

论文题目：知识型档案管理系统建设及其关键技术研究

专业：计算数学

硕士生：肖丹

指导教师：李小福 副教授

## 摘要

企业档案管理经历了从数字化到信息化发展的历程，近年来随着档案管理信息化的深入发展，档案管理知识化的研究成为热门的课题。

档案管理是一项专业性很强的业务工作，数字化、信息化和知识化均必须满足业务管理的需求。虽然信息型档案管理与知识型档案管理均满足档案管理业务的需求，但在档案管理的服务方向、服务内容及其作用等方面均存在一定的差异。目前信息型档案管理系统建设，主要依据国际上公认的开放档案信息系统参考模型 OAIS (Reference Model of Open Archival Information System) 进行规划和设计，相关的系统建设技术也很成熟。而知识型档案管理系统建设目前仍然处于探索阶段，没有公认成熟的模型可供借鉴。本文从研究知识管理系统出发，结合档案管理业务的专业特点，对 OAIS 进行知识化扩充，提出了知识型档案管理系统参考模型 KAMS (Reference Model of Knowledge Archives Management System)。基于 KAMS 模型，本文对大型电力企业的知识型档案管理系统建设及其关键的知识表示技术作了相应研究。论文的主要工作如下：

1. 提出知识型档案管理系统中档案本体建设的方法和流程，并给出某大型电力企业档案本体建设的实例，对实际系统中档案本体建设具有指导意义。
2. 在研究知识管理系统的知识表示与知识应用相关理论和技术的基础上，结合档案管理业务的专业特点，提出知识型档案管理系统参考模型 KAMS。
3. 以 KAMS 为依据，规划设计了某大型电力企业的知识型档案管理系统，并成功应用于指导该企业知识型档案管理系统建设项目。

**关键字：**档案管理系统 知识管理 档案本体 知识型档案管理系统

---

Title: Research on the Construction and the key Technologies of  
Knowledge Archives Management System

Major: Computational Mathematics

Name: Xiao Dan

Supervisor: Li Xiaofu Associate Professor

## ABSTRACT

Enterprise archives management experiences a development from digitalization to informatization. With the in-depth development of archives management informatization, the research on knowledge archives management has become a hot topic in recent years.

Archives management is a highly professional business, therefore digitalization, informatization and knowledge construction demands for meeting the business management needs . Although both the informative and knowledge archives management meet the needs, there are still some differences among the service direction, service content and its role in certain aspects. At present, the construction of informative archives management system is planned and designed according to OAIIS(Reference Model of Open Archival Information System) which is internationally accepted. And related system construction technique is also very mature. However, the knowledge archives management system construction is still at an exploratory stage and there is not a widely accepted model. According to knowledge management systems and the professional traits of archives management business, we have enriched the knowledge of OAIIS, and came up with Reference Model of Knowledge Archives Management System (KAMS) in this paper. Based on KAMS model, we have done relative research on knowledge archives management system construction and its knowledge representation techniques in a large electrical enterprise.

We have present the main work as follows:

---

1. In this paper, we concluded the method and process in the construction of archives ontology, and showed the example of construction of archives ontology in the large electrical enterprise. All mentioned above makes significant guidance in practical construction of knowledge ontology.

2. Based on the theory and techniques in knowledge representation and application, combined with the professional traits of archives management business, we proposed Reference Model of Knowledge Archives Management System (KAMS).

3. According to KAMS, we have planned and designed a knowledge archives management system for a large electrical enterprise. It is successfully applying in the guidance for the construction project of the knowledge archives management system.

**Key words:** archives management system, knowledge management, archives ontology, knowledge archives management system

## 论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：肖丹

日期：2010年5月29日

## 学位论文使用授权声明

本人完全了解中山大学有关保留、使用学位论文的规定，即：学校有权保留学位论文并向国家主管部门或其指定机构送交论文的电子版和纸质版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆、院系资料室被查阅，有权将学位论文的内容编入有关数据库进行检索，可以采用复印、缩印或其他方法保存学位论文。

学位论文作者签名：肖丹

日期：2010年5月29日

导师签名：李松

日期：2010年5月29日

# 第一章 绪论

## 1.1 研究背景及意义

20 世纪 90 年代以来, 计算机技术和网络通信技术的迅速发展和成熟应用, 推动了信息技术的发展和信息化应用的普及与深化, 信息的收集与利用已经成为各行各业的一个重要竞争点。“信息化是当今世界发展的大趋势, 是推动经济社会变革的重要力量”<sup>[1]</sup>。有史以来, 人们从来没有像今天这样重视信息, 研究信息, 利用信息, 对于“信息化”内涵的理解也在不断提升, 正朝着“知识化”的方向发展。现阶段“信息化”, 以“数字化”为基本特征, 传递内容是数字化资源, 使网络化工作方式成为可能, 同时以“知识化”为目标, 为人们提供智能化的个性服务, 把信息资源以知识形式进行组织、管理, 把信息变成知识, 把知识变成决策, 从而提高组织机构的竞争力。在信息化的基础上提出的知识管理, 是一种以人为中心、以信息为基础、以知识创新为目标, 将知识看作可开发资源的思想理念。它的目标是寻求信息处理能力与知识创新能力的最佳结合, 在管理过程中最大限度实现知识共享。

档案作为重要的信息资源和知识载体, 怎样实现知识化的组织、存储和利用受到越来越多的关注。对于档案的定义, 目前公认的是《中华人民共和国档案法》中的表述: 过去和现在的国家机构、社会组织以及个人从事政治、军事、经济、技术、文化、宗教等活动直接形成的对国家和社会有保存价值的各种文字、图表、声像等不同形式的历史记录<sup>[2]</sup>。由此可见, 档案具有真实性、原始性和不可替代的依据和凭证作用, 能够提供历史资料, 并为组织、管理、决策、科研等工作提供服务。而档案管理工作伴随着档案生产、组织、保存、利用、销毁的全生命周期, 其管理模式对档案产生了很大的影响。目前, 传统的档案管理模式, 正面临着计算机技术、通信技术、网络技术以及多媒体技术的巨大冲击, 已经很难满足信息化管理的需求。因此, “档案信息化”的呼声也越来越高, 特别是在信息化活跃的企业, 以知识管理为战略目标的“档案信息化”已经成为各机构档案部门面对的重要课题。

随着“档案信息化”的深入研究，传统意义上的档案工作也在发生历史性变革，档案服务的方向、内容及其作用面临新的调整；信息技术发展导致档案爆炸性增长和档案载体形态的多元化，使传统的档案管理理论与方法越来越难以适应新形势新任务的要求<sup>[1]</sup>。一方面，由于信息的快速增长，各种形式的档案资源呈几何递增趋势，传统手工管理方式效率太低，无法在时间和空间上满足使用者的需求。另一方面，由于信息技术的广泛应用，越来越多的档案不再以纸质或胶片等传统形式为载体，而是以磁盘、光盘等新型材料为载体直接存放于计算机之中，由于其表现形式为多媒体、音像、动画等，也很难将其直接转换为传统档案进行收集和管理，从而迫切需要计算机辅助进行档案管理。在这种情况下，传统档案管理面临着极大的挑战，档案管理信息化、知识化成为一种趋势。

在“档案信息化”趋势明显以及“知识管理”热门研究的大环境下，本文将“知识管理”的理念与“档案信息化”相结合，主要研究了知识型档案管理系统建设，该系统应既能满足传统档案管理的自动化和信息化，包括实现档案资源的规划、收集、归档、利用、销毁等，又能将档案作为知识组织和管理起来，使其能够与日常业务管理对接，为管理者提供辅助。

“知识型档案管理系统”的提出，既源于档案界学者提出的“数字档案馆”，同时结合了“知识管理”的理念与技术，对其研究目前还处于探索阶段，从管理模式、法律法规、人文思想、技术路线等各方面都有很多亟待解决的问题，本文主要从信息技术的角度来研究其相关问题。目前“数字档案馆”和“知识管理系统”的建设中主要存在的技术包括：

1. 数字化技术，包括传统档案的数字化技术(主要采用 OCR 技术)和多媒体信息的数字化技术。
2. 档案获取技术，包括如何从已有的应用系统中抽取需要的档案资源。
3. 档案的知识表示和知识组织技术，包括将档案资源以知识的形式表示和组织，新的档案知识如何添加到已有知识库等等。它为档案检索、查询、利用以及高级服务奠定了基础，是将“数字档案馆”与“知识管理”相结合的关键信息技术。
4. 档案的检索技术，包括分类标准、索引技术等问题。
5. 档案的高级利用技术，特别是提供档案知识化服务，如知识地图导航、知识发现、个性化的知识检索等。

6. 档案共享技术,包括档案资源的传输技术、分布式资源的同步技术、档案知识的协作交流平台建设等。
7. 档案安全技术,由于档案具有法律和凭证的作用,需要长期保存且不可随意更改,所以安全技术的应用尤其重要。主要涉及到数字水印技术、防火墙技术、系统权限和加密、系统检测与监控、数据备份工作等。

以上问题的研究对建设知识型档案管理系统具有重要影响,能否得到有效解决决定了系统建设成功与否,也决定了系统建设的发展方向。只有找到合适的解决方案,才能成功地建设知识型档案管理系统,从而圆满实现档案管理信息化、知识化的转变,使档案馆从封闭走向开放,最大限度地满足组织决策和社会大众的需求。本文分析总结了档案管理系统的模型及其涉及的关键技术,重点研究了知识型档案管理系统建设中的知识表示技术,并对系统建设作了相应研究,虽然还不是十分成熟和完善,但为档案管理工作向知识化阶段的迈进提出一些有意义的研究成果。

## 1.2 档案管理系统的发展历程

档案管理系统的发展与计算机技术的发展密不可分,早在 20 世纪 70 年代中期,国外的部分档案馆已经开始了自动化之路。1994 年 1 月,在第二届国际人文大会上 Margaret Hedstrom 提交了《电子档案馆——网络环境的集成与利用》(Electronic Archives: Integrity and Access in the Network Enviroment),提出了“电子档案馆(Electronic Archives)”的概念。1996 年,美国图书馆研究团体和美国国家档案与文件局合作成立了数字档案的特别工作组——Task Force。1996 年,Task Force 出台了一份关于数字档案的最终报告和建议,阐述了数字档案馆与数字图书馆在信息管理上的显著不同:数字档案馆必须保证信息的长期保管和存取利用,数字图书馆则没有这个责任<sup>[4]</sup>。这个报告勾画了国家数字档案馆,对数字档案信息的长期保存和利用提出了建议。同年,在北京举行的第 13 届国际档案大会上,美国《档案博物馆信息》杂志编辑戴维·比尔曼发表了《虚拟档案》一文,提出了物理档案、实态档案馆和相对应的虚拟档案、虚拟档案馆的概念。从此,电子档案馆、全球档案馆、超级档案馆、数字档案馆等新名称和相关研究成果层出不穷。“数字档案馆”研究的内容不仅包括档案管理信息化技术,还包括组织管理方式、管理人员的

培养管理方面的内容等,本文主要从信息化的角度出发,讨论“数字档案馆”中的“档案管理系统”建设相关问题。

作为现代社会不可或缺的信息机构,档案馆根据其管理结构,可以将“数字档案馆”的发展划分为三个阶段<sup>[5]</sup>,分别是第一阶段的数据型(data)数字档案馆,第二阶段的信息型(information)数字档案馆和第三阶段的知识型(knowledge)数字档案馆。作为“数字档案馆”建设中不可或缺的一部分,我们可以认为“档案管理系统”也经历了数据型、信息型和知识型这三个阶段。

档案管理系统的三个发展阶段是循序渐进的,它们相互渗透相互制约,顺因信息技术的发展而发展。面对第二阶段信息型档案管理系统的光辉成就以及关键技术的不断发展,如何有效组织日益增长的档案知识资源,为公众提供系统、有效的知识化服务与利用,成为第三阶段知识型档案管理系统区别与第二阶段的重要特征。而在档案管理系统从信息型向知识型过渡过程中,怎样将档案以知识形式表示并组织起来,以提供知识服务,成为了关键技术所在。

## 1.3 国内外研究现状

### 1.3.1 国外研究现状

国外对于档案信息化的研究起步较早,早在20世纪90年代初就开始从理论和概念上对其进行研究和探讨,美国、加拿大、日本、英国、比利时等很多国家都有自己的数字档案馆,其中以美国和英国为代表,其数字档案馆的建设起步早,发展快。

美国涉及数字档案管理系统的项目比较多,主要情况如下:2000年3月28日,美国国家档案与文件管理局(the National Archives and Records Administration of the United States, NARA)的局长约翰·卡林宣布,NARA将与美国国家科学基金会合作,共同创建一个电子文件档案馆,并计划在2009年之前实现所有纸质档案的数字化转化。约翰·卡林对建立这个电子文件档案馆的目的作了如下解释:“这个电子文件档案馆将无限期地保存任何类型的电子文件,使其不受生成格式的制约,用户也能够目前在和未来使用的计算机系统上阅读到这些文件。”<sup>[6]</sup>该计划可以看作是美



国整体“数字档案馆”研究与建设的一个组成部分。电子文件档案馆和总统数字档案馆是美国最有代表性的数字档案馆。

英国是世界是实施“电子政务”较早也是较好的国家之一，其国家数字档案馆的建设也起步较早。为了保证电子文件的安全保管和利用，英国公共档案馆实施了两项计划，分别是针对公共网络化的电子文件办公系统 EROS (Electronic Records Office System)和针对结构数据集的联合王国数字档案馆 UKNDA 计划(UK National Digital Archive of datasets)<sup>[6]</sup>。其中联合王国数字档案馆在新的数字资料保存方面发挥了积极的作用。它是比较有代表性的数字档案馆，由伦敦大学计算机中心和伦敦大学图书馆建设和维护，它按内容、时间或档案类型分类等将文件分类管理，并拥有独立的检索系统，能够提供基于关键词检索等检索手段。它的主要功能包括：接受、存储、保管政府部门产生的档案数据，并提供对档案数据的利用。

日本关西数字档案馆是一个专题性的数字档案馆，记录了日本关西地区乃至整个日本历史发展的过程。该项目的将有形的或无形的文化财产以数字形态记录，实现了三个目标：成为传输日本文化信息的基地；成为旅游等文化活动新的休闲地带；最大化利用地区文化遗产并提高文化交流质量。关西数字档案馆扩大了数字档案馆的利用范围，为数字档案馆的发展服务功能提供了一种新思路。

此外，国际上比较成功的数字档案馆案例还有维多利亚电子文件策略项目、加利福尼亚大学历史档案馆、杰斐逊专题数字档案馆、美国 911 数字档案馆等。

### 1.3.2 国内研究现状

在 1999 年由国家档案局发布的《中华人民共和国档案法实施办法》中规定，各级国家档案馆保管的档案应当按照《档案法》的有关规定，分期分批地向社会开放，同时公布开放档案的目录。直至现阶段，我国的各级各类档案馆已经普遍实现计算机辅助管理档案，并在提供档案的社会化服务方面作了很多工作。对档案利用的广度和深度不断加强，同时催生了对档案管理系统的深入研究。

2000 年，国家档案局、中央档案馆印发的《全国档案事业发展“十五”计划》明确提出要“加快档案信息化、加快现有档案的数字化进程”。

2002 年 11 月，国家档案局发布的《全国档案信息化建设实施纲要》中明确提出要“在部分中心城市建设示范性数字档案馆”。

深圳数字档案馆<sup>[7]</sup>，是国家“数字化档案馆系统”工程的第一个试点馆，其应用系统包括：电子档案信息收集、电子档案信息管理、电子档案信息利用、电子档案信息安全维护四个功能模块。它的目标是建成一个可扩展的网络应用系统，其功能包括档案的数字化、档案信息的采集、处理、存储、归档、组织、发布、利用及资源管理全过程。系统采用的关键技术，包括大量并发查询、数据仓库、数据挖掘、海量数据存储、网络安全、图象分类、智能检索、人工语言向自然语言转换、视频点播、虚拟现实技术等，这些技术都属于当前计算机技术、信息技术、人工智能领域中最先进、最有发展前途的前沿科学。

青岛市数字档案馆<sup>[8]</sup>是国内首家运用数字设备和技术建成并第一个正式投入使用的数字档案馆，也得到了国内同行的认可。青岛市数字档案馆已经建成档案信息传输平台(主要由局域网络、政府网络和互联网络三个平台组成)、档案信息采集平台(主要包括电子目录采集和全文数据采集，承担日常信息采集工作)、档案信息存储管理平台和档案信息应用平台(包括馆内应用平台、国际互联网档案信息发布和在线利用平台、政府网络信息服务平台)等四个综合性的软硬件平台。

2002年以来，杭州市档案馆、天津开发区档案馆等也逐步开始试点建设数字档案馆。广东省顺德市、广州市白云区、深圳市福田区和宝安区、浙江省萧山区档案馆等在内的档案馆在数字化、信息化和网络化建设方面也取得了较大成效。

### 1.3.3 企业研究现状

目前以政府部门指导建设的数字档案馆发展迅速，同时也极大促进了企业“档案信息化”的进程。2004年12月21日，国家档案局“十五”重点科研项目——江苏省电力公司数字档案馆通过了国家档案局组织的专家鉴定，标志着国内第一家企业数字档案馆的诞生<sup>[9]</sup>。江苏省电力公司数字档案馆的目标主要是把分散于全省各地的所属单位不同载体、不同类别的档案信息资源，以数字形式采集、存储。公司员工可用带身份识别系统的电子卡通过局域网，在自己的使用权限内从系统中查阅相关的档案信息，以服务于生产、建设和经营等工作。经专家认定，该项目实现了馆藏资源数字化、档案信息结构标准化、数据传递网络化、生成空间虚拟化、信息利用全员化和档案管理自动化；系统在电子档案的永久真实保存、馆藏数字资源的综合利用等方面达到国内领先水平，它采用NAS系统对海量电子数据

集中管理、多格式安全浏览实现数据安全、全文检索技术实现档案全方位便利检索。

2008年12月4日,胜利油田数字档案馆正式启用<sup>[10]</sup>。该数字档案馆目前已开发建设了管理信息系统功能、档案发布网站功能、档案归档接口功能、数据利用功能以及系统辅助工具包功能,使用了较先进的专业档案接口、多级档案授权管理体系、多格式浏览技术和全文检索等技术,数据库中存储基建、设备、实物、照片档案等电子文件1.7万多件,达到行业领先、国内先进水平。胜利油田数字档案馆的功能包括:建立了统一的档案管理体系和平台,实现了广域范围内多个立档单位的统一档案管理,达到馆际间的无缝连接;建成了智能的档案采集系统,构建了档案管理平台与其他信息系统的数据一体化采集接口,可以全面完整、及时高效采集档案相关信息;建立起档案业务的自动化管理模式,提高了档案管理水平;构建了网络化的档案信息服务体系,实现档案信息资源的存储数字化、管理自动化、利用网络化。

此外,韶光钢铁集团档案综合管理系统、广东移动企业档案综合管理系统、山东铝业数字档案馆等的建设也为大型企业档案管理系统建设提供了切实可行的解决方案。

## 1.4 本文主要工作及创新点

### 1.4.1 本文主要工作

本文根据档案管理系统的相关理论,在充分研究知识管理模式后,以大型电力企业为背景,对知识型档案管理系统建设作了研究,并对其中涉及的知识表示技术——档案本体构建作了实际性研究。

本文的主要工作包括:

第一章:研究了档案管理系统的发展历程,以及国内外目前的研究现状,交待了本文的研究背景及意义。

第二章:分阶段研究了档案管理系统的相关理论如系统组成要素、系统模型及其关键技术等。重点研究了第二阶段信息型档案管理系统中提到的开放档案信息

系统参考模型(OAIS), 以及第三阶段知识型档案管理系统中涉及的知识管理相关理论。

第三章: 在研究了知识表示以及本体理论的基础上, 提出了档案本体建设的基本流程, 并以此为依据实际构建了某大型国有电力企业的档案本体。

第四章: 通过对 OAIS 模型和知识管理系统进行分析研究, 对 OAIS 进行知识化扩充, 提出了知识型档案管理系统参考模型 KAMS, 并以此为理论依据, 结合某大型国有电力企业实际情况, 从业务管理模型、系统功能模型、功能实施模型以及系统服务架构多方面对电力企业知识型档案管理的建设作了研究并得到相应成果。

### 1.4.2 本文创新点

目前对于信息型档案管理系统的研究已经达到比较成熟的阶段, 在国内外研究现状分析中提到的“数字档案馆”建设成果中, 大部是信息型档案管理系统, 其相关技术也得到了长足的发展, 现有的数字档案馆建设大多以 OAIS 参考模型为参照进行建设, 而对于知识型档案管理系统的研究目前还处于初级阶段。本文的主要创新点是:

1. 针对知识管理和档案管理的相似性, 将“知识管理”思想与档案管理系统建设相结合, 对 OAIS 模型进行知识化扩充, 提出了知识型档案管理系统参考模型 KAMS。
2. 在电力企业环境下, 以 KAMS 为基础, 以实现档案的知识化管理为目标, 从业务管理模型、系统功能模型、功能实施模型以及系统服务架构几个方面, 对电力企业知识型档案管理系统建设作了相应研究。
3. 以电力企业档案管理为背景, 提出知识型档案管理系统中构建档案本体的方法和流程, 并以此为依据实际构建了该企业档案本体。

## 第二章 档案管理系统相关理论

档案管理经历了从传统的手工管理到计算机辅助管理的过渡，对于计算机辅助管理的档案管理系统又经历了数据型、信息型和知识型三个阶段。知识型档案管理系统的目标是实现档案的知识化管理，它要求既能满足档案的传统管理功能，又能实现对档案信息化、知识化的管理利用，因此它以信息型档案管理系统为基础，以实现档案的“知识管理”为目标。

本章分三个阶段对档案管理系统的相关理论及关键技术作了研究分析，并重点讨论了信息型档案管理系统中的 OAIS 参考模型，以及知识型档案管理系统中涉及的知识管理相关理论，为本文在第三章和第四章对知识型档案管理系统的研究奠定了理论基础。

### 2.1 传统档案管理

#### (一)传统档案与数字档案的差异

档案的本质属性即原始记录性，因此数字档案信息必须明确归属于某一档案机构。数字档案信息的原始性、完整性、可靠性、真实性在很大程度上依靠档案部门的技术、权威和信誉来保证<sup>[1]</sup>。因此，档案管理系统的管理权限较之传统档案馆并未扩大，其核心是将传统档案馆的管理职能实现自动化和网络化。在管理上数字档案与传统档案还存在以下几个方面的差异<sup>[2]</sup>：

1. 数字档案的归档时限与传统档案不同。数字档案可以在结案后随时归档，或定期由计算机自动收集统一归档以保证归档率和完整率。
2. 数字档案的分类、编目等管理方式与传统档案不同。数字档案在归档时即可编制唯一的文件号，文件号中可以包含分类号等其他相关信息，归档同时进行主题词著录。
3. 数字档案的归档方式和保密措施与传统档案不同。数字档案具有易复制性和易更改性，在归档时要进行数字签名、信息认证、身份验证和加密保护确保其在传输、处理和利用过程中的原始性、真实性和准确性。
4. 数字档案在保管方式上和传统档案有差异。数字档案的保管与其格式、存贮介质密切相关。

5. 数字档案查询利用方式和传统档案不同。用户可以通过远程登录的方式进入到资料中心，通过身份确认后根据自己的权限对档案进行查询利用，一份档案可以同时提供给多个用户。

## (二) 传统档案面临的几个问题

目前传统馆藏档案的资源管理普遍存在以下几个问题<sup>[13]</sup>：

1. 档案门类不全，内容单一，利用率低。
2. 馆藏量大，库房面积少，严重超负荷。
3. 重复现象普遍，严重影响了馆藏的优化。
4. 整理工作和档案鉴定工作压力增大。
5. 收集和接收档案资料的力度不够。

档案资源是一个地区政治、经济、科学、文化、教育发展水平的综合反映。而传统档案管理存在的以上问题很大程度上制约了档案资源的丰富与优化，促使档案工作者寻求新的管理模式。利用计算机辅助档案管理，即建立档案管理系统实现对档案工作的自动化、网络化模拟，是在信息技术发展迅速环境下档案管理工作新的出路。

## 2.2 数据、信息、知识与档案

### (一)数据

数据是用一组符号及其组合对客观事物的数量、属性、位置及相互关系进行抽象表示，是对实际事实的记录，是没有经过解释的信号。数据有四种表现形式：数字、文字、声音和图像，这些符号本身没有任何意义，经过加工、存储和处理过程后才能表现其价值。数据经过处理得到的就是信息，即“信息=数据+数据处理”。在档案资源中常见的数据型档案包括：财务凭证、销售数据、物品材料变动记录等。

### (二)信息

信息是由数据在特定场合解释而成，反映行为和现象的状态及其变化方式，是具有意义的数据，是数据负载的内容，也可以说是隐含在数据中的内在逻辑和规则。在档案资源中常见的信息型档案包括：会议纪要、重要事件记录、工作通知，往来函电邮件、输变电质量和安全情况等。

### (三)知识

知识这个概念广泛而抽象，它始于古希腊时期西方哲学的认识论辩论。常见的知识定义有以下几种：

1. 知识是一种得到了证明的、能够提高个人行为有效性的信念。——Nonaka(1944)和 Huber(1991)
2. 知识是经过消减、塑造、解释和转换的信息。——Feigenbaum
3. 知识是由特定领域的描述、关系和过程组成的。——Bernstein
4. 知识是事实、信念和启发式规则。——Hayes-roth
5. 从知识库的观点来看，知识是某领域中所涉及的各有关方面的一种符号表示。
6. 信息是经过解释能够成为一种结构含义丰富的数据，而知识是经过证实并被认为具有真实性的信息。——Vance(1997)
7. 数据是原始的数字和事实，信息是经过处理的数据，而知识是“可发挥作用的信息”。——Maglitta(1995)

由此可见数据、信息和知识之间存在着假定的等级结构和联系。

知识是将有关信息关联在一起所形成的、反映客观事物间关系的信息，是对事物及其运行规律的认识或结论，是数据和信息整合的完备形式，具有目的感和再生能力。在档案资源中常见的知识型档案包括：规划报告、科研报告、设计图纸、工艺文件、部门工作总结等。

### (四)档案

以企业为背景，有专家对企业档案作出了定义，指出企业档案是企业人员状况及其人员工作活动的记载。它记录了企业的人员状况数据及其变化的信息；记录了企业的资产状况数据及其变化的信息；记录了企业的主要工作过程数据和信息；记载了企业人员的工作成果(学习成果、规划成果、设计成果、研究成果、生产成果等)；记录了企业人员的工作看法和认识。由此看出一部分档案直接表现为知识；一部分档案是数据，它们经过认识和统计可转换成信息和知识；一部分档案是信息，它们经过综合和判断可转换成知识。

### (五)数据、信息、知识与档案间的联系

简单来说，数据是信息的载体，信息含在数据之中；信息是知识的载体，知识是信息的内涵，也是能导致有效行为的信息(改善决策或提高效率)。数据可按照

统计学模型挖掘转化成信息、知识；信息可以挖掘转化成知识。档案资源以数据、信息和知识的形式存在，对于档案的知识化管理既要实现传统管理业务，又要注重将数据型和信息型档案转化为知识型档案进行管理。

经分析可以知道档案资源与数据、信息、知识的关系如图 2-1 所示：

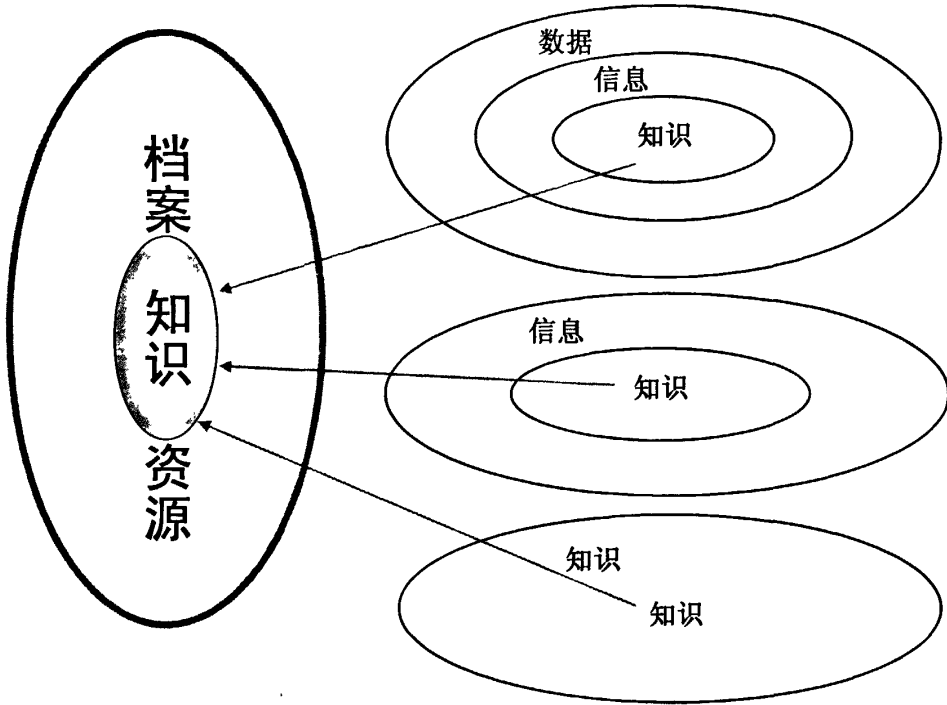


图 2-1 数据、信息、知识、档案之间的关系

## 2.2 数据型档案管理系统

作为档案管理系统发展的第一阶段，数据型(data)档案管理系统的主要工作是：建立计算机辅助档案管理系统、接收业事务系统中的电子文件并做好数据库文件备份。由于数字资源来源不同，档案馆所接收的数据存在异构性，给检索利用带来一定障碍，因此又提出两种不同的管理策略：

### (一) 集中式保管模式

该方案的管理理念与传统档案馆的管理十分相似，将各类档案全部移交档案馆集中保存。即在文件失去现行价值后，由档案馆负责整理、分类、编目和长期保存，电子文件是集中存放在档案馆特定的软硬件环境中。



## (二) 分布式保管模式

这种方案提出档案由文件形成机构自行保管而不向档案馆移交，但是提供通过网络向档案馆进行虚拟归档的功能。此时档案系统的主要工作是对这些档案进行登记管理并加以控制。

数据型档案管理系统建设主要致力于馆藏档案目录的数字化，并在此基础上尝试建立数字档案目录中心，向利用者介绍档案资源的分布情况，这是档案管理系统建设的起步和初级阶段。

## 2.3 信息型档案管理系统

信息型(information)档案管理系统是档案管理系统发展的第二个阶段，这一阶段档案管理系统的建设重点是：建立数字资源标准化体系，对办公自动化(OA)系统和业务系统中的文件进行规范化著录，对已有馆藏档案进行数字化，通过索引、检索等技术，使利用者能够借助网络，对分散在不同档案馆的数字资源进行远程存取或利用。

对于信息型档案管理系统，档案工作者对各类档案信息的创建、获取、存储、管理、传输、发布等做了大量研究，提出了档案一体化管理的理念，比较具有代表的建设成果有：分布式数字档案原型(CEDARS)和开放档案信息系统参考模型(OAIS)。目前在网络环境中实现信息存储与远程服务功能的信息型档案管理系统成功案例主要有：美国的NARA(美国国家档案馆网站)、ARIAS(宾夕法尼亚档案系统)、加里弗尼来在线档案，日本的石川新情报发展书府、加拿大的数字收藏(Canada's Digital Collection)等。

### 2.3.1 系统组成要素

信息型档案管理系统建设包括以下四个方面<sup>[5]</sup>：数字化资源、自动化工作流程、网络化存取和分布式管理。

#### (一) 数字化资源

数字资源的形式包括目录、文摘文本、图形、声音、图像文件等。从信息存储方式的角度，数字资源可分为两种类型，即实体资源(指保存在本馆的数字档案)

和虚拟资源(指存放于异地且需要通过网络获取的数字档案)。在数字化资源方面,主要有以下几个需要讨论的问题:电子文件的管理与保存、馆藏资源数字化、网络采集档案信息、数字资源的存储、档案信息资源的组织与管理。

## (二) 自动化工作流程

档案管理系统的建设促进了档案部门工作方式的创新,通过设置与其他自动化业务系统的接口,提高了档案工作的自动化管理水平,形成了一个协作平台。一般的档案管理系统主要具有以下几方面的功能:

1. 档案采集:包括信息接收、馆藏资源数字化、信息搜寻等。
2. 档案管理:包括信息标准化,档案分类、鉴定、元数据著录、目录管理、信息统计等。
3. 档案利用:包括档案发布、查询检索、编研等。
4. 系统维护:包括数据维护、权限管理、网络安全、加密存储、数字签名、版权保护、日志管理、数据备份与恢复等。

## (三) 网络化存取

利用现代网络技术为用户提供零距离、全天候、全方位的开放式服务是档案管理系统的重要特征。系统要能将分散在各处的数字档案信息资源点的资源有效组织起来,并能够按照用户的需求发送到用户客户端,它具备良好的网络存取功能。

信息资源可分为对象数据库和元数据库。对象数据库是指数字化的文本、图片、声音、影像等数据,元数据是对对象数据的描述性数据,记录有关对象数据的属性,并对对象数据进行管理的数据。元数据库应集中在档案馆中心的大规模服务器上,对象数据库则可分布式存放在各业务应用系统的资源点内,当需要调用对象数据时,档案馆中心通过元数据调度相应的对象数据库中的数据,提供给用户。

## (三) 分布式等第制管理

网络管理系统有集中式和分布式两种管理模式,集中式管理模式是在网络中建立一个全面负责管理所有网络资源的网管中心,而分布式管理模式是按网络层次、区域建立多个网管中心,分别负责管理不同区域和不同层次的网络资源。分布式等第制提出,对局域网络实行分布式管理的同时,对上下级局域网间实行集

中协调的原则。它的原理是，一方面以局域网为独立的节点网，建立节点网的网管中心。另一方面，上级节点网的网管中心负责向属于它管理范围的下级节点网的网管中心提取网管信息，完成集中管理。

### 2.3.2 开放档案信息系统参考模型(OAIS)

20 世纪末，美国空间数据系统咨询委员会 CCSDS(Consultative Committee for Data System)制定并发布了开放档案信息系统参考模型 OAIS (Reference Model of Open Archival Information System)。其目标定位是为具有数字信息长期保存和利用责任的档案机构或非档案机构提供数字信息长期保存和利用的基本概念、术语、基本信息模型、功能模型和指南。在 CCSDS 和 ISO 的推动下，OAIS 参考模型成为了档案管理系统普遍遵从的标准规范，来自图书馆、档案馆、科学数据、文化传承等领域的相关项目和系统都纷纷以其为准，进行系统的规划和设计。因此，本文将从环境模型、功能模型和过程模型几个方面对 OAIS 模型进行介绍<sup>[14]</sup>。

#### 2.3.2.1 OAIS 环境模型

OAIS 将复杂的信息环境抽象为三个角色：信息形成者、管理者和用户<sup>[15]</sup>，如图 2-2 所示。

- (1)形成者：包括提供档案信息及元数据的人员或客户端系统。
- (2)管理者：是管理系统中制定流程、制度的角色，它的职责是控制档案的有效管理，明确档案收集方案，确定档案的保存期限、访问权限等，并对 OAIS 的运行情况进行分析和调整。
- (3)用户：是获取并利用档案信息的人员或客户端系统。

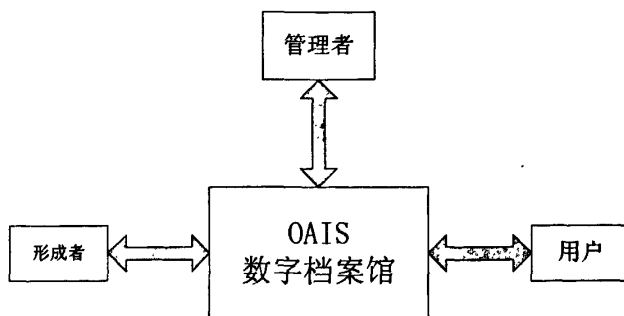


图 2-2 OAIS 环境模型

### 2.3.2.2 OAIS 参考模型

OAIS 参考模型具体可分为：信息模型、功能模型、过程模型。

#### (一) OAIS 信息模型<sup>[16]</sup>

OAIS 信息模型由信息对象、信息包、数据管理信息三部分组成。

(1)信息对象(Information Object)：指由数据对象组成，通过与表征信息结合为用户理解的信息。

- 数据对象：指实体数据或者数字数据。
- 表征信息：将数据对象以一种有意义的方式呈现给用户的信息。

信息对象三者之间的关系如下图所示：

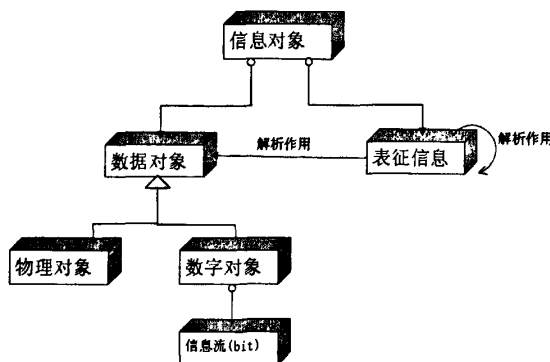


图 2-3 信息对象

#### (2) 信息包(Information Package)

信息包是 OAIS 信息模型中的核心概念，也是功能模型中传递的基本对象。信息包的结构包括：

- 内容信息：被保存的原始信息对象。
  - 保存描述信息：描述内容信息的特征，具体包括来源信息、背景信息、参考信息、固化信息。
  - 打包信息：将内容信息和保存描述信息联系、封装和识别的信息。内容信息、保存描述信息通过打包信息构成一个信息包。
  - 描述信息：是对信息包的特性和属性的描述，它帮助定位和检索用户所需信息。
- 它们之间的关系如图所示<sup>[17]</sup>：

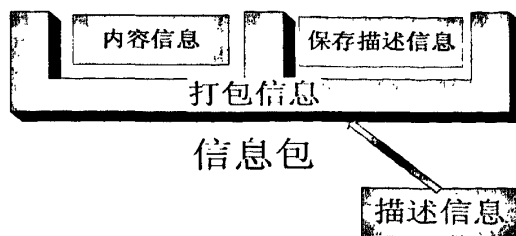


图 2-4 信息包

从信息包的种类来看，信息包分为以下三种：

- 提交信息包(Submission Information Package, SIP)：生产者提交给 OAIS 的信息包，它的形式和具体内容通常由信息生产者和 OAIS 协商决定。
- 档案信息包(Archival Information Package, AIP)：符合永久或长期保存属性的由 OAIS 保存的信息包。
- 分发信息包(Dissemination Information Package, DIP)：OAIS 基于信息用户需求而传递给用户的信息包，可以从一个或若干个 AIP 中提取出来。

(3) 数据管理信息(Data Management Information)：包括政策信息、请求追踪信息、安全信息、基于事件的需求信息、统计信息、保存过程历史信息、用户基本信息、审计信息等，一般由数据管理模块产生和储存。

## (二) OAIS 功能模型

OAIS 功能模型中，定义了 6 项功能，分别为：接收功能、档案存储功能、数据管理功能、行政管理功能、保存规划功能、存取功能。

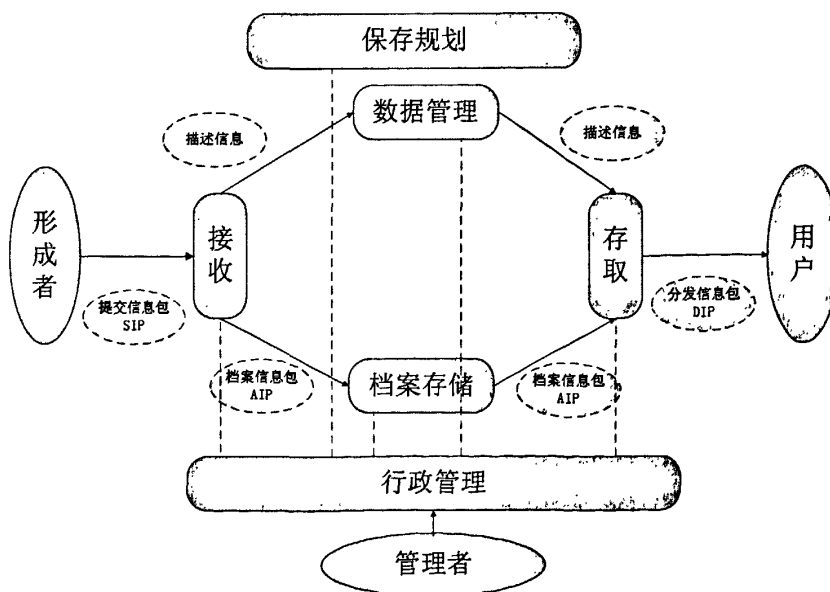


图 2-4 OAIS 功能模型<sup>[18]</sup>

- (1) 接收功能(Ingest), 接收信息生产者提交的提交信息包(SIP), 并将 SIP 转换为档案馆中需要管理和存储的内容。
- (2) 档案存储功能(Archival Storage), 主要是存储、维护和检索档案信息包。
- (3) 数据管理功能(Data Management), 用于识别档案描述信息, 并为档案描述性数据的组配、管理与访问提供相关服务与功能。
- (4) 行政管理功能(Administration), 为档案管理系统的整个运行提供操纵管理等相关的服务。
- (5) 保存规划功能(Preservation Planning), 用于监测 OAIS 的环境, 提供相关建议以确保信息存储的长期保存方案, 即便是在原有计算机环境退化之后, 仍然能够被目标用户所访问。
- (6) 存取功能(Access), 用于支持用户鉴定、定位存储于 OAIS 中的档案信息并对其进行有效访问。

(三) 过程模型<sup>[19]</sup>

过程模型介绍了信息包转换过程, 如图 2-6, 即档案信息如何实现从形成者经由 OAIS 到用户这个生命周期过程中在逻辑上或物理上的相互转换。

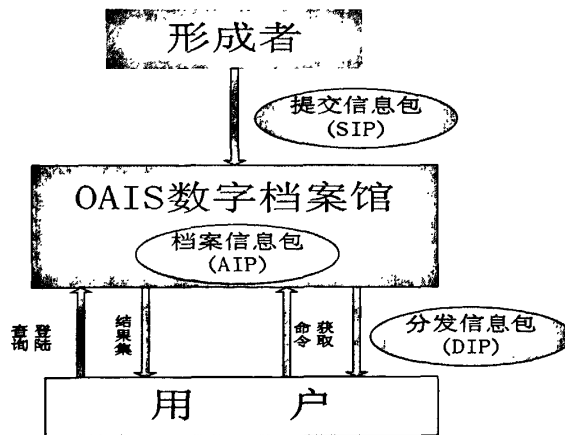


图 2-6 信息包转换

- (1) 形成者中的数据转换: 在将信息存储在 OAIS 之前, 需要与管理者就提交信息包的内容、格式、储存时间等问题达成提交协议。
- (2) 接收功能模块中的数据转换: 接收模块将接收到的提交信息包转换成为长期存储的档案信息包和包描述。

(3) 档案存储功能和数据管理功能模块中的数据转换：档案存储功能将档案信息包转换成长久保存的档案信息，数据管理功能接收由获取功能模块产生的描述信息。

(4) 存取功能模块中的数据转换：存取功能接受到用户的请求并确认是否满足要求，从存储和数据管理模块取得档案信息包和相关的描述信息，将其转换成分发信息包集合，并将分发信息包通过发布过程传递给用户。

OAIS 参考模型是对与数字档案系统相关的环境、功能模块以及信息对象的概念化，它为档案管理系统的设计提供了一个很好的参考模型。它刻画了实际系统中的高层任务、服务以及信息需求。我们遵循此参考模型，进一步结合系统的体系结构、存储或处理过程、数据库设计、处理平台等技术方面的细节，就能构建出一个信息型档案管理系统。

### 2.3.2.3 基于 OAIS 的档案管理系统

信息型档案管理系统是对档案信息内容的管理，它将档案信息的保存、管理、利用结合成一体化管理流程，涵盖档案信息的采集、分类、存储、管理、发布与利用全过程。根据 OAIS 参考模型，档案管理系统框架由四个应用系统(馆藏建设系统、管理系统、发布系统和工作流系统)和两个存储系统(元数据存储系统和档案存储系统)组成，如图 2-7 所示：

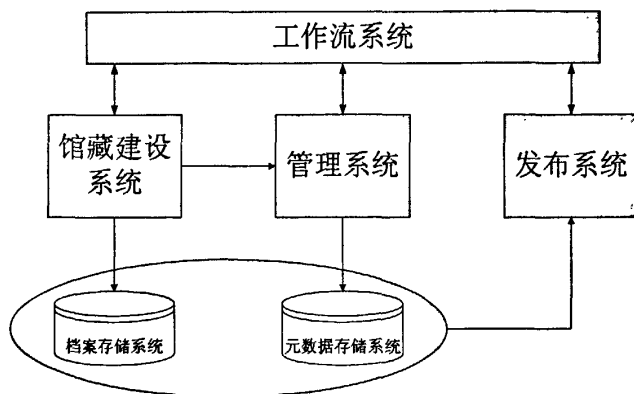


图 2-7 基于 OAIS 的档案管理系统

#### (一) 馆藏建设系统

馆藏建设系统的主要任务是：对外部档案信息的捕获；对传统档案的数字化、标准化和格式转换等工作；对信息进行元数据著录、确定存储方式和保管计划等。馆藏建设系统是整个档案管理系统“内容”的来源。它的建设目标是确定档案信息及

其相关元数据的准确性、完整性和有效性，将档案信息及相关元数据纳入档案存储系统和元数据存储系统中。

## (二) 管理系统

管理系统的功能主要是：管理数据库、执行查询、产生报告和进行数据库的更新工作，使得档案内容被保存、查找、更新等。它本质上是对档案信息对象数据和元数据进行管理，保障档案信息的完整性、有效性、一致性和长期可阅读性。此外，该系统还对存档和访问权限进行监控；对系统运行进行检测更新；执行数据管理、数据查询，生成查询结果。这里涉及的技术包括信息模型的构建、保存元数据体系、内容管理、元数据管理、索引等方面的技术<sup>[20]</sup>。

## (三) 发布系统

发布系统负责将档案信息从存储系统中快速、自动地推送至档案使用者。由于网络技术、电子商务、电子政务的快速发展，门户已成为众所周知的概念，并被作为有效访问 Internet 信息资源的主要窗口。门户发展已经经历了四个阶段：布尔检索、分类导航、个性化和集成化<sup>[21]</sup>。档案管理系统门户应该是一个集成化的门户，既为用户提供统一的内容检索入口，同时也具有内容分类导航功能，比如以内容类型或分类主题建立内容导航。另外，档案管理门户也应整合新的服务功能，如虚拟交流、实时聊天、个性化的搜索引擎界面、基于个性化的内容专题报道等等<sup>[22]</sup>。

## (四) workflow 系统

workflow 系统确保档案信息从捕获、储存到发布的全生命周期过程可以有效、正确地运行。它为档案信息的捕获、组织、维护、鉴定、销毁、发布等提供了流程管理，并使得档案工作流程变得更简洁、责任更明确、内容更准确。

## (五) 档案存储系统

档案存储系统负责档案信息的内容存储与更新维护，并为档案信息的访问和处理提供相应支持。档案存储系统与元数据存储系统共同为档案信息长期保存与访问利用提供规范的、可共享的档案信息资源，负责元数据与档案数据的一体化管理，以保持两者的同步性和一致性。

## (六) 元数据存储系统



元数据是关于数据的数据，它描述档案信息的背景、内容、结构及其整个管理过程并可为系统自动分析、提取和归纳信息提供依据。它用于保障信息的真实性、完整性、关联性和长期有效性；并用于对分布在不同网络环境下的档案信息进行有效管理；协助对分布式档案信息的集成化利用，更是构建信息组织体系必不可少的工具，是数字档案馆建设的基础性标准<sup>[23]</sup>。所谓元数据存储系统是指存储元数据并对其整个生命周期进行管理的系统，是元数据访问的平台。

### 2.3.3 信息型系统关键技术

信息型档案管理系统建设目前处于成熟阶段，其中涉及到的海量数据存储、数据安全、数据检索与访问、数据挖掘、面向服务的架构(SOA)等技术，发展较为成熟。

#### 2.3.3.1 海量数据存储技术

由于各类结构化、非结构化的档案资源需要移交到档案管理系统，海量信息存储的需求不断增强，需要用海量存储设备进行存储，同时提供档案信息的网络化服务利用。

海量数据存储的解决方案主要用到的技术有：直接附加存储(Direct Attached Storage, DAS)、网络附加存储(Network Attached Storage, NAS)、存储区域网(Storage Area Network, SAN)以及内容寻址存储(Content Addressing Storage, CAS)<sup>[24]</sup>

##### 1. 直接附加存储(DAS)

DAS 是一种传统存储方式，其原理是在本地网络上将存储设备与电缆一对一地直接连接到服务器或用户端接口上。DAS 没有独立的操作系统，而是依赖服务器或客户端的操作系统完成对数据的存储与管理及访问控制。服务器和存储设备之间通过独立、专用的通道进行连接，当服务器需要增加存储量时，可以直接在服务器上连接存储器。

DAS 的优点是数据存储速度快，响应速度快，数据时刻在线。它的不足之处在于大量占用服务器资源，当用户增加或运行频繁时，服务器将成为数据存储与访问的瓶颈。DAS 环境容易导致信息孤岛数量的激增，降低资源利用率。受服务

器扩展能力的限制，DAS 的容量会受到限制不能无限扩充。所以 DAS 技术比较适合于信息量较小的档案馆利用。

## 2. 网络附加存储(NAS)

NAS 的工作原理是：使用一个附加层将存储设备从服务器独立出来连接到现有的网络上，通过网络共享方式给客户id提供网络数据资源服务，客户不经过服务器直接访问存储设备上的数据。这个附加层的作用是通过 IP 对存储文件进行寻址。NAS 服务器一般由存储硬盘、操作系统和文件系统几个部分组成。

NAS 的优点是：不同的服务器可以共享存储设备，它实现了异构环境下的数据共享，实现即插即用、支持在线扩容、具有良好的扩展性。服务器只需要支持 IP 协议，不再需要自己的存储设备，可以使用不同的操作系统平台。NAS 的不足之处是使用网络进行备份和恢复过程中带宽消耗较大，因此网络除了处理正常的用户传输外，还要处理包括备份操作的存储磁盘请求。

## 3. 存储区域网(SAN)

SAN 是一种通过光纤集线器、光纤路由器、光纤交换机等连接设备将多台存储设备与相关服务器连接起来的高速专用数据网络，可以提供企业级的数据存储服务，其所有通信在与应用网络独立的专用网络上完成。SAN 的交换式架构将存储设备通过交换机连接到各服务器上，为访问存储单元的路由提供了冗余度，消除个别交换机损坏导致的单点失败。SAN 以数据块的形式提供对共享数据的访问，具有很高的可靠性和很强的连续处理业务的能力。SAN 系统通常由接口、连接设备和通信控制协议三个基本组件，以及附加存储设备和 SAN 服务器组成。

SAN 的优点是：综合了网络的灵活性、可管理性和可扩展性，提高了网络的带宽和存储可靠性。它的运行备份操作不占用服务器网络带宽，其光纤接口提供了 10 公里的连接长度，使物理上分享的存储变得更容易。但是它的初始建设成本比较高，适合于有海量数据且具有良好经济实力的档案管理系统使用。

## 4. 内容寻址存储(CAS)

CAS 设备中组织和存储的数据方式对外部系统是看不见的，对用户来说是存储数据的“黑盒子”，用户不关心它如何存放数据，只需要通过 CAS 提供的应用程序接口 API (Application Programming Interface)来实现对 CAS 设备上的存取或访问。DAS 和 NAS 是基于文件访问的，SAN 是基于数据块寻址的，比较适用于 TB(万

亿字节)级数量的网络应用环境,而 CAS 则是采用寻址方式进行数据存储,主要是针对非结构化、固定内容、静态数据(如文档、电子邮件、影像、Video 流媒体、CAD 图纸及各种数据交易历史记录等)等内容对象的存储而设计的,由于内容数据的大小是没有上限的,因此 CAS 使用一个内容地址来存放和读取此内容对象。

CAS 的优点是:可以确保数据的完整性、有效性和安全性,为每个文件设置保存期,可预防人为的修改和删除,利用结构语义防止技术变化带来的影响。所以 CAS 技术适用于内容固定、永久保存、使用频率不高的最终形式的数字化档案信息的存储。

档案管理系统在解决存储问题时,通常可以考虑将 NAS、SAN 和 CAS 存储技术联合使用,采用分级存储方案,对共享和写作高可用性的文件采用 NAS 存储技术,将访问频率和访问速度要求较高的档案信息存储在 SAN 构建的网络中,将长期保存和访问频率较低的档案信息存储在 CAS 存储设备中。

### 2.3.3.2 信息安全技术

网络环境下的档案管理系统能够充分满足用户的需要的同时,也将整个档案系统置于一个复杂的网络运行环境中,档案信息在形成、处理、传输、存储、维护和利用等各个阶段都有受到安全威胁,而档案作为记录历史的凭证,保证其安全性和真实性是对档案管理系统最基本的要求。所以档案管理系统应该从以下几个方面保证档案信息的安全:

#### (一)保障网络安全

保障网络安全既要保障基础设施和网络设备等硬件的无故障运行,又要保障用户身份的合法性以及操作的合理性。常用的网络安全技术主要包括防火墙、入侵检测技术和虚拟专用网络。

防火墙技术通过在内部网络和互联网之间建立一个安全网关,增强内外网之间的访问控制,防止发生不可预测或潜在危险的入侵。它通过监测、限制经过防火墙的数据流,对外部屏蔽内网信息,来实现网络的安全。常见的防火墙可以归为三类,即包过滤防火墙、双宿网关防火墙和屏蔽子网防火墙<sup>[25]</sup>。防火墙重在对内外网之间的访问控制,可以抵御来自外部的非法访问和病毒入侵,但无法抵御来自内部的攻击。

入侵检测技术是一种主动的网络安全防护措施，侧重于监测、监控和预警，是防火墙的有效补充，该技术有 3 种主要的入侵检测体系结构<sup>[26]</sup>：基于主机的入侵检测系统、基于网络的入侵检测系统和混合分布式入侵检测系统。它通过监视计算机网络或计算机系统的运行，从中发现是否有违反安全策略的行为和被攻击的迹象。一旦发现，能够发出报警并采取相应的措施，如阻断、跟踪和反击等。

虚拟专用网络（Virtual Private Network, VPN）<sup>[27]</sup>的工作原理是利用公用互联网作为信息传输媒介，通过附加安全隧道、用户认证和访问控制等技术从而实现与专用网络类似的安全性能，保证对重要档案信息的安全传输。

## (二)保障系统安全

系统安全包括操作系统、应用软件和数据库系统的安全等。保证操作系统的无漏洞运行，需要使用正版软件，做到及时升级、打补丁，同时定期检查系统的各项设备，建立基于操作系统的防御监控系统。对于计算机应用软件，由于其不可见性、易修改性以及复杂性，使得在其生命周期中存在漏洞，因而特别容易受到攻击。而档案管理系统是特定的应用软件，它的安全主要取决于：是否合法用户在合法权限中进行了合法操作，目前主要用到的技术包括身份认证技术、访问控制、数字水印技术等。

身份认证技术用于确认用户的身份是否被合法授权，它主要通过标识和鉴别用户的身份，防止虚假用户获取访问权限。目前主要的身份认证技术有口令识别(即密码)和生物识别(如指纹识别、虹膜识别等)

访问控制技术的目的是限制用户的操作范围和权限，包括功能权限制和数据权限制,从而防范用户的越权访问,同时禁止未授权的用户访问有关资源。它与身份认证相结合，以确认合法用户对信息的访问范围和操作级别，如对文件的读、写、修改或执行等功能权限制以及对不同安全级别的数据库访问权限的限制。

数字水印技术(又称信息隐藏技术)的目的是：使带有数字水印技术的信息在传输过程中，即使被非法截获，也不能正常使用。含有数字水印技术的数字签章、数字签名可以保障信息安全合法地传递以确保文件内容的真实性与完整性。

## (三)保障数据安全

数据安全是为了保障在传输、存储和利用的过程中档案数据的真实、完整和有效。对于保密和绝密的数据应采用物理隔离，不提供网上操作，同时利用异地

备份避免地震、火灾等自然灾害对档案信息的破坏，建立灾难备份系统。对于网络环境中的档案信息常用的数据安全技术包括数据加密、硬盘加密、文件加密等措施。

数据加密技术是防止档案信息被恶意更改和破坏的主要技术手段，包括数据传输加密技术、存储加密技术。数据传输加密用于防止传输过程中的窃听、泄漏、篡改和破坏，通常包括线路加密(通过对信息传输线路采用加密密钥)和端对端加密(在信息发送端加密使其作为不可识别数据传输，在目的地自动解密为可读数据)等。信息存储加密包括文件加密、数据库加密。

#### (四)病毒防范

网络环境下，病毒会随时产生和传播，而计算机操作系统和应用软件总存在不尽完善的地方，成为病毒入侵的薄弱环节。病毒发作可能造成系统性能降低、数据被破坏等严重后果，因此病毒防范成为档案信息安全技术中重要的一个环节。计算机病毒防治包括病毒预防(通过对系统监视并判断病毒是否存在，阻止病毒对系统的入侵，如引导区保护、加密可执行程序、系统监控与读写控制等)、病毒检测(对病毒特征进行分析判断，如文件长度、关键字等)和病毒删除(对病毒进行分析，自动删除病毒程序并恢复原文件)。在档案管理系统的运行环境中需要从技术和管理等多方面建立完善的防病毒体系，包括建立网络化的病毒防范体系，及时更新升级病毒库，在 workstation 采用无盘 workstation、防毒卡或防病毒软件，在服务器和 workstation 实行查毒扫描、检测、隔离、报警，并及时清除病毒<sup>[28]</sup>。

### 2.3.3.3 信息检索技术

信息检索本质是将信息资源与信息需求进行匹配的过程，它基于用户的查询请求通过检索算法把需要的信息找出来提交给用户。信息检索技术一般分为三类<sup>[29]</sup>：全文检索、数据检索和知识检索。全文检索以关键字来表达信息需求，检索技术主要是基于关键字的匹配而不涉及语义，因此容易返回大量无关信息；数据检索中检索需求信息和被检索信息都以一定的格式或结构来表达，根据数据信息的组织方式不同应采取不同的检索技术即分类查询的方式实现，如顺序结构文件采用顺序检索、分块查找法、两分检索等。由于分类体系的复杂性使其存在一定的局限性。知识检索是基于语义的检索，重点在于知识语义的匹配，它提供给用

户的查询结果是包含语义和知识的，其查准率和查全率都比前两者有更好的保证，将在“知识检索技术”小节中再作论述。

#### 2.3.3.4 服务导向体系架构技术

面向服务的体系架构 SOA (Service-Oriented Architecture, SOA)是一个组件模型，它将应用程序的不同服务功能单元通过定义良好的接口联系起来。采用统一方式定义的接口独立于实现服务的硬件平台、操作系统和编程语言，从而使不同系统中的服务单元可以通过通用接口进行交互。这种没有强制绑定在特定的环境中的接口定义方式称为服务间的松耦合，松耦合系统的优点在于它的灵活性，当各个服务的内部结构和实现发生改变时，整个应用程序可以继续存在。

SOA 是一种新型的能够支持系统架构者从容面对业务需求变化的组件开发和功能部署模式，它使设计者以更迅速、更可靠、更具重用性的方式来架构整个信息系统。SOA 的优点是能够在新增的和现有的应用之上创建应用；能够使客户免于服务需要改变所带来的影响；能够升级单个服务而无需重写整个应用，也无需保留不再需要的现有系统。

基于 SOA 技术架构的档案管理系统，关键在于提供各种“服务”，即在系统中精确定义的、封装完善的、独立于环境和状态的函数。服务层是 SOA 的基础，服务组件可以直接被应用系统调用，从而将应用系统设计为服务的集合，充分考虑服务的重用。基于 SOA 架构的档案管理系统，其主要目标是实现 IT 资产的最大化重用，通过提供服务组件的方式来灵活部署档案管理系统业务功能，使基于统一档案管理系统服务平台的不同实体档案馆方便开展满足各自服务要求的特色综合业务工作。

## 2.4 知识型档案管理系统

知识型(knowledge)档案馆管理系统是档案管理系统发展的第三个阶段，这一阶段是目前档案管理发展的新方向。知识型档案管理系统最显著的特征是能够自动捕捉网络环境下的档案信息，并按照一定的规则将档案信息以知识的方式系统化和自组织化，以可利用的知识形式呈现在利用者面前，使档案管理由原来单纯

的“社会记忆”变为可以思考的“社会大脑”，即具有丰富的知识发现、知识创造和知识应用能力。这一概念的雏形是 1993 年日本的信息处理开发中心在《日本计算机季刊》第 92 期上提出的“知识档案馆项目”及建设知识档案馆应当具备的条件<sup>[5]</sup>：

1. 信息(自动)采集(automation of the acquisition and collection)；
2. 信息自组织(self-organization of knowledge)，使采集的大量信息能够自动储存；
3. 知识(自动)创造功能(creation of new knowledge)，能够在大量检索的基础上根据需要提供可用信息以及相关建议；
4. 信息编译与传输功能(translation and transmission of knowledge)，促进信息的交流与共享。

由此可见，这一阶段档案管理系统建设需要在管理理念和方法上有新的突破和发展，同时需要先进的计算机技术、信息技术和网络技术作为支撑，实现人文科学与信息科学的融合。由于知识型档案管理系统的发展还处于初级阶段，相关理论还不成熟，本文对于“知识型档案管理系统”理解是：对信息型档案管理系统进行知识化扩充，实现档案资源的知识化组织与管理，它可以看作是“知识管理”与“信息型档案管理系统”的结合，与前一阶段的档案管理系统最大的区别在于引入了知识的内涵，因此对档案信息的组织不再以简单的数据库或数据仓库的形式，而是以知识库的形式对档案进行组织。

由于知识型档案管理系统的发展并不成熟，目前还不存在被档案界认定且具有权威性的参考模型，因此本文在此主要介绍与此相关的知识管理相关理论。

## 2.4.1 知识及知识表示

### 2.4.1.1 知识

#### (一) 知识的分类

根据知识能否清晰地表述和有效的转移，通常将知识分为隐性知识和显性知识。

- 显性知识(Explicit Knowledge)：是能用文字和数字规范系统地表达出来，以科学和实验证明等为基础，可以用编码、结构化、格式化的方式表述，易于储存、理解、交流和共享的稳定明确的知识，比如经过编辑整理的程序或者普遍原则。

- 隐性知识(Tacit Knowledge): 是高度个性化难于格式化的知识, 其背后的科学原理不明确, 不易保存、传递、掌握和分享, 比如主观的理解、直觉和预感等。

隐性知识和显性知识之间存在一定的互补性, 可以相互转换, 知识转换如图 2-8 所示, 知识管理应该充分考虑两种知识之间的转换, 既要把显性知识有效地组织管理起来, 又要为充分挖掘、共享隐性知识提供平台。

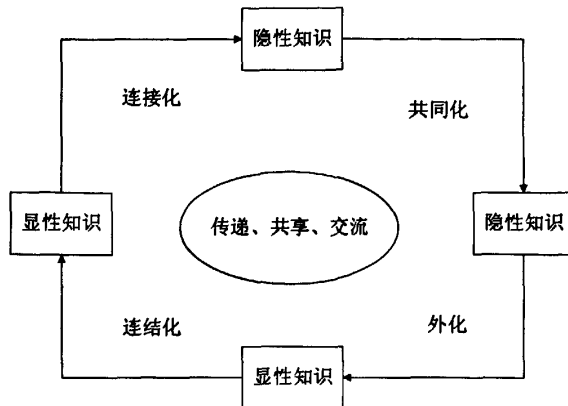


图 2-8 知识转换示意图<sup>[30]</sup>

## (二)知识的全生命周期

知识的生命周期包括以下几个环节:

- 获取: 收集可能抽象为知识的数据、信息;
- 整理: 对数据或信息进行整理加工, 形成初级知识;
- 审核: 对初级知识进行评估、审核与完善, 形成可正式发布的知识;
- 发布: 通过各种途径把知识发布出去;
- 利用: 知识被用户访问辅助于组织的日常工作;
- 更新: 知识在使用过程中不断得以改进、完善、更新和进化;
- 淘汰: 过时的知识被逐渐淘汰。

由于档案的生命周期包括产生、鉴定、归档、利用、处置和销毁, 它和知识的生命周期存在很多相似之处, 这也成为用“知识管理”的理念对档案进行的理论基础。

## (四)知识的特性

知识具有四个方面的特性<sup>[31]</sup>:

### 1. 真假性及其相对性



知识可以通过实践或推理来证明其真假，并且知识的真假是相对于一定条件而言的，前提条件的改变可能影响到知识的真假。

## 2. 不确定性

知识是复杂的，它可能介于真假两种状态之间，只能判断其真的程度，这种不确定性通常表现为不完备性(对客观事物的认识不够全面或不够准确)、随机性(知识所描述的事件发生的可能性)、模糊性(知识描述的客观事情的状态或性质不明确)和经验依赖性(对客观事物的认识依赖于已有的经验)。

## 3. 矛盾性或相容性

矛盾性是指同一知识集中的不同知识之间可能出现相互对立或不一致的情况，可能推出矛盾的结论；相容性是指同一知识集中所有知识之间互相不矛盾，可能无法推出命题的真假。

## 4. 可表示性与可利用性。

可表示性说明知识可以用适当的形式表示出来，这为知识的组织、存储、共享和利用奠定了基础。可利用性是指知识是可以用来解决问题的。知识管理正是以知识这一特性为理论基础的。

### 2.4.1.2 知识表示

知识表示是现实世界的基本代替物，是对现实世界的同态模拟，它是智能化推理的理论基础，是进行思考的有效计算环境，是描述现实世界的语言。由于知识的抽象特征及其复杂的特性，将知识恰当地表示具有一定的难度。知识的表示应该遵循充分性(有充分表示知识的能力)、可利用性、可组织维护性、可理解性和可实现性的原则。常见的知识表示方法主要有<sup>[32]</sup>：

#### (一) 一阶谓词逻辑表示法

定义：谓词由谓词名和个体组成，用于表示事物或概念的性质、状态，以及事物或概念之间的相互关系。谓词的一般形式是  $P(x_1, x_2, \dots, x_n)$ ，其中  $P$  是谓词名， $x_1, x_2, \dots, x_n$  是个体，当个体都是常量、变量或函数时，称为一阶谓词；当谓词个体为一阶谓词的，称为二阶谓词，以此类推。

例：谓词  $Teacher(Zhang)$ ，语义：张是老师；

谓词  $Great(5,3)$ ，语义：5 比 3 大。

定义：连接词是用于连接多个命题或谓词，以表示复杂语义的符号。包括： $\neg$ (非)、 $\wedge$ (且)、 $\vee$ (或)、 $\rightarrow$ (包含)、 $\leftrightarrow$ (等价)，其优先级是从高到低。

定义：量词用于对谓词中的个体做出量的规定。包括： $\exists$ (存在量词)、 $\forall$ (全称量词)

定义：谓词公式是用连接词和量词将谓词连接起来所构成的公式。

现举例说明怎样用一阶谓词逻辑表示法来表示知识。

例：高扬是计算机系的一名学生，但他不喜欢编程。

表示：  $\text{Like}(x,y)$ : x 喜欢 y

$\text{Student}(x,y)$ : x 是 y 系的学生

$\text{Student}(\text{GAo},\text{Computer}) \wedge \neg \text{Like}(\text{Gao},\text{programming})$

一阶谓词逻辑表示法的优点是：自然性，精确性，严密性，充分性，易实现，是后续三种表示法的基础；它的缺点是：不能表示模糊知识，容易组合爆炸，效率低，启发知识难利用。

## (二) 产生式表示法

系统结构的组成要素包括：规则库(知识)、数据库(存放信息)、控制系统(规则的解释或执行程序)

产生式表示法可以描述事实性知识(事实)和过程性知识(规则)。它通常用三元组表示：

1. 用三元组(对象, 属性, 值)表示事物的属性, 如(雪, 颜色, 白)表示“雪的颜色是白的”。
2. 用三元组(关系, 对象 1, 对象 2)表示事物之间的相互关系, 如(热爱, 王, 祖国)表示“王热爱祖国”。

产生式规则的基本形式是： $P \rightarrow Q$  或  $\text{If } P \text{ then } Q$  (可信度)

产生式表示法的优点是：自然性、有效性、一致性；它的缺点是：效率较低、不能表示结构性知识。

## (三) 框架表示法

框架表示法是在框架理论的基础上发展起来的一种结构化知识表示方法。框架通常由若干个称为“槽”的结构组成，每一个槽根据实际情况拥有若干个“侧面”，每一个侧面可以拥有若干个“侧面值”。

框架的一般形式是

<框架名>

槽名(Slot)1:	侧面(Facet)名 1:	值 1:
	...	
	侧面名 m1:	值 m1:
	...	
槽名 n:	侧面名 1:	值 1:
	...	
	侧面名 mn:	值 mn:
约束:	约束条件 1:	
	...	
	约束条件 n:	

常用的槽名有:

ISA 槽 (is a...): 下层可以继承上层, 表下是上的特例;

AKO 槽 (A kind of...): 类属关系, 下层可以继承上层;

Subclass 槽: 子类与类, 子集与集, 下层可以继承上层;

Instance 槽: 指出下层框架有哪一些, AKO 槽的逆, 可继承;

Part-of 槽: 部分与全体, 通常不可继承 (如教室与黑板);

Infer 槽: 两个框架间的逻辑推理关系;

Possible-Reason 槽: 结论与可能的原因关系。

基于框架表示法的推理是通过对框架的匹配与填槽来实现, 如果框架可以找到合适的填充值, 就把值填入框架, 否则要选择新的框架。如果两个框架相对应的槽没有矛盾, 可以认为它们可以匹配。

#### (四) 本体表示方法

本文将在第三章重点讨论本体理论, 并以电力企业为背景, 实际构建其档案本体。

## 2.4.2 知识管理的组成要素

为了将知识有效地进行管理,必须先获取知识,再将知识适当组织起来存储在计算机中,然后知识才能用于解决实际问题。所以,知识的获取、知识的表示和知识的应用成为知识管理的三大主要研究对象<sup>[33]</sup>。

### 1. 知识的获取

对知识获取的研究主要包括:从已有的经过组织后的知识载体(数据库、知识仓库、知识地图等)检索知识的过程,主要涉及检索技术;从未经组织的非结构化数据(网络资源、媒体资源等)中进行知识发现的过程,主要应用信息抓取和数据挖掘技术。

### 2. 知识的表示

知识表示研究的主要是知识如何形式化、如何组织的问题。要将知识在计算机中以逻辑形式进行表示,并最终形成编码应用到计算机中去。经过恰当表示的知识能够形成统一描述的知识模型,应不受用户改变的影响,达到知识共享的目的。知识的表示应该既能表示事物间层次结构的静态知识,又能表示对事物操作处理的动态知识,它包含语义,且不易形成知识孤岛。本体的引入满足了知识表示的需求,本文将在第三章重点讨论如何用本体表示档案领域知识。

### 3. 知识的利用

对知识的利用以产生各种效益是知识管理的最终目的,也可以理解为知识服务或知识共享。它的研究内容主要涉及隐性知识和显性知识的转换,知识地图导航,知识库建设、共享环境或平台的搭建等。而知识库的建设是基础,知识表示是前提,它们决定了知识利用的宽度和深度。

## 2.4.3 知识管理系统

知识管理系统(knowledge management system, KMS)是为了实现知识管理的而设计的计算机应用系统, KMS 是一个具有知识管理能力的系统,它是实现知识管理的工具。根据前文所讨论的知识管理相关理论,本文设计了基于本体的知识管理系统模型,如图 2-10,它由知识获取层、知识组织层和知识应用层组成:

- (1) 知识获取层：将结构化、半结构化和非结构化的信息从异构环境中抽取出来，经过分析鉴定，在领域专家的指导下，通过本体编辑器自动或半自动地编辑成本体传递给知识组织层。
- (2) 知识组织层：接收经编辑后的知识，并将新增本体与已有的本体库进行匹配、推理等操作后添加到本体库中，对本体库进行更新和进化。
- (3) 知识应用层：接收并分析用户的服务要求，利用查询检索、推理、推送等算法对知识库进行操作，将相关的知识组织成有机统一体，并传递给需求用户，以满足用户对知识服务的需求。

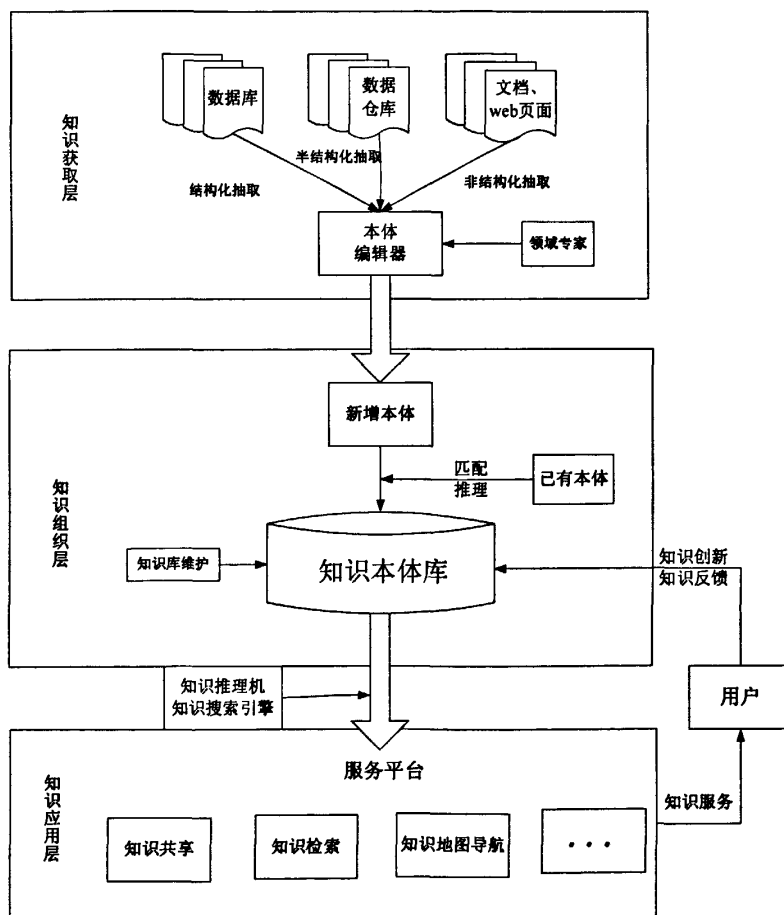


图 2-10 基于本体的知识管理模型结构图

#### 2.4.4 知识型系统关键技术

目前在实施知识管理方面存在的主要问题包括：知识的分散，知识可能分散在各个管理系统甚至管理人员的头脑中，怎样实现知识的共享与交流变得很困难；

知识的无序存储，这使得当需要知识时，对知识的检索十分困难；新知识的累积，由于知识的产生速度相当快，如何将新的知识与原有知识进行匹配也很重要。从技术角度出，考虑到知识管理的关键技术<sup>[34]</sup>，可以分析出知识型档案管理系统中主要涉及的关键技术包括：

#### 2.4.4.1 知识获取

业务管理系统是对企业日常管理工作的支持，其中包含并产生大量信息资源，是档案知识的重要来源，业务管理系统和知识管理系统的融合将以分布式网络环境为支撑，以知识库的建设为前提，以数据采集和知识集成为保障，实现知识化的管理。为了将业务管理系统中产生的档案资源采集到知识型档案管理系统中去，可以通过设置接口的方式实现。

##### ● 与馆藏档案的接口

馆藏档案馆收藏了企业许多具有重要价值的档案资料如企业发展规划、工作总结报告、企业重大活动的记录等等，这些档案资料也需要在档案管理系统中收藏。需要开发数字化接口，实现馆藏档案的数字化。

##### ● 与业务职能部门接口

业务职能部门管理人员在日常工作中积累的工作经验或相关知识，是企业重要的档案资源，需要开发相关的接口，以方便他们使用规范的方法，将显式或隐式的工作经验或相关知识转化为档案并管理和保存起来，实现知识的积累和共享。

##### ● 与数据中心接口

企业的数据中心是知识型档案管理系统中主要的档案资源来源，通过企业数据服务总线提供的数据服务，实现与数据中心的数据接口，实现档案的自动获取，提高企业档案获取的自动化水平。

##### ● 与业务应用系统接口

对尚未整合到数据中心的业务应用系统数据，知识型档案管理系统需要开发专用适配器接口获取业务应用系统的归档数据，并将档案知识传递给需要档案支持的业务系统，从而实现与业务系统归档及档案利用的自动衔接。

##### ● 与企业系统内/外接口

与企业系统内接口是指为上下级管理单位之间或同级单位之间实现档案资源

共享而开发的通用数据接口。与企业系统外接口是指为企业在生产经营活动中与系统外界往来的、具有归档价值的电子邮件开发的专用邮件数据接口，以实现邮件数据归档的自动化或半自动化。

#### 2.4.4.2 知识表示技术

利用计算机实现知识管理、知识服务需要强大的知识库支持，因此知识库建设是基石，只有建立具有关联性的知识库，知识检索、智能问答等知识利用才会成为可能。知识的表达内涵丰富，需要领域专家和信息化工作者的合作，将领域内的知识进行深层的分析、表达和组织，并整理归结成概念网络，形成知识本体达到知识共享的目的。

#### 2.4.4.3 知识地图

知识地图是一种有效的知识管理工具，它反映知识间的关联，可以清晰了解组织内知识的大致分布情况以及与组织相关的内外部知识线索。因此，知识地图是知识管理的重要部分，它为寻找知识提供线索，能够根据知识需求指向知识源，知识源的内容不包含在知识地图本身，而是固化在一定载体上的文献、数据库(显性知识)或者某领域工作者的大脑中(隐性知识)。知识地图提供给用户的包括知识存储地、与其相关联的新知识或新关系，因此可以帮助用户实现知识的定位和共享，将知识与其需求者相关联。

知识地图的内容包括两个方面：一是知识资源的目录或者层级，二是各目录、各层级间资源间的关系。因此，一份完整的知识地图通常能够提供知识资源的来源地、关键词、访问权限等和包含该资源的所有类型的知识，如文档、政策、经验等，以及它与其他知识资源的关联。

知识地图分为静态地图和动态地图，静态地图绘制成后在未来一段时间内其组织信息和状态不会改变和更新，动态地图可以根据知识的变化随时更新其中的知识层次或相互关系。

#### 2.4.4.4 知识检索技术

为了将恰当的知识传递给知识需求者，知识地图提供了一个导向和基础，知识检索提供了寻找知识的工具。基于关键词匹配的全文检索和搜索引擎是目前

主流技术，技术成熟稳定但查准率不高，由于它的匹配是非语义匹配，因此可能包含一些无用信息。而知识检索技术<sup>[35]</sup>是以知识组织为基础的，知识组织中充分表达了知识的内容及其相互之间的关系，包括静态的语义、逻辑关系和动态的操作、控制关系，由于知识检索实际上是模拟了人类关于知识处理和利用的行为，因此它可以实现知识服务，向用户提供隐性知识，提供智能和个性化的服务。目前主要的知识检索模型包括分类检索模型、认知检索模型和分布式检索模型。

#### ● 分类检索模型

分类检索的核心思想是利用事物之间的本质关系对其进行抽象分类，分类后的概念逻辑是知识检索的基础。知识的组织采用类层次结构，类或对象之间的从属或属性关系是对其语义的表示。因此该模型综合应用对象归类和类结构的查找方法，利用对象之间的等级关系和继承原理，可实现快速的自顶向下的检索。它最大的优点是有语义继承特性，提供多维组合信息给用户，如对象的邻节点对象、父节点对象的信息，适合于领域知识和用户知识等大规模的知识背景。

#### ● 多维认知检索模型

认知检索模型的理论基础是人工神经网络，它以模拟人脑结构和神经系统为目标，将知识资源组织为语义网络结构。这种模型利用启发式和传递式技术，形成语义推理，实现认知式联想搜索，利用学习规则和反馈技术对知识的组织不断修改和完善。它适合于专业概念知识的检索，不适合于大规模知识库。

#### ● 分布式检索模型

分布式检索主要用于对网络异构分布式信息资源的检索，它通过建立元知识库对知识进行集成、管理和存储，元知识库包含用户专家知识、控制调节知识、优化检索协议以及描述知识内容的元数据等。它综合应用神经网络、并行推理、数据挖掘等技术，对资源及其相关性进行评估，选择最优资源，再利用了聚类、综合分析等方法，将最优结果提供给用户。

### 2.4.4.5 知识发现技术

知识发现(Knowledge Discovery in Databases, KDD)，又称数据库中数据挖掘(Data Mining, DA)<sup>[36]</sup>，是从大量的、不完全的、有噪声的、模糊的、随机的实际



应用数据(数据库、文本文件、图形、图像、音频和视频等电子文件)中提取隐含的、事先未知的但有潜在应用价值的信息和知识的过程。

数据挖掘的步骤:

- 确定对象: 即明确问题, 确定挖掘目标。
- 数据准备: 包括数据选择、数据预处理、数据转换。
- 数据挖掘: 利用合适的挖掘算法对转换后的数据进行挖掘。
- 结果分析: 解释并评估结果, 根据挖掘目标使用合适的分析方法, 得到用户既定形式的结果, 如可视化结果。
- 知识同化: 将分析得到的知识集成到业务系统的组织结构中。

数据挖掘的分类:

由于不同的技术用于解决不同的问题, 所以人们在研究数据挖掘技术时根据数据挖掘研究对象的数据结构形式, 通常将数据挖掘分为结构型数据挖掘、Web 数据挖掘和文本数据挖掘。

- 结构型数据挖掘: 主要面向结构化的数据即 SQL Server、Oracle 等数据库或数据仓库。常用的技术和方法主要有: 统计方法、基于事例的推理、规则推理、关联分析和序列模式算法、神经网络、决策术、遗传算法、可视化技术、模糊集、粗糙集等。数据挖掘主要应用于预测、回归、分类、聚类、关联分析、探索性数据分析、异常检测等方面。
- Web 数据挖掘: 主要面向的对象是半结构化的 Web 数据, 可分为: Web 内容挖掘、Web 结构挖掘、Web 使用记录挖掘, 其目的是从海量的 Web 数据提取有益的信息和知识, 以便改进站点设计、更好的开展信息服务, 常用的技术和方法除了上述的方法还有路径分析、序列模式等。
- 文本数据挖掘: 主要面向的对象是巨量的文本信息, 它包括特征提取、文本摘要、文本分类与聚类等工作, 所用到的技术和方法主要有词串表示法、贝叶斯分类算法、词集合算法、文本聚类算法、最近邻参照分类算法等。

## 第三章 电力企业知识型档案管理系统的知识表示

企业经常把知识当作一项新的“生产要素”，是除了人力、财力、物力以外的企业第四大财产，而档案作为知识资源的重要来源，在企业管理中倍受关注。如何将档案以知识的方式表示并组织管理起来，是企业知识管理的重要内容，也是建设知识型档案管理系统的基礎。

由于知识型档案管理系统区别于传统档案管理系统的关键在于将档案以知识形式表示，因此知识表示技术成为建设知识型档案管理系统的关键技术之一。知识表示是一种数据结构与控制结构的统一体，要同时考虑知识的存储与使用，近年来，在知识表示领域，对本体的研究越来越受到重视，而基于本体的知识查询和检索在查全率和查准率方面都有存在很大优势，也是其他知识服务如知识地图导航、知识协作等的开发基础。

本章从本体理论出发，提出档案本体建设的方法和流程，并以此为依据，给出某大型电力企业知识本体建设的实例，对实际系统中档案知识本体建设具有指导意义。

### 3.1 本体理论

知识管理系统中，知识表示是最基本、最关键的技术，知识的表达及组织是否合理，关系到系统建设能否达到预期标准，以及知识服务能否有效进行。而本体是从语义层面对知识进行规范性描述和组织，具有概念层次结构且支持逻辑推理，成为知识表示的良好选择。

#### 3.1.1 本体的概念

本体论(ontology)又称为实体论，源自于形而上学的哲学分支，它对客观世界的事物进行分解，发现其基本的组成部分，进而研究客观事物的抽象本质。1993年，Gruber给出了本体的一个定义，即“本体是概念模型的明确的规范说明”<sup>[37]</sup>。后来，Borst在此基础上，给出了本体的另外一种定义：“本体是共享概念模型的形式化规范说明”。Studer等对上述两个定义进行了深入的研究，认为本体是共享概

念模型的明确的形式化规范说明。这包含 4 层含义：概念模型(conceptualization)、明确(explicit)、形式化(formal)和共享(share)。

- 概念模型：是指通过对某个客观事物或现象的概念进行辨析或提取从而获得的关于该事物或现象的抽象模型，其含义独立于具体的环境状态；
- 明确：是指对所使用的概念类型，以及概念之间的约束或关系给予明确的说明；
- 形式化：以精确的数字或编码来表示本体并以计算机可读的形式存在；
- 共享：本体中体现的是共同认可的知识，反映的是该领域公认的概念集，不针对个体特殊环境，从而具有交流共享的能力。

总之，本体的目标是捕获表示相关领域的知识，它反映了给定领域的知识，是该领域中的概念以及这些概念间关系的集合。一般来说，本体具有两个特征：静态性和动态性——静态性指的是它反映的概念不涉及动态行为；动态性指的是它的内容和服务对象可以不断变化和更新的。

### 3.1.2 本体的分类

由于本体的研究和开发是在不同层次上进行的，因此根据本体的研究层次，本体的类型大致可以分为如下几种类型<sup>[38]</sup>：

- 通用本体(Generic Ontology)：研究非常通用的概念，如时间、空间、事物、对象、事件、行为等，他们独立于特定的问题或领域。
- 领域本体(Domain Ontology)：研究与特定领域相关的术语或词汇，如医学、教育、企业、档案等，它包含特定领域的相关知识，提供特定领域的概念定义和概念之间的关系，提供该领域中的主要理论和基本原理等。一般来说，针对不同的领域可以定义和构造不同的本体。
- 任务本体(Task Ontology)：主要定义通用任务或推理活动，涉及动态知识。任务本体主要研究可共享问题的求解方法，其实质是从推理和问题求解的角度刻画领域知识。
- 应用本体(application Ontology)：描述依赖于特定领域和任务的概念及概念之间的关系。它主要用于描述特定的应用，既可以引用特定领域本体的概念，又可以引用任务本体中的概念。

### 3.1.3 本体的组成

Perez<sup>[39]</sup>归纳了本体的基本建模元语(Modeling Primitive), 分别是: 类(classe)、关系(relation)、函数(function)、公理(axiom)和实例(instance)。通常把类称为概念(concept)。

- 类或概念: 概念的定义非常广泛, 可以指代任何事物, 如描述、功能、行为、策略和推理过程等, 本体中的这些概念一般构成一个分类层次。
- 关系: 领域中概念之间的相互作用。例如: 父子类关系和属性关系。
- 函数: 一类特殊的关系, 在这种关系中, 前  $n-1$  个元素可以唯一决定第  $n$  个元素。例如 Father-of 关系就是一个函数, 其中 Father-of( $x,y$ )表示  $y$  是  $x$  的父亲, 显然  $x$  可以唯一确定他的父亲。
- 公理: 代表永真命题, 是定义在“概念”和“属性”上的限定和规则, 比如概念甲属于概念乙的范围。
- 实例: 属于某概念类的基本元素, 即某概念所指的实体。

### 3.1.4 本体的应用

本体正在应用于人工智能、信息管理、知识管理、知识工程、自然语言处理、信息检索、语义 Web 等各个领域。典型的本体应用包括以下几个方面<sup>[40]</sup>:

1. 基于语义的信息检索, 如网络搜索引擎和数字化图书馆<sup>[41]</sup>。
2. 基于本体的数据集成、机器学习等<sup>[42]</sup>。
3. 领域本体的应用。比如, 在生物信息学中已经建成的 Gene Ontology, 它对生物信息学界已经有巨大的影响。
4. 语义 Web 服务<sup>[43]</sup>。

由于本体通过对概念、术语及其相互关系的规范化描述, 可以对各种信息作出语义清晰的表示, 可以将非结构化、半结构化的信息转化为结构化的知识进行存储, 并且本体的层次结构模型可以方便地进行存储。本体通常表现为一棵概念化的树, 越往树的下层越具有领域相关性, 它适用于企业知识共享和重用, 当领域知识结构变化时, 可以通过本体进化并调整, 保持系统的一致性。它所提供的语义可以很大程度上提高检索的查全率和查准率, 使用本体来对知识进行表示和组织成为知识管理的首选, 也成为知识管理系统建设中的关键技术所在。

## 3.2 档案本体构建

### 3.2.1 构建原则

出于对各自学科领域和具体工程的不同考虑，构建知识本体的原则各不相同。目前没有统一的标准。但 Gruber 从实际出发，于 1995 年提出了构造本体的 5 条规则<sup>[44]</sup>，比较得到大家的认可：

- **明确性(Clarity)**：本体必须有效的表达领域术语的内在含义。本体概念的定义应该是客观的，独立于背景的。同时，它应该是形式化的，可以采用逻辑公理描述的。所有客观的、形式化的语义定义都应该使用自然语言描述。
- **一致性(Coherence)**：本体应该是一致的，即支持与其定义相一致的推理，使推理不会产生矛盾。它所扩展的公理以及用自然语言进行说明的信息都应该具有一致性。
- **可扩展性(Extendibility)**：本体应该为可预见的共享词汇提供概念基础，可以满足可预见的任务，以扩展现有的概念体系。在向本体中添加专用术语时，不需要修改已有的内容定义。
- **最小编码倾向(Minimal encoding bias)**：本体的概念模型应该被描述为知识层次，而不依赖于符号层次的编码。在本体高计中，编码应该倾向最小化，使得在不同编码或不同类型编码描述的系统中满足知识共享。
- **最小本体承诺(Minimal ontological commitment)**：本体应该要求最小本体化承诺以支持预期的知识共享行为，只要能够满足特定的知识共享需求即可。这可以通过定义约束最弱的公理以及只定义通讯所需的词汇来保证。

### 3.2.2 构建方法

目前为止,本体构建中常用的几种方法包括 IDEF-5 方法、TOVE 法、骨架法、METHONTOLOGY 法、七步法、KACTUS 法、SENSUS 法等。

#### 1、IDEF-5 方法<sup>[45]</sup>

IDEF 是在结构化分析方法的基础上发展起来的，它通过使用图表语言，获取客观存在的概念、属性和关系，并将它们形式化表达成本体。IDEF-5 创建本体的 5 个主要步骤是：

- (1)组织和范围：确定本体构建项目的目标、观点和语境，组织课题并为组员分配角色。
- (2)收集数据：收集本体构建需要的原始数据。
- (3)分析数据：为抽取本体作准备。
- (4)本体初步开发：从收集的数据当中建立一个初步的本体。
- (5)本体优化与验证：完成本体构建过程。

## 2. TOVE 法<sup>[46]</sup>

TOVE 本体建模方法是由 TOVE 项目开发出来的方法，主要包括以下几个步骤：

- (1)激发场景：明确设计动机，应用领域的某些场景激发了本体的构建，在场景中更好理解本体构建的需求。
- (2)非形式化的系统能力问题：在本体没有形式化之前对系统能够解决什么问题以及如何解决用术语表示，其答案用公理和形式化定义回答。
- (3)术语的规范化：从非形式化能力问题中提取非形式化的术语，然后用本体形式化语言进行定义。
- (4)形式化的系统能力问题：本体构建后非形式化的能力问题便可以用形式化的术语定义出来；
- (5)将规则形式化为公理：本体中的公理指定了术语的定义以及约束，将公理用逻辑形式表示出来，是对本体的规格说明。
- (6)通过对本体的检验调整方案，使本体趋于完备。

## 3. 骨架法<sup>[47]</sup>

“骨架法”研究的是独立于本体领域的企业本体构建方法，它建立在企业本体基础上，是相关商业企业间术语和定义的集合。它主要包括如下四个主要步骤：

- (1)确定本体应用的目的和范围，根据所研究的领域建立领域本体时，领域越宽所建本体越大，因此需要先限制本体的领域范围。
- (2)本体分析，确定本体中概念的意义及概念之间的相互关系，对该领域了解的越多，所建本体就越完善，因此一般需要领域专家的参与。
- (3)本体表示，一般用语义模型表示本体。
- (4)本体的评价，建立本体的评价标准是清晰性、一致性、完善性、可扩展性。

(5)本体的建立,对所有本体按上述标准进行检验,符合要求的以文件形式存放,否则重新进行本体分析,如此循环往复,直到对所有步骤的检验结果均达到要求为止。

#### 4. 七步法<sup>[48]</sup>

七步法是斯坦福大学提出的基于 Protégé 本体构建工具的一种领域本体构建方法。一共包括 7 个步骤:

- (1) 确定知识本体的专业领域和范畴;
- (2) 考查复用现有知识本体的可能性;
- (3) 列出本体中的重要术语;
- (4) 定义类和类的等级(层次)体系;
- (5) 定义类的属性;
- (6) 定义属性的分面;
- (7) 创建实例。

本文对档案的特点进行充分分析后,参考以上几种本体构建方法,提出了适合档案特征的本体构建方法包括以下几个步骤,如图 3-1 所示:

1. 确定档案本体的领域和范围,即确定哪些内容是需要作为档案收集的信息。
2. 对档案领域本体的分析建模,根据档案收集的范围,首先确定档案涉及的重点概念及其概念之间的关系,然后将这些概念及其关系用基本的框架表示出来,最后形成档案本体的概念模型。
3. 由概念模型导出形式化编码,这是因为基于档案本体的管理利用需要建立在形式化编码基础上,因此要将本体概念模型以编码的形式集成到相关算法中。
4. 对档案本体进行检验与评价,通过在整个系统中利用本体进行操作,从而检验本体的设计是否合理,是否需要进化或修正。如果需要修正则重复第 2-4 步,否则认为本体构建成功。

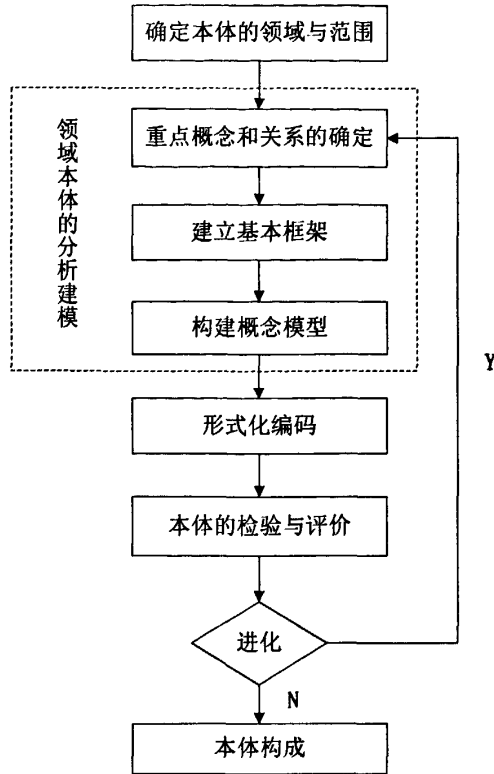


图 3-1 档案本体构建流程

### 3.2.3 构建工具 protégé

随着对本体的应用和研究越来越深入，本体构建这项庞大的知识工程也需要一些工具来帮助完成其开发任务。目前在国内常见的本体构建工具包括 OntoEdit、WebOnto、WebODE、KAON 和 Protégé 等，本文采用的是 protégé 工具。

Protégé<sup>[49]</sup>是斯坦福大学为知识获取而开发的本体构建工具，应用于知识的获取和现存本体合并和排列，因其可以免费下载并公开源代码而成为目前国内使用最为广泛的本体编辑工具。它主要有以下几个特点：

- 使用 java 语言编写，具有强大的功能插件体系和开放的模块化风格，使系统开发者可以通过生成恰当的插件以增加新的功能，方便二次开发。
- 知识模型可扩展性，使用户可以根据需要对模型进行扩充和修改。
- 导入导出功能强大，兼容性好，可以从 RDFS、带 DTD 的 XML 文件、XML Schema 等文件中导入本体，也可以将本体转化为多种形式化语言描述，如 RDFS、OWL 等进行输出。
- 具有图形化的用户界面，操作简单方便。



- 可以与数据库相连，将知识概念模型导入数据库进行存储。

### 3.3 电力企业档案本体构建

#### 3.3.1 确定本体的领域与目标

企业档案本体涉及的范围相当广泛，要建立一个完备的档案本体，其工作量很庞大，而完整性和客观性也难以得到保证，通常可以采用从部分领域作为切入点的办法，再逐步向外延扩展。因此，可以先针对部分档案建立本体，再进行扩展，逐步完善。档案本体的建设目标是将企业档案中的概念及其相关属性之间的关系进行关联，形成知识，从而实现档案领域知识化管理，为知识型档案管理系统建设奠定基础。

#### 3.3.2 重点概念和关系的确定

在确定了本体的目标和范围以后，需要通过学习档案领域的相关知识，对档案管理模式进行研究，同档案馆工作人员进行交流，从而抽取出关键的名词术语、概念，并确定它们之间的关系。根据 Gruber 提出的本体构建原则，可以采用核心扩展(middle-out)的方法建立领域知识概念模型，其表现形式为：由具有本体雏形的一组核心概念入手，不断扩展本体<sup>[50]</sup>。对于电力企业来讲，定义档案领域的核心概念集包括档案学科概念集和档案属性概念集两部分，然后根据本体的需要和评估结果进行适当扩展或修改。

##### 1. 档案学科概念集

档案学科的组织即《中国档案分类法》，是中国档案分类标引与检索标准化的工具。根据《中国档案分类法》以及电力企业独特的经营管理模式，可以将档案分成以下 17 个基本大类，再根据需要设计子类，形成等级分明、次第清楚的分体系，见附录一。

##### 2. 档案属性概念集

档案属性是用于描述档案特征的概念或类，根据档案的特征信息，定义主要的属性概念如下：

- (1) 档案分类：按照前面档案学科概念的分类，可以将电力企业档案分成 17 类，再根据需要设置子类，本文只设计到第三类分级类目，还可以继续细分，分类信息可以作为档案编码的前缀部分，使得档案的唯一标志符中包含分类知识。
- (2) 内容特征：主要描述了与档案内容相关的知识，一般包括题名、责任者、关键字、发文日期、文种、内容摘要等，可以根据知识模型的需要进行扩展。
- (3) 结构特征：由于电子文件的特殊性，使用结构特征对其进行描述也是有必要的，它包括电子档案大小、档案格式、建立时间、属性等。
- (4) 管理过程：它将档案知识的静态利用与动态利用进行关联，主要描述档案的生命过程，包括鉴定信息、封装信息、权限管理、归档处置、备份历史以及利用历史等。

### 3.3.3 建立基本框架

通过列举领域中的概念并分析概念之间的基本关系，得知档案本体中所包含的主要类、属性和关系，并设计档案领域本体的基本框架如图 3-2：

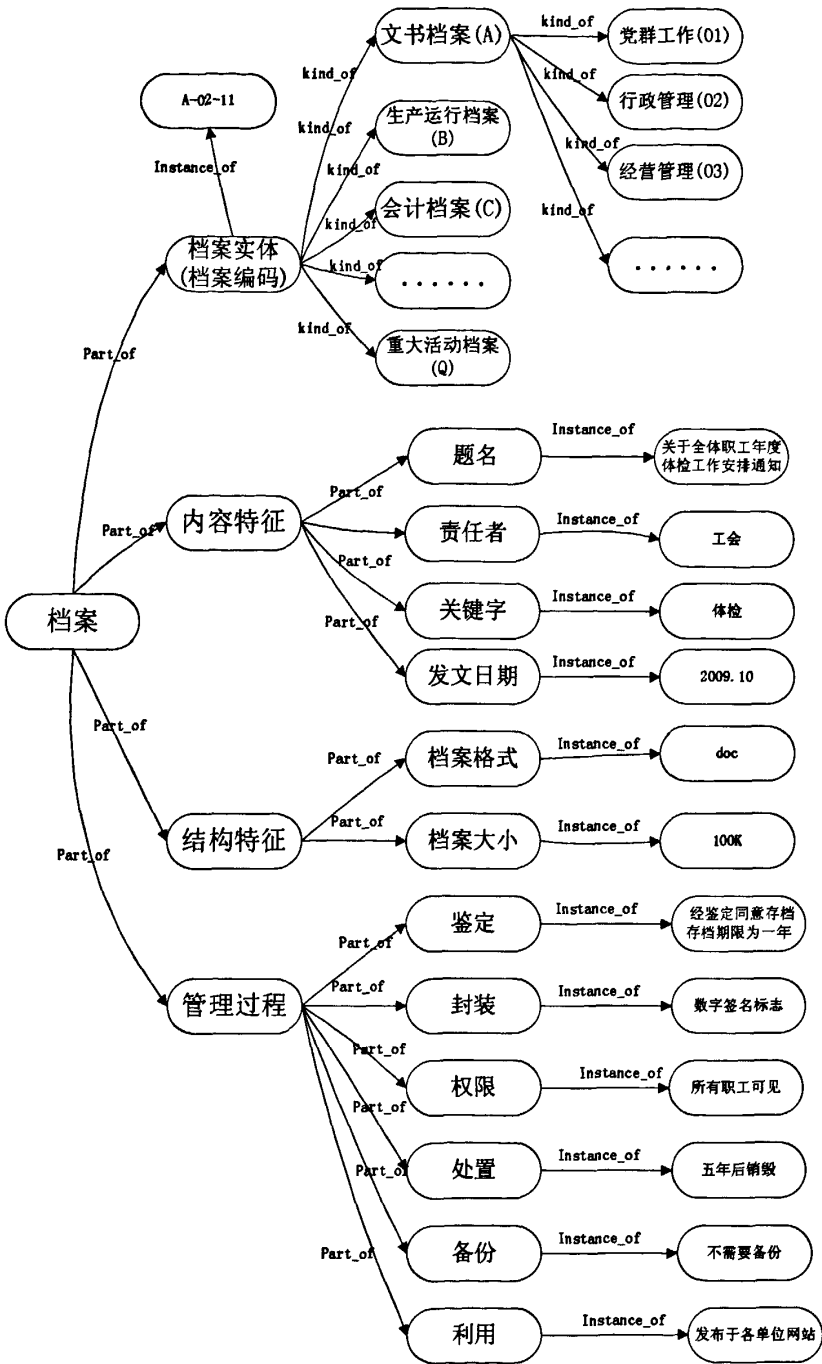


图 3-2 档案本体基本框架

通过对档案系统所包含的主要概念间关系的分析，我们发现概念之间主要有如下四种关系：

(1) part-of: 表达概念或类之间部分与整体的关系。

例如：档案是整体，内容特征是部分。

(2) kind-of: 表达概念之间的继承关系, 也即上下位的关系。

例如: 文书档案是档案的一个基本类, 而党群工作又是文书档案的一个子类。

(3) instance-of: 表达概念的实例与概念之间的关系。

例如: A-02-11 是档案编码的一个实例, 表示文书档案-行政管理类的 11 号档案。

(4) attribute-of: 表达某个概念是另一个概念的属性。

例如: 结构特征是档案的一个属性。

### 3.3.4 在 protégé 中构建概念模型

根据已有的本体框架, 可以利用 protégé 工具构建概念模型, 考虑到整个电力企业档案本体涉及范围广, 很难保证其完整性和正确性。本文以档案分类中的部分类别(会计档案、建设项目档案、生产运营档案、文书档案、信息化建设档案)及部分关系为基础建立电力企业档案本体的概念模型, 由于 protégé 具有很强的扩展能力, 根据类似的方法扩展成为企业完整的档案本体概念模型。

在 protégé 中构建本体, 一般包括以下几个步骤: 建立文件(project)、建立类(class)及类层次、建立属性(property)及属性值、添加实例。使用 protégé 工具建立电力企业数字档案本体的结构如图 3-3:

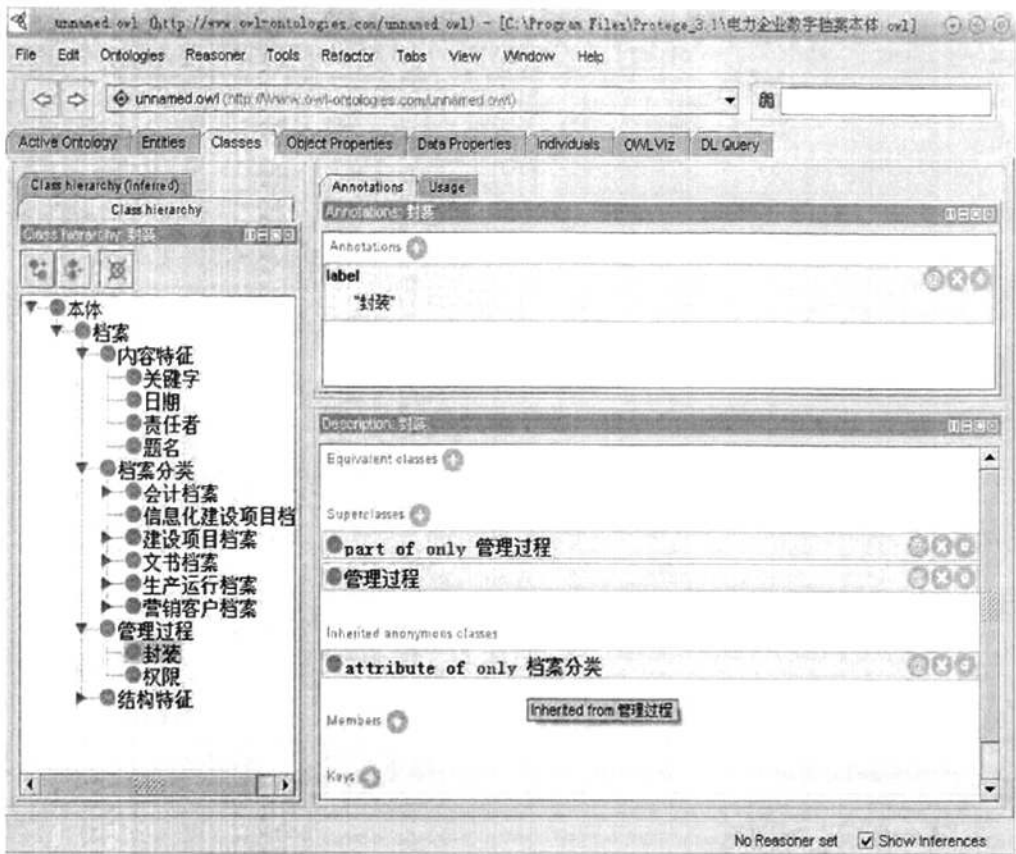


图 3-3 利用 protégé 定义的电力企业档案本体界面图

由 protégé 工具生成的电力企业档案本体概念模型的部分截图如下：

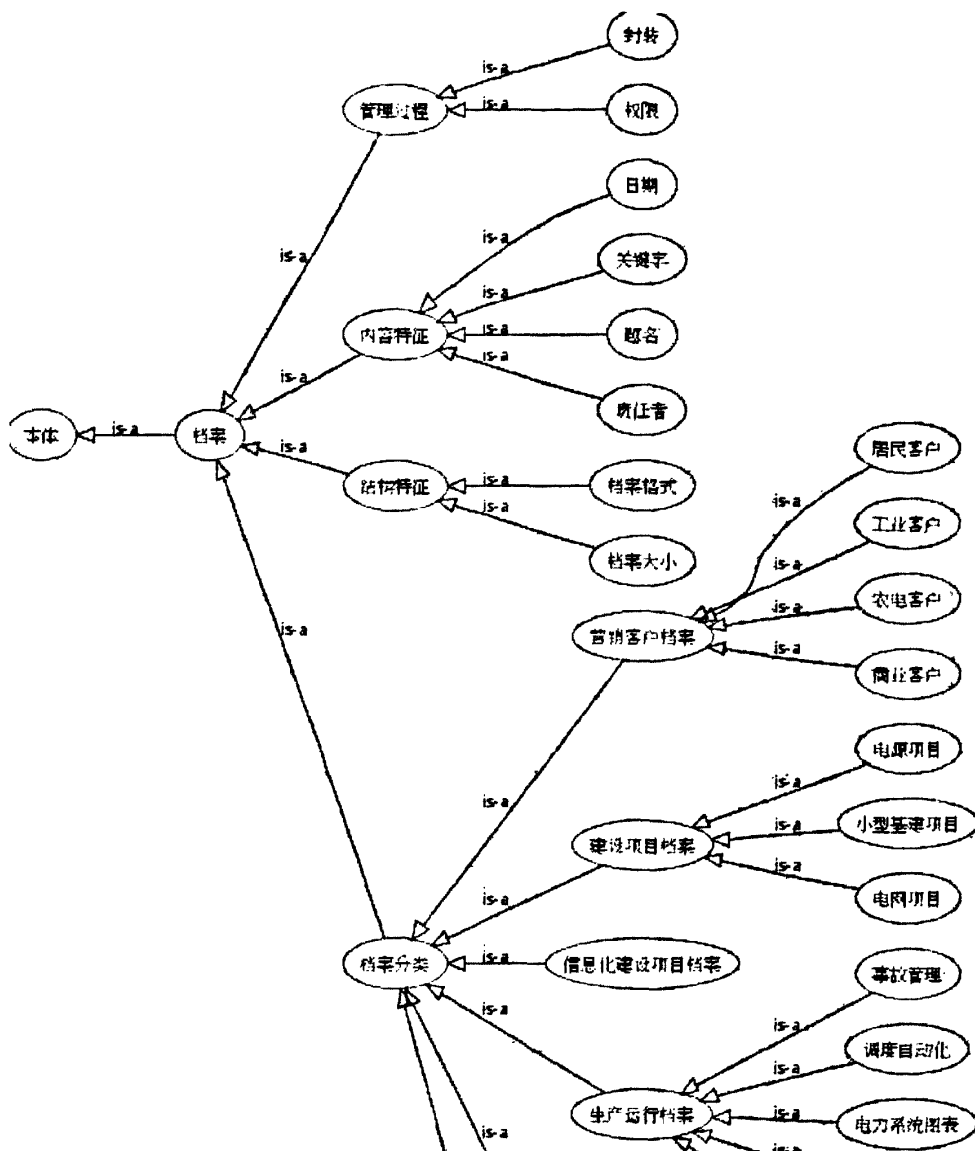


图 3-4 protégé 工具生成的电力企业档案本体关系图

### 3.3.5 形式化编码

常用的本体描述语言是 OWL (Web Ontology Language)，它是 W3C 开发的一种网络本体语言，用于对本体进行语义描述。由于 OWL 是在 DAML+OIL 的基础上改进而开发的，所以一方面要保持对 DAML+OIL/RDFS 的兼容性，另一方面又要保证更加强大的语义表达能力，同时还要保证描述逻辑 (DL, Description Logic) 的可判定推理。

使用 protégé 工具构建的电力企业档案本体，可以生成 OWL 文件，部分 OWL 文件见附录二。

## 第四章 电力企业知识型档案管理系统建设的研究

随着信息化技术的不断发展，企业在信息化发展的道路上展开了激烈的竞争，信息化目标的定位不再仅仅是提升企业经营管理业务的自动化水平，更需要不断提升企业经营管理决策分析的智能化水平。而决策分析是以企业经营管理过程的历史与现状作为信息支撑的，档案是记录历史与现状的信息载体，所以档案管理系统成为决策分析的基础。企业档案管理系统建设不仅要传统档案管理工作自动化，更需要结合企业的知识化管理，从档案资源知识化的角度考虑以实现对档案资源的知识化组织与管理，使其与业务系统对接，实现对日常生产经营管理工作的知识支持。

第二章中讨论了基于 OAIS 的数字档案管理系统，它由四个应用系统(馆藏建设系统、管理系统、发布系统和 workflow 系统)和两个存储系统(元数据仓储系统和档案存储系统)组成；同时讨论了基于本体的知识管理系统，包括知识获取层、知识组织层和知识应用层。

本章以大型国有电力企业的档案管理业务为背景，结合知识管理理论，提出了知识型档案管理系统的参考模型 KAMS (Reference Model of Knowledge Archives Management System)，并以此为基础，从业务管理模型、系统功能模型、功能实施模型和系统服务架构几个方面对电力企业知识型档案管理系统建设作了相应研究。

### 4.1 知识型档案管理系统参考模型(KAMS)

#### 4.1.1 系统目标定位

在知识经济时代，知识作为一个组织的核心竞争力，将知识有效地管理起来成为维持组织竞争力的关键。“知识”和“知识管理”这两个词语已经成为学者和管理者们使用频率最高的词语之一。知识型档案管理系统目标是实现档案资源的知识化管理，是建立在“知识管理”基础上的对信息型档案管理系统的发展和提升。什么是“知识管理”呢？



知识管理 (Knowledge Management, KM) 是指在组织中构建知识系统, 使信息和知识通过获取、整理、审核、发布、利用、更新等过程, 达到知识发现和更新的目的, 并回馈给知识系统成为组织的智慧资本, 从而为个人与组织的决策提供经验支持。知识管理为企业实现显性知识和隐性知识的转换与共享提供了途径, 是信息管理的进一步发展, 第一代信息化管理的是数据; 第二代信息化管理的是信息; 而知识管理将信息化推进到第三阶段, 第三代信息化管理的对象是知识。由此可见, 档案管理系统的开发与信息化管理的发展是息息相关的, 而知识型档案管理系统将成为档案管理发展的趋势。

结合“知识管理”以及信息型档案管理系统理论, 可以定位知识型档案管理系统的建设目标主要包括:

1. 实现档案的自动化采集、归档、存储、检索等, 用计算机辅助传统档案管理工作使其自动化。
2. 实现档案的长期保存, 由于档案存放方式由档案柜架变成了光盘、磁盘等电子设备, 存储设备和信息安全两个方面要满足档案长期保存的要求。
3. 实现档案的常规利用, 能够根据访问者的权限对其提供不受时间和空间的限制的网络化档案利用, 如查询、检索、借阅、编研等。
4. 实现档案的知识化服务, 主要包括提供知识共享服务平台、知识地图导航和对档案知识的推理学习等。

#### 4.1.2 参考模型(KAMS)

##### (一) KAMS 环境模型

KAMS 将环境抽象为四个角色: 档案生产者、专家知识提供者、管理者和用户。

- (1) 档案生产者: 是指提供档案的人员或客户端系统。
- (2) 专家知识提供者: 是指将个人工作经验、工作总结等隐性知识提供给档案系统的人员。
- (3) 管理者: 制定档案管理政策的角色, 它的职责是通过制定政策控制档案的有效管理, 明确档案收集的范围, 确认档案的保存期限, 对 KAMS 的运行情况进行分析和调整。

(4) 用户：是指通过与档案管理系统服务交互，获取并利用档案知识的人员或客户端系统。

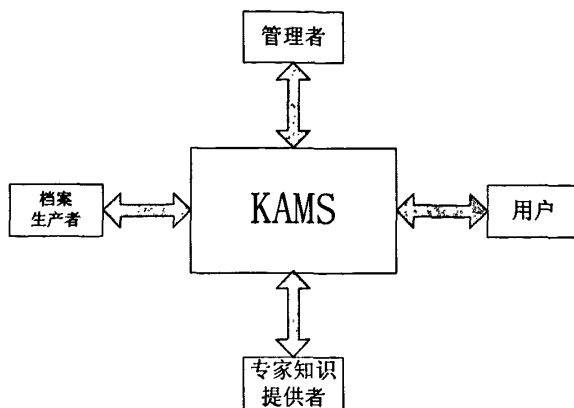


图 4-1 KAMS 环境模型

## (二) KAMS 功能模型

本文通过对 OAIS 参考模型以及知识管理模型的研究，综合考虑知识型档案管理系统的目标定位，对 OAIS 进行知识化扩充，提出了知识型档案管理系统的参考模型(KAMS)，如图 4-2 所示。KAMS 功能模型中，定义了 7 项功能，分别为：档案接收功能、知识档案存储功能、数据管理功能、行政管理功能、保存规划功能、知识协作功能和知识服务功能。

- (1) 档案接收功能，接收信息生产者提交的提交信息包(SIP)，并将 SIP 转换为档案馆中需要管理和存储的内容。
- (2) 知识档案存储功能，主要将接收到的档案以知识的形式进行组织存储，并提供知识检索功能。
- (3) 数据管理功能，用于识别档案描述信息，并为档案元数据的组配、管理与访问提供相关服务与功能。
- (4) 行政管理功能，为档案管理系统的整个运行提供操纵管理等相关的服务。
- (5) 保存规划功能，用于监测 KAMS 的运行环境，规划档案长期保存方案，即便是在原有计算机环境退化之后，仍然能够被目标用户所访问。
- (6) 知识协作功能，提供知识交流协作平台，接收知识型档案(工作经验、专业知识等)并组织存储，为知识及其需求者提供关联平台。

(7) 知识服务功能，用于支持用户鉴定、定位需求档案信息，并以知识的形式推送给用户，包括常规档案利用如查询、借阅、编研等和知识服务如知识检索、知识推理和知识地图导航等。

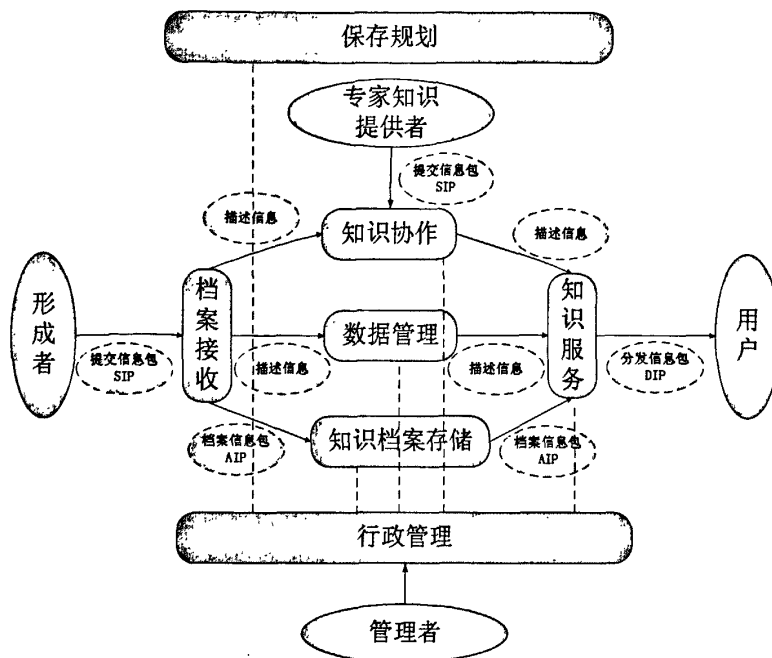


图 4-2 KAMS 参考模型

基于 KAMS 参考模型，设计知识型档案管理系统如图 4-3。它主要包括六个主要子系统组成，包括档案知识获取系统、档案业务工作系统、档案知识组织系统、档案存储系统、档案知识服务系统和工作流系统。

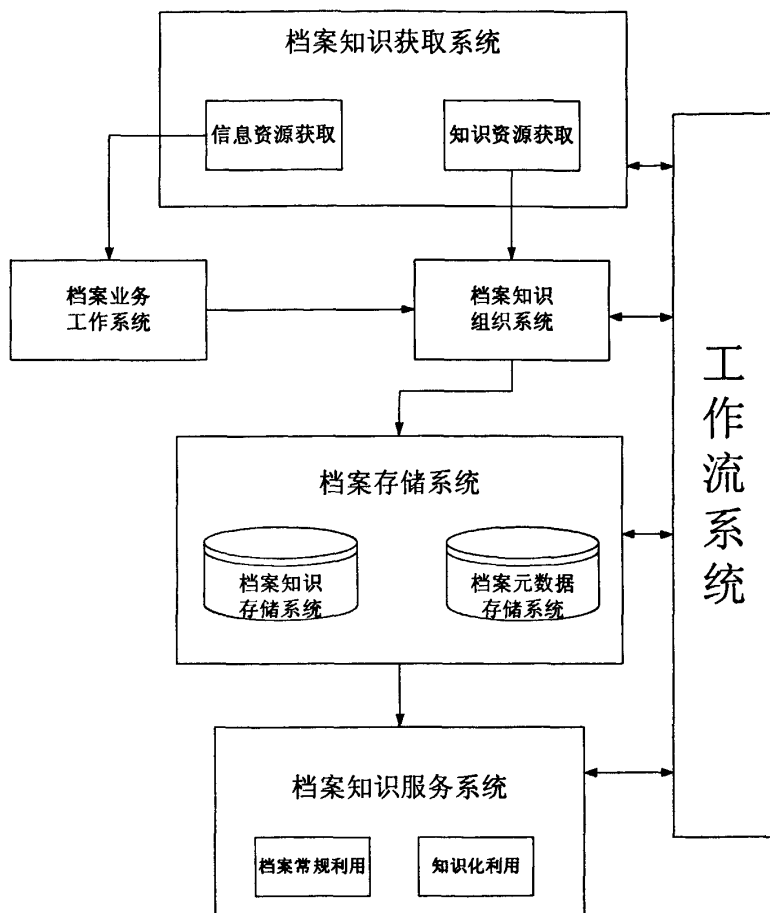


图 4-3 基于 KAMS 的档案管理系统

### (一)档案知识获取系统

档案知识获取系统的主要任务包括：获取各种形式的档案资源，包括从业事务系统中自动或半自动抽取、档案工作人员手动录入、业务人员通过系统提交等信息档案资源；以及从领域专家、管理者、普通员工处获取专家知识、员工经验、工作总结等以知识档案资源。

### (二)档案业务工作系统

档案业务工作系统的主要任务包括：将以数据、信息形式存在的档案资源经过鉴定、审核、著录、编目、分类等工作流程后，形成可归档的档案，并提交给知识组织系统。

### (三)档案知识组织系统

档案知识组织系统的主要任务包括：将待归档的档案资源和知识资源添加到档案知识库中，同时将与之关联的元数据存入到元知识库中，通常需要经过“概念确定→关系确定→知识匹配”流程将档案知识组织到知识库中成为新知识。

#### (四)档案存储系统

档案存储系统包括档案知识存储系统和档案元知识存储系统，档案知识存储系统中主要存储档案知识源，如会议记录、生产运营报表、财务审计报告以及总结经验等。元知识是帮助有效获取知识或建立知识之间联系的知识，可以看成知识的索引。档案元知识存储系统中包括管理档案知识的元数据、描述知识库组织的数据等。

#### (五) workflow 系统

workflow 系统是确保档案信息从捕获、储存到发布，可以有效、正确地运行的整个流程。就数字档案馆而言，workflow 系统为档案数字信息的捕获、组织、维护、鉴定、销毁、发布等提供了全面的流程管理，并使得档案管理工作流程与传统档案管理工作相比变得更为简洁、责任更为明确、内容更为准确。

#### (六) 档案知识服务系统

档案知识服务系统提供常规档案服务和知识化服务两种类型的服务。常规档案服务利用主要包括档案查询检索、借阅、编研等，而知识化服务包括个性化检索、知识地图导航、知识交流协作等。该系统负责接收服务请求，并根据需求将档案知识从存储系统中快速、自动的组织并推送给档案用户。

本文通过对 OAIS 的知识化扩充，得出了知识型档案管理系统的参考模型 KAMS。在以下章节中，本文以此为理论依据，设计了某大型国有电力企业知识型档案管理系统的建设方案。

## 4.2 系统核心业务管理模型

核心业务管理模型定义了档案管理业务的范围与边界，规范了档案处理流程，是系统功能模型、功能实施模型和系统服务架构的物理基础。它以该企业现有的档案工作流程为基础，以建设知识型档案管理系统为目标，在对档案管理工作的需求进行分析后对档案管理核心业务流程进行了前瞻性描述，如图 4-4 所示。

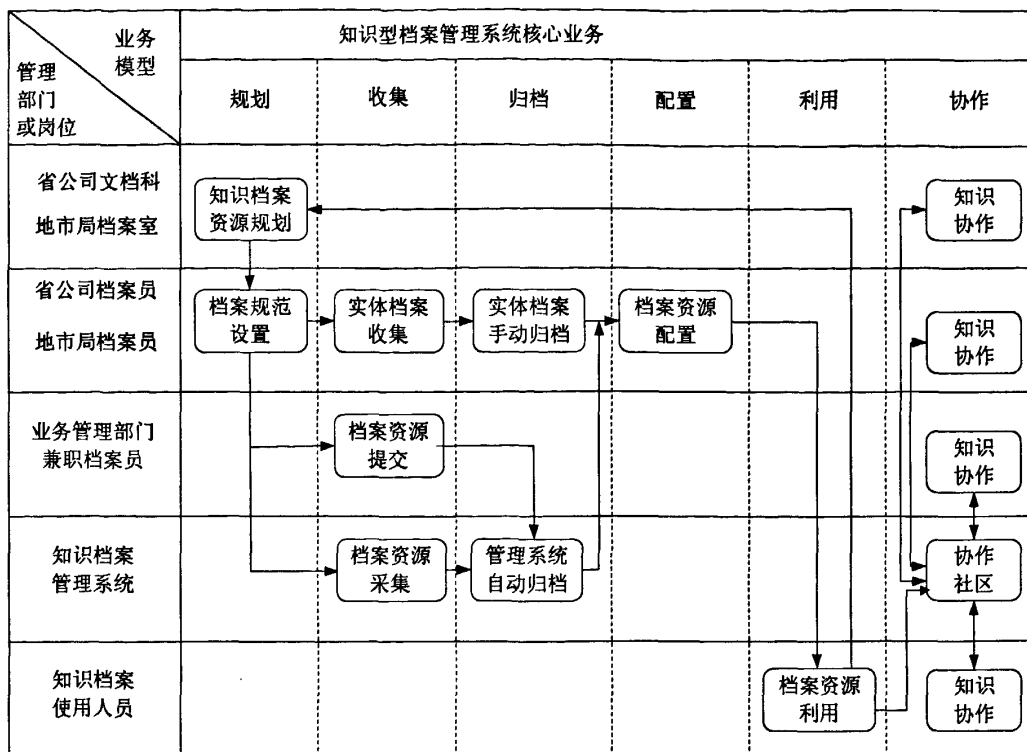


图 4-4 知识型档案管理系统业务模型总图

对核心业务的需求分析是系统设计的前提和出发点，经过对电力企业档案管理工作进行分析调研并结合考虑档案的生命周期，可以将主要的核心业务分为六个部分：档案规划、档案收集、档案归档、资源配置、档案利用、知识协作。

#### 4.2.1 档案规划业务

档案规划业务包括档案资源规划和档案规范设置：

- (1) 资源规划业务：对档案资源进行资源规划的目的是界定档案的管理范围与管理方式，包括需要归档的资源来源，档案名称、内容和格式，归档时间或周期，档案资源的分类编码，负责单位等，从而形成资源规划方案。
- (2) 规范设置业务：执行资源规划方案并设置元数据以远东档案资料的收集与整理工作。档案元数据主要用来描述档案的背景、内容、结构及管理过程信息，用于档案全生命周期的管理。元数据包括档案来源、时间、分类编码、结构、内容特征(关键字、作者、摘要)等。

## 4.2.2 档案收集业务

档案收集包括实体档案收集、数字档案提交、数字档案采集。

- (1) 实体档案收集：实体文档主要是以纸张、图片、磁盘、证章等实物载体记录的档案，对于实体文档的收集需要向系统提交档案的查询记录，包括档案名称、存放地点、内容、特征属性等。对于重要数字档案的实体化也是实体档案收集的内容。
- (2) 数字档案提交：主要由档案工作者对日常工作中归档内容、形式和时间不确定的部份档案进行整理归档，一般采用提交、审核、批准、电子签名的流程。
- (3) 数字档案采集：通过档案采集接口，根据采集调度表，将企业其它业务应用系统中的数字档案自动、半自动采集到档案管理系统中实现企业档案的一体化管理。

## 4.2.3 档案归档业务

归档业务包括实体档案手动归档和管理系统自动归档。

- (1) 实体档案手动归档：主要由各级档案工作者根据相关规定完成对档案的鉴别、著录登记、整理编目、装桢上架等工作，填写并提交档案的相关描述信息到档案管理系统进行统一管理。
- (2) 管理系统自动归档：自动归档要求计算机辅助完成档案鉴定(将待归档文件与资源规划方案中的指标参数进行匹配形成鉴定意见)、著录登记(根据编码规则对档案进行唯一性编码)、整理编目(对档案进行分类管理)、归档处置(自动存储、备份，并生成相关文件)、移交接收、信息发布(对已经形成档案的知识资源进行发布以提供给用户服务功能)等功能。

## 4.2.4 资源配置业务

资源配置业务的目的是：确定各级管理人员中哪些业务需要档案资源提供支持，需要怎样的档案支持，并以自动或半自动的方式根据用户权限将所需档案资源推送给相关联的业务，实现档案管理对业务管理工作的自动化决策支持。

### 4.2.5 档案资源利用业务

档案资源的利用包括常规利用和知识服务：

- (1) 档案资源的常规利用：是指利用计算机辅助实现传统档案馆的功能，如检索、查阅、借用、编研、管理应用等。
- (2) 档案资源的知识服务：是指将档案知识化后利用计算机网络技术提供给用户知识服务，如档案订阅、知识地图导航、知识发现等。

### 4.2.6 档案知识协作业务

档案管理系统的一个重要目标是实现隐性知识的显性化，将团队人员大脑中的经验知识发掘出来进行共享，协作业务包括协作团队建设和知识社区建设。

- (1) 协作团队建设：建立电力企业知识协作团队，从管理制度上鼓励愿意为组织贡献个人工作经验或知识的人员。
- (2) 知识社区建设：建立电力企业知识协作社区，在社区中开展知识共享和知识交流活动等。
- (3) 知识地图建设：需要根据档案知识的组织建立起知识地图，以便于知识的关联与查找。

## 4.3 系统功能模型

系统功能模型由档案数据的来源与宿、数据流向、数据处理以及数据存贮等基本元素构成，是设计知识型档案管理系统的重要依据。根据电力企业经营模式的特点，设计知识型档案管理系统功能模型如图 4-5：



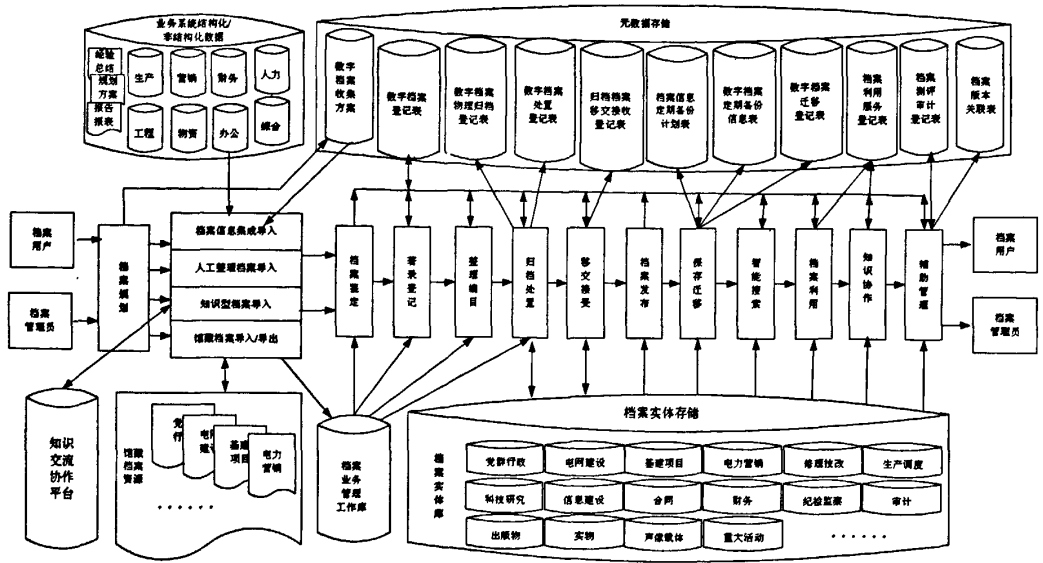


图 4-5 知识型档案管理系统功能模型

### 4.3.1 档案数据来源与宿

档案数据的来源主要包括：馆藏的档案数据、核心业务应用系统（生产、营销、财务、人力资源、工程、物资、办公以及综合）数据、档案管理人员整理的的数据等。

档案数据的宿指的是档案数据的目的地，主要包括：档案知识库、业务接口、档案管理者以及各级档案利用者（档案用户）。

### 4.3.2 档案数据处理

从档案管理业务模型中可以分解出档案数据处理过程，主要包括：档案规划、档案征集、档案鉴定、著录登记、整理编目、归档处置、移交接收、档案发布、保存迁移、智能搜索、知识地图、协作服务以及辅助管理（测评审计、知识统计与版本管理）等。

### 4.3.3 档案数据存储

档案数据存储包括：档案实体知识资源存储和档案元知识存储两类。档案元知识定义了档案的属性信息、管理规则、生命流程、安全权限等内容以及档案知识与知识库其它知识间的关联关系。档案实体知识资源存储的内容是组成档案的

实际内容，按档案的管理分类方法进行组织存储和管理。

#### 4.3.4 档案数据流

档案数据流是用于关联数据来源与宿、数据处理和数据存贮的数据实体，主要包括：流入元数据库的数据实体、流入档案实体库和元知识库的数据实体、流入馆藏档案馆的数据实体以及流向系统外部(档案管理人员以及档案利用人员)的数据实体(如统计分析报表以及档案检索输出的数据)等。

### 4.4 功能实施模型

根据电力企业知识型档案管理系统的系统功能模型，可以将系统功能分为三个平台进行实施建设，分别是：档案本体库维护平台、档案管理业务服务平台和档案知识协作服务平台，如图 4-6 所示：

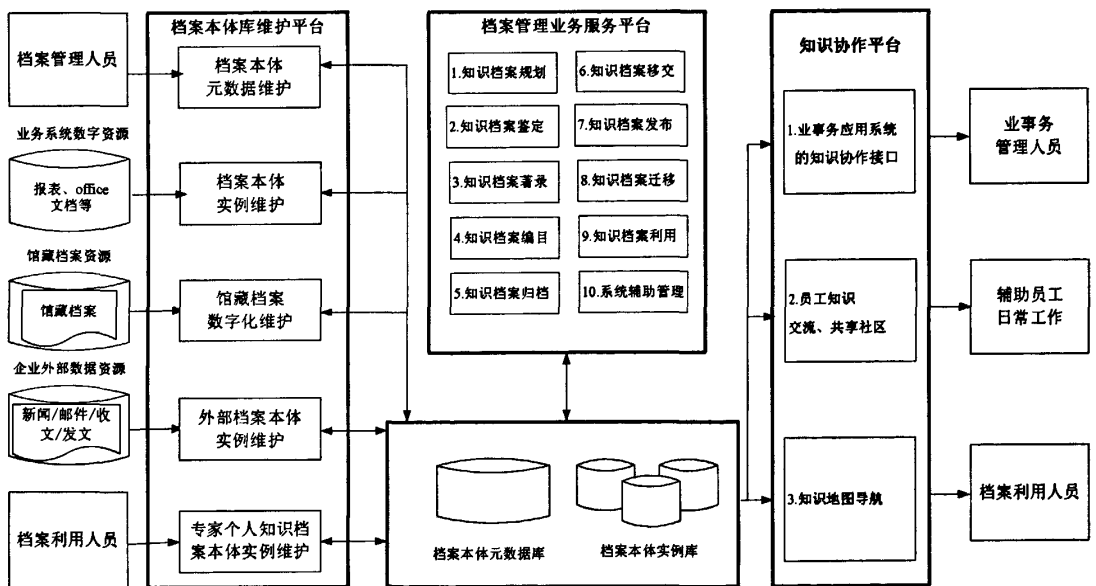


图 4-6 知识型档案管理系统功能实施模型

#### 4.4.1 档案本体库维护平台

本体库的维护包括档案本体元数据维护、档案本体实例维护、馆藏档案的数字化、外部档案本体实例维护和专家个人知识档案本体实例的维护。

(一) 档案本体元数据库包括共享元数据库和个性化元数据库, 这些档案本体元数据主要包括: 服务定义、过程映射、业务规则、实体和关系、编排、转换、参考数据模型、业务活动和事件、审计要求、角色和授权映射以及治理规则和策略等等。系统提供本体元数据的维护功能, 包括录入、浏览、增加、修改、删除等, 为档案资源规划的提供方案。

(二) 档案本体实例主要指来源于业事务应用系统中的档案实体, 系统在本体元数据的指导下完成档案实体采集工作, 包括档案的自动或半自动导入/导出、人工整理档案的导入以及业事务系统中档案的集成/导入等, 并提供浏览、增加、修改、删除等维护功能。

(三) 馆藏档案的数字化维护是将馆藏档案数字化、格式标准化, 导入到档案管理系统中进行统一管理。

(四) 外部档案本体实例维护是指系统将与企业外部进行交流的新闻、邮件、文件等外部档案本体实例数字化后导入到档案管理系统中, 并提供浏览、增加、修改、删除等维护功能。

(五) 专家个人知识档案本体实例维护是将存储在专家个人大脑或电脑上的可用于交流共享的知识, 包括专业知识、个人工作经验等数字化后导入到档案管理系统中, 进行统一管理并用于知识协作。

#### 4.4.2 管理业务服务平台

根据核心业务管理模型和系统功能模型可以得出, 在档案管理业务服务平台中的功能包括: 资源规划、收集、鉴定、著录、编目、归档、移交、发布、迁移、利用以及辅助管理等功能。

#### 4.4.3 知识协作服务平台

知识协作服务平台实现的功能包括: 与业事务应用系统的知识协作接口功能、员工共享知识社区功能、知识地图导航功能。

##### (一) 与业事务应用系统的知识协作接口功能

根据业务模型中的资源配置业务, 对于在业事务应用系统中需要利用档案知识资源的业事务提供知识协作接口, 实现档案管理系统与业事务应用系统关联对接的辅助决策支持。

(二) 员工共享知识社区功能

在企业范围内构建共享知识社区，将愿意共享个人专业知识和工作经验的员工组织和管理起来，促进企业的共享文化建设。

(三) 知识地图导航功能

建立能够反映企业档案架构的全局知识地图以及企业的岗位知识地图，为企业员工利用档案知识提供导航功能。

### 4.5 系统服务架构

对于大型国有电力企业省公司来说，知识型档案管理系统不仅面向于省一级的档案管理业务，还要面向下属地市级的档案管理业务，在部署过程中还要实现某些特有的面向各级管理业务的服务功能。因此使用面向服务(SOA)的技术架构，将知识型档案管理系统的功能封装成各类服务，并在服务库中进行分层管理，是作为大型企业档案管理系统的良好选择。本文中设计的知识型档案管理系统正是采用面向服务(SOA)的技术架构，如图 4-7 所示：

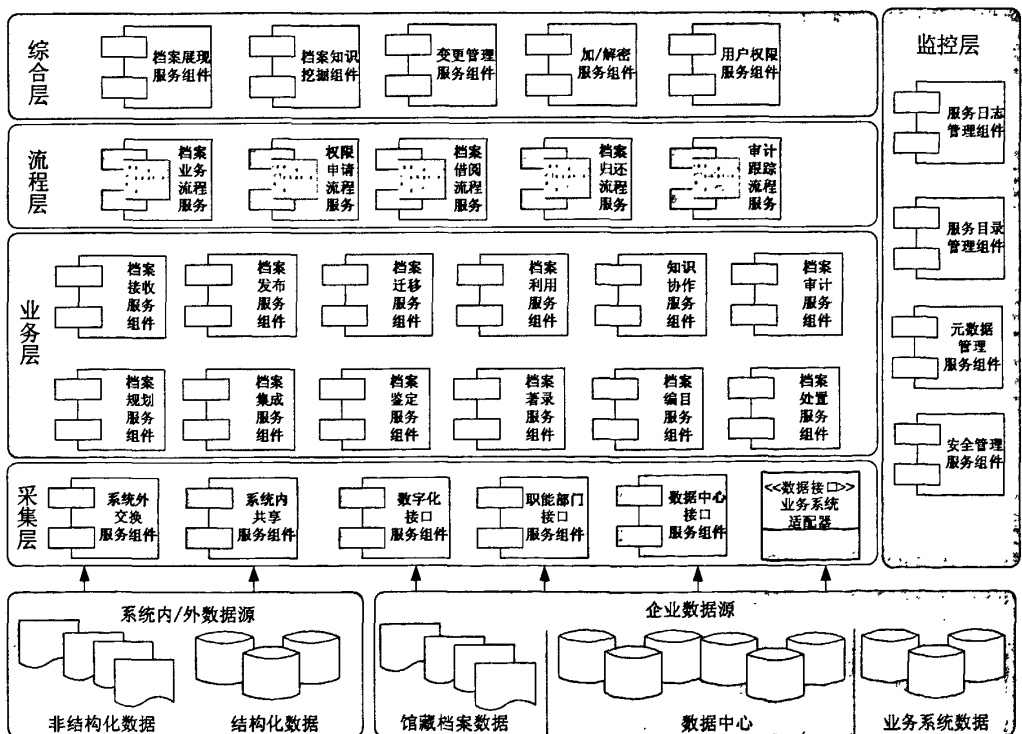


图 4-7 知识型档案管理系统服务体系架构

电力企业知识型档案管理系统的服务体系架构将服务分为 5 个层次,它们分别是:档案数据采集服务层、档案管理业务服务层、档案管理流程服务层、档案综合服务层以及管理监控服务层。其中管理监控服务层是贯穿于前 4 层的服务层。

#### (一) 档案数据采集服务层

档案数据采集服务层封装从不同渠道采集档案数据的功能,包括:与系统外交换数据的服务、与系统内共享数据的服务、与数字化接口的服务、与职能部门接口的服务、与数据中心接口的服务以及与业务系统接口的服务等。

#### (二) 档案管理流程服务层

档案管理流程服务层封装档案管理与流程相关的业务处理功能,包括:档案业务流程服务、权限申请流程服务、档案借阅流程服务、档案归还流程服务以及审计跟踪流程服务等。

#### (三) 档案管理业务服务层

档案管理业务服务层封装档案管理的业务处理功能,包括:档案规划业务服务、档案集成业务服务、档案鉴定业务服务、档案著录业务服务、档案编目业务服务、档案处置业务服务、档案接收业务服务、档案发布业务服务、档案迁移业务服务、档案利用业务服务、知识协作业务服务以及档案审计业务服务等。

#### (四) 档案管理综合服务层

档案管理综合服务层封装档案管理的高级业务处理功能,包括:档案展现服务、档案知识挖掘服务、档案变更管理服务、档案加/解密服务以及用户限管理服务等。

#### (五) 管理监控服务层

档案管理监控服务层封装系统管理功能,包括:系统服务日志管理服务、系统服务目录管理服务、元数据管理服务以及系统安全管理服务等。

## 第五章 总结与展望

### 5.1 论文总结

本文在研究了不同阶段的档案管理系统后，结合知识管理相关理论，以实现档案“知识化管理”为目标，研究了知识型档案管理系统的建设及其档案本体表示技术。

文对档案管理系统的发展历程及研究现状进行了详细的总结与分析，目前对于档案管理系统的研究基本上处于信息型阶段，其模型设计和关键技术发展已较为成熟，而对知识型阶段的研究处于萌芽状态。

本文在对知识管理相关理论作了研究后，发现档案作为知识的重要载体，以知识的方式进行管理具有可行性和必要性。基于将档案管理“知识化”的思想，本文研究了知识型档案管理系统建设的关键技术——档案知识表示技术，提出了档案本体构建的流程，以此为依据，实际构建了某大型国有电力企业档案本体。

本文在对档案管理业务及知识管理理论进行详细研究后，提出了知识型档案管理系统的参考模型 KAMS，并以此为理论依据，在某大型国有电力企业背景，从业务管理模型、系统功能模型、功能实施模型和系统服务架构多个角度，对其知识型档案管理系统的建设作了实际研究。

### 5.2 展望

知识型档案管理系统的建设目前发展还不是很成熟，从它的概念定义、管理范围、功能界定、关键技术等各个方面都需要深入研究。本文对知识型档案管理系统的参考模型和知识表示技术作了详细研究，得出一些研究理论成果，以此为依据，以某大型电力企业为背景，设计了该企业知识型档案管理系统的建设方案，并构建了企业档案本体。

本文在以下几个方面还需要继续研究：

1. 本文主要讨论了档案本体的建设，基于本体的推理未作深入探讨，该问题有待进一步研究。

2. 基于本体的知识型档案管理系统,其知识化利用和服务有别于传统档案管理系统,涉及到的技术如基于本体的检索查询技术、知识地图导航技术、知识协作交流技术等,在理论和实施上都有待深入研究。
3. 在对知识型档案管理系统建设的研究中,本文提出了 KAMS 参考模型,并以此为基础展开了结合实际的研究,但对于整个系统建设的全面性、完善性、可维护性等方面,在实际应用过程中,需要根据实际需求再做深入探讨。

## 参考文献

- (1) 2006-2020 年国家信息化发展战略.  
<http://www.cast.org.cn/n435777/n435794/n457817/37079.html> .2006.
- (2) 中华人民共和国档案法.2005 年颁布.
- (3) 徐义全.电子文件系列讲座之十二电子档案的鉴定销毁.北京档案,2001,  
2001(12):16-17
- (4) Waters D& Garrett J. Preserving Digital Information. Washington DC  
Commission on Preservation and Access .
- (5) 朱小怡.数字档案馆建设理论与实践.华东师范大学出版社.2007 年
- (6) 李音.美国将创建电子文件档案馆.中国档案,2000(7):50
- (7) 李国庆.深圳数字档案馆建设的理论架构及阶段性成果.中国档案.2003(3):  
11-14
- (8) 法奕.青岛市城建档案数字化建设.城建档案.2008(9)
- (9) 张怀廉.季雪岗.李慎民.企业数字档案馆的建立——江苏省电力公司数字档  
案馆概况.中国档案.2007(10)
- (10) 吴大勇.胜利油田企业数字档案馆项目.紫光档案报.2009(1)
- (11) 赵屹.陈晓晖.对“数字档案馆”的简单认识.档案学通讯.2003(5):38-41.
- (12) 冯万恒.电子档案与传统档案管理上的差异.兰台世界.2005(4)
- (13) 柳琰.馆藏的控制与优化.中国档案.2001(3)
- (14) Reference Model for an Open Archives Information System(OAIS).CCSDS  
650.O-B-1.MBLUE BOOK. January 2002
- (15) 屠跃明.黄永文.对 OAIS 参考模型的研究.档案学研究.2007(2):46
- (16) 章燕华.刘霞.OAIS 参考模型.数字资源长期保存的概念框架.浙江档案.  
2007(3)
- (17) 徐周亚.镇锡惠.许绥文.OAIS 参考模型与中文元数据方案.现代图书情报技  
术.2003(4):9
- (18) 朱德红.基于 OAIS 的数字档案馆元数据体系构建.硕士论文.黑龙江大学.17.  
2009



- (19) 王琳.OAIS——海洋数字档案馆的神经中枢.海洋信息.2009(1):10
- (20) 王平.皮介郑.郝春云. 数字敦煌方案设计.现代图书情报技术. 2006 (11):24.
- (21) Reynolds. H & Koulopoulos. T. Enterprise knowledge has a face. Intelligent Enterprise. 2(5):29-34.
- (22) 金更达.基于 OAIS 的数字档案馆系统框架研究.浙江档案.2007(4)
- (23) 何嘉荪.金更达.电子文件管理元数据规范.浙江档案.2005(4)
- (24) 薛四新.档案馆信息化与档案管理变革.北京:机械工业出版社.2008.136
- (25) 陈兵.电子政务技术与安全.北京:北京大学出版社.2003.100
- (26) 樊如霞.郑志荣.影响数字档案信息安全的因素与对策.档案学通讯.2007(6): 76
- (27) 张勇.数字档案信息安全保障体系研究.硕士论文.苏州大学. 26.2007
- (28) 庾宁.杨春秀.数字档案长期安全保存问题分析.华南金融电脑.2009(9):81
- (29) Guarino N.Masolo C.Vetere G.OntoSeek: Content-Based Access to the Web. IEEE Intelligent Systems.1999.14(3):70-80.
- (30) 苏新宁.任皓等.组织的知识管理.北京:国防工业出版社.2004. 58
- (31) 刘峡壁.人工智能导论——方法与系统.北京:国防工业出版社.2008. 60
- (32) 张仰森. 人工智能原理与应用. 北京:高等教育出版社. 2005.50
- (33) Guus Schreiber. 知识工程和知识管理. 北京:机械工业出版社. 2002
- (34) 张桂平.尹宝生.蔡东风.知识管理综述.沈阳航空工业学院学报. 25(5)
- (35) 张玉峰.李敏.晏创业. 论知识检索和信息检索.中国图书馆学报.2003(5):23
- (36) Jiawei Han& Micheline Kamber. 数据挖掘概念与技术. 北京: 机械工业出版社. 2006. 3
- (37) Borst W N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. PhD thesis. University of Twente. Enschede.1997
- (38) 缪涵琴.基于本体的档案信息检索系统的设计与实现.硕士论文.苏州大学. 2007
- (39) Perez A G. Benjamins V R. Overview of Knowledge Sharing and Reuse Components. Ontologies and Problem Solving Methods. Proceedings of the IJCAI99 workshop on Ontologies and Problem-Solving Methods.1999:1-15.

- (40) 张艳.具有本体特征的档案管理系统设计与实现.硕士论文.苏州大学.10.2008.
- (41) Davies J & Weeks R. QuizRDF. Search Technology for the Semantic Web. Proceedings of the 37th Hawaii International Conference on System Sciences. 2004
- (42) 杜剑峰.网络信息集成系统的研究.广州.中山大学. 2002:6
- (43) Miller E. Weaving Meaning. Semantic Web Applications. Presented at INTAP Interoperability Technology Association for Information Processing. Tokyo. Japan. 2003.11
- (44) Gruber T R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies. 1995(43):907-928
- (45) 陈禹. IDEF 建模分析和设计方法.北京:清华大学出版社.2000.
- (46) Gruninger M & Fox M S. Methodology for the Design and Evaluation of Ontologies. Workshop on Basic Ontological Issues in Knowledge Sharing. Montreal. Canada. 1995 . IJCAI-95
- (47) Uschold M & Gruninger M. Ontologies: Principles, Methods and Applications. Knowledge Engineering Review. 1996.11(2):93-155
- (48) Natalya F. Noy & Deborah L. McGuinness. A Guide to Creating Your First Ontology .Ontology Development 2001(8):101
- (49) 程显毅.刘一松.晏立.面向智能体的知识工程. 北京:科学出版社.2008.80
- (50) 董慧.陈亮. 数字图书馆历史领域资源本体构建. 信息化与信息资源管理学术研讨会文集. 武汉: 湖北人民出版社. 2005.213-226

## 附录一 档案实体分类体系

序号	一级类目	二级类目	三级类目
1	文书档案 rapporteur_archives	党群工作	党群党务
			组织工作
			工会工作
			共青团青年
			协会学会工作
		行政管理 administration	行政事务
			医疗卫生
			后勤福利
			外事工作
		经营管理 operation_ management	计划统计
			物资管理
			用电供热营业
			农网农电管理
			产品销售
		生产技术 Production_ technologies	应急管理
			规划基建管理
			生产管理
			施工管理
			安全环保
			科技管理
质量管理			
标准计量			
财务审计 Financial_ardit	财务管理		
	会计帐务		
	审计工作		

		人事劳资 Personnel_ Department	机构、编制 人事管理 劳动管理 工资保险 教育培训
2	生产运行档案 production_run_achives	电力系统运行 Power_system_run 电力系统图表 Power_system_chart 调度通信 Attemper_communication 调度自动化 Attemper_automation 事故管理 Accident_manage	
3	会计档案 accounting_achives	机关帐 Department_account 工程帐 Engineering_account 科研帐 Scientific_account 生产帐 Produce_account 基建帐 Capital_account	
4	建设项目档案 construction_project	电网项目 Power_grid_project	架空线路（交、 直流工程） 电力电缆（包括 海底电缆） 换流站工程

			交流变电工程
			申补站
			光缆通讯工程
		电源项目 Power_supply_project	抽水蓄能电站项目 水电站建设项目
		小型基建项目 Small Projects of Capital construction	
5	审计档案 auditing_archives	专项审计	工程审计 资产审计 资金审计 大修技改审计 营销审计 电费审计 经济效益审计
		经济责任审计	离任审计 届中审计
		年度报表审计	
		工程竣工决算审计	
6	照片档案 photo_archives	新闻照片	
		工程照片	
		专题照片	
7	合同档案 contract_archives	资产管理合同	
		技术及劳务合同	
		工程合同	
		营销合同	
		其它合同	
8	科研档案 scientific_archives	重大技术攻关	
		新技术、新工艺开发	
		新产品、新材料试制	

		软课题、应用理论研究	
		信息系统研究	
		引进技术消化、吸收与创新	
		科技成果转化与推广应用	
		技术标准编制与修订	
		其它	
9	荣誉实物档案 honour_entity	荣誉类	
		纪念品类	
		印信类	
		证照	
10	声像电子档案 vedio_achives	声音文件	
		图像文件	
		影视文件	
		网页文件	
		电子邮件	
		多媒体文件	
		程序文件	
11	营销客户档案 customer_marking	工业客户 Industrial_consumer	
		商业客户 Business_consumer	
		农电客户 Country_consumer	
		居民客户 Residents_consumer	
12	信息化建设项目档案 informatization_ construction_project	项目分类	

13	招标投标档案 bidding_archives	建设项目	
		变电设备	
		输电线路设备材料	
		低压、配电网设备	
		电网检测、试验设备	
		管理信息系统建设	
		综合设备、材料、物资	
14	修理、技改项目档案 reforming_and_repairing	电网项目	架空线路（交、直流工程）
			电力电缆（包括海底电缆）
			换流站工程
			交流变电工程
			串补站
			光缆通讯工程
		电源项目	抽水蓄能电站项目
	水电站建设项目		
	小型基建项目		
15	纪检监察档案 supervision_and_investigation	纪检工作	
		监察工作	
16	出版物档案 publication_archives	公开出版刊物	
		内部刊物	
17	重大活动档案 important_activity	领导视察	
		企业庆典	

## 附录二 档案本体部分 OWL 代码

```

<Ontology
  ontologyIRI="http://www.owl-ontologies.com/unnamed.owl">
  <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#" />
  /**定义各种类**/
  <Declaration>
    <Class IRI="#Accident_manage"/>
  </Declaration>
  ... ..
  <Declaration>
    <Class IRI="#achives_format"/>
  </Declaration>
  /**定义类与类之间的关系**/
  <Declaration>
    <ObjectProperty IRI="#attribute_of"/>
  </Declaration>
  <Declaration>
    <ObjectProperty IRI="#instance_of"/>
  </Declaration>
  <Declaration>
    <ObjectProperty IRI="#kind_of"/>
  </Declaration>
  <Declaration>
    <ObjectProperty IRI="#part_of"/>
  </Declaration>
  /**定义类与类之间的上下级关系**/
  <SubClassOf>
    <Class IRI="#Accident_manage"/>
    <Class IRI="#production_run_achives"/>
  </SubClassOf>
  ... ..
  <SubClassOf>
    <Class IRI="#Power_system_run"/>
    <Class IRI="#production_run_achives"/>
  </SubClassOf>
  /**定义类与类之间 kind_of 的关系**/
  <SubClassOf>
    <Class IRI="#Small_Projects_of_Capital_construction"/>
    <ObjectAllValuesFrom>
      <ObjectProperty IRI="#kind_of"/>
      <Class IRI="#construction_project"/>
    </ObjectAllValuesFrom>
  </SubClassOf>

```



```
... ..
/**定义类与类之间 part_of 的关系*/
  <SubClassOf>
    <Class IRI="#achives_size"/>
    <ObjectAllValuesFrom>
      <ObjectProperty IRI="#part_of"/>
      <Class IRI="#structure_character"/>
    </ObjectAllValuesFrom>
  </SubClassOf>

... ..
/**定义类与类之间 attribute_of 的关系*/
  <SubClassOf>
    <Class IRI="#content_character"/>
    <ObjectAllValuesFrom>
      <ObjectProperty IRI="#attribute_of"/>
      <Class IRI="#achives_class"/>
    </ObjectAllValuesFrom>
  </SubClassOf>

... ..
/**kind_of 属性具有传递性*/
  <TransitiveObjectProperty>
    <ObjectProperty IRI="#kind_of"/>
  </TransitiveObjectProperty>

... ..
</Ontology>
```

## 致谢

在本文完成之际，谨在此向论文编撰以及研究生学习期间一直给予我支持、关心和鼓励的人表示衷心的感谢。

本文的工作是在导师李小福副教授的严格要求和悉心指导下完成的，从选题、构思到具体的细节方面，都凝聚着李老师的关心与指导。在研究生学习阶段，李老师宽厚待人的品德、一丝不苟的工作态度、深厚的学术修养和创新精神，给了我深刻的启发，并为我的学习和工作树立了榜样。在此，衷心地感谢导师一直以来对我的支持和鼓励。

同时，感谢研究所武珍珍、秦荣芝、周植诚等同学，以及郑会师姐，是他们在学习和生活上对我的关心和帮助陪我顺利走过这两年，我会珍惜这份宝贵的情谊并珍藏这份美好的回忆，更感谢你们在学术上给我的启示和意见。

再次，我要感谢我的父母，是他们给了我克服困难的勇气和前进的动力。在此，我特致以你们最衷心的感谢和最真挚的祝福。

最后，感谢在校期间曾经关心和帮助过我的研究生院和数学与计算科学学院领导、老师和同学。也感谢为评阅本文而付出辛勤劳动的专家和学者，同时非常感谢在学位论文答辩中提出宝贵意见的各位专家。