

Web コンテンツの巡行経路に基づくアクセス管理

清 光 英 成[†] 田 中 克 己^{††}

本論は、ハイパーテキスト文書の意味的なまとまりを制作者自身が定義し、制作者の意図どおりに利用者を閲覧させることを目的とする。そのため、制作者が予め意味的なまとまりとして意味的情報単位を定義でき、制作者の意図どおりに利用者を閲覧させることができるリンク機構、状況依存リンクを提案する。また、意味的情報単位内のどこから閲覧を開始しても等価な情報をえられるようなハイパーテキストシステムを考察する。そこでは意味的情報単位内で閲覧を省略されたハイパーテキスト文書の情報を補完する枠組を提案すると共に、利用を許可できるアクセスかそうでないアクセスかを、キーページ概念を導入して判別する。これにより Web コンテンツのアクセス管理を利用者の巡行経路に基づいて行うことができるようになる..

Access Management for Dynamic Web Contents Based on User's Navigation over Hypertext Documents

HIDENARI KIYOMITSU[†] and KATSUMI TANAKA^{††}

In this paper, we propose the idea of changing the view of Web pages and their linking structure based on trace of user's link navigation. The major objective of the proposed methodology is to reflect Web-page author(creator)'s intention over the web documents and its linking structure. The view of each Web-page including its hyperlinks is variable according to which navigation path the user took to reach the Web-page. If user takes a navigation path which is intended by the author, the system will provide necessary and sufficient view of the Web-page including enough hyperlinks. In order to realize the proposed navigation-path-dependent Web-view, first we will introduce and formalize the syntax and semantic of "semantically meaningful paths", which are usually defined by the author. Then, we will define the equivalence relationships between two semantically meaningful paths. The equivalence relationship is useful to find redundant paths defined by authors. Also, we investigate a methodology to make up for the document information missed when user's navigation was against the semantically meaningful paths defined by author.

1. はじめに

ハイパーテキスト文書は柔軟なリンク機構により文書間の関係を容易に表現できる。またその閲覧もリンクを辿るだけでなく、ブックマークや URL の入力による直接アクセスも許す。一方、作者は複数のページをリンクで連結して一連の意味を表現しようとする。そのため、利用者が作者の意図どおりにリンクを辿らなければ情報が十分に伝わらないだけでなく誤解も生じる可能性がある。そこで、本研究は

- ハイパーテキスト文書の意味的なまとまりを、制作者が予め定義できる意味的情報単位概念を

導入

- 利用者の巡行経路に応じて、動的に航行可能なリンクを有効にする状況依存リンクを提案
- 意味的情報単位の semantics を到達可能経路間の関連から等価性を導き形式化した¹⁾。さらに本論では、意味的情報単位内で到達可能経路集合の元とならないアクセス、つまり意味的情報単位の先頭ページ以外からの閲覧に対して、動的に情報を補完する手法を提案する。

複数のハイパーテキスト文書から構成された、一連の情報を検索する研究が行われているが、それらは、WWW 上から論理的なまとまりを発見することを目的としている⁷⁾。そのため制作者の提供したい意図には焦点をあてておらず、既存のデータから情報単位を発掘するというアプローチであるといえる。

一方、本論は制作者が予め意味的なまとまりとして意味的情報単位を定義でき、状況依存リンクを用いて

[†] 神戸大学経済学部

Faculty of Economics Kobe University

^{††} 神戸大学大学院自然科学研究科

Graduate School of Science and Technology Kobe University

制作者の意図どおり利用者が閲覧できる機構を提案している。また、利用者の閲覧順序に対する制約という観点からだけでなく、見てほしい情報を利用者に対して過不足なく配信することを目的としている。これにより、制作者側からの経路補完サービスが実現でき、利用者が検索に費やした労力を有効に反映するような情報提供が可能になる。このアプローチは利用者が膨大な経路候補を情報検索システムから得て、その梗概から次の情報資源を選択するのではなく、制作者側から経路候補を提供するものである。

2. 意味的情報単位

コンテンツ作者は利用者が A, B, C の順にリンクを辿って閲覧することで一連の意味を理解できると考えれば、図1の(i)のような AB, BC のリンクを定義する。また、 B または C を A の閲覧の後に閲覧させたい場合には図1の(ii)のように AB, AC のリンクを定義する。

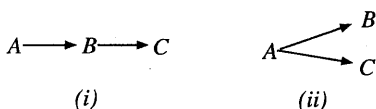


図1 既存のリンク構造

図1(i)では閲覧順序が決まっているが、図1(ii)では B が先に閲覧されても C が先に閲覧されてもよい。しかしながら、 B, C の閲覧は A の閲覧の後でなければならない。

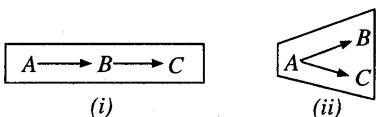


図2 意味的情報単位

図2(i)は A から閲覧を開始すれば B へのリンクが有効になり、その後 B を閲覧すれば C へのリンクが有効になるような意味的情報単位を多角形で示している。一方、利用者が B または C から閲覧を開始すれば、利用者の巡行経路に A が存在しないので状況依存リンク BC が利用者に見えない。本論における意味的情報単位とは状況依存リンクを用いてこのような性質を表現できる有向非巡回グラフである。

ここで問題となるのが、利用者は隠されたリンク先のコンテンツの存在を知る権利を奪われるということである。本論では、意味的情報単位の先頭ページ以外

からの閲覧開始に対して、それ以前に閲覧されるべきであったのに閲覧されなかったページの情報を利用者に知らせる機構を提案する。

3. 到達可能経路

意味的情報単位は制作者がハイパーテキスト文書で一連の意味を表現するために定義する有向非巡回グラフ構造をもつページ集合であるが、議論を簡単にするため、本論では意味的情報単位を経路に限って整理する。

意味的情報単位 $p = v_1 v_2 \dots v_n$ において意味的情報単位 p 中のコンテンツ v_i ($1 < i \leq n$) に到達可能とは v_1 から順に辿り、 v_{i-1} に到達していることをいう。このとき $v_1 v_2, \dots, v_1 v_2 \dots v_n$ を到達可能経路と呼び、 $Path(p)$ と記述する。また、 v_i に既に到達していれば v_i に到達可能であるとすると。つまり、到達可能経路は利用者がコンテンツ作者の意図通りに v_i ($1 \leq i \leq n-1$) に到達すれば v_{i+1} への経路 $v_i v_{i+1}$ が航行可能になるということの意味する。

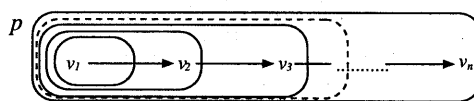


図3 到達可能経路

逆に、利用者の巡行経路に p の到達可能経路が存在しなければ、状況依存リンクが既存のリンクを隠蔽し、制作者が定義した閲覧順序を維持する。

3.1 到達可能経路間の関係

意味的情報単位 $p_1 = v_1 v_2 \dots v_n$ が他の意味的情報単位を部品として持つ場合、例えば他の意味的情報単位を $p_2 = v_k v_{k+1} \dots v_{k+m}$ ($1 \leq k \leq n, m \leq n-k$) とすると、 p_2 の到達可能経路は $Path(p_2) = v_k, v_{k+1}, \dots, v_{k+m}, v_k v_{k+1}, \dots, v_k v_{k+1} \dots v_{k+m}$ である。意味的情報単位 p_1 と p_2 の到達可能経路 $Path(p_1), Path(p_2)$ の間には $k=1$ または $k=n$ において、

$$Path(p_2) \subset Path(p_1)$$

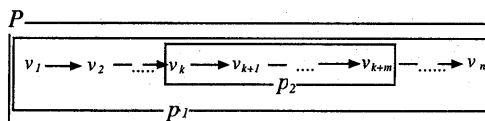


図4 二つの意味的情報単位

の関係があるが、 $k \neq 1, k \neq n$ ではこの関係が成立しない。しかしながら、 p_1 は p_2 上の経路であり、意

味的情報単位 p_1 上の全ての意味的信息単位を P とすると, $p_1 \in P, p_2 \in P$ が成り立つ. ある意味的信息単位 p_1 中に他の意味的信息単位 p_2, p_3, p_4 が図5のように連結がなく含まれる場合, p_1 の到達可能経路上の全ての意味的信息単位集合 P の到達可能経路は

$$Path(p_1) \cup Path(p_2) \cup Path(p_3) \cup Path(p_4)$$

であり,

$$Path(P) = \cup_{p_i \in P} Path(p_i)$$

と記述することができる.

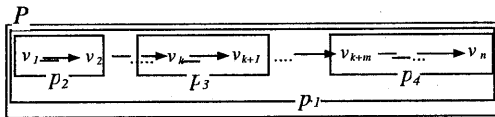


図5 複数の意味的信息単位

一方, 二つの意味的信息単位において図6のような p_1 に p_2 の先頭コンテンツが含まれる場合, p_1 は p_2 と連結しているという. つまり, 利用者は p_1 上を巡回している間に p_2 の先頭コンテンツを閲覧するので, p_2 上の経路を巡回できる.

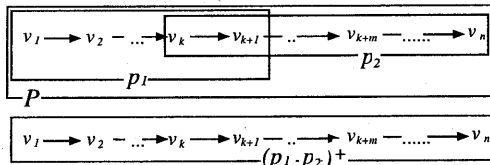


図6 意味的信息単位の連結

このとき, 意味的信息単位集合 P の到達可能経路は $Path(p_1)$ と $Path(p_2)$ の和だけでなく, p_i と p_j 連結して得られる $v_1 v_2 \dots v_k \dots v_m$ という経路の到達可能経路も含んでいる. p_1, p_2 を連結してできる到達可能経路を $Path((p_1, p_2)^+)$ とかく. 本来, 記号 R^+ は閉包であるが, 本論では連結によって導出される到達可能経路という意味で用いている. したがって, 意味的信息単位集合 P の到達可能経路は

$$Path(P) = Path(p_1) \cup Path(p_2) \cup Path((p_1, p_2)^+)$$

と記述できる.

さらに一般化するために, 意味的信息単位集合 P 中の意味的信息単位 p_i, p_j, \dots, p_m が連結しているとは, 各意味的信息単位の先頭コンテンツを $p_{i_1}, p_{j_1}, \dots, p_{m_1}$ としたとき,

$$p_{j_1} \in p_i \wedge p_{j_1} \neq p_{i_1}$$

$$\wedge p_{(j+1)_1} \in p_j \wedge p_{(j+1)_1} \neq p_{j_1}$$

$$\wedge \dots$$

$$\wedge p_{m_1} \in p_{m-1} \wedge p_{m_1} \neq p_{(m-1)_1}$$

を満たすことをいう. また, 意味的信息単位集合 P 中の全ての連結を P^+ と表す. これにより, 意味的信息単位集合 P の到達可能経路は

$$Path(P) = \cup_{p_i \in P} Path(p_i) \cup Path(P^+)$$

と一般化できる¹⁾.

4. 状況依存リンクによる情報補完

意味的信息単位は状況依存リンク機構を用いてコンテンツ作者の制作意図を利用者の閲覧順序に反映させることが目的である. しかしながら, 作者の意図に反した閲覧に対してリンク先のコンテンツをただ隠蔽するだけでは利用者の利益にはならない. そこで, 意味的信息単位の先頭コンテンツ以外へのアクセスにより欠如した情報を補完する機構を提案する.

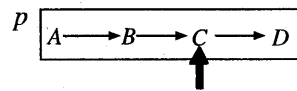


図7 先頭ページ以外へのアクセス

図7は意味的信息単位の先頭コンテンツ以外へのアクセスを示している. 図7の意味的信息単位 P の到達可能経路は AB, ABC, ABCD であるから, 意味的信息単位 P に対して C への直接アクセスは状況依存リンクが CD という経路を存在させない. ここで, 例えば図8のように C を表示している間 A, B のコンテンツを補助ウィンドウに表示させることで省略された閲覧を補うことができる.

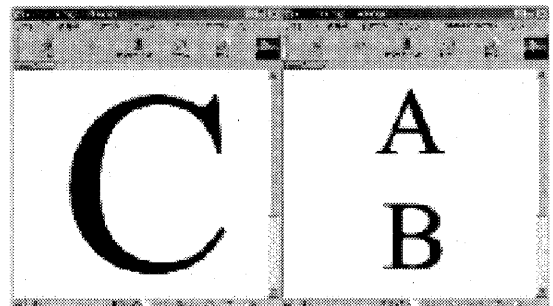


図8 情報補完の例

また, A, B の情報を補完された C の閲覧は, 状況

依存リンクによって D への経路 CD を有効にする。

4.1 複合的な意味の情報単位の情報補完

意味の情報単位は任意に制作者が定義できる。したがって、図 7 のような単純な場合だけでなく複雑な構造の意味の情報単位の情報補完も議論する。

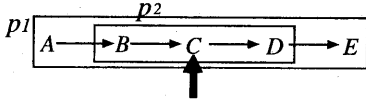


図 9 意味の情報単位の先頭ページ以外へのアクセス

図 9 は二つの意味の情報単位 p_1, p_2 に対して利用者が C から閲覧を開始しようとしていることを示している。意味の情報単位上の情報補完は先頭ページ以外でアクセスされたページを理解できるようにするものである。したがって、上述の p_2 では C を理解するためには B の理解が必要である。同様に p_1 では C を理解するためには B の理解が必要かつ B を理解するためには A の理解が必要である。また、利用者が意味の情報単位 p_1 にアクセスしたのか p_2 にアクセスしたのかシステムが知る由もない。更に、C を経路上に持つ全ての意味の情報単位の C より前のページ情報を補完することは適切でない。そのため、本論では C を理解するために必要な最低限のページ情報を補完することにする。この例の場合、C を理解するためには最低限 B の閲覧が必要になる。換言すれば、C を理解するために極小の情報補完をするためには C より前に極小な意味の情報単位 p_2 の状況依存リンクによる情報補完を行う。

【定義 1】

利用者の閲覧開始ページを v_e としたとき、 v_{e-1} を要素として含む全ての意味の情報単位集合を P とする。また、 $length(p_i, v_e) \{p_i \in P\}$ を意味の情報単位 p_i の先頭ページから v_e までの長さとする。このとき、 $length_{min}(p_i, v_e)$ となるような意味の情報単位を v_e より前に極小な意味の情報単位と呼び、 $P_{min}(v_e)$ と書く。またこのとき、 $length(p_i, v_e) \leq length(p_j, v_e) \{p_j \in P\}$ ならば、 p_i の v_e 以前の経路は p_j の v_e 以前の経路上に存在しなければならない。

【定義 2】

意味の情報単位 p_i の v_e 以前の経路が p_j の v_e 以前の経路上に存在するとは、

$$p_i = v_{i1} \dots v_{ik} v_e \dots v_{im}$$

$$p_j = v_{j1} \dots v_{jl} v_e \dots v_{jn} \{k \leq l\}$$

において

$$v_{ik} = v_{jl}, v_{ik-1} = v_{jl-1}, \dots, v_{i1} = v_{j1-k+1}$$

が成り立っていることをいう。

またこのとき、 $k=l$ ならば、 p_i, p_j ともに v_e より前に極小な意味の情報単位である。

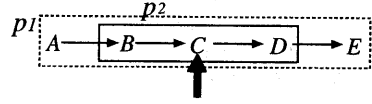


図 10 補完後の navigation

図 10 のように情報補完後は C より前に極小な意味の情報単位の状況依存リンクが機能し、それに従った巡行を利用者に促す。

一方、図 11 のような二つの意味の情報単位が C で交わっている場合はどのように処理すれば良いのだろうか。

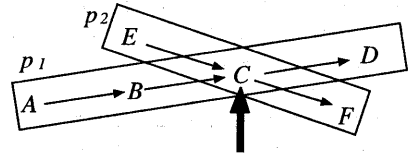


図 11 交差する二つの意味の情報単位

C を要素として含む意味の情報単位集合の元は p_1, p_2 である。 $length(p_1, C) > length(p_2, C)$ であるが、 p_2 の C 以前の経路は p_1 の C 以前の経路上にない。この場合 p_1, p_2 共に C より前に極小な意味の情報単位と判別する。つまり、作者は p_1, p_2 に異なる意味を与えており、利用者が p_1 でも p_2 でも巡行できるように C 以前のページ情報を補完する。これにより、C の閲覧以降は意味の情報単位 p_1, p_2 の状況依存リンクが機能し、それに従った巡行を利用者に促す。

もう少し複雑な例を示す。図 9, 11 の合成のような図 12 の場合、

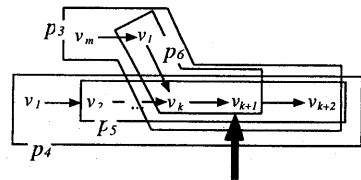


図 12 極小な意味の情報単位

各分岐の意味の情報単位集合 P_i で v_e より前に極小

な意味的情報単位 $P_{i,min}(v_e)$ の集合

$$P_{min}(v_e) = \cup_{P_i \in P} P_{i,min}(v_e)$$

が補完すべき極小のページ情報となる。このとき、 v_e 閲覧後の利用者の巡行経路は全ての $P_{min}(v_e)$ の状況依存リンクが提供する。

4.2 情報補完の例

ここでは小規模な組織を例として情報補完が行われる例を示す。ある組織が Web で情報発信するとき“www.econ.kobe-u.ac.jp”などのホームページを根とする木のような構造で各ページをリンクが結ぶのが一般的である。

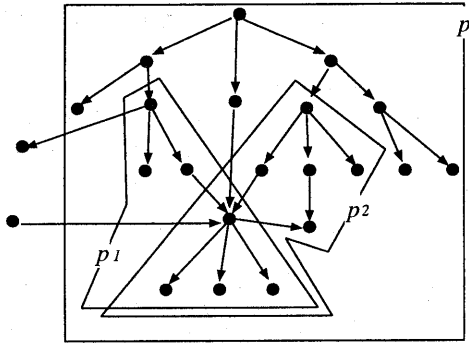


図 13 情報補完の例

図 13 はある組織全体を意味的情報単位 p と定義している。またその中で幾つかの意味的情報単位が定義される。外部からリンクされるページが教官 T と学生 S の共同研究文書であり、研究内容を表現する意味的情報単位 $p1, p2$ を定義している。ここで、共同研究文書に外部からリンクがつけられ、 $p1, p2$ の先頭ページ以外にアクセスされる状況を考える。このとき、意味的情報単位 $p1, p2$ の状況依存リンクはそれぞれの先頭ページから共同研究文書ページ直前までの経路上に存在するページ情報を補完し、利用者の閲覧を補助する。この例から、意味的情報単位が木構造となっても情報補完が可能であることがわかる。

次に三回で終る講義を意味的情報単位 *Lecture 1* として定義し、各回を意味的情報単位 *Section 1, 2, 3* と定義した例を示す。

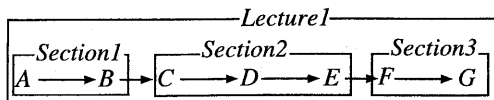


図 14 情報補完の例 2

ここに受講者が C から受講を開始するときを考える。C は意味的情報単位 *Section 2* の先頭ページであり、状況依存リンクによって D へのリンクを航行することができるようになる。しかしながら、このアクセスに対して C より前に極小な意味的情報単位 $P_{min}(C) = \text{Section1}$ である。そこで、意味的情報単位 *Section2* の状況依存リンクは意味的情報単位 *Section1* の要素 A, B のページ情報を補完し、意味的情報単位 *Lecture 1* の到達可能経路 ABC を得る。一般に *Lecture 1* の講師は *Section 2* の冒頭で前回の内容、つまり *Section 1* を概観するのでこの情報補完処理は自然である。

一方、D から受講を開始した場合、D より前に極小な意味的情報単位 $P_{min}(D) = \text{Section2}$ であるから、C の情報を補完し、到達可能経路 CD を得る。従って、*Section2* の状況依存リンクは E へのリンク DE を有効にする。このとき、受講者の巡行経路は CDE であるから *Lecture 1* の到達可能経路を含まないので、F へのリンク EF は有効にならない。このように、状況依存リンクは利用者の閲覧開始ページから補完すべき情報を判別し、更に巡行経路に制限を設けることができるようなアクセス管理に利用できる。

5. 議 論

閲覧を省略されたページの情報を補いながらハイパーテキスト文書を利用者に巡行してもらうために意味的情報単位概念を示した。そこでは制作者が複数ページで構成したハイパーテキスト文書の一連の意味を利用者に理解してもらうために状況依存リンクを導入している。

本論で意味的情報単位を経路の側面からのみ議論したが、本来は有効非巡回グラフである。例えば、ある節点から複数の枝が出ているような場合、各枝の間に順序や選択、択一のような意味を持たせることができる。表現力は高まるが、意味的情報単位の定義が複雑になり、状況依存リンクを機能させるための到達可能経路の計算がそれと共に複雑になる問題がある。

更に、意味的情報単位をグラフとして論じるとき、図 14 の例を階層構造グラフと捉えることができる。図 14 の例では一次元上に *Lecture 1* が展開されたが、意味的情報単位 $\text{Lecture1} = \text{Section1Section2Section3}$ というように定義されることが考えられる。つまり、現存する Web コンテンツも線形にリンク構造を定義されても階層的に定義されても、ハイパーテキスト文書制作者の意図は等価であるからである。

意味的情報単位の先頭ページ以外へのアクセスによ

る情報補完について議論しているが、補完すべき場合とすべきでない場合が存在する。意味的情報単位中のページ v_k を閲覧していなければそれ以降のページの閲覧を許さないという要求が考えられる。このとき v_k を本論ではキーページ (key page) と呼ぶ。状況依存リンクによる情報補完機構は v_k より前に極小な意味的情報単位を探索する工程でキーページを発見したときキーページを含む分岐中の意味的情報単位の情報補完を行わず、それらの意味的情報単位で v_k 以降の巡回経路を隠蔽する。これは例えば「専門科目 S の受講および聴講は共通科目 G の単位取得者に限る」というような制約的な閲覧条件を記述するために利用できる。

本論で議論した意味的情報単位は、ある一連の意味を表現するためのページ集合を状況依存リンクで繋いだグラフである。そのため、リンク情報を各ページコンテンツ内に持たせるのではなく、一元的に管理したい。ここで問題となるのが、利用者の巡回してきたページの履歴を既存のハイパーテキスト言語のリンク記述で十分に表現できないことである。UNIX 上の Netscape Navigator では `/.netscape/history.db` に巡回の履歴が保存されているが、これは利用者権限でしか利用できない。そのため、状況依存リンクは各ページに埋め込んだ行内リンクではなく、利用者のクライアントプログラムから呼び出すような行外リンクを利用して実装することが現実的である。

6. ま と め

ハイパーテキスト文書制作者の情報伝達意図をリンク構造に反映する意味的情報単位を提案し、利用者の巡回してきた履歴から次に利用者に見えるページを動的に提供する状況依存リンクの概念を示した。また、利用者が制作者の意図に反した閲覧を行っても意味的情報単位が発信する一連の意味を誤解なく伝達するために利用者が閲覧を省略したページ情報を状況依存リンクによって補完する方法を述べた。そこでは、ただそのページよりも前に極小な意味的情報単位のページ情報の補完だけでなく、講義の例で議論したとおり文脈を意識した情報補完も有効であることを示した。また、情報コンテンツへのアクセスが著しく制作者の情報伝達意図に反している場合にはキーページと利用者が閲覧を開始しようとしているページの関係からアクセスの制限を行うことも有効ではないかと考えられる。

これらにより、ハイパーテキスト文書制作者の情報伝達意図をリンク機構に反映するだけでなく利用者の利益も損なうことがない。そこでは状況依存リンクに

よる閲覧を省略されたページの情報を補完しつつも、著しく制作者の情報伝達意図に反する利用者のアクセスを排除することで、意味的情報単位を用いて表現したい一連の意味をより誤解なく伝達することができるようになる。

< 謝辞 >

本研究を進めるにあたり、多くの御意見と助言を頂いた本学工学部自然科学研究科田中研究室の皆様へ深く感謝致します。また、研究環境において御支援頂きました、本学経済学部 Keizo Nagatani 教授、三谷直紀教授と玉岡助教授、流通科学大学情報学部高橋秀行教授に深く感謝致します。

本研究の一部は、日本學術振興会未来開拓學術推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」(プロジェクト番号 JSPS-RFTF97P00501) および、文部省科学研究費重点領域研究「高度データベース (No. 275)」(課題番号 08244103) による。ここに記して謝意を表します。

参 考 文 献

- 1) 清光英成, 田中克己, Web グラフにおける意味的情報単位の状況依存リンク, 情報処理学会研究報告, DBS-99-118, Vol. 99, No. 39, pp. 41-48, 1999.
- 2) P. David Stotts, Cyrano Ruiz Cabarrus, "Hyperdocuments as Automata: Verification on Trace-Based Browsing Properties by Model Checking" *ACM Transaction of Information Systems*, Vol. 16, No. 1, pp. 1-30, 1998.
- 3) Jean-Hugues Réty, "Structure Analysis for Hypertext with Conditional Linkage", *Proc. of Hypertext '99*, pp. 135-136, 1999.
- 4) Frank M. Shipman III, Catherin C. Marshall, Mark LeMere, "Beyond Location: Hypertext Workspace and Non-Linear Views", *Proc. of Hypertext '99*, pp. 121-130, 1999.
- 5) Keishi Tajima, "Querying Composite Object in Semistructured Data", *Proc. of 5th International Conference on Foundation of Data Organization*, Nov. 1998.
- 6) Kenji Hatano, Ryouichi Sano, Yiwei Dan and Katsumi Tanaka, "An Interactive Classification of Web Documents by Self-Organization Maps and Search Engines", *Proc. of DAS-FAA '99*, pp. 19-25 1999.
- 7) Wen-Syan Li and Yi-Leh Wu, "Query Relaxation by Structure for Web Document Retrieval with Progressive Processing", *Proc. of Advanced Database Symposium '98*, pp. 19-25, Dec. 1998.