

未来互联环境

诸葛海

1960年,马文·明斯基(Marvin Minsky)¹曾经预言:在3~8年的时间里计算机将和人一样聪明。但近半个世纪过去了,计算机系统还是无法通过“图灵测试”。除了在机器人技术、数学定理验证、科学分类和高级用户界面等几个方面取得了进展之外^[1],智能科学仍旧不尽如人意。

科学家和工程师们几乎已经实现了万内瓦尔·布什(Vannevar Bush)²利用互联网和WWW技术创建通用多媒体数据处理机的梦想。如果将这个视野延伸到未来,图灵奖获得者吉姆·格雷(Jim Gray)³预测:高安全、高可用、自编程、自管理和自复制的计算机和网络将得到发展^[2]。格雷设想了一个系统,类似于布什所提出的Memex⁴(可以自动组织、索引、分类、评价和提取信息)。但建立能够自编程、自管理和自复制的智能网络仍然是一项巨大的挑战。

中国知识网络研究组(<http://kg.ict.ac.cn>)成立于2001年,致力于探索未来互联环境的自然法则。

如今,网络已遍及自然、社会和虚拟世界,规定了各种资源和行为的结构与功能。探索与规范未来互联环境的准则成为一大挑战。(色块,放在第1页,4号字)

面向全新的计算环境(2级标题)

Web的出现为人工智能的研究和应用提供了一个前所未有的平台。由于Web可以帮助人们访问存放在任何已经接入互联网的数字内容,它给全球的商务、科研、政府和公共信息服务带来了重大变革。

然而,由于机器无法有效地理解为方便人类阅读而设计的网页,目前的Web还不能充分地支持智能应用软件。这些应用软件需要一种全新的互联网应用平台,以便智能地适应全球分布的、基于“Web服务描述语言”等开放标准的服务(e-services)所需要的开发、调度、交互和管理。

科学家们正在使用符号推理、文本挖掘、信息抽取与恢复、及其他一些技术,来改善和扩展Web。例如,IBM的WebFountain(www.almaden.ibm.com/webfountain),可以将Web页面、电子邮件、消息框和聊天等在线内容转换成基于XML技术的格式,并且对其进行分析,以确定其商业价值。语义Web^[3]的目的是通过使用本体和逻辑等机制,以及开发标准的标记语言,例如资源描述框架(Resource Description Framework, RDF),让网上内容的意义得到更好的表达和理解。

此外,下一代互联网Internet2⁵将比当前的互联网快几百倍,并且更加安全。通过为高级应用提供更加丰富的地址空间,下一代互联网将推动互联网应用平台的变革,而这种变革反

¹ MIT教授,人工智能专家,1969年图灵奖得主。

² 曾任美国国家科学研究与开发办公室主任,他提交给罗斯福总统的一份题为《科学—无尽的战线(Science, the Endless Frontier)》的报告,对于美国的科技政策产生过极为深远的影响。

³ 1998年图灵奖得主。

⁴ 1945年,万内瓦尔·布什提出了一个Memex(存储扩充器)的系统。Memex建立如下思想基础上:机械化图书馆将各种信息存储在缩微胶片中,各书目之间的链结可以自由转移。尽管这一思想从未得到实现,布什仍被认为是超文本的鼻祖,他在1945发表的论文“As We may Think”一直是一个参考标准。

⁵ Internet2是由美国许多大学、研究机构、公司和政府组织共同建设的网络,目的是为美国的大学和科研单位建立一个技术领先的网络。该项目始于1996年。

过来将产生更多的新应用。

研究人员正在研究其他人工网络的先进功能。例如，网格技术 (www.gridforum.org) 更加关注共享、管理、协同和控制分布式计算资源，包括机器、网络和任意接入的数字设备中的数据。开放网格服务架构 (Open Grid Services Architecture, OGSA⁶) 试图结合网格的优势和语义 Web 与 Web 服务的优势，但它并不太适用于大规模、非稳定性的动态网络。

对等网络 (P2P⁷) 在支持计算机资源共享方面已得到了广泛应用。尽管对等网络具有天然的自治和可扩展性，但它缺乏必要的理解、协调和调度能力来支持高级应用^[4]。

未来的互联环境必须吸收 AI⁸和分布式系统的精髓，继承 Web、语义 Web、网格和对等网络等技术的优点，并且通过采纳新的原理超越这些技术范畴。

原则、系统参数和挑战 (2 级标题)

计算环境已经从个人计算或集中式计算发展为分布式网络计算，再发展为人机环境，未来将发展为大规模人机环境。如图 1 所示，未来的互联环境将是一个大规模的人机环境，其中整合了以下三个世界：

物理世界 (physical world) — 包括自然界、自然和人造的物体、物理设备和网络；

虚拟世界 (virtual world) — 主要由视觉 (例如文字、图像、色彩、图形等等) 和听觉 (某种程度上也包括触觉、嗅觉和味觉) 创建的可感知的环境；

精神世界 (mental world) — 全部源于思维、感情、创造和想象力^[5]的观念、信仰、道德、文化、艺术、智慧和科学知识。

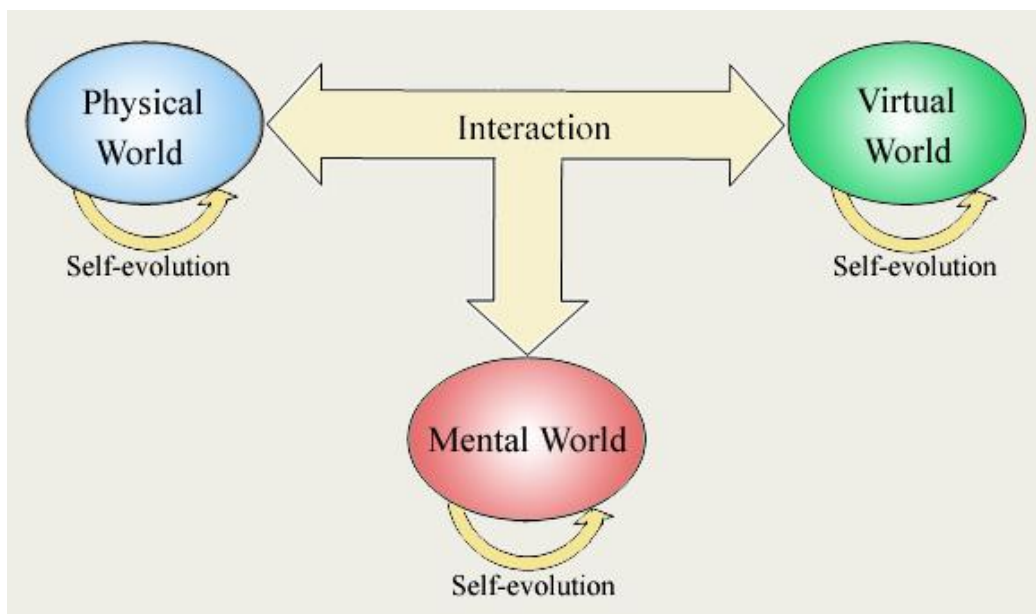


图 1 未来互联环境。物质世界、虚拟世界和精神世界将彼此作用自我演化、协同发展。

在理想状态下，这个环境将是一个自治、活跃、可持续的，并且是智能的系统。在这个系统内，社会和自然协同发展。这个环境将收集和组织资源，使其具有语义丰富、机器和人都可以方便使用的格式。在这个环境下，广泛分布在不同地理位置的用户将可以通过网络积

⁶ 见 www.globus.org/ogsa

⁷ Peer-to-Peer

⁸ Artificial Intelligence 人工智能

极促进物质、能量、技术、信息、知识和服务的流动，来合作完成任务，解决问题。

原理（3级标题）

未来互联环境的发展遵循以下原理。

开放性 使环境开放，以避免停滞。标准是保证系统开放的关键，但标准也必须随环境的演化而不断更新。

增量式发展 环境将从小规模发展为大规模、从简单发展为复杂，其增量可能发展为指数级。与此同时，开发人员的数量随之增长，技能也逐步提高。从应用的发展前景看，发展应该平衡继承和创新这两个方面。工作环境和工作模式的平稳升级将确保新技术的有效应用。

经济性 经济性为参与者、资源和环境带来合理、公平的分配。市场使参与者理性地调整他们的决策和行为，无论是生产者还是消费者都在寻求满意的、而不是最佳的结果，因为他们之间必须达成协议。这一简单的机制避免了复杂的计算。

生态性 未来互联环境将孕育一个复杂的生态，从而引发对自然世界、虚拟世界和人类社会之间相互作用与影响的深入探索。

竞争与合作 环境中的资源方必须为了生存、权益和声誉展开竞争。与此同时，他们还得彼此合作、互相约束，以保障其赖以竞争的服务的功能和价值。

动态可扩展性 参与者和资源必须能够在不影响环境整体功能的前提下加入或离开环境。网络及其关系和组织层应该支持这种动态可扩展性。

完整与简洁性 环境的魅力在于它自身、个体、种群和社会的底层结构的完整与简洁性。

未来互联环境将是一个可持续的和谐系统。（色块，4号字）

系统参数（3级标题）

未来互联环境将是一个可持续的、和谐的系统，这体现在以下几个方面：

空间 容纳多种个体和共享资源，包括物体、信息、知识、服务和自然环境中的物理空间；

时间 进化和衰退的过程；

结构 环境本身的构成和环境各种资源的构成；

关系 过程之间的关系、资源之间的关系，以及过程和资源之间的关系；

价值 资源、过程、及其关系的状态及发展趋势。

爱因斯坦的广义相对论揭示：物理世界中空间和时间都是具有延伸性的实体。在最大的尺度上，空间是动态的，随时间扩展或收缩。

未来互联环境将通过在不同层面（从物理层到人机社会层）、在不同学科领域支持社会行为，来促进知识资源的增长。作为社会的一种自然产物，知识的演化伴随人类的发展，而不仅仅是个体的一生。

人类的社会行为在很大程度上依赖于自然语言语义学。未来的社会行为将依赖于一种新型语义学，这种语义学可以在人类和无生命资源之间建立理解。人机语义学将使服务与知识得到有益的、安全的应用和控制。

挑战（3级标题）

未来互联环境的多样性和复杂性制约了用单一理论来建模的可行性。这需要增强洞察力，跨越传统学科界限，以迎接下面一系列的挑战。

资源重构 准确和完整的资源管理需要一种组织手段。关系模型虽然成功地确保了数据管理，但它并不能管理动态、分布环境中复杂、语义丰富的资源。我们需要一种新的理论来

支持语义丰富的资源组织形式，并在保持完整性的条件下使用资源。第二代互联网的先进特性使这一理论切实可行。

协调规范化和自组织 规范化反映稳定和秩序，而自组织则反映了无组织现象中的序。

“小世界”⁹现象说明了无尺度网络¹⁰中的稳定性。资源组织的规范化保证操作的准确性；自组织则保证了自治、平等和适应性。一种调和规范与自治的方法就是在某一层规范组织、在另一层实现自组织、并且在层次之间保持映射和一致性。

语义互联 在多语义层上连接各种资源以支持智能应用。关键在于建立一个语义计算模型，该模型既适用于显式语义，又适用于和感觉、感情相关的隐含语义。自组织资源的“感知和感性”在语义互联中同样具有重要的作用。

智能聚融 智能化服务依赖于聚类、重聚异构资源的能力。由于当前的被动资源模型并不支持主动聚类，因此非常有必要建立一个智能的资源模型。有了这个模型，就可以在各种实体或者各种内容之间产生融合。

网络退化 研究人员已经对 Web 的增长和分布状况作了大量研究⁸，然而却在很大程度上忽略了其退化。但是，在现实世界中任何事物的发展终究都会达到一个极限。因此，弄清退化是否有可能影响或限制未来互联环境的发展、以及如何影响或限制未来互联环境的发展是非常重要的。

抽象流建模 寻找适用于物质、信息、知识和服务流的共同规则和原理，揭示它们间互相支持的特性是另一个挑战。我们需要使用优化和控制等方法进行抽象过程建模。

域论 类似于物理世界，在未来的互联环境中资源从高能量的节点流向低能量的节点。这自然需要恰当的、按需供应的后勤特性，因为能量的差异反映了对流的需要。但是，现实世界中的能量守恒法则在此并不适用，因为复制、产生和删除数据几乎无须物理成本。适用于该领域的基本法则和原理需要进一步探究。

抽象资源 抽象是理解、思考和解决问题的基础。从资源中自动获取语义，并且在一个统一的语义空间中进行推理和解释是一项挑战。环境需要一个单一的语义映像机制⁹，该机制可以为各种资源建立一种共同理解。其约束和规则保证了在语义空间中资源的有效使用。

生态学 未来互联环境中的资源归为种群。种群的演化依赖于物种间的流。类的发展取决于样本之间的流动。其挑战在于应用自然生态学的方法和原理、帮助理解和探索未来互联网生态学。

动态继承 继承是有机世界中的一个普遍现象，也是面向对象方法论和系统中实现复用的机制。如何在一个演化的环境中，在演化的资源之间实现继承机制是一个挑战。继承应该遵循生态学、生物学和社会学的相关原则。

生物界面 传感器网络正逐渐使互联网络无所不及。科学家们已在尝试在动物的身体内植入传感器，以整合生物数据和电子数据，而且他们正在把人体作为网络中的一部分为计算机提供能量。这些传感器将成为未来互联网的一个有机组成部分。然而，从传感器所搜集来的低层信息，与能够被自动理解和智能化处理的高层信息之间还存在鸿沟。

有机体 一个真正的动态网络应该拥有组织特征，例如自保护、自治疗、容错、动态适应、自复制、自激励和自推动力。这就需要我们建立一个系统架构，类似于解剖学中的免疫系统、神经系统、消化系统和循环系统。

方法论 一个大规模的、动态的、智能的互联网需要跨学科的系统方法论，以及用于指

⁹ 所谓小世界现象，或称“六度分离（six degrees of separation）”，是社会网络（social networks）中的基本命题，即每个人只需要很少的中间人（平均 6 个）就可以和世界上另外任意一个人建立起联系。

¹⁰ 相对“随机网络”的一个概念。在随机网络中，各节点的连接数目是接近均匀的；而在“无尺度”网络中，少数节点的连接数目大大高于平均值，而多数节点的连接数目非常小。互联网就具有这样的特征：少数网页承担了绝大部分连接，而 80% 以上的网页平均连接数低于 4。

导开发、运行和维护网络及其应用的认识论。

架构和互联规则（2级标题）

图2是根据增量开发原则所提出的未来互联网的分层参考结构。最底层是一个介于现实世界和虚拟世界之间的界面。科学家们利用传感器网络收集信息、利用各种设备来控制小规模的自然环境，例如人工降雨^[10]。自然界赐予技术革新最直接的灵感，如基因计算、神经计算、集群智能和生物分子计算等。

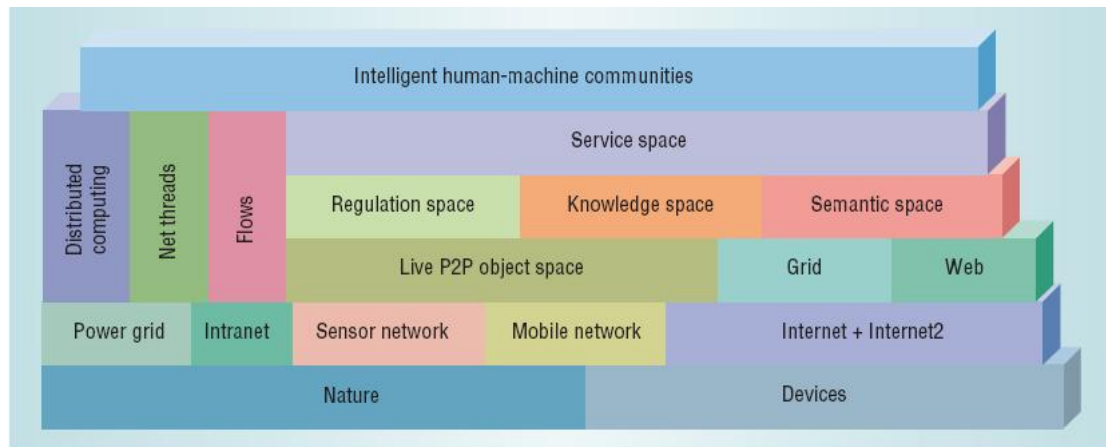


图2 未来网络互联参考架构。自然界赐予技术革新最直接的灵感。

活的对等网络对象空间包括对于各种环境资源和每个对象生命跨度的抽象表示。网络线程运行应用程序。规则空间中的对象实现自主资源管理。知识和语义空间覆盖了对等网络物体空间，并且支持服务空间。服务可以找到由规则和资源所提示的需求。人们在智能人机社区中根据规则和社会原则工作、休闲、贡献和享受各种服务。个人的角色可通过物质、信息、知识和服务的流从一个社区迁移到另一个社区，这些流可以连接社区和利用计算资源。

物体空间增长（3级标题）

活的对等网络对象空间由语义连接和资源节点组成。网络中的每个节点都有一个生命周期，经历从“出生（加入网络）”到“死亡（离开网络）”的过程。管制网络增长的规则必须同时考虑节点与连接的生和死。

作为一个案例，中国知识网格研究组的研究人员对比了Web超链分布和两种现实语义网的连接分布模型，一种是随机增长模型，另一种是有向进化图模型，且这两种模型都获得了同样的无尺度（scale-free）分布规则。

图3所示是入度和出度分布，分别显示了利用这两种模型和Web数据所得到的结果^[11]。模型中连接和节点的损失说明了结果的差异：离开的速度越快，不一致性越大。进一步分析得知，所研究模型的曲线斜率始终要比Web图形模型的要小。连接的数量代表了节点的价值：富节点拥有更多的连接。“富者越富”的现象使富者和贫者间的差距越来越大。

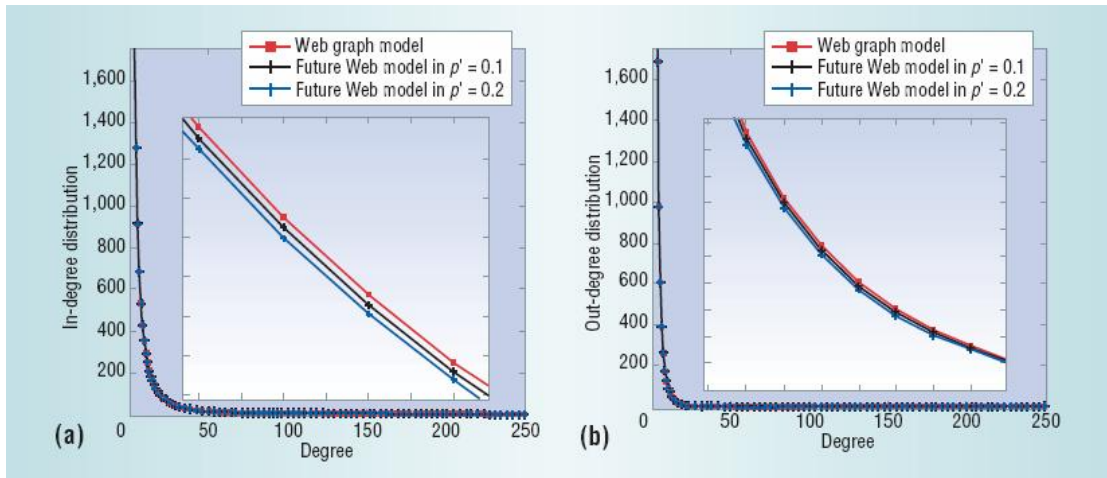


图 3 活的对等网络对象空间增长。(a) 入度和 (b) 出度分布状态分别表示随机增长模型和有向图模型，并与 Web 数据进行对比。

阻尼效应 (3 级标题)

经验表明，富节点要比贫节点持续时间更长，但财富的积聚在一定条件下会受限，阻止节点获得更多连接。图 4 所示与 power-law 不同的分布，尾部的增长接近极限，这意味着相对富节点分享了财富。

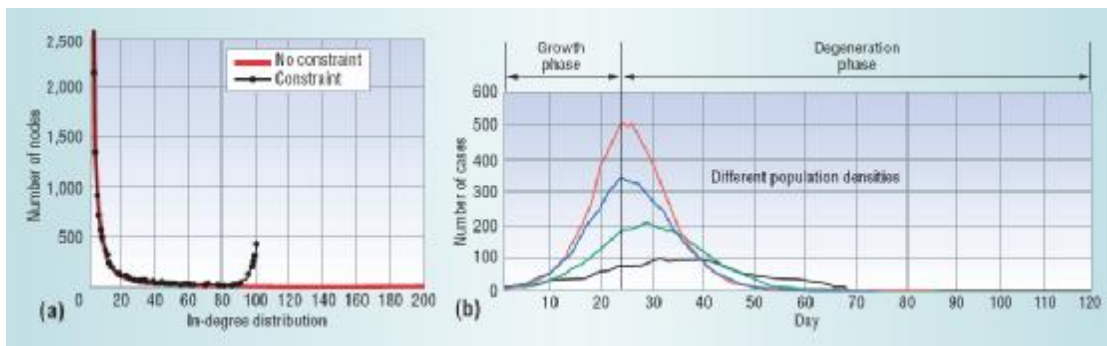


图 4 阻尼效应。(a) 是入度受限下的分布；(b) 是 SARS 传染扩散网络。

阻尼效应同样在很多现实网络中存在，使他们由兴转衰。例如，在传染病扩散网络中，被感染后，患者作为节点加入了该网络；而患者治愈或者死亡后，这些节点又离开了网络。图 4 (b) 所示的是 2003 年发生在国内的严重急性呼吸系统综合症 (Severe Acute Respiratory Syndrome, SARS) 传播中因抗体的发展及社群自我防护措施而形成的阻尼效应的实例。

在未来互联网中，资源也在类似的阻尼效应下开展竞争。富节点发出更多的信息、知识和服务；而贫节点则更容易变成富节点。一些社会和自然法则也将在互联环境中起作用。

压缩和扩展 (3 级标题)

图 5 是未来互联网的时空模型。收缩和膨胀如同潮涨潮落，对发展起到了牵动的作用。收缩强化了技术的竞争力，如第一代互联网、第二代互联网、Web、网格和对等网络推动了一些技术出现，产生导致膨胀的新技术，例如语义网格和知识网格领导了目前的膨胀。

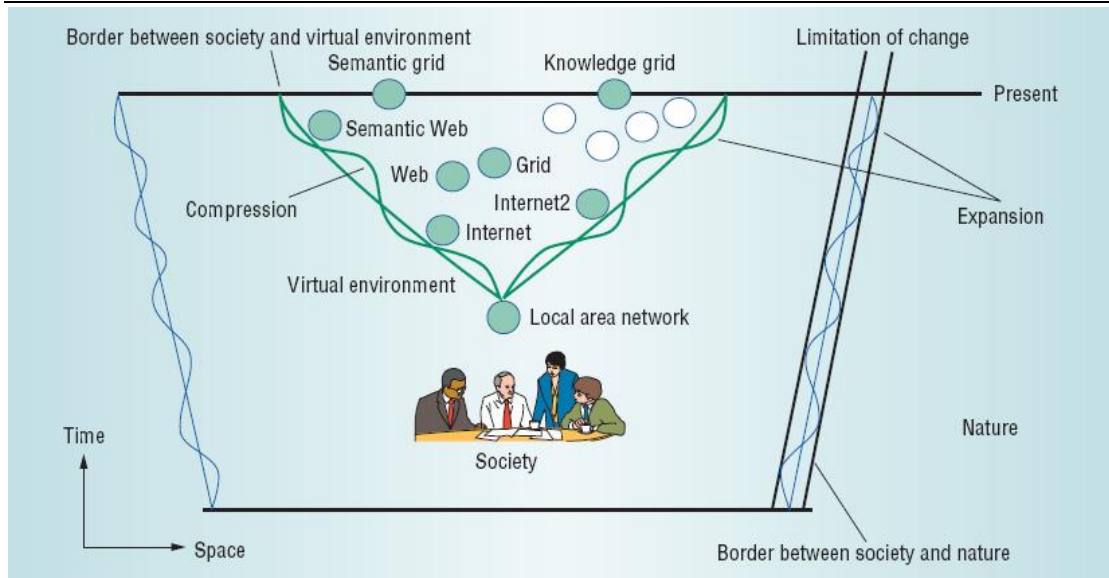


图 5 时空模型。收缩和膨胀对未来互联环境的发展起到了阻止和推动的作用，如同潮起潮落。

收缩和膨胀的范围影响可持续性。自然生态系统中，可持续发展的研究成就，对于未来互联网环境可持续性研究具有启示意义^[12]。

在互联环境中各种资源间相互竞争以求生存^[5]。一些资源在竞争中成为主导，正如 Web 中“富者越富”现象^[8]。

收缩和膨胀改变了空间与自组织行为。社会的收缩和膨胀影响了虚拟环境的发展，而虚拟环境的发展将逐渐和社会发展相融合，并与社会发展保持同步。

社会环境和虚拟环境间的边界将从屏幕和键盘逐渐演进为各种移动设备、传感器和生物界面。

演化的 E-SCIENCE 环境（2 级标题）

中国知识网络研究组正在设计 e-Science 知识网络环境，作为一种试验性的未来互联环境的微观世界^[9]。设计这个演化、动态、自组织、自我管理且可扩展的系统是为了支持多样的、分布的、智能的信息、知识和服务的发展。

如图 6 所示，这个架构包括以下五个方面的元素：

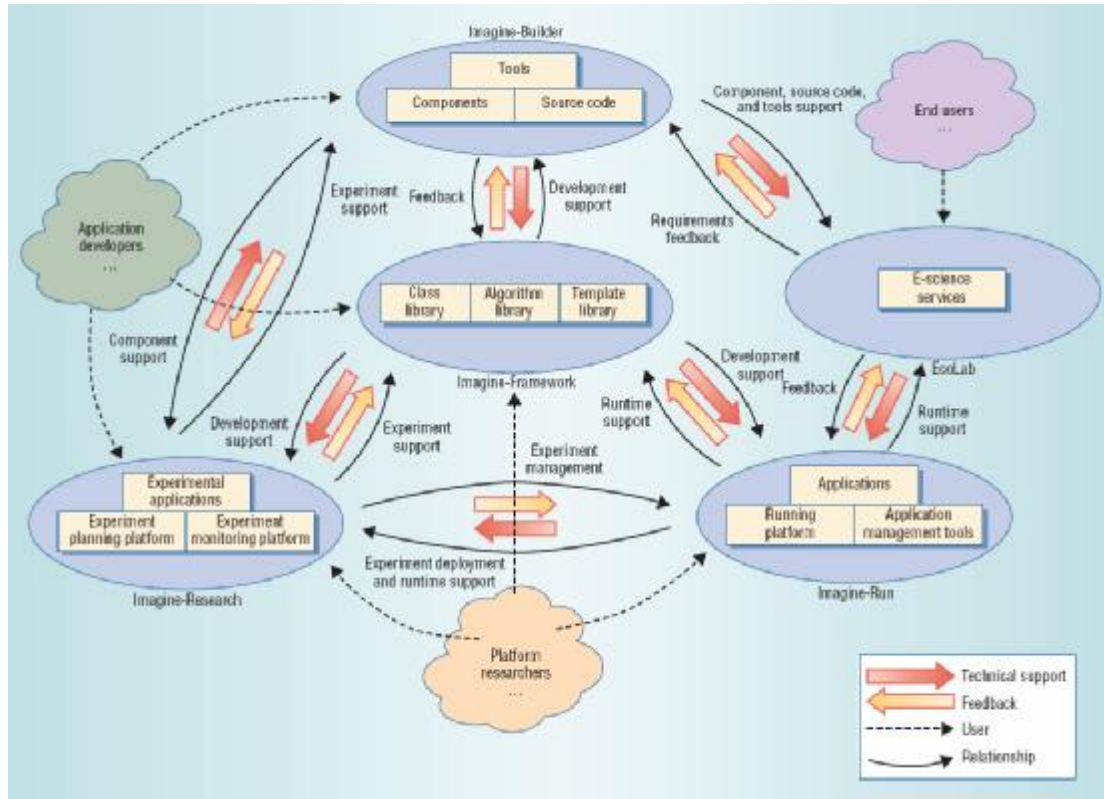


图 6 中国 e-Science 知识网格环境。是一个演化、自组织、自管理、可扩展的系统，其设计目的是在对等网络中支持分布式生态学应用的开发。

平台框架 (IMAGINE-Framework) 是核心的构件，用以支持系统的开发。该基本应用的发展框架以一种分散和自治的方式在对等网络中组织和管理资源。其中包括用来开发和管理高级别应用程序的类库、运算法则库和模板库。

运行平台 (IMAGINE-Run) 是一个支持底层对等网络和高级应用 (包括网络维护、应用程序设计、系统配置和执行控制) 实时管理的嵌入式平台。

研究平台 (IMAGINE-Research) 是一个虚拟的网络实验室，用于支持监控、调试、配置、测试和验证，以促进环境的演变和发展。研究人员和用户可以在其中开展全面的交流，以形成一个需求和技术的正反馈循环。

建造平台 (IMAGINE-Builder) 是一个包含工具、源代码和虚拟组件的平台。这些工具、源代码和虚拟组件可以促进大规模网络分布应用的开发。

生态实验室 (EcoLab) 是一个虚拟的科学研究实验室，地理分散的生态学者可以使用这个实验室有效地发布、共享、管理和使用分布的资源，包括大型对等网络中的分布计算能力、数据、信息和知识。该实验室可以反馈用户的需求，以改善平台和应用。

在运行平台和研究平台开发完毕之后，我们将以这两者为基础来检验和改善各种技术和软件，且将开发框架扩展到各类大规模动态网络中。平台中的研究人员、应用开发者和最终用户合作使用平台框架、研究平台和运行平台，最终改善整个环境。

中国知识网格研究组一直在探索如何实现理想的未来互联环境，正在进行中的工作包括探索互联语义、研究先进的高层机制 (如动态软设备的继承)，以及它们在 e-science、e-culture、跨学科跨文化研究系统开发中的应用。

致谢

本文研究受到国家重点基础研究发展 973 计划（2003CB317001）和国家自然科学基金的资助。

诸葛海 中国科学院智能信息处理重点实验室主任（<http://www.knowledgegrid.net/~h.zhuge>），是知识网格领域的先驱之一。

参考文献

- [1] D.L. Waltz. "Artificial Intelligence: Realizing the Ultimate Promises of Computing". AI Magazine, Vol. 18, No. 3, 1997, pp. 49-52.
- [2] J. Gray. "What Next? A Dozen Information-Technology Research Goals". J. ACM, Vol. 50, No. 1, 2003, pp. 41-57.
- [3] T. Berners-Lee, J. Hendler, and O. Lassila. "Semantic Web". Scientific American, Vol. 284, No. 5, 2001, pp. 34-43.
- [4] H. Balakrishnan et al. "Looking Up Data in P2P Systems". Comm. ACM, Vol. 46, No. 2, 2003, pp. 43-48.
- [5] H. Zhuge and X. Shi. "Eco-Grid: A Harmoniously Evolved Interconnection Environment". Comm. ACM, Vol. 47, No. 9, 2004, pp. 79-83.
- [6] J. Kleinberg. "Navigation in a Small World". Nature, vol. 406, 24 August 2000, p. 845.
- [7] J. Kumagai and S. Cherry. "Sensors and Sensibility". IEEE Spectrum, July 2004, pp. 18-24.
- [8] L.A. Adamic and B.A. Huberman. "Power-Law Distribution of the World Wide Web". Science, vol. 287, No. 24, 2000, p. 2115.
- [9] H. Zhuge. "China's E-Science Knowledge Grid Environment". IEEE Intelligent Systems, vol. 19, No. 1, 2004, pp. 13-17.
- [10] S.K. Moore. "Electronic Rainmaking Technology Gets Mexico's Blessing". IEEE Spectrum, Apr. 2004, pp. 14-15.
- [11] H. Zhuge. The Knowledge Grid. World Scientific, 2004.
- [12] A.J. McMichael, C.D. Butler, and C. Folke. "New Visions for Addressing Sustainability". Science, Vol. 302, No. 5652, 2003, pp. 1919-1920.

原文 H.Zhuge. The Future Interconnection Environment. IEEE Computer, 38(4), 2005, pp. 27-33.
根据中国计算机学会与 IEEE Computer Society 的合作协议发表该文的中文版。