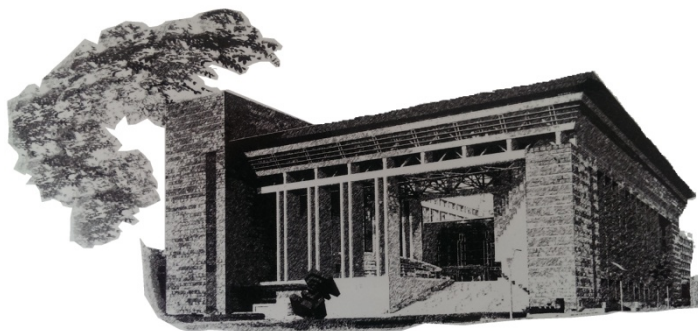




面向语义信息环境的嵌入式图书馆服务



孙 坦

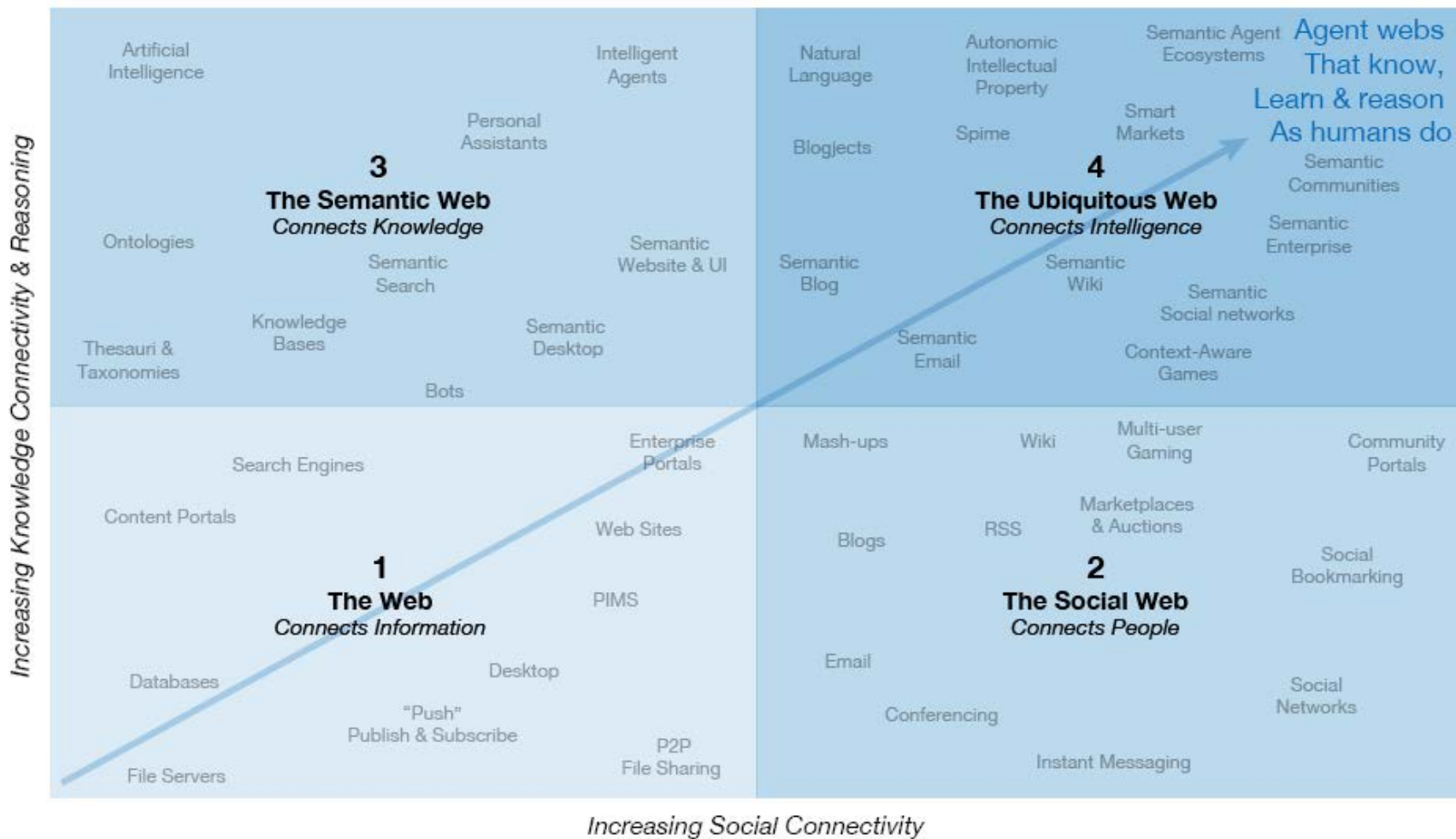
2015年8月10日 Monday 中国·舟山





学术图书馆20年间的变化

2020年网络发展4阶段



于容芬





1

基于系统的服务变化

2

基于馆员的服务变化

3

再造数字信息基础设施





- 1st generation DL: 基于资源的数字图书馆

Resource-based digital library: = digital collections?

- 2nd generation DL: 基于集成服务的数字图书馆

Integrated digital resources and related services

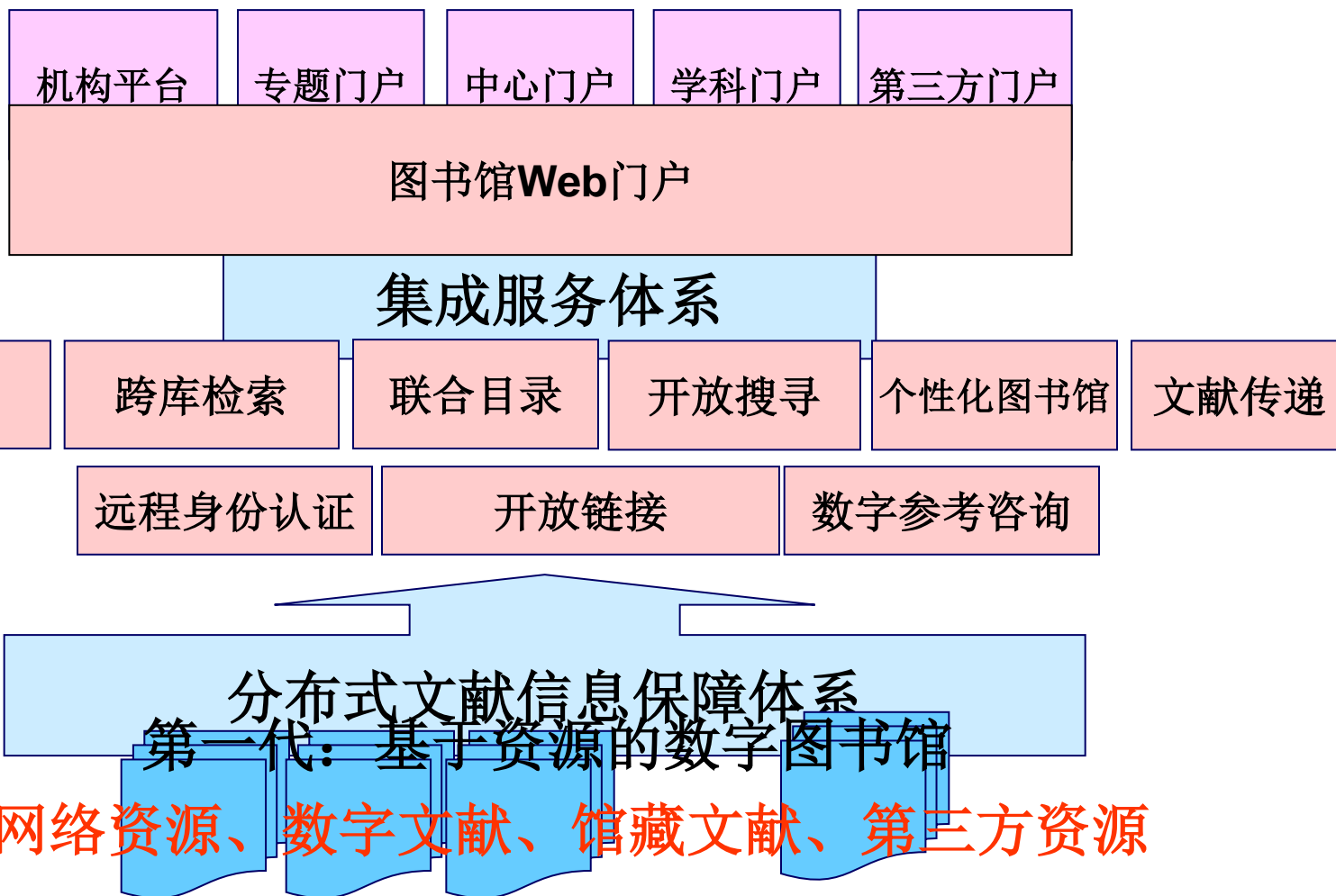
- 3rd generation DL: 基于用户的数字图书馆

Digital resources and services personally organized and embedded into user application environment

数字图书馆范式演变



第二代：基于集成服务的数字图书馆





Bioscience Researchers

- **Print is dead, really dead**
- **The library seen as e-journal provider**
- **Faculty don't come to the physical library**
- **Students come to the physical library**
- **Article databases greatly underused**
- **Buy books from Amazon**



Moving to the Network (Global) Level

- **What can/should we do at the global/network level that we no longer need to or should do at the local level?**

Lorcan Dempsey, OCLC




Local to Global: WorldCat.org

① User "Finds in a Library"



Search for an item in libraries near you:
Enter title, subject or person



Signed into WorldCat as lacrossegrri (Sign Out) My WorldCat Account

UNIVERSITY OF WASHINGTON Libraries

Your Library Account CUSTOM LINK CUSTOM LINK Feedback

Search for Items: Search

[Advanced Search](#)

② Locates Services in WorldCat.org

Find in a Library with WorldCat

Search for Items in Libraries: marathon running

Ultra running with Scott Jurek by Jim Whiting

Language: English Type: Book Juvenile audience Internet Resource
Publisher: Hockessin, Del. : Mitchell Lane, 2007
ISBN: 1584164845 9781584164845 OCLC: 64876077 [Cite This Item](#)
Subjects: Jurek, Scott | Runners (Sports) -- Juveniles -- Biography, I Meation
[Table of contents only](#)

Libraries Details Subjects Reviews

1. [Anchorage School](#) \$19.95

Clicks to a library service





UW View of WorldCat.org

You are **not signed in WorldCat** ([Sign In](#) or [Register](#))

[My WorldCat Account](#)



[Your Library Account](#) [Feedback](#)

Search for items:

Search

[Advanced Search](#)

Search results for **'guns, germs and steel'**

Sort by: **Relevance**

Refine Your Search

Author

[Jared M Diamond](#) (17)

[Jared Diamond](#) (7)

[Diamond Jared](#) (4)

[Steve Olson](#) (2)

[Block Walter](#) (1)

[Show more ...](#)

Content

[Sociology](#) (16)

[Anthropology](#) (1)

[Biological Sciences](#) (1)

[Business & Economics](#) (1)

[History & Auxiliary ...](#) (1)

Format

[Article](#) (35)

[Book](#) (23)

[Visual Material](#) (6)

[Sound Recording](#) (3)

Results **1-10** of about **65** (.20 seconds)

<< First < Prev 1 2 3 Next >

1. **Guns, germs, and steel : the fates of human societies**

by Jared M Diamond

Language: English Type: Book

Publisher: New York : W.W. Norton & Co., ©1998.

Held by: University of Washington Libraries

2. **Guns, Germs, and Steel: The Fates of Human Societies**

by Diamond; R M Meyer

In: JAMA : the journal of the American Medical Association. 284, no. 22, (2000): 2933

Language: English Type: Article

Publisher: Chicago : American Medical Association, 1960-

Held by: University of Washington Libraries

3. **Guns, germs, and steel**

by Jared M Diamond; Tim Lambert; Cassian Harrison; Peter Coyote; Lion Television Ltd.; National Geographic Television & Film.; National Geographic Society (U.S.)

Language: English Type: Visual Material : Videorecording : DVD video

Publisher: [United States] : National Geographic, [2005]

Held by: University of Washington Libraries



Local to Group to Global

Architecture in the digital age : design and manufacturing

by Branko Kolarevic

Language: English Type:  Book

Publisher: New York : Taylor & Francis, 2005.

Held by: University of Washington Libraries

UW First

The new feminine brain : how women can develop their inner strengths, genius, and intuition

by Mona Lisa Schulz

Language: English Type:  Book  Internet Resource

Publisher: New York : Free Press, ©2005.

Held by: Summit

Then Group

Defining intuition : a framework for understanding intuition in psychology

by Julie M Sprengle

Language: English Type:  Book : Thesis/dissertation/manuscript  Archival Material

Publisher: 2005.

Held by: WorldCat Libraries

Then Rest of World

- 20% increase in use of local collections
- 40% increase in interlibrary loan requests
- 50% increase in borrowing within geographic region



基于集成服务数字图书馆的问题

■ 几个



□ 请



□ 我



□ 我



□ 微生物科学数据与科技文献的天联



□ 空天研究对大装置信息的需求



□ 理论物理研究所和计算所对预印本和会议信息的需求



□ 越来越多对人的查找需求

第三代数字图书馆

基于用户的数字图书馆



VIVO - Microsoft Internet Explorer

文件(E) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 前进 刷新 地址(D) http://vivo.cornell.edu/ 转到 链接 >>

Google G 开始 收藏夹 已拦截 180 个 拼写检查 翻译 发送至 设置

Cornell University
Library

SEARCH: go

VIVO Cornell

VIVO life sciences

Home People Education & Training Research Tools Facilities

VIVO Index About Contact Us

VIVO - A Service of Cornell University Library

Use VIVO to discover who's working on a particular research topic; what they've taught or published recently; where facilities might be and what online tools are available to expedite research.

Searching VIVO: use the shortest term you can (e.g., "rice" instead of "rice research", or the last name only for people searches); see [search tips](#) for more help.

Search examples:

<input type="text" value="nanofab*"/>	<input type="button" value="try it!"/>	(an asterisk matches variations at the beginning or end of words)
<input type="text" value="'climate change'"/>	<input type="button" value="try it!"/>	(a leading quote indicates you want to search for a phrase, not individual words)
<input type="text" value="crystallography"/>	<input type="button" value="try it!"/>	

VIVO is an ongoing project of the **Life Sciences Working Group** in the Cornell libraries, and coverage is not yet complete. At present we are adding content from the College of Engineering and Weill Cornell Medical College. Please [contact us](#) if you or your unit should be included in the site, or if you have any corrections, comments, or suggestions for improvements. You may also speak with Medha Devare at 255-2199.

完毕

开始 Microsoft PowerP... 综合科技信息资... 综合科技信息资... VIVO - Microsoft I... 10:49



Institut AIFB - People - Microsoft Internet Explorer

文件(E) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

地址(D) http://www.aifb.uni-karlsruhe.de/Personen/english

Google 开始 书签 已拦截 180 个 拼写检查 翻译 发送 设置

People

Find person

with research-area

with project

with research-group

[Members of Institute](#) / [Scientific Staff](#) / [PhD Students and Scholarship Holders](#) / [Visiting Lecturers](#) / [Admin. Staff](#) / [Guests](#) / [Alumni](#) / [Student Workers](#) / [Org. Map](#)

Agarwal	Sudhir	Tel.: +49 (721) 608 7946	Office: 246	
Allerding	Florian	Tel.:	Office:	
Bartsch	Michael	Tel.:	Office:	
Betz	Stefanie	Tel.: +49 (721) 608 4553	Office: 1A-14 (Geb. 05.20)	
Bloehdorn	Stephan	Tel.: +49 (721) 608 6558	Office: 252	
Blohm	Sebastian	Tel.: +49 (721) 608 7363	Office: 251	
Bonn	Matthias	Tel.: +49 (721) 608 6034	Office: S121	
Branke	Jürgen	Tel.: +49 (721) 608 6585	Office: 223	
Brockmans	Saartje	Tel.: +49 (721) 608 4754	Office: 258	
Bulander	Rebecca	Tel.: +49 (721) 608 6588	Office: 233	
Bungert	Peter	Tel.: +49 (721) 608 3924	Office: 250	
Campelo	Eulálio	Tel.: +49 (721) 608 4283	Office: 233	
Che	Haiying	Tel.: +49 (721) 608 4063	Office: 242	
Cimiano	Philipp	Tel.: +49 (721) 608 3705	Office: 222	

任务栏: 开始 | 会议门户 | Microsoft PowerP... | 综合科技信息资... | 综合科技信息资... | Institut AIFB - Pe... | 10:41



Welcome to SEKT — SEKT Portal - Microsoft Internet Explorer

文件(E) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

地址(D) <http://www.sekt-project.com/>

Google 开始 收藏夹 媒体 搜索 已拦截 180 个 拼写检查 翻译 发送至 设置

site map accessibility contact

project partners resources r&d events news sekt project research & development resou

log in

you are here: home

navigation

- Home
- Events
- News
- SEKT Project
- Research & Development
- Resources
- Members
- What's new?
- Demo Videos
- News

Welcome to SEKT

by [sekt](#) — last modified 2006-12-13 04:27 PM

The EU IST integrated project **Semantic Knowledge Technologies (SEKT)** developed and exploited semantic knowledge technologies. Core to the SEKT project has been the creation of synergies by combining the three core research areas **ontology management, machine learning and natural language processing.**

- SEKT Info:
 - Co-funded by the EU [6th Framework](#) programme.
 - Start date:** 1.1.2004, Duration: 36 months.
 - Project Nr:** EU IST IP 2003-506826
 - Coordinator:** John Davies, British Telecommunications Plc., UK
- SEKT Downloads:
 - SEKT In a Nutshell:** [SEKT Flyer \(PDF\)](#) & [Brochure \(PDF\)](#)
 - SEKT Components:** [Components developed in the SEKT project](#) and the [SEKT component demos](#)
 - SEKT Deliverables:** [List of deliverables](#), PDF versions of public deliverables
 - SEKT Demo:** [The SEKT concept demonstrator](#) has been created to show the business potential of SEKT technology. The scenario is set in the financial services sector, but the ideas can be easily extrapolated to a range of sectors and application

news

开始 会议门户 Microsoft Po... 综合科技信... 综合科技信... PowerPoint 幻灯... Welcome to ... 10:55



Who's who in the environment - 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 刷新 地址(D) http://www.swed.org.uk/swed/servlet/Entry?action=v 转到 链接 >>

Google 开始 书签 已拦截 180 个 拼写检查 翻译 发送 设置

swed semantic web environmental directory

go

all items within current results [help?](#)

Home Administration

Recently visited:

RDF Powered

Browse

Organisations and projects are categorised using six characteristics or facets which are shown below. The numbers in the brackets show how many results are in that facet. To start browsing pick an option from one of the facets. Click **+** to see full lists of options for each facet.

+ Topic of interest [What is this facet?]
[Animal Welfare](#) (13) | [Archaeology](#) (2) | [Arts and Crafts](#) (9) | [Biodiversity](#) (5) | [Biotechnology and Genetic Engineering](#) (1) | [Built Environment](#) (19) | [Business and Commerce](#) (12) [More options](#)

+ Organisation type [What is this facet?]
[Advisory](#) (5) | [Campaign](#) (1) | [Commercial](#) (9) | [Development \(World\)](#) (2) | [Educational](#) (9) | [International](#) (2) | [Membership Organisation](#) (2) [More options](#)

+ Activity [What is this facet?]
[Animal Welfare](#) (7) | [Awards Accreditation](#) (6) | [Campaigning and Lobbying](#) (74) | [Conservation Cultural Heritage](#) (2) | [Cultural Activities](#) (8) | [Education and Training](#) (79) | [Employment](#) (2) [More options](#)

+ Project type [What is this facet?]
[Advisory Project](#) (1) | [Campaigning Project](#) (1) | [Conservation Project \(Heritage\)](#) (1) | [Conservation Project \(Wildlife\)](#) (1) | [Economic/Social Development Project](#) (1) | [Educational Project](#) (3) | [Pilot Project](#) (1) [More options](#)

Operational area [What is this facet?]
[International Areas](#) (5) | [Non UK Counties](#) (6) | [United Kingdom](#) (81) | [Worldwide](#) (26)

+ Name [What is this facet?]
[A*](#) (5) | [B*](#) (9) | [C*](#) (15) | [D*](#) (2) | [E*](#) (4) | [F*](#) (7) | [G*](#) (3) [More options](#)

开始 Microsoft Po... 综合科技信... 综合科技信... Who's who in... 10:57



SCIENCE
INFRASTRUCTURE
GRID
NETWORK
COLLABORATION
SCIENTIFIC COMMUNITIES



- Home
- About D4Science
- Activities
- Project organisation
- Scientific communities
- News and events
- Resources
- Documentation

gCube:
the middleware that powers D4Science

- Login
- Print this page
- E-mail this page
- Contact us
- RSS Feed

Home

BELIEF DL Submission Form

Prepare a public document | Prepare a private document

Public Document Submission

Title

.....

Authors - Lastname, Firstname

.....

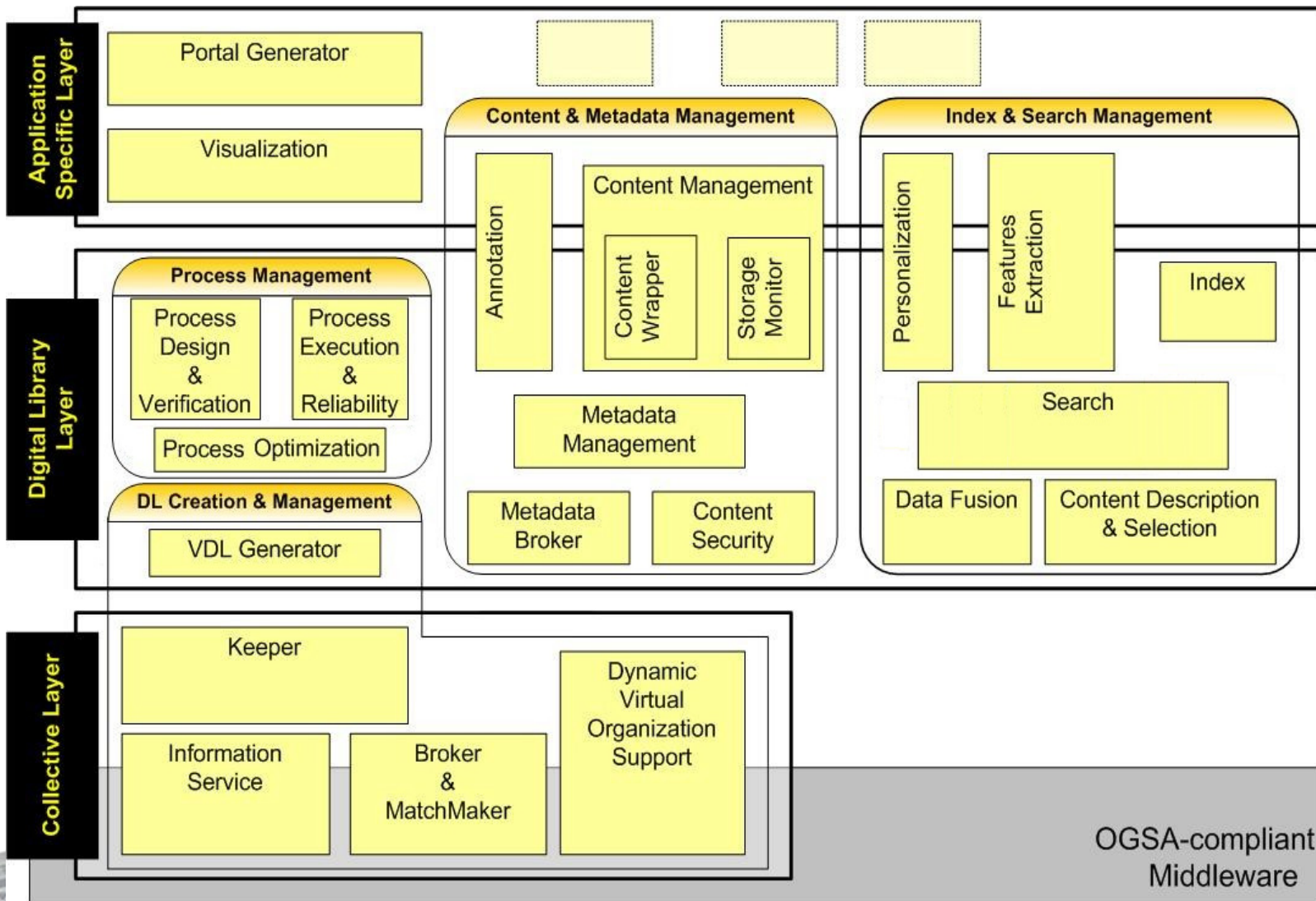
Abstract

Date of publication - YYYYMMDD

.....

Subject

基于用户的数字图书馆





数字图书馆的概念本质

- **Not about books! Never was!**
- **Also not about information!**
- **But it IS about people – connecting people to resources of all kinds needed for learning and critical thinking and (hopefully) knowledge.**
- **No longer just “organizing” knowledge containers, but helping to create them and mold them, host them, etc.**



Global Research Library

- **Being Global**
- **Global Infrastructure**
- **Scholarly Communication and Peer Review**
- **Data, data, and more data**
- **Global Regulatory and Policy framework**



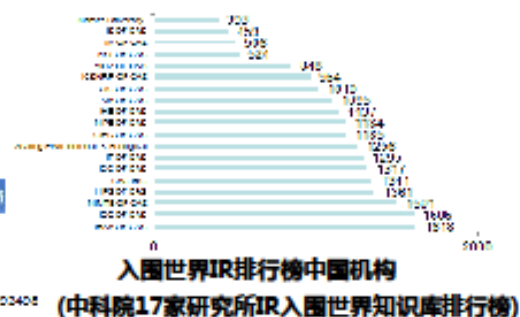
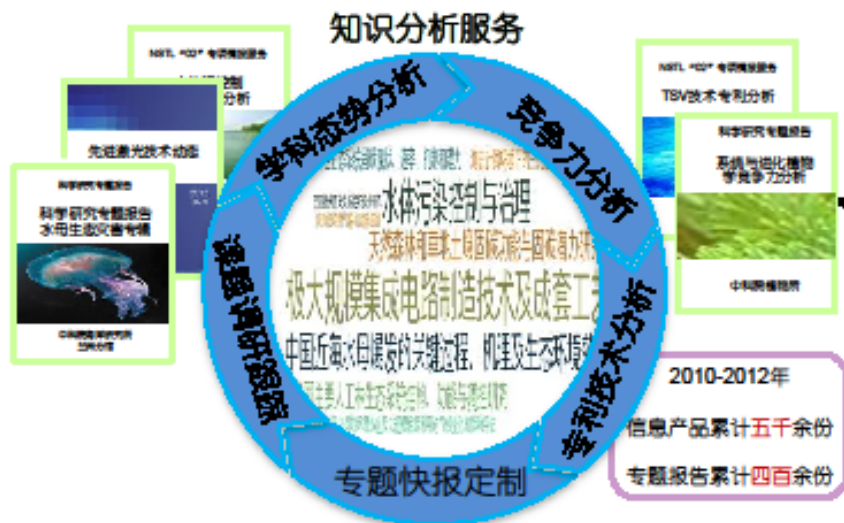


2

基于馆员的服务变化

——从嵌入式服务谈起





群组知识环境 136

植物科学领域科学家网络
 国家自然科学基金项目数据库

科技创新服务





嵌入式图书馆服务的缘起

1978年Dickinson提出了经典学科馆员定义

1. 专业参考帮助
2. 院系联络

1. 需求驱动、个性化嵌入式成为服务目标
2. 与用户关系从合作协同走向伙伴式关系

1950-1990年代
初始期

1990-2000年代
变革发展期

2000-2006年代
成熟创新期

2007年以来
转型发展期

1. 建立对话机制
2. 构筑协作关系
3. 加强服务宣传
4. 创新服务能力

2007年David Shumaker 和Mary Talley: 有关嵌入式学科服务发展、管理以及实践最初调查
《嵌入式图书馆员模式》 Models of Embedded Librarianship

嵌入式学科服务是用于描述一个研究型图书馆员持续地参与研究型课程并且在课程中教授信息素养技能。这也包括馆员需要将工作场地从以图书馆为中心转移到用户身边，便于更好沟通与合作





2012年1月，英国大学图书馆委员会Mary Auckland 《重新确立支撑科研所需的技能》（Re-skilling for Research）对当前和今后2-5年学科馆员所需要的潜在技能和知识进行了研究，提出32个方面的技能和知识

专业学科	1) 具有用户的学科/专业的深度知识；2) 具有与学科/专业相关的内容方面（在各种相关载体）良好的知识；3) 具有在该学科/专业领域的书目和其他查找工具的良好知识；4) 具有就本地和其他地方相关档案和特藏提供咨询的知识
研究过程	5) 了解本单位当前和变化的研究兴趣；6) 了解有代表性的科研人员的经历，包括其工作流，科研人员在一个学科/专业和科研人员的职业生涯的不通过阶段怎样获取和利用信息；7) 具有获取个人/项目需求的能力，包括有效的倾听技能；8) 拥有研究资助信息源的知识，帮助科研人员找到潜在的资助者
沟通合作	9) 具有与科研人员和校内其他专业人员建立良好合作关系的技能，并与外部建立协同关系，管理好客户关系；10) 意识并能够由国家和国际合作计划和机构提供服务价值和机会；11) 能够有效地参与研究项目，包括确认图书馆在项目中的角色，辅助投标和报告撰写；12) 能够能动地向科研人员提供咨询和宣传推广图书馆的服务。
信息	13) 娴熟的信息发现、文献检索等技能；14) 具有综合、分析和提供所发现信息摘要的能力；15) 具备提供科研人员信息管理咨询的知识，包括可携带性，特别是参考文献管理工具（如Endnote）；16) 具备对科研人员信息的处理与展示进行咨询的能力；17) 具备对引文和参考文献提供咨询的知识，并能使用文献管理软件。
研究数据	18) 具有良好的该学科/专业数据源的知识；19) 具备提供有关数据管理和监管咨询的知识，包括萃取、发现、获取、传递、保存和可携带；20) 具备对该学科/专业所使用的潜在数据处理工具提供咨询的知识；21) 具备对数据挖掘提供咨询的知识。
信息素质	22) 具备良好的设计信息素质教育的技能（面对面和在线），以满足不同类型科研人员的不同需要。



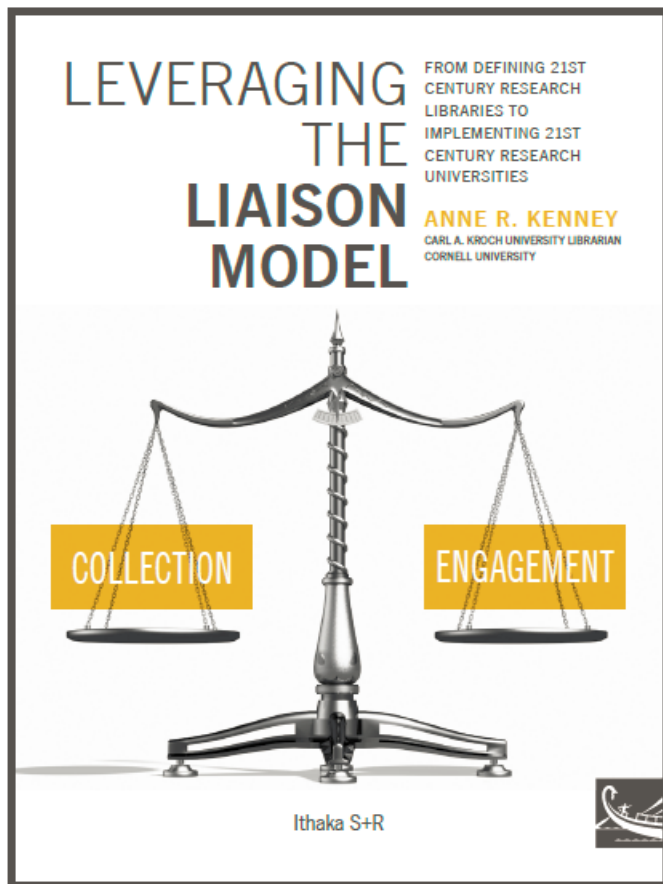


学术交流	23) 能提供有关国内外研究成果出版与传播方法和模式的当前趋势、最佳实践和可能的选择等方面的咨询, 包括学术交流和开放获取出版; 24) 能提供有关研究成果保存方面的咨询; 25) 能提供有关项目记录(如信函)保存方面的咨询。
资助机构要求评估	26) 具有足够的支持科研人员遵守资助机构各种要求的知识, 包括开放获取的要求; 27) 了解国家和本单位科研评价过程, 以及可再生能源基金会(REF)的要求; 28) 了解科研影响因子和绩效指标以及怎样使用, 并能够提供引文分析、文献计量学等方面咨询; 29) 了解作者权利、版权法和知识产权问题以及剽窃, 必要时提供咨询或指导。
元数据	30) 具备元数据的宣传、咨询、使用方面的知识; 31) 能够制订元数据架构, 并为每个科研项目提供学科/专业标准和实践的咨询。
新兴和Web2.0技术	32) 能够提供有关移动技术、web2.0和其他通信工具(如开源免费软件Mendeley、虚拟研究环境)的价值和利用方面的咨询。

当前和在今后的2-5年将继续并越来越重要的9大方面的技能和知识: 1) 在该学科/专业具有良好的书目和其他发现工具的知识; 2) 具有良好的设计信息素质教育的技能(面对面和在线), 以满足不同类型科研人员的不同需要; 3) 具有良好的信息发现、文献检索等方面的技能; 4) 具有对引用和参考文献提供咨询的知识, 并能利用文献管理软件; 5) 能够能动地向科研人员提供咨询并宣传推广图书馆的服务; 6) 具有良好的该学科/专业数据源的知识; 7) 具有良好的有关该学科/专业内容的知识; 8) 了解本单位当前和变化的科研兴趣; 9) 能够获得每个科研人员/项目的需求, 包括有效的倾听技能。



协同服务思考



Ithaka S+R研究报告：发挥学科馆员服务模式效能

Anne R. Kenney著

当研究型图书馆树立从“馆藏资源为核心”到“用户参与为核心”的领导地位时，图书馆需要思考如何衡量这种“参与”为教员、学生和其他群体带来的积极影响，这种思考会很有意义和益处。发挥学科馆员模式的影响将对图书馆展示新的形象至关重要，图书馆需要展示出自己并不仅仅是一个资源内容的供应商，图书馆所有专业知识和专业技能会应该是延展至校园内外的整个学术知识体系的重要组成部分。学科馆员将继续拓展他们所在机构和所在院系的成功程度和成功的疆界。上述的转变将让我们不再是去定义一个21世纪的“优秀图书馆”，而是去参与和投身于建设21世纪“优秀的研究型大学”。





从关注图书馆员的工作到关注学者的工作，并基于学者的需求和成果指标制定图书馆参与的战略，这是学科馆员服务模式发展过程中新出现的转变。学者 **Jaguszewski** 和 **Williams** 最近出版了一本关于学科馆员角色转变的专著，他们认为，“嵌入式的学科馆员服务，旨在提高科研人员的学术生产力、提高学习人员的综合能力，并参与和融入到研究教学和学习的整个生命周期。”在2013年 **ARL** 举办的主题为“科研事业总动员”（**Mobilizing the Research Enterprise**）的秋季研讨会上，**Karla Hahn** 介绍了俄亥俄州立大学对于研究全过程的支持活动，包括制定了覆盖广泛的“利益相关者地图”（**Stakeholder Map**）。

过去十年间，图书馆学科馆员发展模式从传统上的全职参与馆藏建设和参考服务，转型到为教师和学生提供增值和组合的延伸服务。《研究图书馆议题》（**Research Library Issues**）于2009年8月发表了学科馆员的新角色相关案例研究报告，**Karla Hahn** 在报告引言中提出，“构建新的合作关系，尤其是构建与教职员工的合作，对有效发挥学科馆员服务的功能至关重要。”**Karen Williams** 从岗位职责描述上，阐明了学科馆员从以馆藏为中心到以参与为中心的巨大变化。尼苏达州大学图书馆对学科馆员主要的岗位描述框架进行了调整，包括了学术交流、在线学习、数字工具、延伸服务、资金申请等方面。新的服务样式也随之产生，包括：数据管理、学术简历、数字科研、研究 workflow、学术出版新形式、创造性呈现、公共倡导、数据驱动学术、影响力评估、语义网开发、联邦资助授权和托管、全球化合作及研究态势分析等等。

协同服务思考

学科馆员模式发展的第三个主题是日益注重推广有助于教师和学生参与的工具和模板。

- 研究型图书馆通过可用的数字功能，如机构知识库、引文管理软件，以及最近通过提供诸如数据管理计划工具（**DMPTool**）这样的系统，帮助研究人员建立数据管理的计划。
- 最近，大家把的重点关注在了整合信息碎片，特别是自动获取及处理（例如通过**Symplectic Elements**工具从各大学和出版商的数据库获取数据）
- 充分利用跨平台的程序和工作流（例如在机构及学科知识库存储或提交年度报告以及完善教师信息系统）
- 在用户空间嵌入工具/模板，保证更高的利用率（例如在美国国立卫生研究院的**PubMed**中心，突显**SPARC**作者附录或提供一个说明信模板作为手稿提交的一部分，或通过课程管理系统上传课程**Libguides**）
- 通过链接数据、持久数字标识符、以及语义网的发展丰富整个研究系统。





协同服务思考

- 主仆关系
- 合作关系
- 伙伴关系

协同服务
参与式设计

新场景
新定位
新角色

- 购买服务
- ↓
- 协同创新

服务创新
创新系统构建

用
数

- 文献保障
- ↓
- 数据工具

Participatory Design in Academic Libraries

New Reports and Findings

Nancy Fried Foster, Editor
February 2014



COUNCIL ON LIBRARY AND INFORMATION RESOURCES





3

再造数字信息基础设施



再造数字信息基础设施的动力与愿景

➤ 吉姆·格雷 (Jim Gray) 认为互联网能把所有科学数据与文献联系在一起，创建一个数据和文献能够交互操作的世界，以提高科学的“信息速率”，促进研究人员的科学生产力。

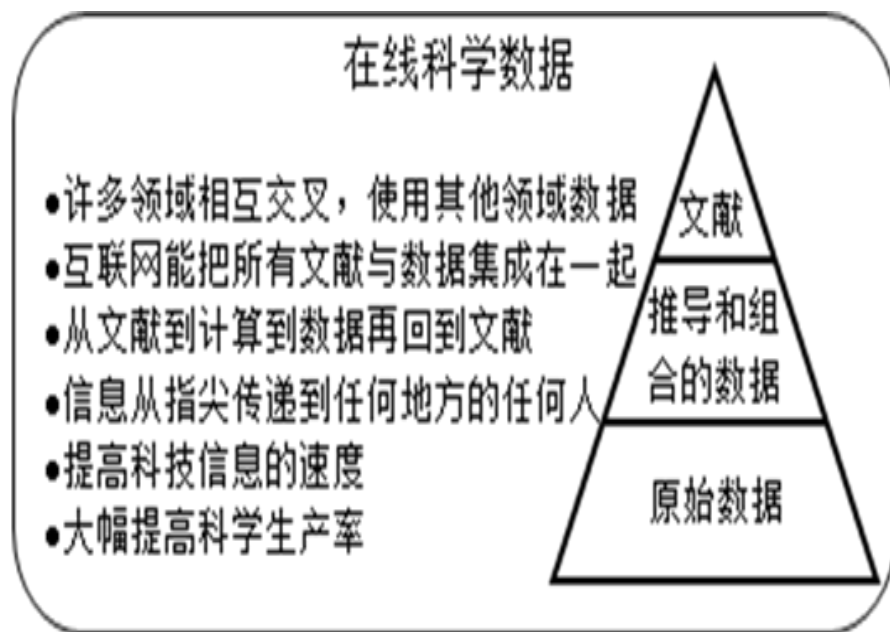


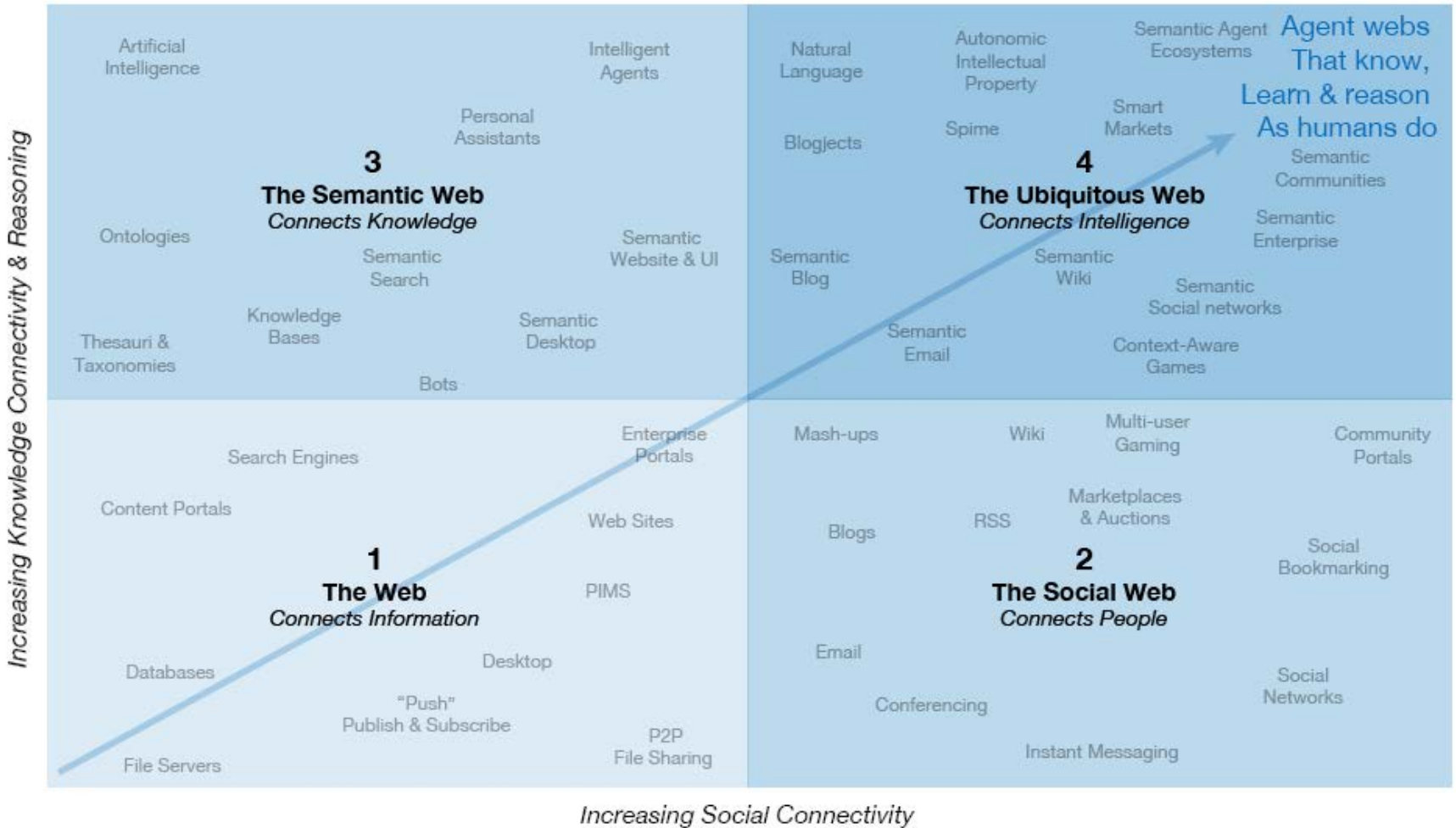
图1：科学数据与文献的交互操作

Tony Hey, Stewart Tansley, Kristin Tolle 著；潘教峰，张晓林等译
[M]. 第四范式：数据密集型科学发现. 科学出版社，2012



再造数字信息基础设施的动力与愿景

2020年网络发展4阶段



再造数字信息基础设施的动力与愿景

➤ 增强出版、语义出版和融汇型期刊的挑战

- 自然出版集团 (NPG) 在学术交流未来方向中明确提出了语义出版的模式，信息的发现是通过结构化的、交互化的、可查询的图表与文本实现，服务已更多地转为科学信息内容的分析、可视化与分发功能

- RSC期刊Molecular Biosystems通过增强HTML标记机制，对文本中被化学术语数据库收录的词汇和来自基因、序列和细胞知识本体的词汇进行高亮显示，并提供外部链接

- Elsevier启动了“Article of the Future project”，力图实现科学论文格式的彻底变革，并已经推出了SciVerse和科学工作流解决方案Reaxys Xcelerate

- Elsevier和Thomson Reuters为代表出版机构和数据库商致力于对自身海量信息的开发利用，通过关联、计量和挖掘直接面向用户提供科研评价、分析和战略咨询服务，对学术图书馆的传统文献服务带来巨大挑战



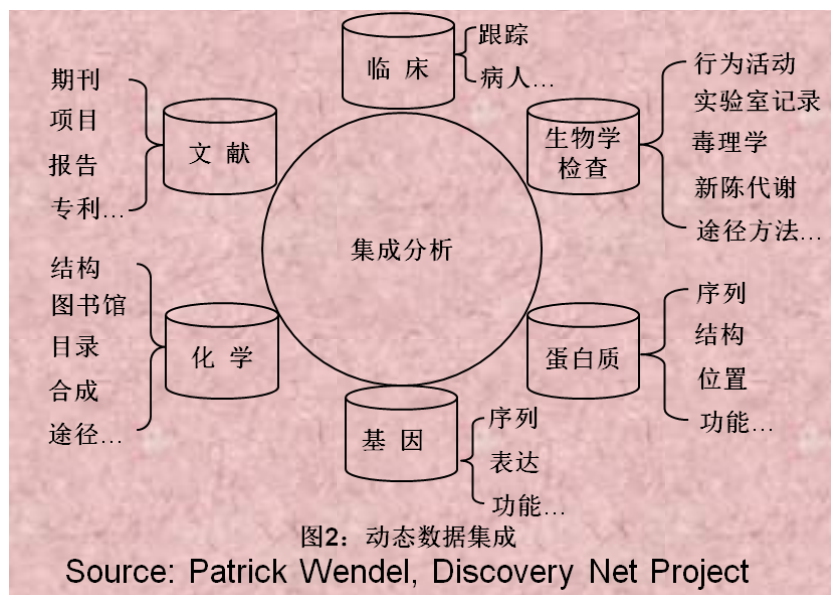
再造数字信息基础设施的动力与愿景

- 第三方关联检索、语义丰富与知识发现服务的挑战
 - 在W3C为代表的各类结构化、数据化、语义化标准的支持和语义网技术推动下，以知识关联、重组和语义互操作为特征的新型知识产品与知识服务首先集中呈现在生命科学领域，如PubMed Entrez、NxOpinion、NCIt、NCBO、VIVO等
 - VIVO通过一个广泛适用于科研领域的科研本体和生命科学领域本体支持对海量科研人员、仪器设备、科研活动和项目、科学数据、科研成果以及文献资源等进行结构化、语义化的知识表示与可视化展示，支持科研人员的关联发现
 - NCBO 基于NCI Thesaurus（参考术语本体）和连接到UMLS中的200多个本体，实现了对自由文本大规模语义标注，使非结构化的自由文本数据变得结构化、标准化。为研究人员提供在线工具及基于网络的资源（如BioPortal），使研究人员可以在生物医学研究及临床实践的各个方面实现对不同本体的访问、审查与集成，搜索与特定的本体术语相关的生物医学资源



再造数字信息基础设施的动力与愿景

- 开放信息环境中，新的科学研究范式和学术交流模式对信息的利用要求是：在e-Science基础设施（网格、工具和中间件等）支撑下，对开放综合、异构互联的各类信息对象进行基于知识单元和语义互操作的集成分析与计算
- Patrick Wendel 在 Discovery Net中提出了“信息网格：动态数据集成（InfoGrid: Dynamic Data Integration）”的信息利用模式，支持用户按需访问各类数据源和结构化信息，并创建自己的数据集进行集成分析



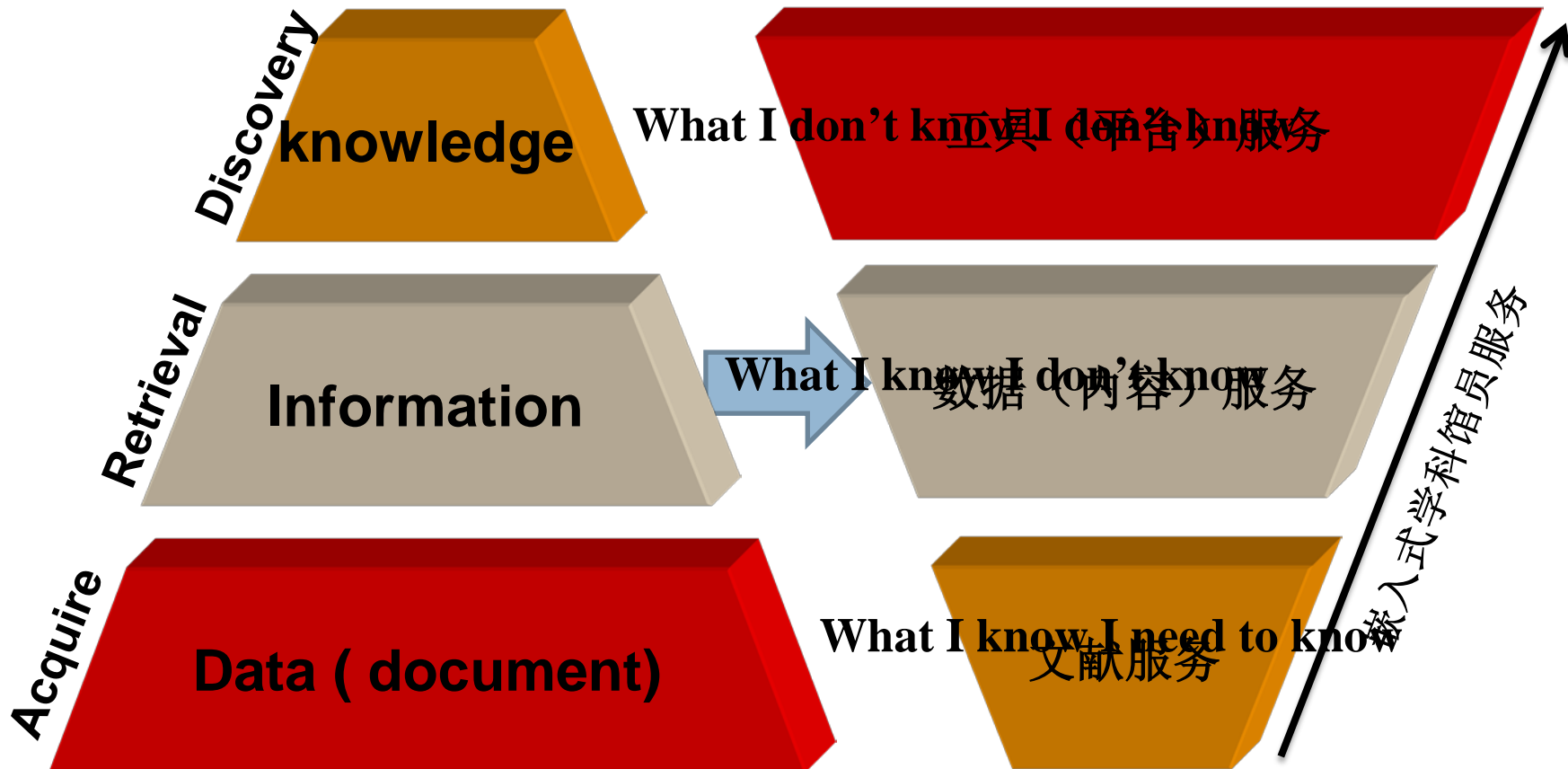


再造数字信息基础设施的动力与愿景

馆藏信息资源体系

嵌入/植入

用户知识环境



数字知识表示应用需求

开放信息环境中海量异构信息与碎片化需求的矛盾：

- 海量数据分散在各种书籍、报纸、不同格式的音视频媒体以及互联网和其他媒介中，并快速的数字化。这些数据的指数增长导致人们利用计算机查找专门数据和有用的相关信息的能力迅速降低。
- 搜索引擎的算法主要依赖关键词匹配，而不考虑关键词的各种不同含义以及位于短语中组合关键词的复杂含义，甚至是关键词和短语被用于不同上下文中的不同含义。
- 多数知识呈现出要么非结构化要么是异构化的，致使其不能被进一步处理，**根本性问题**：如何从海量数据源中有效提取需要的数据？如何找到那些代表着问题答案或有助于解决问题的必要和潜在可用的却又未知的信息片段？如何集成那些潜在的记录（信息片段）？

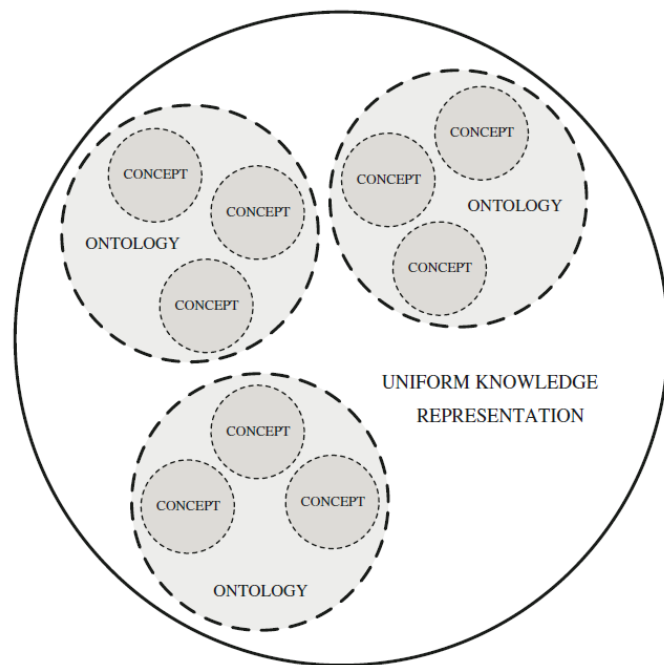
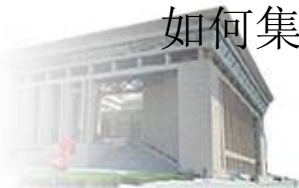
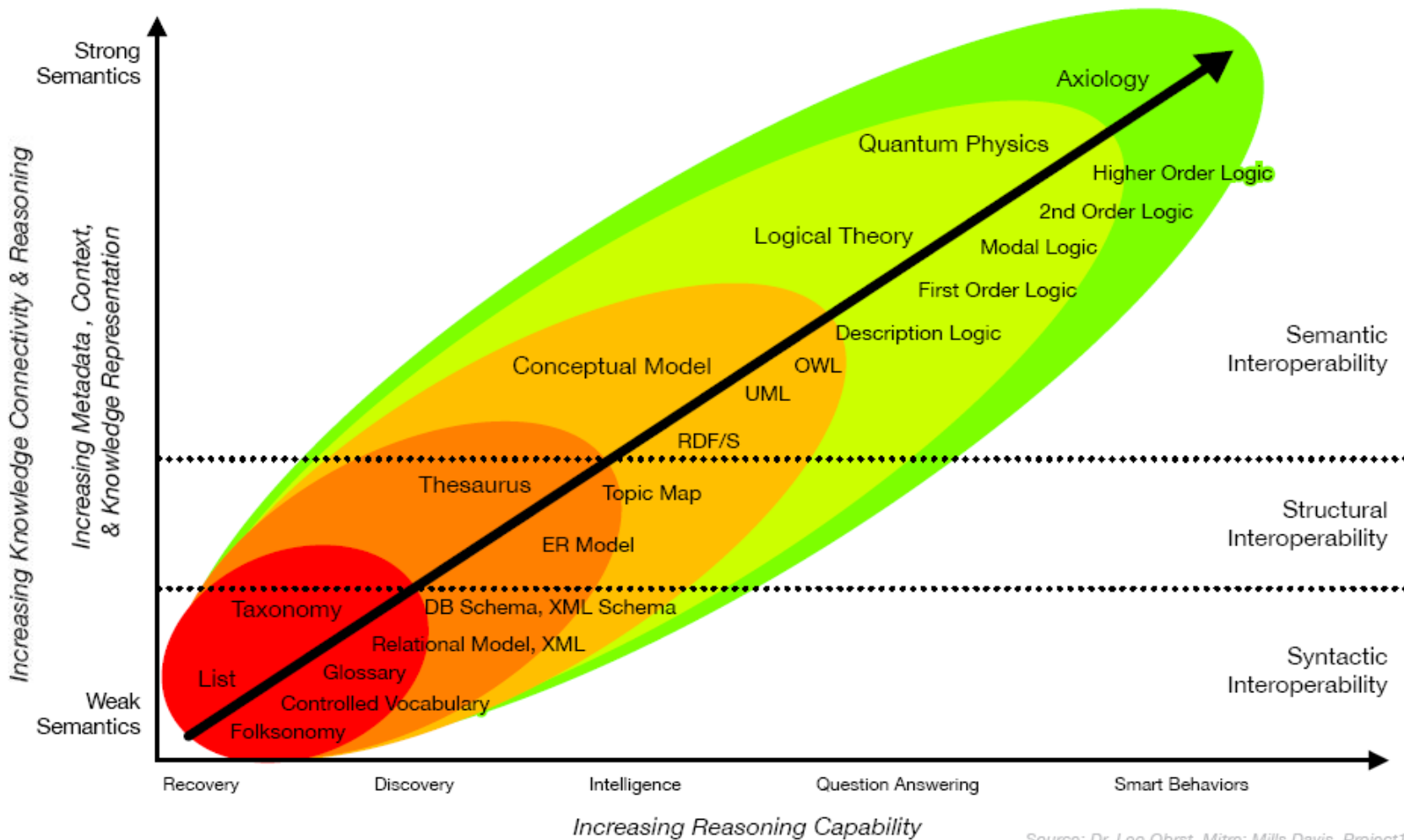


Fig. 1.1 Uniform knowledge representation consisting of ontologies populated by concepts



语义网络与知识表示技术

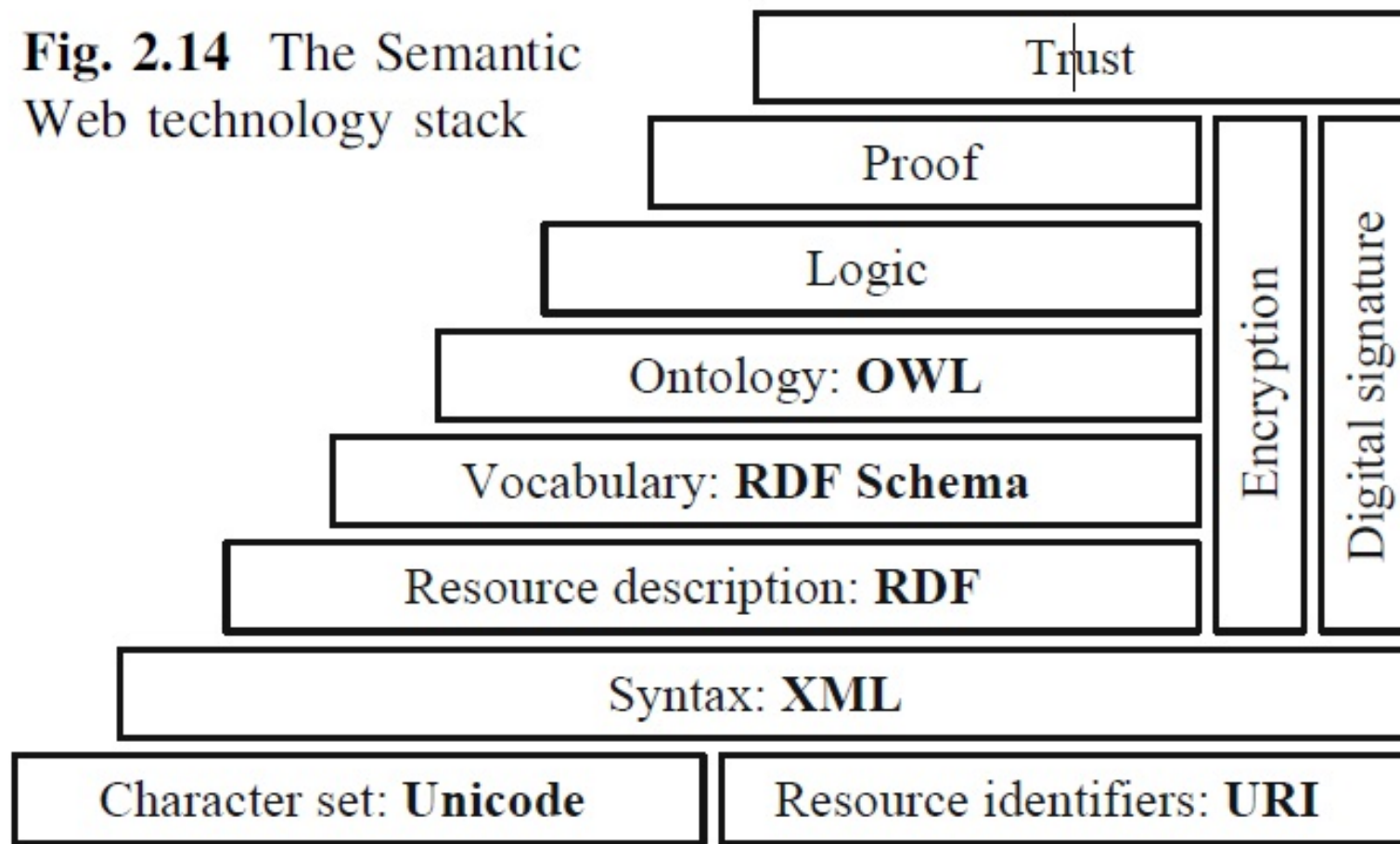
知识表示发展光谱
从简单列表到复杂推理



Source: Dr. Leo Obrst, Mitre; Mills Davis, Project10X

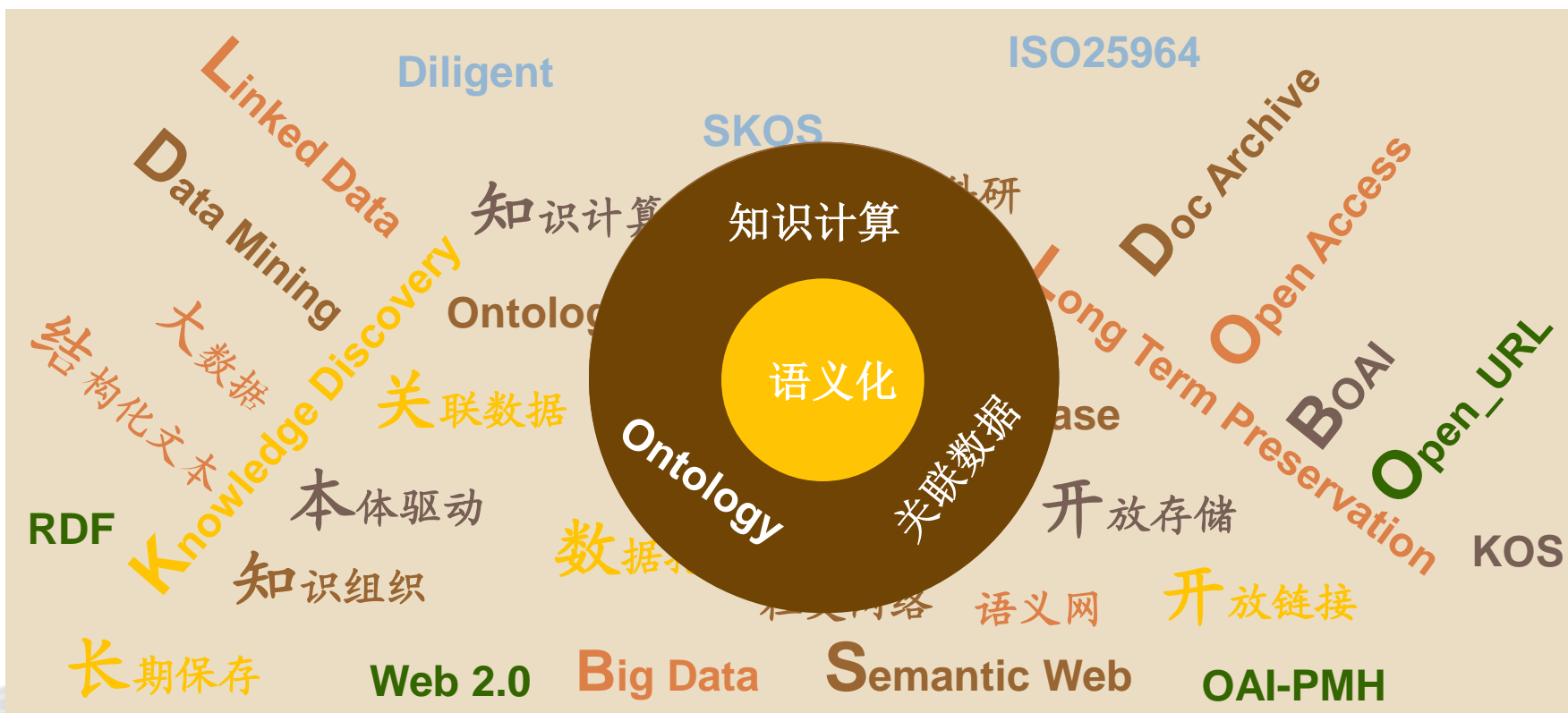
数字知识表示框架

Fig. 2.14 The Semantic Web technology stack



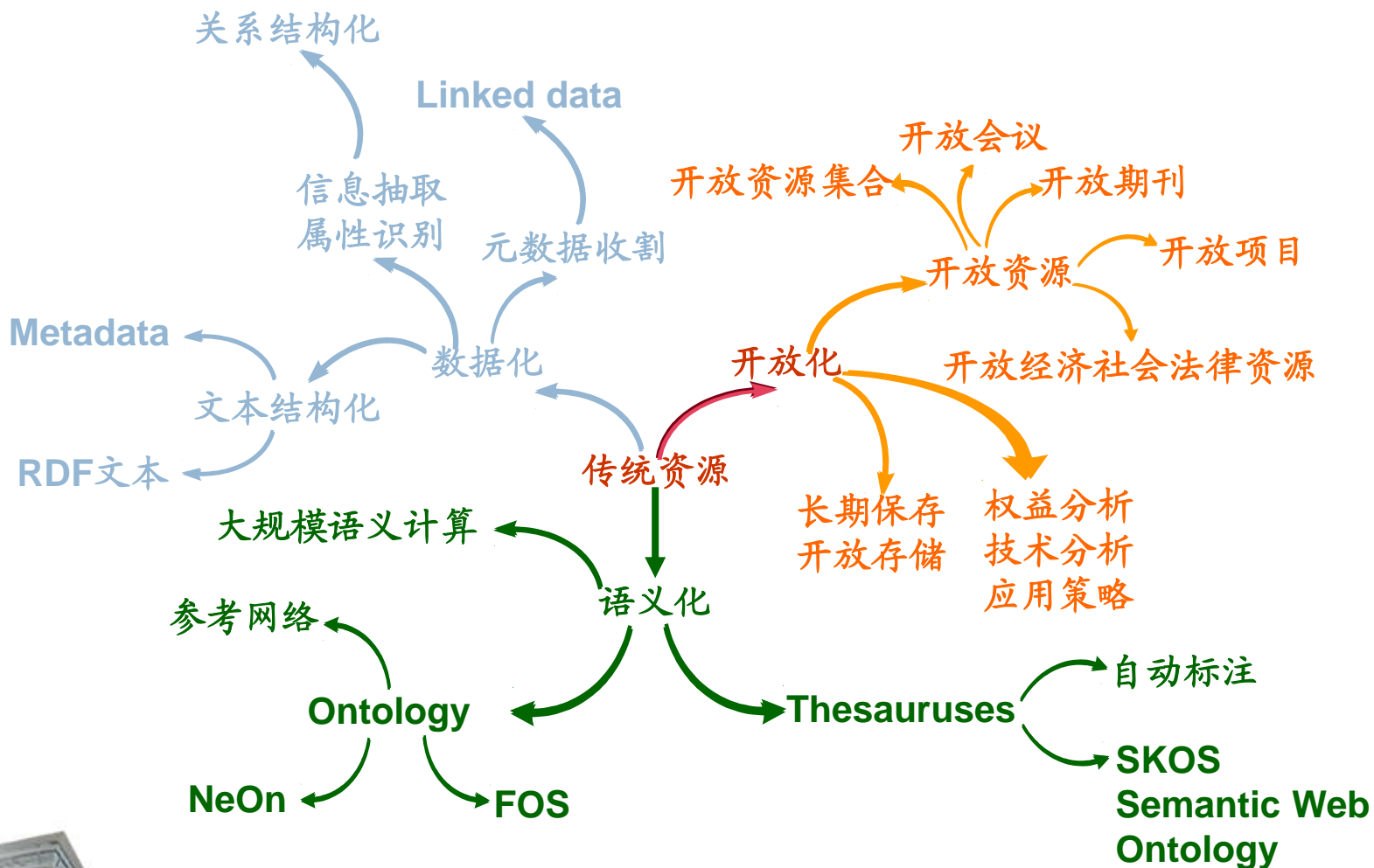
信息资源知识网络实现路径

知识表示与语义化路径包括：利用**开放关联数据**实现语义关联、利用**本体**实现开放信息语义化、利用**实体识别**实现开放信息语义化和利用**文本挖掘**实现开放信息语义化等。（引自钱庆）





信息资源知识网络实现路径



数字知识表示与语义网络应用

➤ 关联检索、语义丰富与知识发现服务层出不穷

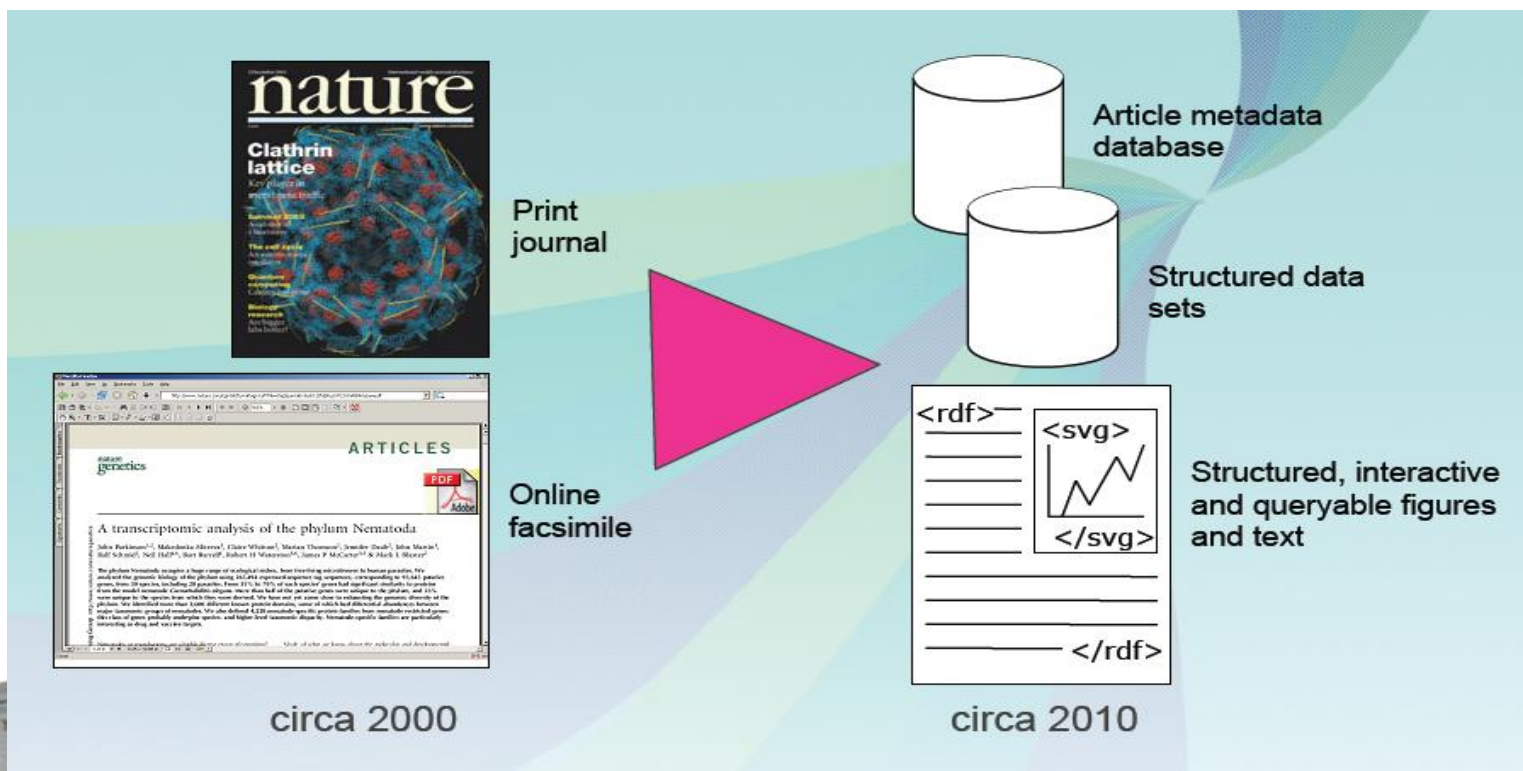
- 在W3C为代表的各类结构化、数据化、语义化标准的支持和语义网技术推动下，以知识关联、重组和语义互操作为特征的新型知识产品与知识服务首先集中呈现在生命科学领域，如PubMed Entrez、NxOpinion、NCIt、NCBO、VIVO等
- VIVO通过一个广泛适用于科研领域的科研本体和生命科学领域本体支持对海量科研人员、仪器设备、科研活动和项目、科学数据、科研成果以及文献资源等进行结构化、语义化的知识表示与可视化展示，支持科研人员的关联发现
- NCBO 基于NCI Thesaurus（参考术语本体）和连接到UMLS中的200多个本体，实现了对自由文本大规模语义标注，使非结构化的自由文本数据变得结构化、标准化。为研究人员提供在线工具及基于网络的资源（如BioPortal），使研究人员可以在生物医学研究及临床实践的各个方面实现对不同本体的访问、审查与集成，搜索与特定的本体术语相关的生物医学资源



数字知识表示应用案例分析

➤ Nature关联数据

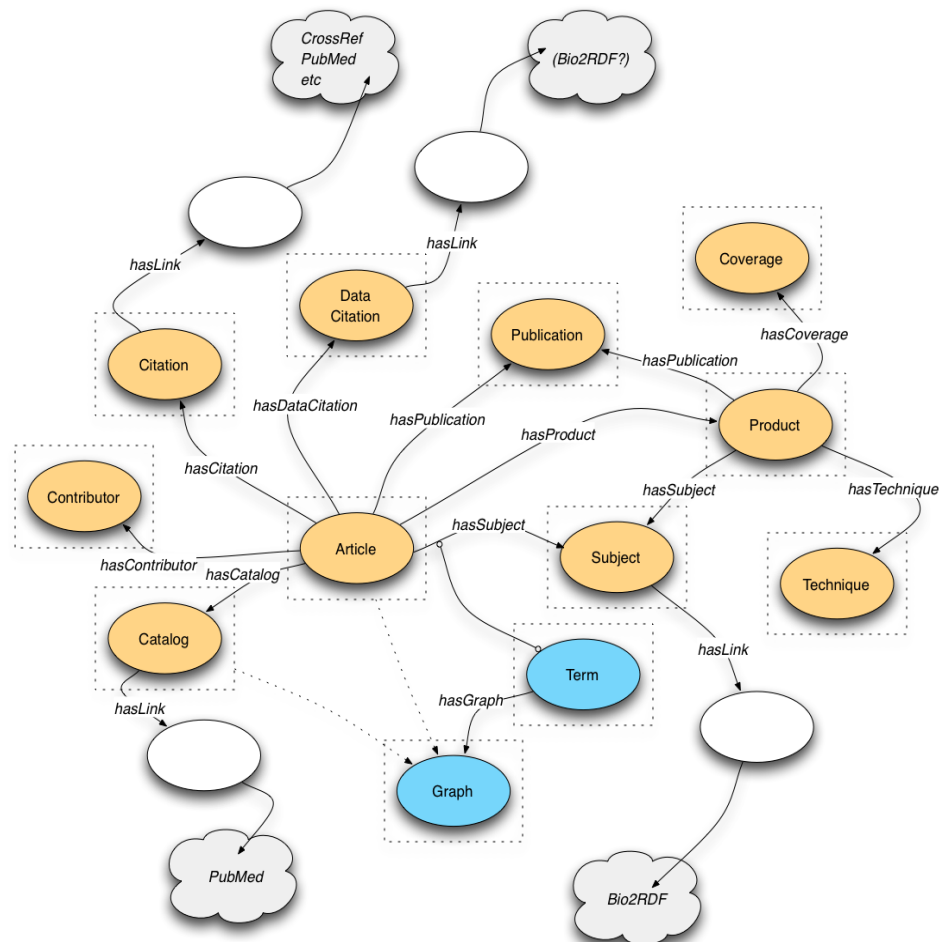
- 自然出版集团 (NPG) 明确提出了语义出版的模式, 信息的发现是通过结构化的、交互化的、可查询的图表与文本实现, 服务已更多地转为科学信息内容的分析、可视化与分发功能



开放关联数据 (Open Linked Data)

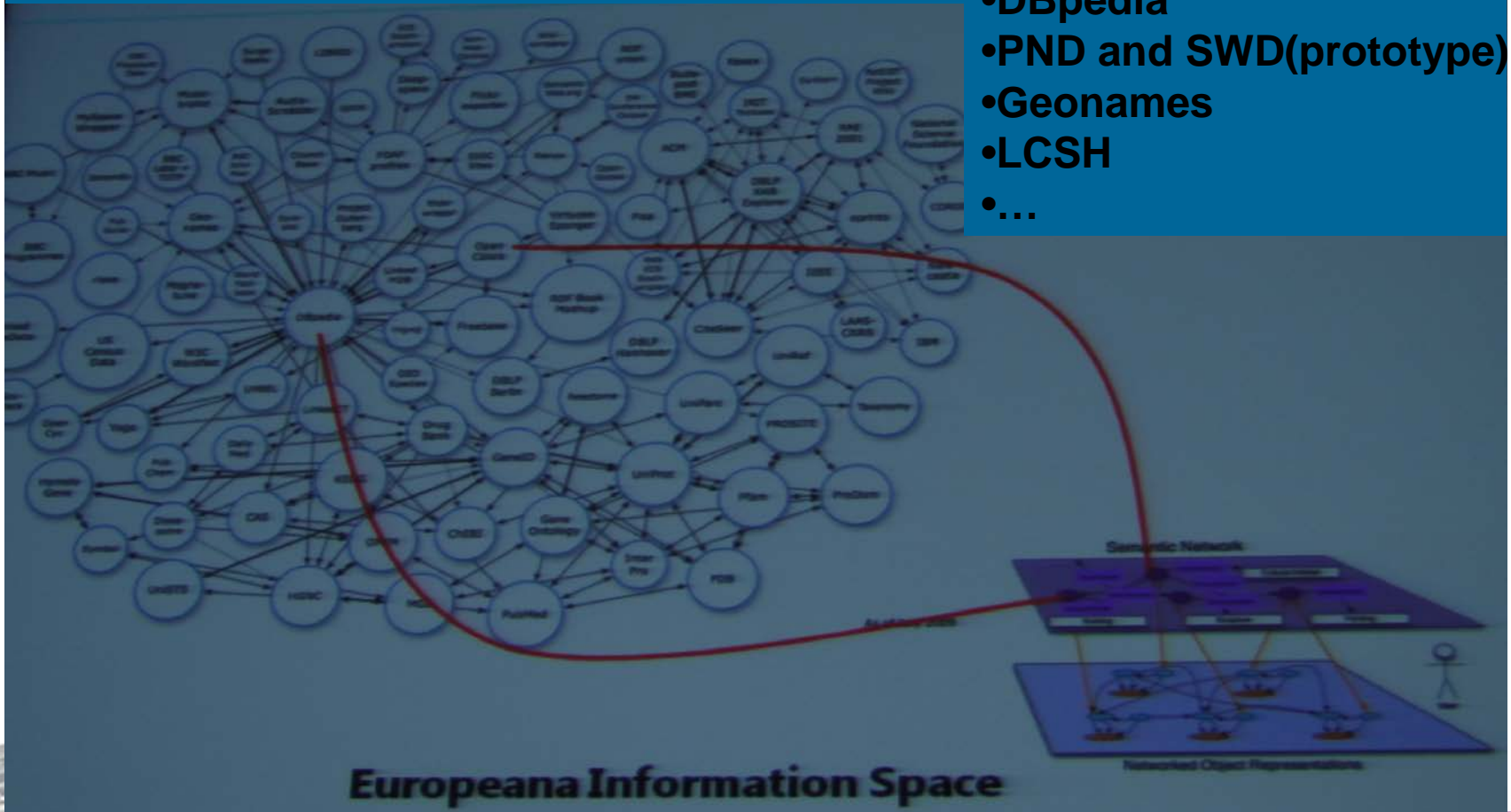
<http://dx.doi.org/10.1038/ndigest.2012.120202>

bibo:contributorList	_:b0
bibo:issue	2
bibo:pageEnd	2
bibo:pageStart	2
bibo:volume	9
dc:identifier	info:doi/10.1038/ndigest.2012.120202
dc:identifier	doi:10.1038/ndigest.2012.120202
dc:publisher	Nature Publishing Group
dc:title	Detectors home in on Higgs boson
npg:articl	ndigest.2012.120202
npg:doiHash	108bqvzywvd1t
npg:hasContributor	http://ns.nature.com/contributors/eugenie-samuel-reich-108bqvz
npg:hasProduct	http://ns.nature.com/products/ndigest
npg:hasPublication	urn:issn:1880-0556
npg:title	Detectors home in on Higgs boson
prism:aggregationType	issue
prism:copyright	Nature Publishing Group
prism:doi	10.1038/ndigest.2012.120202
prism:endingPage	2
prism:genre	News
prism:number	2
prism:publicationDate	2012-02-01
prism:startingPage	2
prism:url	http://dx.doi.org/10.1038/ndigest.2012.120202
prism:volume	9
rdf:type	http://ns.nature.com/terms/Article
rdf:type	http://purl.org/ontology/bibo/Article
rdfs:label	Detectors home in on Higgs boson



数字知识表示应用案例分析

EDM and Linked Open Data





数字知识表示应用案例分析

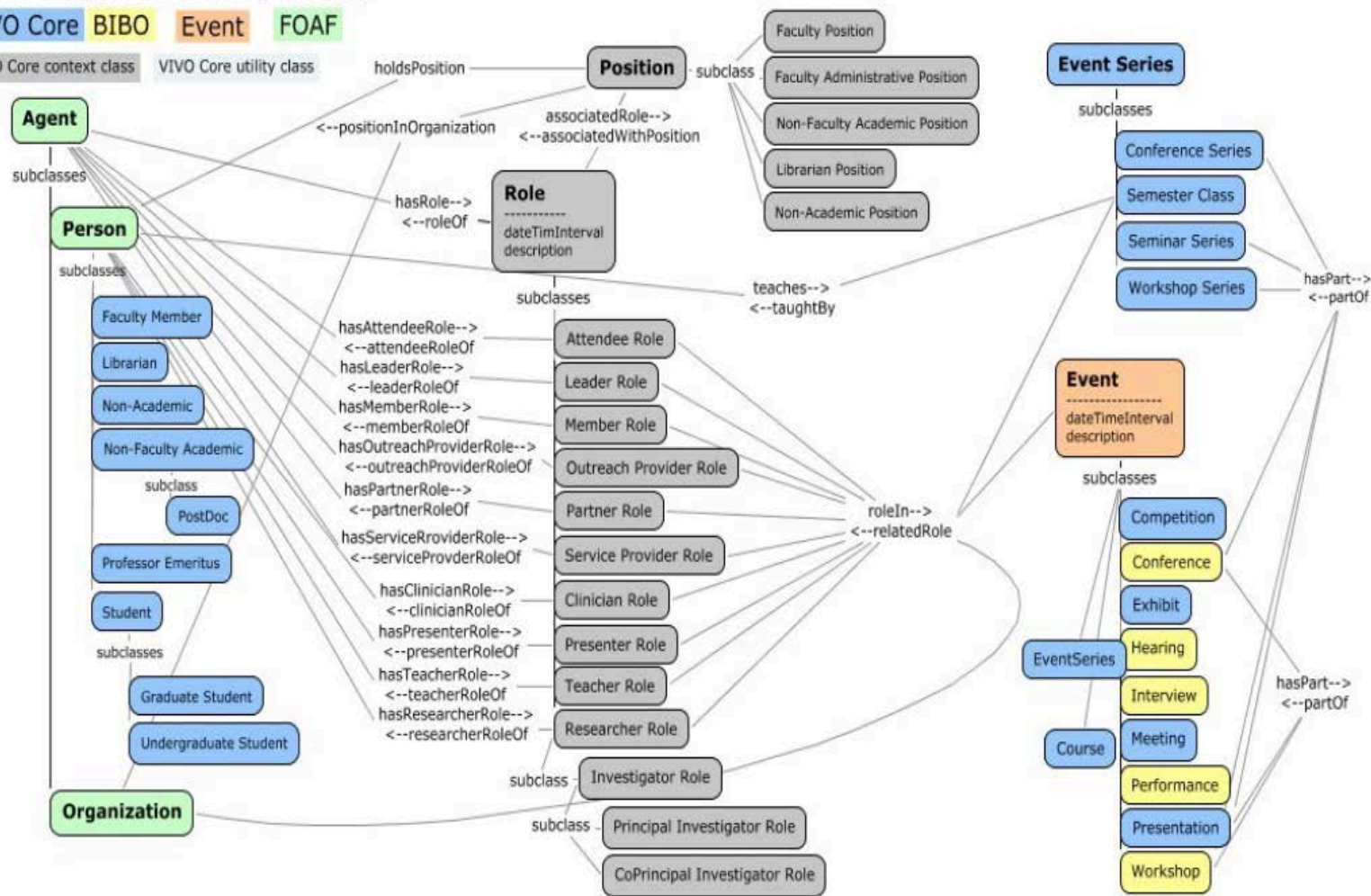
The screenshot shows the VIVO website interface. At the top, there is a red header with the VIVO logo and the text "Research & Expertise Across Cornell". To the right of the header are links for "Index" and "Log in", and a search bar with a "Search" button. Below the header is a navigation menu with links for "Home", "People", "Organizations", "Research", and "Events". The main content area is divided into several sections. On the left, there is a "Welcome to VIVO" section with a description of the service and a search bar. In the center, there is a vertical list of categories with their respective counts: "People (34,717)", "Activities (6,376)", "Courses (19,787)", "Events (29,905)", "Organizations (8,725)", "Equipment (101)", "Research (140,167)", "Topics (2,001)", and "Locations (1,063)". On the right, there is a list of categories under the heading "基于本体的数据结构" (Data Structure Based on Ontology). The categories listed are: "Cornell Academic Appointee", "Cornell Academic Staff", "Cornell Affiliated Person", "Cornell Emeritus Professor", "Cornell Faculty Member", "Faculty Member", "Faculty Member Emeritus", and "Graduate Student". At the bottom right, there is a green box with the text "事物关联，而不是字符串检索" (Thing Association, not String Search).

数字知识表示应用案例分析

VIVO 1.1 (Partial focusing on role)

VIVO Core BIBO Event FOAF

VIVO Core context class VIVO Core utility class



数字知识表示应用案例分析

Google



乌尔姆

乌尔姆是德国巴登-符腾堡州的一座城市，位于多瑙河畔。它附近的大城市有东南部约70公里远的奥格斯堡和130公里远的慕尼黑和西北约100公里远的斯图加特。乌尔姆与其隔岸相望的属于巴伐利亚州的新乌尔姆一起是这个地区的文化和商业中心。 维基百科

人口: 12.37 万 (2011 年 12 月 31 日)
面积: 118.7 平方公里
成立时间: 1181 年
天气: 13°C, 风向西, 风速 3 米/秒, 湿度 77%
当地时间: 星期一 上午 9:51

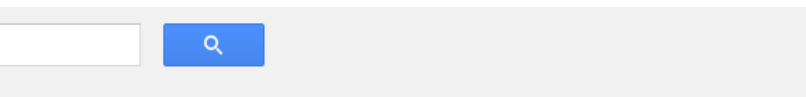


苏黎世大学

路线

苏黎世大学是瑞士的一所州立大学，位于瑞士德语区苏黎世。该校成立于1833年，现有425位教授、两万四千余名学生分布于7个学院、百四十多个研究所，是瑞士最大的综合大学。苏黎世大学在分子生物学、神经科学、人类学等领域享有世界声誉，产生了爱因斯坦等12位诺贝尔奖得主。 维基百科

地址: Rämistrasse 71, 8006 Zürich, 瑞士
电话号码: +41 44 634 11 11



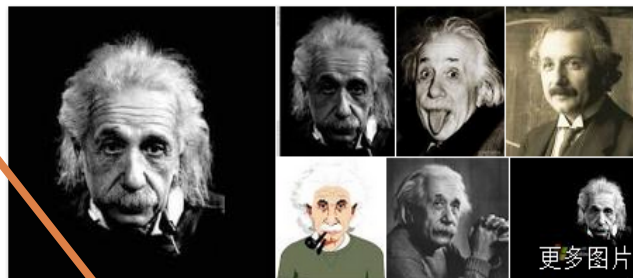
得者,

edia

20世

是著

每年颁



阿尔伯特·爱因斯坦

阿尔伯特·爱因斯坦，20世纪犹太裔理论物理学家，创立了相对论，现代物理学的两大支柱之一。虽然爱因斯坦的质能方程 $E = mc^2$ 最著称于世，他是因为对理论物理的贡献，特别是发现了光电效应而获得1921年诺贝尔物理学奖。爱因斯坦总共发表了300多篇科学论文和150篇非科学作品。 维基百科

生于: 1879 年 3 月 14 日, 德国乌尔姆

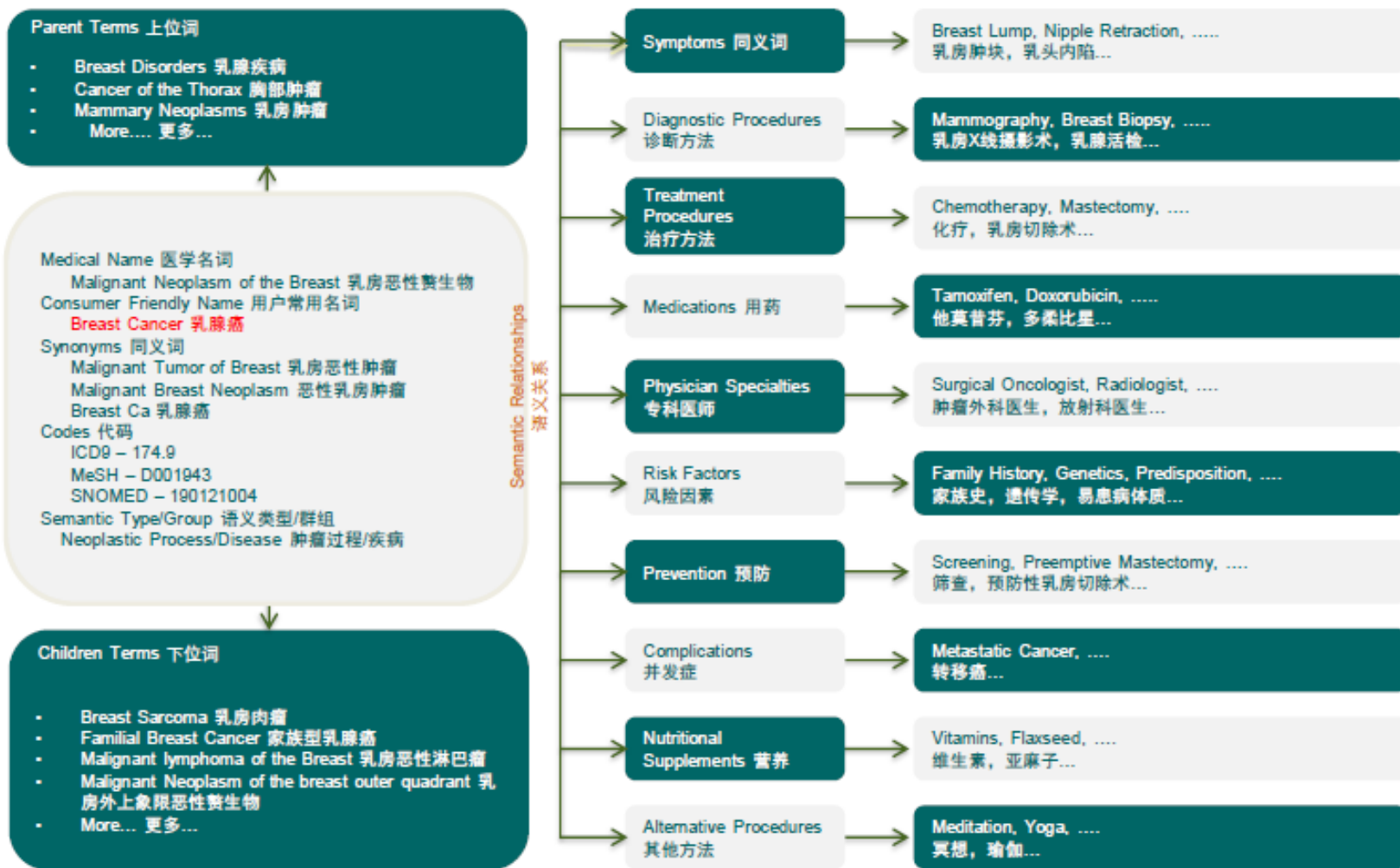
死于: 1955 年 4 月 18 日, 美国新泽西州普林斯顿

教育背景: 苏黎世大学 (1905 年), 苏黎世联邦理工学院 (1901 年), 阿劳州立中学 (1895 年-1896 年), 路易博德文理中学

子女: 汉斯·阿尔伯特·爱因斯坦, 爱德華·爱因斯坦, 麗瑟爾·爱因斯坦, 伊爾莎·爱因斯坦, 馬格特·爱因斯坦

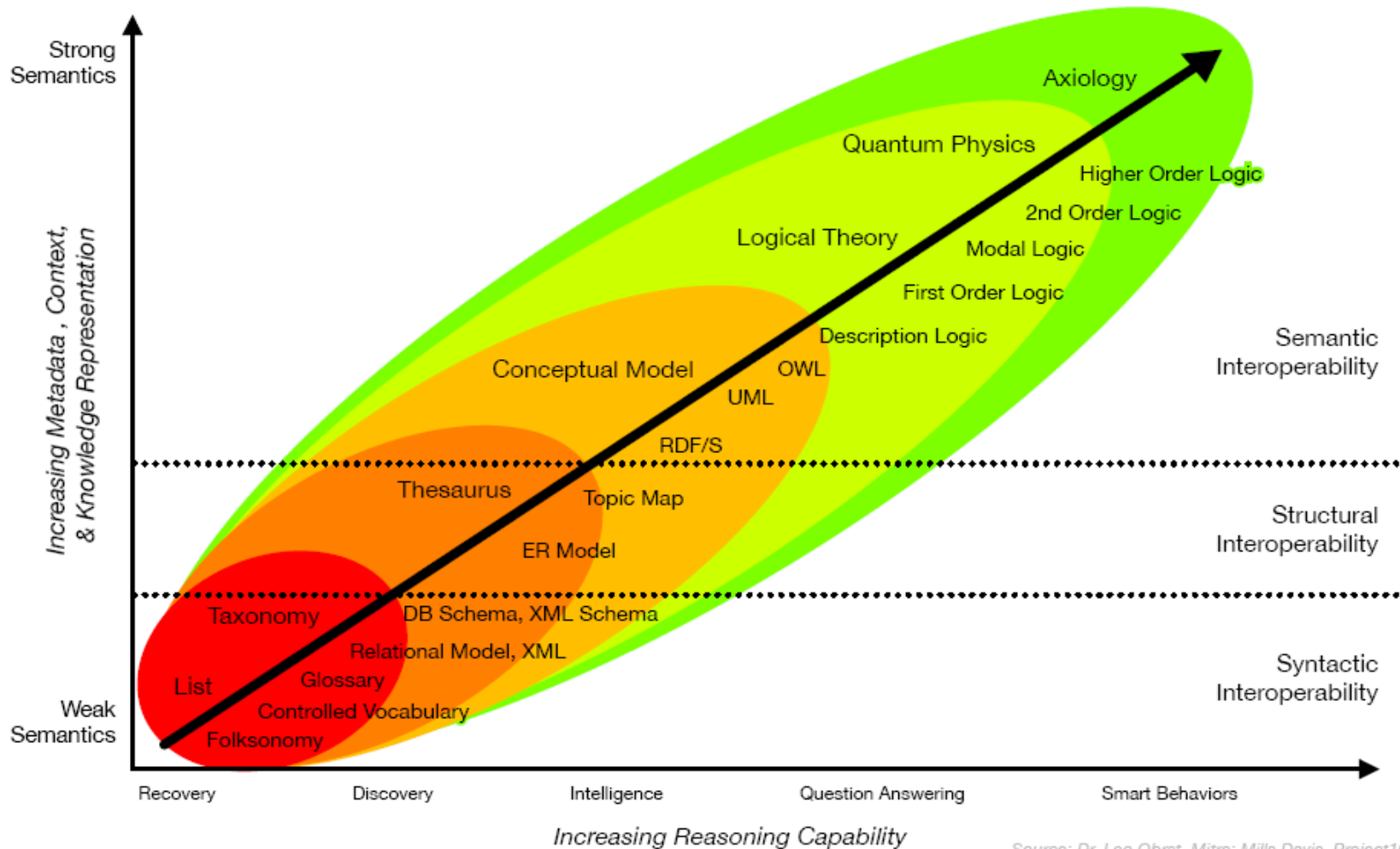
数字知识表示应用案例分析

EMMeT: Elsevier 合并医学分类法





知识表示发展光谱



Source: Dr. Leo Obrst, Mitre; Mills Davis, Project10X



词表体系的缺陷与应对

➤ 关联检索、语义丰富与知识发现服务层出不穷

- 在W3C为代表的各类结构化、数据化、语义化标准的支持和语义网技术推动下，以知识关联、重组和语义互操作为特征的新型知识产品与知识服务首先集中呈现在生命科学领域，如PubMed Entrez、NxOpinion、NCIt、NCBO、VIVO等
- VIVO通过一个广泛适用于科研领域的科研本体和生命科学领域本体支持对海量科研人员、仪器设备、科研活动和项目、科学数据、科研成果以及文献资源等进行结构化、语义化的知识表示与可视化展示，支持科研人员的关联发现
- NCBO 基于NCI Thesaurus（参考术语本体）和连接到UMLS中的200多个本体，实现了对自由文本大规模语义标注，使非结构化的自由文本数据变得结构化、标准化。为研究人员提供在线工具及基于网络的资源（如BioPortal），使研究人员可以在生物医学研究及临床实践的各个方面实现对不同本体的访问、审查与集成，搜索与特定的本体术语相关的生物医学资源



以湿地的一种类型：关键词“mire、mires”，时间“1900-2014”，在WOS上检索相关湿地论文，经过机器和人工判读，获得论文2451篇文献，提取作者关键词4709个（去重）。

•分析DDC类目表中与湿地相关的类目，从水文学和水生态学等下的51个类目获得术语2642个；随后进一步扩展，获取环境保护、环境监测等下的DDC类目，共152个类目，术语6910个。

•与DDC提取的湿地叙词表（包括概念优选和非优选术语）比对，匹配上了419个，仅占8.9%。

•与STKOS所有术语（包括概念优选术语和非优选术语）数据比对，匹配2151个术语，比对上的STKOS术语占作者关键词总数的45.68%。（备注：1990年之前文章没有作者关键词和文摘）。

词频	关键词	机器自动匹配数量	百分比
1	3490	1336	38.28%
2	558	321	57.53%
3	230	154	66.96%
4	120	92	76.67%
5	68	52	76.47%
6	43	35	81.40%
7	33	24	72.73%
8	23	17	73.91%
9	21	17	80.95%
10	17	14	82.35%
11到20	62	47	75.81%
21到30	21	20	95.24%
31到40	7	7	100.00%
41到50	6	5	83.33%
50以上	10	10	100.00%
总数	4709	2151	45.68%



(1) 以上分析可以得出，不适合采用DDC分类表确定湿地词表的基础范围。由于湿地研究涉及范围广，从类目分析来确定范围，会大大影响词表的内容范围。因此以作者关键词和综合词表整体比对的数据源作为创建专业词表的基础更为合理。

(2) 在匹配数据中，词频高的数据匹配率高。从下表统计可以看出，词频在50次以上的数据，匹配率为100%，词频为1次的数据，匹配率为38.28%。总体上，STKOS词表中的概念基本能覆盖主要研究内容，在数据补充时，也可以根据词频来确定补充顺序。

(3) 作者关键词中未匹配的词汇分析：需要进一步处理复合词，拆分成多个叙词，或者对形容词加上叙词的表示形式进行拆分；未匹配高频词，需要进一步规范增加到湿地叙词表中。

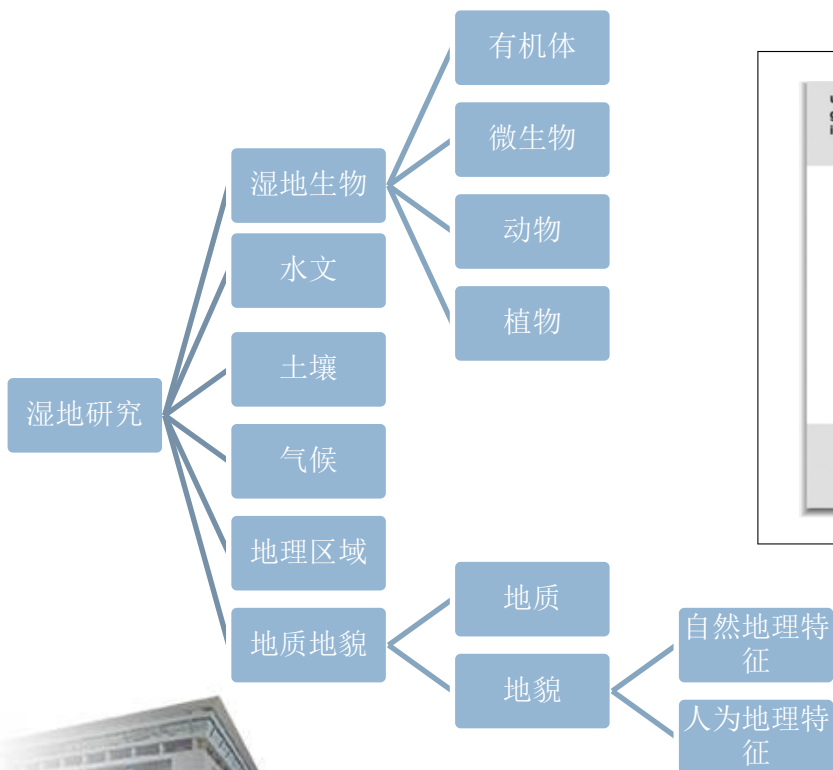
(4) 湿地叙词表结构检查和重构：作为一部叙词表，除了基础的术语外，更重要的叙词表所表示的用代属分参的数据结构，因此对所有术语，在继承来源叙词表的属性基础上，需要进行入口词检查、语义逻辑关系检查，以满足叙词表标准规定的要求。

主题识别（主题选择：湿地植物和动物、湿地地理位置）

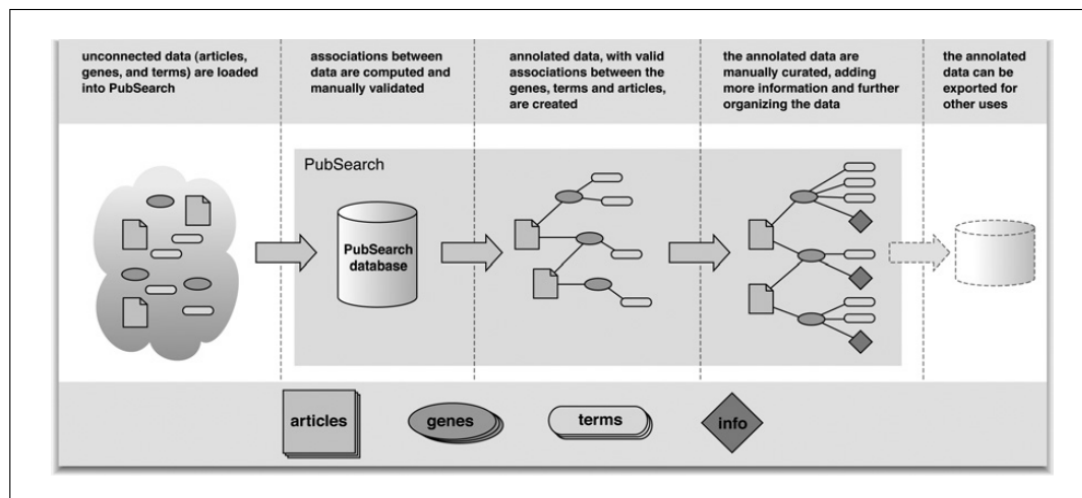
由于作者关键词涵盖内容的广泛性，以此形成的叙词表，仅是散乱的单点，很难论文、数据、术语相互关联的网络，也难以支持湿地数据的领域分析和研究。



如何能够突破词形匹配，深入语义，揭示湿地研究相关的各个方面，将术语与所属实体类型相结合成为了难点问题。对于湿地研究而言，主要涉及湿地生物、水文、土壤、气候、地理位置、地质地貌等多个领域，如下图。

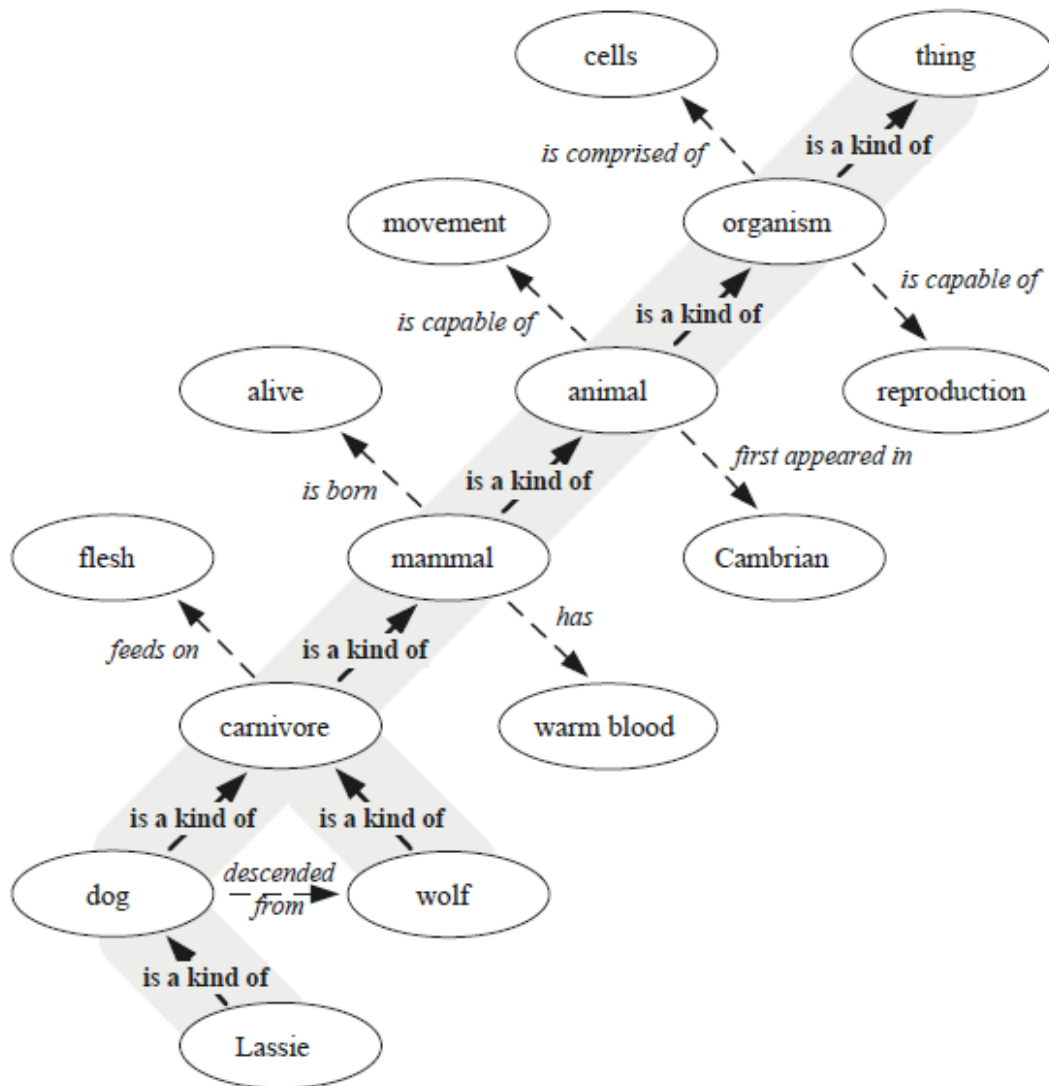


下图为TAIR数据库采用Pubsearch系统，实现拟兰芥论文中基因、论文、基因本体和植物本体、信息的标注和关联，引自Canny Yoo, Iris Xu, ect. pubsearch and pubFetch. Current protocols in bioinformatics,2006:9.7.1）。



本体与语义网络 (Ontology/Semantic Web)

Ontology



Ontology Use

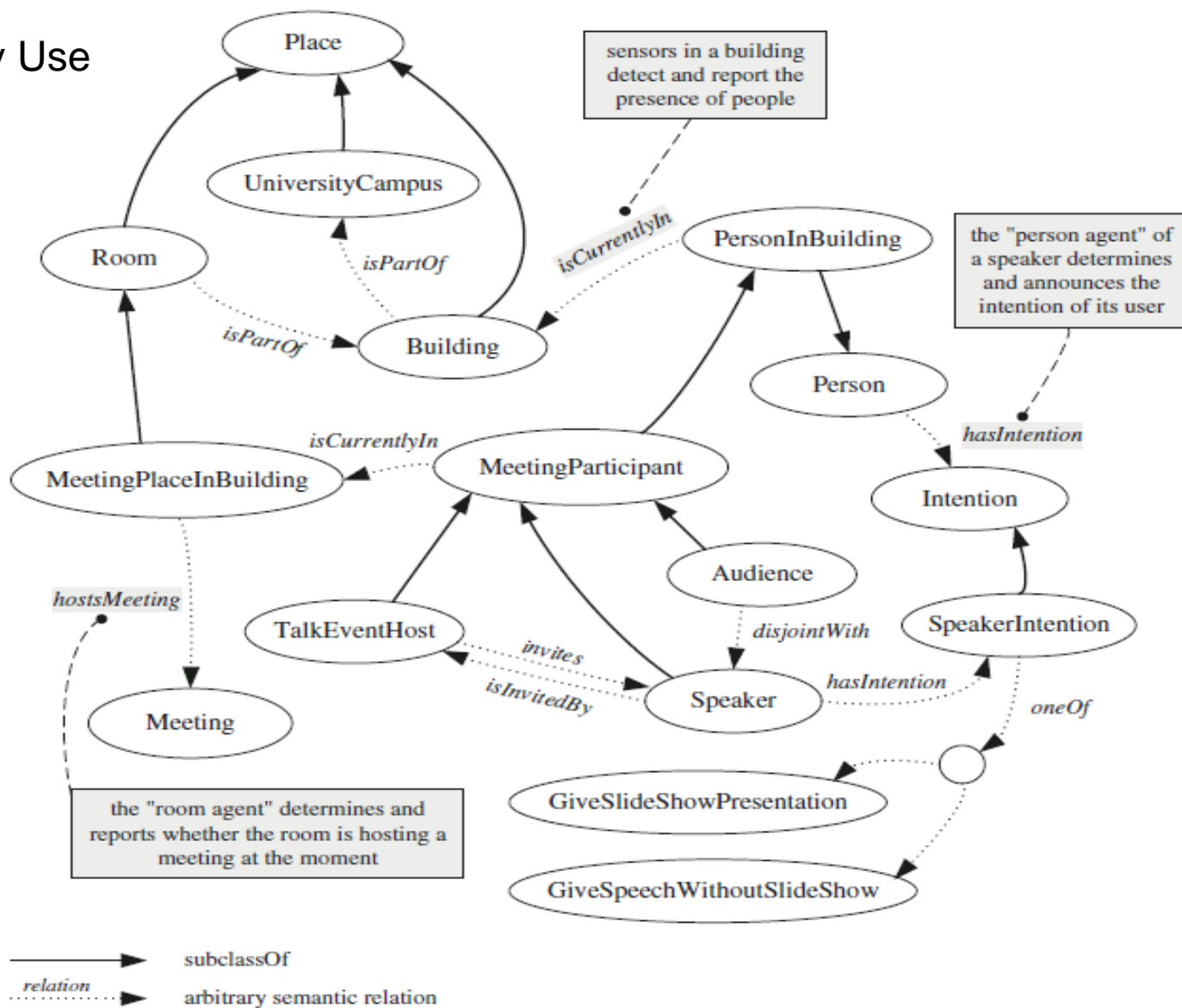


Fig. 3.3 A part of the ontology supporting the intelligent context broker (Chen et al. 2003a)



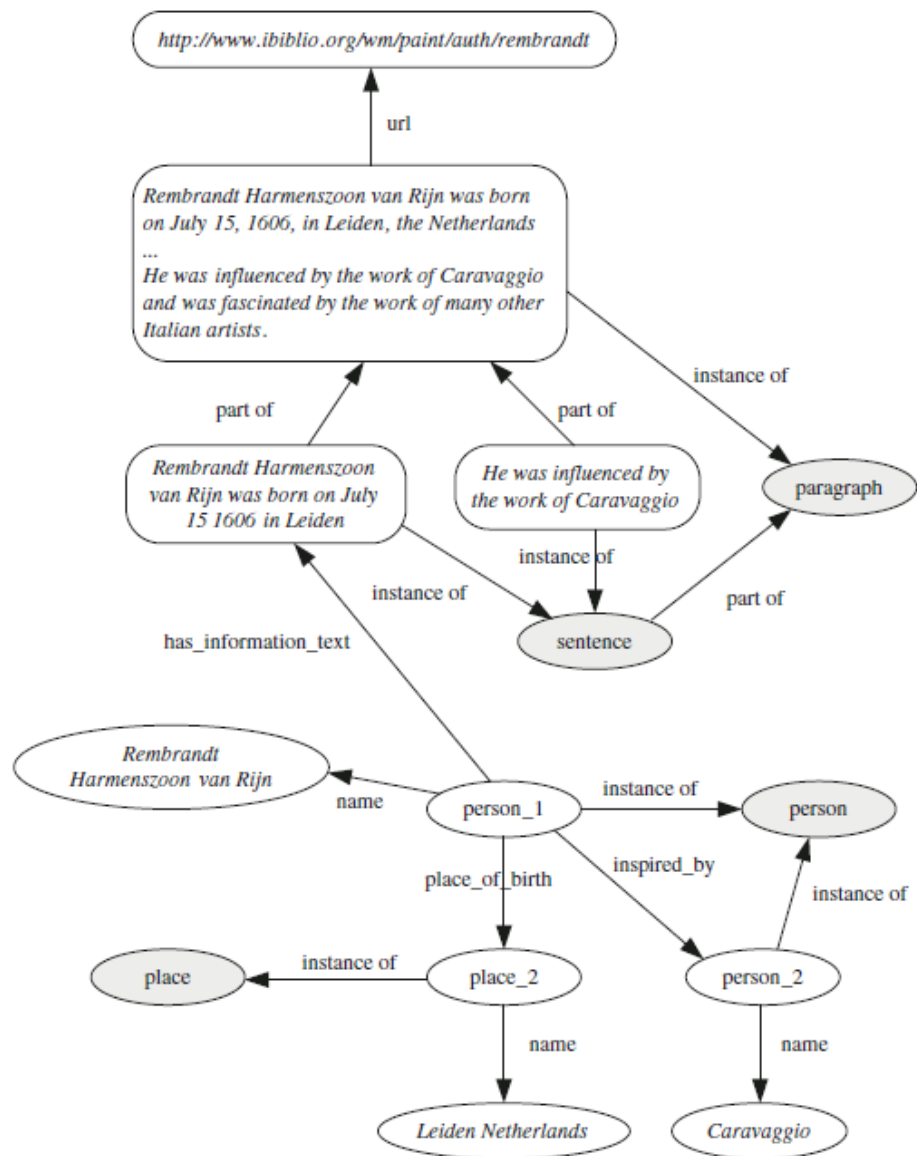


Fig. 3.11 Knowledge base populated with knowledge presented according to the ontology domain representation (Alani et al. 2003)



Knowledge Extraction

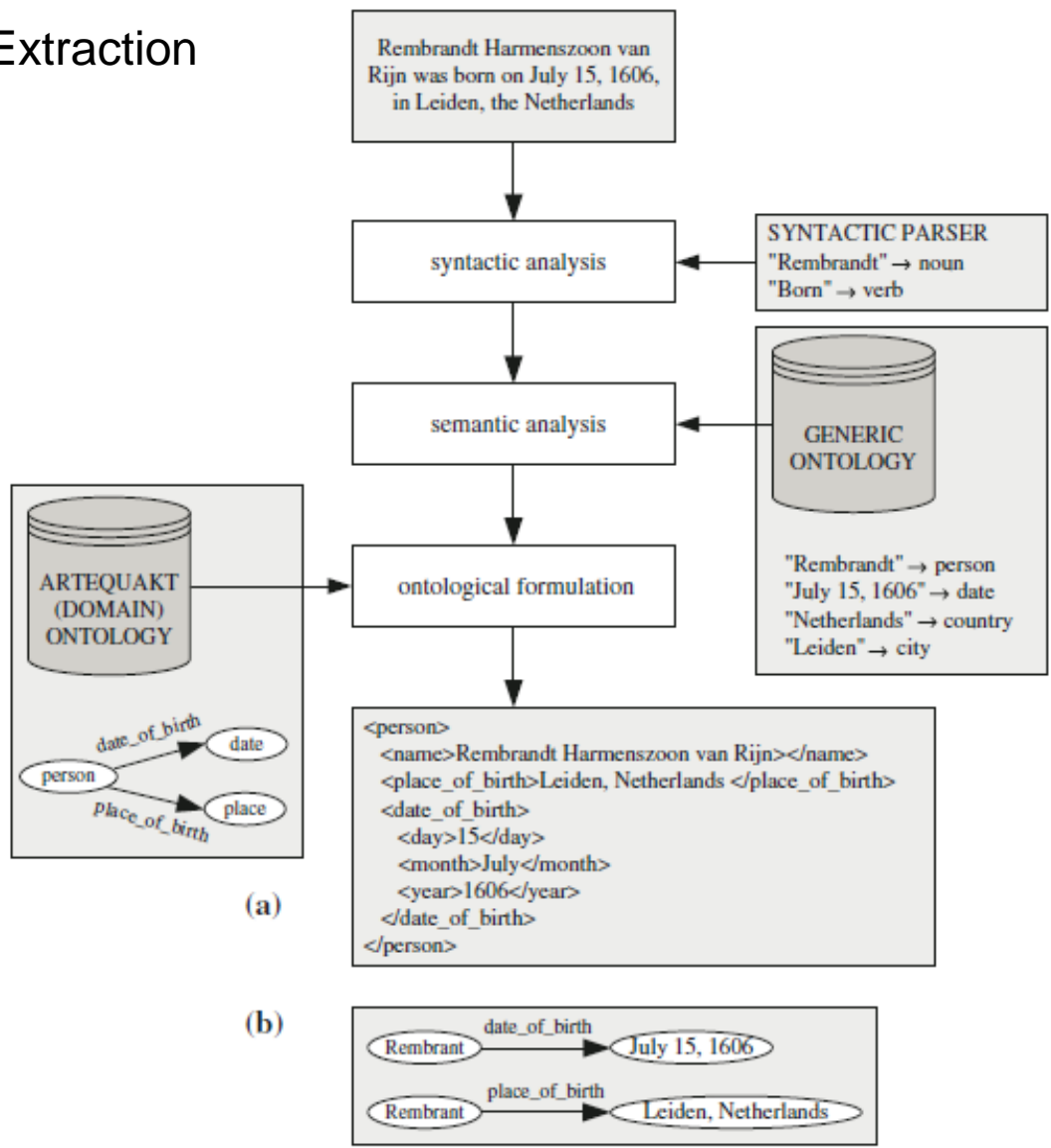
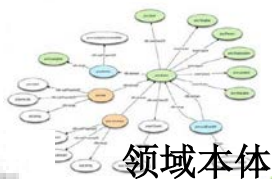


Fig. 3.10 An example of knowledge extraction from a Web page using Artequakt (Alani et al. 2003)





领域本体

LarKC: The Large Knowledge Collider
The aim of the EU FP7 Large Scale Integrating Project LarKC is to develop the Large Knowledge Collider (LarKC, for short, pronounced "lark"), a platform for massive distributed incomplete reasoning that will remove the scalability barriers of currently existing reasoning systems for the Semantic Web. LarKC is sponsoring the following international scientific events:
11th International Conference on Principles of Knowledge Representation and Reasoning (KR 2008)



- 物种名录
- 图片库
- 标本库
- DNA条码库
- 知识组织体系

知识抽取

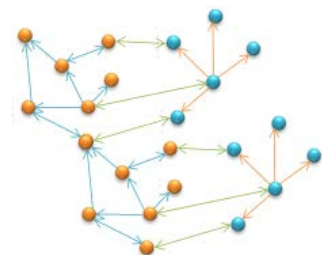
知识集成



文献库

知识检索与知识关联

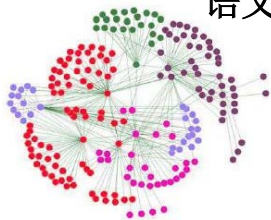
从关键词检索到语义检索
从单纯知识对象到异构对象关联与集成
从孤立文献揭示到基于知识对象的跨界融汇
从信息检索获取到知识探索利用



语义检索

知识服务

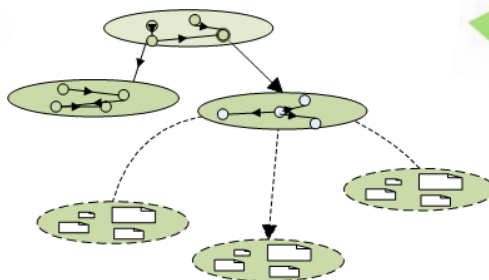
知识索引



知识探索



可视分析



知识表示与关联实例

植物多样性复杂数据汇聚



物种名录



标本库



文献库



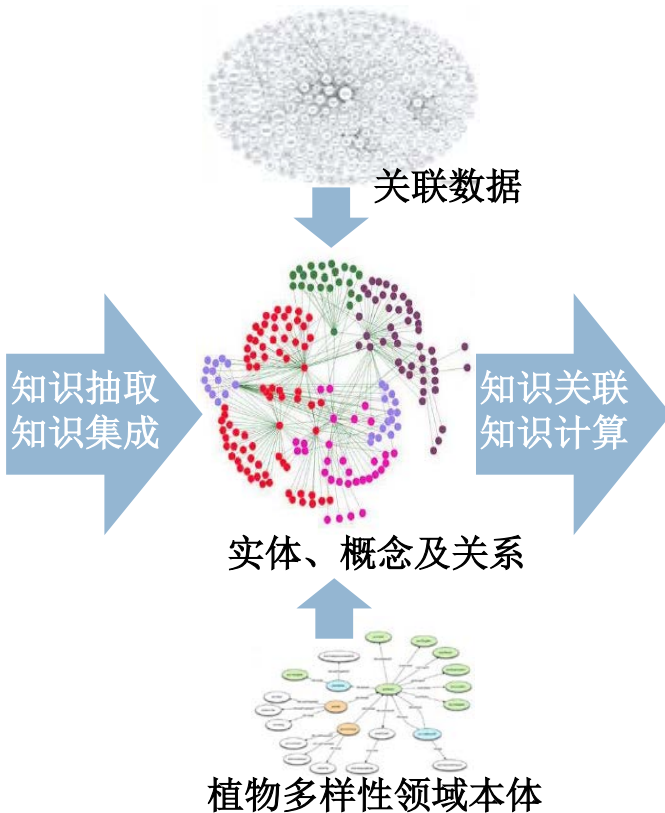
图片库



地名库



DNA条码库



知识检索与关联呈现



语义检索

- 语义索引
- 分面检索
- 本体导航
- 个性化知识
- 多维语义揭示



可视化分析

- 主题聚类
- 热点刻画
- 趋势分析
- 前沿探测



开放关联与集成

- 数据集成融汇
- 自动推理
- 关联发现





Thank You!

