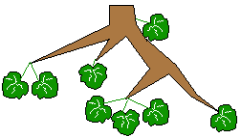


木構造

総合科目「IT革命を解き明かす」、2004年10月04日

電子・情報工学:
新城 靖
<yas@is.tsukuba.ac.jp>

このページは、次の URL にあります。
<http://www.hlla.is.tsukuba.ac.jp/~yas/gen/it-2004-10-04/>
あるいは、次のページから手繰っていくこともできます。
<http://www.hlla.is.tsukuba.ac.jp/~yas/gen/>
<http://www.hlla.is.tsukuba.ac.jp/~yas/index-j.html>
印刷配布資料
<http://www.hlla.is.tsukuba.ac.jp/~yas/gen/it-2004-10-04/it-2004-10-04.pdf>



■今日の重要な話

- 情報を整理する時に、もっとも基本的で重要な方法として、木構造を使う方法がある。
- 木構造には、様々な表現方法がある。

木構造の例を、大学の組織を使って説明する(図1)。

- 木構造は、節と枝からできている。
- 特別な節として、**根(root)**が1個ある。
- 末端の節を**葉(leaf)**と言う。
- 木構造では、ある節から別の節までの道が1通りしかない。
- つながっている節と節で、根に近いものを**親**、そうでないものを**子**という。

木構造という名前は、本物の木が、一度枝分かれした後は決して交わらないことに似ていることによる。ある節から別の節までの道が2通り以上あるは、グラフ構造と呼ばれる。

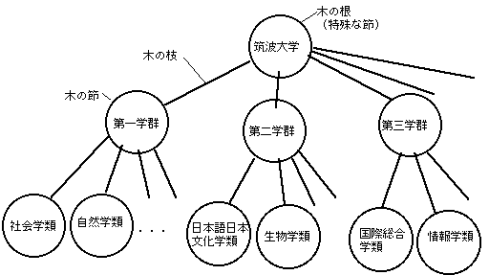


図1 大学組織に見られる木構造

図1では、筑波大学と書いてある所が木の根にあたる。根からは、何本かの学群の枝が出ている。このように、コン

- 節と枝。根も葉もある。根を上を書く。
- 領域的(集合のベン図)
- 段付け (字下げ)
- (区切り文字を挟んで)単語を並べる
- 木構造は、コンピュータであつかうものの名前を管理するために使われている。
- 木構造は、英語の読解や作文に役立てることができる。
- 木構造を使うと要素の数が増えても対応することができる。
- 木構造だけでは、情報の整理はうまくいかない。木構造を補う仕組みが必要になる。
- World Wide Web では、ハイパーテキスト(グラフ構造)が使われている。
- URL (Uniform ResourceLocator) には2つの木の表現が含まれている。

■木構造

木構造 (tree structure) というのは、コンピュータ・サイエンス (情報学類で学ぶ学問) でよく使われる用語である。分野によっては、同じものを**階層構造 (hierarchical structure)** という言葉で表現することが好まれる。ドイツ語語源の、ヒエラルヒー(Hierarchie)という言葉が使われることもある。

ピュータ・サイエンスでは、木の根を上を書く習慣がある。第三学群の節には、情報学類、社会工学類などの子の節がある。第三学群の親は、筑波大学である。

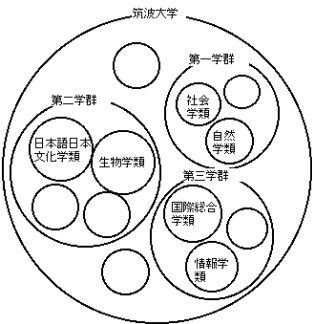
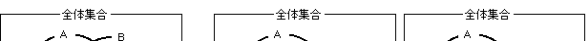


図2 大学組織に見られる木構造 (領域的な見方)

A、B、2つの集合があると、一般には、次の4つに分割される。

1. Aにしか含まれない部分
2. AとBの両方に含まれる部分
3. Bにしか含まれない部分
4. AにもBにも含まれる部分

領域 (木構造) の場合、完全に含まれるか、無関係かのどちらかになる。



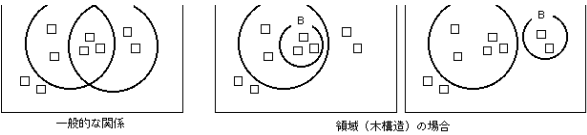


図3 2つの集合の関係

◆字下げによる木の表現

木構造を字下げで表すことがある。

筑波大学

- 第一学群
 - 自然科学類
 - 社会学類
- 第二学群
 - 生物学類
- 第三学群
 - 情報学類
 - 情報科学主専攻
 - 計算機システム主専攻
 - 知能情報メディア主専攻
 - 社会工学類

◆区切り文字入り表記

「情報学類」という節を、次のように表記する。

筑波大学第三学群情報学類

コンピュータの中で、文字列（文字の並び）で木構造上の位置を表現する時には、節が分かりやすくために、はっきりと区切りを入れて表現することがよく行われる。

筑波大学．第三学群．情報学類

筑波大学／第三学群／情報学類

情報学類．第三学群．筑波大学

区切り文字としては、「.」（点）、「/」（スラッシュ）、「\」（バックスラッシュ）、「¥」（円記号）などがよく使われる。単語を並べる時に、木の根に近いほうから書く流儀と遠い方から書く流儀がある。

◆木の例

コンピュータでは、次のような場所で木構造が使われている。

- ファイルの名前
- インターネット上のコンピュータの名前
- 電子メールのアドレス
- World Wide Web のアドレス (URL)
- ネットワーク・ニュースのニュース・グループの名前
- さまざまなメニュー

コンピュータ以外では、次のような場所で木構造が使われている。

- 分類（生物、図書館の本）
- 住所
- 会社組織、官僚機構
- 1つの文
- 1つの文書

◆文の構造

1つの文も、木構造で表される。同じ単語の並びでも、組み立てられる木構造が違っていると違う意味になる。英文の解釈は、木構造を組み立てることである。

Time flies like an arrow.

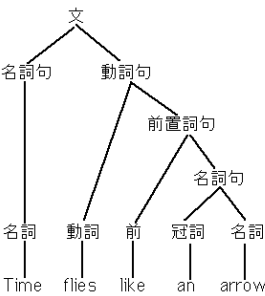


図4 「Time flies like an arrow.」の木(その1)

光陰矢のごとし。

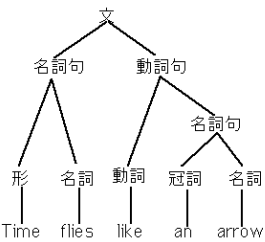


図5 「Time flies like an arrow.」の木(その2)

時蠅、矢を好む。

◆英語の文法

基本 5 文型は、木構造を意味する。

- 1. S V
- 2. S V C
- 3. S V O
- 4. S V O O
- 5. S V O C

これを <clause>とする。英語の文は、次の形式で <clause>をいくつかつなげ、最後に「。」をつけたたものである。

```
<sentence> ::= <simple sentence> '।'
              または <compound sentence> '।'
              または <compound-complex sentence> '।'

<simple sentence> ::= <clause>

<compound sentence> ::=
  <clause> ',' <FANBOYS> <clause>
  または <clause> ';' <transition word> <clause>
  または <clause> ',' <conjunctive> <clause>

<complex sentence> ::=
  <dependent clause> ',' <independent clause>
  または <independent clause> <dependent clause>

<compound-complex sentence> ::=
  <clause> <dependent clause>

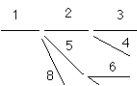
<dependent clause> ::= <subordinate> <clause>
```

- 章立て。章、節、項。
- 箇条書き。箇条書きの中に箇条書き。
- パラグラフ

◆パラグラフの木

英語国民に対する英語の教育では、パラグラフ(段落)の内部も、きちんと木構造で書くことが求められる。

- パラグラフは、1つの topic を語る (0個、2つ以上は不可)。
- 1つのパラグラフには、1つの topic sentence (パラグラフの内容を一言で語った文)を置く。
- topic sentence は、多くの場合、パラグラフの先頭に置かれる。
- パラグラフに含まれる他の文は、次のいずれか (宙に浮いているものは不可) 。
 - topic sentence を展開させたもの (topic sentence を木の根として、その子供の節になる)
 - (他のパラグラフとの関係を示すもの)
- 木構造を明確にするために、signal wordsを使う。
 - First, Second, Third, ...
 - 原因・結果
 - 対句、対比



```
<subordinate> ::=
  <subordinate-adj> または <subordinate-adv>
  または <subordinate-noun>

<subordinate-adj> ::=
  who または whom または that または which

<subordinate-adv> ::= after または before または beca
  または although または since

<subordinate-adv> ::= that または whether

<clause> ::= 基本 5 文型で表せるもの

<FANBOYS> ::=
  for または and または not または but または
  or または yet または so
```

```
<conjunctive> ::=
  <subordinate conjunction> または <coordinate conj
  または <conjunctive adverb>

<subordinate conjunction> ::= as, if, that など
<coordinate conjunction> ::= and, but, or, for など
<conjunctive adverb> ::= however, nevertheless,
  still, then など
```

よい英文は、1つの <sentence> には、2つの<clause>、つまり、S V が2つが現れることが多い。

◆文書の木

文章の中にも、木構造が現れる。

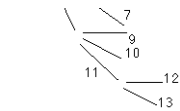


図6 topic sentence を根とする木構造

(よく書かれた) 英語の長い文章を斜めに読むには、パラグラフの先頭の topic sentence だけを読めばよい。

◆日本語の段落の構造:逆茂木型

逆茂木: よろい、かぶとの時代に、敵の侵入を防ぐために樹を斬り倒して尖った枝を外に向けて並べた障害物。

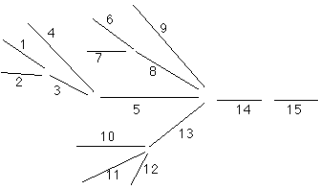


図7 逆茂木型の段落

日本語の特徴

- 漢字仮名混じり文。重要な言葉が漢字で書かれている。
- 文の構造。動詞が最後。

(日本語ではなくて)「国語」の教師は、木構造(生成文法、文脈自由文法、キョムスキー(学者の名前))を知らないこともある。

日本語でも、文学作品をのぞいて、英語的に木構造になっていると読みやすい。

◆is-a関係

This is-a pen の意味。

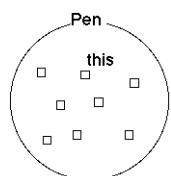


図7 「This is-a pen」の集合的な意味

is-a関係

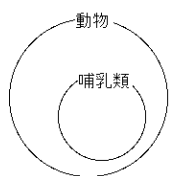


図8 is-a関係の例(哺乳類 is-a 動物)

◆ディレクトリの木

ディレクトリは、全体では木構造(tree structure)になっている。階層化ディレクトリ(hierarchical directory)と呼ばれることもある。

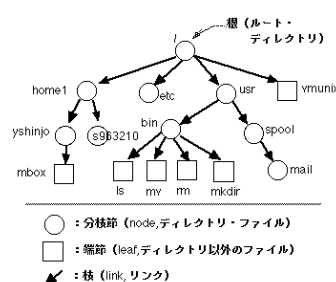


図9 ファイルとディレクトリの木

ファイルは、葉(leaf)になる。ディレクトリは、節(node)になる。特殊な節として、根(root)がある。これを、ルート・ディレクトリ(root directory)という。

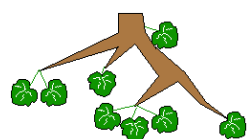


図10 自然の木

◆パス名

ファイルの名前の表現には、「パス名」がよくつかわれる。パス(path)というのは、道の意味である。パス名では、どの道を通ればよいかの道順を示すことでファイルの名前を表現する。

木構造では、節、または、枝(道)に名前がついている。ファイル名は、区切り文字で区切られた、節、または、枝の名前の並びになる。ファイルの名前を表現する時の区切り文字としては、次のものがよく使われる。

- /
- 𐄧
- :

ファイル名で「.」は、木構造の区切りとしては使われない。

パス名の例:

- /home6/s991001
- /usr/local/bin/mnews

◆絶対パス名と相対パス名

パス名には、次の2種類がある。

絶対パス名 (absolute path name)

絶対パス名というのは、ディレクトリの木の根(ルート・ディレクトリ)から出発する道順

相対パス名 (relative path name) 現在着目しているディレクトリ(カレント・ワーキング・ディレクトリ)から出発する道順

例: 絶対パス名 /usr/bin/awk

1. ルート・ディレクトリから出発する
2. usrという枝へ進む
3. binという枝に進む
4. 最後に(awk)という枝に進む。

よって、/usr/bin/awk は、こういう手順で見つかるファイルを意味する。

例: 相対パス名 bin/awk (カレント・ワーキング・ディレクトリが /usrの時)

1. カレント・ワーキング・ディレクトリから出発する
2. binという枝に進む
3. 最後に(awk)という枝に進む。

◆ホーム・ディレクトリ

複数の人が使うコンピュータで、個人のファイルを保存する時の出発点となるディレクトリを、ホーム・ディレクトリ

リと呼ぶ。

たとえば、icho という名前のコンピュータで、新城のホーム・ディレクトリは、絶対パス名では、/home1/yshinjo/ である。

- ルートディレクトリ (/) から出発する。
- home1 という名前の節に進む。
- yshinjo という名前の節に進む。

◆世界最大の木構造:DNS(Domain Name System)

電子メールを送ったりWorld Wide Web のページを閲覧する時には、データの送り先やデータを持っているコンピュータを指定する必要がある。インターネットで使われている、コンピュータの名前を管理する仕組みは、DNS(Domain Name System) と呼ばれている。DNS では、膨大な数のコンピュータの名前を含む名前空間を階層的にドメイン (領域) に分割して管理している。

たとえば、次のような名前を考える。

adonis1.coins.tsukuba.ac.jp

このように、インターネットでのコンピュータの名前は、「.」で区切られた文字列 (文字の並び) である。この文字列で使える文字は、アルファベットと数字、ハイフン

- 最後に、coins ドメインに adonis1 という名前が登録されている。

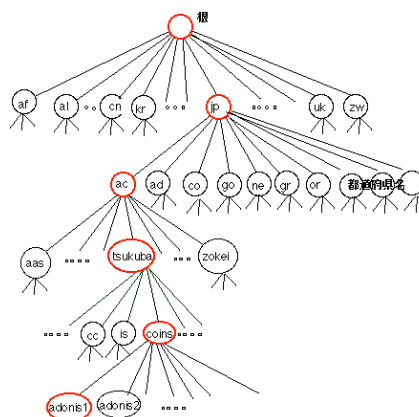


図9 名前空間の木構造としての見方

「adonis1.coins.tsukuba.ac.jp」を 図9 で考えると、次のようになる。

- まず根から出発する。
- 根の下 jp という節を選ぶ。
- その下の ac という節を選ぶ。
- その下の tsukuba という節を選ぶ。
- その下の coins という節を選ぶ。
- 最後に、その下の adonis1 という節(葉)を選ぶ。

(マイナス) である。DNS で名前は、右から左に向かって解釈される。

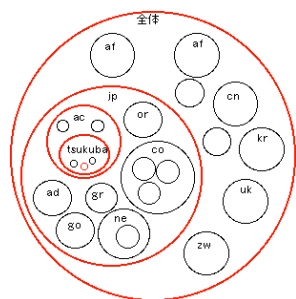


図8 名前空間のドメインへの分割

「adonis1.coins.tsukuba.ac.jp」を 図8 で考えると、次のようになる。

- 名前空間全体は、まず、af, al, ..., jp, ..., uk, zw というドメインに分割されている。その中の jp を選ぶ。
- その中の jp ドメインは、ac, ad, co, ne, gr, or などのドメインに分割されている。その中の ac を選ぶ。
- その中の ac ドメインは、ass, ..., tsukuba, ..., zoeki 等のドメインに分割されている。その中の tsukuba を選ぶ。
- tsukuba ドメインは、さらに cc, is, coins などのドメインに分割されている。その中の coins を選ぶ。

◆DNSができるまで

初期のインターネットは、コンピュータの名前 (ホスト名) は、フラットな名前空間が使われていた。

問題

- 名前の衝突。同じ名前を取り合いになった。
- 中央の管理組織の負荷が増えた。
- 最新のホスト名の一覧表の維持が困難になった。コピーがネットワークを重たくした。

1986年、3100の公式名と6500の別名。

1990年、6400の公式名。DNS に以降。この時点で、137,000。

2004年7月ごろの

名の統計

ドメイン名数

.com 2932万

.net 479万

.org 303万

.biz 100万

.info 121万

.jp 59万

.jp は地域型含む。

第二レベルのドメインの数。ホストの数はもっと多い。

<http://jpinfo.jp/stats/> JPドメイン名に関する統計 by JPRS <http://www.zooknic.com/Domains/counts.html> by Zooknic.

◆ドメイン名の種類

ccTLD (country code Top Level Domain)

ISO (国際標準化機構, International Standardization Organization) が定めた2文字による国別コード (country code) ([ISO 3166](#)) を使う。日本の国別コードは、jp。

gTLD (generic Top Level Domain)

ccTLD 以外。 .com, .edu, .gov, .net, .org, .mil など が現在でも使われている。

jp の下には、次のような枝 (領域、ドメイン) がある。

ac
学校関係(主に大学)、学術研究機関

ad
ネットワーク管理、JPNICの会員

co
会社

ed
児童、生徒などの教育・育成を行う組織。

以前は、IANA (Internet Assigned Numbers Authority) 。

2000年に初めての理事選挙があった。

◆オルタネート・ルート問題

インターネットのドメイン名の根は、1つしかない。13個のサーバにコピーが世界各地にある。

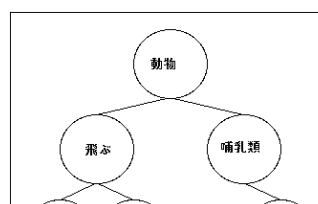
もし、別の根の情報を持つサーバがあれば、どうなるのか。

オルタネート・ルート (alternate roots) 。

■木構造の性質

大量の情報を保存するには、木構造を使うしかない。しかし、、、

◆こうもりの分類問題



go
政府機関、国立の施設

gr
任意団体

ne
インターネット接続サービス・プロバイダ

or
法律に基づく団体

都道府県の名前
地方自治体、個人。

◆汎用JPドメイン名

jp の下に、上のような属性を持たない第2レベルドメイン名 (SLD: Second Level Domain) も使えるようになった。

[汎用JPドメイン名\(JPNIC\)](#)

◆ICANN

ICANN (Internet Corporation for Assigned Names and Numbers)

インターネットにおけるドメイン名や IPアドレスを調整するための非営利法人。1998年に設立。



図13 こうもりの分類 (1)

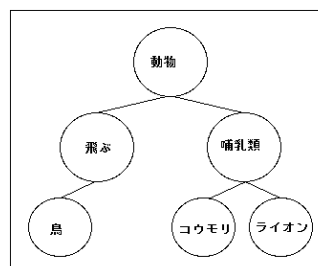


図14 こうもりの分類 (2)

◆木構造の制約を緩和する仕組み

◆シンボリック・リンク/エイリアス/ショートカット

木構造は、ファイルを整理するのに非常に強力な構造である。しかし、それだけでは、ファイルを整理するには不都合が起きる。それを解消するために、次のような名前と呼ばれる仕組みが用意されている。

- エイリアス (alias, Macintosh)
- ショートカット (shortcut, Windows)

- シンボリック・リンク(symbolic link,UNIX)

2つの節に、「別名」をつけて、2つの道からたどり着けるようにする。(木構造では、1つの節にたどり着く道は、ただ1つしかない。)

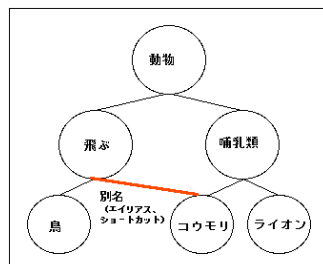


図15 こうもりの分類 (別名つき)

◆官僚制度の2つの見方

- 整理棚。人間を系統的に分類する。事実を系統的に分類する。(保存)
- 情報が流れる経路。(通信)

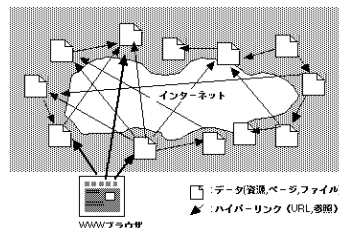
情報の流れには予算の流れも関連している。

◆官僚制度に見られる木構造の問題点

- 大きさ、細分化の問題。縦割り行政(木構造行政)。

から、関連している情報を含んでいるテキストのある部分を引き出すことが簡単になる。

ハイパーテキストを拡張し、テキスト・データだけでなく、音声や画像などのデータを扱えるようにしたものを、ハイパーメディア(hypermedia)という。World Wide Web は、(木構造ではなく)ハイパーメディアに基づいて作られている情報提示のための仕組みである。



HTML や SGML では、「<」と「>」で括られた範囲がタグになる。タグには、「<name>」という形式と「</name>」という形式の2つの種類があり、前者を開始タグ 後者を 終了タグ といいます。開始タグと終了タグに囲まれた部分が、マークが付けされたテキストになる。

ハイパーリンクを実現するためのマークアップ言語には、次の2つの機能が必要になる。

- データにラベルを付ける(参照の目標となるマークを付ける)。

◆URL

HTML では、他のデータへの参照を実現するためにURL (Uniform Resource Locator) という形式を使う。次に、URL の例を示す。

<http://www.tsukuba.ac.jp/education/college.html>

http

HyperText Transfer Protocol。WWWのデータを保持しているプログラムと、WWWを表示するプログラムの間でデータをやり取りするときの形式を定めた約束。

www.tsukuba.ac.jp

そのデータを持っているコンピュータの名前。

/education/college.html

そのコンピュータの中での資源の名前(ファイルの名前)。最後の.html は、その資源がHTML で書かれている事を表わしている。

◆URL中の2つの木

URL には、2つの木構造の表記方法が混じっている。

- 木の根が左(ファイル名)
- 木の根が右(コンピュータの名前)

- 時間的に古くなる。環境変化についていけない。

中間管理職の意味==横方向に情報が流れない。

木構造でしか情報が流れないような組織は、潰れる。木構造を補う意味で、会社組織では、裏チャンネルや同期会が重要となる。

◆外務省機密費問題

会計法違反。外務省に分かれた予算を首相官邸に流した。国会の審議の意味がなくなった。

◆領土問題

木構造のどちらに付くか。

■ハイパーテキストとハイパーメディア

木構造を補う方法として、ハイパーテキストと呼ばれる方法を使うことがある。

ハイパーテキスト(hypertext)とは、内部に他のテキストへの「参照(reference)」が埋め込まれているテキスト(文書、文字だけから構成されるデータ)である。ハイパーテキストという仕組みを使えば、テキストのある部分

◆URLを間違えた時/URLが変更された時

2 種類のエラー

- Not Found. The requested URL **XXX** was not found on this server.
- Unable to locate the server URL **XXX**. Please check the server name and try again.

長い URL のどの部分が怪しいかを区別できるようにする。

エラーが出た時には、木構造で親を探してみる。