

HyTime 文書解析エンジンの試作

3 A C - 2

青山 ゆき[†] 高橋 亨[†] 星 幸雄[‡]
 (株)日立製作所 [†]情報・通信開発本部 [‡]ソフトウェア開発本部

1. はじめに

HyTime は、ハイパーリンクや時変データを持つ文書を SGML(ISO 8879)で記述するために、その DTD(文書型宣言)の構成方法と意味を規定した国際規格(ISO/IEC 10744)である[1][2]。

HyTime は、SGML が表現する文書の論理構造に加えて、ハイパーメディアの本質的な構造情報として、文書内および文書間の相互結合を表すハイパーリンク、並びにマルチメディア情報の時間的及び空間的な結合を表現する手段を提供する。

本発表では、この HyTime で記述された文書を規格に沿って解釈し、リンク情報を抽出するエンジンの試作結果と、HyTime エンジンの応用の検討について報告する。

2. HyTime の概要

2.1 Architectural Forms

HyTime は、DTD や文書構造を定めるのではなく、DTD を作成する形式を規定するものである。この形式のことを Architectural Forms(AF)と呼ぶ。

例えば、文書中のある要素に、HyTime のハイパーリンクを表す構造である“clink”という意味を持たせるには、次のように DTD を構成する。

[HyTime DTD の例]

```
<!ELEMENT myclink - - (#PCDATA)>
<!ATTLIST myclink
    HyTime NAME #FIXED "clink"
    linkend IDREF #REQUIRED >
```

この DTD は、HyTime の“clink AF”に従っているため、インスタンス中の要素“myclink”は、ハイパーリンクを表すことになる。

2.2 ハイパーリンクの記述方法

HyTime では、テキストやマルチメディアデータ等の様々な種類のデータオブジェクトやその一部、あるいはそれらの中の位置を locate し、ID 付けを行うための location AF を提供する。ハイパーリンクやマルチメディア情報の結合は ID を指定して表わされるため、ハイパーメディアシステムを構成する際には、ID 付けの柔軟さが重要となる。

locator には、例えば、次のようなものがある。

- (1) **nameloc** … 構造やデータ実体を表す Entity 名等に対して ID を付与する。
- (2) **dataloc** … n 文字目から m 文字目といった形式で、データ中の一部に対して ID を付与する。画像中の一部なども指定できる。
- (3) **treeloc** … SGML の tree 構造の n 番目の子要素といった形式で、構造に ID を付与する。

これらは、一般に location ladder(梯子)として使用される。例えば、文書を locate(nameloc) し、その中の構造を locate(treeloc) し、構造中のデータ位置を locate(dataloc)するように locator を組み合わせて使用することにより、文書の外部からでも任意の箇所を locate し、そこに ID を付与することができる。

また、HyTime の規定するハイパーリンクには、次のようなものがある。

- (1) **clink(contextual link)** … 文書中に直接リンク情報を記述する。
- (2) **ilink(independent link)** … リンクを張りたい箇所以外の場所(別文書も可)に、そのリンク情報を記述する。

ilink を用いてハイパーリンクを記述する例を図 1 に示す。locatorA によりテキスト中の一部に

ID_A を付け、locatorB により画像中の一部に ID_B を付けて、これらの ID を linkends 属性として記述することで外付けのリンクが実現される。

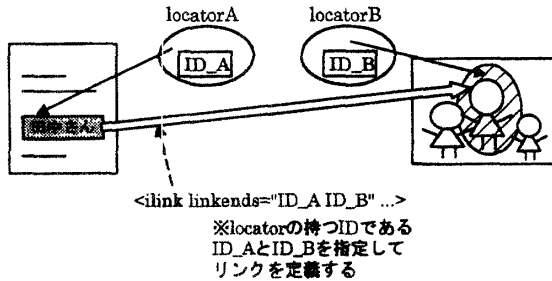


図1 locatorによるリンク定義

3. HyTime エンジンの構成

試作した HyTime エンジンの構成を図 2 に示す。入力した HyTime 文書を SGML パーサによって解析し、次にその結果を利用して HyTime の規格に従って解釈を行う。HyTime エンジンは、解釈した HyTime 情報をテーブル形式で内部に保持する。HyTime エンジンを利用する AP は、エンジンに問合せを行って、リンク情報等を取得するよう構成した。

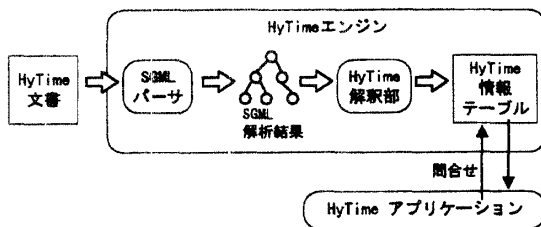


図2 HyTime エンジンの構成

上記、HyTime 解釈時の処理手順を示す。

[Step1] HyTime 適合チェック

入力した文書が HyTime の規格に従うかをチェックする。HyTime 使用宣言¹に記述された情報と不整合はないか、各 AF に従って正しく DTD が構成されているかをチェックする。

[Step2] HyTime 情報の抽出

入力文書のインスタンス中の HyTime 要素を抽出し、HyTime 要素テーブルに登録する。

[Step3] location 情報のトレース

HyTime 要素テーブル中、locator 要素をトレースして、ID と絶対位置（文書名+文書中の位置）の組みに変換し、ID テーブルに格納する。

[Step4] リンク情報の抽出

HyTime 要素テーブル中、ハイパーリンクである要素を解釈し、リンク元の ID とリンク先の ID の組みをリンク情報テーブルに格納する。

HyTime AP は、エンジンに問合せを行い、各テーブルに登録された情報を取得する。今回は、リンク情報を取得して HTML に変換するサンプル AP を作成し、エンジンの動作確認を行った。

4. 応用の検討

HyTime の応用は、次の二通りに分けられる。

- (1) HyTime 文書の流通を前提とするもの … HyTime ビューア、エディタ等
- (2) システムの内部形式に HyTime を利用するもの … リンク情報を扱う文書管理システムや AP 等

流通している HyTime 文書は、まだ少なく、(1) の応用の普及は先になると思われる。(2) の応用は、HyTime の柔軟なリンク機構を利用でき、また、リンク情報が構造化されているため変換が容易であり、将来そのまま文書配布することも可能になると思われるため、有効な応用方法と考える。

5. まとめ

HyTime を、ハイパーメディア情報の統一された交換の手段として利用するために、HyTime エンジンの試作を行った。今後、本エンジンを組み込んだ文書管理システムを開発する予定である。

参考文献

- [1] ISO/IEC 10744 : “Information technology - Hypermedia/Time-based Structuring Language(HyTime)”, ISO, (1992)
- [2] Steven J. DeRose, David G. Durand : “Making Hypermedia Work”, KLUWER ACADEMIC, p385(1994)

¹ HyTime のどのバージョンやモジュールを使用するかを文書ごとに宣言したもの