第 34 卷 第 1 期 计 算 机 工 程 2008 年 1 月
Vol.34 No.1 Computer Engineering January 2008
·软件技术与数据库· 文章编号：1000—3428(2008)01—0075—03 文献标识码：A 中图分类号：TP338

基于生物多样性的分布式计算软件资源分类

张加口[1,2,3]，曾国荪[1,2,3]

(1. 同济大学计算机科学与工程系，上海 201804；2. 国家高性能计算机工程技术中心同济分中心，上海 201804；
3. 同济大学嵌入式系统与服务计算教育部重点实验室，上海 201804)

摘 要：从生物多样性角度出发，借鉴了生物学的分类方法，结合分布式计算软件资源的特点，提出了一种分布式计算软件资源的分类技术。该技术可以形式化地描述和区分分布式计算软件资源，实现需求与分布式计算资源的最优匹配。分类试验结果表明，该分类技术可以识别 70%~80%的 ftp 软件，初步解决了异构资源难以区分的问题。

关键词：分类；异构性；生物多样性；分布式资源

Classification of Distributed Computing Software Resources Based on Biodiversity

ZHANG Jia-kou[1,2,3], ZENG Guo-sun[1,2,3]

(1. Department of Computer Science and Engineering, Tongji University, Shanghai 201804;
2. Tongji Branch, National Engineering & Technology Center of High Performance Computer, Shanghai 201804;
3. Key Laboratory of Ministry of Education for Embedded System and Service Computing, Tongji University, Shanghai 201804)

【Abstract】 After benefiting from the biodiversity characteristic and biology taxonomic classification, this paper proposes a classification scheme for the Distributed Computing Software Resources (DCSR) referring to the characteristic of the distributed computing software resources. DCSR can be clearly described and classified, also is matched to personal demand optimally. The result shows that the technology can identify 70%~80% ftp software, and preliminary solution is given for distinguish heterogeneous resources problem.

【Key words】 classification; heterogeneous; biodiversity; distributed resources

分布式计算环境是由不同异构计算机为实现资源共享而组成的虚拟计算机环境[1-2]。分布式计算环境通过虚拟组织技术，实现参与者之间资源的共享。分布式计算资源的异构性增加了信息系统之间资源存取的复杂程度。如何有效地发现资源、利用资源的异构性成了研究的热点[1]。传统的资源发现方法，如Netsolve、泛洪、分布式哈希表等[1-3]，可以发现异构性资源，即当给出一个所需资源的描述，资源发现机制就返回一个与描述匹配的资源位置。但是传统的资源发现方法仅仅实现了资源的可用，没有对分布式计算资源进行规范、统一的命名。用现有的不规范的资源标识名，很难准确地描述资源，对异构性资源也不能清楚地加以区分，不能最大程度地利用异构资源。本文从生物多样性角度出发，探讨了最大化利用生物多样性对自然界发展的意义，并引申出分布式计算资源异构性对分布式计算发展的意义。分布式软件资源是现有资源发现的重点和难点，本文主要研究分布式计算软件资源的发现。在论证分布式计算软件资源和生物资源相似性后，借鉴生物学的分类方法，提出一种针对分布式计算软件资源的分类技术。利用该分类技术可以对分布式计算软件资源进行规范、唯一的命名，并把分布式计算软件资源解释清楚，真正实现需求与分布式计算软件资源的最优匹配。

1 生物多样性及其分类方法

1.1 生物多样性

生物多样性是指所有动物、植物、微生物物种以及所有生态系统及其形成的生态过程的总和[4-6]。生物多样性的自然和社会价值是不可估量的，它是一种不可替代的资源，不仅是人类生存发展的基础，也是自然界持续发展的支撑力量。地球上现存的生物估计有 200 万～450 万种，如何对这些多样性的生物进行分类、整理，实现统一、规范的命名，这是生物分类方法所做的工作[6]。

1.2 生物分类方法

瑞典植物学家林奈——生物分类方法奠基人，在《自然系统》中提出了最初分类方法和生物命名法，但他认为物种是不会改变[4]。

在达尔文进化论的影响下，逐步形成了现今的生物分类方法。生物分类方法主要根据比较解剖学和胚胎学的研究结果，利用相似程度和亲缘关系进行分类[5-6]，经过 200 多年的发展，已经形成了比较成熟的分类体系。生物分类方法避免了不同种类间的混淆，识别了生物界中现存绝大部分的生物，为人类利用和研究生物界提供了基础。

生物分类方法(BTM)描述如下：

$$BTM=(\Omega,\Phi,\Theta,A)$$

其中，Ω 表示宏观分类系统的集合；Φ 表示包含分类阶元的

基金项目：国家自然科学基金资助项目(60673157)；教育部科研基金资助重点项目(105071)；上海高校网格技术 E-研究院基金资助项目(200301-1)

作者简介：张加口(1984－)，男，硕士研究生，主研方向：异构计算；曾国荪，教授、博士生导师

收稿日期：2007-06-04 **E-mail：**jiakouzhang@gmail.com

分类集合；Θ 表示同功器官和同源器官的集合；A为找出同源器官后，继续分类参照的标准集合。四元组 (Ω,Φ,Θ,A) 表示在给定宏观分类系统 Ω 基础上，利用 Θ 的数据并结合一系列的标准A，不断扩充现有包含分类阶元的分类集合 Φ，最终实现分类的唯一性。$\Omega=\{\Omega_i \mid i=0,1,2,3,4\}$={原核界,原生界,植物界,动物界,真菌界[6]}；$\Phi=\{\Phi_i\cup\Phi_{i+1}\cup\cdots \mid i\in classlist\cap classlist \supseteq$ {种,属,科,目,纲,门,界}$\cap\Phi_i\subseteq\Phi_{i+1}\}$。此外还有：

$$\Theta=\{\Theta_{ij} \mid i=0,1,2,\cdots,m;\ j\in\{af,hf\}$$

$$\Theta_{kj}=\begin{cases}1 & \text{if } k\in\{0,1,2,\cdots,m\}\cap j==af\\ 0 & \text{if } k\in\{0,1,2,\cdots,m\}\cap j==hf\end{cases}$$

$$A=\{A_{ij} \mid i,j\text{与}\Phi\text{中}i,j\text{相对应}\}$$

给定某个生物 B，则 $\exists\Omega_i\in\Omega$ 使得 $\Phi_{界}=\Phi_{界}\cup(\Omega_i)$，$B\in\Phi_{界}$。若 $\Theta_{ij}==1$，执行 i++ 直到 $\Theta_{ij}==0$ 为止，取 $(\forall A_{ij}\in A)\cap(A=A-A_{ij})$ 执行这个标准，$\exists\Phi_i\subset\Phi$ 使得 $\Phi=\Phi\cup\Phi_i$；如此循环操作直至集合 A 为空，若分类阶元层次>7，在分类过程中对 classlist 列表进行扩充，直到能够准确的描述分类结果。

图 1 显示了人类在生物学中的分类层次。

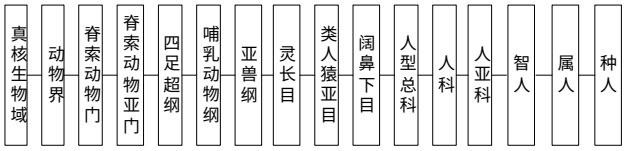

图 1　人类的分类层次

2　生物资源与分布式计算软件资源相似性

虽然生物与分布式计算软件是属于 2 个不同的学科领域，但是笔者发现了它们之间的相似性，可以用类比推理分析方法来证明。

纵观分布式计算软件资源发展的历程，发展规律可以概括为继承、创新、功能过剩、生存竞争。而分布式计算软件的生存演化本质上也是为了适应环境的变化。

给定一个三元组 $SP=<A,\beta,F>$，描述了在 A 领域内具有 β 属性特征的对象可以用 F 分类方法进行分类。其中，F 是生物的分类 $BTM=(\Omega,\Phi,\Theta,A)$。令

A={生物}

$\beta=\{\beta_1,\beta_2,\beta_3,\beta_4\}$

$\beta_1=$ {数量庞大}

$\beta_2=\{keyword=$(遗传,变异,繁殖过剩,生存斗争)}

$\beta_3=\{virtue=$适应环境变化}

$\beta_4=\{aim=$规范命名}

又令

A={软件}

$\beta'=\{\beta'_1,\beta'_2,\beta'_3,\beta'_4\}$

$\beta'_1=$ {数量庞大}

$\beta'_2=\{keyword=$(继承,创新,功能过剩,生存竞争)}

$\beta'_3=\{virtue=$适应环境需求}

$\beta'_4=\{aim=$规范命名}

不难发现，软件属性特征集 β' 与生物属性特征集 β 极其相似，可以得到新的三元组 $SP'=<A,\beta',F>$，即可以用生物学的分类方法对分布式计算资源软件进行分类。

3　基于生物分类方法的分布式计算软件资源分类

3.1　基本定义和命题

分布式计算软件资源分类方法形式化定义为

$STM=(\Omega,\Phi,\Theta,A)$

定义 1 分布式计算软件资源宏观分类系统为

$\Omega=\{\Omega_i \mid i=0,1\}$ ={开源软件界，非开源软件界}

定义 2 分布式计算软件资源的分类阶元，即

$\Phi=\{\Phi_i\cup\Phi_{i+1}\cup\cdots \mid i\in classlist\cap classlist\supseteq$ {组,股,群,族,种,属,科,目,纲,门,界}$\cap\Phi_i\subseteq\Phi_{i+1}\}$

定义 3 同源软件指用同种语言开发核心技术的软件。

定义 4 同功软件是指两种软件应用的功能是相同的。

定义 5 协同演化软件是指具有同种功能的两种软件，在演化上具有相互独立同时又相互适应的特性。协同演化核心是两种同功的软件之间相互竞争。

定义 6 趋异演化软件是指虽然二者属于同源软件，但在演化过程中由于不同的环境条件，使其具有较大的差异。

定义 7 Θ 表示同功软件、同源软件、协同演化软件和趋异演化软件的集合。

定义 8 A 为找出同源软件和协同演化软件后，继续分类参照的标准集合。

3.2　分布式计算软件资源的分类方法

分布式计算软件资源分类 $STM=(\Omega,\Phi,\Theta,A)$ 的目标是建立类似公司的人才库、医院病例库，其基本思想为：从“是否是 Θ 中的同源软件”思想出发，利用 A 中软件演化历史，根据 Θ 中不同演化阶段的趋异演化得到“同功软件”；根据软件架构和体系结构理论，从 A 中的体系结构模式和设计模式出发进行分类；以竞争关系为出发点，利用 A 中协同演化的特点对同功软件进行分类，最后成为使用者所需的个性软件。

本文分别利用结构体形式描述演化阶段和趋异演化分支的关系，以及协同演化和协同演化分支的关系。演化阶段和趋异演化分支的关系结构体定义如下：

```
typedef struct evol {
  EvolutionName en; //演化阶段名称
  struct evol *p;//上述演化阶段的趋异演化分支
} *pevol;//pevol 指针指向包含演化阶段的结构体动态数组
```

pevol = &结构体数组{(语言,指针),(平台,指针),(网络,指针),(服务,指针),(功能,指针),……}；

协同演化和协同演化分支的关系结构体定义与之类似。

以“语言”这个演化阶段来演示趋异演化分支的结构形式，其余阶段与此相同。

```
pevol[0].en = 语言；
pevol[0].p[0]->en = 面向对象门；
pevol[0].p[0].p[0]->en = Java 亚门；
…
```

依次类推；

分类方法伪代码表示如下：

```
function classify (pointer *pevol, pointer *pcomp, string targetsoftware) {
  int i = 0;
  while (pevol[i] != null) {
        select next step according targetsoftware;
        classify (pevol[i].p, pcomp, targetsoftware);
        i++;}
  if (pevol[i] == null) {
        int j = 0;
        while (pcomp[j] != null) {
              select next step according targetsoftware;
              classify (pevol, pcomp[j].p, targetsoftware);
              j++;}
```

}}

应用上述分类方法对 wu-ftp 和 vs-ftp 两种软件分类的结果见图 2。

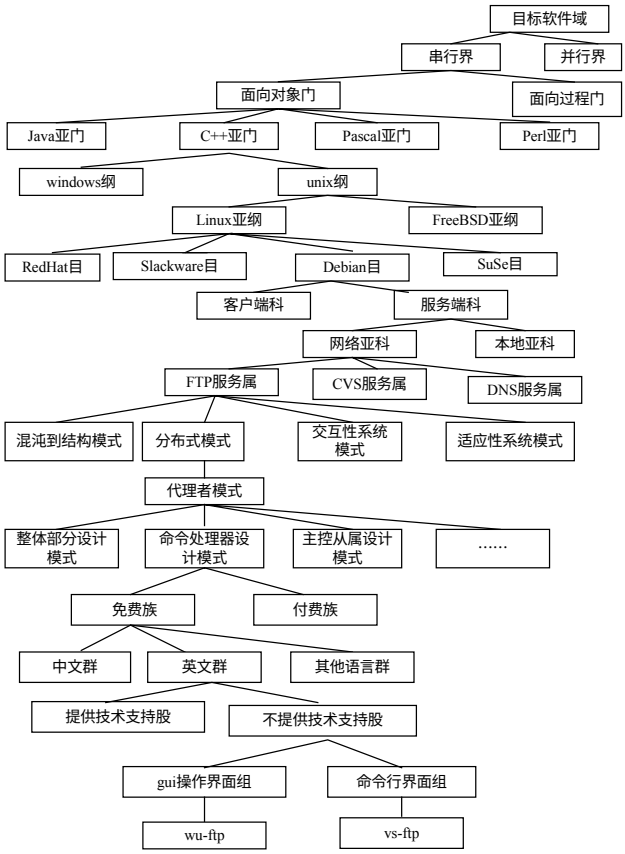

图 2　分类结果

4　实验仿真及结果分析

本文以 PlanetLab 网络平台为实验环境。作为计算服务“覆盖网络”，PlanetLab 是一个开发全新互联网技术的开放式全球性测试平台。以 ftp 软件作为分类的例子，利用不同 Linux 操作系统的软件包管理工具计算出现有版本下的 ftp 软件的数量，因此，上述实验节点安装了 Linux 操作系统，并以覆盖网络形式连接在一起。从实验结果(表 1)可以得出，本文提出的分类技术对各个不同 Linux 操作系统下 ftp 软件识别准确率为 70%~80%。该分类技术在一定程度上可以解决异构性资源难以区分的问题，并可以形成一套行之有效的资源命名规范。

表 1　实验结果

节点	操作系统	版本	软件数量(识别数量)	准确率/(%)
thu1.6planetlab.edu.cn	Debian	etch	329(265)	80.55
tongji1.6planetlab.edu.cn	Fedora Core	5	215(155)	72.09
tongji2.6planetlab.edu.cn	Slackware	9.0	254(179)	70.47
sjtu1.6planetlab.edu.cn	Mandrake	8.0	197(142)	72.08
sjtu2.6planetlab.edu.cn	Gentoo	2006.1	87(49)	56.32

5　结束语

本文提出的基于生物多样性的分布式计算软件资源分类技术是有效的，但该分类方法还不够完善，需要不断的发展。接下来的工作是要不断优化“演化历史”、不同演化阶段下的“趋异演化”分支和“协同演化”分类，更加合理、完整、全面地将分布式计算软件资源进行分类，提高分类结果的有效覆盖率，并在此基础上对分布式计算软件资源进行资源发现、查找、管理等方面的研究。

参考文献

[1] Foster I, Iamnitchi A. A Peer-to-Peer Approach to Resource Location in Grid Environments[C]//Proc. of the 11th Symp. of High Performance Distributed Computing. [S. l.]: Kluwer Publisher, 2002.

[2] Gupta A, Agrawal D, Abbadi A E. Distributed Resource Discovery in Large Scale Computing Systems[C]//Proceedings of Symposium on Applications and the Internet. Tokyo: [s. n.], 2005.

[3] Yang B. Improving Search in Peer-to-Peer Networks[C]//Proc. of the 22nd Int'l Conference on Distributed Computing Systems. Washington: IEEE Computer Society Press, 2002.

[4] 田清涞. 生物学[M]. 北京: 北京大学出版社, 1990.

[5] 陈阅增. 普通生物学-生命科学通论[M]. 北京: 高等教育出版社, 2003.

[6] 胡玉佳. 现代生物学[M]. 北京: 高等教育出版社, 2005.

（上接第 74 页）

虚拟现实系统和虚拟现实建模工具都是当前 VR 的热点话题，建模工具的提高有助于 VR 系统的更进一步发展。未来的 VR 建模工具将会朝着自动建模、静态建模、人机交互建模三者两两结合，甚至三者结合的方向发展，在三维建模的大规模化、精细化及可交互性之间寻找更加合适的平衡点。

参考文献

[1] Winterbottom S J, Long D. From Abstract Digital Models to Rich Virtual Environments: Landscape Contexts in Kilmartin Glen, Scotland[J]. Journal of Archaeological Science, 2006, 33(10): 1256-1367.

[2] Bishop I D, Gimblett H R. Management of Recreational Areas: Geographic Information Systems, Autonomous Agents and Virtual Reality[J]. Environment and Planning B: Planning and Design, 2000, 27(3): 423-435.

[3] Counsell J. 3D Built Form and Landscape from 2D Maps[J]. Habitat, 1998, 6(5): 41-43.

[4] Parsons S, Leonard A. Mitchell P. Virtual Environments for Social Skills Training: Comments from Two Adolescents with Autistic Spectrum Disorder[J]. Computers & Education, 2006, 47(2): 186-206.

[5] Takase Y, Sone A, Hatanaka T. A Development of 3D Urban Information System on Web[C]//Proc. of Processing and Visualization Using High-Resolution Images. Thailand: [s. n.], 2004.

[6] Takase Y, Sho N, Sone A, et al. Automatic Generation of 3-D City Models and Related Applications[C]//Proc. of International Workshop on Visualization and Animation of Reality-based 3D Models. Switzerland: [s. n.], 2003-02.

[7] Huang B, Lin H. GeoVR: a Web-based Tool for Virtual Reality Rresentation from 2D GIS Data[J]. Computers & Geosciences, 1999, 25(10): 1167-1175.

[8] Whyte J, Bouchlaghem N, Thorpe A. From CAD to Virtual Reality: Modelling Approaches, Data Exchange and Interactive 3D Building Design Tools[J]. Automation in Construction, 2000, 10(1): 43-55.