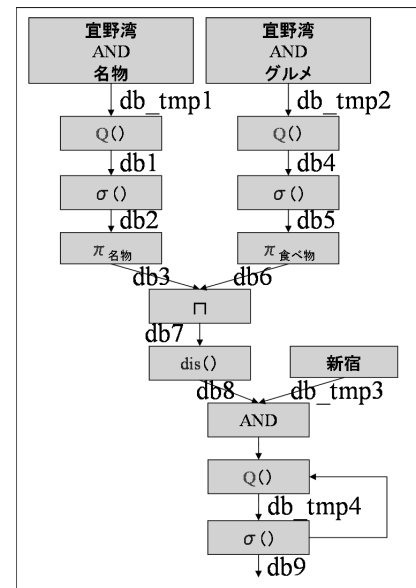


図 8 実行例 2 の式の処理の流れ

$\sigma(Q(\text{con}(\pi_{50} \text{食べ物}(\sigma(Q(\text{宜野湾 AND グルメ})))) \text{ AND 新宿}))$

となる。式の処理の流れ及び各々の入力出力とデータベースの関係を図示したものが図 8 である。

選択インターセクションの部分で「ターウム」「泡盛」「沖縄そば」「シークワサー」を選択した結果、一番外側の Q 関数では「新宿 AND (ターウム OR 泡盛 OR 沖縄そば OR シークワサー)」という query により検索が行われ、最終的な結果として得られた URL のページから「かりゆし」「島の人」といった店についての情報を得られた。

1 章で述べた様に、この情報要求に対して「宜野湾 AND 新宿」や「宜野湾 AND 名物 AND 新宿」と検索しても、知りたい情報はほとんど得られなかった。

6. 考察

本章では、まず従来手法との比較を行う事により WSA に基づく情報検索支援システムの有効性について述べ、次に情報探索行動と WSA 式との対応について述べる事により本システムの今後の課題を述べる。

6.1 従来手法との比較

従来の Web 検索手法との比較を行う事により本システムを利用する事による有効性を示す。

本システムの手法が従来手法よりも優れている最大の点は、ある検索結果を次の検索に簡単に利用できる点である。従来手法では、ある検索結果を次の検索に利用するためには一旦検索結果を記憶するかどこかに一時的にコピーしておき、次の検索をする際先程コピーしたものを再びコピーするなどの手間がかかる。それに対しては本手法では、検索結果を登録するために始めの 1 回はコピーや入力を行わなければならないが、登録先となるデータベースは自動的に作成され、登録されたものを簡単に次の検索結果に利用することができるため、従来手法と比較して、コピーを行う回数および新たな検索クエリを検索エンジンに与えるために入力する量を少なくすることができる。また検索結果を関数 1 つで簡単に OR, AND で結合することができるので、従来の検索手法ではユーザ自身で複数の単語を組み合わせなければ作成する事ができなかった複雑な検索クエリにおいて、関数、演算子を利用する事により複数の単語を組み合わせる部分をシステムの方で行ってくれる点も有効な点である。

従来手法では検索結果を 1 ページずつ URL を選択してそのページを見るという事を行うのに対して、本システムでは検索結果から先にまとめて URL を選択し、後でまとめて Web ページを見ていく。この点に関しては複数のページの情報を集める必要のない情報要求の場合、検索結果の上位数ページで答えが得られるため、本システムの手法の様に前もって URL をまとめて選択するという方法は従来手法と比較して無駄な作業が多くなってしまおうと言える。従って、非常に単純な情報要求の場合には従来手法の方が効率が良い。しかしながら、答えを得るために複数のページの情報を必要とする情報要求に対してはその無駄な作業は少なくなる。さらに、ユーザは URL の選択をしていてこれでもう十分と思った時点ですぐに”完了”をクリックすれば、すぐに次の処理へ進むため、それ程大きな負担ではないと考える。この部分に関して本手法の方が優れている点としては、URL を開く段階では URL を選択するだけで、従来手法の様に”戻る”ボタンを押す必要なくページを次々に開く事ができるため、ある程度多いページ数を見る必要のある様な情報要求に対しては本システムの方が手間が少なくなる点である。

以上より、答えを得るために複数のページの情報を必要とせず、検索エンジンで検索してその検索結果の上位の数ページを見るだけで答えが得られる様な情報要求に対しては従来手法の方が手軽に検索を行えるためユーザの負担が少なくなるが、そうでない場合には、検索結果を次の検索に容易に活かせる点、複数のページを見る際ページに戻る必要が無い点からユーザの負担は少なくなると言える。また複雑なクエリも関数を利用する事により簡単に作成する事ができるため、従来手法の時には様々な準備をしなければ出来なかったり、複数の単語を組み合わせなければクエリを作成する事ができなかったため、行われなかったような検索を本システムを利用する事により簡単にを行う事が出来る様になり、情報探索行動の幅が広がったと言える。

6.2 WSA 式と情報探索行動モデルとの対応

WSA に基づく情報検索支援システムが特定の情報要求に対して有効である事を例により示したが、まだどの様な情報要求に対しても十分に対応する事ができているとは言いがたい。そこで 2 章で定義した WSA 式による情報探索行動の能力を情報探索行動モデルの観点から検討してみる事により、WSA がより一般的な情報要求にも有効とするためにはどうすれば良いか、今後の課題を示す。情報探索行動モデルは様々なものが存在するがここでは Ellis による情報探索行動モデルを採用する事にする [4] [5]。Ellis による情報探索行動モデルは元々研究に関する情報を探す際の情報行動パターンを抽出し、それらを分類した結果であるが、企業のエンジニアに対しても有効である事が示されており、Web における情報探索行動にも多くの部分で対応すると考え採用した。

Ellis は情報探索行動を以下の 8 つに分類した。そこでその 8 つの観点から述べる。

- Starting : 検索の開始

WSA においては自分の情報要求を WSA 式に変換する事が対応している。現時点ではこの部分はユーザ自身で行う必要があるが、将来的にはこの部分の自動化も行えればと考えている。ただ、そのためには自然言語処理の技術なども必要になると思われる。

- Chaining : 既にある情報から新たな情報を見つけ出す

Web においては、検索エンジンで検索する事により新たな Web ページを得る行為に該当すると言え、WSA では Q 関数で対応している。またリンクを辿る行為も該当するが、これも π 演算子で Web ページを見ていく際リンクを辿る事も可能であるため、WSA はこの点に関しては十分に対応していると言える。

- Browsing : 漠然と情報を探す

Web においては Web ページを漠然と見ていき、興味がある内容があれば詳しく文章を見たり、リンクを辿ったり、新たに検索エンジンに検索クエリをかける行為に対応する。WSA は情報要求がはっきりしている検索に対しては効果的であるが、この様に目的のはっきりしない情報探索行動に対しては対応しているとは言いがたい。その場の思いつきなどにも対応できるようにするためには WSA 式の変更を途中で自由に変更出来る必要があると言える。

- Differentiating : 情報源の選択をする

Web 検索においては、検索クエリのキーワードを選択する事、検索結果から URL を選択する事、開いているページからリンクを辿る事が該当すると言え。これらについては WSA ではそれぞれ Q 関数、 σ 演算子、 π 演算子により十分に対応している。

- Monitoring : 最新情報を保持する

検索エンジンはクロールする事により最新の情報を保っているため、検索エンジンを利用する場合その範囲で最新であると言える。もしも、情報の更新すなわち Web ページの更新ごとに情報を最新に保つ為には、指定した Web ページの更新を常にチェックする必要がある。それを WSA で行う為には、最終更新日をチェックする関数を用意し、WSA 式を自動的に発

行・処理するシステムを開発する必要がある。

- Extracting: 情報を抽出する

Web においては Web ページから有用な情報を探し出す事に該当する。WSA においては π 演算子でメモに登録する事で対応できている。

- Verifying: 情報の正確さを確認する

情報が正確かどうかを最終的に判断するのはユーザ自身であるが、Web における判断材料は情報が載っているページ数と各 Web ページの信頼度である言える。従って WSA において対応するためには、該当する情報を自動的に集める関数や、そのページの信頼度を測定する関数というのを用意する事により対応できる様になると考えられる。

- Ending: 検索の終了

ユーザが十分に情報が集まったと判断した時点で検索の終了となる。WSA においては式の処理が全て終わった段階または途中で十分な情報が集まり式の処理を中断した場合が該当する。なお、実行例を実際にやっている最中に自分が何をやっているか見失いやすいという弱点が存在していた。そのため WSA ナビゲーション [6] の様に、現在 WSA 式のどの部分を処理しているかユーザに提示する必要性があると考えられる。

7. おわりに

本研究では WSA に基づく情報検索支援システムを提案した。そして WSA の関数、演算子を実装し、それらを利用して情報要求の答えを得る例を示す事により、従来の検索手法では答えを得る事が難しかった情報要求の答えを本システムを利用する事により得られる事を実証し、WSA 及び WSA に基づく情報検索支援システムの有効性を示した。

今後は 6 章で述べた点を考慮してさらに情報支援システムを強化していきたいと考えている。

文 献

- [1] L. Page, S. Brin, R. Motwani, T. Winograd, "The PageRank Citation Ranking", *Stanford Digital Library Technologies, Working Paper SIDL-WP1999*, 1999.
- [2] Natalie S. Glance "Community Search Assistant", *Proceedings of the 6th international conference on Intelligent user interfaces*, pp.91-96, 2001
- [3] Ronald Fagin, Ravi Kumar, Kevin S. McCurley, Jasmine Novak, D. Sivakumar, John A. Tomlin, David P. Williamson "Searching the workplace web", In *Proceedings of the twelfth international conference on World Wide Web*, pp.366-375, 2003
- [4] Ellis D. "A behavioural approach to information retrieval design", In *Journal of Documentation*, pp. 318-338 1989
- [5] Ellis D. , M. Haugan "Modelling the information seeking patterns of engineers and research scientists in an industrial environment" In *Journal of Documentation*, pp.384-403, 1997
- [6] 片山聡一郎, 遠山元道 "Web surfing Algebra における関数および演算子の性質について", *データベースと Web 情報システムに関するシンポジウム (DBWeb2005)*, pp.197-204, 2005.
- [7] Yahoo!デベロッパネットワーク: <http://developer.yahoo.co.jp/>