网络知识组织系统的研究现状和发展趋势

王军*、张丽
（北京大学信息管理系，北京，100871）

文摘：网络知识组织系统（NKOS）是网络环境下知识组织体系的电子化描述，包括术语表、地名录、分类法、主题词表、本体等不同类型和层次的知识组织系统。NKOS 是随着传统的知识组织工具在网络环境下的发展和演变，而逐渐形成的一个新的研究领域。NKOS 的发展将为图书馆知识组织的传统带来生机，推动网络环境下的信息组织、知识表示、基于内容的信息检索等应用的发展。本文全面综述了 NKOS 这一领域的研究现状，包括定义与类型、发展历程、标准与规范、生成与维护、现有的应用和未来发展。以此为国内 NKOS 和数字图书馆的研究和建设提供一定的参考。

关键词：网络知识组织系统　NKOS　语义工具　本体　综述

An Overview on the Networked Knowledge Organization System

Jun Wang, Li Zhang
Dept. of Information Management, Peking University

ABSTRACT: The networked knowledge organization system (NKOS) is the representation of knowledge organization system (kos) in digitized and networked circumstances, and it includes a rich variety of controlled vocabularies with different complexity, such as term lists, gazetteers, classification schemes, subject headings and ontologies. The NKOS has emerged as a new and hot research area in information science. It is believed that the NKOS will intrigue new development of the knowledge organization tradition in libraries, and will facilitate the application of information organization, knowledge representation and content-based information retrieval. This paper surveys this hopeful domains, and introduces the definition, variety, conversion, standard, creation, maintenance and application of NKOS.

KEYWORD: networked knowledge organization system, semantic tools, ontology

* 王军，副教授，北京大学信息管理系。HTTP://KVision.pku.edu.cn

1. 引言

知识组织，是在图书馆领域形成的、管理和利用信息资源的传统方式和基本手段，是图书馆工作的重要基础。但是随着人类信息环境向网络平台的迁移，数字资源成为主流，搜索成为泛在的工具，传统的知识组织理论、方法和工具因囿于图书馆的范围、未能在网络信息环境下得以普遍应用而受到质疑，进而动摇了图书馆的存在基础。这给我们带来了两个问题，一是对搜索的过分依赖导致了网络信息资源利用的单一化；二是图书馆内井井有条的馆藏与外部泛滥无序的网络信息之间泾渭分明的界线无形中似一道壁垒，把图书馆变成了信息海洋中的孤岛。网络知识组织系统(NKOS)，既是在这一背景下出现的试图在网络环境下重振知识组织传统而付诸的努力。

NKOS 的出现，是由两个方面的共同作用所致：一、随着人类信息活动由纸制环境向数字环境迁移，传统知识组织系统(KOS)的数字化、网络化势在必行；二、网络环境下信息量急剧增长，需要相应的语义工具实施对网络信息资源的组织，改进检索性能，实现对网络资源的深度挖掘和智能利用。这两个方面分别来自于图书馆界和网络社区，其中后者是 NKOS 的主要应用方向和推动力，目前较为成熟的 NKOS 技术多数都与语义网（Semantic Web）计划有着密切的关系。语义网旨在将现有网络发展成为一个数据交换与集成、知识化利用与管理的基础环境，其架构需要新的信息组织机制的支持。

本文所讨论的网络知识组织系统(NKOS，Networked Knowledge Organization System)，是指**应用于网络环境下的、对知识结构进行系统化描述、解释和说明的、用于支持网络信息的表示与检索等活动的知识组织系统**。NKOS 有两种类型：一是传统的知识组织系统（KOS）在网络环境下的延伸和发展，如采用美国国会分类法(LCC)的十个基本大类组织网络资源的 CyberStacks、美国国会主题词表(LCSH)的网络应用 INFOMINE、亚历山大数字图书馆地名录（ADL Gazetteer)；二是那些在网络环境中产生和成熟起来的语义工具，如网页目录(如 Yahoo! Directory)、本体、语义网络(如 WordNet)等。NKOS 的最终目标是**实现机器可理解的、可应用的知识体系描述**。在这一终极目标和现实应用之间，有不同层次、不同成熟程度、不同技术框架下的知识组织工具。很难在 NKOS 和 KOS 之间划出一个泾渭分明的界限。

NKOS 已经逐渐成为信息科学领域一个新的、重要的研究领域，一些有代表性的国际会议成立了 NKOS 的专门工作组。国际数字图书馆联合会议(JCDL)已经召开过 7 次 NKOS 的研讨会；欧洲数字图书馆会议(ECDL)举办过 4 次；都柏林和元数据国际会议 2003 和 2005(DC-2003, DC-2005)也分别设立了 NKOS 的相关专题。国内也出现了一些相关的研究，涉及到的问题包括：本体的描述及其在检索中的

应用[1,2]、不同 NKOS 间的转换与映射等[3]。综合目前国际、国内的研究现状，NKOS 研究领域涉及如下几个方面：

1) NKOS 的表示：以结构化或半结构化的方式存储，以机器可理解的形式描述，是 NKOS 区别于 KOS 的一个特征，是 NKOS 在网络环境下自动应用的基础。
2) NKOS 的互操作：不同组织结构、不同领域、为不同应用开发的 NKOS 在表示、结构、内容等方面有很大的差别。在这些异构的 NKOS 之前实现交换、共享和集成，是降低 NKOS 开发成本、促进 NKOS 应用要解决的关键问题。
3) 标准化问题:标准化贯穿于 NKOS 的各个方面,NISO(美国国家信息标准组织)、W3C 等国际组织已经发布了制定、表示、使用 NKOS 的一些标准。
4) NKOS 的生成和维护：包括传统 KOS 的改造、NKOS 的演化和 NKOS 的自动更新与维护等。
5) NKOS 的应用：应用是 NKOS 研究的最终目的。支持网络信息资源的知识组织和智能信息检索是 NKOS 的主要功能。

本文试图从以上几个方面对 NKOS 目前的研究现状作全面综述，分析当前存在的问题和未来的发展趋向，以推动国内 NKOS 的研究和应用。

2. NKOS 的类型和表示

2.1 NKOS 的类型

根据语言的受控程度和结构化程度，Gail Hodge 将 NKOS 分为术语表、分类法和词汇关系网三类[4]。术语表是对术语的定义和解释，一般是线性结构的，例如规范档（Authority Files)、专用名词词典（Glossary)、字典（Dictionary)、地名词表（Gazetteer）等；分类法则注重主题/学科集合的形成，对术语之间关系的揭示重点在于等级关系，一般是树状结构的，例如主题词表、分类法、专类分类表（Taxonomy)、类目系统（Categorization Schema）等；词汇关系网对术语之间关系的揭示更复杂、更细致，除了传统词表中的“用、代、属、分、参”，还可以有整体-部分、蕴含、因果等关系，呈现出网状结构，如词汇数据库(lexical database)、语义网络（Semantic Network）和本体。图 1 是 Zeng 和 Salaba 根据 NKOS 的结构化和语言规范程度给出的现存各类 NKOS 的分布空间[5]。纵轴也表示了描述能力、推理性能和机器可处理能力从底向上的增强。

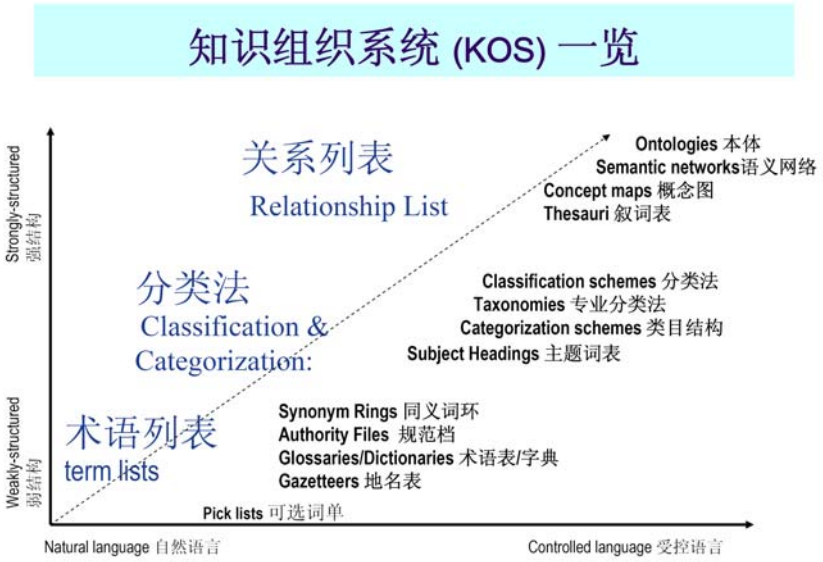

图 1. KOS 类型分布图

2.2 NKOS 的表示

将 NKOS 所描述的概念、概念间的关系和知识结构以机器可理解的形式表示出来是 NKOS 在网络下应用要解决的首要问题。KOS 在网络环境下的表示基本上可以划分为三个阶段：1）KOS 的电子化；2）HTML 表示的 KOS；3）用语义万维网（Semantic Web）的相关技术表示 KOS，例如 XML、RDF、OWL 以及 W3C 最新推出的 SKOS。

1） **KOS 的电子化**：NKOS 发展的前期阶段是 KOS 的电子化，包括 KOS 的 MARC 描述和数据库化。用数据库存储和表示便利了对 KOS 的管理和访问，也便于将它们与相应的电子资源集成在一起。例如《中国分类主题词表》的电子版就是用数据库技术来制作的，在 INSPEC（英国科学文摘）和 EI（工程索引）数据库中分别集成了 INSPEC 词表和 EI 词表，以方便查询词的选取、扩检和缩检等操作。MARC 格式是电子化管理、发布类表、词表、人名、机构名等规范档的标准方式。例如 LCSH、LCC 都提供 MARC 版本并可以在网上查询。用 MARC 格式表示的 KOS 可以植入 OPAC 系统中与书目数据统一管理。

2） **基于 HTML 的 KOS**：NKOS 的先声是在网络环境下呈现传统的知识组织工具。随着万维网的普及应用，众多的 KOS 都把 Web 作为展示的窗口，通过 HTML 网页提供基本的浏览和查询功能。这也是目前 NKOS 在网络上表现的主要方式，澳洲学者 Middleton 收集了百余个在网络上可以访问的词表和类表[6]，包括 AAT(艺术与建筑词表)、NASA（美国航空航天局词表)、ERIC（教育资源信息词表)、GEMET(通用多语种环境词表)、MeSH(美国医学主题词表)、MSC(数

学主题分类)、EI、LCC、UDC 等。HTML 是一种描述网页显示格式和布局的语言。KOS 的 HTML 表示，相当于 KOS 在网络上的翻版，不同 KOS 在体例上、结构上、内容上的异构性依然存在，也不便于计算机的自动处理和利用。

3) **基于语义网技术表示 NKOS**：在语义网框架下发展出来一系列的语义描述语言，包括描述结构的 XML、表达语义的 RDF 和表示本体的 OWL 等，其目的是实现机器可理解的信息描述。这些工具被用来表示 KOS 标志着 NKOS 的真正产生。应用 XML 描述大型词表的例子有 DDC(杜威十进制分类法)和 MeSH。OCLC 提供的 DDC 网络版就是用 XML 实现的；NLM（美国医学国家图书馆）制作的 MeSH 的全部 XML 文件可以从 NLM 网站自由下载。一些学者也致力于 KOS 的 XML 表示的规范化，Voc-ML 就是这样的一个规范草案，它用元数据描述 KOS 的基本信息，用 DTD 定义 KOS 的结构。

RDF(S)比 XML 更进一步，能够描述资源（概念可被视为一类特殊资源）并表达概念之间的复杂关系。采用 RDF 描述 KOS 的例子有欧洲中心实验室委员会 CCLRC 支持的 LIMBER[7]项目中制定的《语义网下的词表交换格式》标准，联合国粮农组织（FAO）用 RDF 表示的多语言词表 AGROVOC。

在语义网框架下的系列语言中，最复杂的是用于描述本体的 OWL，它内嵌了描述逻辑的功能，能表达逻辑并进行推理。本体是网络环境下一种新型的知识系统，用于支持智能代理进行资源组织、智能查询、知识发现等活动。OWL 提供的丰富的描述逻辑用于表示非本体类型的 NKOS，有些过于复杂。为了在语义网框架下简明地表示和使用各类简单概念系统，W3C 于 2005 年发布了简约知识组织系统表述语言（SKOS）标准草案[8]，它以 RDF 为基础，是一种描述 KOS 基本结构和内容的语义标记语言。NKOS 的最终目标，是提供机器可理解/利用的 KOS，以推动基于自动知识组织的智能信息服务。SKOS 的出现标志着 NKOS 朝着这个方向迈出了关键的一步。目前已经采纳 SKOS 描述的词表有：欧盟开发的通用多语种环境词表（GEMET）、大英档案词表（UKAT）、联合国粮农组织词表（AGROVOC）等。

3. NKOS 的互操作

NKOS 互操作主要解决两个问题：多语言和异构。跨语言的互操作问题在欧洲很受重视，这源于它的多语言、多文化的背景，相关的研究项目有 MACS、Merimee、Renardus 等。我国也有一些关于双语 NKOS 互操作的研究，例如《汉语主题词表》与 LCSH 之间的转换研究[9]、《中图法》与 DDC 类目设置比较[10]、《中图法》与 DDC 对照系统的研制等[11]。国外也有学者尝试自动建立中、英文词表之间的映射[12]。异构 NKOS 间互操作的目的是实现不同系统间的知识交换、共享和重用。例如：

OCLC 通过人工和统计等方法建立 LCSH 主题词与 DDC 类号之间的映射；Renardus 以 DDC 为映射中心实现多个不同语言、异构的 NKOS 间的映射[13]；利用计算机建立分类法与主题词表转换系统的可行性研究等[14]。

3.1 互操作的实现方式

Marcia Lei Zeng 和 Lois Mai Chan 两位学者总结了 KOS 互操作的八种实现方式[15]：

- 继承/仿建：以现有的复杂的词表为原型，创建专业的、或简单的词表。
- 翻译/改编：从其他语言的词表翻译、改编形成自己的词表。许多非英语主题标目都是从 LCSH 翻译并发展而来的。
- 卫星子表：对现有词表的某个主题进行扩展，形成的新的子表称为原表的卫星。通常是对一个通用的 KOS 的某个学科在其基础上添加相对专业化、具体化的术语，并细化术语间的关系，形成针对专业领域的比较详尽的 KOS。
- 直接映射:直接在不同 KOS 的词语之间或者词语与分类号之间建立等价关系。
- 共现映射：通过 KOS 词语在元数据记录中的共现关系建立术语间的映射。
- 中心转换：将参与互操作的多个 KOS 映射到一个共同选定的中心 KOS 上。这样，两个 KOS 之间的互操作可以通过中心 KOS 的转换实现。
- 临时列表：根据查询词，临时从不同的 KOS 提取相匹配的对象，组建临时对应列表。这种映射方法是一种散点式的映射方案。
- 协议连接：通过建立 KOS 服务协议供其它应用程序访问，创建连接环境，实现 KOS 的互操作。

在这八种方式中，卫星子表实际上也可归入继承/仿建方式——以原 KOS 的一部分为基础的扩展的继承。以这种方式实现互操作的应用，通常是以一个跨学科的通用 KOS 中与相应学科领域相关的部分作为原型，在其基础上添加相对专业化、具体化的术语，并细化术语间的关系，形成针对专业领域的比较详尽的 KOS。例如，LIV (Legislative Indexing Vocabulary)就是对 LCSH 中立法相关的部分进行扩展而创建的。而对于共现映射，我们并不认为它是一种与直接映射和转换并列的映射方式。事实上，无论是在直接映射还是在间接转换方式中，具体到每两个 KOS 之间的映射，都可以采用基于 KOS 自身的体系结构或基于元数据中共现关系建立映射两种方法。基于以上认识，我们将 KOS 间互操作的实现方式归纳为演化、映射、协议和临时连接四种，如图 2 所示：

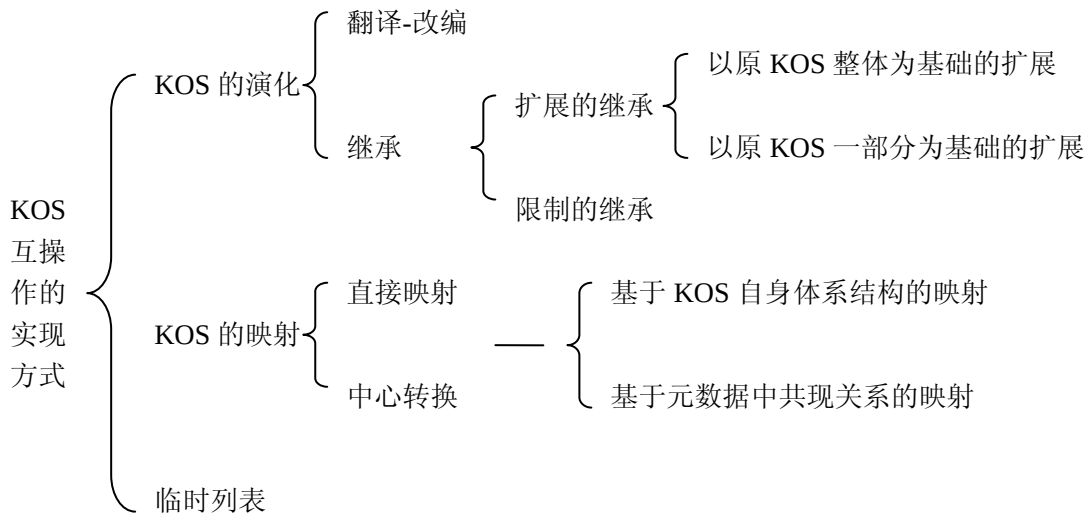

图 2.　KOS 互操作实现方式分类

虽然 KOS 的演化并不以互操作为目的，但它为满足特定需求对原有 KOS 进行改造，新建的 KOS 与原有 KOS 之间形成了对应关系，客观上支持了互操作。对于独立创建的 KOS，映射和服务协议是实现 KOS 互操作的主要方式。在参与互操作的 KOS 比较明确时，例如在特定的几个机构间进行信息共享，映射方式比较适用；而在参与互操作的 KOS 并不明确时，例如 KOS 的拥有者只是希望提供一种知识组织服务，并不明确自身的 KOS 要与哪些 KOS 进行互操作，协议方式较为合适。临时联合列表是基于对查询提问的字面匹配的，互操作的效率和准确性不是很高，但实现起来比较简单。可见，各种互操作方式有其各自的特点和适用范围，在具体的信息资源共享活动中需要从实际出发选择合适的方式。各种互操作方式的比较如表 1 所示。

表 1.　KOS 互操作实现方式比较

实现方式	**KOS 演化**	**KOS 映射**	**服务协议**	**临时联合列表**
以实现互操作为目的	否	是	是	是
参与的 NKOS 是否独立生成	否	是	是	是
是否保存映射或连接	是	是	是	否
参与的 NKOS 个数是否固定	是	是	否	否
实现的自动化程度	手工	人机	协议	机器
适用范围	KOS 的改造和利用	互操作	知识组织服务	简单联合检索

4. 相关的标准与规范

1. **Z39.19 和 BS8723**：经过较长的发展历史，创建和发展传统的 KOS 有很成熟的标准和规范可以参考。例如《ISO 2788：单语种词表的创建和开发》,《ISO 5964：多语种词表的创建和开发》,《NISO Z39.19：单语种控制词汇的创建、格式和管理》,《BS 5723：创建和开发单语种词表指南》等。这些标准大多制定于上世纪 80-90 年代。进入本世纪，要求制定数字环境下相关标准的呼声越来越高。

   1999 年 NISO 组织了一次全美的研讨会讨论开发电子词表标准的问题，并形成了推荐草案[16]。以此为基础，2005 年 NISO 发布了 Z39.19 第四版[17]。Z39.19 是关于词汇控制工具最主要的标准。它提供了单语种词汇控制工具（包括同义词环、专类类表和词表等）的内容、显示、构建、维护和管理等方面的原则和规范，充分考虑了标引非传统纸质文献的要求，如专利、化学分子结构、地图、多媒体资料等。也提出了在网络环境下的显示要求。BS 5723 是英国制定的关于单语言词表的标准，发表于 1987 年。2005 年发布的《BS 8723：用于信息检索的结构化词汇》全面取代了 BS 5723。BS 8723 包括 5 个部分，第 1、2 部分是关于词表的，第 3、4、5 部分还在制定中，它们是关于分类法、商业分类法、本体等其他控制词汇的，以及不同的控制词汇工具间的互操作、协议、格式等问题。BS8723 提供了对电子词表的功能设计、词表管理软件、在电子环境下的显示和分面分析等诸多规范和建议。

2. **Zthes**：Zthes 是一个基于 XML 的规范，用来描述和传输词表、类表等 KOS。Zthes 的主要目的是实现和 KOS 相关的应用间的互操作性。Zthes 的核心规范是一个描述词表词汇的抽象模型，它表示了一个词汇集主要的三部分：关于词汇集的信息、词汇的信息、词汇间的关系。基于该模型，对符合 Zthes 规范要求的词表的访问和查询有 SRU、SOPA 和 Z39.50 三种方式。SRU (Search-retrieve by URL)定义了一种将查询转换成 URL 的方式，通过 HTTP 协议向对远程词表数据库提交查询，结果包装在 XML 中返回。Zthes 还定义了一个词表的查询语言 CQL。Zthes 试图提出一整套描述、访问、查询词表的标准方法，但是 Zthes 是基于 ISO 2788 制定的，而形成于 1987 年的 ISO 2788 已经不能适应电子环境下词表的要求，这使得 Zthes 的应用有限。

3. **SKOS**：上述标准为词表的电子化和网络环境下的应用起到了积极的推动作用。但是，这些标准还是没有脱离开传统词表的模式，也未能为不同 KOS 的共享和交互提供一个解决方案。W3C 在参考了多种现存的 KOS 标准后于 2004

年发布 SKOS 推荐标准，它是一个基于语义网技术表示受控词表及其它知识工具的概念框架。SKOS 目前仍处在发展阶段，它具有简洁、通用、易扩展、与语义网及传统图书情报领域联系紧密的特点，在促进受控词表在网络环境下的使用等方面具有非常重要的意义。有趣的是，SKOS 是欧洲学者提出的，他们大力研究并积极推广；但是北美学者对此似乎反应冷淡。这不知是由于学术背景的差异还是欧美的学术对峙。

SKOS 包括三个主要部分：核心集(SKOS Core)、映射(SKOS Mapping)和扩展(SKOS Extensions)。SKOS 核心集是一个表示概念体系基本结构和内容的模型，这里的概念体系是指“一个概念的集合，(不)包括概念间的语义联系”。SKOS 映射用于描述概念间的映射，SKOS 扩展用于描述 SKOS 的特定应用。SKOS 核心集基本发展成熟，已形成了相应的语法标准和应用标准，后两者目前还处于研发阶段。SKOS 核心集由 31 个词汇构成，其中绝大多数是由 RDFS 定义的，重要的有：描述概念的 Concept、prefLabel/altLabel；描述概念间关系的 broader、narrower、related；描述概念体系的 ConceptScheme；描述标引关系的 SubjectOf 等。

鉴于 W3C 在网络语言标准化方面的权威地位，SKOS 十分值得重视。事实上，尽管还是一个推荐标准，SKOS 已经被用于若干大型词表的表示了，包括欧洲的多语言环境词表 GEMET、英国的档案词表 UKAT、澳大利亚公共事务信息服务词表 APAIS 等。北京大学信息管理系 KVision 研究小组采用 SKOS 描述了中国分类主题词表的一个片断，并基于此实现了一个语义检索系统[18]。

5. NKOS 的生成和维护

传统知识组织系统的生成和维护有赖于专家的手工劳动，本文第 4 节介绍了相关的标准和规范。网络环境下，手工更新的方式远远滞后于爆炸性的信息增长和知识更新。自动的生成与更新是 NKOS 发展和应用的一个关键所在。

1. **词表的自动生成：**从上世纪 70 年代起，就有学者探索从全文语料中自动构建词表，并测试它们对信息检索的作用。其中包括信息检索领域著名的学者 K. Sparck Jones[19], Gerard Salton[20]等。Carloyn Crouch[21]和 Hsinchun Chen[22]在这一领域的工作也比较重要。本质上所有的词表自动构造技术都以对文中词汇的同现统计分析为基础，这样构造的词表中词汇间仅有相关关系，其实质是一些在给定的文献集合中具有某种同现模式的相关词汇的集合，主要目的是为了提高检全率。随着 Web 的发展和网络信息资源的激增，检准是人们关心的主要问题。不少搜索引擎（如 Google、Yahoo!、百度）通

过挖掘用户日志，分析词汇的构成和相互间的包含关系，建立相关词汇列表，以帮助用户明确检索需求并挑选合适的检索词。严格说来，上述这些自动构造的词汇工具还称不上词表，词表是以等级关系为主干的。

要全自动地生成像词表这样的需要高智力投入的知识产品，自然语言处理和人工智能技术要达到一个非常的水平才有可能。比较现实的途径是对现有的词表自动或半自动地丰富。王军在中国分类主题词表和美国国会主题词表上进行了深入的研究，提出：从已标引语料（如书目数据）中挖掘新词，并通过分析它们和标引词间的关系，确定新词所对应的规范词，最终将新词作为对应规范词的下位词添加到词表中。这一方法的实质是挖掘已标引语料中的领域知识和标引知识，并建立它们之间的对应关系[23,24,25]。大量的试验证明这一方法是有效的、可行的，这些研究成果在网络上提供公开评测[26]。

2. **传统分类法的改造：**传统分类法（如 DDC、UDC、LCC）一般都具有几万个类目，深度达十余层。采用这些分类体系实现自动分类，最大的困难是稀疏数据和错误传播。稀疏数据是指单个类下已分类样本太少，不够分类器学习该类的特征；错误传播是指上层的分类错误向下传播，过深的类层次使得底层类目的分准率太低。王军以美国国会图书馆十年的书目数据为样本，根据书目记录在 DDC 类体系中的分布特性，通过收缩、合并、截枝等操作重构 DDC，使之易于机器学习[27]。并引入有限次数的人工辅助判断，以实现一个可应用于实际分类工作的 DDC 交互分类系统[28]。

3. **从词表向本体的演化：**Soergel 等学者尝试将 AGROVOC 转化一个 Ontology[29]。遵循 UMLS 的设计方法，他们在词表的基础上设计了一个三层概念模型，包括概念、术语和词串。这样做一方面能继承词表中丰富的词汇，另一方面提高了等级结构的结构化程度，并添加更细致的关系，以满足概念推理的需要。受该项目启发，联合国粮农组织已经开始着手实施了一个“农业本体服务/概念服务”[30]。Wielinga 等人在 AAT 词表上也作了类似的工作[31]。王军提出将分类主题词表和元数据集成在一起（相当于元数据上架），构造一个知识网络并在其上实施知识浏览和概念检索[32]。其不足之处在于知识网络的构建方法不够形式化，缺乏推广性。近期，他们基于 DC 元素集形成一个书目数据本体，其中采用 SKOS 表示分类主题构架。本体用 OWL 描述并装入 RDF 数据库 Sesame 中进行管理。以其为后台服务，可提供高级语义查询[18]。

6. NKOS 的应用

1. **术语服务**：通过 Web 服务（Web Service）技术在网络上提供分布式的词汇服务是目前 NKOS 服务的一种主要形式,Web 服务是一个基于 HTTP 协议和 XML 的协议标准。已提供这类服务的词表有：AGROVOC、AAT、CSA/NBII 生物复杂性词表(Biocomplexity Thesaurus)、美国国家农业词表（NAL）、亚历山大数字图书馆项目中的 ADL 地名表协议等。

   欧盟语义网计划中的 SWAD-Europe 词表工作组基于 SKOS 设计了 SKOS API，方便了开发基于 SKOS 的网络应用程序和访问用 SKOS 表示的各类知识组织系统。SWAD-Europe 还特意开发了基于 SKOS API 的客户端和服务器，以展示 SKOS API 的功用。

2. **术语映射服务**：OCLC 术语服务项目的目标是通过开放标准和各类 Web 协议向 Web 应用提供对 NKOS 开放的、模块化的词汇服务，包括规范档、主题词表、叙词表、分类法等。目前提供的是不同 KOS 间的词汇映射。有两种：一是直接映射，这是在来自不同的 KOS、具有等价关系的词汇间进行的，包括叙词、主题词、甚至类号；二是共现映射，这是通过分析在同一元数据记录中同时出现的来自不同 KOS 的标引词汇实现的。目前，该项目已经在 LCC、DDC、LCSH、MeSH、ERIC 等 KOS 间进行了试验。映射的结果还可以通过 OAI 协议来访问和下载。

3. **检索辅助**：信息检索是 NKOS 的关键应用之一。在此举两例说明：跨语言检索和分面检索。多数的跨语言检索系统是建立在对同一资源的多语标引之上。通过 NKOS 的互操作可以避免这一限制。例如 Renardus 项目以 DDC 为中心转换语言，通过 KOS 之间的映射，将其他语言表示的检索式转换为 DDC 表示的检索式，以此实现跨语言检索。FACET 系统使用分面叙词表来辅助用户构造检索式[33]。它将词表集成到系统的检索界面和搜索功能中，减少了用户在选择检索词时的负担。利用分面提供的上下文知识，来自动/交互地扩展或精化词汇。在分面查询编辑器中，用户可以根据词表中分面的语义作用生成复杂的查询。FACET 采用 AAT 词表，以美国科学与工业博物馆的数字馆藏为实例，提供比关键词查找更广泛更深入的访问。

   KVision 是北京大学信息管理系研发的一个语义检索系统[18]。KVision 将《中国分类主题词表》计算机应用(TP39)部分用 SKOS 表示，并与千余条用 DC 表示的书目数据集成在一起，创建了一个领域本体，并基于此提供语义检索和概念浏览的功能。KVision 还将 NKOS 和搜索引擎结合起来，实施

Web 资源查询结果的分类/聚类。展示了 NKOS 在检索方面的应用。

4. **词汇注册**：“分类法仓库”(Taxonomy Warehouse)站点展示了控制词汇的另一个应用。该站点提供 KOS 的注册服务，任何机构都可以将自己创建的 KOS 提交到该站点注册。目前在该站点登记在册的 KOS 达 660 之多，这些 KOS 由 261 个不同机构创建，归类在 73 个主题领域下，涉及到 39 种语言。其中 65% 的 KOS 是以数字方式存储的。类似的项目还有：Becta Terminology Studio, HILT Terminology Service, XMDR Extended Metadata Registry, NSDL Metadata Registry 等。词汇注册服务可以作为数字图书馆体系结构中的一个关键组件来实现，它的主要功能有：登记和管理创建者提交的各类 NKOS；发布和发现关于术语的信息；证实术语的真实性和状态；发现术语间的关系；支持推理、映射等功能；提供对相关资源的导航；促进不同控制词汇系统间的互操作等。词汇注册服务要求采用开放标准和通用结构（如 Zthes, SKOS, MARC 等）描述登记在案的 KOS。它还可以提供编程接口，同时向用户和职能代理提供服务。

7. 结论

NKOS 代表了 KOS 的发展趋势：数字化、网络化、语义化、协议化和自动化。NKOS 的发展已经超越了“KOS 在网络环境中的应用”这一初级阶段，成为网络环境下新生的“语义工具”。NKOS 将为图书馆知识组织的传统带来生机。基于内容的检索是 Web 信息检索的必然发展趋势，从当前的关键词查找发展到内容检索，这其中 NKOS 的作用十分关键。可以预见，NKOS 将是数字图书馆领域的一个发展热点，它也是图书馆领域的从业人员在 Web 资源利用的战场上能有所作为的一个阵地。从第五届欧洲数字图书馆会议(ECDL)上召开的 NKOS 分会，可以看出未来 NKOS 研究和应用的若干关键方向是：

1) NKOS 的表示和服务协议：制定和推行基于语义网技术的 NKOS 表示与服务标准、确定 Web/网格环境下提供 NKOS 服务的粒度、NKOS 的可扩展性和持续性管理等。
2) NKOS 的互操作：支持跨系统浏览和检索的分布式 NKOS 服务，术语间的映射，从 NKOS 到 Ontology 的映射与演化，语义互操作的实现，与元数据的集成应用等。
3) 随着 Web2.0 的兴起，参与式的、基于用户的知识组织和标引方式得到越来越多的重视。社会化标签、大众分类法等非正式知识组织结构对 KOS 的影响及其作用。

4） 其它的问题包括： NKOS 的可视化、NKOS 在数字图书馆体系结构中的集成、NKOS 的自动更新与维护、NKOS 和搜索引擎的结合等。

信息组织和信息检索是信息利用与服务的两把利刃。从历史来看，搜索技术大行其道之前，信息组织一直为人们所倚重。随着搜索时代的到来，信息组织失去了往日的光彩。在 Web 上，Yahoo!最终放弃了对 Yahoo! Directory 的大规模和系统化维护；在图书馆，使用 OPAC 系统的读者很少利用分类体系和主题标引进行查询。搜索技术一统天下，一方面便利了对网络资源的利用，另一方面，它给网民带来的快餐式信息消费习惯，把网络推向了信息大集市的发展方向。一个成熟的信息环境更需要纵深的发展，需要对网络信息资源的过滤和整理、对信息资源的系统性和整体性把握、对信息对象间联系的揭示、需要提供基于内容的处理和更精确更深入的服务。这些需求都呼唤知识组织功能重整旗鼓。但是，企图仿照传统图书馆的模式，创建一个统一的井然有序的分类帝国，将所有的信息资源装入一个知识体系，在今天张扬多元化个性化的社会文化背景下，是不合时宜的。这是图书馆业界在面对网络信息资源组织时，常常误入的一个死胡同。在 NKOS 领域呈现出来的实用性、多样化、形式化、自动化、与搜索技术相结合的特点，给我们很多启发。另外，为 Web2.0 环境下的个人信息资源组织服务的语义工具，也将是一个需要关注和深入研究的问题。

参考文献

[1] 曹树金,马利霞. 论本体与本体语言及其在信息检索领域的应用. 情报理论与实践, 2004(6).

[2] 徐海涛, 汪方胜, 蒋馥. 基于本体的信息检索机制研究. 情报杂志, 2005(10).

[3] 程鹏. 《中图法》和《杜威法》对照系统的研制. 图书馆学研究, 2001(02): 55-57.

[4] Gail Hodge, Systems of knowledge organization for digital libraries, DLF, 2000

[5] Marcia Lei Zeng & Athena Salaba, Toward an international sharing and use of subject authority data, FRBR Workshop, OCLC 2005

[6] Controlled Vocabularies, http://www.imresources.fit.qut.edu.au/vocab/

[7] Language Independent Metadata Browsing of European Resources，http://www.limber.rl.ac.uk/

[8] Simple Knowledge Organization System，http://www.w3.org/2004/02/skos/

[9] 张广钦, 吴辉. 两种受控语言间兼容转换问题研究.情报理论与实践. 1997(04): 239-243

[10] 黄燕芳. DDC20 版与《中图法》第 3 版计算机类目设置的比较研究. 广西大学学报(哲学社会科学版) . 1997(05): 105-108

[11] 程鹏. “《中图法》和《杜威法》对照系统”的研制. 图书馆学研究, 2001(02): 55-57.

[12] C. C. Yang and K. W. Li. Automatic Construction of English/Chinese Parallel Corpora. Journal of the American Society for Information Science and Technology, vol.54, no.8, June, 2003, pp.730-742

[13] T Koch, H Neuroth. Renardus: Cross-browsing European Subject Gateways via a common

classification system. IFLA Satellite Conference "Subject retrieval in a networked world", 2001

[14] 李波，戴秀梅，侯汉清．计算机建立分类法和主题词表转换系统的尝试．现代情报 2003(06): 112-115

[15] Zeng, M. L. & Lois Mai Chan. 2004. Trends and issues in establishing interoperability among knowledge organization systems. Journal of American Society for Information Science and Technology (JASIST) 55(5): 377 - 395.

[16] Meeting Report: NKOS Group Reviews Draft DTD for Thesauri, D-Lib Magazine, 6(12), 2000

[17] ANSI/NISO Z39.19 – 2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. NISO, 2005

[18] http://kvision.pku.edu.cn/kvision

[19] K. Sparck Jones. Automatic thesaurus construction and the relation of a thesaurus to indexing terms. Aslib Proceedings, 22, 1970, 26-28.

[20] G. Salton. Experiments in automatic thesaurus construction for information retrieval. Proceedings IFIP Congress 1971, TA-2, 43-49 (1971).

[21] Carolyn J. Crouch, et al. Experiments in automatic statistical thesaurus construction. SIGIR’92

[22] H. Chen, et al. Automatic Thesaurus Generation for an Electronic Community System. Journal of American Society of Information Science, 1995, 46,175-193

[23] Automatic thesaurus development: Term extraction from title metadata, Jun Wang, JASIST 57(7): 907-920 (2006)

[24] Jun Wang. Automatic feature thesaurus enrichment: extracting generic terms from digital gazetteer, JCDL’06 : 326 – 333 (2006)

[25]王军．词表的自动丰富--从元数据中提取关键词及其定位．中文信息学报. 2006.5

[26] http://kvision.pku.edu.cn/enrich-lcsh；http://kvision.pku.edu.cn/gazetteer

[27] Jun Wang, Meng-Chen Lee, Reconstructing DDC for Interactive Classification, ACM 16th Conf. on Information and Knowledge Management (CIKM 2007), Lisboa, Portugal

[28] http://kvision.pku.edu.cn/auto-ddc

[29] Dagobert Soergel. Reenginnering thesauri for new application: the AGROVOC example. Journal of Digital Information,4(4). 2004

[30] http://www.fao.org/aims/aos.jsp

[31] B. J. Wielinga, et al. From thesauri to Ontology. 1st International Conference On Knowledge Capture, 2001, 194 - 201

[32] 王军，基于分类法和主题词表的数字图书馆知识组织，中国图书馆学报，2004.30(3):41-44

[33] Douglas Tudhop, etc., FACET: thesaurus retrieval with semantic term expansion, JCDL’02