

# WWW 情報統合のための動的アクセスプランニング手法

A dynamic access planning method for WWW information integration

野田 知哉 北村 泰彦 辰巳 昭治

Tomoya Noda, Yasuhiko Kitamura, and Shouji Tatsumi

大阪市立大学 工学部 情報工学科

Department of Information and Communication Engineering

Faculty of Engineering Osaka City University

Abstract

WWW information integration solves problems by integrating information from multiple WWW resources. While we take time to access WWW resources, their contents are frequently updated and the freshness of cached data is degraded. In this paper, we propose a dynamic access planning method to deduce a good solution considering the freshness of data even when the number of network access is limited.

## 1 はじめに

WWW 上には多くの WWW 情報資源があり、多種多様な情報が公開されている。インターネット人口や発信情報の増加が予想される今後、WWW 上に分散した情報を用いて問題解決を行う情報統合は、ますます重要になるであろう [1]。

本稿では、図 1 に示す mediator を用いて、WWW 情報源が公開している情報を用いた情報統合を試みる。利用者から要求を受けると、mediator は WWW 情報源にアクセスして必要な情報を取得し、それらを用いて要求に応える。しかし、このような mediator には以下の 2 つの問題が発生する。

1 つは、WWW 情報源の公開する情報は非同期に更新される点である。WWW 情報源は個別に運営されており、更新頻度はそれぞれで異なる。そのため mediator の蓄える情報と、WWW 情報源の公開する情報に食い違いが生じる可能性がある。更に、更新されたかどうかは、実際にその WWW 情報源にアクセスしないと分からない。

もう 1 つは、インターネットを介するので、WWW 情報源のアクセスに時間がかかる点である。利用者の要求に素早く対応するには、情報統合に必要な情報をすべて集めるのは困難である。そのため、限られた時間でより適切な解を求める手段が必要になる。

これらの問題に対処するために、本方式では mediator の蓄える情報に、その情報の内容が信頼のおけるものかを表す信頼度を与える。信頼度は、より適切な解を提示

〒558-8585 大阪市住吉区杉本3-3-138  
{tnoda,kitamura,tatsumi}@kdel.info.eng.osaka-cu.ac.jp  
<http://www.kdel.info.eng.osaka-cu.ac.jp/>

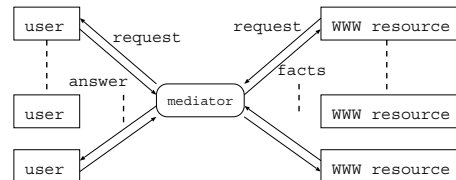


図 1: mediator

するために再検索する情報を決定したり、利用者に提示する解を選ぶのに用いられる。本稿では、ネットワークへのアクセス回数が制限された状況下においても、信頼度を考慮してできるだけ適切な解を得るためのアクセスプランニングアルゴリズムを提案する。

## 2 アクセプランニング

mediator は、WWW 情報源の公開する情報を取得すると、それを今後の問題解決のために蓄えておく。これを事実  $f$  とする。

WWW 情報源で公開されている情報は、時間と共に更新される。そのため、mediator が蓄えている事実と、WWW 情報源の公開する情報に食い違いが生じる可能性がある。情報を取得してから時間が経つ程、食い違う可能性は高くなる。そこでシステムが蓄える事実に対し、その事実がどの程度信用できるかを表す信頼度関数  $r(f)$  を定義する。WWW 情報源から情報を取得した直後が最も信頼でき、時間が経つにつれてその信頼は失われると言える。そのため、信頼度関数は初期値 1 で 0 に収束する単調減少関数になる。ただし、減少の傾向は WWW 情報源によって異なる。例えば、株価や為替相場の様に頻繁に更新される情報は、その信頼度はすぐに失われる。あるいは、ある事実が全く更新されないと分かっている

とき、その事実の信頼度は常に1である。以降、更新される可能性のある事実を動的な事実、更新されない事実を静的な事実と呼ぶ。

利用者から要求を受けると、mediatorは自身の蓄えた事実の集合  $\{f\}$  を用いて、利用者の要求に対する解  $A$  を求める。解の求め方は、それぞれの問題によって異なるので、本稿では解の求め方については扱わず、事実を用いて解が出せるという前提で議論を行う。要求に対する解は、その解を求めるのに用いた事実から構成されるので、

$$A = \{f_a, f_b, \dots\}$$

とする。

解は mediator の蓄えている事実を用いて求めるが、WWW 情報源の公開する情報は更新されている可能性があり、用いた事実を WWW 情報源より再検索すればより適切な解を提示できる可能性がある。しかし、限られた時間内で解を改善するには、再検索する事実の選び方が重要になる。そこで、質、提示スコア、再検索スコアといった関数を用いてそれぞれの解の質を評価し、再検索を行う事実を決定する。以下にアルゴリズムを示す。

STEP1: 利用者の要求に対して、蓄えている事実を用いて解を求める。

STEP2: 決められた時間内で以下を繰り返す。

STEP3: 求めた解の中で、最も再検索スコアの高い解を  $A_i$  としたとき、 $A_i$  の使用した事実の中で最も信頼度の低い事実  $f_j$  を再検索の対象とする。

STEP4: WWW 情報源より  $f_j$  を取得する。

STEP5: 提示スコアの高い解から利用者に提示する。

解を改善するために事実の再検索を繰り返すが、そのために必要な関数について述べる。まず、それぞれの解がどの程度利用者の要求を満たしているかを表す質を定義する。質には2種類あり、静的な事実のみを用いて求めた質を静的な質  $Q_s(A)$  とし、動的な事実を用いて求めた質を動的な質  $Q_d(A)$  とする。それぞれの質の評価方法は、問題や利用者の要求によって異なる。また静的な質は不変である。元の事実内容が変化するので、動的な質は変化するのである。

mediator は、蓄えている事実を用いて解を求めるが、使用した事実の信頼度が低いと、その事実を用いた解も信頼度は低いといえる。解の信頼度は、用いた事実の中で最も低い事実の信頼度を解の信頼度とする。ある解  $A$  を求めるのに用いた事実を  $f$  として、信頼度関数を、

$$R(A) = \min_{f \in A} r(f)$$

とする。

最新の事実を取得するとより適切な解を提示できるが、時間的な制約のために効率良く解を改善する事実が

表 1: 事実  $\{f\}$

事実 $f$	出発地	目的地	発着時刻	席	信頼度
$f_0$	札幌	羽田	0800, 0900	空	0.70
$f_1$	札幌	伊丹	0830, 0930	空	0.60
$f_2$	羽田	伊丹	0930, 1030	満	0.30
$f_3$	羽田	伊丹	1200, 1300	空	0.40
$f_4$	羽田	鹿児島	0930, 1030	空	0.40
$f_5$	羽田	鹿児島	1300, 1400	満	0.30
$f_6$	伊丹	鹿児島	1100, 1200	空	0.70
$f_7$	伊丹	鹿児島	1500, 1600	空	0.80

ら再検索しなければならない。質はよいが信頼度の低い解に関して、信頼度を改善できるとその解の提示スコアも改善される。また、質が著しく低い解や、信頼度の高い解は再検索の順位は低くてよい。そこで、再検索の順序を決定するために、再検索スコア関数を、

$$S(A) = Q_s(A)(1 - R(A))$$

とする。

こうしてそれぞれの解の再検索スコアが決定される。再検索スコアが最大となる解が、最も提示スコアが改善されると期待できる (STEP3)。よってその解を求めるのに用いた事実の中で、最も信頼度の低い事実を再検索する (STEP4)。

STEP2 から STEP4 は利用者の要求を受けてから行われるので、この間は利用者を持たせていることになる。そのため利用者の指定した検索時間が経過した時点で、STEP2 から STEP4 の繰り返しは終了する。

複数の解の中から利用者に提示する解を選ぶために、提示スコアを決める (STEP5)。単に質だけで解を評価すると、信頼できない事実を使用した解を優先的に利用者に提示する可能性がある。そのため、質に加え信頼度も考慮して利用者への提示順序を決定する。より要求を満たし、信頼度の高い解を利用者は望むと考えられるので、提示スコア関数は、

$$P(A) = Q_s(A) \times Q_d(A) \times R(A)$$

とする。利用者には提示スコアの高い解から順に提示する。

### 3 例題

#### 3.1 問題

航空会社は WWW 上に空席状況を含めた時刻表を公開していて、利用者が目的地などの要求を入力すると、要求を満たす旅程を提示する。ここでは例題として、複数の航空会社の公開する時刻表を用いて、利用者の要求を満たす旅程を検索する問題解決を行う。

利用者の要求は、出発地、目的地、希望到着時刻、検索時間から構成される。検索時間とは、利用者が要求を入力してから、mediator が解を提示するまでの待ち時間である。ここでは利用者の要求を「札幌から鹿児島に14:00に着きたい。ただし検索時間は15秒」とする。この要求を満たすような便の組合せが例題における解になる。航

表 2: 解  $\{A_i\}$ 

解 $A_i$	$Q_s$	再検索回数 0 回				1 回				2 回			
		$Q_d$	$R$	$P$	$S$	$Q_d$	$R$	$P$	$S$	$Q_d$	$R$	$P$	$S$
$A_0 = \{f_0, f_2, f_7\}$	0.50	0	0.30	0.00	0.35	0	0.30	0.00	0.35	0	0.30	0.00	0.35
$A_1 = \{f_0, f_3, f_7\}$	0.50	1	0.40	0.20	0.30	1	0.40	0.20	0.30	1	0.40	0.20	0.30
$A_2 = \{f_0, f_4\}$	0.69	1	0.40	0.28	0.41	1	0.40	0.28	0.41	0	0.70	0.00	0.21
$A_3 = \{f_0, f_5\}$	1.00	0	0.30	0.00	0.70	1	0.70	0.70	0.30	1	0.70	0.70	0.30
$A_4 = \{f_1, f_6\}$	0.83	1	0.60	0.50	0.33	1	0.60	0.50	0.33	1	0.60	0.50	0.33
$A_5 = \{f_1, f_7\}$	0.50	1	0.60	0.30	0.30	1	0.60	0.30	0.30	1	0.60	0.30	0.30

A: 解,  $Q_s$ : 静的な質,  $Q_d$ : 動的な質,  $R$ : 信頼度,  $P$ : 提示スコア,  $S$ : 再検索スコア

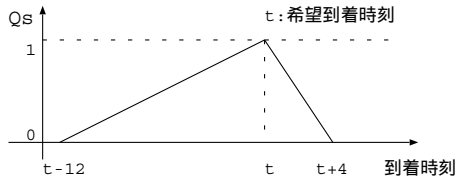


図 2: 解の質

空会社の運営する WWW 情報源からは、出発日、便名、出発地、目的地、出発時刻、到着時刻、空席状況から構成される時刻表を取得でき、mediator はこれらを事実として蓄える。また、空席状況は変化するので動的な事実であり、その他は静的な事実である。

信頼度関数  $r$  と質  $Q_s, Q_d$  は、以下のように設定して実験を行った。予約やキャンセルのために空席状況は刻々と変化するので、信頼度は徐々に低下する。そこで、空席状況を取得してから  $t$  時間後における信頼度は、

$$r(f) = \frac{1}{t/24 + 1}$$

とする。

利用者の希望する時刻に到着する解が最もその質がよいと言え、静的な質  $Q_s$  は希望到着時刻とその解での到着時刻で評価する (図 2)。例えば、14:00 に鹿児島に着きたいという要求に対しては、 $f_5$  を利用した  $A_3$  が最も要求を満たしている。

空席状況が動的な質であり、満席の便がを使用する解は提示できないので、動的な質を、

$$Q_d(A) = \begin{cases} 1 & \text{すべての便が空席のとき} \\ 0 & \text{その他のとき} \end{cases}$$

とする。

### 3.2 解を導出する流れ

例題での本アルゴリズムの各ステップでの動作を説明する。

再検索回数 0 回: それぞれの解の信頼度、提示スコア、再検索スコアを求める。例えば、 $A_0$  の信頼度は、

$$R(A_0) = \min(f_0, f_2, f_7) = 0.30$$

となる。また、 $f_2$  が満席なので、 $A_0$  の動的な質、提示スコアは共に 0 である。再検索スコアの最も大きい  $A_3$  が再検索の対象となり、 $F_3$  の中で最も信頼度の低い  $f_5$  を再検索する。

表 3: 例題での解の提示順序

再検索回数	提示順序
0	$A_4, A_5, A_2, A_1$
1	$A_3, A_4, A_5, A_2, A_1$
2	$A_3, A_4, A_5, A_1$

再検索回数 1 回:  $f_5$  を再検索すると、キャンセルにより空席ができていた。そして、再び提示スコアを計算すると、表 2 の 1 回目の欄のようになる。 $A_3$  の信頼度  $R(A_3)$  と提示スコア  $P(A_3)$  が改善され、再検索スコア  $S(A_3)$  が下がっているのがわかる。また、次の再検索すべき事実は  $A_2$  で用いた  $f_4$  となる。

再検索回数 2 回:  $f_4$  を再検索すると、満席になっていた。すると、 $f_4$  を用いた解  $A_2$  の提示スコアは 0 になる。このように、再検索を繰り返して解の信頼性を改善して、より適切な解を提示するが、利用者の制限時間である 15 秒が経過した時点で再検索は終了する。各段階での提示解を表 3 に示す。

## 4 まとめ

WWW 情報資源の公開する情報を用いた問題解決について紹介した。情報を mediator に蓄積することにより、WWW 情報源に頻繁にアクセスできない状況下においても、解を求めることができる。また、信頼度を用いることで再検索の効率がよくなる。

他の例題としては、コンピュータ部品の組合せを提案するシステムがある。複数のバーチャルショップからパソコン本体やディスプレイの性能や値段を取得し、利用者の要求を満たす商品の組合せを提示する。

今後は、例題に挙げた航空機の旅程プランニングシステムの実装を行い、解の質や信頼度などの妥当な計算方法を調査する。

他には、再検索する事実を決定するアルゴリズムの改良がある。今回の例題の場合、 $f_0$  は 4 つの解で使用されている。そのため、 $f_0$  を再検索すると、4 つの解すべての信頼度が改善される。このように、より効率的な再検索のために、再検索する事実を決定するアルゴリズムを改良する。

## 参考文献

- [1] 野田知哉, 北村泰彦, 辰巳昭治, WWW 情報統合のための協調型アーキテクチャ, 人工知能学会研究会資料, SIG-KBS-9803-8 (1999).