

自動リンク付けの研究

泉 寛幸 松岡秀達 松井くにお

(株)富士通研究所マルチメディアシステム研究所

情報共有化の研究調査の一例として、研究所内の毎月の研究報告(月報)を文書データベースに蓄積し、ユーザの自由なときに検索・閲覧できる月報管理システムを試作している。本報告では、リンク元となる語句と、リンク先の文書(手続き)とが確定している場合について、リンク先とリンク元との対を記載したリンク付け表を用いて、月報文書内に参照タグ(リンク)を付与する手法の適用事例を示す。その際のリンク付け表の応用として、月報閲覧時にSGML文書からHTML文書に変換する際にリンクを動的に付加する手法、リンク先としての検索手続きの起動においてのSGMLの構造タグの利用法、構造タグを利用しての文脈に応じたリンク付け表の切替え法などを示す。

A Study of Auto linking

Hiroyuki IZUMI, Hidesato Matsuoka, Kunio Matsui

Multimedia System Laboratories, FUJITSU LABORATORIES LTD.

For a study of information sharing, we are developing a monthly report management system which stores monthly research reports of our laboratories with SGML forms in the document database, and enables the users to read and look up them at any time. This report shows a method of adding reference links to the monthly report documents automatically, using the tables which include the pairs of the phrase(i.e. link source) and the URL(or procedure)(i.e. link destination) of every link, where each source and destination of links were already given. Moreover, we show some applications of this linking tables, a method of dynamic linking on the transformation of the reports from SGML format to HTML form when the user reads them, a method of information retrieval using the tags of SGML form, and a method of changing the linking tables using the tags of SGML.

1. はじめに

情報共有化の研究調査の一例として、研究所内の毎月の研究報告（月報）を月報データベースに蓄積し、ユーザの自由なときに検索・閲覧できる月報管理システムを試作している。

情報共有においては、蓄積されている文書の検索と再利用のために、文書の内部構造を解析して、章・節・パラグラフなどの構造タグおよび、参照情報への参照タグ（リンク）を付与しておくことが有効視されている。我々は、月報を一例として、参照タグ（リンク）の自動付与の研究を行なっている。

本報告では、特に、リンク元となる語句と、リンク先の文書（手続き）とが確定している場合について、文書内に参照タグを付与する手法を、月報文書に適用した結果を報告する。

2. 月報管理システム

2.1. 概要

今までは、各研究のグループリーダーは、毎月の研究の進展状況を、月報と呼ばれるA4一枚の書式にまとめて印刷して上司に提出していた。上司の承認印の承認印の押された紙文書は、まとめて保管されていた。従来方式では、紙ベースのため月報の写しを関係者に配布したり、回覧する手間が大きく、社内での情報共有という点で不満足なシステムであった。そこで、我々は、情報共有を押し進めるべく、研究所内の毎月の研究報告（月報）を月報データベースに蓄積し、ネットワークとWWWブラウザを通じてユーザの自由なときに検索・閲覧できる月報管理システムを試作している。

月報作成者は[作成者]や[年月度]や[項目]などの簡易タグを付加した平文で月報を記載して、メールで月報管理システムに送る。システムは、その月報を承認者にメールとして回送する。承認者から月報管理システムに承認メールが送られたならば、システムは、その月報平文をSGML形式に変換して蓄積する。蓄積された文書は、月報閲覧者の希望に応じて、

- ・HTML形式に変換され、参照情報へのHTMLリンクを付けられてWWWブラウザで閲覧される。
- ・平文に戻されて、関係者にメールで送られる。 ・Latex形式に変換されて印刷される。
- ・SGMLタグを利用して、範囲をしばりこんだ様々な検索が行われる。

などの操作が行われ、情報の共有化がなされる。月報はSGML化されることによって、HTMLやLatexへの変換が容易になった。また、パラグラフ単位にタグ付けされているので、文書の検索や再利用も容易になり、数ヶ月毎のマトメの作成支援などにも利用できる。以下、SGML内の月報の参照情報について説明する。

2.2. 月報の入力形式

従来の月報は、報告事項や特記事項や部長意見などを記載する罫線枠の用意されたワードプロセッサテンプレートが用意されており、その枠の中に、報告文を記載して、紙として印刷出力されていた。月報文書を処理するシステム開発において、従来のテンプレートでは、罫線枠は電子メールで送りづらい。かつ、罫線枠を処理する文書処理ツールは少なく、開発も難しい、といった問題点があった。

罫線枠は、全体を読みやすく印刷することと、月報の各部分の「報告事項」「部長意見」といった役割を示す機能がある。全体を読みやすく印刷するためには、Latexのような文書整形ツールを用いることである程度解決できる。

重要なのは、文書を「報告事項」「部長意見」のように構造化しておく機能である。文書がある程度構造化されていれば、文書整形ツールの入力もしやすいし、様々な観点からの検索もしやすい。文書の構造化手法としてSGML形式があるが、一般のユーザにはSGML形式での入力は難しい。そこで、月報作成

者に下記のような【報告事項】【部長意見】などの簡易タグを用いて、月報平文を記載してもらうことにした。簡易タグは、SGMLを(ましてや DTDを)なるべく意識させずに、月報の作成が行えるように留意して作成した。簡易タグにより平文の段階である程度の構造化がなされており、SGML化を容易にしている。また、文章途中の空行はパラグラフの終了とみなされ、パラグラフ開始タグ<p> と</p>とが自動的に付加されるようにした。今のところ作成者にとって、簡易タグそのものは、あまり心理的負担が大きくないという感触を得ており、エディタやメールの操作に慣れている作成者には好評である。 —

簡易タグ付き月報平文	SGML化された月報
【月報】	月報<
【月報作成者】 ○○○○	ヘッダ<
【年月度】 97年2月度	作成者<NO>830287<姓> ○○ <名> ○○
【研究題目】 文書情報処理	<所属> ○○部 <役職> 主任研</作成者> <
【報告事項】	承認者<NO>830287<姓> ○○ <名> ○○
[項目] 文書検索	<所属> ○○部 <役職> 部長 </承認者> <
.....	年月度> ○○年○○月度</年月度>
.....	発行年月日> ○○○○、○○○○</発行年月日>
.....	送付先> ○○○○、○○○○、..... </送付先>
.....	研究題目> ○○○○</研究題目>
[項目] 文書構造化	</ヘッダ>
.....	本文<
【特記事項】	報告事項
.....	<項目><名> 文書検索</名>
【部長意見】	<P>..... </P>
.....	<P>.....
【以上】 </P></項目>
	<項目><名> 文書構造化</名>
	<P>.....
 </P> </項目>
	特記事項> </特記事項>
	部長意見> </部長意見>
	</報告事項> </本文>
	</月報>

2.3. データベースとの連携

月報管理システムは、以下の3つの種類の簡単なデータベースをもっている。

- (a) 月報関係者のデータベース：月報作成者(承認者)の、従業員番号(ID)、姓、名、メールアドレス、所属、役職など
- (b) 研究題目のデータベース：各研究題目ごとの月報に対しての、研究題目ID、研究題目名、月報作成者のID、承認者のID、月報の送付先など
- (c) 組織のデータベース：各部のID、部の名前、部のホームページアドレスなど。

これら(a)(b)(c)の情報は、月報作成者から送られてきた月報文書の簡易タグをもとに検索され、

- (1) その月報の研究題目から、(b)により月報承認者を調べ、(a)により承認者のメールアドレスを調べて、月報を承認者に転送するとき、あるいは、
- (2) 研究題目から、(a)と(b)により月報作成者と承認者の役職やアドレスや月報の送付先を調べ、それらの情報を必要に応じてWWWブラウザで表示したり、Latex 印刷に用いるときなどに使われる。

基本的に上記の3種の情報は、月ごとに変化しないものとみなされる。それゆえ、データベース化され毎月同じ情報が使われる。従って、当初の月報管理システムにおいては、WWWブラウザでの表示や印刷のときに、上記のデータベースから参照されて表示や印刷情報として用いられていた。

しかしながら、実際には、組織変更により月報作成者の役職や所属、月報の送付先などは、数ヵ月ごとに変わることがあった。そのため、半年前の月報を表示したのに、最新の月報作成者の役職や所属や月

報送付先が、古い月報文書と共に表示されてしまうという問題があった。

従って、現在の月報管理システムにおいては、数ヶ月ごとに変化する可能性のある月報作成者・承認者の役職や所属、月報の送付先などは、データベースから読み込んで個々の月報SGML文書のヘッダ内に対応するタグとともに埋め込むことにした。

すなわち、月報の参照情報として、内部情報としてSGML文書内に埋め込んでしまうものと、リンクによって外部参照されるものがある。以下、外部参照情報へのリンク付けについて説明する。

3. HTML文書のリンク付け

3.1. 月報からの参照情報

月報閲覧者にとっては、月報とあわせて下記のような情報をも閲覧したい。

- (1) 報告書内に記載されている下記の種類の語句に関してより詳細な解説情報を見たい。すなわち、専門用語、研究計画、工程、関連研究、関連部署名（組織データベースへのリンク）、顧客名、論文名（論文データベースへのリンク）、特許名（特許データベースへのリンク）、より詳細な図や表、先/次月の月報 など
- (2) 月報に記載されているデモプログラムやデモビデオを起動・閲覧してみたい。
- (3) 専門用語の他の出現箇所を（索引として）提示してほしい。
- (4) 文書内の明記されたリンク先（URL）に容易に移動してみたい。

3.2. HTMLのリンク付け

月報文書をHTML形式で表示することにより、上記の情報の参照はHTML文書のリンク機能により実現可能となる。仮に文書内の語句XXXをリンク元としリンク先の文書を示すアドレス（URLと呼ぶ）を <http://www.xyz/file1.html> であるとしたとき、HTML文書内では、

```
月報文書例（リンク付けの例） -----  
作成中の<A href="http://www.xyz/file1.html">XXX</A>に関して、担当部署と・・・
```

というように、リンク先のURLを含む<A ..>ととで用語XXXを囲む。HTML文書においては、これでXXXにリンクが付加されたという。このとき、

```
月報文書例（リンク付けの例） -----  
作成中のXXXに関して、担当部署と・・・
```

WWWブラウザの画面上では、このXXXの部分の色が他と異なり、下線が引かれる。この色が異なる部分をクリックすると、リンク先（URL）である<http://www.xyz/file1.html> の（詳細情報の）文書がWWWブラウザに表示される。もし、URLが図やビデオやデモプログラムや検索を起動・表示するような手続きの名前であったならば、語句をクリックすると、その手続きが起動された結果がWWWブラウザに表示される。

月報作成者は、2の(1)～(3)に対しては月報文書内の語句に対して適当なURLをリンクとして付加することになる。(4)に対しては文書内に明記されたURLを語句とリンク先の両方に該当するとみなして前節のリンク付けを行えばよい。

3.3. リンク付けの半自動化

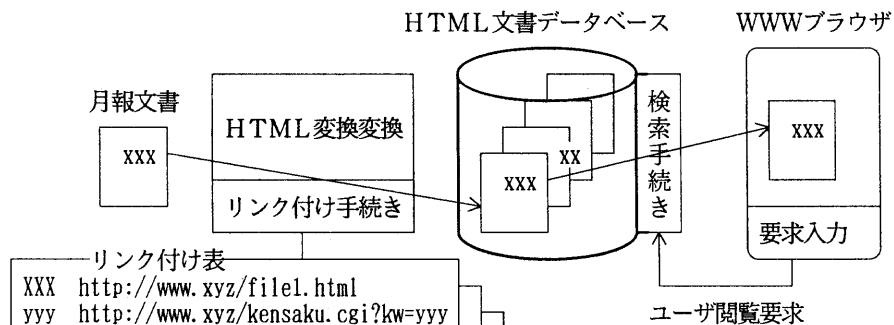
月報作成者が、外部情報への参照のために付与するリンクには、2種類考えられる。一つは、その月

だけの月報に付加しておくリンクである。例えば、その月の報告内容の詳細情報へのリンクである。この場合、月報作成者は詳細情報へのリンクURLを直接月報平文内に記載してもらうことにした。

月報管理システムは、月報文書内にURLが明示されている場合には、そのURL自体に前記のリンクをはってHTML文書をWWWブラウザで表示する。例えば、「詳細は、<http://www.xxx.zzz/...>を参照されたい。」というような文章を月報作成者が月報文書内部に記載したとき、月報管理システムは「詳細は、`http://www.xxx.zzz/...`を参照されたい。」というようなHTML文書に変換する。表示上は、「詳細は、<http://www.xxx.zzz/...>を参照されたい。」と下線でリンクが付与されていることが示される。

もう一つの参照情報は、例えば、2.0(1)に出現する語句に対するリンク先、すなわち用語解説文書へのリンクのような場合である。毎月の月報にその用語が出現するたびにその解説文書へのリンクをはらなければならない。リンク付けをすべて手動で行っていたときは、月報作成者は決まりきった語句のリンクを毎月、新しい月報文書内に埋め込まなければならないという問題点があった。

その解決法として、特定の語句とそれに対応するリンクとの対を記載したリンク付け表を、月報作成者に用意してもらうことにした。月報管理システムは、HTML文書に変換する際に、そのリンク付け表を見ながら月報文書内の語句にリンクを付加していく。



月報作成者は、あらかじめ一度だけリンク付け表に特定の語句とそれに対応するリンクとを記載しておけばよい。リンク付け表は、用語説明集用、検索手続き付加用、プログラム起動手続き用、その他任意のURL付加用など、複数種類用意することができる。

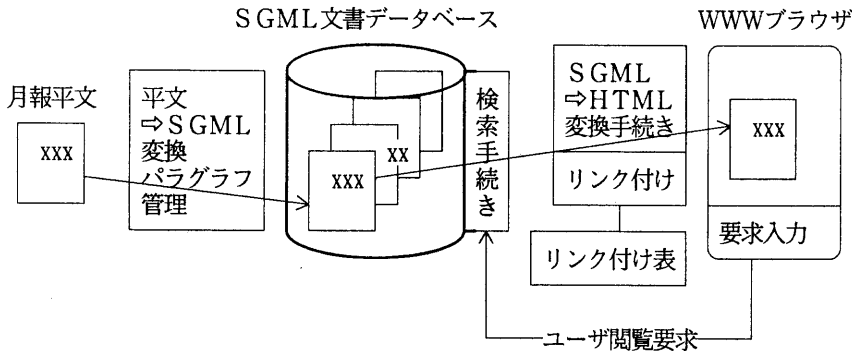
3.4. HTML文書の動的生成

基本的には、月報本文は公式報告であるから、月報が承認されて公開された後は、変更するのは望ましくない。しかし、・リンクの付け方には共通の付け方が存在していない、・どの語句にリンクを付加すればいいのか、月報作成者も経験をつんでいない、といったことから、リンクの付けかたは個々の作成者の試行錯誤にまかしている。したがって、月報作成者として、リンク付けの経験をつむにつれ、古い文書に対しても、語句にリンクを追加したり、リンクを削除したい、リンク先を変更したい、といった要望が存在する。リンクは、公式の報告とまだ見なされていないので、公開後であっても月報作成者が変えたいと思うならば、(できるだけ)変更可能なほうが望ましい。

当初の月報管理システムは、前図のように月報の毎月の分をそれぞれHTML文書化して、保存していた。この形式のときには、もし、これらの月報内の特定の語句のリンクを追加や変更しなくなったときには、個々のHTML文書ごとにリンクを張りなおさなければならなかった。例えば、HTML文書データベース内に、4月、5月、6月、7月と4ヵ月分の月報が存在し、どの月報にも含まれている「文書認識」という語

句のリンク先を変えたい場合、少なくとも4回リンクの書き直しをしなければならない。これは面倒である。

そこで、現在の月報管理システムは下図のように、月報の毎月の分をそれぞれSGML文書化して保存している。個々のSGML文書では（特に指定されていない限り、原則として）語句にリンク付けをしておかない。



そして、月報閲覧者が閲覧を希望したときに、SGML文書からHTML文書に変換して、同時にリンク付け表をもとにリンクを付けるべき語句にリンクを付加した形式で、月報閲覧者にその表示を見せている。

この場合に、月報閲覧者は、リンク付け表内の「文章認識」という語句のURLを一回だけ変更すればよい。さらに月報閲覧時に毎回リンク付け表を用いてリンクを付加するので、月報作成者は適当な時にリンク付け表内のリンクを更新できる。

4. リンク付け表の応用例

4.1. 複数のリンク

一つの語句に対して、その解説文書や用語説明や検索手続きやプログラム起動などの、複数のリンク（URL）を付加したくなることもある。

その解決法として、月報HTML文書においては、語句の末尾に複数の脚注にあたるマークを付加し、その脚注マークからリンク先へ飛べるようにした。

月報文書例（リンク付けの例）
 作成中のXXX#\$に関して、担当部署と・・・

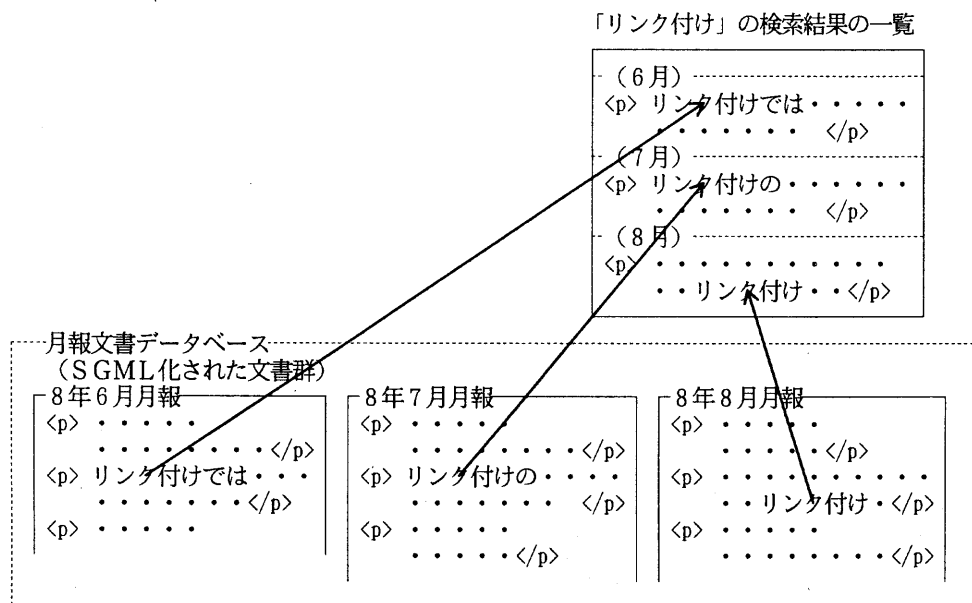
たとえば、上記の画面表示例において、XXXをマウスカーソルでクリックすると、XXXの解説文書へ飛び、\$をクリックすると、XXXの説明用デモプログラムが起動され、#をクリックすると、下記のようなXXXの用語検索手続きが起動される。

4.2. 用語検索機能（検索結果の一覧）

月報閲覧者は、特定の研究内容が過去にどのように進捗したかを知るために、特定の語句が、過去数ヶ月分の月報のどこに出現したかを検索したくなることもある。

この解決法として、特定の語句の検索要求が月報閲覧者から出されたときには、システムが下図のように、その語句の出現したパラグラフ（または項目）だけを数ヶ月分集めて表示する機能を持たせた。これによって月報閲覧者が特定内容に関して時系列情報や関連情報を容易に閲覧できるようになった。

これは月報管理システムが、月報をSGML形式でパラグラフ単位（項目単位）で蓄積しているために可能となった。



索引手続きの検索結果を示す例

この索引手続きは、基本的に全文検索であるが、SGML形式の各月報の<本文>と</本文>のタグの間の文書のみが検索され、<ヘッダ>～</ヘッダ>の間は検索されない。また、キーワードが<項目名>～</項目名>の間にあるときは、<項目>～</項目>の範囲を出力し、キーワードがパラグラフ内にあるときには、パラグラフ開始タグの<P> から終了タグの</P>までの文書を出力する。このように、SGMLタグを用いて、全文検索における検索の範囲や検索の出力形式を制御できる。

また、検索範囲を同一研究グループ内に限ったときには、そのキーワードに関する研究の時系列情報が得られ、検索範囲を研究所内に広げたときには、そのキーワードに関する研究所内の関連情報が得られる。抽出された情報は、パラグラフまたは項目の一覧として示されるので、そのキーワードに関する研究の全体像を見るのに都合がよい。

4.3. 文脈に応じたリンク付け

特定の語句に対するリンク先が文脈に依存せずに確定している場合には、前述のリンク付け表を用意することで、文書リンクを付加することができる。しかしながら、マニュアルなどの文書では、「主装置」と「周辺装置」とのそれぞれの章では、「オプション構成」という言葉に対してそれぞれ別のリンク先を設定したい。

また、月報文書の実例としても、下記右図のような例が、散見されたが、この研究項目XXXのときの「事業部」と研究項目〇〇〇の「事業部」とは異なる事業部であり、リンク先として参照するときには、違うリンクをはりたい。すなわち、同一の語句に対して文脈に依存したリンク先を設定したい。

しかしながら、現在のリンク付け表では、文脈は考慮されていないから、同じ語句がどこに出現しても同じリンク先が付けられてしまう。文脈を考慮して、同じ語句でも違う位置にある語句には、違うリンクを張ることはできなかった。例えば、マニュアル全体では、「語句1」は「リンク先」へリンクす

るが、特定の章か節でだけは、「語句1」は「リンク先2」へリンクする、といったように、全体とその部分としての章や節に応じて、リンク付け表の切替えをすることもできなかった。

<p>1. 研究題目XXX 事業部から、.....</p> <p>2. 研究題目OOOを事業部に.....</p>	<pre><章, 番号=2, リンク付け表="Table2"> <節, 番号=2.1, リンク付け表="Default" > ..事業部から..... </節> <節, 番号=2.2, リンク付け表="Table2.2"> を事業部に..... </節> </章> <章, 番号=3, リンク付け表="Table3"> </pre>
---	--

その対策として、上記左図のように、文書全体がSGMLで構造化されている場合、個々の章や節のタグにリンク付け表を指定して、切り替えることを検討している。

5. おわりに

我々は、情報共有における情報ナビゲーションのために、リンク元となる語句と、リンク先の文書（手続き）とが確定している場合について、文書内に参照タグ（リンク）を自動的に付与する手法を報告した。さらに、リンク先の検索手続きが起動され、構造タグにより検索範囲がしぼられて、検索内容の一覧ができるように、リンク付けとタグ付けと検索とを融合する手法の有効性を示した。ここで開発した手法は、他の文書データベースにおいても適用可能である。

さらなるリンクの自動付与のために、リンク元となる語句の自動的な選択法、リンク先の文書の容易な構成法、両者の簡単な結合法などを検討している。

参考文献

- [1] 松岡他：「SGMLを利用した情報共有とナビゲーションI－情報共有－」，情処学会第53回（平成8年後期）全国大会予稿 2K-4，分冊4 pp.233-234，1996.
- [2] 泉他：「SGMLを利用した情報共有とナビゲーションII－ナビゲーション－」，情処学会第53回（平成8年後期）全国大会予稿 2K-5，分冊4 pp.235-236，1996.
- [3] ロー・リメイ 著、武舎他訳：HTML入門，プリンティスホール，1995.