Practices and Open Problems of Library Collection Digitization

Xiaohui Zheng
Tsinghua Univ. Library
Oct. 2007

Facts of Organization

- In 1999, THUL began the digitization testing
- In 2001, First DL project was funded by Univ. which was named “Tsinghua Architecture Digital Library”. Consequently several DL projects were planned.
- In 2003, Digital library Research Group was established to plan projects and strengthen research on theories and technologies.
- March of 2003, Xianzhi Resource Digitization Center was established. 3T data was produced till now.

Digital Library Research Group

Main Task

- Organizing, coordinating and constructing digital initiatives planned in library
- Planning, Coordinating, Organizing digital library projects collaborated with other partners
- Collection retrodigitization and reformatting
- Research on resource organization and management
- Research on intellectual property
- Research on web presentation and infrastructure standards.

Digital Library projects planned in Lib

The Digital Library of Chinese Science & Technology History
- Architecture
- mathematics
- Machine and mechanics

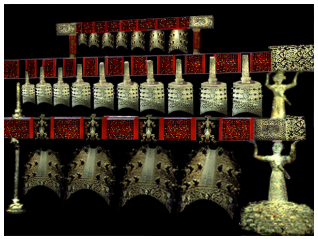
Teaching resources DL
- THU Dissertation Databases
- course information and reference
- textbook

THU special resources DL

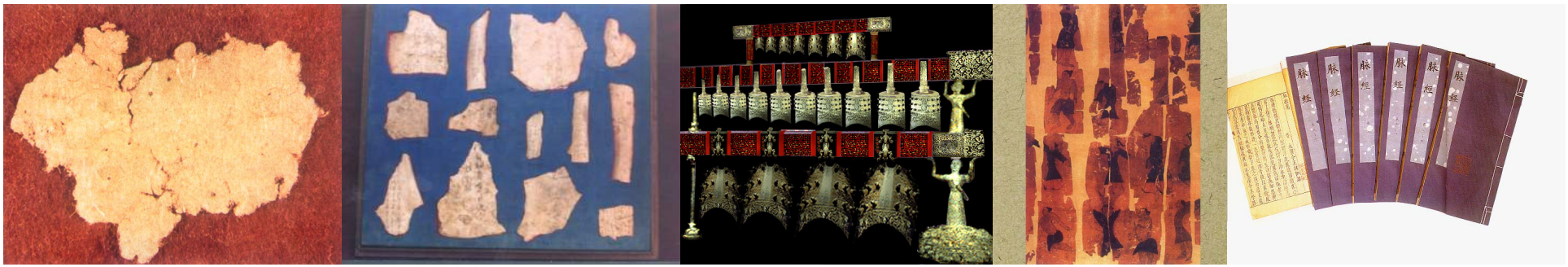
The Digital Library of Chinese Science & Technology History

Main subjects on history of Chinese sci. and tech.

Agriculture, Medicine, Astronomy, Mathematics, Biology, Geology, Physics, Chemistry, Architecture, Hydraulics, Fabric, Machine and Mechanism, Metallurgy, Printing,Weapon,…



Origin of the Project



Main types of the contents：

Culture relics, Oracle, Bamboo books, Ancient printing books, manuscripts, cards, pictures, drawings, etc.

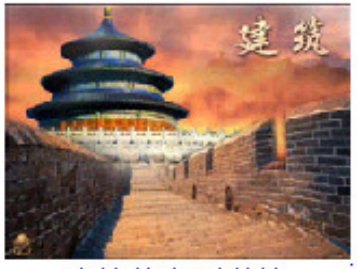
中国科技史数字图书馆 - Microsof 西文 科能软件技术有限公司...

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 前进 停止 刷新 主页 搜索 收藏 历史 邮件 字体 打印 编辑 讨论

地址(D) http://dbgroup.cs.tsinghua.edu.cn/millionbook/start.htm 转到 链接

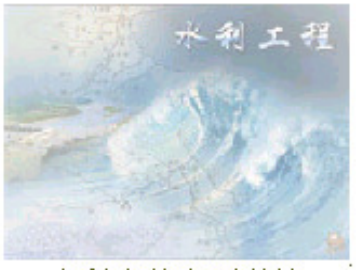
中国科技史数字图书馆



建筑数字图书馆
Architecture DL

数学史数字图书馆
Mathematics History DL

机械史数字图书馆
Mechanics History DL



水利史数字图书馆
Hydraulics History DL

工艺史数字图书馆
Arts and Crafts History DL

陶瓷史数字图书馆
Chinaware History DL

http://dbgroup.cs.tsinghua.edu.cn/digital Internet

开始 En 18:24

Tsinghua Architecture DL

Tsinghua Architecture Digital Library （THADL） was funded by Tsinghua Research foundation.in Nov. 2000, and finished in Oct. 2002.

--A Cross-disciplines project collaborated by Library, Computer Science dept. and Architecture Dept.

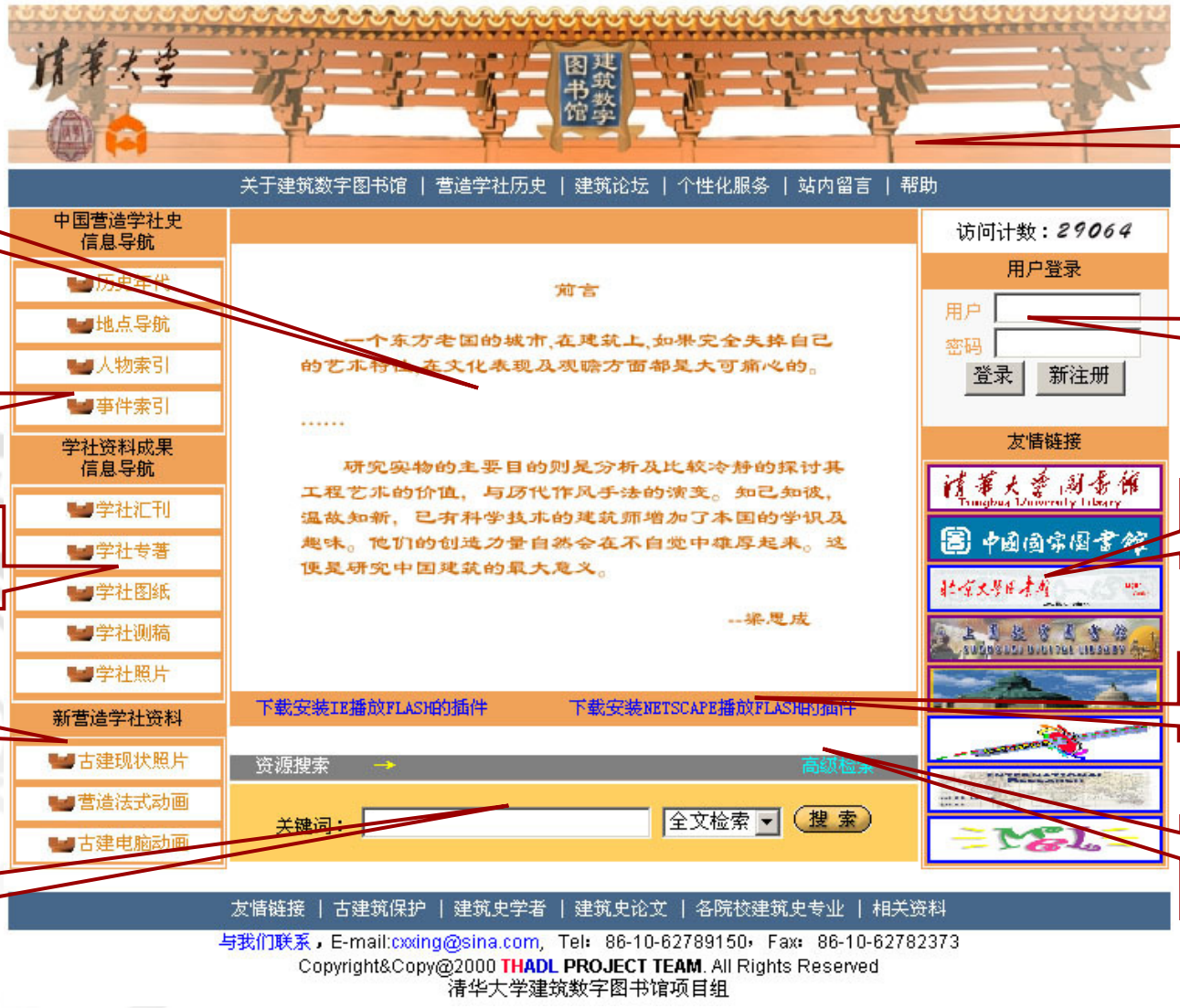
Tsinghua Architecture DL

Achievements:

- About 20000 pieces was digitized.
- DC based metadata schemas were drafted for several types such as monograph, paper, photo, drawing and animation.
- Prototype system was built which includes the content creation and indexing, information organization, search and retrieval, web publishing and rights management functions.
- Learning environment was provided for Architecture History courses.
- THADL website:

http://dbgroup.cs.tsinghua.edu.cn/digital/

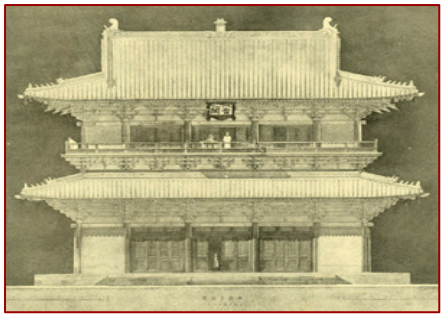
Homepage of THL Architecture DL



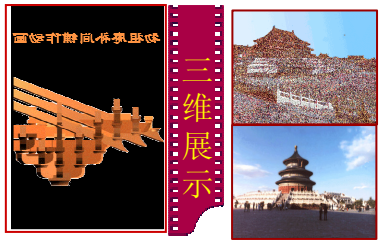
Guanyin pavilionin built in the year of 984

Photo in 1931

Photo in 2000



Drawing in 1931



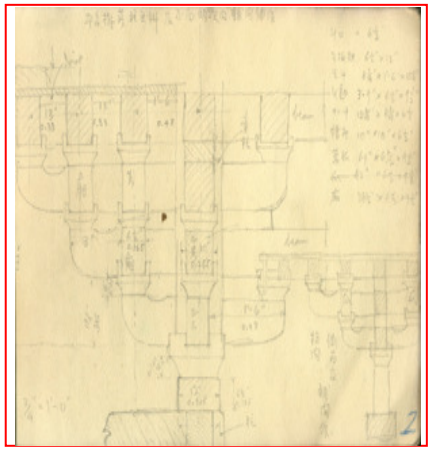
Video and Audio

Dou gong

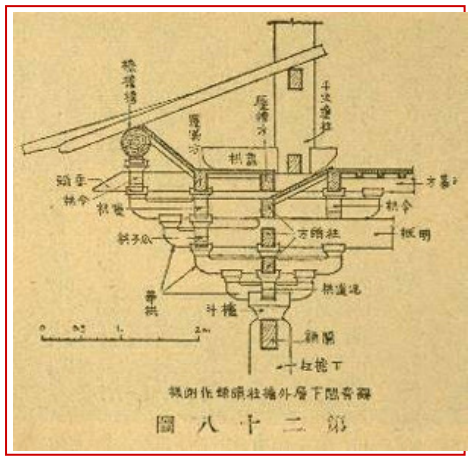
Photo in 1931

Photo in 2000

Photo in 2000

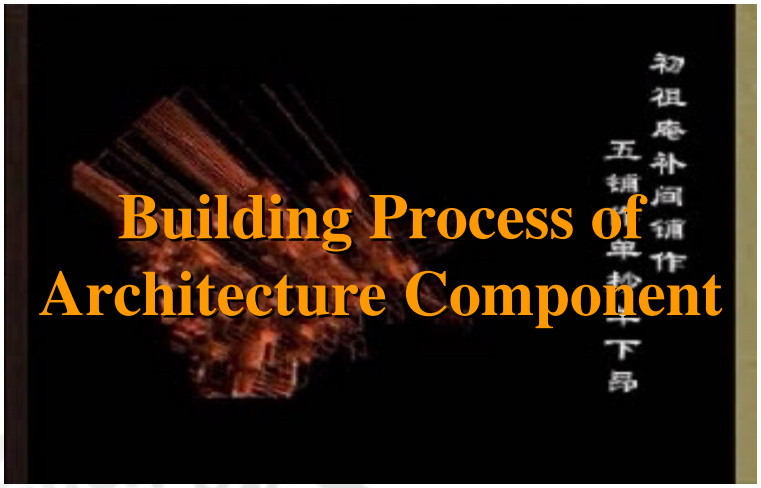


Sketch in 1931

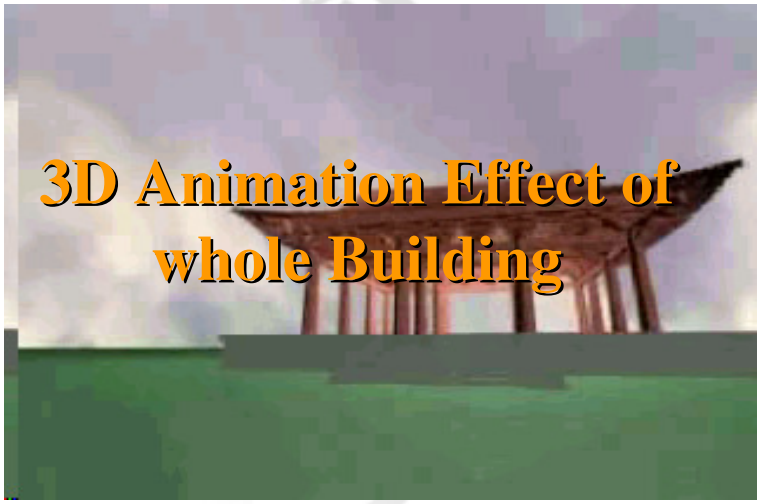


Drawing in 1931

Tsinghua Architecture DL



Building Process of a part of Architecture



Digital Library of Chinese Math. History

EMANI

Electronic Mathematics Archiving Network Initiative

Main Goals：

To retrodigitise printed publications in mathematics.

To care about the long-term preservation of this content in readable form.

To improve the usability of the retrodigitised publications by introducing advanced linking and searching facilities.

Partners:

The content provider Springer-Verlag and EMIS proposed the project.

Members:

Cornell Univ., USA	Goettingen State Univ., Germany
Orsay Library, France	Tsinghua Univ., China

In Nov. 2001, Tsinghua Univ. Library as invited partner involved in:

- Digitization of the Chinese Math material.
- To participate in establishing technical standard and specification.
- As a mirror site to store the digital content in mathematics from partners.
- Provide convenient and affordable access to the stored content for mathematicians and professionals using mathematics world-wide.

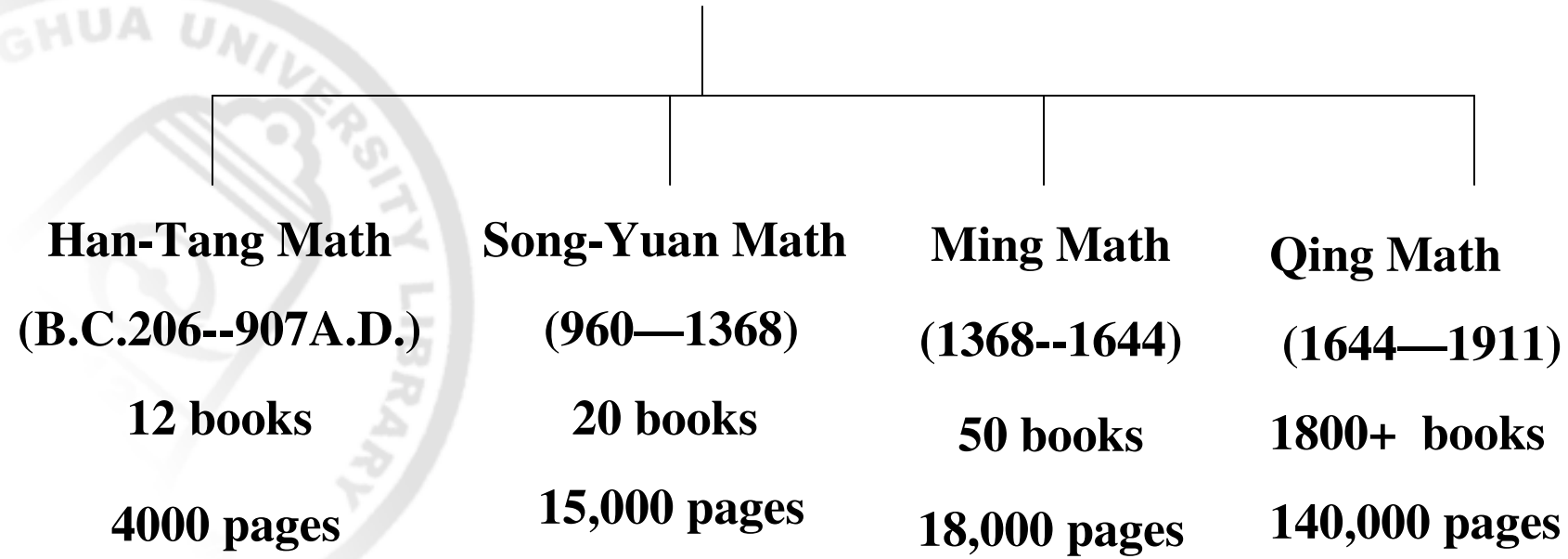
Digital Library of Chinese Math. History

Main Achievements:

- **Investigation of the materials:**

Existing Ancient Chinese Mathematical works is about 1910

—Statistics from "Series of the history of mathematics in china App. Vol. 2"



❑ Digitization progress

☐ Finished 150 E-books, 70000 pages digitizied.

☐Finished 3 indexing databases: Union Catalog on Ancient math., Research paper index on Chinese Math History. Tsinghua Catalog on Ancient Math.。

☐ Designed 20 animations on ancient calculation processing, such as Pythagoras' theorem, Sphere's computing method, Cube's arithmetic Multiplication.

CMDL Website:
http://166.111.205.16:8081/mathdl/default.htm

Interoperability

- METS based encoding cataloging toolkit was published.
- 52 Chinese Math Journals were localized and published together with other EMANI partners on website www.emani.org

Homepage of DLCMH

详细信息 - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 搜索 收藏 历史

地址(D) http://166.111.68.83/millionbook/search/Record.asp?QueryString=&PageNo=1&URL=九章卷 转到 链接

返回到搜索结果

正题名：	九章算术九卷
主要责任者：	（晋）刘徽注
次要责任者：	（唐）李淳风注释
序跋者：	御制武英殿聚珍版十韻有序；御制题九章算术有序；刘徽九章算术注原序
收藏印记：	刘复所藏；江阴刘氏；镏家书库；刘半农藏书；清华大学图书馆藏；
版本：	清乾隆三十九年武
出版者：	（清）武英殿
出版日期：	清乾隆三十九年
载体特征：	4册1函；线装；
来源：	原为刘半农藏书；现藏于清华大学图书馆
馆藏信息：	清华大学图书馆

阴：
hidden;
lunar;
negative;
overcast;
shade;
in intaglio

Internet

开始 PPT 详细信... DL(V7) BEIJINGL 古今数... CH 16:51

Digital Library of Chinese Machine and Mechanisms History

In 2003, the project DLMMH got fund from the library academic committee.

- Famous scholar Prof.Liu Xianzhou's study on Chinese machines and mechanisms history provide a navigator on the collection organization.
- In 2004, Investigation of the collection was started, Metadata schema was drafted.

Contents in the project：

-- 40 related monographs ；

Such as *Pictures of Plowing and Weaving* (Rare book, 2 vol. 46 pictures, 1662-1722)

-- 12554 Index Cards collected by the library in 50’-60’s

-- Prof. Liu Xianzhou’s monographs, papers, Manuscripts, pictures, book index, etc.

-- Animations and pictures on reconstructive models of ancient machines and mechanisms.

DLMMH Website is:

http://elib.lib.tsinghua.edu.cn/techlibrary/htm/index.htm

Homepage of Chinese Machine and Mechanisms History DL

http://elib.lib.tsinghua.edu.cn/techlibrary/htm/index.htm

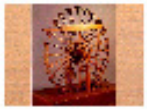
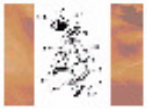
机械史数字图书馆

机械技术典籍
机械工程简史
研究论著索引
研究论文选载
机械史图像库
工程技术史料

检 索

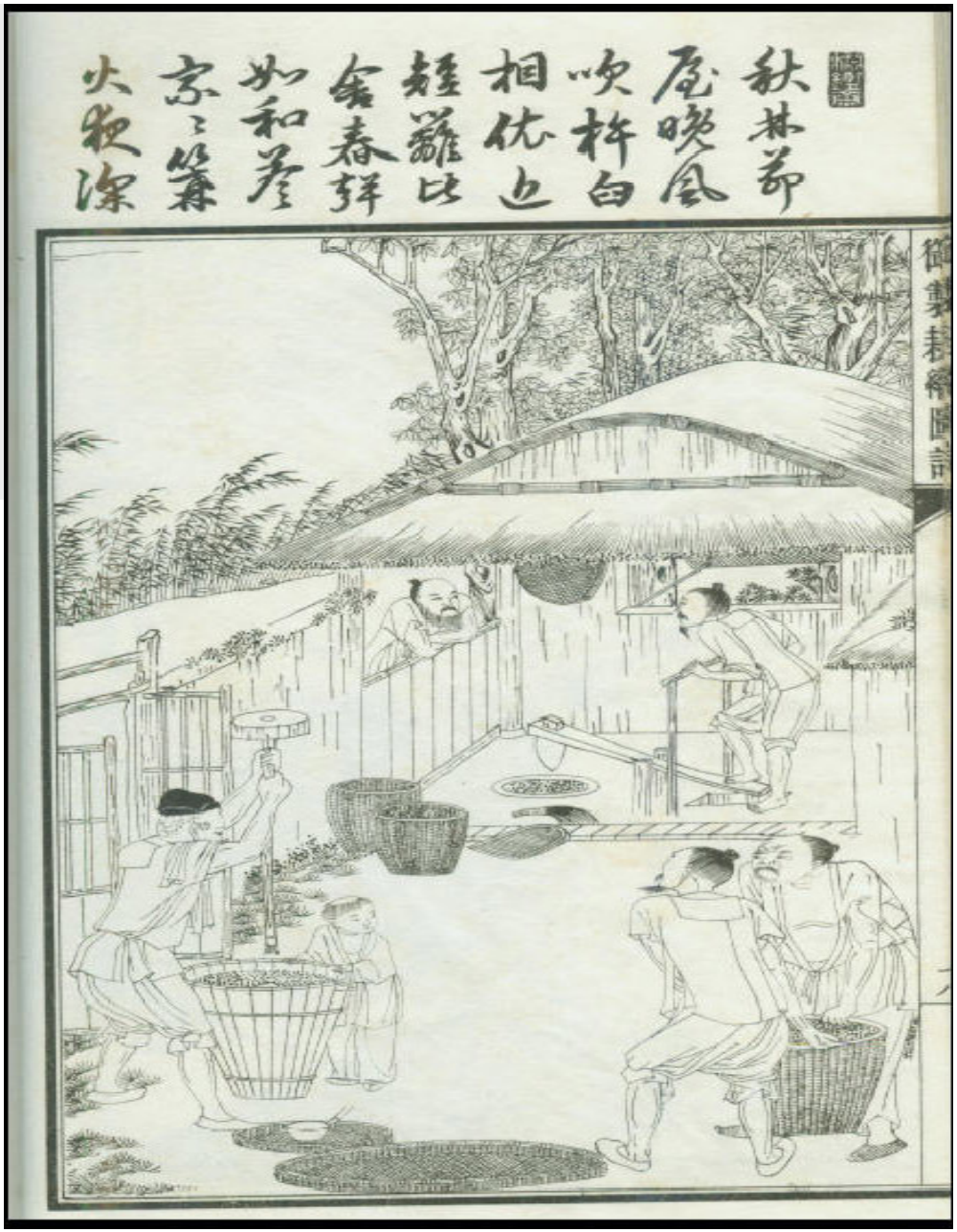
最后更新：2004年3月18日

机械史图像库

农业机械	纺织机械	车辆	船舶	军事机械	天文仪器	其他机械
江东犁	纺车	独轮车			水运仪象台	被中香炉
		铜车马				风箱
						
		陶牛车				

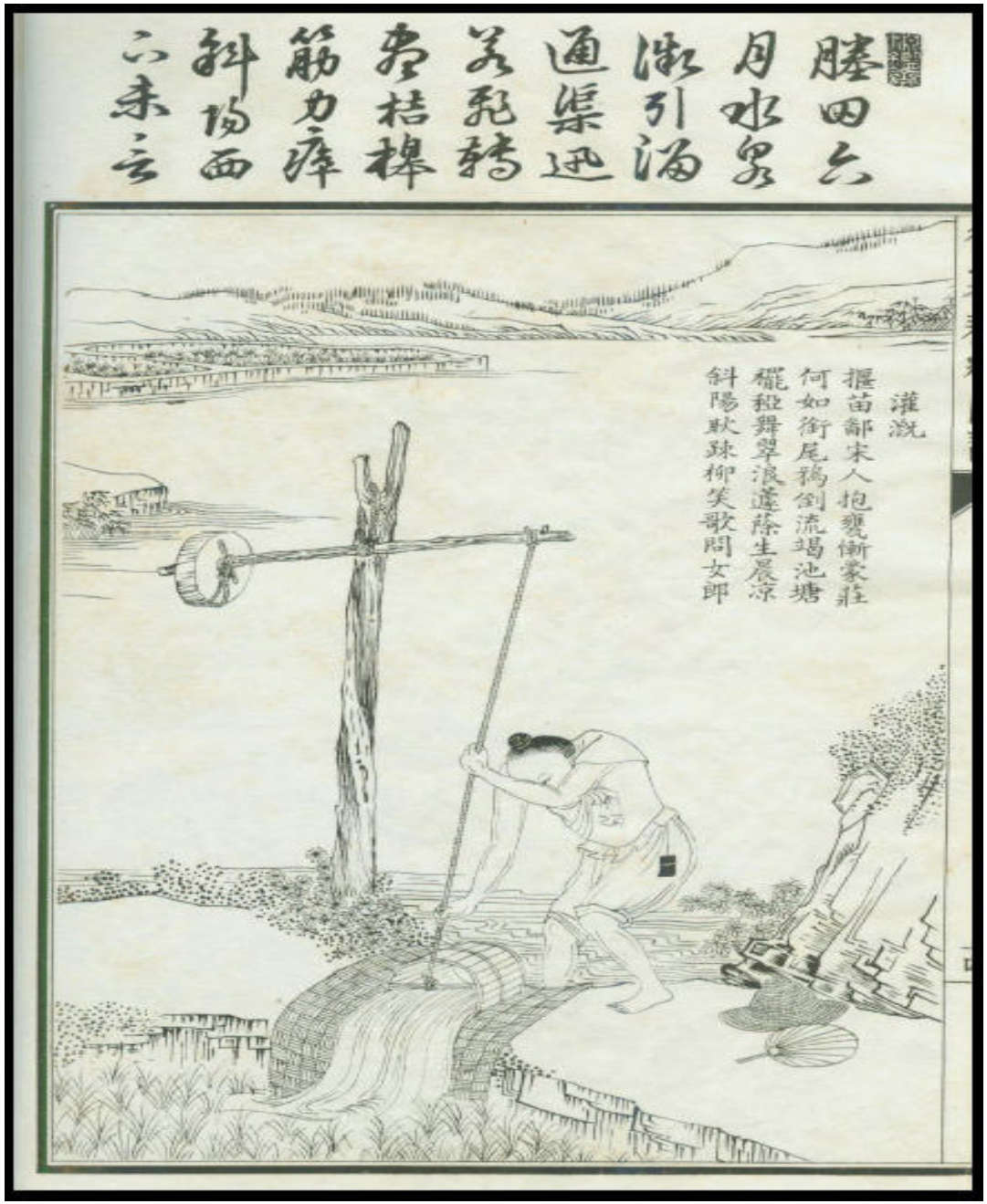
A kind of agricultural instrument: pestle and mortar

***Pictures and Poems for Plowing and Plaiting*, published in 1827**

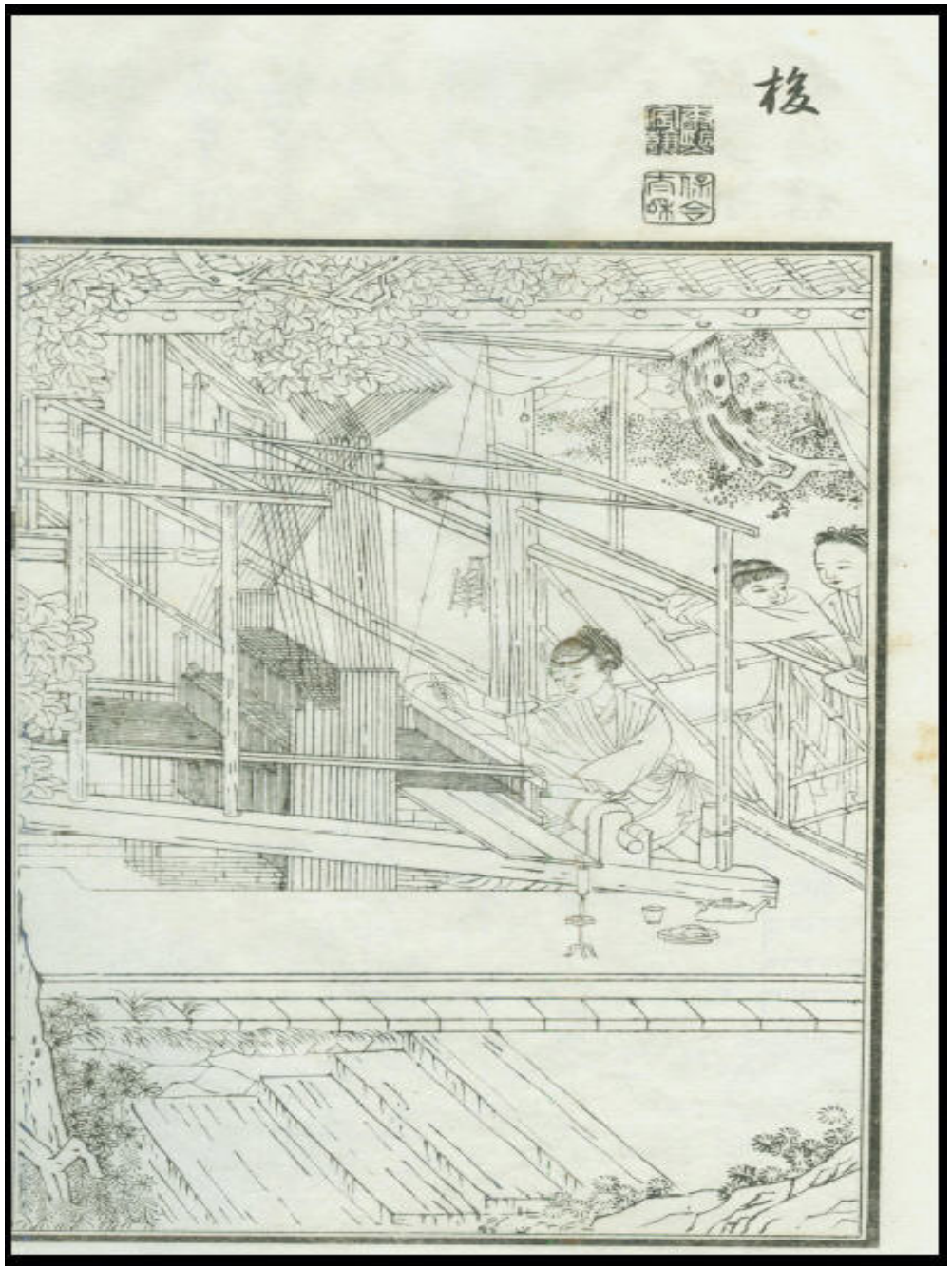


TSINGHUA UNIVERSITY LIBRARY

An irrigating instrument



A weave machine

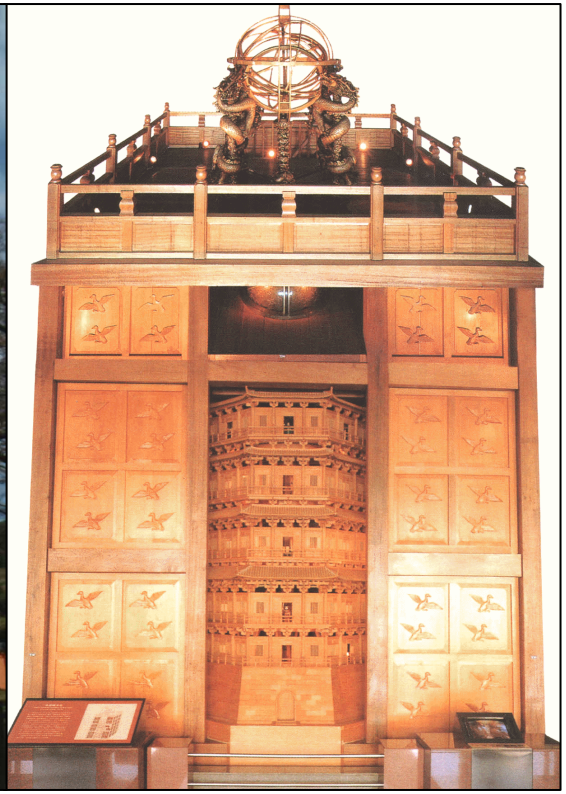


TSINGHUA UNIVERSITY LIBRARY

Reconstructive Models of the Astronomical Clock Tower

The earliest model in the National Museum of China, 1958

Model in Japan



Model in Taiwan

Objective in near future: to build a "Digital Library of Chinese Science & Technology History" in which more and more academic disciplines can be gradually involved.

Tsinghua University Teaching Resourse DL

2,000 courses information of China and other country universities.

10,000 E-reference books

http://www.lib.tsinghua.edu.cn/textbook/index.html

清华大学图书馆教参信息系统

清华大学图书馆教参信息系统

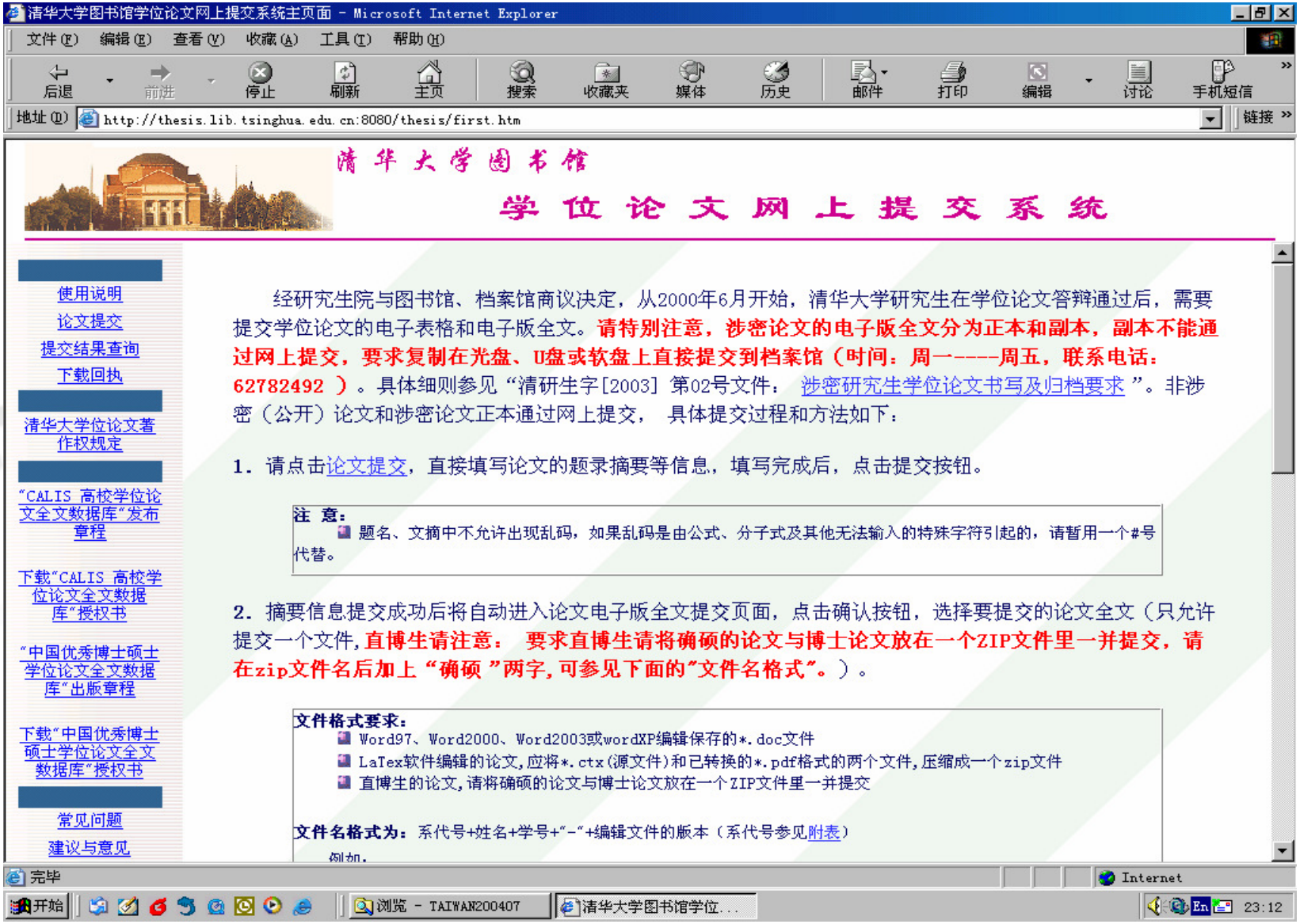
电子教参 E-Reserves

电子图书分类浏览

自动化技术、计算机技术
无线电电子学、电信技术
机械、仪表工业
能源与动力工程
环境科学、安全科学
航空、航天
水利工程
交通运输
建筑科学
金属学与金属工艺
电工技术
原子能技术
石油、天然气工业
化学工业

数理科学和化学
自然科学总论
生物科学
语言、文字
文学
一般工业技术
经济
文化、科学、教育、体育
政治、法律
艺术
军事
历史、地理
医药、卫生
综合性图书

TH University Dissertation Databases



清華大學圖書館

清华大学图书馆学位论文网上提交系统主页面 - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 前进 停止 刷新 主页 搜索 收藏夹 媒体 历史 邮件 打印 编辑 讨论 手机短信

地址(D) http://thesis.lib.tsinghua.edu.cn:8080/thesis/first.htm 链接

清 华 大 学 图 书 馆

学 位 论 文 网 上 提 交 系 统

使用说明
论文提交
提交结果查询
下载回执

清华大学位论文著作权规定

"CALIS 高校学位论文全文数据库"发布章程

下载"CALIS 高校学位论文全文数据库"授权书

"中国优秀博士硕士学位论文全文数据库"出版章程

下载"中国优秀博士硕士学位论文全文数据库"授权书

常见问题
建议与意见

学位论文提交表单（*为任选项，其余为必选项）

研究生姓名： (姓名间不准空格)	
学号： (学号不足10位,在学号后面加0补足10位)	
密码(至少6位)：	(供查询修改时用，务请牢记!!!)
请再次输入密码：	
培养单位：	
专业：	
联系电话：	
Email*：	
论文保密级别：	公开 内部 秘密 机密
论文答辩日期：	- -
学位级别：	
论文题名(中文)：	
论文题名(英文)：	
第一导师姓名：	
第一导师单位：	

完毕 Internet

开始 浏览 - TAIWAN200407 清华大学图书馆学位... En 23:13

欢迎访问学位论文检索系统 - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹
地址(D) http://thesis.lib.tsinghua.edu.cn:8001/xwlw/index.jsp
转到 链接
清華大學 学位论文服务系统
Tsinghua University Library
首 页 基本检索 高级检索 帮 助 FAQ
欢迎访问学位论文服务系统!
您是IP访问用户，无需再输入账号和密码!
本地认证
帐号: tsinghua2
密码:
登录 取消
摘要 查看学科树
与 作者
与 论文题名
从 年 到 年
查 询
说明：1. 本数据库包括清华大学1980年以来的所有公开的学位论文文摘索引。
2. 本数据库收录2003年以来的论文全文，回溯论文加工正在进行中，几乎每月都有论文全文更新。
3. 本数据库默认为IP登录，如果您的IP在许可范围内，则可以直接检索。
4. 如果您使用学科检索，请点击检索框旁边的"查看学科树"，将对应的学科名输入检索框。
5. 您在阅览全文时，需下载插件，请参见"下载说明"。
6. 关于本数据库的使用方法，请参见"使用帮助"。
6. 本数据库检索结果的排序方式默认为"相关度排序"。
Internet
开始
文章(2)
清华大学图书...
清华大学图书...
清华大学信息...
欢迎访问学位...
清华大学校园...
2:10

Tsinghua Special Resources DL

Contents in the project:

-- Tsinghua Memory

Campus Culture, Class events, Organization Origin etc.

-- Tsinghua famous professors

Manuscripts, Photos, Monograph, Letters etc.

-- Tsinghua Alumni directory

--Tsinghua published journals

-- Tsinghua oral history

……..

The Project is under construction. (celebration of 100 anniversary)

China-US Million Book Digital Library Project

- Leader: Zhe Jiang University
- Participate members: 16 Academic Libraries including Tsinghua
- Chinese Government invests 70,000,000 RMB（2003-2005）, US NSF offers the equipments and software valued around 3000,000$
- Achievements: (Up to Mar. 2007)

  No. of Digitized books:1,225,654

  No. of web published book: 861,591

  Data storage space:50 TB

  16 digitization centers established

China-US Million Book Digital Library Project

Access Policy:

- Around 140,000 Ancient e-books (published before 1911) accessed worldwide
- Around 180,000 Ming Guo e-books (published between 1911-1949) accessed among “211” Universities
- Modern Books and Desertation will be opened among participating members after getting property permittion.

Average per day hits: 150,000,

Average per day downloads:120 GB, (Equal with 4000 e-books)

Tsinghua provides 50,000 Ebooks and Dissertation for CADAL

高等学校中英文图书数字化国际合作计划 - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
中文(中国)
后退 搜索 收藏夹
地址(D) http://www.cadal.zju.edu.cn/Index.action
转到 链接
高等学校中英文图书数字化国际合作计划
CHINA-US MILLION BOOK DIGITAL LIBRARY PROJECT
CADAL
首 页 | 特色服务 | 使用帮助 | 关于CADAL | 登 录
快速检索 | 高级检索 | 图像检索 | 视频检索 | 书法字检索
检 索
分类浏览
古籍 民国图书 民国期刊 现代图书 学位论文 绘画 视频 英文
“登录 / 注册”让本站为您提供更好的服务。
项目其他网址：CADAL管理中心 | CALIS
浙江大学 | DCD实验室 | CADAL开发团队 | 帮助
Copyright 2005 China-US Million Book Digital Library Project
完毕
Internet
开始
文章(2)
清华大学图书馆概...
清华大学图书馆介...
5 Internet Expl...
清华大学校园网用...
2:30

Resource Digitization Center

Facility:

- Three flatbed AVA3 AVISION scanners
- Three FB6000E AVISION flatbed scanner
- Minolta PS 7000
- High speed AVISION AV3800

Staff:

- 1 manager, 1 technical supervisor, 10 temp. staff

Current Capacity: 3,000,000 page/year

Budget: One million RMB

Digitization Process

- Preparation (Selection， Identifier assignment)
- Scanning
- Image processing
- Metadata creation and packaging
- Quality control
- Data storage and backup

Related Software

Scanning: QuickScan…

Image processing: Bookshop, ACDSee, XnView, UltraEdit, Scanfix, DjVuerPro,…

Cataloging and Packaging: CADAL Cataloging Tool, OEBEditor, CMDL Cataloging Toolkit,…

Data transferring: DResManages

Practice Experience

1. We walked three steps in these years:

Step 1: Simple digitization and web presentation
"Journal of Tsinghua Weekly"
Step 2: Research on some key theories, technologies and cooperative model
"Tsinghua Architecture Digital Library"
Step 3: Pay more attention to interoperability, preservation and National and International cooperation
" EMANI "
" Millionbook Project"

Practice Experience

2. Emphasis on the resource collecting and organizing

   The coverage of subject oriented collection is limited and need to completed cooperated with other potential partners

3. Create a test bed for pursuing research areas, such as platform, storage formats, image processing and organization.

4. Open the service interface to google scholar

5. Infrastructure like DOI system and metadata knowledge base need to be prepared for future extensibility.

6. Easy batch output with OAI and METS encoded format

Problems

1. Digital format duplicates
2. Government funding or company support
3. Collecting mechanism need to be study (enlarge scale)
4. Possibility to cooperate with database provider (sustainability)
5. Technical problem such as Ancient character OCR and other organization and service technology.
6. Copyright issues. (Fair use area)
7. Integration of different projects in different platforms.

Thank you for your attention!

Other collaborative efforts are expected!

International cooperation

- **Subscription Collaboration:** Cooperated with scientific database providers for organizing subscription consortia and seting up mirror sites
- **Academic exchange:** First China Digital Library international Conference held in Tsinghua cooperated with Dr. C.C. Chen (USA)
- **Project cooperation:**EMANI project with partners from Germany, USA, France

- **Other collaborative Efforts are expected**