

信息存取原理与技术

Principle and Technology of Information

(PIM)

孙 平

文献检索教研组

2002. 3

科技文献检索与利用

Sci-Tech Information Access and Application

Information, Info, 情报, 资讯

LI: Library Instruction

BI: Bibliographic Instruction

UI: User Instruction

Library Literacy

Information Literacy(信息素养):

Information Literacy Defined

A set of abilities to

- Recognize when information is needed
- Locate and evaluate information
- Effectively and responsibly use information

Information Literacy includes

- Library literacy
- Media literacy
- Computer literacy
- Internet literacy
- Research literacy
- Critical thinking skills

Five Standards (ACRL, USA)

- Determines the nature and extent of the information needed
- Assesses needed information effectively and efficiently
- Evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system
- Individually or as a member of a group, uses information effectively to accomplish a specific purpose
- Understands many of the economic, legal and social issues surrounding the use of information and accesses and uses information ethically and legally

Five Standards (ACRL五个标准)

- 确定所需信息的范围
- 有效地获取所需的信息
- 鉴别信息及其来源, 将检出的信息融入自己的知识基础
- 有效地利用信息去完成一个具体的任务
- 了解利用信息所涉及的经济、法律和社会问题, 合理、合法地获取和利用信息

Expected Outcomes of Information Literacy

Information literate people have learned how to learn

They know how to find and use information

They are life-long learners

They can find information for any task and use it to make decisions

前言 Preface

- 课程特点 Feature

目的意义：应用科学 能力培养

教学方式：讲课

自学(多媒体教学软件)

实习(计算机实习报告)

- 课程内容Content

检索基础知识

计算机检索技术

网络信息资源检索利用



两个应用网站

- 清华教育在线-网络教学: <http://166.111.37.253>
进入：多媒体教学软件,“网络教材”（课件）
- 清华大学图书馆主页：
<http://www.lib.tsinghua.edu.cn/>
进入：本馆电子资源，选择数据库检索

教科书与参考书

《电气与电子工程文献检索与利用》孙平，大连理工大学出版社，1999

- 参考书：
 - 《科技信息检索》孙平等，清华大学出版社，1997.
 - 《科技文献检索》赖茂生等，北京大学出版社，1994.
 - 《计算机信息检索》李莹等，机械工业出版社，1997.
- 检查考核：作业、练习

重要检索工具(数据库)

四大重要检索评价工具：Ei, SCI, ISTP, ISR

ISR: Index to Scientific Research

- 《科学引文索引》《Science Citation Index》(SCI)
- 《工程索引》《The Engineering Index》(Ei)
- 会议文献的检索：《Index to Scientific & Technical Proceedings》(ISTP)
- 《科学文摘》《Science Abstracts》(SA)
- 科技报告的检索：
《Government Reports Announcements and Index》(GRA&I) / 《NTIS》
- 专利文献的检索：《World Patent Index》(WPI) /
《DERWENT Innovations Index (DII)》《IBM Patent Server》《USPTO
Patent Databases》
- 学位论文的检索
《PQDD》《中国学位论文库》



第一讲 基础知识

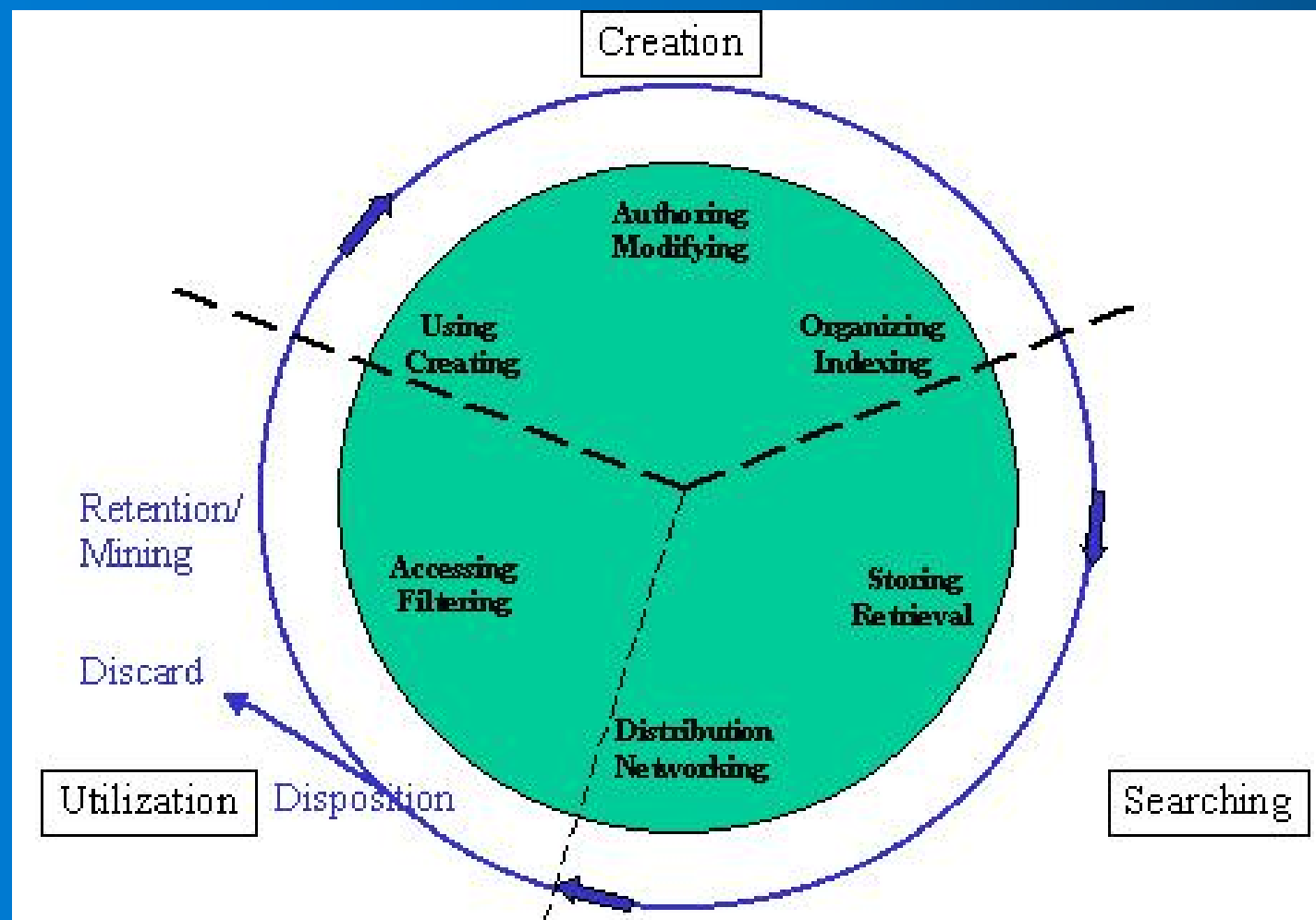
1 信息

1. 信息的结构



- Data—信息的原材料
- Information—由人提供的有组织的数据
- Knowledge—对信息的理解与认识
- Wisdom—精炼的知识

2. 信息的循环和增值



3. 信息量

- Library of Congress: ~20 Terabytes
- Dialog: ~9.2 Terabytes
- Web: ~4 Terabytes (98)

Gigabyte 10^9

Terabyte 10^{12}

Petabyte 10^{15}

Exabyte 10^{18}

4. 信息量过载的问题(Information Overload)

- 网络环境下，越来越多的电子信息的生产，并为信息共享提供了优越的条件。
- 信息资源质量的严重不均衡。
- 过量的信息与没有信息引起的困惑同样是问题。
- 人们必须在有限的时间范围，有效地获得自己所需要的信息，排除不相关的信息。
- 需要专门的工具来帮助人们完成这些工作。

二. 文献

1. 定义

文献：记录有知识的一切载体

媒体 media

2. 功能：传递, 评价 (财富, 凭证)

3. 状况：数量 分布情况 运行环境

information highway, paperless society,

virtual library, digital library

• 数量

(Textbook p14, p252)

世界文献出版情况：Book in Print

Ulrich's International Periodicals

我国的文献出版：中国出版年鉴

世界重要文献收录情况：

四大重要检索评价工具：Ei, SCI, ISTP, *ISR*

ISR: Index to Scientific Research

清华大学的文献信息资源

- 清华大学图书馆建馆于1912年，有近90年的历史
- 总藏书量 260万册
- 线装古籍书30万册；
- 期刊合订本35万余册；
- 年订购现刊5千种；
- 本校博、硕士论文14,000篇；
- 缩微资料7万余件；音像资料5千余盒；
- 多媒体光盘2千余种；
- 各类数据库150余种；
- 中外文电子期刊10,500种

清华大学的文献信息资源

- 近年来电子资源（网上服务）大幅度增加，全文电子期刊总量达8,700多种（外文期刊：5,200种；中文期刊3,500种）
- 其它数字文献：会议录、科技报告、专利等购买的中外文文摘索引数据库150多个，基本覆盖全校的设置的所有学科
 - 电子/电气、计算机、自动化、生物、材料、化学/化工、地球与环境科学、汽车、机械、医学/医药、农业、经济、核科学、人文、金融、管理、商业... ..

清华大学的文献信息资源

- 1998年11月国家发展计划委员会正式批准立项：中国高等教育文献保障体系（CALIS），主要起到文献信息保障基地的作用
 - 全国性四个文献信息中心：文理、工程、农学、医学
- 清华大学是四个中心之一：全国工程文献中心。承担引进数据库、自建数据库、文献保障、培训等主要任务。

4. 形式

- 印刷型 Printed form
- 缩微型 Microform
- 声像型 (视听型) Audio-Visual form
- 机读型 (电子型) Machine Readable form

印刷型 Printed form:

易携带性(portability)

易阅读性(readability)

做标记性(markupability)

可支付性(affordability)

存档性(archivability)

占有性(ownability)

持久性(durability)

缩微型 Microform:

- 形式: Microfilm 缩微胶卷 microfiche 缩微平片
- 缩微: 国际标准型 1: 24 $14 \times 7 = 98$ 页
- 存储: 存储空间1/10, 存储密度高, 易保存和流通
价格低、管理方便。
- 处理: Microphoto microcopy 纸介文献的还原
- 阅读: 阅读机 (reader) 还原机(reader-printer)
- 缩微公司: UMI(University microfilm International)
2.7万报刊缩微品

声像型（视听型） **Audio-Visual form:**

- sound & music
- image
- animation
- audio & video
- **multimedia**

Compact audio/video form

机读型 (电子型) Machine Readable form Electronic Publication :

电子版本:

floppy, Magnetic tape, CD-ROM, online, network

可存取性 (accessibility) 可获得性(availability)

即时性(immediacy) 及时性(currency)

可研究性(researchability) 广博性(encyclopedic)

学科交叉性(interdisiplinarity) 动态性 (dynamism)

合作性(collaboration) 可链接性(linkability)

交互性(interactive) 多媒体性(multimedia)

5. 级别:

一次文献/信息 primary Literature/ info

二次文献/信息 secondary literature /info

三次文献/信息 Tertiary literature/ info

一次文献/信息 **primary Literature/ info:**

概念: 第一次书写的、第一次报道的或第一次出版的那些文献; 以作者自己的研究成果、工作经验和实践体会为原始依据, 信息内容具体、丰富。

特点: 新颖性、创造性、系统性

如: Journal literature Sci-Tech report,
 Proceedings Patent specification
 Paper Thesis Dissertation ...

二次文献/信息 **secondary literature /info:**

概念: 对一次文献的加工、提炼和浓缩

特点: 浓缩性、汇集性、有序性

条目 entry, reference, citation, record

- 文摘 Abstract
- 索引 Index
- 目录 Catalog
- 题录 Bibliography

三次文献 Tertiary literature/ info:

概念: 对一次文献和二次文献进行汇集、综合、
分析、详述

如: Reference book

文献加工的有序化过程

6. 类型与识别

科技图书 Sci-Tech book

Reading Book: Monograph, Anthology, Textbook...

Reference book:

- 科技期刊 Sci-Tech Journal (magazine*, serial, periodical...)
- 科技报告 Sci-Tech Report
- 专利文献 Patent Document: Patent specification/bulletin/aids
- 会议文献 Conference paper: Proceedings, Compilation, Collected paper, Records...
- 学位论文 Dissertation, Thesis
- 标准文献，产品资料，技术档案，政府文献

举例：

例1.

Gong, Victor, and Norman Rudnick, Eds. AIDS: Facts and Issues. New Brunswick: Rutgers UP, 1986.

例2.

Schware, R. “Climate Change in Water Supply: How Sensitive is the Northeast?” Climate Change and Water Supply. Washington, D. C.: National Academy of Sciences, 1986: 82-98 .

举例：

例3.

Chun, Lu, and K. Y. Lan, “ Dynamic Analysis of Clamped Laminated Curved Panels.” Composite Structures. 30.4 (1995): 389-396.

ISBN: International Standard Book Number 0-471-81086-X

ISSN: International Standard Serial Number 1003-3513

CN11-2856/G2

举例

例4-1

Duan, G, et al, Program for Control System Simulation.
Report NIST/TN-2740, 1994.

例4-2

Wang, C. M., et al, Software for Performing Gray-Scale
Measurements of Optical Fiber End Faces, Report
NIST/TN-1370, Washington, DC: Nat. Inst. Stand. &
Technol, 1994.

举例：

例5.

Gough, P. A., et al, “ Electrothermal Simulation of Power Semiconductor Devices.” Proc. of the 3rd International Symposium Power Semiconductor Devices & ICs. Ed. Shbib, M. A. New York: IEEE, 1991: 89-94.

例6.

Zaccarin, Andre and Liu Bede, Method and Apparatus for Determing Motion Vectors for Image Sequences. U. S. Patent 5210605, 1993.

举例：

例7.

Dong, Siyi Terry, The Modeling Analysis and Synthesis of Communication Protocols. Diss. Michigan U, 1991, Ann Arbor: UMI, 1992. 9228854.

例8.

Champa, K. S., “Mordrian in Theory and Practice.” Rev. of The New Art-The New Life, by Mondrian, P. New Criter 6 Sept. 1987: 47-47

规则：

顺序： **Author(s), Title, Source**

参照：《MLA Handbook for Writers of Research Papers》

MLA: Modern Language Association

变通：30. 4 = 30(4) = 30: 4 = Vol. 30, No. 4
= v30, n4

三. 文献信息检索

Information Storage & Retrieval / Access

查询方法

- 直检法：专业核心期刊跟踪
- 常规法：顺查法, 倒查法, 抽查法
- 追溯法：参考文献, 引用目录
- 综合法：分段法, 循环法, 交替法

1. 检索的概念：

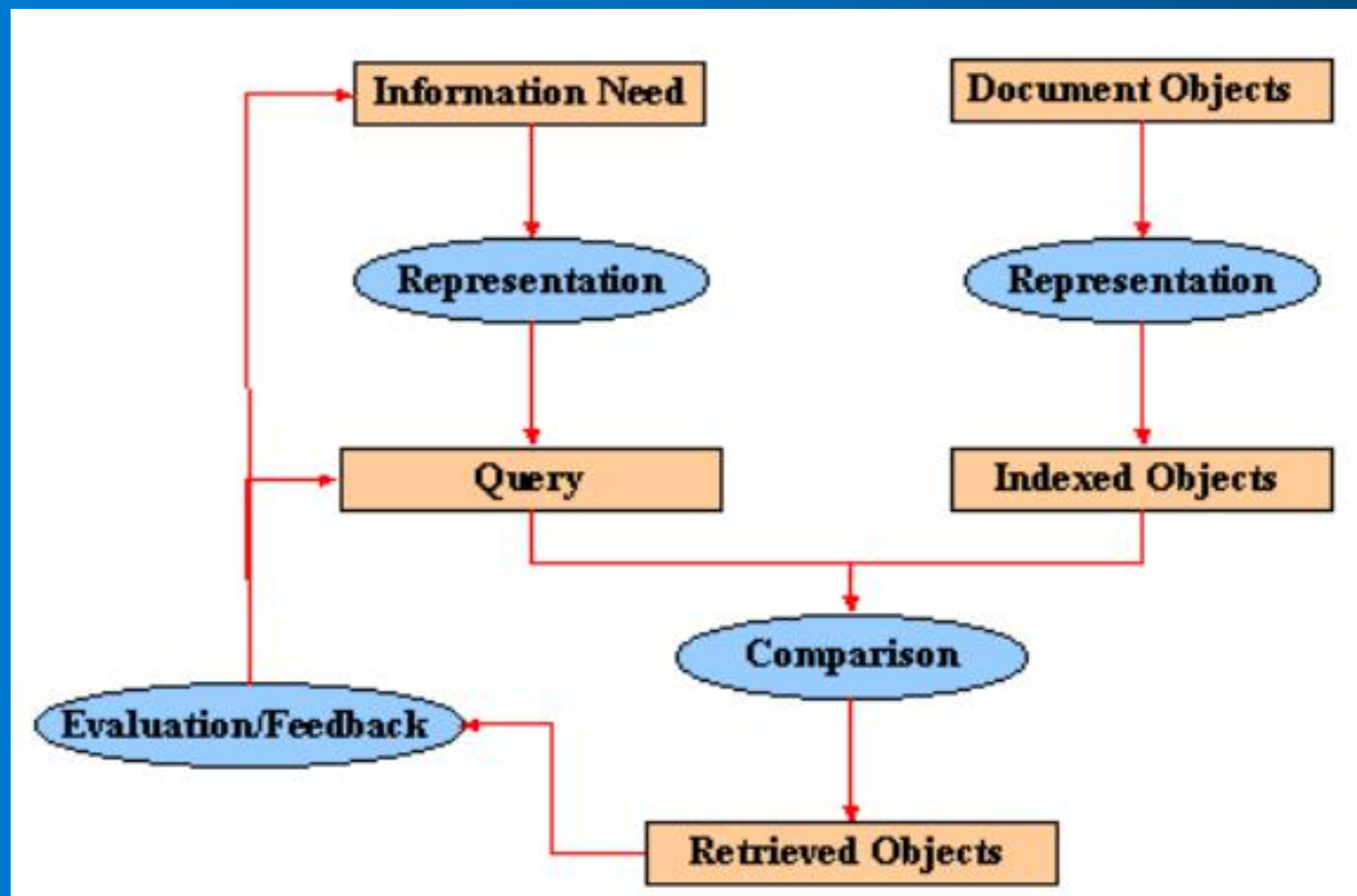
(1) 检索的本质 匹配：信息需求 ~ 信息集合

(2) 检索的手段 手检：书本型工具(刊)
 机检：数据库系统

(3) 检索点(途径)：Access Point

用文献各种内外特征,作为检索出发点,从不同角度来检索文献信息.

信息存取/比较匹配

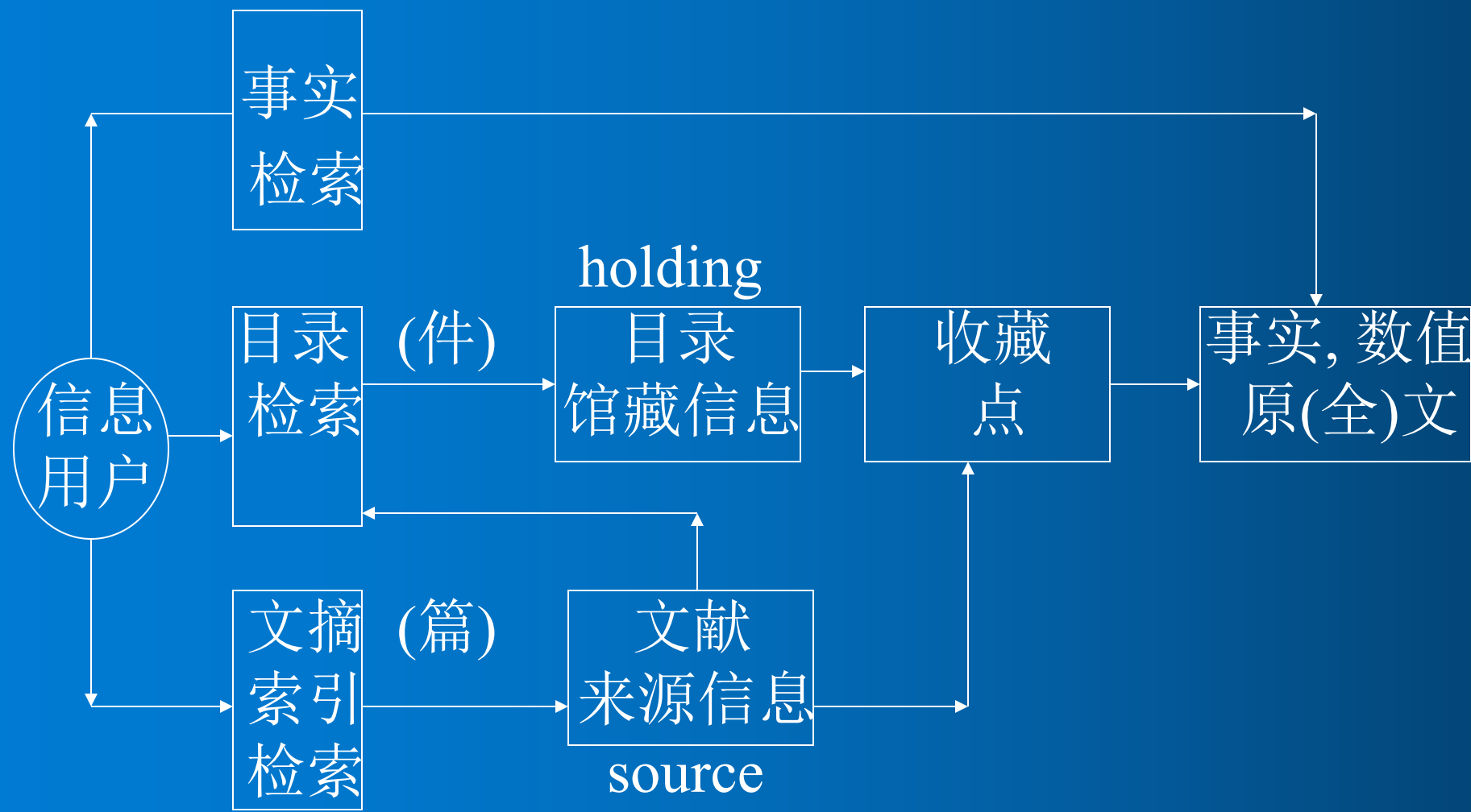


(3) 检索点:

- 分类(Classified): 所属学科性质、类别 分类体系
检索点: 分类类目对应的类名、类号。 现期检索
- 主题(Subject): 所属主题内容
主题词的字顺(音、形) 回溯检索
- 作者(Author)
- 团体著者(Corporate Author/ Authors' Affiliation)
- 名称(Title)
- 代码(Code/Coden)...

2. 检索的类型

事实	fact	}	直接, 确定
数据	data		
全文	full-text		
目录	catalog	}	间接, 相关
文摘	abstract		
索引	index		



(1) 事实检索

Reference book

Database:

Numeric Database

Property Database

Textual-numeric Database

Full-text Database

Terminological Database

Graphic Database ...

(2) 目录检索 (收藏, 出版) 图书/报刊/资料目录

出版物:

- 《中国国家书目 1995》 北图编 1997出版, 3卷, 有网络版 60万种图书
- 《全国新书目》《全国总书目》 中国版本图书馆编 1994
- 《在版书目》(Book in Print) 美国 R. R. Bowker 9卷
每年一版, 1996第49版, 美国在版书目总汇: 2.7 万
出版社的80多万种图书。有缩微版, 光盘版, 联机版
- 《乌利希国际期刊指南》(Ulrich's International Periodicals Directory)
美国R. R. Bowker公司1932年出版, 1996年第35版年, 收录
世界 200多个国家的15.6万种期刊
- 《中文核心期刊要目总览》 北京大学出版社1996年, 第2版,
收录我国出版的1578种核心期刊

收藏：馆藏目录

联合目录(Union Catalog)

媒体：印刷型目录

机读型目录 MARC

联机公共检索目录 OPAC

(Online Public Access Catalog)

OCLC(Online Computer Library Center):

(2001年4月报道)使用OCLC产品和服务的用户已有
70多个国家和地区的39,517个图书馆和教育科研
机构

目录检索点：

Author	Title	Call Number
Subject	Keyword	ISBN ISSN...

例：Call Number

书次号 →

N2-62
FS41

 N2—62 FS41
CIP: Cataloging In Print

国图书馆图书分类法(中图法)

利用OPAC检索

- [清华大学OPAC](#)

检索点：

[作者](#) [索书号](#)

[题名](#) [文献号](#)

[主题](#) [ISBN/ISSN号](#)

[关键字](#)



图书分类法：

- 《国图书馆图书分类法》（中图法）
- 《中国科学院图书馆图书分类法》（科图法）
- 《中国人民大学图书馆图书分类法》（人大法）
- 《Dewey Decimal Classification System》
(杜威十进分类法)
- 《Library of Congress Classification》
(美国国会图书馆图书分类法)
- 《Universal Decimal Classification》
(国际十进分类法)



3. 原文获取：

ILL/DD:

ILL: Interlibrary Loan (馆际互借, 清华大学图书馆ILL)

DD: Document Delivery (文件传递)

Mail, Fax, E-mail, Network...

联机检索/借阅/订购/电子传送

End-user, intermediate

一步到位(one-stop)

ohioLINK: User-initiated interlibrary borrowing

4. 特种文献收藏

专利文献：专利局文献馆

科技报告：中国科技信息研究所(ISTIC) (中情所)
北京文献服务处(国防科技信息中心)

会议文献：北京图书馆
大学/研究型图书馆(academic library)
中科院图书馆
邮电大学图书馆

学位论文：北图，中情，授予机构/院校

5. 检索系统/工具的类型

目录	catalog	(item)	外部特征
题录	bibliography	(article-level)	外部特征
文摘	abstract		内、外部特征
	Information / indication	/review	
索引	index		
	信息及其物理位置的指示物, 提供指示系统, 使用户准确找出信息集中的特定信息.		
要素:	标目, 说明语, 储存地址		

6. 检索语言与词表

Searching Language & Thesaurus

描述检索系统中信息特征与表达用户信息提问的一种专门语言

自然语言 natural Language 非规范

人工语言 Artificial Language 规范

描述文献信息内容特征的语言：

标题词语言 Subject Heading

单元词语言 Uniterm

叙词语言 Descriptor

关键词语言 Keyword

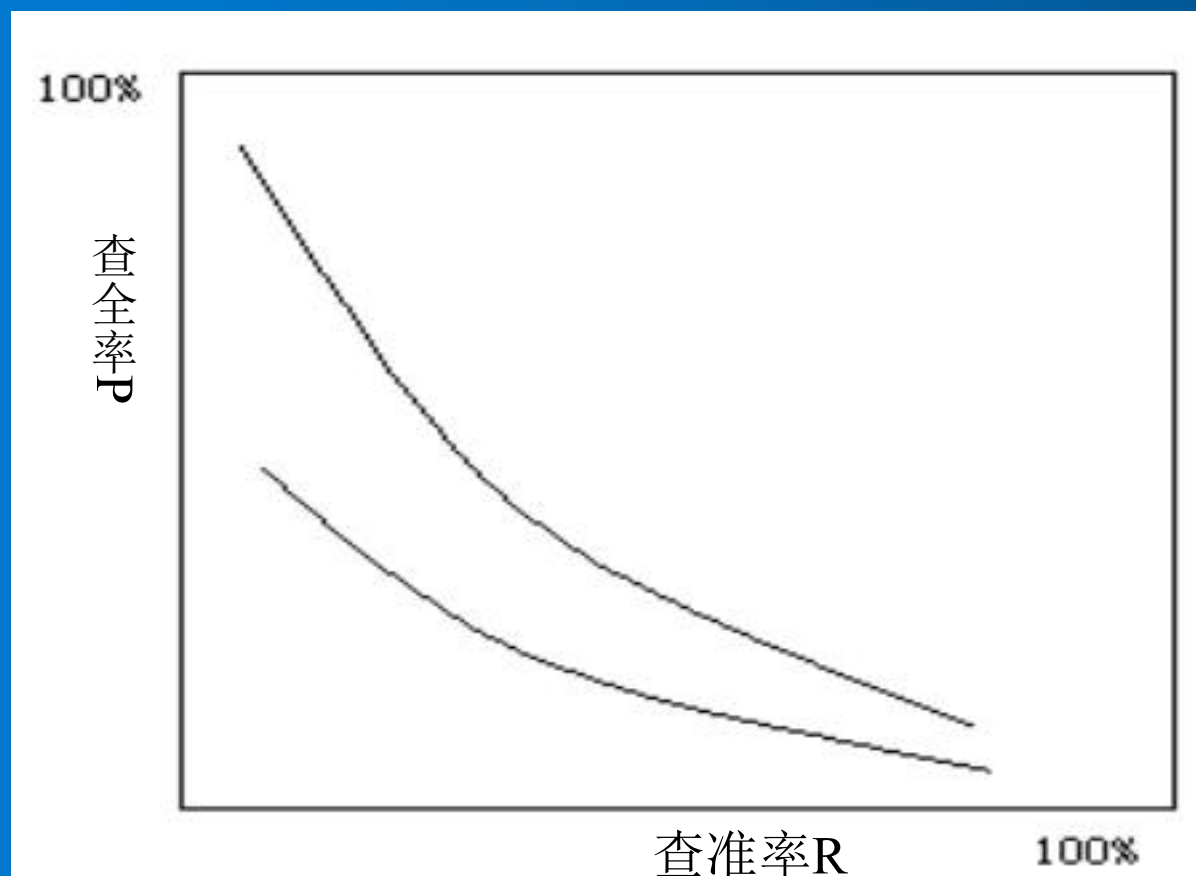
分类语言

7. 检索效果

	相关文献	关文献	总计
检出文献	A	B	A+B
未检出文献	C	D	C+D
总计	A+C	B+D	A+B+C+D

查全率(Recall Ratio): $R = A / (A+C) * 100$

查准率(Precision Ratio): $P = A / (A+B) * 100\%$



提高措施：工具质量，检索技能，指标调节