

在数字图书馆结构中融入知识组织系统*

Linda Hill, Ph.D.¹ Olha Buchel, MLS.¹ Greg Janée, MS.¹ 曾蕾, Ph.D.²

¹(美国 加利福尼亚州 加州大学圣巴巴拉分校 亚历山大数字图书馆项目)

²(美国 俄亥俄州 肯特州立大学 图书情报学院)

[摘要] 以亚历山大数字图书馆项目将数字图书馆馆藏、知识组织系统、服务设施在数字图书馆环境中互相配合的实践为例, 讨论在数字图书馆结构中将知识组织系统作为有机的组成部分的意义和具体方法, 提出数字图书馆和分类学研究两方面应该探索的新课题。

[关键词] 知识组织系统 数字图书馆

Integration of Knowledge Organization Systems into Digital Library Architectures

Linda Hill, Ph.D.¹ Olha Buchel, MLS.¹ Greg Janée, MS.¹ Marcia Lei Zeng, Ph.D.²

¹(Alexandria Digital Library Project, University of California, Santa Barbara,
California, USA)

²(School of Library and Information Science, Kent State University, Ohio, USA)

[Abstract] Discusses the importance of integrating knowledge organization systems (KOS) into digital library architectures and the approaches used at the Alexandria Digital Library (ADL) project, in which the collections of the ADL, knowledge organization systems, and services have been integrated. The agendas for digital library and classification research in relating to KOS are also proposed.

[Keywords] Knowledge Organization Systems Digital Library

1. 引言

为数不少的数字图书馆研究与发展基本上始终围绕着建设数字化馆藏 (collections)和服务设施 (services)两方面 (Arms, 2000), 在某种程度上, 另外的一个焦点是用户(users)和使用 (uses) 以及数字图书馆的结构怎样满足用户的需要 (Borgman, 2000)。这种数字图书馆的发展中漏掉了一组结构, 即我们称之为“知识组织系统”(knowledge organization systems, KOS) 的资源, 也就是我们用来组织和定义用于表述和组织真实世界物体的术语和符号的各种既熟悉但又处于变革中的知识组织系统。就像数字图书馆中的馆藏和服务设施能够按通用的、支持整个结构的框架作出模型来一样, 我们也完全有可能在数字图书馆中融入各种知识组织系统, 将数字图书馆的总体结构延伸扩大, 以便在发展和使用其馆藏和服务设施的过程中同时考虑知识的表述和组织问题。在数字图书馆结构中将知识组织系统作为有机的组成部分, 这对数字图书馆和分类学研究这两方面的人来说都是新的课题。

2. 数字图书馆中的知识组织系统

知识组织系统主要有以下一些类型(以下划分是根据 NKOS 网站上对 KOS 的划分和 Hodge 2000 文章修改而成的):

- 系统分类和大致分组的模式
 - 大致分组归类的类表 (categorization schemes): 十分松散的结构, 可以是任何分组归类用的大纲。
 - 系统分类的分类表 (classification schemes): 类表中将用于表达泛指主题的类号(数字或字母型)按照等级或分面形式排列。
 - 标题表 (subject headings): 提供一系列用以表达一个馆藏中各文献主题的受控词, 以及一套将标题组配成复合标题的规则。
 - 知识分类表 (taxonomies): 根据事物的某种特征将事物分成有序的分类组, 例如生物分类学的严格分类体系。
- 元数据式的系统模式
 - 指南 (directories): 名称及相关信息的列表。

- 地名辞典 (gazetteers): 含有名称和类型划分的有关地点的地理参考字典, 辞典中将地理位置词间的相互关系通过地理表达方式以及明确的关系类型 (例如“属于 XXX 的一部分”) 来表示 (Hill, 2000)。地名辞典可以扩展到包括事件 (如飓风) 和有命名的时间阶段, 在第二种情况下地理参考数据变成时间跨度。
- 关系模式
 - 实用分类系统 (ontology) (又可以理解为概念空间 concept spaces): 用以表述十分复杂的事物间的相互关系的特定概念模型, 其中包括语义网络模式中所缺乏的规则和公理。
 - 语义网络 (semantic networks): 表达概念的词汇集, 按照在复杂多变的关系网络中的节点的模式建造。
 - 叙词表 (thesauri): 表达概念及其等同、等级、相关关系的词汇集。这种词表的结构是根据美国国家标准组织 ANSI/NISO (1993, R2003) 和国际标准组织 ISO (1986) 所制定的词表标准而建的。另一种类型的词表 (如 Roget's Thesaurus) 则只表现词间的等同关系 (同义词), 以及词汇的归类。
- 词汇单
 - 规范文档 (authority files): 用于控制用于同一人或事物的不同名称, 或者为某个特别领域中专用名词的词汇单。
 - 字典 (dictionaries): 按字母顺序排列的词单, 提供词的定义, 其中包括对每个词的各种词性的解释。
 - 术语表 (glossaries): 按顺序排列的词汇, 通常带有定义。

在一个数字图书馆环境中, 知识组织系统可以通过其共同的元素和一系列相同的功能来判断, 一般来说, 所有这些词表类表都含有标识 (labels, 包括语词和类号), 其含义及其相互之间的关系。知识组织系统所支持的功能包括:

- 描述: 受控的标识集, 用以描述事物
- 定义: 标识的意义

- 翻译：等价表达形式之间的匹配
- 导航：在一个有组织的表达结构中的联接

除此之外，各种知识组织系统都带有自己的观点，往往以某学科知识体系为模式，为特定目的而设计，所以，每个知识组织系统都有一个总体结构、范围、目的，理解这一点对诠释其内容十分重要。

与之相反，数字图书馆中的馆藏 (collections)可以被看作是“一组的物件”，它们由各种以单个文件为单元的元数据来描述，不管是自行积累的还是专门正规建立、正规管理的馆藏都是一样 (Hill, et.al.,1999)。在数字图书馆中，通常说来馆藏支持以下功能：

- 经选择的内容：从可获得的材料中挑选的文献子集
- 组织：应用一致的有序化原则
- 文献记录：依照上下文的、固有的、以及行政管理的元数据
- 存档：长期管理

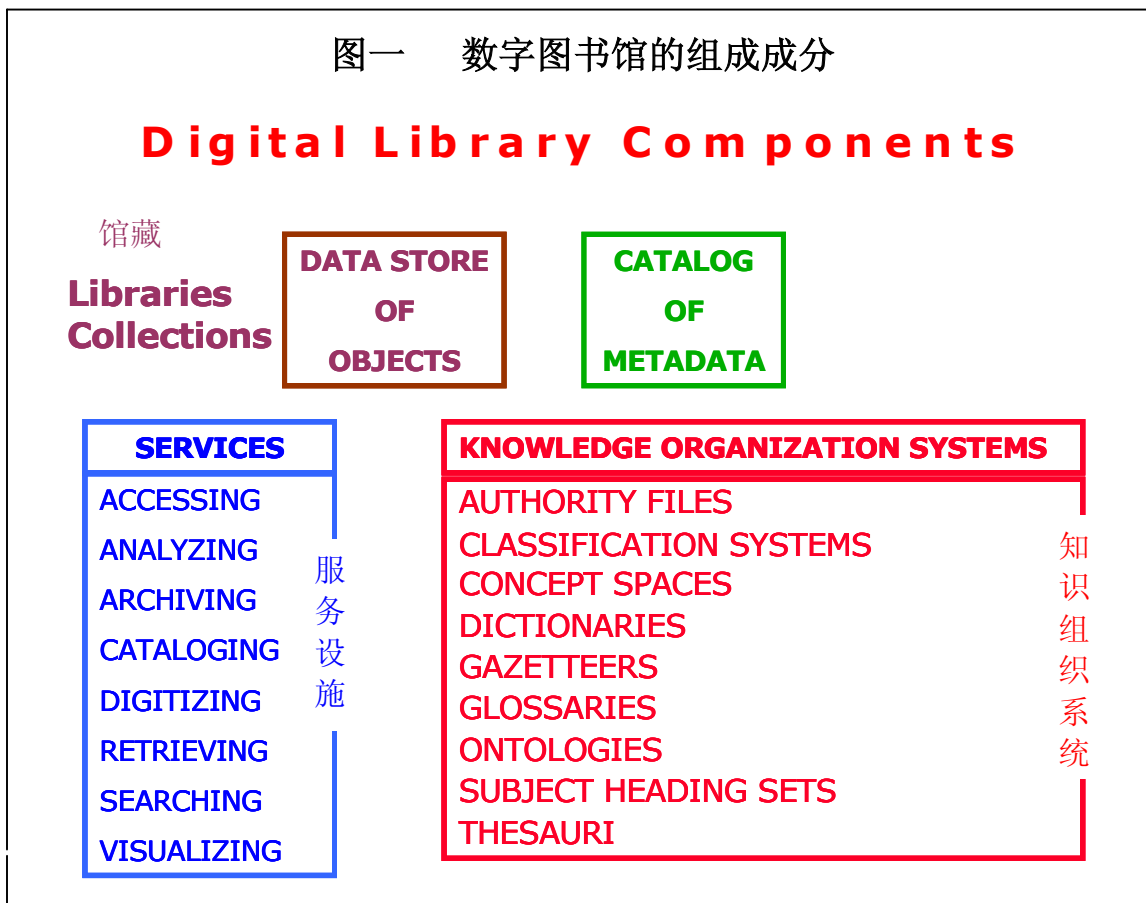
馆藏发展虽也是按特定范围和目的来决定的，但与知识组织系统不同，馆藏基本上是不封口的，总是不断进行。了解一个馆藏的范围可以帮助有效地使用其资源。我们应该想象一个知识组织系统是靠内部填充来扩展的，而一个馆藏是靠不断购进来扩展的。也就是说，知识组织系统在其内部的记录中增加新的条目，而馆藏则不断增加其资源。

数字图书馆的服务设施 (services)是与馆藏、知识组织系统、用户（包括机器和人）相互作用的。一般来说服务设施趋向于模块式，是按特定目的来设计的，设计中都考虑到了网络环境下的工作条件，为便利机器间的通讯而做到与网络标准的兼容，在数字图书馆中它们所支持的功能包括：

- 采购与编目：馆藏建设，元数据编制，以及维护
- 查寻与检索：分布式的查询提问和反馈，查询提问的修订，查询方法等。
- 分析与评价（包括可视化）。

（图一. 数字图书馆的组成成分: 馆藏、服务设施、知识组织系统）

图一 数字图书馆的组成成分



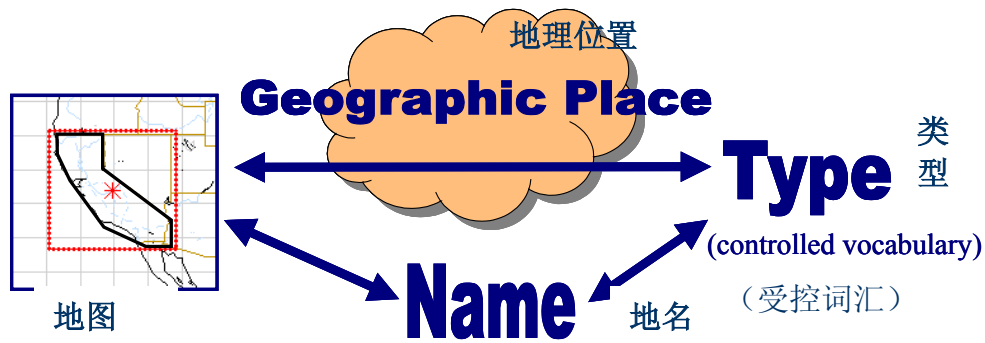
3. 亚历山大数字图书馆项目

加州大学圣巴巴拉分校的亚历山大数字图书馆（Alexandria Digital Library, 简称 ADL)项目的重点在于设计并运行其颇有特色的地理参考数字图书馆，自 1994 年国家自然科学基金会资助立项以来，该工程已经涉及到馆藏、服务设施、知识组织系统各个方面的设计和组建，除了具体馆藏外，已经完成的部分还有：

- 以文件集为单位（相对于以单个文件为单位）著录的元数据结构（Hill,et al., 1999)
- 表述计算机模型的元数据 (Hill, Crosier, Smith, & Goodchild, 2001)
- 地名辞典的设计和运行(*ADL Gazetteer Development Page*)，包括：
 - 地名辞典的内容标准(见 *ADL Gazetteer Development Page*)
 - 地名辞典服务设施协议(Janee & Hill)

- 文本-地理信息的一体化(Frew & Smith, 2001)
- 叙词表编制
 - ADL 地貌特征叙词表(Hill, 2002)
 - ADL 物件类型叙词表(*ADL Object Type Thesaurus*)
 - ADL 词表服务设施协议(Janéé, Ikeda, & Hill, 2002)
- 查询篮 (search bucket) 结构, 为跨越不相似馆藏检索而设计(Janéé & Frew, 2002)
- 概念空间的设计与建制, 用于自然科学教学(Smith, Zeng, & ADEPT Knowledge Team, 2002)
- 地理空间和概念空间的可视化(Ancona & Smith, 2002)

地名辞典一直是亚历山大数字图书馆 (ADL) 中的重要成分, 曾经有一段时期, 地名辞典被当作与其它地理参考资料一样的馆藏处理, 像空中俯视照片、遥感图像、地图等等。但是地名辞典的作用在 ADL 的宏观结构中又是十分特别的, 地名辞典能回答如“贝克斯(地名)在哪?”的问题, 同时又在处理提问中起翻译作用, 例如对“数字图书馆中有什么样的遥感图像上有贝克斯”的问题, 其中“贝克斯”这个地名要被翻译成经纬度, 以便找到涵盖该地区的遥感图。地名辞典的数据可以被强加到地理参考图和地图上, 对图中的地貌进行辨识并加上标识, 这等于对评价地理数据提供了必要的上下文。地名辞典还支持对文本文献的地理切分, 在这种情况下, 一篇文献中关于某地区的内容被表达成经纬度, 以此将这些文本文件转变成适合于数字图书馆的物件, 从这个意义上来说, 地名辞典已成为 ADL 结构中的一个组成部分。(图二)



图二 数字地名辞典的组成部分及相互关系

为了提供方便查找各种不同结构的地名辞典的通用程序，ADL 制定了地理辞典服务协议（ADL Gazetteer Service Protocol），以此支持按照地理名称的基本原则性款项（名称、足迹、类型、关系）进行查询，并且将查到的资料按照标准格式送到用户端，所有这些所要求的就是要安装一个地理辞典服务器，该服务器要能接受特定的 XML 提问式并能发送按特定标准结构做的报告。

叙词表和规范文档在 ADL 的宏观结构中也各有应用。地理辞典的条目就是根据《ADL 地貌特征叙词表》来组织的，ADL 查询篮特别要求在馆藏的物件类型和文件格式描述成分中使用带有等级结构的词集，这样做的目的在于将这些术语用于物件元数据和地理辞典条目的描述内容。知识组织系统的结构既能帮助用户查找所需信息，又在数字图书馆馆藏中做导航工具。为了支持这些功能，本项目又制定了 ADL 叙词表服务协议（ADL Thesaurus Service Protocol），作为自动联网查询各种不同词表的通用协议。就像地理辞典服务协议一样，所有这些所要求的就是要建一个叙词表服务器，该服务器能识别 XML 提问式并向用户端发出按特定标准结构制作的报告。

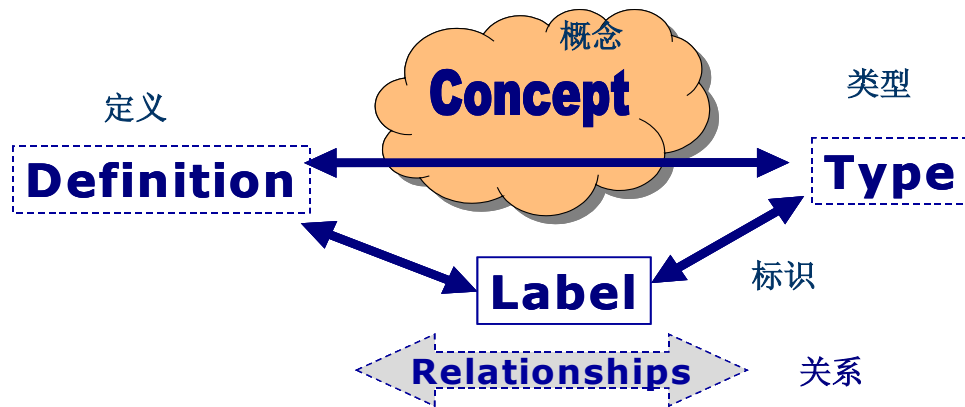
最近的一项亚历山大数字地球模型系统（ADEPT）项目中的一个研究，是建立一个概念空间模型，(Smith, Zeng, & ADEPT Knowledge Team, 2002) 以展示学科领域中的概念及其关系，并以此作为本科生教学中的一个最基本的方式。概念空间模型实际上扩展了叙词表模式，对知识领域中概念集的表达更加全面和完善，这种

模式的意义在于将概念相关关系本身单独作为一组组织成分——其本身就是一个知识组织系统。

这种对将相关关系作为分开的可以定义的知识组织系统的组成成分的认识，在地理辞典内容标准（Gazetteer Content Standard）的制定中也得到体现。地理辞典条目可以通过对绝对的关系的陈述（也可以通过地理参考关系）来将各个条目联系起来，比如说，一个事物（例如，一个县）是另一个事物（例如，一个州）的部分（PartOf）。地理辞典条目之间的关系可以依叙词表的模式来建立，譬如说，建立更详细的属分关系类型。

既然关系可以被做成单独的组成成分模式，那么地理辞典、叙词表、以及概念空间的共同性就一目了然了：它们都是用某种标识（labels）表达概念的，其中有一些为了方便参考而被定为‘选定的（preferred）’形式，表达概念的语词都通过有关特征以及与系统中的其它语词的关系来定义，或者通过指向外部资源来定义。

（图三）



图三 知识组织系统的组成部分以相互关系

另外一个值得注意的问题是地名辞典的表现形式，第一种是以叙词表形式出现的，最明显的例子是 Getty 的《地理名称叙词表（Thesaurus of Geographic Names）》（Getty Information Institute, 1997），第二种是以象元数据一样的单个数据记录的形式出现的。叙词表模式的地名辞典将地理条目之间的关系完全显示在叙词表结构中，（通常按照行政区划的等级），另一方面，ADL 地理辞典的内容标准

和美国联邦政府发行的地理表等都是按照单个条目记录的形式建制的，对每个地方都建有一条记录，地方之间的关系成为记录中的属性。叙词表形式可以很容易地转换成类似元数据的形式，只要将关系当成条目里的属性，而且不要考虑叙词表本身的结构就可以了。

4. 数字图书馆和分类学研究议程

数字图书馆和分类学研究议程中应该探讨将重新铸造的知识组织系统牢固地纳入数字图书馆中的问题，因为各种类型的知识组织系统都有相同的核心功能，所以这样一些研究和发展是完全能够做的：

- 知识组织系统的互操作性以及纳入数字图书馆的服务设施
 - 编制有关知识组织的类型表，对每一类型都指出其内容和功能上怎样与数字图书馆服务协议相互作用和联接。
 - 开发知识组织的登记注册库，编制以知识组织系统为对象的元数据来描述它们。
 - 制定知识组织系统内容的 XML/RDF 表达形式，这种表达形式应该可以针对不同类型的知识组织系统进行改编。
 - 找出不同应用环境中的关于编制和使用知识组织的说明书中的分歧，找出妨碍知识组织系统之间互操作的结构成分。
 - 编制能适用于所有知识组织系统的一套关系模型。
 - 探索知识组织系统在数字图书馆宏观结构和服务设施中的融合问题。
- 数字图书馆服务设施：
 - 制定一个通用的知识组织系统服务协议，从中可以再产生针对各种类型知识组织系统的协议。
 - 开发一个有活力的联接模式，这样使数字图书馆的实体（馆藏、物件、服务设施等）可以指向相应的知识组织系统的实体（概念、标识、关系），二者可以有机的结合起来，支持咨询、翻译、同义词匹配等。
 - 开发能够完全采用和显示知识组织系统中丰富语义的可视工具。

5. 结束语

数字图书馆馆藏、知识组织系统、服务设施需要在数字图书馆环境中互相配合。知识组织系统在馆藏发展、知识发掘和查询、评价、可视化等环节中都有特别的作用，我们需要有一套正式的、统一的关于知识组织系统的定义，一套正式的、统一的方法来识别、查找、参考应用具体的知识组织系统资源，以及一套正式的、统一的知识组织系统使用协议，这些将帮助我们将这些有用的资源融入数字图书馆的大环境中。不同的用户群体所采用的知识组织系统将能够为外界所使用，这是符合日益明显的跨学科专业的情报需求的发展趋势的，那些现存的各自为阵的孤立的知识组织系统则是不符合将之纳入特定的元数据标准和服务协议的大潮流的。

鸣谢

ADL 工程实施和知识组织小组的成员给本文提出了十分有用的意见，特别是 Michael Freeston, James Frew, Terence Smith. 在此特表示感谢。

* 本文根据作者在 2002 年美国情报学会分类法研究研讨会上的发言撰写。原发言稿全文（英文）PDF 版见：<http://alexandria.sdc.ucsb.edu/~gjanee/archive/2002/kos-dl-paper.pdf>

参考资料

- ADL Gazetteer Development Page. Alexandria Digital Library Project. [On-line]. Available: <http://www.alexandria.ucsb.edu/gazetteer> (Last accessed Nov. 8, 2003.)
- ADL Homepage. Alexandria Digital Library Project. [On-line]. Available: <http://www.alexandria.ucsb.edu> (Last accessed Nov. 8, 2003.)
- ADL Object Type Thesaurus. Alexandria Digital Library Project. [On-line]. Available: <http://www.sdc.ucsb.edu/~mary/objtype.tes/index.htm> (Last accessed Nov. 8, 2003.)
- Ancona, D., & Smith, T. R. (2002). Visual explorations for the Alexandria Digital Earth Prototype. Paper presented at the Second international workshop on visual

- interfaces to digital libraries, at the ACM+IEEE Joint Conference on Digital Libraries, July 18th, 2002, Portland, Oregon, USA.
- ANSI/NISO (American National Standards Institute/National Information Standards Organization). (1993, R2003). Z39.19-2003 Guidelines for the construction, format, and management of monolingual thesauri. Bethesda, MD: National Information Standards Organization.
- Arms, W. Y. (2000). Digital libraries. Cambridge, Mass.: MIT Press.
- Borgman, C. L. (2000). From Gutenberg to the global information infrastructure: access to information in the networked world. Cambridge, Mass.: MIT Press.
- Frew, J., & Smith, T. R. (2001). Textual-geospatial integration services for the National SMETE Digital Library. University of California, Santa Barbara. [On-line]. Available: http://www.ehr.nsf.gov/gpra/gpra_award_info.asp?AWD_ID=0121578
- Getty Information Institute. (1997). Thesaurus of Geographic Names. [On-line]. Available: http://www.getty.edu/research/conducting_research/vocabularies/tgn/
- Hill, L. L. (2002). Feature Type Thesaurus. Alexandria Digital Library Project. [On-line]. Available: http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/FTT_metadata.htm
- Hill, L. L., Crosier, S. J., Smith, T. R., & Goodchild, M. F. (2001). A content standard for computational models. D-Lib Magazine, [On-line], 7(6) (June). Available: <http://www.dlib.org/dlib/june01/hill/06hill.html>
- Hill, L. L. (2000). Core elements of digital gazetteers: Placenames, categories, and footprints. In J. Borbinha & T. Baker (Eds.), Research and Advanced Technology for Digital Libraries: Proceedings of the 4th European Conference, (ECDL 2000) Lisbon, Portugal, September 18-20, 2000. (pp. 280-290). Berlin: Springer.
- Hill, L. L., Dolin, R., Frew, J., Janée, G., & Larsgaard, M. (1999). Collection metadata solutions for digital library applications. Journal of the American Society for Information Science, 50(13), 1169-1181.
- Hodge, G. (2000). Systems of knowledge organization for digital libraries: Beyond traditional authority files: Council on Library and Information Resources. [On-line]. Available: <http://www.clir.org/pubs/reports/pub91/contents.html>.

- International Organization for Standardization (ISO). (1986). Documentation - Guidelines for the establishment and development of monolingual thesauri (ISO 2788:1986).
- Janée, G., & Frew, J. (2002). The ADEPT digital ILibrary architecture. Paper presented at ACM+IEEE Joint Conference on Digital Libraries, July 18th, 2002, Portland, Oregon, USA. [On-line]. Available: <http://alexandria.sdc.ucsb.edu/~gjanee/archive/2002/jcdl-adept.doc>.
- Janée, G., & Hill, L. L. Gazetteer Service Protocol. Alexandria Digital Library Project. [On-line]. Available: www.alexandria.ucsb.edu/gazetteer/protocol (Last accessed Nov. 8, 2003.)
- Janée, G., Ikeda, S., & Hill, L. L. (2002). The ADL Thesaurus Protocol. Alexandria Digital Library Project. [On-line]. Available: <http://www.alexandria.ucsb.edu/thesaurus/protocol/>
- NKOS (Networked Knowledge Organization Systems/Services). Webpage. [On-line]. Available: <http://nkos.slis.kent.edu/> (Last accessed Nov. 8, 2003.)
- Smith, T. R., Zeng, M. L., & ADEPT Knowledge Team. (2002). Structured models of scientific concepts for organizing, accessing, and using learning materials. In Lopez-Huertas, et al. (Eds.): Challenges in knowledge representation and organization for the 21st century. Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain (pp.232-239).