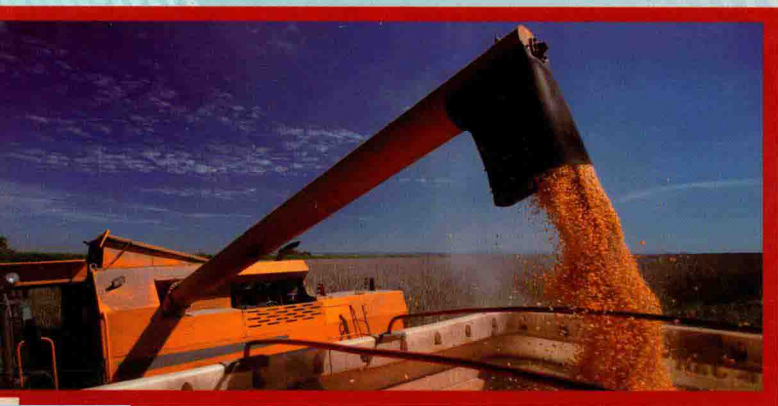




新型职业农民培训通用教材

现代农业信息获取辨析与利用



覃红梅 主编

天津出版传媒集团

 天津科学技术出版社



现代农业信息获取辨析与利用

覃红梅 主编

天津出版传媒集团

 天津科学技术出版社

图书在版编目 (CIP) 数据

现代农业信息获取辨析与利用 / 覃红梅主编. 一天
津: 天津科学技术出版社, 2016. 10
ISBN 978-7-5576-1871-1

I. ①现… II. 覃… III. ①农业—信息资源—信息
获取—研究 IV. ①F302.4

中国版本图书馆CIP数据核字(2016)第261462号

责任编辑: 吴 颀

责任印制: 兰 毅

天津出版传媒集团 出版
 天津科学技术出版社

出版人: 蔡 颀

天津市西康路35号 邮编300051

电话 (022) 23332399 (编辑部)

网址: www.tjkjcs.com.cn

新华书店经销

三河市恒彩印务有限公司

开本 710×1 000 1/16 印张 11 字数 163 000

2016年10月第1版第1次印刷

定价: 27.50元

前言

农业是国民经济的基础，在国民经济中占有重要地位。随着科技的进步，农业正在向现代化发展。要想实现现代化，让农业的生产力实现质的飞跃，必须大力发展农业信息技术这一管理和传播手段。

农业信息技术，是指利用信息技术对农业生产、经营管理、战略决策过程中的各种信息进行采集、存储、传递、处理和分析，为农业研究者、生产者、经营者和管理者提供资料查询、技术咨询、辅助决策和自动调控等多项服务的技术的统称。可以说，这是利用现代高新技术改造传统农业的重要途径。

党的十六大明确指出“信息化是我国加快实现工业化和现代化的必然选择”，并将大力推进信息化作为21世纪经济建设和改革的一个主要任务。2006年，中央1号文件强调要“积极推进农业信息化建设”，并将其作为现代农业和社会主义新农村建设的一项重要内容。党的十六届六中全会进一步强调要进一步加强农业信息化建设，通过信息技术改造传统农业、装备现代农业，通过提高信息化水平缩小国内外农业发展的差距。

基于此，我们在借鉴和吸收国内外学者研究成果的基础上，编写了《现代农业信息获取辨析与利用》一书，从各个方面介绍农业信息的相关内容，引导人们认识和了解现代农业的信息技术。

本书共分为七个章节。

第一，介绍了农业信息技术，其中包括农业网址数据库管理系统，从信息服务向知识转变，农业信息技术应用展望等；

第二，介绍了国内外农业信息搜索引擎现状，其中包括国内外农业相关的信息搜索引擎，面对主题的专用搜索引擎系统核心技术研究等；

第三，介绍了垂直搜索引擎的原理与技术，其中包括垂直搜索引擎系统架构特点，垂直搜索引擎开发关键技术等；

第四，介绍了农业信息主题网页采集技术，其中包括农业信息主题网页的特点分析，数据采集与更新模型等；

第五，介绍了农业信息辨析与宏观农业，其中包括我国五种主要农产品消费趋势，中国粮食与食物安全早起于景系统研究等；

第六，介绍了农业信息资源管理，其中包括农业数据库建设的现状及发展趋势，数字图书馆的评价体系初探等；

第七，介绍了信息资源的利用，其中包括信息资源的收集、整理与分析，科技论文写作，科技论文的投稿与评价等。

编者希望广大读者能够通过阅读本书，对现代农业信息有一定的了解，从而更好地发展现代农业，促进现代农业发展。因编者水平有限，加之编写时间仓促，书中难免出现一些错误之处，恳请广大读者批评、指正。

目 录

第一章	现代农业信息技术	001
第一节	农业网址数据库管理系统	002
第二节	基于光盘的农业信息传播	004
第三节	从信息服务向知识服务转变	008
第四节	农业数字化图书馆集成服务系统框架	015
第五节	农业信息技术应用展望	019
第二章	国内外农业信息搜索引擎现状	025
第一节	国内外农业相关的信息搜索引擎	026
第二节	面向主题的专用搜索引擎系统核心技术研究	031
第三节	服务于电子商务的搜索引擎在专业领域的应用	032
第四节	价格搜寻理论	033
第五节	农业信息搜索引擎开发目标与技术路线	035
第三章	垂直搜索引擎的原理与技术	037
第一节	垂直搜索引擎系统架构特点	038
第二节	垂直搜索引擎开发关键技术	041

第三节	检索质量评估标准	051
第四章	农业信息主题网页采集技术	053
第一节	农业信息主题网页的特点分析	054
第二节	数据采集与更新模型	057
第三节	性能测试与评估	065
第五章	农业信息辨析与宏观农业	069
第一节	中国粮食与食物安全早期预警系统研究	070
第二节	我国农业发展为食品行业提供的机遇与需求	077
第三节	我国五种主要农产品消费趋势	082
第四节	宁夏吴忠市利通区奶业发展及对农民增收的研究	086
第五节	中国农民食物消费与营养状况	092
第六节	饲料原料价格波动, 饲料价格风声水起	101
第七节	我国有机农业的发展及对策分析	106
第六章	农业信息资源管理	111
第一节	农业数据库建设的现状及发展趋势	112
第二节	数字图书馆的评价体系初探	119
第三节	人力资源管理的创新发展——知识管理	127
第四节	现代信息环境下国家农业图书馆的知识服务	131
第五节	美国大学图书馆对资源的高效合理利用	137
第七章	息资源的利用	143
第一节	信息资源的收集、整理与分析	144
第二节	科技论文写作	152
第三节	科技论文的投稿与评价	161

第一章

现代农业信息技术



第一节 农业网址数据库管理系统^①

随着我国计算机网络基础设施建设的迅速发展,一大批农业网站先后出现。由1998年的200多个发展到2000年底的2000多个。但目前我国农业网站和农业信息资源的建设仍然存在着总体规模小、地域分布不均衡,设计不够精细和规范,缺乏个性和专业特色,信息资源缺乏多样性,信息开放性和共享程度低、信息的时效性差等问题。

为了了解和跟踪我国农业信息网站的发展,为建立科学性较高的农业搜索引擎和搜索策略打下基础,以便更有效地从互联网浩瀚的无序的信息海洋中收集和整理加工有效和有序的信息,1998年农业科学院内部出版了《农业黄页》一书,并在全中国农业科技工作会议上发放。该书分为国内信息源、搜索引擎、生物技术、生物技术、国外农业信息源、世界各国农业站点等六大部分,共提供国内外农业及与农业有关的网址6000多条。内容已成为中国农业专家热线(<http://www.agric138.com.cn>)的组成部分。随之我们进行了农业信息网站分类体系的研究,初步提出了农业网站信息分类方案。在方案设计中,采用面分类法,以地理区域、主办者、主要内容、涉及行业等为主面,每个主面下设若干个类,每个类下设若干个子类,形成多维空间信息的分布关系。从不同角度揭示网站的主要特点,满足用户对多属性信息网站的检索要求,使之更加科学合理。该分类体系已应用于农业部信息中心网站(<http://www.agri.gov.cn>)。在这些工作的基础上,进一步研制了农业网址数据库及其管理系统。

一、农业网址数据库

农业网址数据库是在《农业黄页》的基础上建立起来的,其中除包含《农业黄页》的全部网址外又新增网址六万余条。农业网址数据库由网址分类表和网址数据两个表组成,由于目前对于网络信息如何分类尚在探讨之中,现在所使用的分类方案也将不断地改进,因此在系统中建立了网址分类表及其维护表单。网址数据表用于存放网址数据的主要信息及其分类。

1. 网址分类表 网址分类表是为管理网站的分类而设计的,由于网站的分类采用树型结构,因此要把这种树型结构表示出来就不能采用一般的主题标引将关键词罗列放在一个字段中的办法建立数据库结构。本网址分类表的结构设计采用

^① 钱平、苏晓鹭,农业科学院农业信息研究所研究员。

关系型数据库的基本结构,将分类的树型结构分为父类与子类两个字段,而树型结构间的关系由字段的值体现,分类表设有键值、上级的键值和分类三个字段。

2. 网址数据表 网址数据表中设立了包括键值、是否连通、中文名称、英文名称、域名、地理区域、省名、国名、国际组织、网站属性、网站内容描述、关键字、科技教育、政策与管理、市场信息、分析预测、气象与环境、涉农行业、其他涉农行业、种植业、其他种植业、特种种植业、其他特种种植业、畜牧业、其他畜牧业、特种养殖业、其他特种养殖业、种业、其他种业、渔业、其他渔业、农垦、其他农垦、花卉园艺、其他花卉园艺、农业生产资料、其他农业生产资料、工程设备、其他工程设备、农副产品加工、其他农副产品加工、植物保护、其他植物保护、主办单位、电子邮箱、联系人、电话等47个字段。其中在带有“其他”两字的字段中包含的内容为本分类中没有列出的类名,而此类名被认为是需要加进分类中以便查询使用的。

二、农业网址数据库管理系统

农业网址数据库管理系统由分类表的维护、网址数据的录入、查重、查询等主要模块以及数据预处理、数据批量导入等程序组成。

1. 分类表的维护模块 分类表的维护模块将Treeview与Listview控件结合使用,在Treeview控件中装入分类表, Listview控件显示所点击的类及其子类。其使用方法类似于Windows操作系统下的资源管理器,在任一节点可进行类的增加、删除、修改和移动。由于网站信息分类中的某个类所在的位置是有意义的,因此在节点移动时必须确定准确的将要移动到达的位置。同时在其以下的所有同级的节点均要被修改顺序下延,被移动节点的原位置也将进行相应的处理。

2. 网址数据的录入模块 网址数据的录入模块将网站的分类全部列在界面上,并采用check控件便于录入者输入,这样既可以提高速度又减少出错率。因受屏幕大小的限制将整个表单分为四页,前三页放置网站的基本信息和分类信息,第四页以表格的形式显示数据库的全部内容,用户可用此页浏览数据库,另外,录入数据表单具有网站名称查重和网站域名查重的功能,输入人员可根据这两项的查重结果确定是否录入此条记录。

3. 网址数据的查重模块 网址数据的查重模块为批量装载数据而设计,表单由三页组成,第一页调入要进行查重的数据库并可选择进行网站名称或网站域名的查重,第二页显示查重结果,此时的显示结果为逻辑删除重复的记录,第三页将重复的记录以分组的形式显示,显示顺序可升可降,使用者可在重复记录组中重新确定选择保留和删除的记录,在确认无误后可进行记

录的物理删除。经查重后的数据库可覆盖原来的数据库,也可另存为新的数据库。数据的查重可在单个库内进行也可进行多个数据库联查。

4. 网址数据查询模块 网址数据查询模块提供选择查询及组合查询两种方式。在选择查询方式下,用户通过在查询选项中选择自己所需要的选项进行查询,选择查询的好处在于无需输入汉字,所选择的选项均为数据库中的规范数据,所查结果准确。组合查询除具有选择查询的功能外,还可通过输入关键词方法对网站名称、网址、网站描述等字段进行逻辑组配检索。这两种查询方式都可从不同的角度、多个方面对数据库进行查询。

5. 查询结果浏览、输出模块 查询结果的显示分为网站名称的浏览与详细内容的显示两部分,网站名称前用不同颜色的图标表示网站的联通状态;通过点击网站名称可直接进行互联网登录,调用浏览器访问该站,同时网站名称改变颜色以标记此操作;查询结果可选择全部或部分方式,以文本文件的格式保存到磁盘或硬盘。

6. 其他程序模块 系统还提供了一些辅助的程序,如批量装载数据、选择某目录下的某类文件入库、入库数据的预处理等等。这些辅助的程序不仅节省了批量数据处理的时间而且减少错误,保证数据质量。

7. 系统的安装和环境要求 系统包含自动安装程序,可完成自动安装过程。系统环境为windows98,硬件环境在586或以上,内存64M,1G以上硬盘的一般的计算机上均可运行,无特殊要求。

三、结语

在采用面分类法设计的农业网站信息分类方案基础上研制的农业网址数据库及其管理系统,从多维空间信息的分布关系出发,从不同角度揭示我国农业网站的主要特点。同时提供了方便实用的模块,既满足了用户对多属性信息网站的检索要求,又提供了数据查重、录入等数据管理手段,更加容易实现对网站数据库的维护。利用这个系统,我们较好地完成了农业部信息中心开展的我国农业信息网站的普查工作,同时在实践中进一步完善了系统的功能。

第二节 基于光盘的农业信息传播^①

农业信息化建设的内容很广,信息资源的建设是其中的一个重要方面,

^① 周国民,中国农学会计算机农业应用分会副秘书长;郑彦妍,中国农业科学院农业信息研究所多媒体技术研究室副研究员。

如何利用信息技术加快信息资源的建设和传播是当前的研究热点。中国农业科学院科技文献信息中心多媒体室于1998年开始探索使用光盘为载体传递农业信息,在分析农业信息的特点的前提下,使用不同类型的光盘传递不同类型的农业信息,构建一套完整的农业信息传播模式,提出基于光盘的农业信息传播系统,编辑出版各类农业实用技术光盘20余种。

一、光盘及其种类

光盘是用聚焦的氢离子激光束处理记录介质的方法存储和再生信息的介质。英文名称为Compact Disc,简称CD。光盘是一种近代发展起来的不同于磁性载体的光学存贮器,它是继纸张、缩微品、磁存贮器之后问世的又一种新颖的信息存贮载体,是当今用于存贮声音、图像、图表、照片、数据、文字等各种信息的最为理想的介质。

光盘具有存贮容量大、读取速度快、稳定性好、坚固耐用、寿命长和使用方便的特点。一张CD-ROM的存贮容量为550~600兆字节,每只盘片的存贮量在6亿个数据字符以上,随着激光存贮技术的不断发展,同样体积的光盘的容量还在不断增加。光盘存取速度平均每个字符的响应时间约0.5秒。采用激光扫描盘面信息槽进行非接触式读取,多次使用不会损坏盘面,数据稳定性好。另外,盘面上涂有防护层,密封性好,不易受磁场、尘土和手印损害,抗误操作能力也强。理论上光盘可永久保存,对环境无苛刻要求,最少保存寿命达30年以上。

光盘的种类大致可分为电脑光盘、音乐光盘(称作Music CD、CD或Audio CD)、电影光盘(VCD)和数码视像光盘(DVI));而从功能上光盘又可分为只读光盘(CD-ROM)、可记录光盘(CDR)和可重写光盘(CDRW)。当然,光盘的种类也在不断发展,如我国自行研制的拥有自主知识产权的高密度激光视盘系统EVD。

Audio CD光盘的存储对象只能是顺序播放的音乐数据,而不是图形、图像、文本或其他多媒体数据,CD-ROM系列只读型光盘存储器就是适应这种需要而出现的,其光盘外形与CD光盘相同,《黄皮书》中详细规定了它的物理规格及数据格式。对CD-ROM光盘驱动器来讲,数据传输率、缓存大小、寻道时间等是衡量其性能的重要指标。在CD-ROM的基础上,又发展了一系列新的只读光盘存储器。如CD-I、Photo-CD、VCD等,其中CD-I、Photo-CD分别属于飞利浦和柯达的专有技术,在市场上并未普及起来,而VCD由于一些独特的优点,在中国得到了广泛的普及,但VCD也有一些固有的缺点而不能令人满意,如图像的清晰度不够,交互性相对较弱,每张盘至多储存74分钟的活动视频图像。针对这一情况,一些公司和企业纷纷开发出了几种新的VCD格式,如CVD、SVCD、VCD6.0等。

目前,以DVD为代表的新一代光盘存贮技术正在迅速发展。DVD-Video是DVD系列的典型代表,DVD-Video光盘可在DVD-Video播放机和DVD-ROM驱动器上播放。DVD-Video主要有以下几个特点:针对不同的视频制式,有相应的分辨率和帧率;支持标准电视(4:3)和宽屏幕电视机(16:9);支持8种语言的声音及32种语言的字幕;支持多角度功能,重播时可选择多达9种的角度来欣赏;具有多种情节功能,在录入光盘时,可录入多重故事情节,并能在屏幕菜单上显示有关剧情的选择;此外,还具有重锁功能和数字音频输出,支持PCM立体声和杜比AC3。

二、基于光盘的农业信息传播模式

农业信息包括的内容非常广泛,如植物病虫害知识及预防、防治方法和措施;植物栽培技术;植物繁殖、引种、耕作等常识、知识;动物养殖、繁育技术知识;动物传染病预防治疗知识等等。有效开发利用农业信息,通过多种有效途径,建设渠道,传播普及农业信息,将对我国农业生产发展产生直接的推动作用。

基于光盘的农业信息传播模式如图1-1。模式中显示了两条信息传播路径,两条路径的最大区别是光盘载体在传递信息过程中是否利用其他载体(媒体)。这是由于农业信息传播过程中的特殊性决定的。我国农村经济还不发达,能够使用光盘的信息终端还不健全,因此在信息传播时需要一个中转站(服务站)。我们开发的光盘系统中,VCD光盘可以直接被农户使用,这是由于VCD机器的价格比较低,在农村的普及率高。CD-ROM光盘由于需要计算机才能阅读,这类光盘适合农村推广机构使用,并在通过其他方式将信息传递到农户,例如,针对近几年农村电话普及率在提高的现状,我们还开发了将CD-ROM光盘中的信息与电话系统相连的接口,CD-ROM光盘放在推广机构的计算机中,通过计算机与电话系统的接口,农户在家中通过电话访问CD-ROM光盘中的信息。

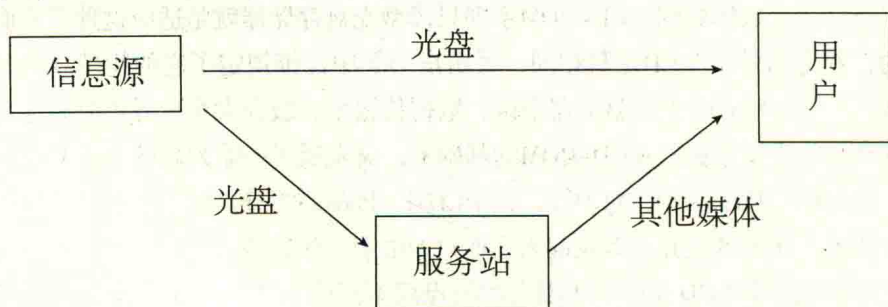


图1-1 基于光盘的农业信息传播模式

三、基于光盘的农业信息编辑出版系统

在农业信息传播模式中，使用CD-ROM/VCD/电脑VCD等形式，如何实现跨格式的信息编辑出版是我们过去几年一直着力解决的问题，并逐步发展了如图1-2的农业信息编辑出版系统，该系统运行于WINDOWS平台，其最大的特点是构建一个综合的多媒体数据库，并且针对不同类型的光盘设计不同的编辑写作系统，编辑写作系统从多媒体数据库中调入相应的素材，编辑生成各种特殊文件格式的发行库，发行库加上专用的浏览器就形成了相应的光盘系统。

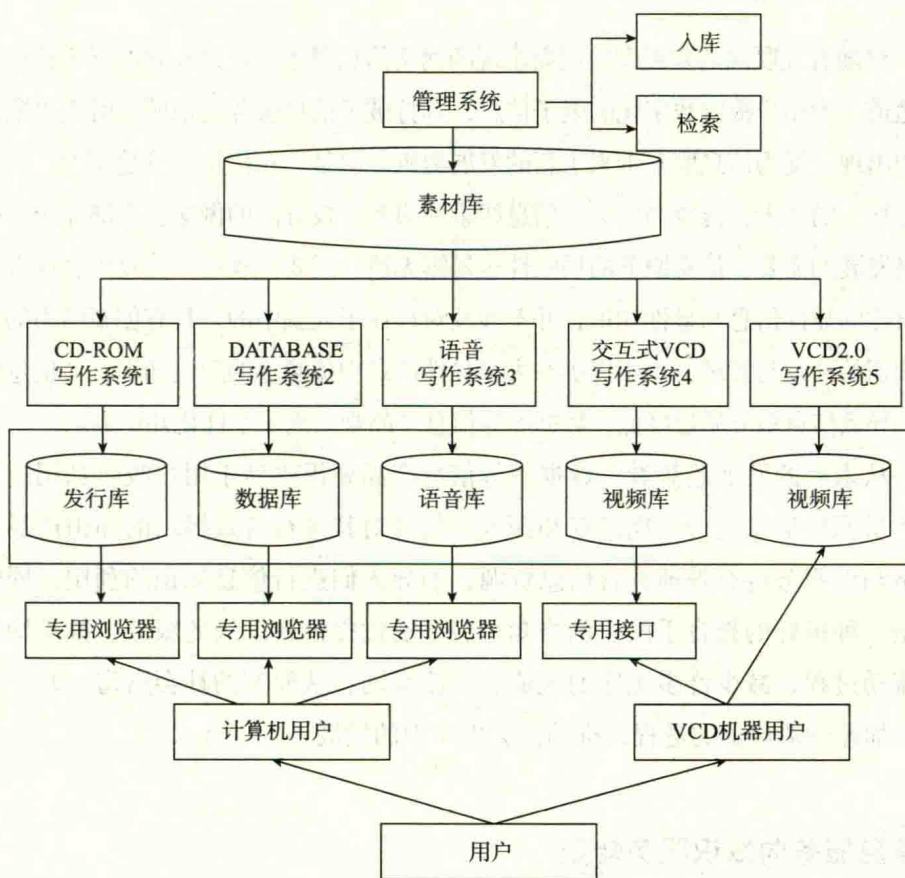


图1-2 基于光盘的农业信息编辑出版系统

四、结语

在农业领域，由于农业信息技术的拓展应用，信息与知识已成为农业生产活动的基本要素和发展动力，信息和技术咨询服务业也正在上升为整个农业结构的基础产业。随着信息和智力活动对农业增长的贡献越来越大，农业

生产方式经历着由资源依附型向信息、知识依附型——信息农业生产方式转变。在这种情况下，一些专家认为研究和开发基于光盘的农业信息传播系统具有非常大的现实意义。他们也将及时跟踪光盘技术的发展趋势，推动信息传播模式的创新，不断完善基于光盘的农业信息编辑出版系统。

第三节 从信息服务向知识服务转变^①

伴随着互联网的发展和应用而出现的网络信息服务，是以互联网络为传递信息通道，为用户提供数字化的电子信息，其打破了信息服务的地域、时空界限，它的出现一度为信息服务带来了新的发展契机。但是，在信息量日趋庞大、“信息爆炸”的今天，传统的以文献信息线索为服务主攻方向的做法已不能完全适应网络发展的需要，信息服务的局限性也暴露无遗，主要体现在：信息内容局限于素材性的显性信息与显性知识，并不涉及对存在于人脑中的、具有创新活力的隐性知识的开发与管理，也不提供在大量显性信息中蕴涵着的隐性知识；“信息爆炸”导致信息服务质量降低；数据库等信息产品难以满足个性化知识需求。

从未来的发展趋势看，数据库等信息产品要提供易于用户吸收利用的、对大量原始信息进行长期跟踪和深入分析并对其进行有效揭示的知识产品。网络知识服务将有效地整合信息资源，引导人们进行信息知识的利用，同时还是一种极好的教育手段，网络知识服务通过提供知识级的服务，推动知识的流动过程，减少许多无序的人流、物流流动，从而节约社会财富，减少浪费，加速创造价值的进程，推动社会生产力的发展。



信息服务向知识服务转变

目前国内网上信息资源的开发利用重点是提供信息服务，很少关注增值（Value-Added）信息服务方面。但是，从信息服务向知识服务的转变，是社会信息化发展的时代要求。社会信息化的发展经历了三个主要的发展阶段：第一阶段是数字化，就是将各种信息用0和1的组合来表示，使信息能够通过设备进行有

^① 王文生，中国农业科学院农业信息研究所副所长。

效的传输；第二阶段是网络化，Internet的诞生使计算机相互联系，共同对信息进行处理，实现信息共享；第三阶段是智能化，这一阶段将是面向用户的而不是面向技术和设备，也就是说把信息服务放在第一位，其后才是产品和技术。在这一阶段，信息服务将向着个性化、本地化、精品化、专业化、层次化、客户化的方向发展。为此，要求Internet信息服务提供者必须向知识服务转型。

一、信息与知识的联系和区别

信息与知识既有密切的联系也有本质的区别。数据、信息、知识是信息增值链上的不同层次，每个层次代表信息加工的不同阶段。一般来说，数据是指分散的、孤立的、无含义的、互不关联的客观事实、文字、数据、图形和符号的集合；信息是经过系统地采集、筛选、组织、排序、归类和分析的数据，它是可视的、物理存在的、经处理后形式各异的、易于传播和复制的，并独立于行动、决策和现实环境；而知识是人们通过实践对客观事物及其运动过程和规律的认识，是有组织的描述推理过程或实验结果且通过一定的通讯介质以系统的形式与他人交流的事实和想法的集合。将信息提升为知识需要根据用户的实际需求，对信息内容进行提炼、比较、挖掘、分析、概括、判断和推论。因此，知识是不可视的，是和用户能力和经验紧密结合的，是与行动、决策、现实环境紧密相关的，能够用于解决问题或创造新知识的信息。信息是知识的基础，知识是信息的核心，是信息升华的结果，也是一种浓缩的系统化了的信息，知识只有通过学习才能够传播和复制。

分析客观的知识存在，知识具有如下特点：

1. 社会实践性

知识源于社会实践，并随着社会实践和科学技术的发展而发展，社会实践也是检验知识的标准。

2. 增值性

知识的有效组织将导致知识大幅度增值。

3. 共享性

借助一定的语言形式或物化为某种知识产品的形式，可以实现知识共享，成为人类共同的精神财富。

4. 关联性

知识点间的关联度是知识的一个重要特征。各知识点间具有关联特性，且不同的知识点间关联度不同。构建知识库时，需要设计相应的方法以表示

各知识点间的关联度。

5. 可增长性

知识随着社会和技术的发展而增长，并不断出现新的学科知识。

6. 时效性

随着科学技术的不断进步，有些知识会过时并落后于时代，因此应该不断地更新知识，使其成为当今最新的、最符合时代要求的知识。

二、知识服务的概念

知识服务 (Knowledge Services, 即KS) 是指从各种显性和隐性信息资源中, 针对人们的需求将知识提炼出来、传输出去的过程。它是面向特定实践应用, 通过采取多种途径与方法, 从大量文献资源的宝库中开发、加工出可以利用的知识资源, 包括那些能被沟通、共享的认识和经验, 以及所组织起来的信息等, 通过对事物本身及内在联系的有序揭示, 提高知识的可见度, 以使用户的吸收、利用和从中获得启迪。它是以资源建设为基础的高级阶段的信息服务。

三、知识服务与信息服务的区别

通常认为, 信息服务是以向人们提供有用的显性信息为内容的信息传播过程, 而知识服务是指从各种显性和隐性信息资源中, 针对人们的需要将知识提炼出来的过程。与传统信息服务的最大不同是, 知识服务首先是一种观念, 一种认识和组织服务的观念, 是以资源建设为基础的高级阶段的信息服务。它不只停留在简单的信息积累、加工和传递的低层次上, 而是转向了高层次的知识开发与利用。在实现这些服务功能中, 人们充分重视信息服务中的智力参与, 增加服务中的知识因素, 使服务成为发现和培育新想法、新思维的过程。知识服务为数字化、网络化的环境实现从文献层面的服务转变为知识层面的服务提供了机会和可能。它是根据用户的需求, 为他们的学术研究和产品开发, 提供专项的、动态的、持续的信息服务, 这些信息被用户接受之后, 往往会产生一定的学术价值和经济效益, 而且是具体可见的甚至可以用经济指标来衡量的价值, 因此有人也称知识服务为效益性服务。

知识服务与信息服务之间的最大差异在于: 信息服务通过图书馆书目、联机检索、搜索引擎、网上浏览等提供信息, 这些信息对用户可能有用, 也可能无用甚至有害; 而知识服务则要求对用户所需的主题更贴切、更准确, 对问题求解更直接有用, 能够为人们带来时间、财富、效率、效益的知识。