卒業論文

シストリックアルゴリズムライブラリの構築に関する研究

東北大学工学部阿曽研究室４年 山口貴士

目 次

第1章 論文 **1**

1.1 背景 1

1.2 目的 3

第2章 記述形式について **4**

2.1 SGML とは 4

2.1.1 SGML の当初の目的 4

2.1.2 SGML の新しい利用法 5

2.2 SGML の構成 5

2.3 シストリックアルゴリズム用ＤＴＤ 6

2.3.1 要素「systolic」 8

2.3.2 要素「論文」 8

2.3.3 要素「問題」 9

2.3.4 要素「セル」 9

2.3.5 要素「結線」 10

第3章 の検索 **15**

3.1 検索プログラムについて 15

3.2 操作方法 15

3.3 テキスト viewer 19

第4章 結論 **21**

参考文献 **23**

図 目 次

1.1 内積ステップ・プロセッサ 2
1.2 行列 E=B × C + D を計算するシストリック回路 3

2.1 SGML の構成 5
2.2 シストリックアルゴリズム用のＤＴＤ 7
2.3 要素「systolic」 8
2.4 要素「論文」 8
2.5 要素「問題」 9
2.6 要素「セル」 9
2.7 要素「結線」 10

3.1 検索プログラムの操作フロー 18
3.2 SGML ビューア 19

第1章

序論

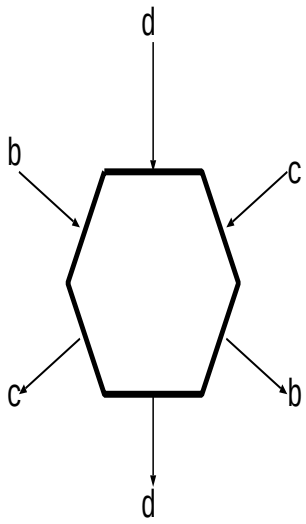
1.1 背景

最近のコンピュータの処理速度は非常に速い。10 年前までは大学や企業の研究室で使っていたコンピュータと今のパソコンの速度が同じになる程に小型化・集積化が進んでいる。これまでの高速化に主に寄与してきたのはデバイス技術であったが、近年では複数の演算器を同時に動かすことで高速化をはかる並列処理も注目されている。その並列処理方式の一つとして、シストリックアレーがある。
シストリックアレーとは 1978 年に Carnegie-Mellon 大学の H.T.Kung によって提案された並列処理アーキテクチャである。

その特徴としては、

- 少数のセルにより構成
- セル間の通信はローカルで規則的
- パイプライン処理方式を採用

などの点が挙げられる。このようなセルの単純さ、データの流れの一様性、単純な接続関係等の利点は VLSI 化に適している。これを用いてたたみ込み、行列演算など、全体の計算時間が入出力時間よりも処理のための計算量に依存する問題を高速に計算するための方法が多数提案されている。
シストリックアルゴリズムの例として、行列の積を求めるものを次に示す。まずセルを図 1.1 に示す。



b(t+1)=b(t)
c(t+1)=c(t)
d(t+1)=b(t)xc(t)+d(t)

図 1.1: 内積ステップ・プロセッサ

このセルを空間上に配置して 2 × 2 行列の積を求めるアレイを作ると図 1.2 のようになる。

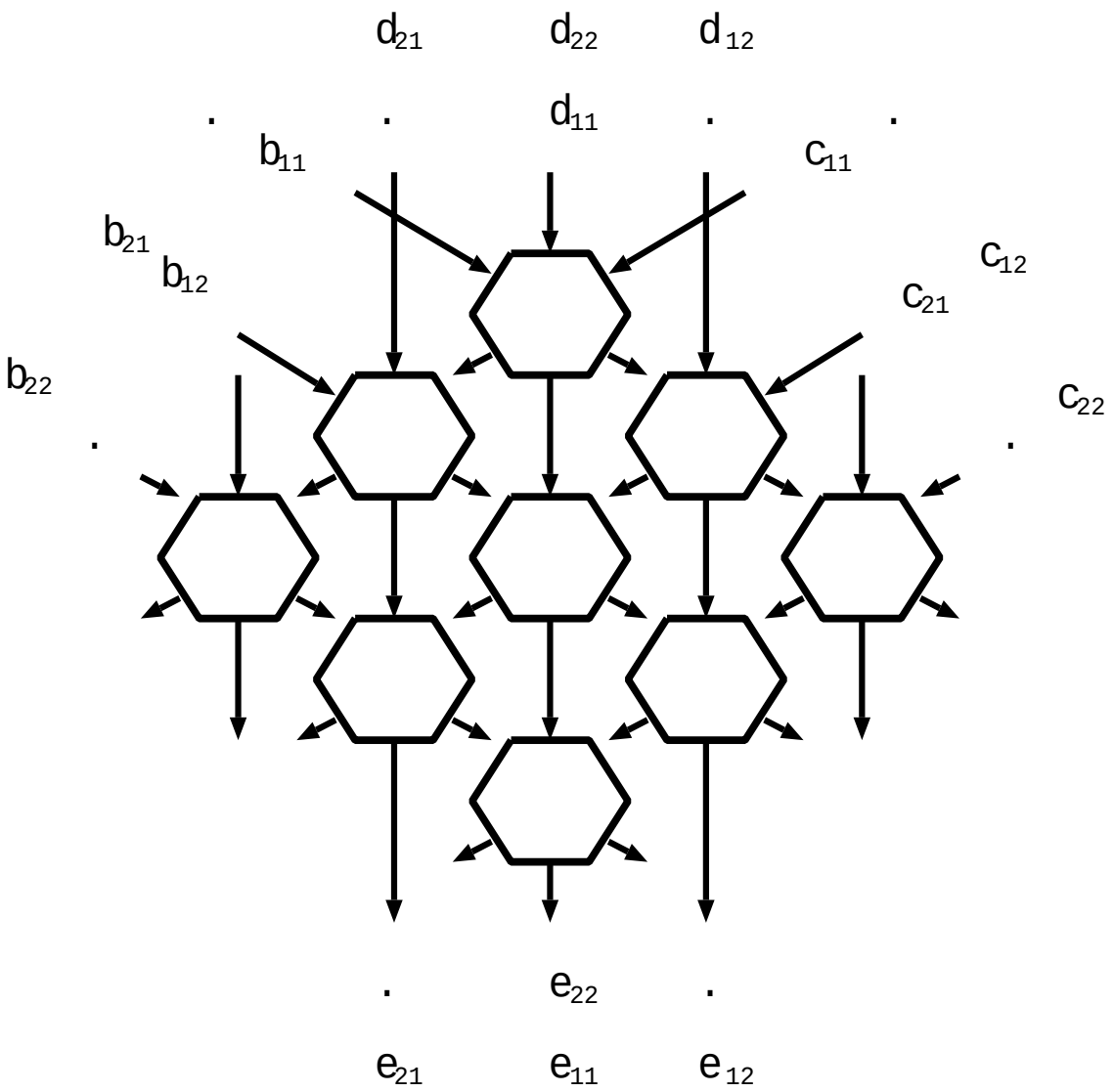


図 1.2: 行列 E=B × C + D を計算するシストリック回路

1.2 目的

シストリックアルゴリズムは直観的に分かり易い方法で並列処理を実現できる反面、ある特定の問題専用のアーキテクチャを個々に設計する必要があるというデメリットもある。設計方法を覚えてもいろいろな問題用のアレーを必要に応じて設計することは容易ではないため、既存の設計されているアレーを見て学習することになる。そのとき作りたいアレーについての資料が必要となる。そこで本研究では資料を探す労力を減らすためにシストリックアルゴリズムのライブラリの設計について考える。

第2章

データ記述形式について

前述したシストリックアルゴリズムの設計方法をライブラリ化するために、まずどのような形式で記述するか、ということが問題となる。必要なテキストを効率良く参照するために、SGML という規格に沿って記述することにした。この章では SGML という規格についての解説とそれを用いたテキストの例を示す。

2.1 SGML とは

2.1.1 SGML の当初の目的

SGML (Standard Generalized Markup Language) は、さまざまなアプリケーションの間で文書の交換をするために提案された国際規格 (ISO8879) である。
現在でもさまざまなワープロが存在し、異なるワープロソフト間にはデータの互換性がない。そのため、あるワープロソフトから別のワープロソフトへの変換ソフトなどが必要となる。しかしそれぞれのワープロでは機能も異なり、レイアウト情報などの定義付けの方法もさまざまであるから 100%変換するということは事実上不可能である。そこで SGML はソフト固有の情報を必要とするレイアウト情報を取り除き、文字をすべてテキスト形式 (制御コードのない基本的な文字コードによって構成されるもの) で表現することによってそれらの問題を解決しようとしている。テキスト形式とはいえ、「見出し」や「段落」のような文書の識別情報は「タグ」と呼ばれる特別な形のテキストを入れることによって伝達できるようになっている。
SGML では「文書の意味を伝える」ということに用件をしぼる割り切りによって、文書交換を可能にした。また「見出し」や「段落」のような文書の識別情報をユーザーが自由に

定義付けできるため、文書データを表現する自由度が高い。

2.1.2 SGML の新しい利用法

SGML は文書データを全て識別情報 (タグ) 付きのテキスト形式で表現し、そのタグをユーザーが自由にカスタマイズできるようになっている。このタグによってマルチメディアの識別情報を容易に作ることができる。また、文書を「木構造」で表現しているため、文書の一部を「部品」として扱うことができる。これらをデータベースで管理すれば文書データの「部品」を共有し再利用することも可能であり、それに応じた取り組みもなされている。

2.2 SGML の構成

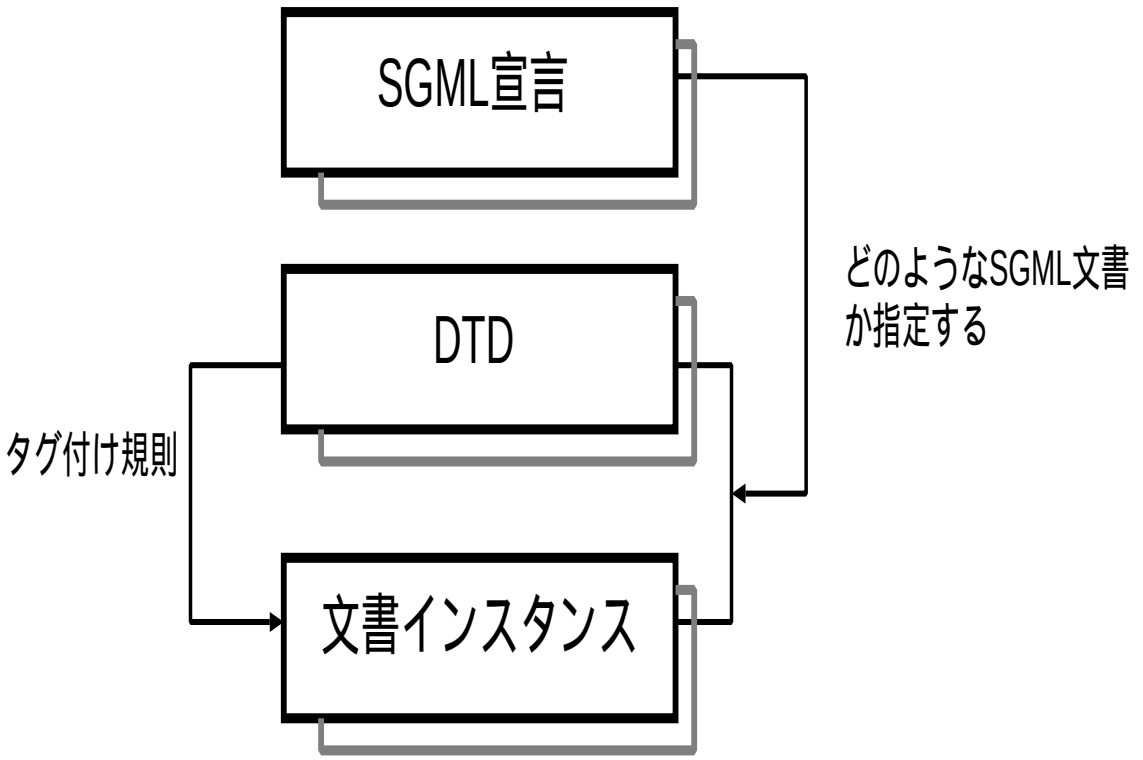


図 2.1: SGML の構成

ＳＧＭＬ文書は次の３種類のテキストによって構成される。

ＳＧＭＬ宣言 使用する文字集合、処理に必要な記憶容量、具体構文の有効範囲、使用されている具体構文、使用されているＳＧＭＬ機能、処理に必要なアプリケーション固有機能などを宣言する。

ＤＴＤ　ＤＴＤは Document Type Definition の略で、ＪＩＳでは文書型定義と訳している。ＤＴＤは文書テキストで使われるタグの構造・定義などを、ＳＧＭＬの構文を使って定義・記述したもの。本、論文、レポートなどの文書の種類によって、その構造が異なるので、種類ごとにＤＴＤも異なる。

文書インスタンス　ＤＴＤに従ってマーク付けしたテキストで、これが本文となる。普通のテキストの中にＤＴＤによって定義されたタグを埋め込んだ形になっている。

SGML 文書を書くときに最も重要なのがＤＴＤの記述である。シストリックアルゴリズムを記述するためには、それに最適なＤＴＤを記述する必要がある。尚、ＳＧＭＬ宣言については本研究と関係ないので触れないこととする。

2.3　シストリックアルゴリズム用ＤＴＤ

図 2.2 はシストリックアルゴリズム用に作成したＤＴＤの木構造を図示したものである。シストリックアルゴリズムの設計法を記述した論文から重要と思われる部分のみを取り出すために、このように文書構造を定義している。

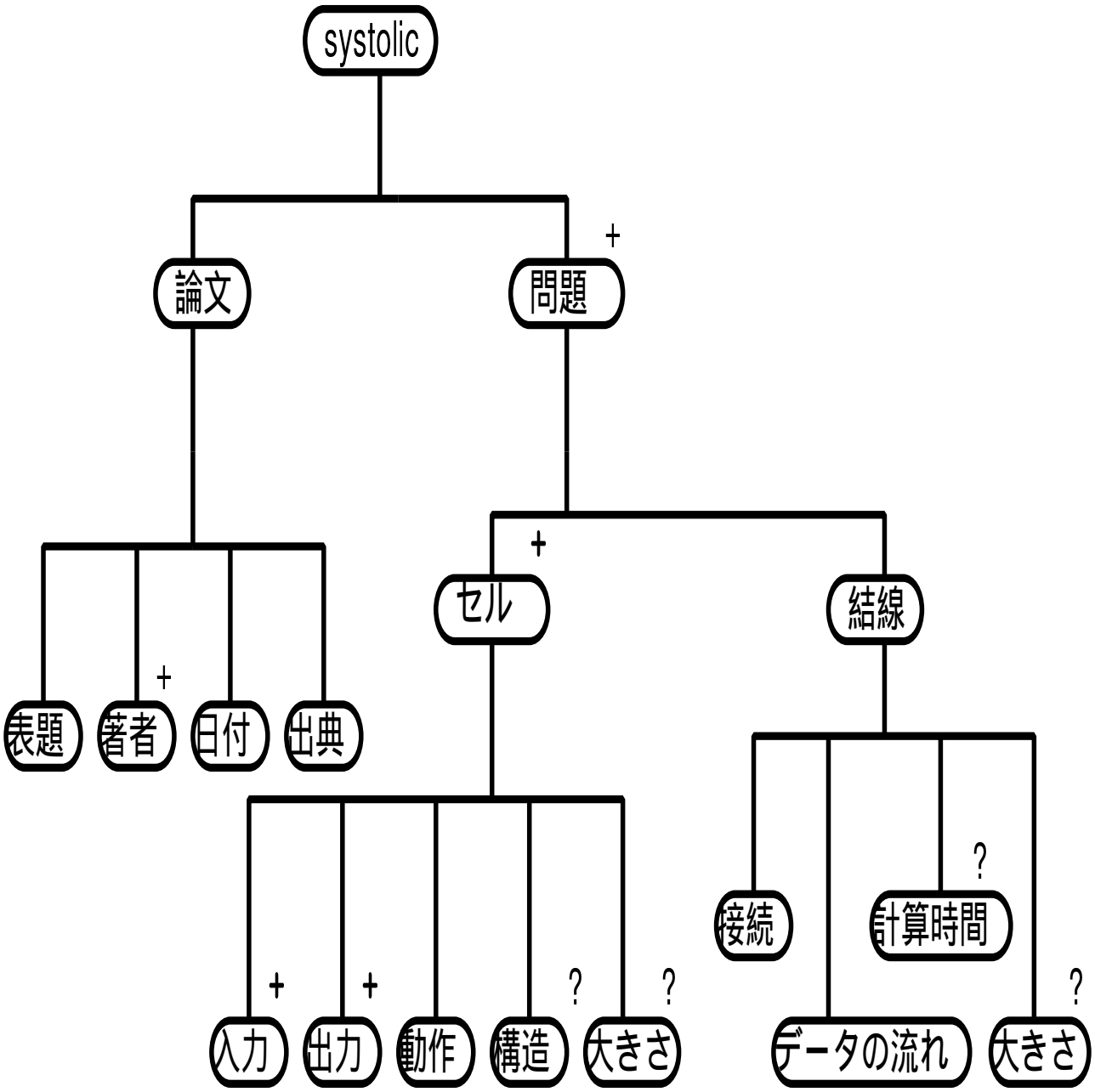


図 2.2: シストリックアルゴリズム用のＤＴＤ

まず、図 2.2 の読み方を解説する。要素間の親子関係を考えると普通子供は複数存在するので、その順序を指定する必要がある。SGML では次の 3 つのパターンで出現順序を定義することができる。

- 順番に出現
- 順不同で出現
- どれか一つが出現

しかしこのＤＴＤでは順不同だけしか使っていないので、特に記号はない。また SGML では要素の出現回数について、次の 4 つのパターンを許している。

- 必ず 1 回出現
- 0 回もしくは 1 回出現
- 1 回以上出現
- 0 回以上出現

各パターンを表すマークの種類は順に“　”、“？”、“＋”、“＊”で、「必ず 1 回出現する」場合はマークを付けない。図 2.2 では要素の右上にマークを付けている。
以下では各要素について説明する。

2.3.1　要素「systolic」

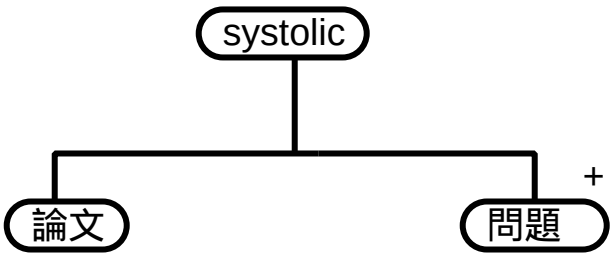


図 2.3: 要素「systolic」

このＤＴＤ文書は大きく 2 つの部分に分かれる。シストリックアレーの設計に直接関係する部分とそうでない部分である。前者はアレーが解くべき「問題」で、後者はアレーを発表している「論文」となる。

2.3.2　要素「論文」

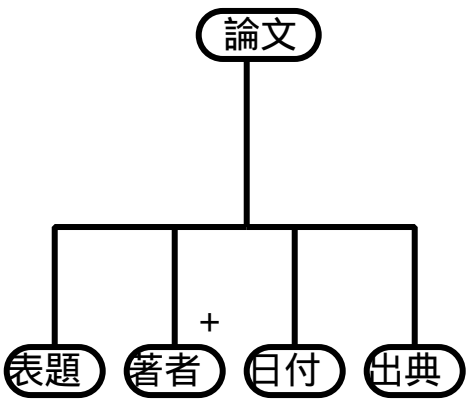


図 2.4: 要素「論文」

始めの部分は、論文のタイトル、著者、発表年月日などの情報が記述してある。ここで「著者」は複数である場合があるので「1 人以上」という意味の “+” が付いている。

2.3.3 要素「問題」

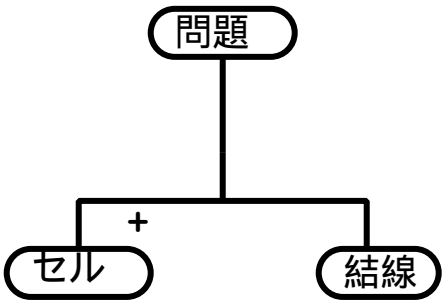


図 2.5: 要素「問題」

もうひとつの部分は、シストリックアレーの設計に関する部分である。これがセルに関する部分とその結線に関する部分からなる。1 章で述べたようにセルは少数であって複数のセルが使われる場合もある。よって “+” マークがセルに付く。

2.3.4 要素「セル」

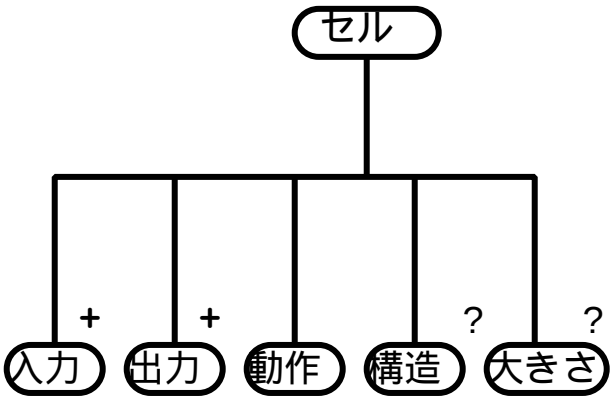


図 2.6: 要素「セル」

セルの部分はその入力と出力、そこで行われる処理などの情報が記述される。入出力は 1 章の例のように複数あるのが自然なので “+” が付く。また「構造」「大きさ」という要素は論文中に示されていないことがあるので「0 回か 1 回」を意味する “？” が付いている。

2.3.5　要素「結線」

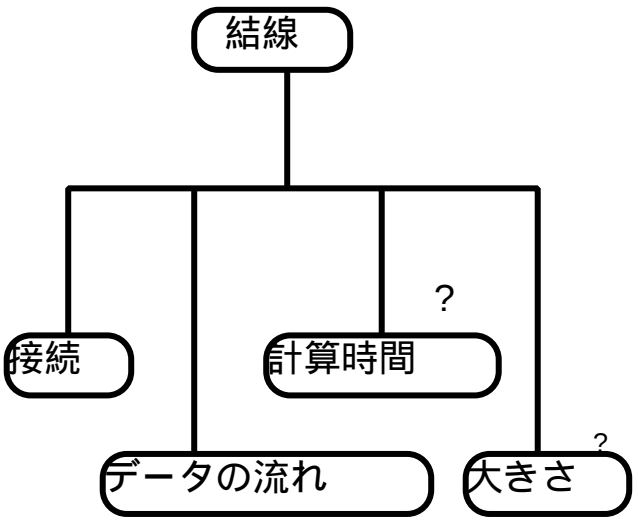


図 2.7: 要素「結線」

結線に関する部分は、セルの接続方法やアレイに対するデータの入出力などの流れと、計算時間、大きさから成る。ここでも「計算時間」と「大きさ」は“?”が付いている。

このように定義された文書構造に従い、書き起こしたテキスト (文書インスタンス) のサンプルを次に示す。

<!DOCTYPE systolic SYSTEM "wani1/usrs/taka/B-ron/systolic.dtd">

<systolic>

<論文>

<表題 id="JO24-4">VLSI 用の行列計算法

<著者>都倉信樹

<著者>萩原兼一

<日付 name="1983">1983 年

<出典>情報処理

</論文>

<問題 name="matrix">行列計算 (乗算 TH 法)

<セル name="naiseki">

<入力><p>b(t)

<入力><p>c(t)

<入力><p>d(t)

<出力><p>b(t+1)

<出力><p>c(t+1)

<出力><p>d(t+1)

<動作>

<p>b(t+1)=b(t)

<p>c(t+1)=c(t)

<p>d(t+1)=d(t)+b(t)*c(t)

</動作>

</セル>

<結線 name="rokkaku">

<大きさ>

<p>n*n 行列に対して (2n-1)*(2n-1) 個のセルが必要

<接続>

<p>六角網状 (ひし形)

<データの流れ>

<p>バンド行列をパイプライン的に入力する

<計算時間>

<p>n*n 行列に対して 3n-1 単位時間

<p>また, パイプライン時間 (1 つの行列積を計算するための入力開始時刻と次のそれとの時間差) は,n 単位時間である

</結線>

</問題>

</$systolic$>

実際のＤＴＤは次のようになっている。詳しいＤＴＤの読み方は参考文献 [6,] を参照して欲しい。

```
<!DOCTYPE systolic [
<!ENTITY % doctype "systolic" >
<!-- ENTITYS -------->
<!ENTITY % para "#PCDATA">
<!ENTITY % common "id ID #IMPLIED">
<!ENTITY % cdata "name CDATA #IMPLIED">
<!----- エレメント -------->
<!--見出し語 宣言要素 タグの有無 下位要素（順序と回数）-------->
<!ELEMENT %doctype;  - -（論文, 問題+）>
<!ELEMENT 論文 - -（表題&著者+&日付&出典）>
<!ELEMENT 表題 - O（%para;）>
<!ELEMENT 著者 - O（%para;）>
<!ELEMENT 日付    - O  （%para;）>
<!ELEMENT 出典    - O（%para;）>
<!ELEMENT 問題   - -（セル+, 結線）>
<!ELEMENT セル    - -（入力+&出力+&動作&構造?&大きさ?）>
<!ELEMENT 入力 - O（p+）>
<!ELEMENT 出力 - O（p+）>
<!ELEMENT 結線 - -（接続&データの流れ&計算時間?&大きさ?）>
<!ELEMENT 動作    - -（list*）>
<!ELEMENT 構造   - -（p+）>
<!ELEMENT 大きさ - O（p+）>
<!ELEMENT データの流れ  - O（p+）>
<!ELEMENT list    - O（p+）>
<!ELEMENT 計算時間 - O（p+）>
<!ELEMENT 接続 - O（p+）>
<!ELEMENT p - O（#PCDATA）>
<!----- 属性リスト -------->
<!ATTLIST 表題 %common;>
<!ATTLIST 結線 %cdata;>
```

```
<!ATTLIST 日付 %cdata;>
<!ATTLIST 問題 %cdata;>
<!ATTLIST 構造 %cdata;>
<!ATTLIST セル %cdata;>
<!ATTLIST 動作 steps NUMBER #IMPLIED>
<!ATTLIST 参照 refid IDREF #CONREF >
]>
```

第3章

データの検索

2章で述べたようにSGMLは標準規格なので、SGML形式で書かれたテキストの変換や表示、印刷などをするためのソフトがいろいろと開発されている。その中でも手に入り易いものとしてフリーウェアを選び、検索などができるソフトを探したが、あらかじめ決まったＤＴＤに対して検索や変換を行うもののみであった。日本語を扱えてＤＴＤの解析をして検索や変換をするようなものは見つからなかった。よって本研究の目的に合った検索用のプログラムを開発した。

3.1　検索プログラムについて

おおまかな流れは、実行ファイルがデータファイルを読み込む、次にユーザーが「問題」要素の中から解きたい問題を選ぶ、その問題を解くアレーを設計するために必要なテキストを表示する、という単純なものになっている。

3.2　操作方法

実行ファイルにデータファイルの名前を付けて次のように実行する。

sless systolic.sgm

すると、次のように表示される。

0:最短路問題　1:コンボリューション　2:多項式除算

3:ＬＵ分解　4:ＬＵ分解　5:行列計算 (乗算 TH 法)

番号を入力し、return key を押して下さい

ここで 0–5 の中からシストリックアレーを使って解きたい問題を選び (問題名が同じでも番号の違うものは別なアレーについて書かれている) 、その番号を入力すると 2 章で示したようなテキストから要素「問題」以下の要素とテキストを加工して表示する。ここではUNIX のコマンドの less を使っているので、less の操作でテキストを読んだり検索したりすることができる。「Ｑ」を押すと less が終了し次へ進む。次はそのアレーを提案している論文について表示される。ここも同様に less を使っている。この less が終了すると、プログラムも終了する。
次に実際に表示されるものの例を示す。

セル name="cel"-
入力-Ai
入力-Bi
入力-Ci
出力-Ai
出力-Bi
出力-Ci
構造-
５個のレジスタを持つ
そのうち４つのレジスタ (R1-R4) はそれぞれサブレジスタ (Ai,Bi) に分割される
残りのレジスタ Rc とする
各レジスタの機能は次のとうり
R1,R2:データ保持
R3,R4:右隣接セルへのデータ伝達のためのパイプライン
Rc:コンボリューションの部分積和の記憶レジスタ
動作 steps="6"-

セル i がデータを含まない場合
セル i は Rc に対してなにもしない

セル i が 1 対のデータを含む場合

```
R(t,c,i)= A(t,1,i)* B(t,1,i)
```

セル i が 2 対のデータを含む場合

```
R(t,c,i)= A(t,1,i)* B(t,2,i)+ A(t,2,i)* B(t,1,i)
```

⋮

まず表示されるのはこのようなテキストで、less によって表示しているので読みたい部分をそのまま検索することもできる。この less を終了させると (“Q” を押すと)、そのアレーを発表している論文の情報が less で表示される。

<表題 id="AL85-13">最適実時間シストリック多項式除算アルゴリズム
<著者>梅尾　博司
<著者>杉岡　俊幸
<日付 name="1985">1985 年
<出典>電子通信学会技術報告　オートマトンと言語
</論文>

これを終了するとプログラムも終了するので、もう 1 度検索するときはまた

sless systolic.sgm

と入力し直し、再度実行する。

最後に操作の流れ図を示す。

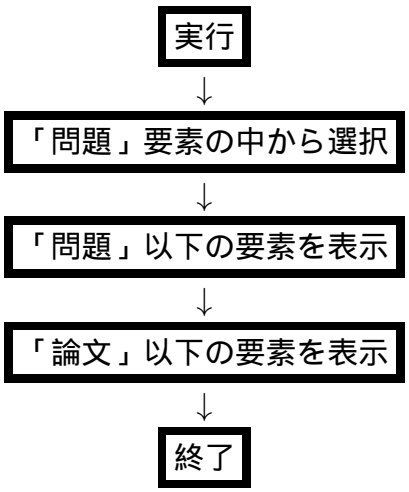


図 3.1: 検索プログラムの操作フロー

3.3 テキスト viewer

ある形式に整形されたテキストファイルを見るためのフリーのビューアを使っている。2章で示したシストリックアルゴリズムのテキストを変換してそのソフトで扱える形に整形した。このビューアは Tcl/tk で書かれている。

実行方法は

$jfx2.tk$

である。すると次の図のようなウィンドウが開くので、そのウィンドウ上のボタンをクリックすればSGML から変換されたテキストを見ることが出来る。

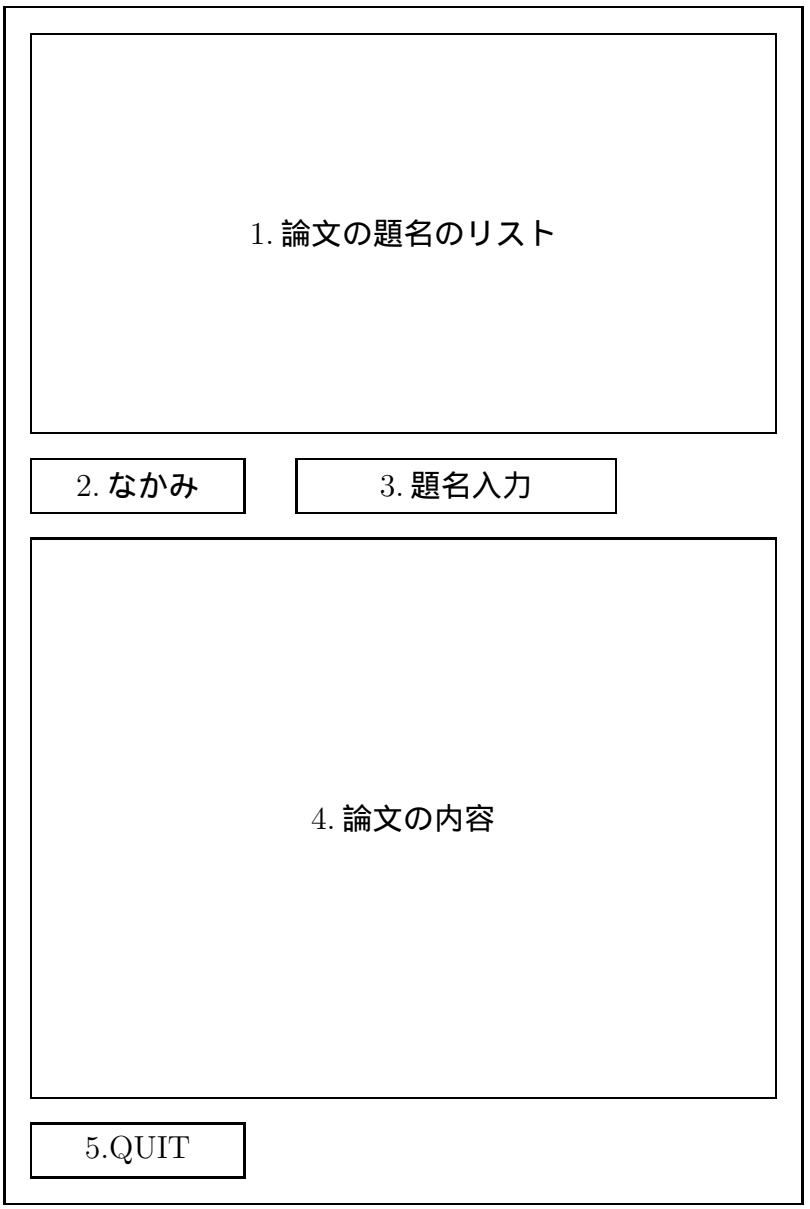


図 3.2: SGML ビューア

図 3.2 の解説

1 · · · 図の通り。

2 · · · ボタンになっていて、クリックすると 3 のところに入力した題名の内容が 4 に表示される。

3 · · · 題名を入力するための部分で、キーボードから入力することもできるが、1 の題名をクリックするとそれがここに表示される。

5 · · · ボタンになっていて、クリックするとウィンドウが閉じる。

第4章

結論

本論文では、シストリックアレーに関するライブラリを構築するためのデータ記述の形式としてSGMLを採用し、それによる記述方法を提案した。また単純で制約が多いものだがSGML形式で記述されたテキストの検索用プログラムを書き、実際に検索できるようにした。

今後の課題は、ＤＴＤの拡張、検索用のソフトの検討または改良、SGML以外のものでライブラリを構築した場合との比較・検討である。

謝辞

本研究を進めるにあたり、全般的な御指導、御助言を戴いた東北大学工学部阿曽弘具教授に深く感謝致します。また、日頃から本研究の御指導、御意見を賜わった大町真一郎氏、黒岩丈介氏、森大毅氏、並列グループの菅谷至寛氏に深く感謝致します。最後に、日頃よりお世話になった阿曽研究室の皆様に心より感謝致します。

参考文献

[1] 村松 晃　宮岡 伸一郎　船橋 もとひさ:　「Systolic Array によるネットワーク最短路問題の並列解法」, 電気学会論文誌 C (1983)

[2] 梅尾 博司 :　「パラレル・コンボルバについて」, 電子通信学会技術報告 オートマトンと言語 (1984)

[3] 梅尾 博司　杉岡 俊幸 :　「最適実時間シストリック多項式除算アルゴリズム」, 電子通信学会技術報告 オートマトンと言語 (1985)

[4] 堀池 聡　西田 正吾　坂口 敏明 :　「疎行列ＬＵ分解のシストリックアレイ」, 信学技報 CAS(1984)

[5] 都倉 信樹　萩原 兼一 :　「VLSI 用の行列計算法」, 情報処理 (1983)

[6] (株) 日本ユニテックＳＧＭＬサロン (佐藤 譲　田中 稔　梅野 謙一　奥井 康弘): 「はじめてのＳＧＭＬ」, 技術評論社