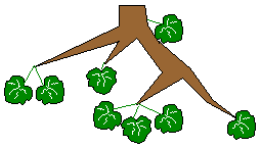
木構造

平成15年度 筑波大学第三学群 情報学類 大学説明会 ミニ講義B、 20

電子・情報工学系
新城 靖
<yas@is.tsukuba.ac.jp>

このページは、次の URL にあります。
http://www.coins.tsukuba.ac.jp/~yas/oc-2003
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.is.tsukuba.ac.jp/~yas/
http://www.coins.tsukuba.ac.jp/~yas/

印刷配布資料 http://www.coins.tsukuba.ac.jp/~yas/oc-2003/tree.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。
  - 節と枝。根も葉もある。根を上に書く。
  - 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。

■参考書

山口和紀, 古瀬一隆, 中村敦司, 新城 靖, 西山博泰, 林 謙一, 金谷英信, 鈴 木孝幸, 端山貴也: "新The UNIX Super Text 上/第34章 木構造とグラフ構造", 技術評論社 (2003年3月25日). ISBN: 4774116823.

■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。

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。

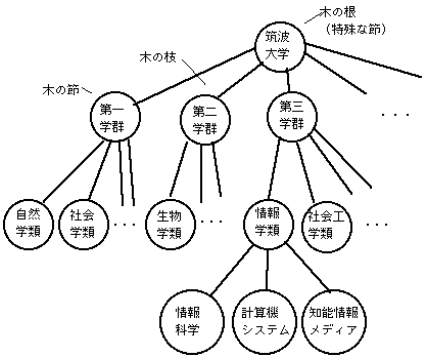


図1 大学組織に見られる木構造

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、コンピュータ・サイエンスでは、木の根を上 に書く習慣がある。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。

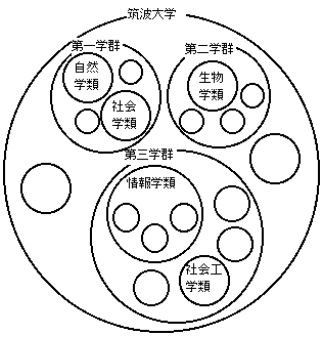


図2 大学組織に見られる木構造（領域的な見方）

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。

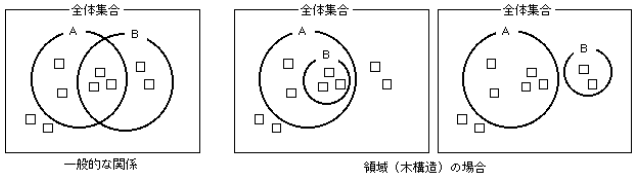


図3 ２つの集合の関係

◆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 が違うと違う意味になる。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.

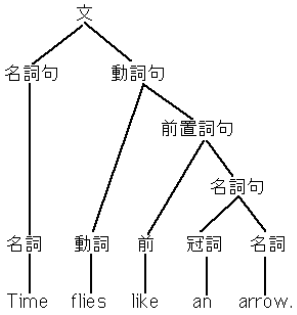


図4 「Time flies like an arrow.」の木(その１)

光陰矢のごとし。

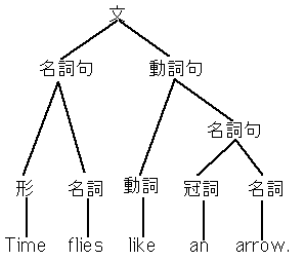


図5 「Time flies like an arrow.」の木(その２)

時蝿、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>

<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 because
    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 conjunction>
    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
```

受験生へのヒント：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。

- 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

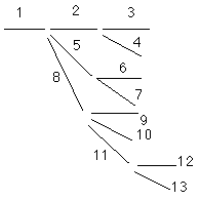


図6 topic sentence を根とする木構造

受験生へのヒント：（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。

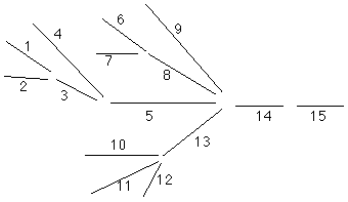


図7 逆茂木型の段落

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。木構造を知らない人が試 験問題を作る場合、木構造の知識が応用しくにい。

(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。)

◆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（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。

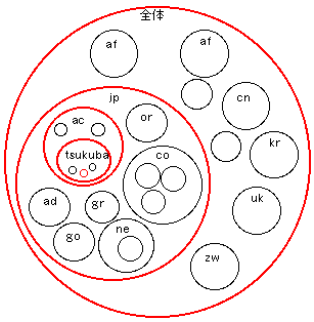


図8 名前空間のドメインへの分割

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

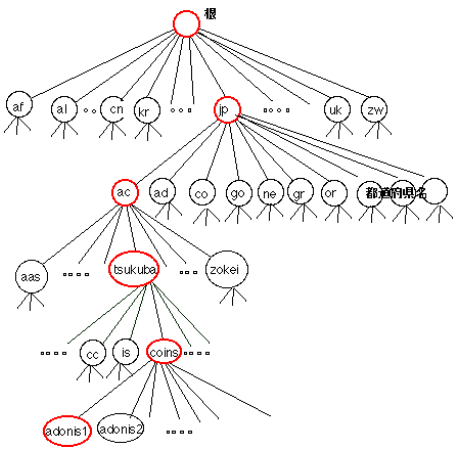


図9 名前空間の木構造としての見方

「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。
- 根の下の 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。

◆木構造の利点

- 分類をする時に木構造を使うことは自然である
- 要素数が増えても扱える
- 永遠に循環しないで全ての節を訪れることが簡単にできる

◆木構造の問題点

木構造では、ある節から別の節にたどり着く道は、ただ１つしか許されていない。

◆こうもりの分類問題

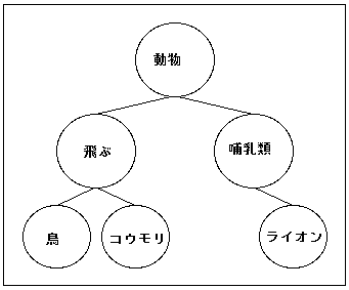


図10 こうもりの分類（１）

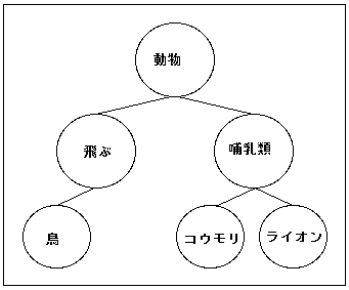


図11 こうもりの分類（２）

類似の問題は、領土問題、外交機密費問題、縦割り行政（木構造行政）問題にもある。

◆部分的にグラフ構造を取り入れる

１つの節に、１つの本名の外に、いくつかの「別名」をつけて、２つの道から たどり着けるようにする。検索する時には、本名と別名を区別する。

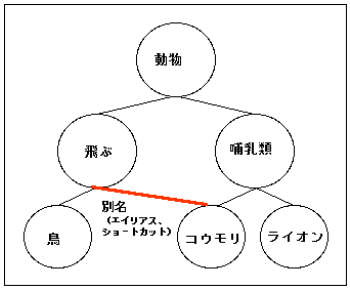


図12 こうもりの分類（別名つき）

■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(グラフ構造) と呼ばれる方法を 使うことがある。

(単なる)テキストとは、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(hypertext)とは、内部に他のテキストへの「参照 (reference)」（ハイパーリンク）が埋め込まれているテキストである。埋め 込まれた参照を使えば、テキストのある部分から、関連している情報を含んで いるテキストを簡単に引き出すことができる。

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（マルチメディア・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。

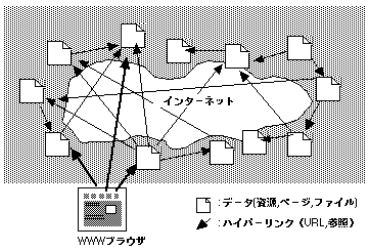


図13 インターネットの資源とハイパーメディア

ハイパーメディアやハイパーテキストのデータを作成するためには、次の２つ の事が必要になる。

1. 差されるデータに印(mark,label)を付ける。
2. 差すデータに、参照を埋め込む。

文書（テキスト）に、「ここは表題」、「ここは箇条書」といった、文書の構 造を示す目印（マーク）を付けることをを付けることを、マークアップすると いう。 ハイパーメディアを記述するためには、上の２つのことができる人工の言語を 使う。このような言語を、マークアップ言語(markup language)という。WWW では、マークアップ言語として HTML (HyperText Markup Language) と呼ばれ ている言語が使われている。

◆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。

- コンピュータの名前。区切り文字「．」。木の根が右(右から左に解釈)。
- コンピュータ内の資源の名前。区切り文字「／」。木の根が左(左から右に解釈)。

◆URLを間違えた時/URLが変更された時

次の URL をWWWブラウザで開くことを考える。

http://www.aaa.bbb.ccc/ddd/eee/fff.html

次の２種類のエラーが出る可能性がある。

1. Not Found. The requested URL **/ddd/eee/fff.html** was not found on this server.
2. Unable to locate the server URL **www.aaa.bbb.ccc**. Please check the server name and try again.

エラーが出た時には、木構造で親になる節を探してみればよい。

- /ddd/eee/fff.html が違うと言われた時
  - http://www.aaa.bbb.ccc/ddd/eee/fff.html
  - http://www.aaa.bbb.ccc/ddd/eee/
  - http://www.aaa.bbb.ccc/ddd/
  - http://www.aaa.bbb.ccc/
- www.aaa.bbb.ccc が違うと言われた時
  - http://www.aaa.bbb.ccc/ddd/eee/fff.html
  - http://www.bbb.ccc/ddd/eee/fff.html
  - http://www.ccc/ddd/eee/fff.html

www は、標準的な名前。