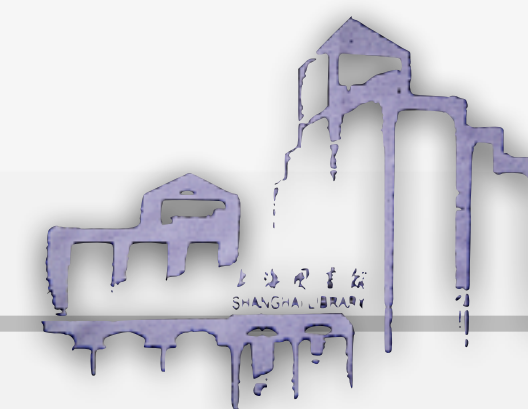




关联数据与信息组织

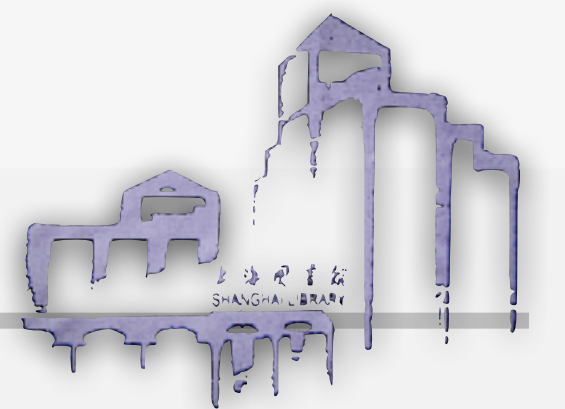
刘炜 kevenlw@gmail.com





智慧时代的信息组织

刘炜 kevenlw@gmail.com



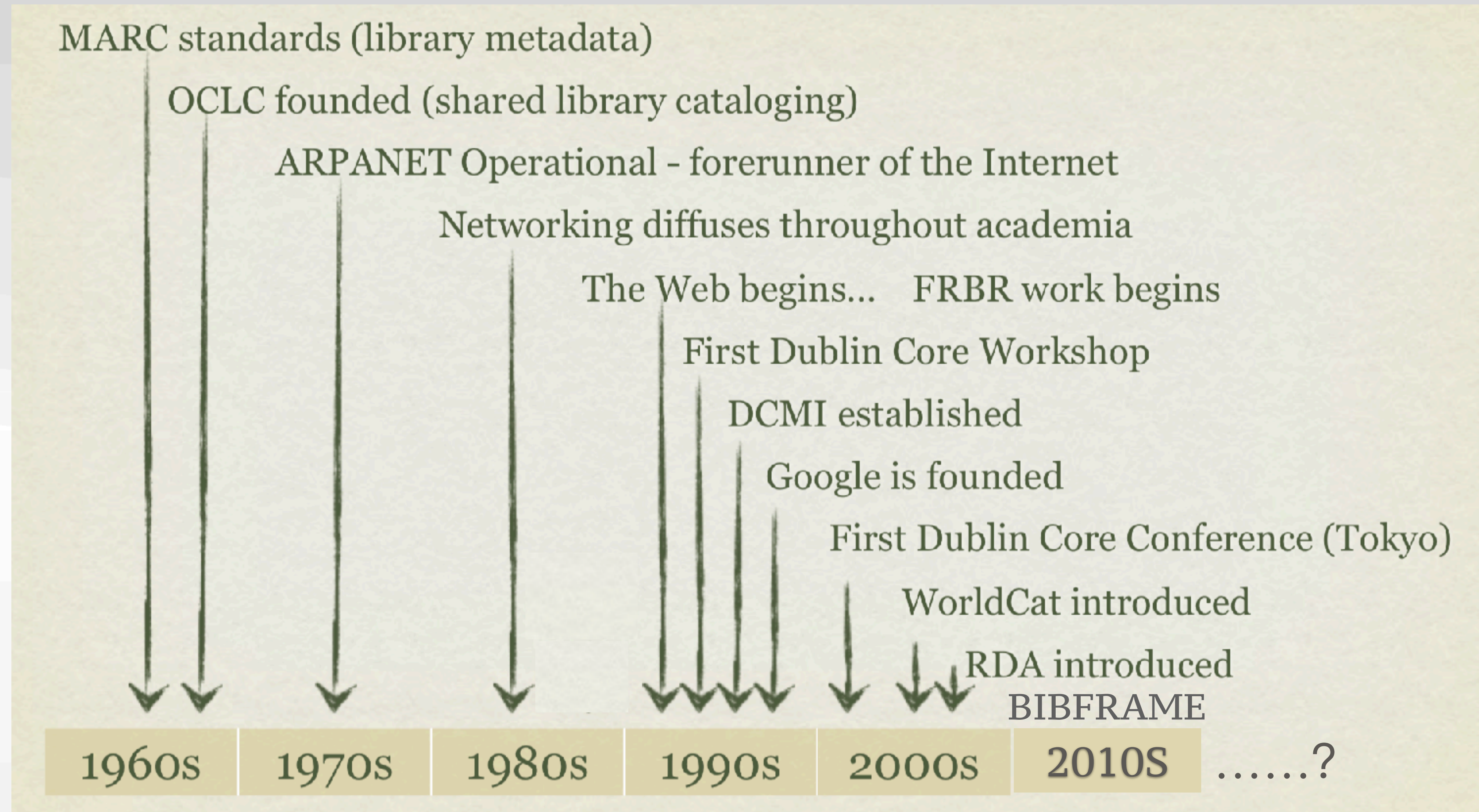
大数据时代的文献编目

信息组织的历史和未来

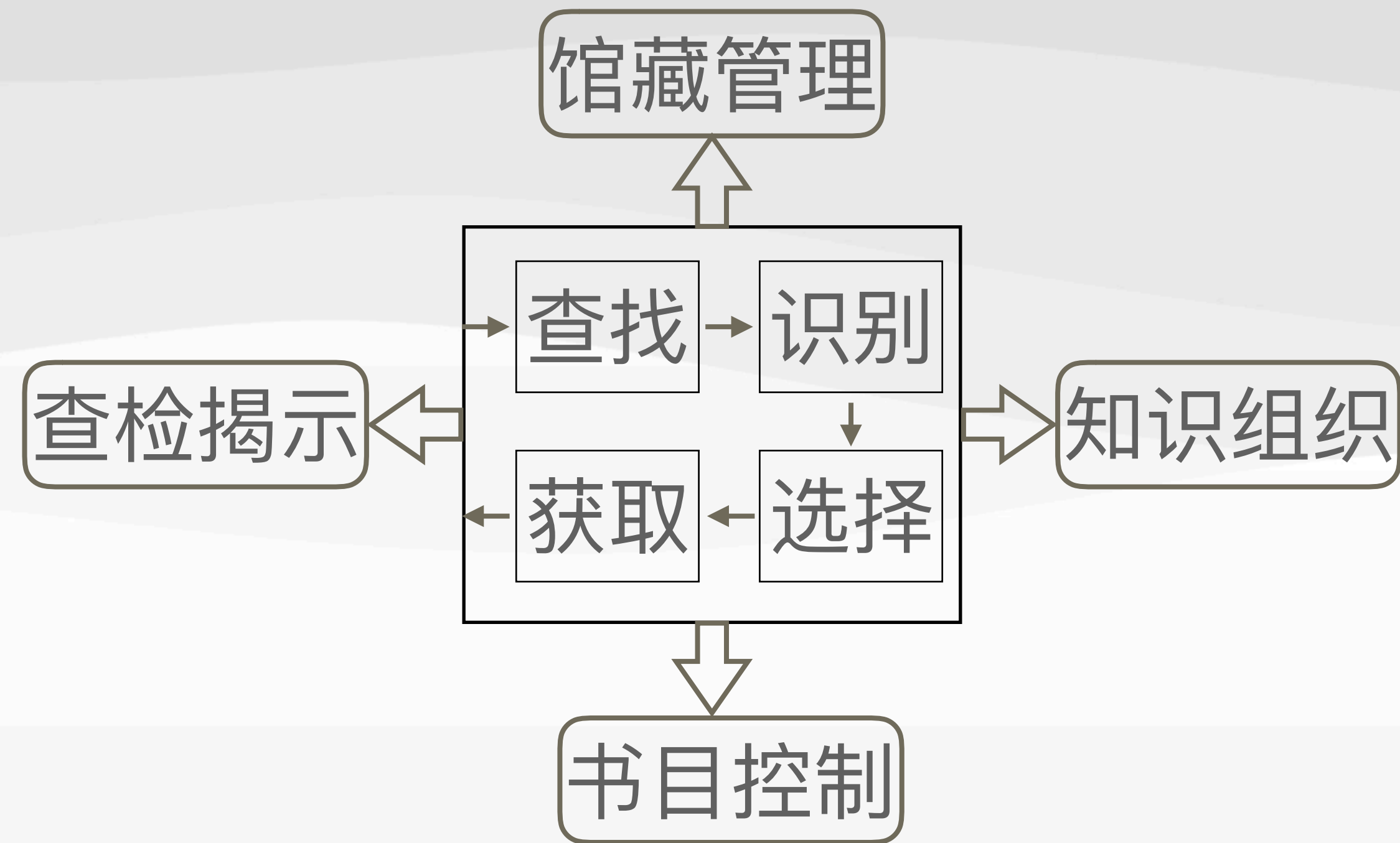
从关联数据到智慧数据

大数据时代的文献编目

历史



为什么编目？文献编目的功能需求



几个概念：文献、信息、知识、语义

编什么目？描述万物

Group 1	Group 2	Group 3
Work	Person	Concept
Expression	Corporate body	Object
Manifestation		Event
Item		Place

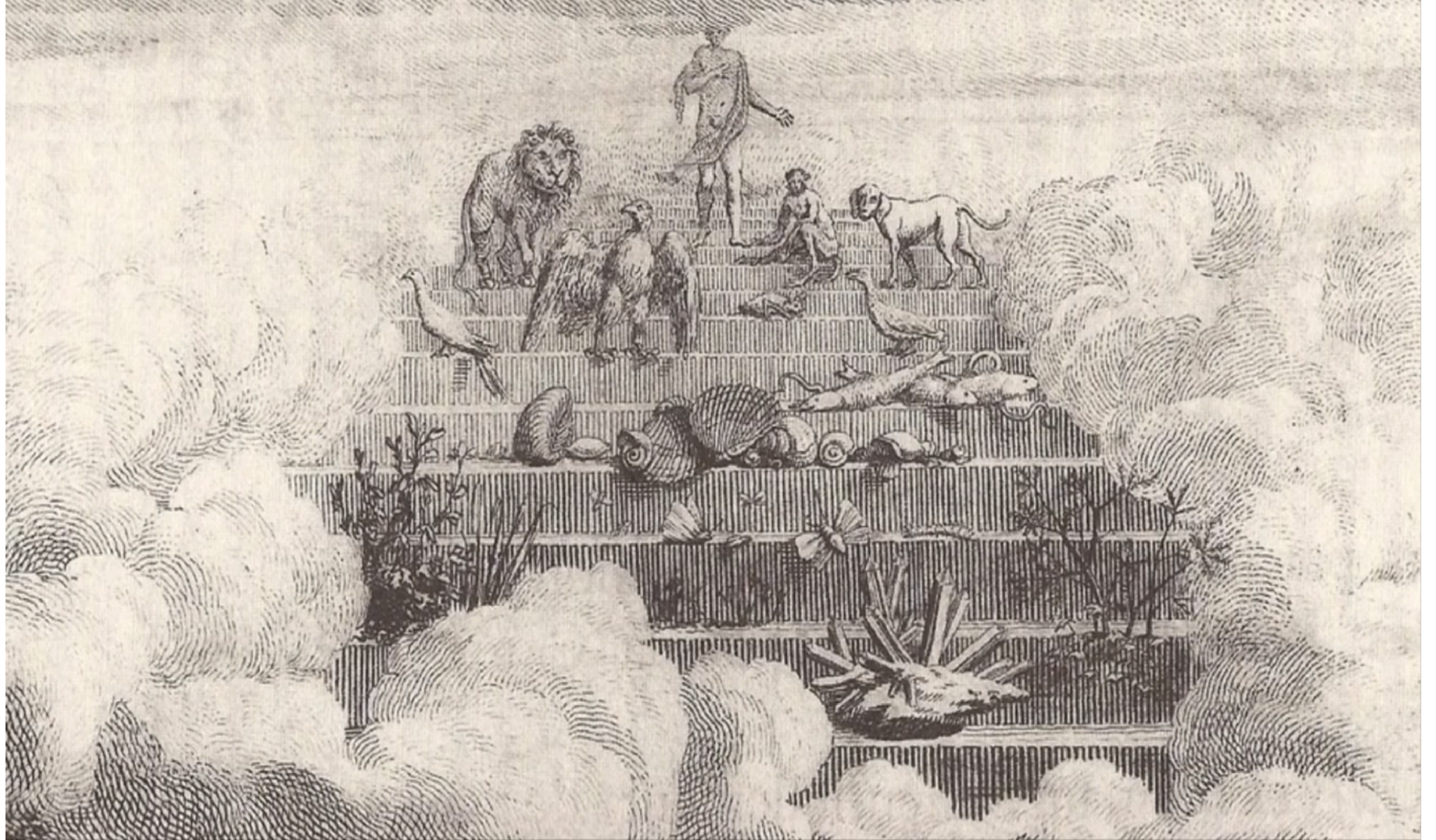
文献编目的新挑战

- 来自资源：数字化、数据化；馆藏已不再是那些馆藏，知识已不再与载体捆绑，如何建设？
- 来自用户：碎片化、媒体化；用户驱动型、业务驱动型、数据驱动型？
- 来自技术：语义化、知识化；深度、广度、粒度、态度？
- 来自模式：智慧化、虚拟化；外包、内包、众包、机包？
- 来自应用：自描述、自学习；一切都可以自动生成吗？

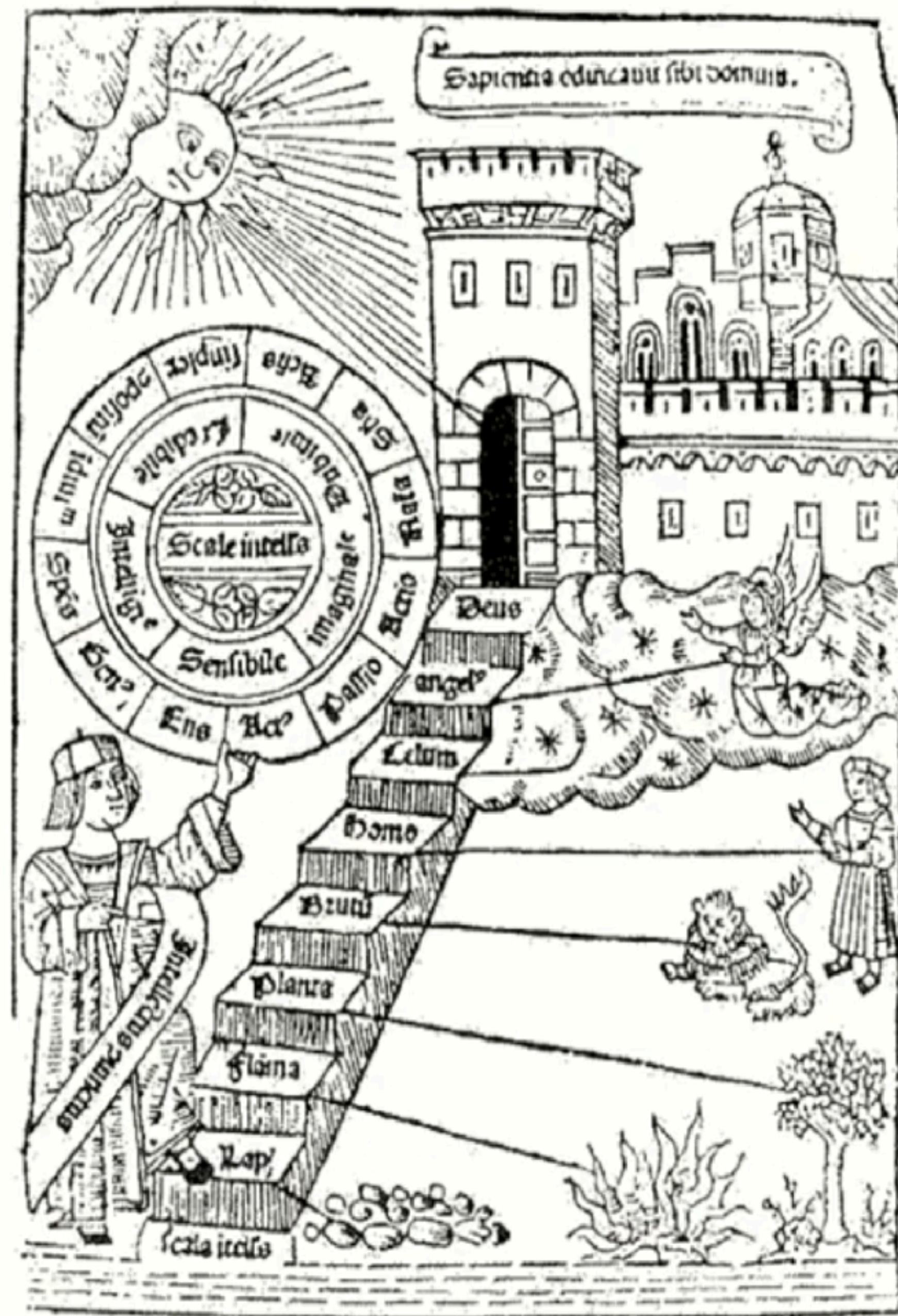
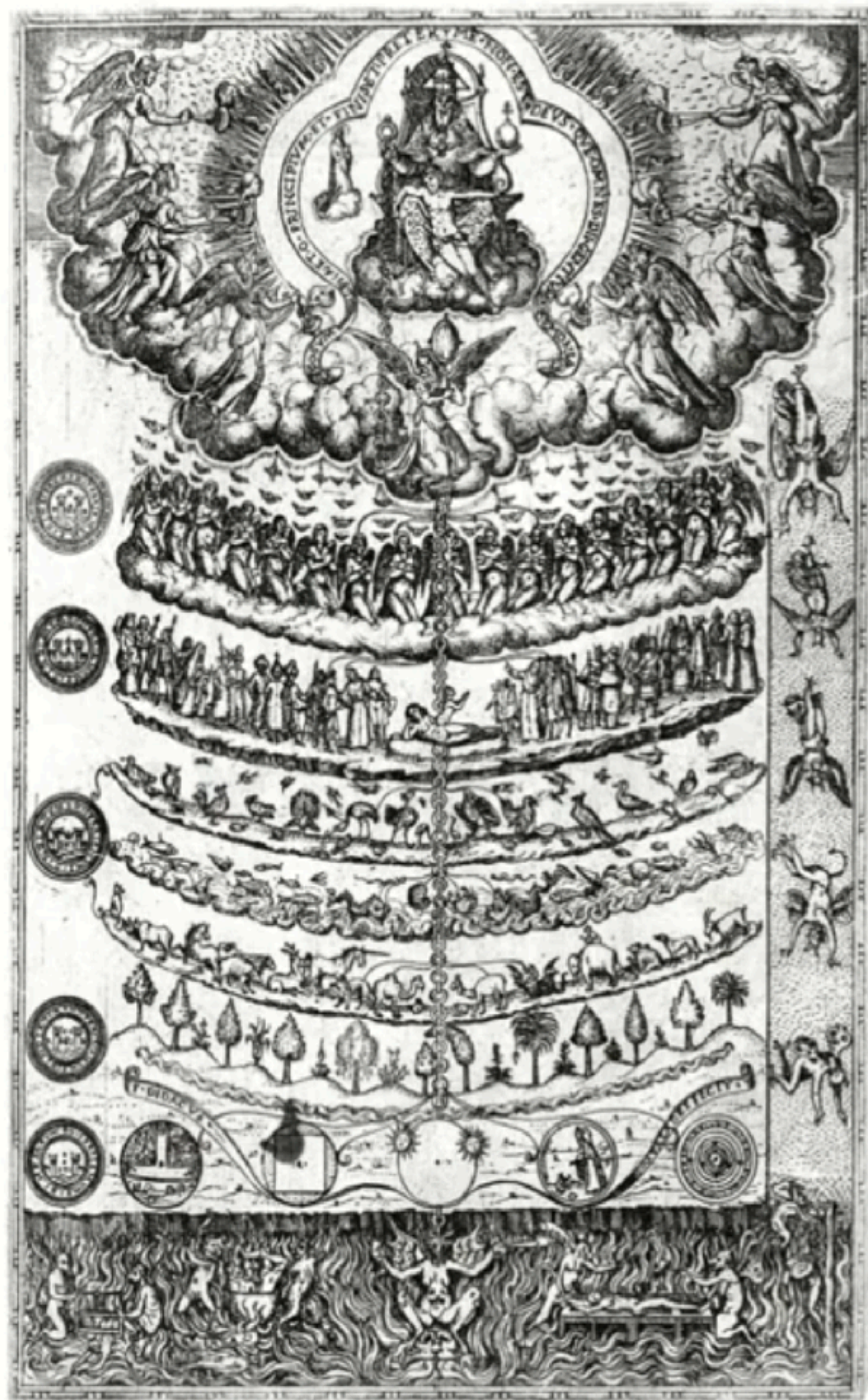
大数据时代文献编目的十个方向

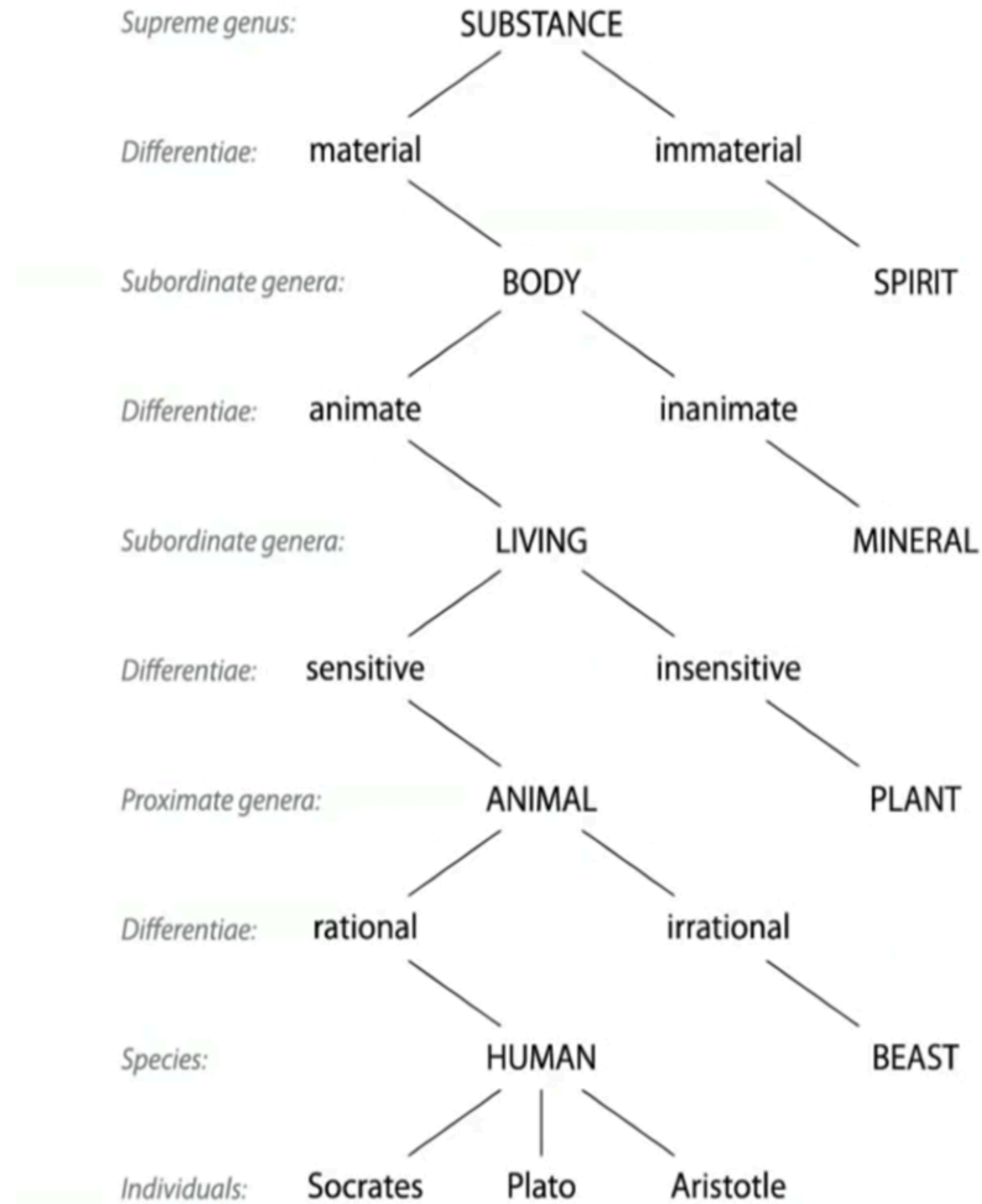
- 描述对象：从载体到知识
- 描述内容：从文献到语义
- 描述原则：从专业到通用
- 描述规则：从静态到动态
- 描述方式：从人工到机器
- 描述目标：从揭示到模拟
- 描述规模：从机构到网域
- 描述过程：从集中到始终
- 描述主体：从专家到众包
- 描述标准：从统一到够用

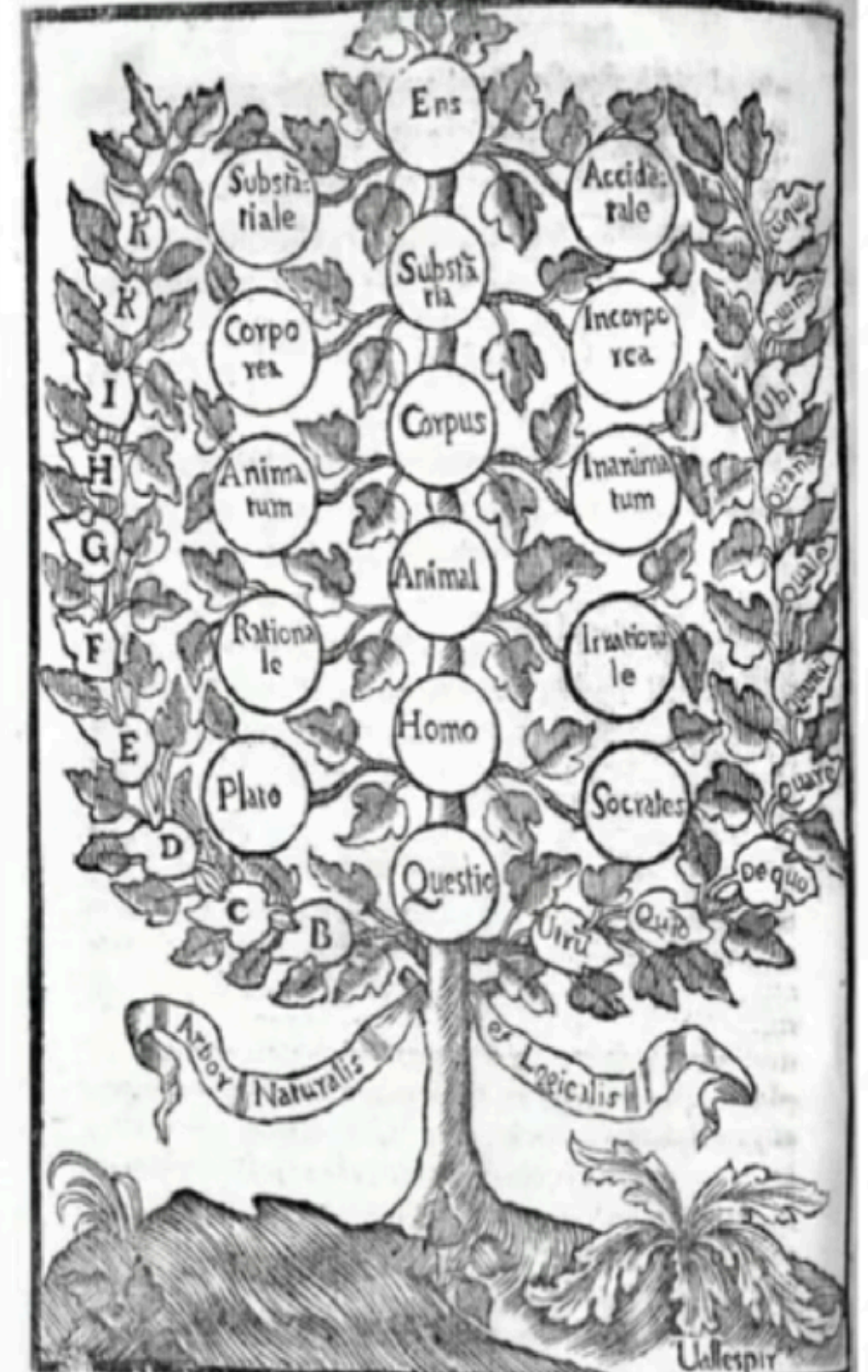
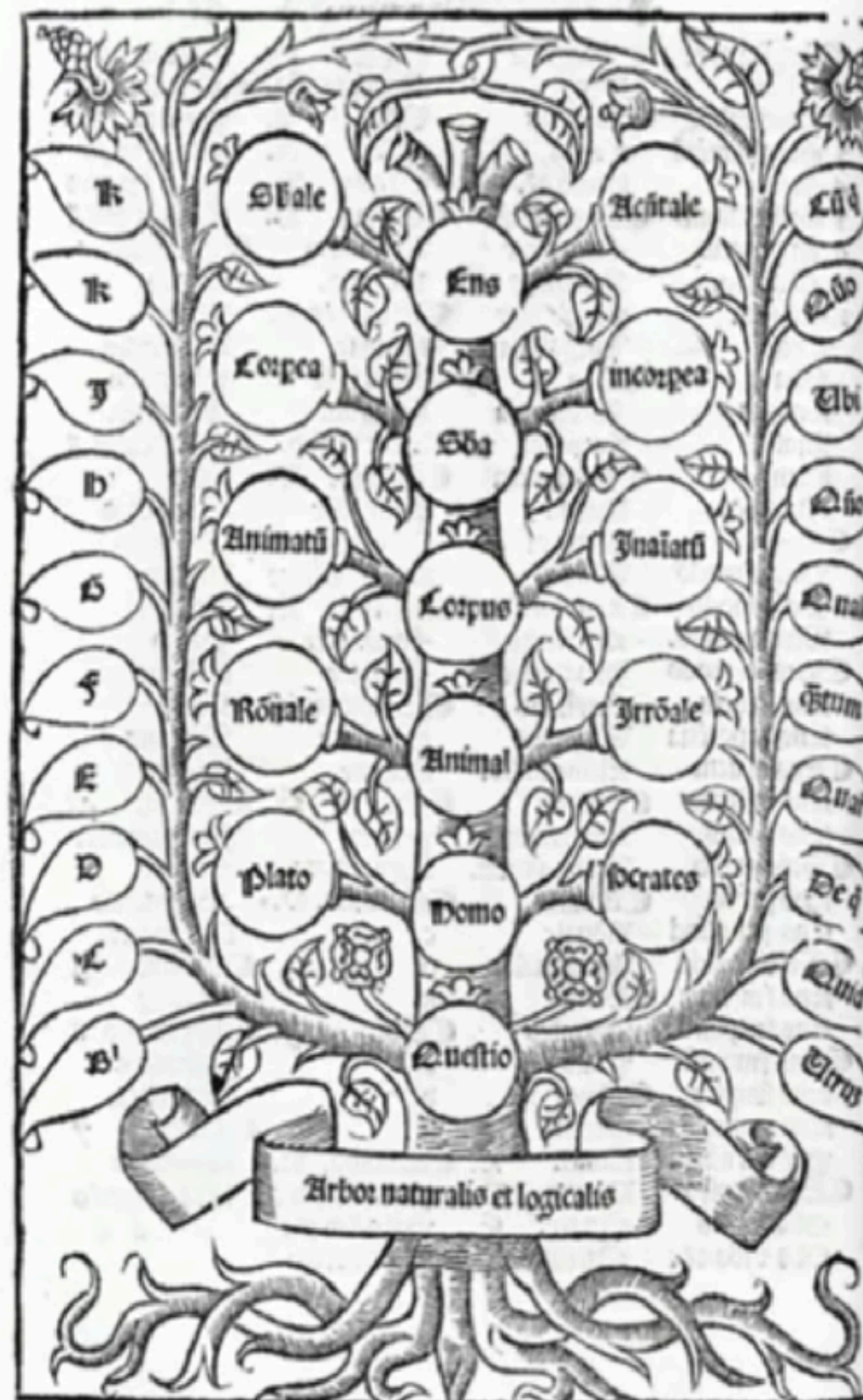
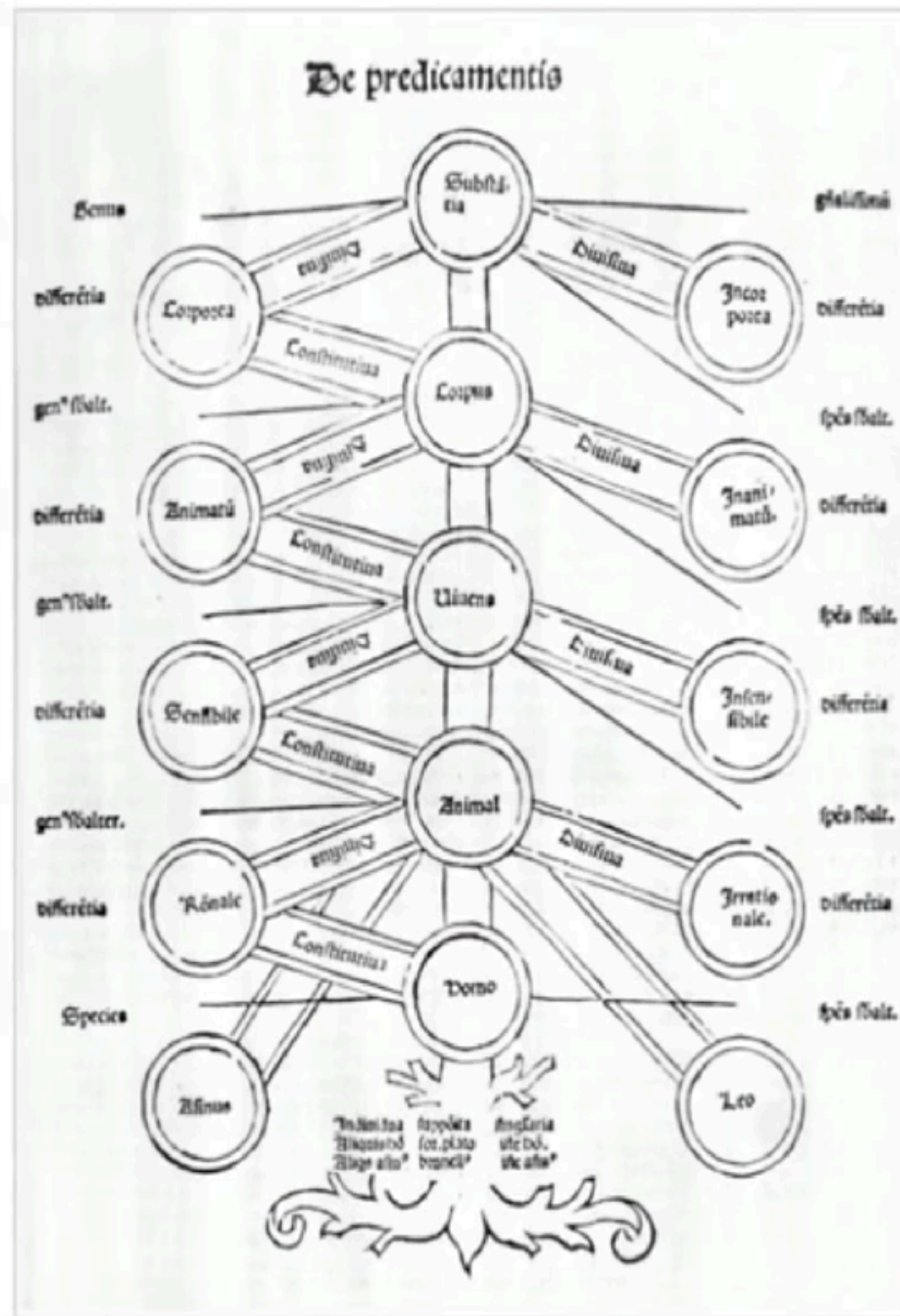
信息组织的历史和未来



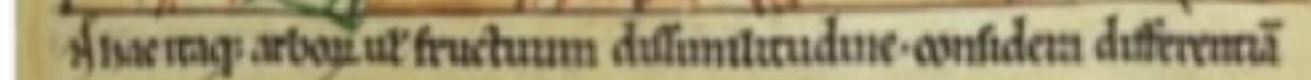
The Chain of Being, Charles Bonnet, 1781



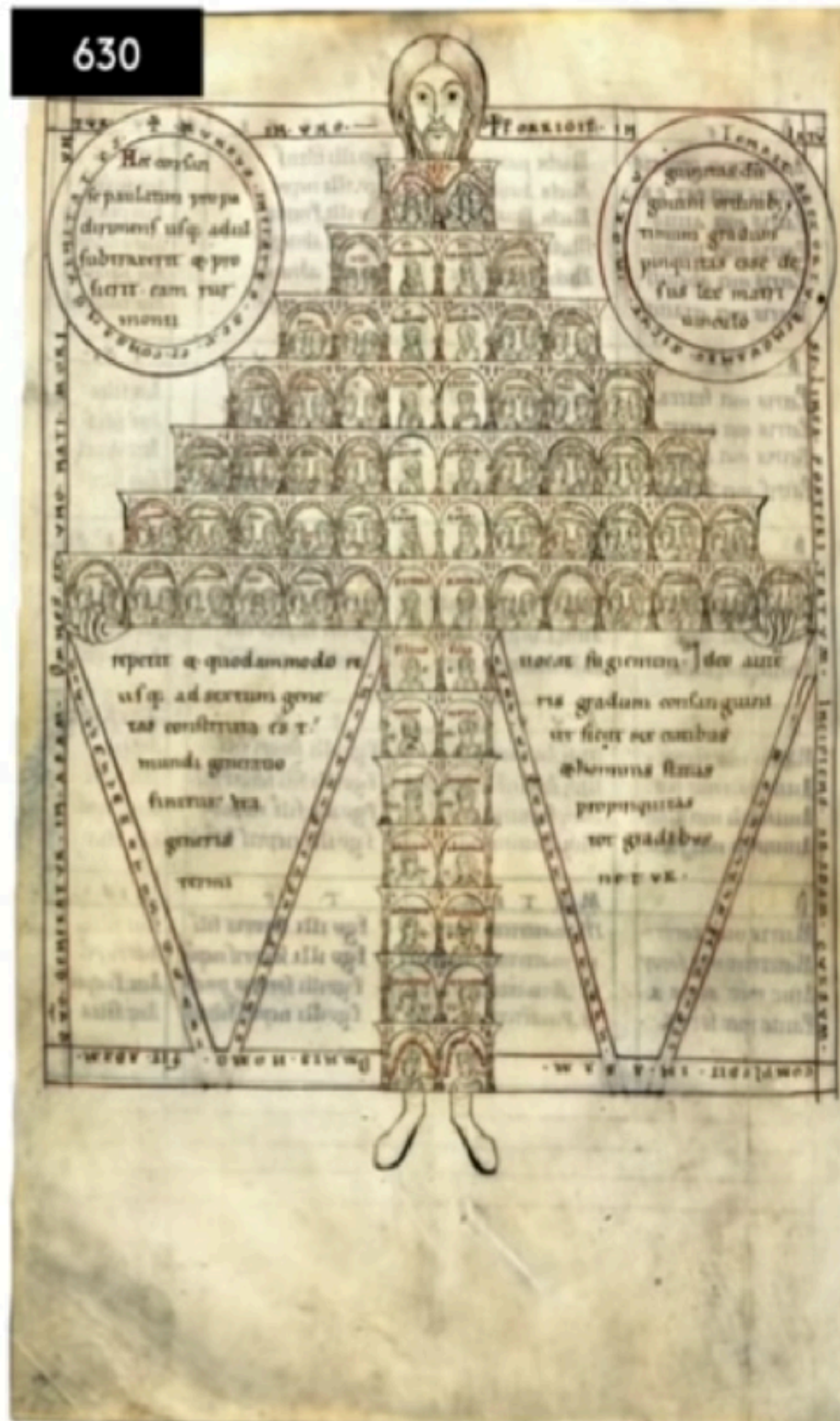




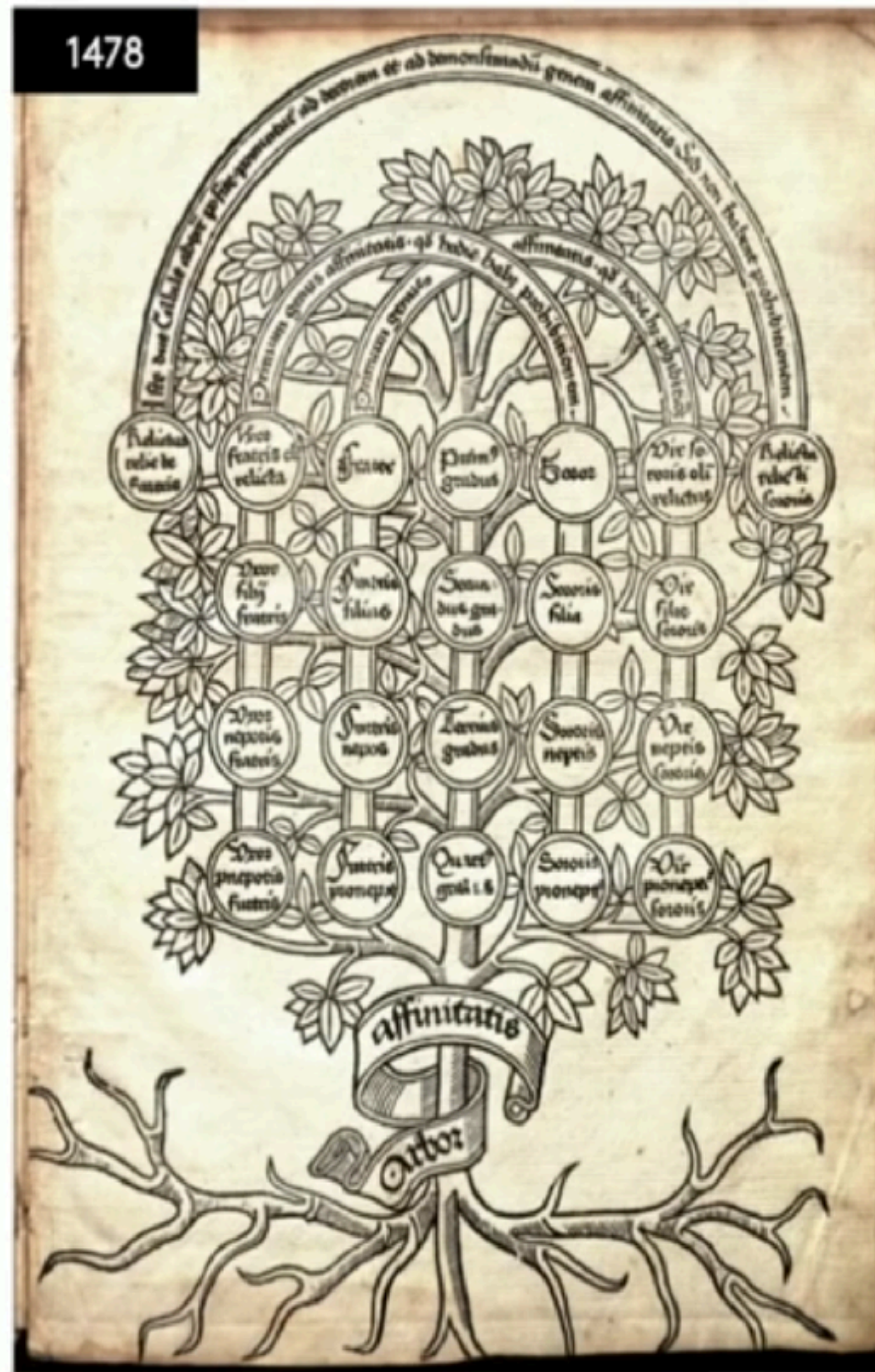
Porphyrian Tree (oldest Tree of Knowledge)



630



1478



1471

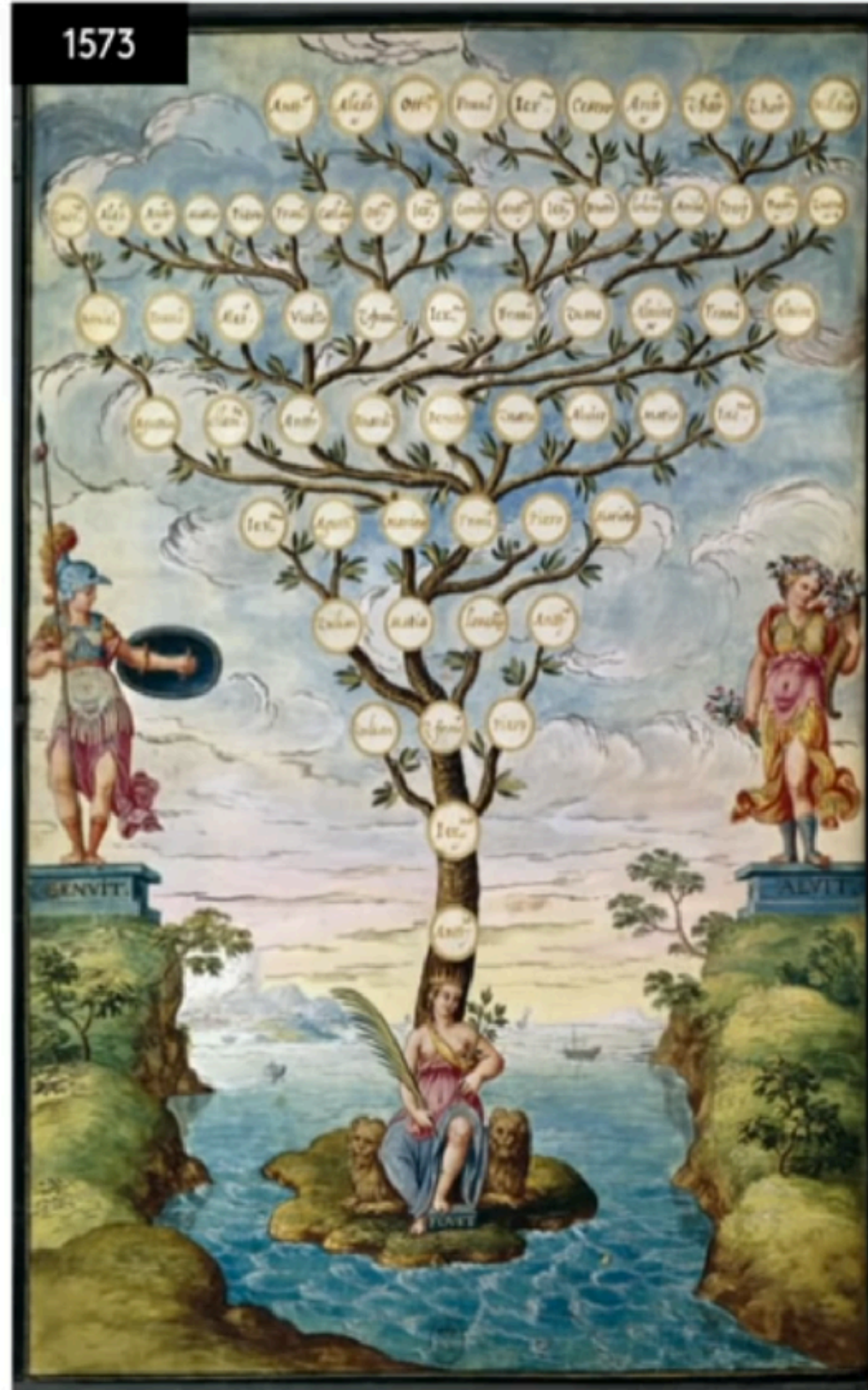


Consanguinity (Blood ties between people)

1632



1573



1493

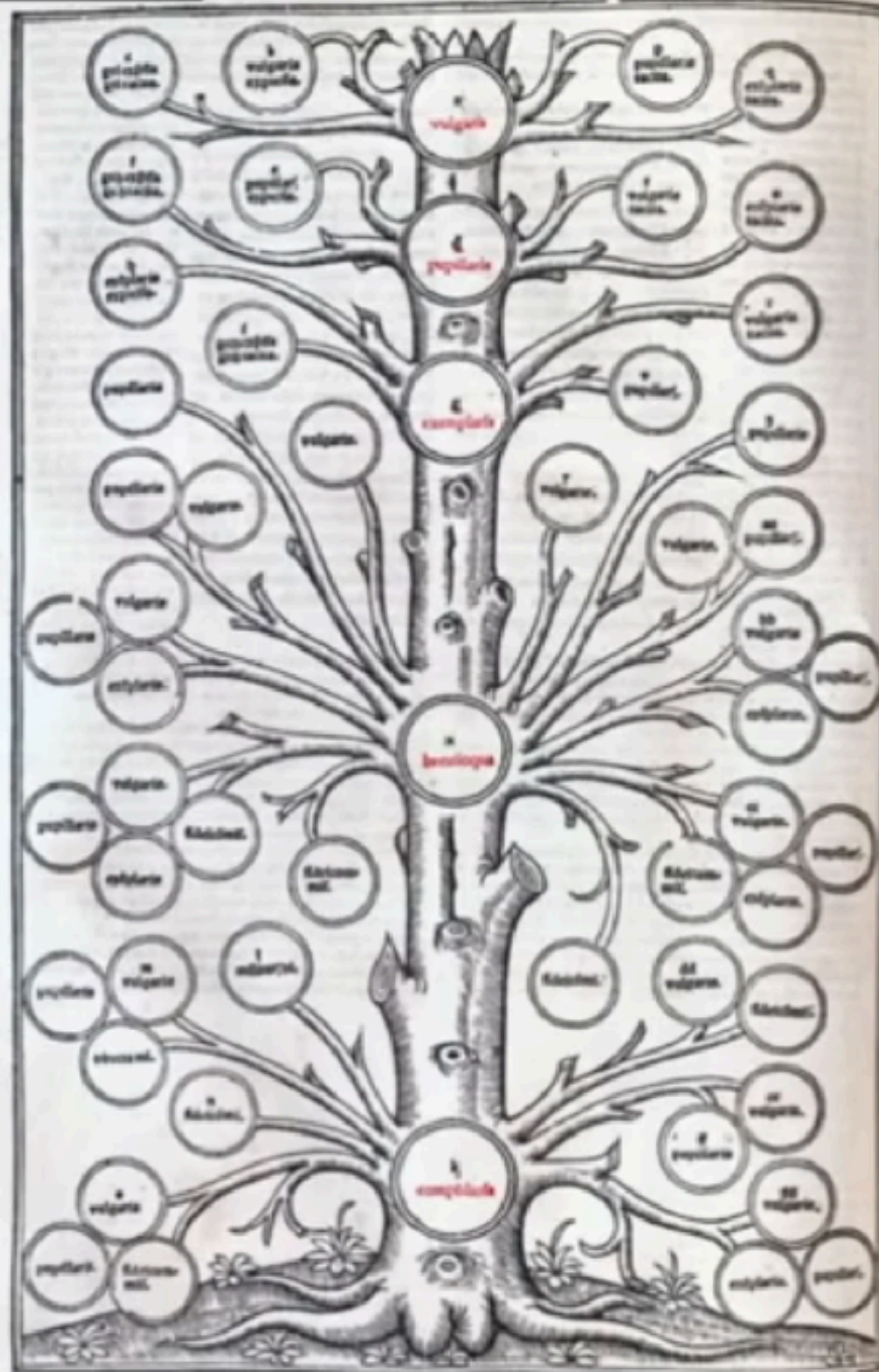
der het vij. sijn vij. besasse europaz vij. ward vō water vmb 8 edelst welli gebenedeyt. vij. upsteyng genamsche. vnd sind vō im entspangt. xv. geschlecht. aber durch die sijn 'Ja'ngschafft. Gomer der erst gepoon isphas sam in europam vnd richtet Gomerias das von di frecht galat: vnd das land nach in galacia genant sind. das selb lād stisset an haf pamen vij. lufstamen die hat es von mittertag. vnd vom ndergang vnd mitternacht das wendlinet. vnd vom auffgang den stuf signant vnd die land der teitschen. Gomer der erst sijn isphet daron galacy. vnd het vier sijn der erst Alanes. daron samacia 8 schysit lewet in 8 neff 8 meothidische pfürsche wende. vij. ist en elide geit. vol vnsiliger parren. Der ander Daphat oder raphaa. daron paslagones. vij. 8 land paplagoma 8 flennen asie Der dert Thogoma. daron friga vij. 8 land friga 8 mindern asie vromals darbana darnach troya genant. Der ander sijn isphet Magog daron schyt vij. von in Schia vnd Gethia. Der dert medar oder madeus. dant komē medr vij. vō in media das lād nabit bey affria vij. persia. Der viert Janaan. dant greci vij. hat vier sijn. der selb janaan machet ionas die frecht. Jona was en ge- gent 8 frechen zwische ca- ra vij. cold mit thampa ge- nū. 8 erst lēhsan dant he- lse. daenach Hely genā. vij. colia 8 isel sol. xvij. schet vō weidhe lād sijn 8 ander tharja. 8 in 8 mu- dem asia tharfos machet. die erren elices. vij. nach



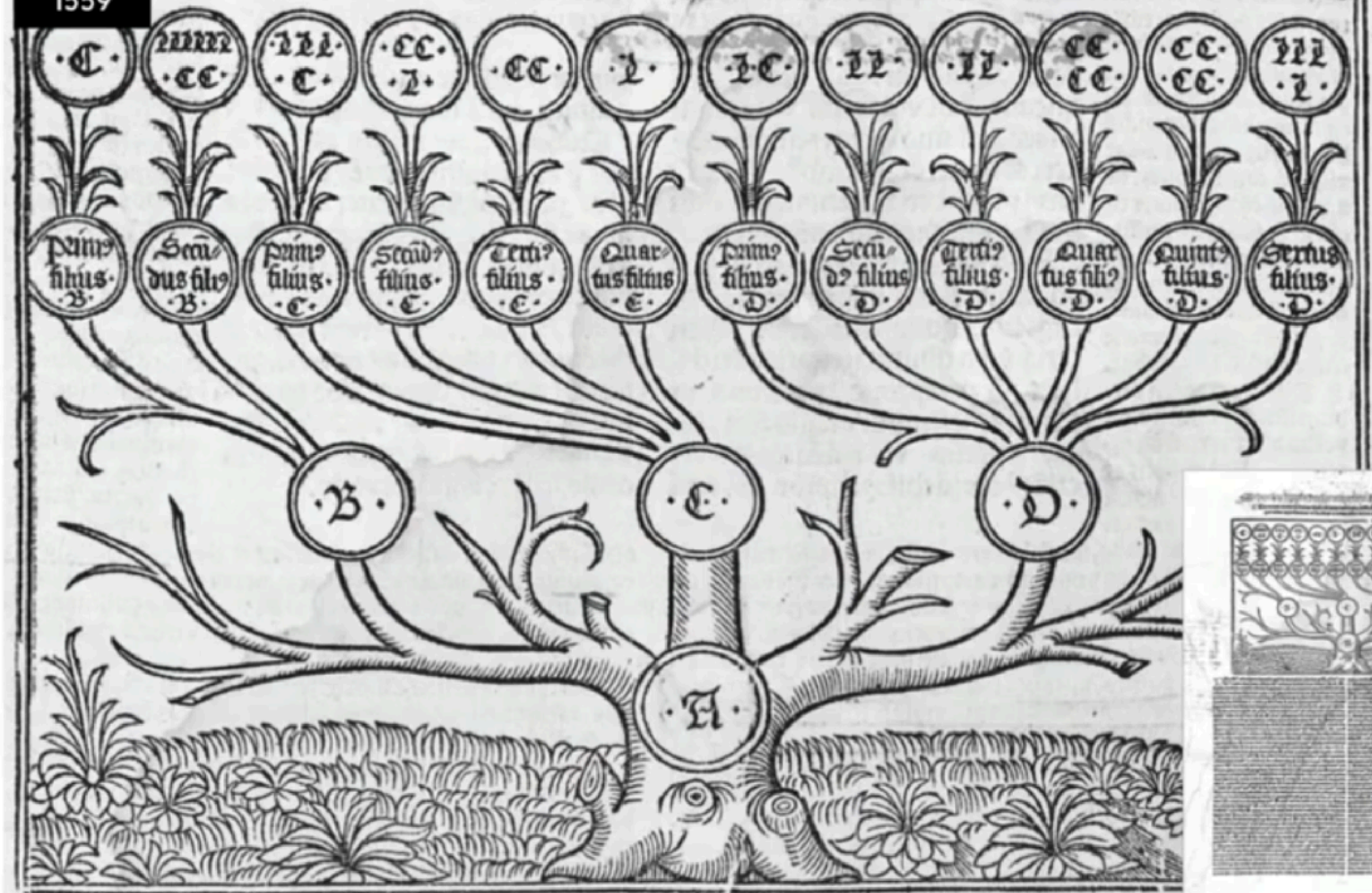
Genealogy (Family trees)

1550

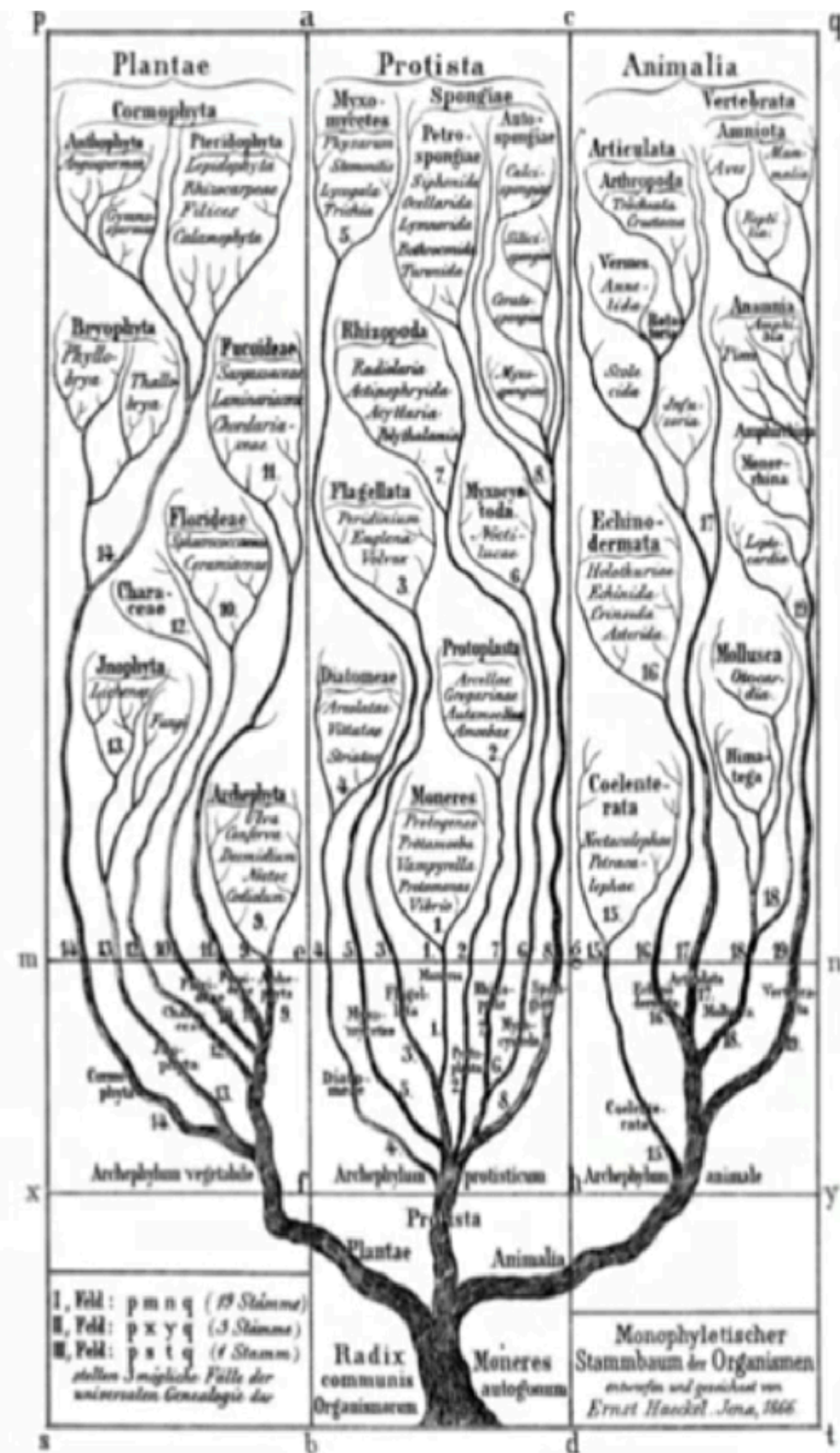
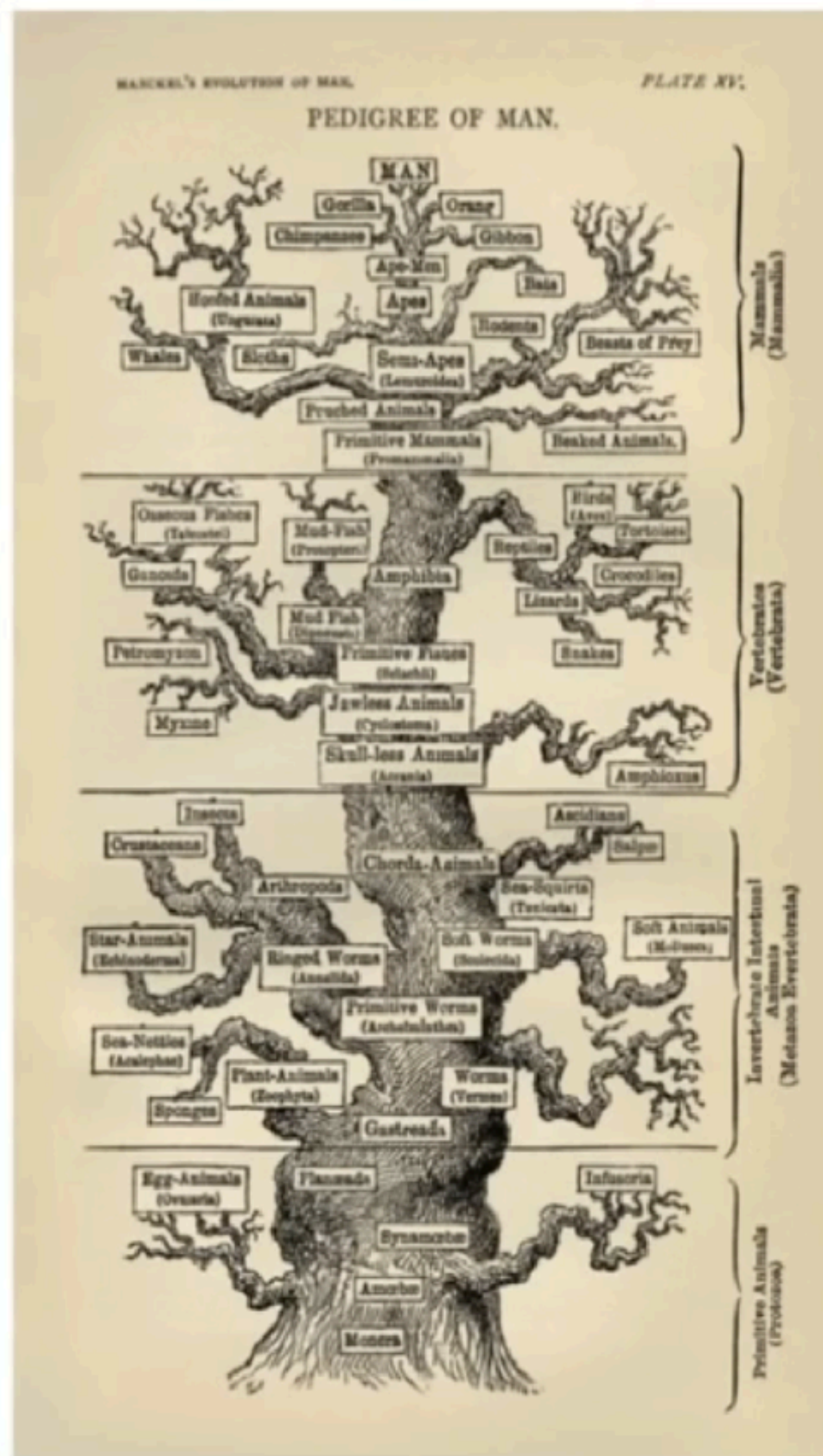
Arbor substitutionum



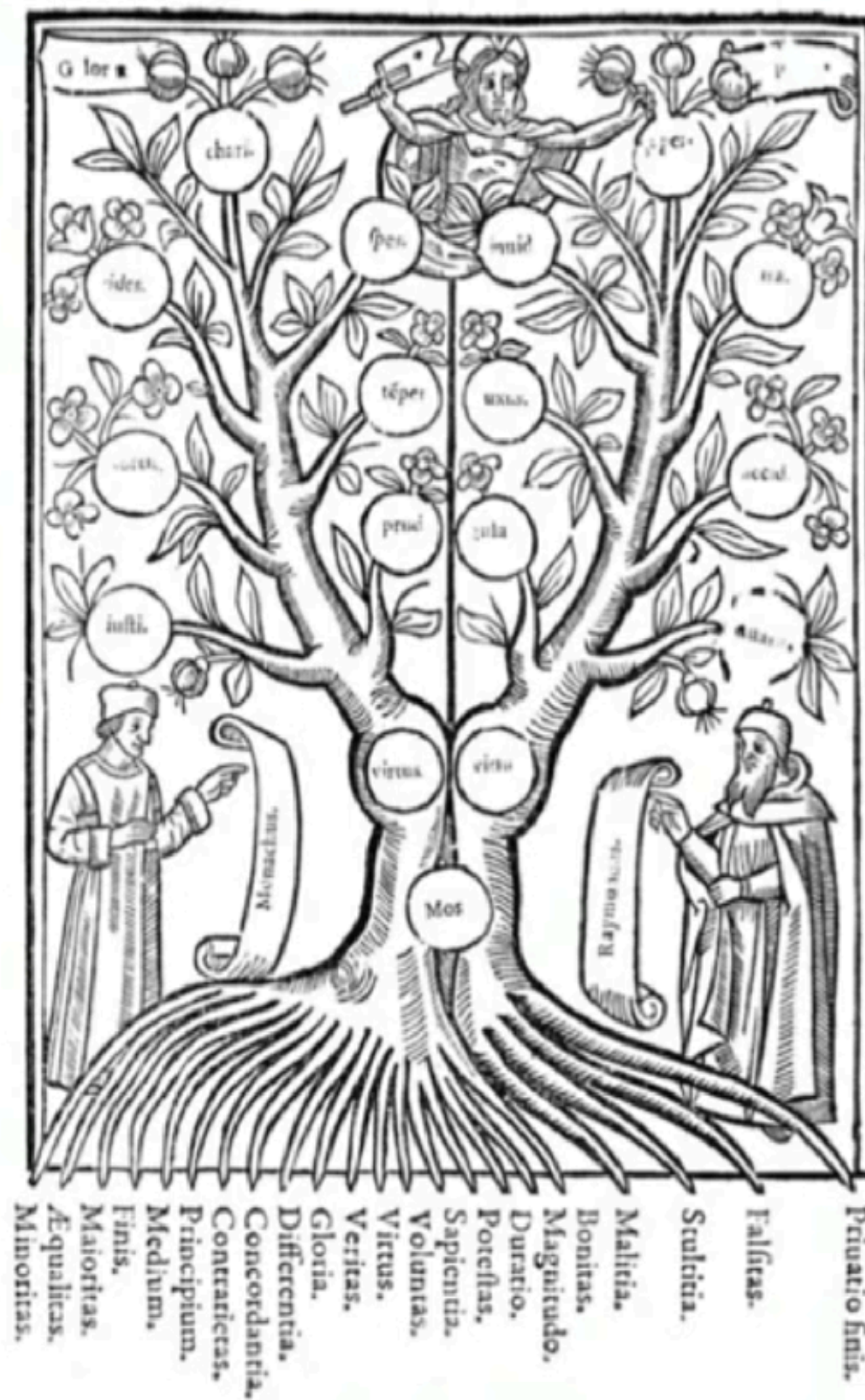
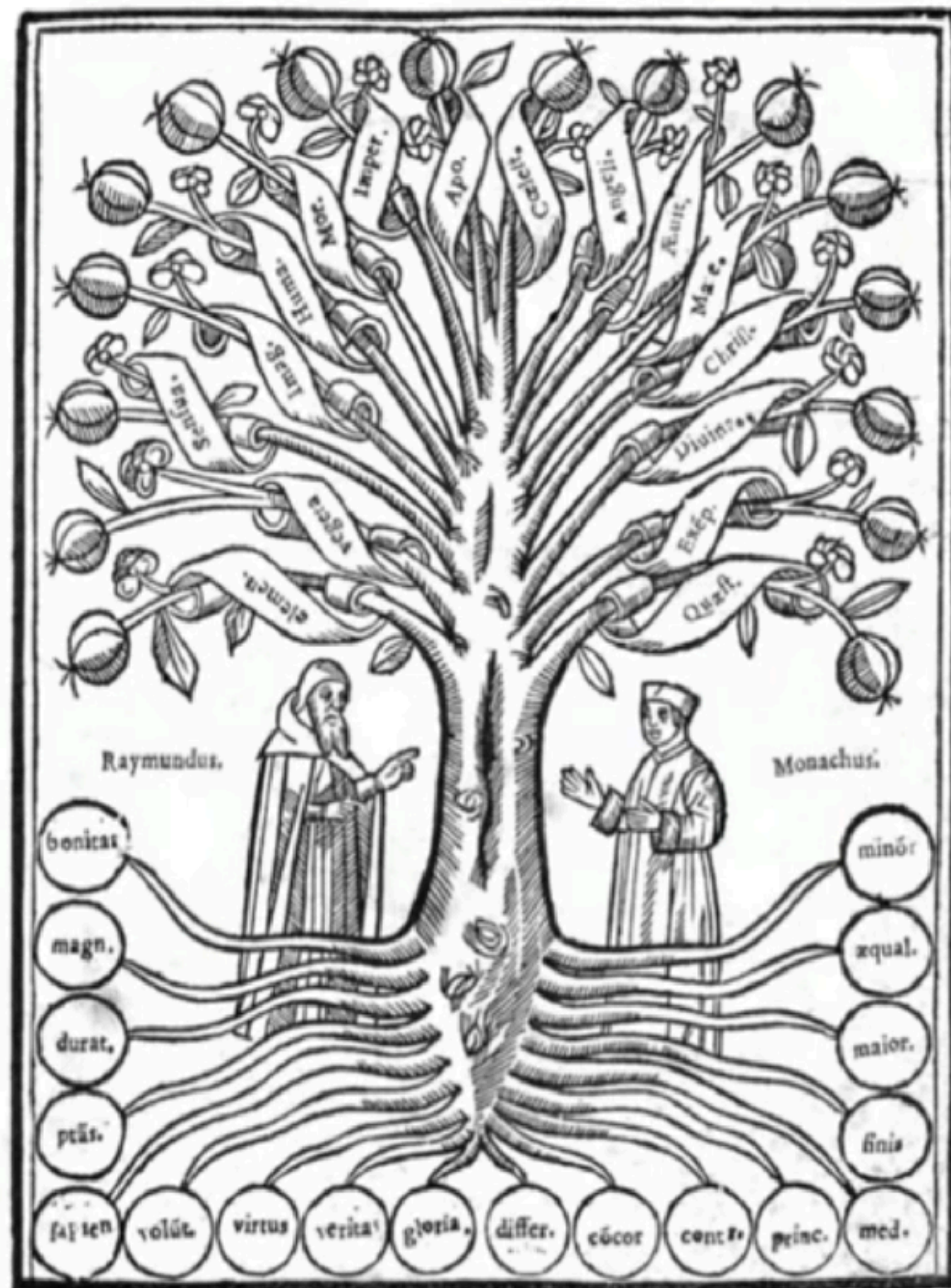
1559



Systems of Law

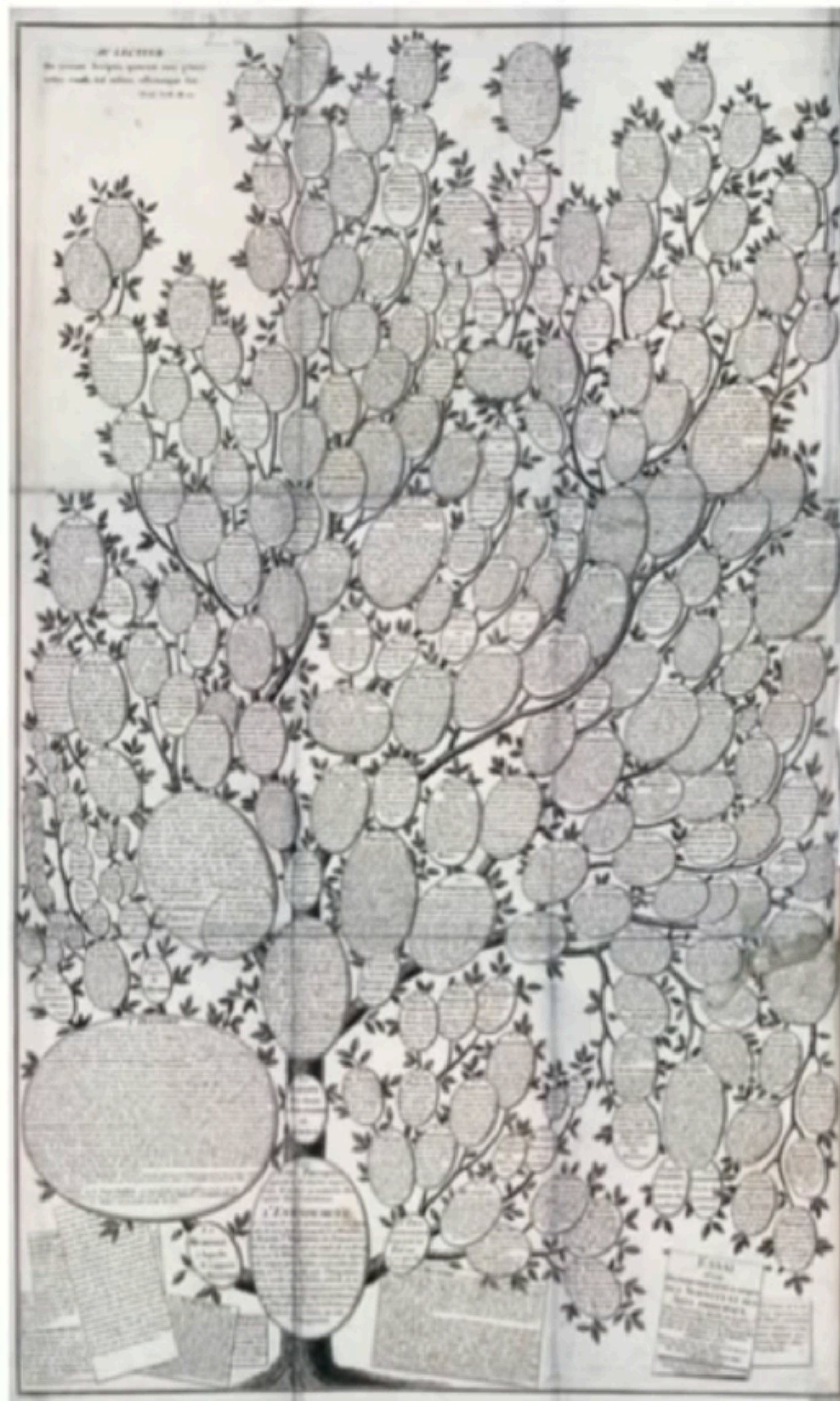


Mapping Species



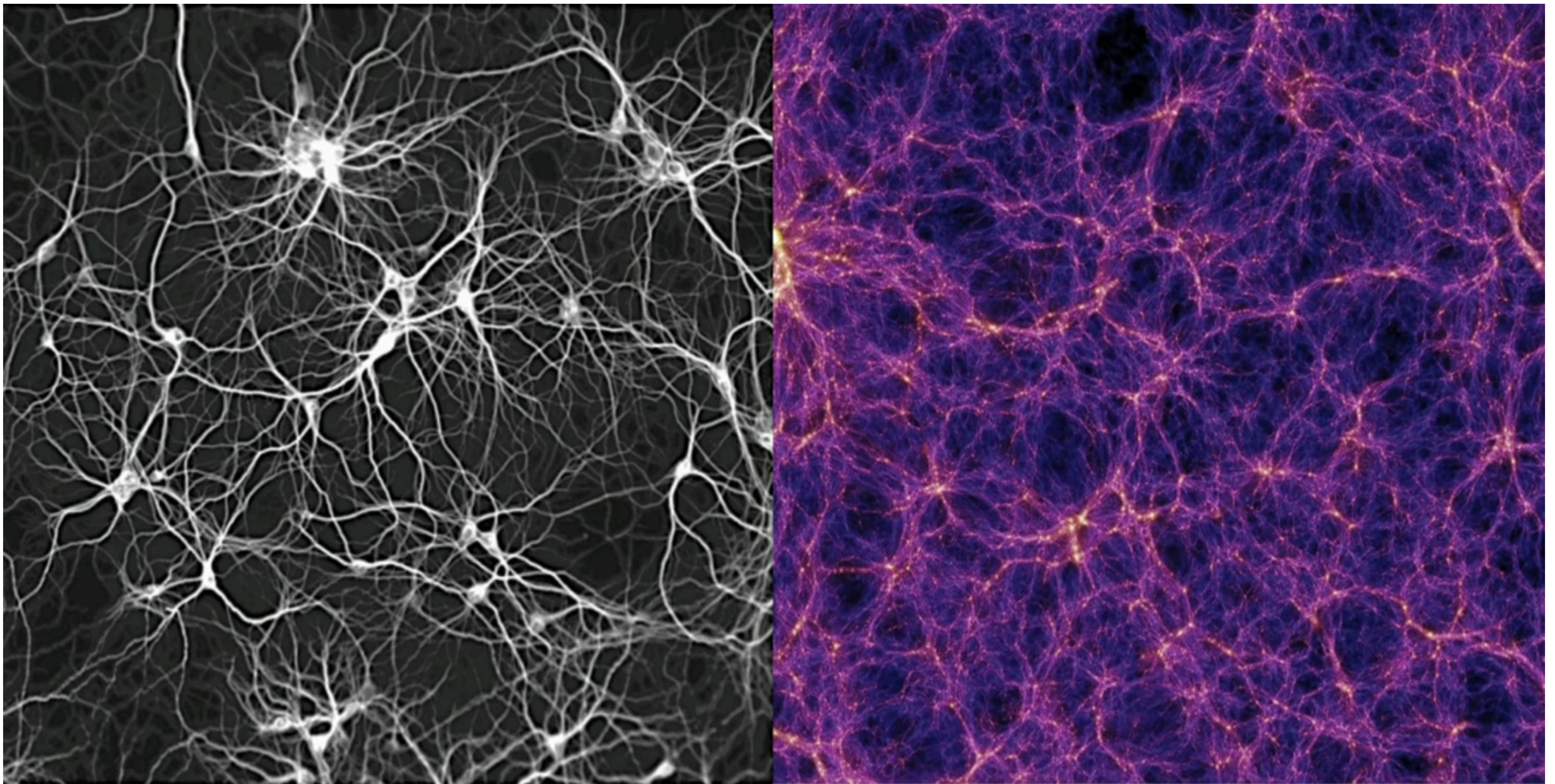
Knowledge

Tree of Science (1296)
Ramon Llull



Knowledge

Encyclopédie (1751)
Diderot and D'Alembert

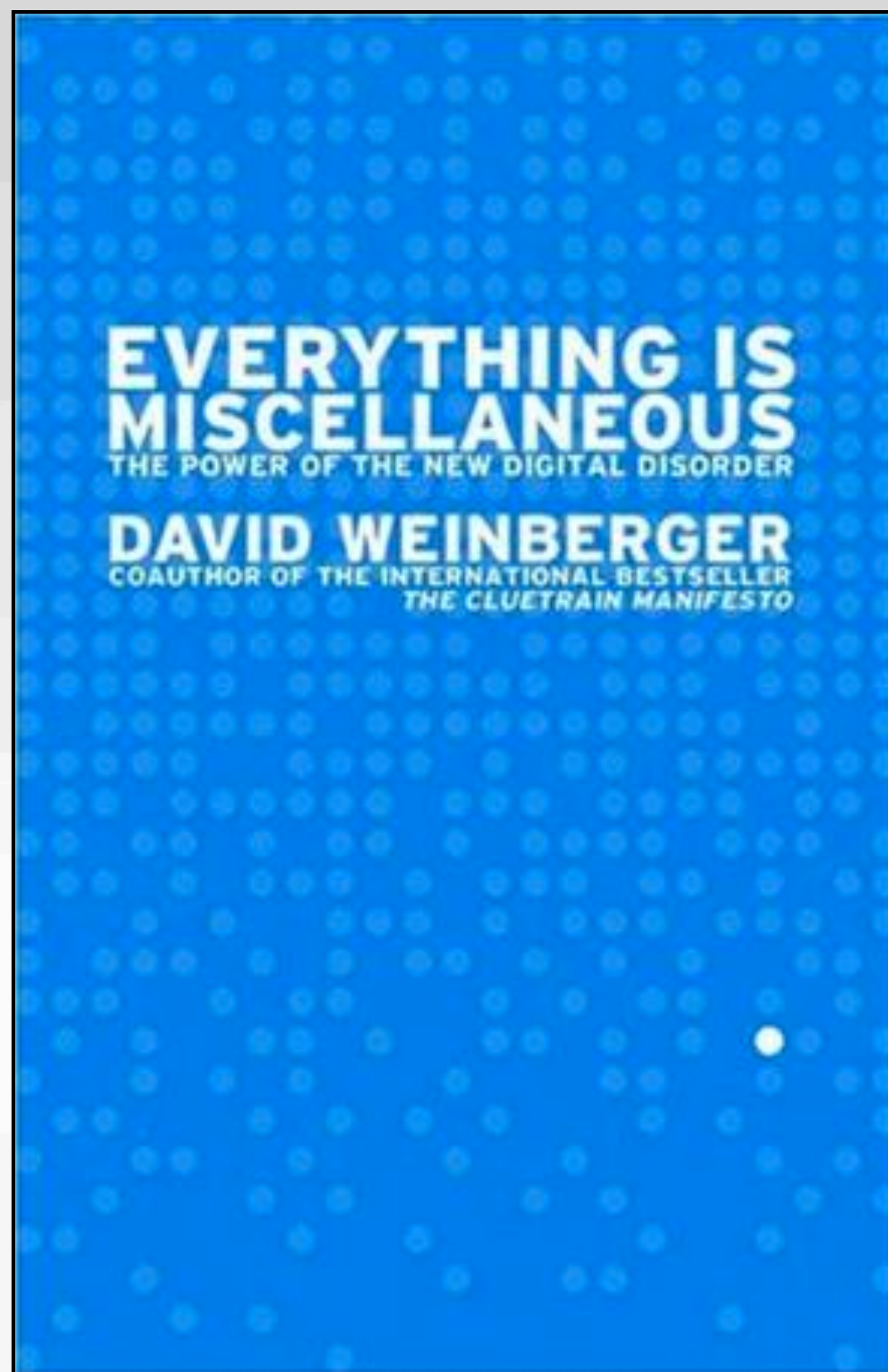


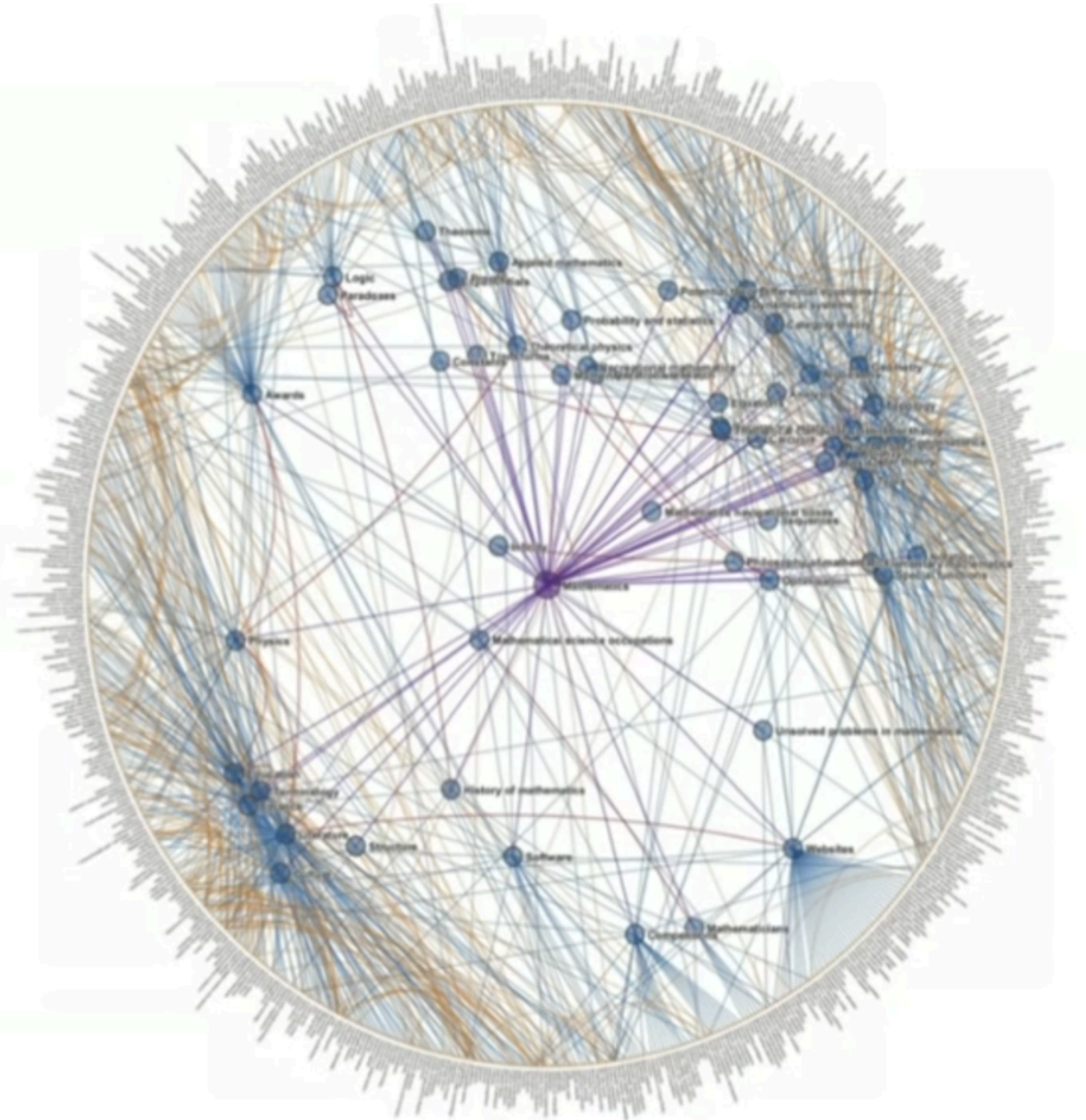
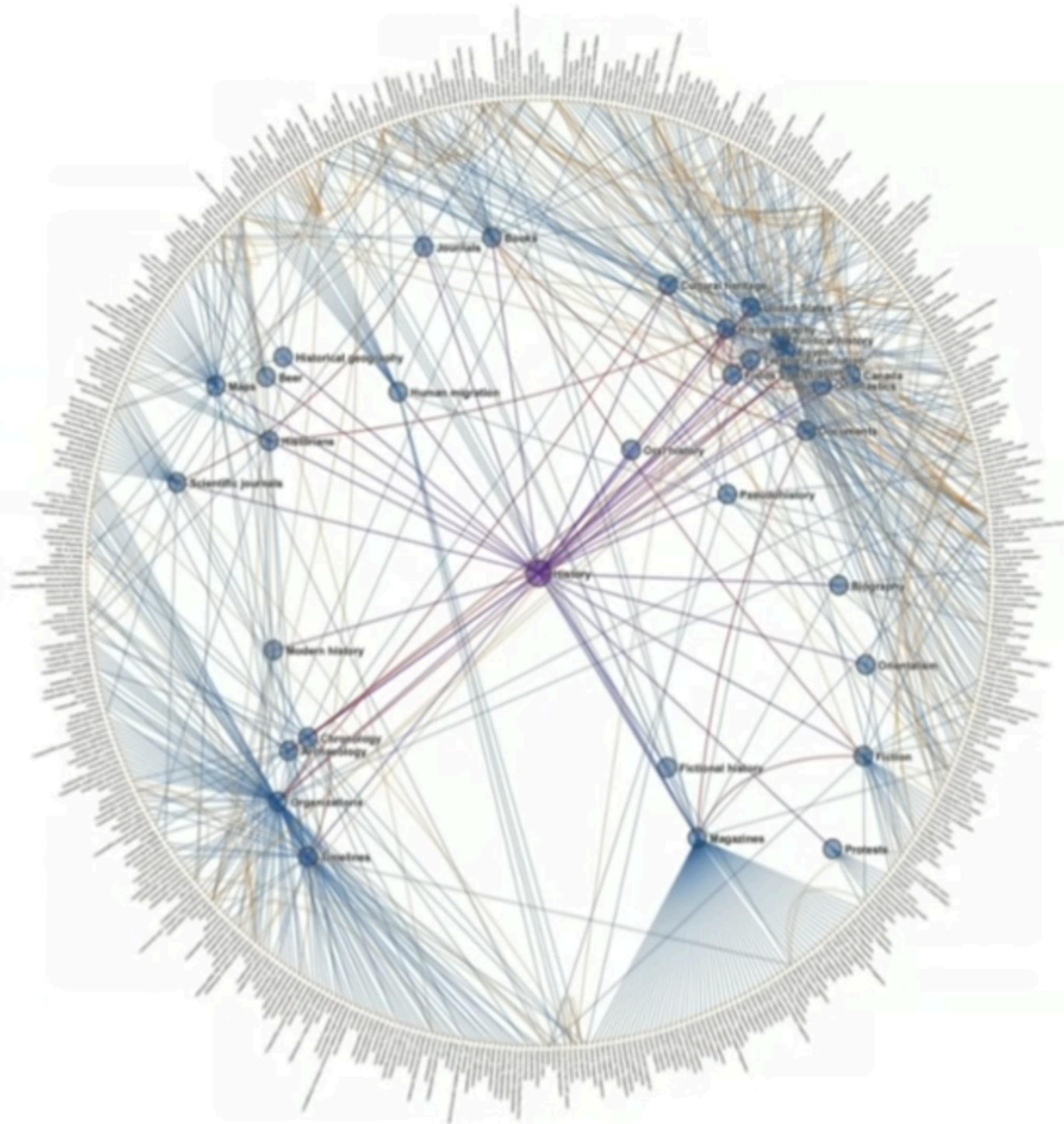
Brain

Mouse's Neuronal Network
Paul de Koninck, Laval university

Universe

Millennium Simulation
Max Planck Society

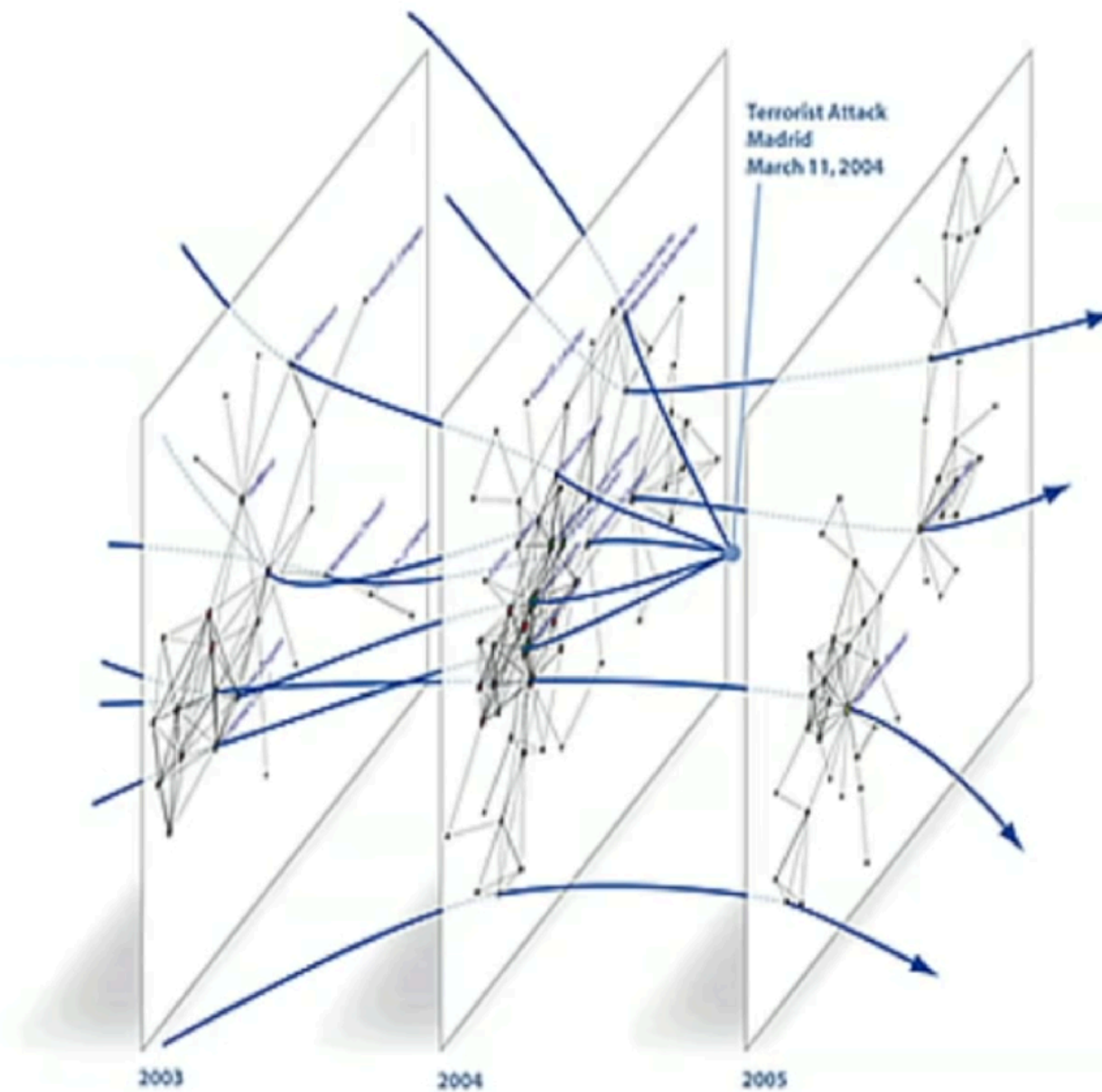




Knowledge

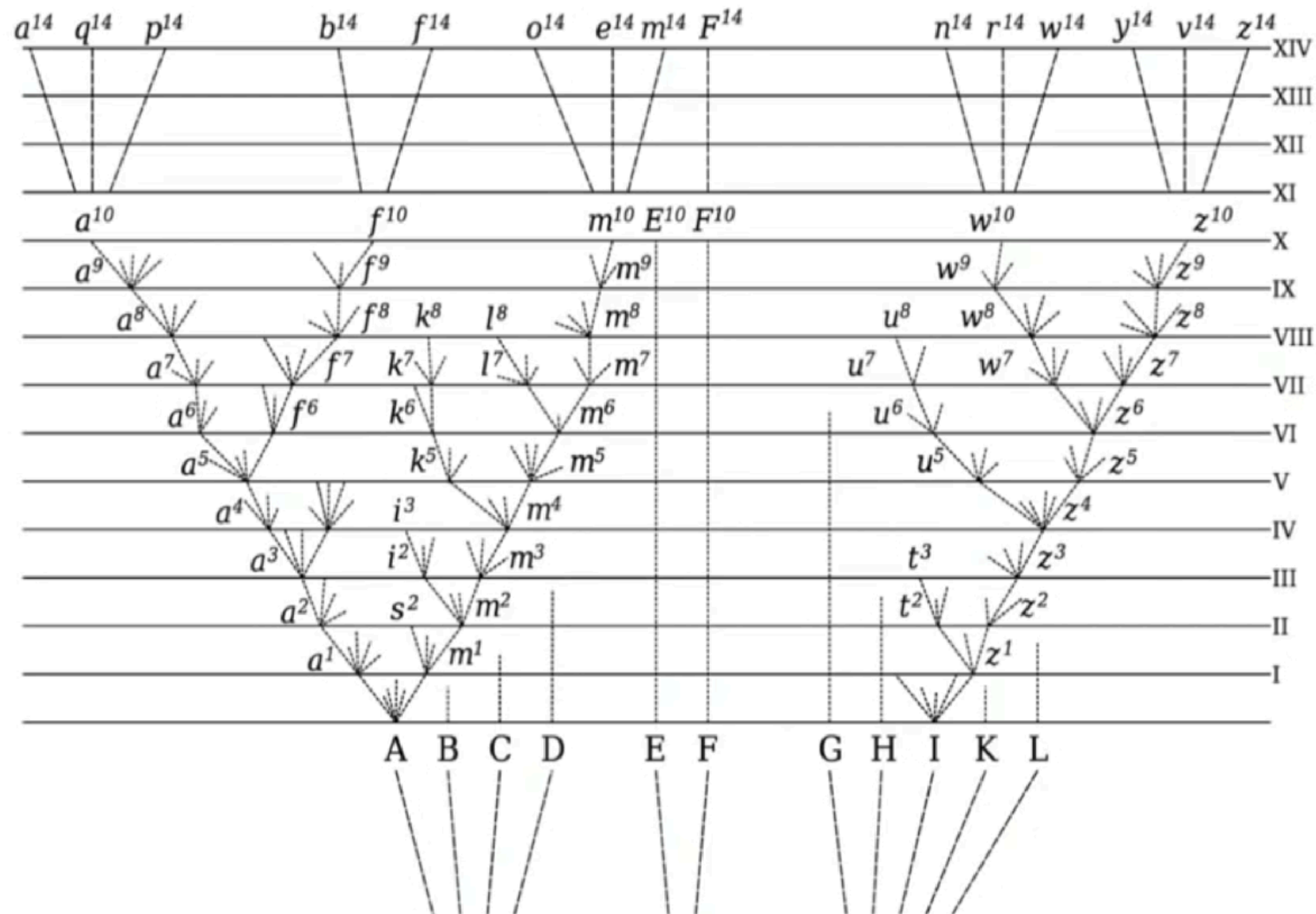
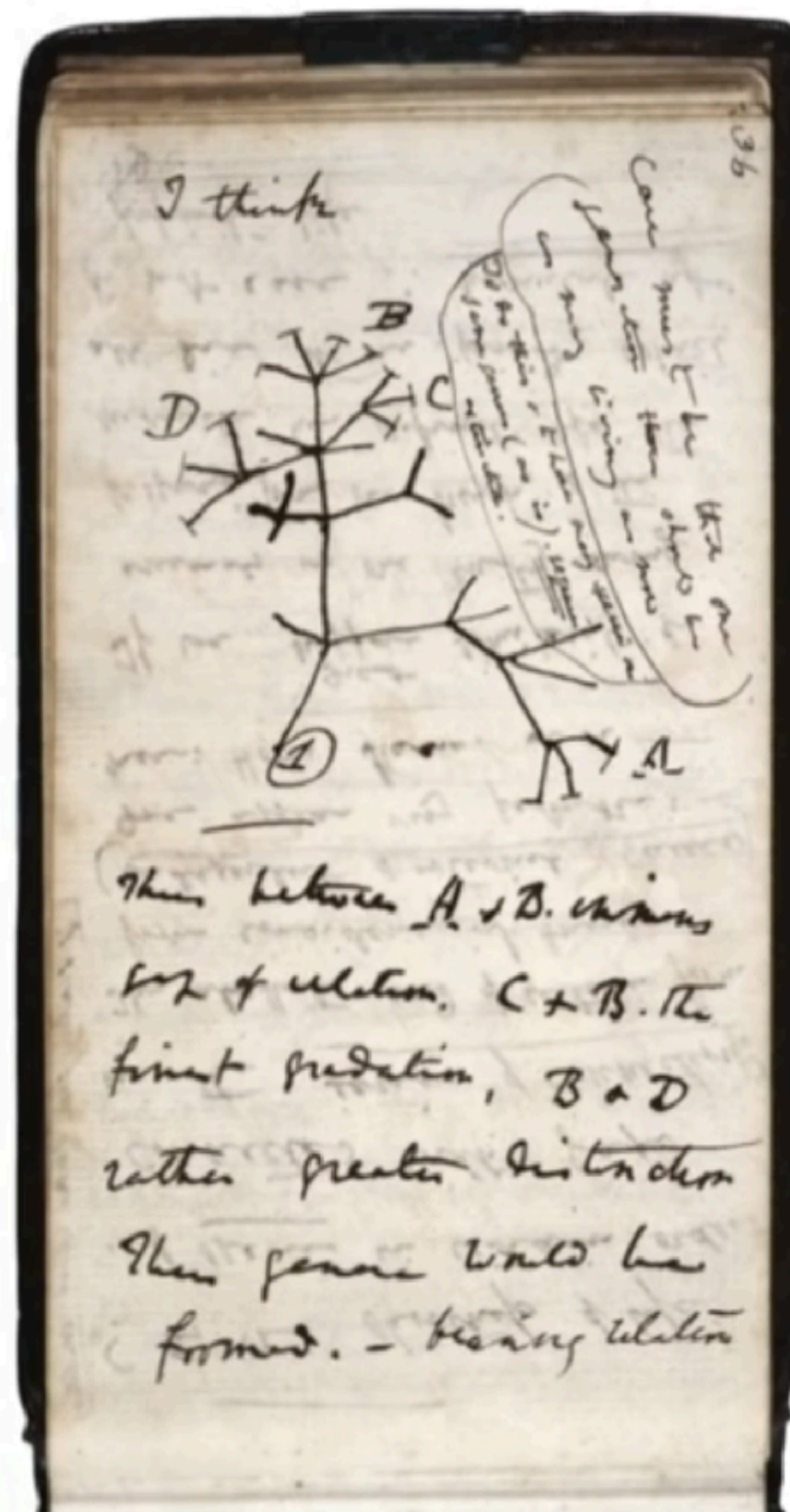
ClusterBall (2007)
Chris Harrison

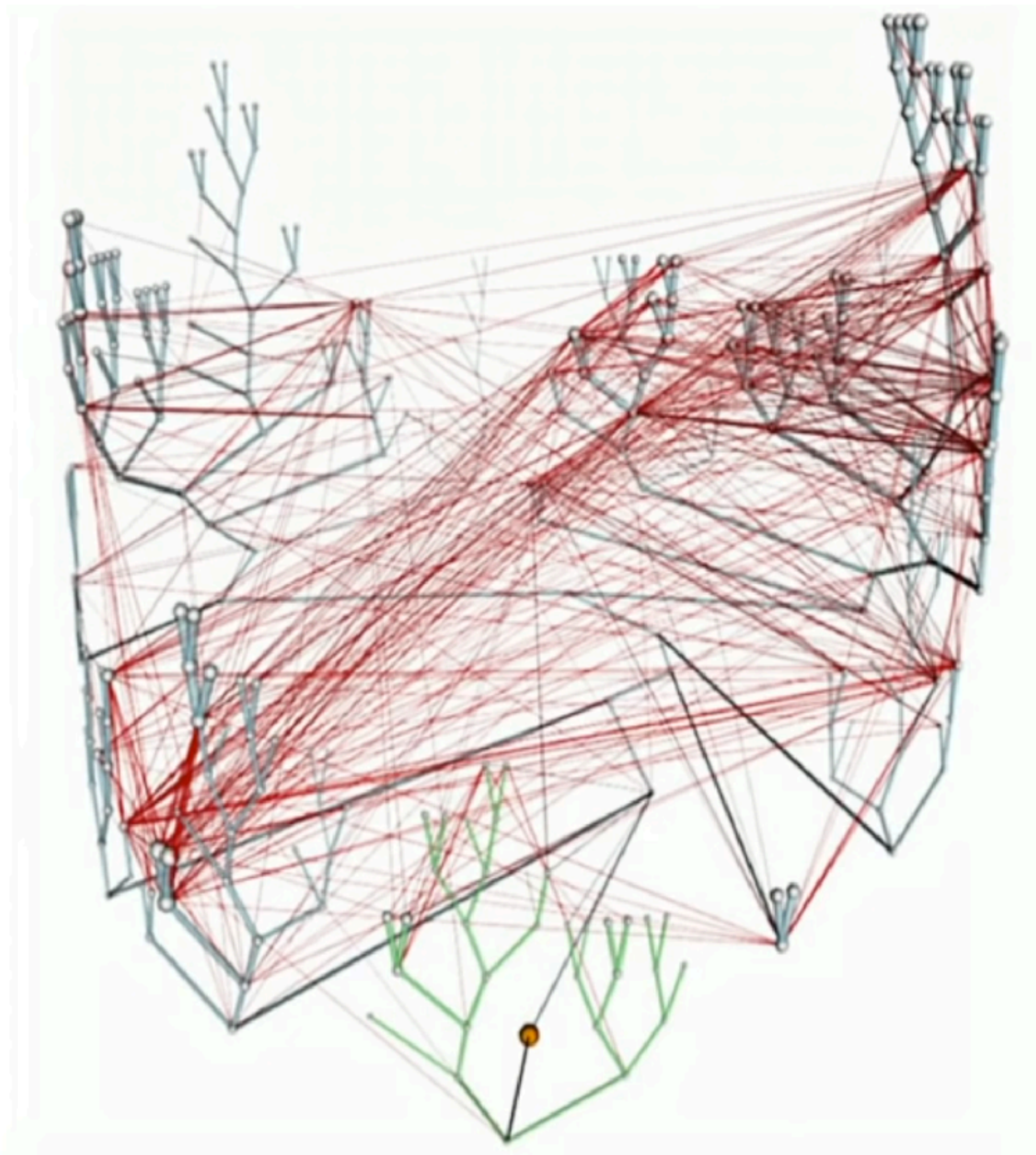




Social Ties

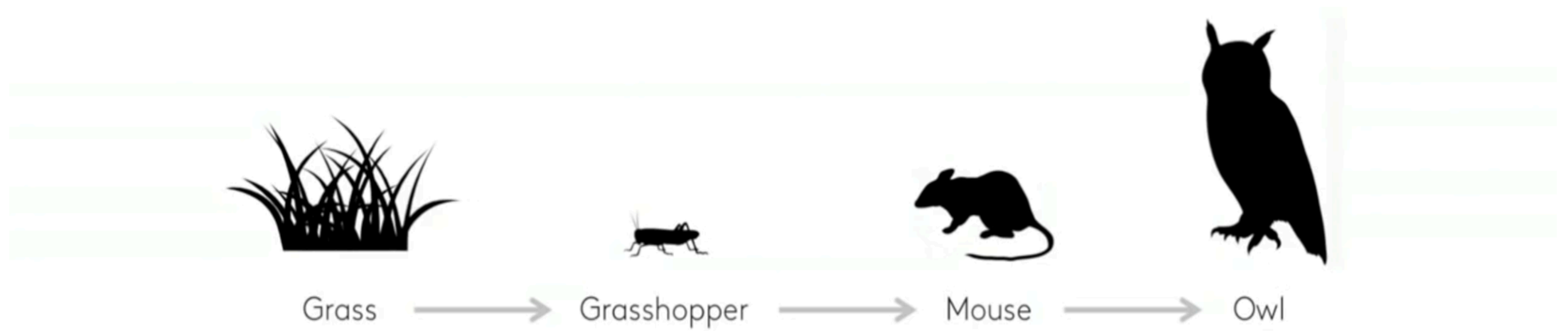
Madrid 2004 Attack – Terrorist Network
ATA - Advanced Technology Assessment



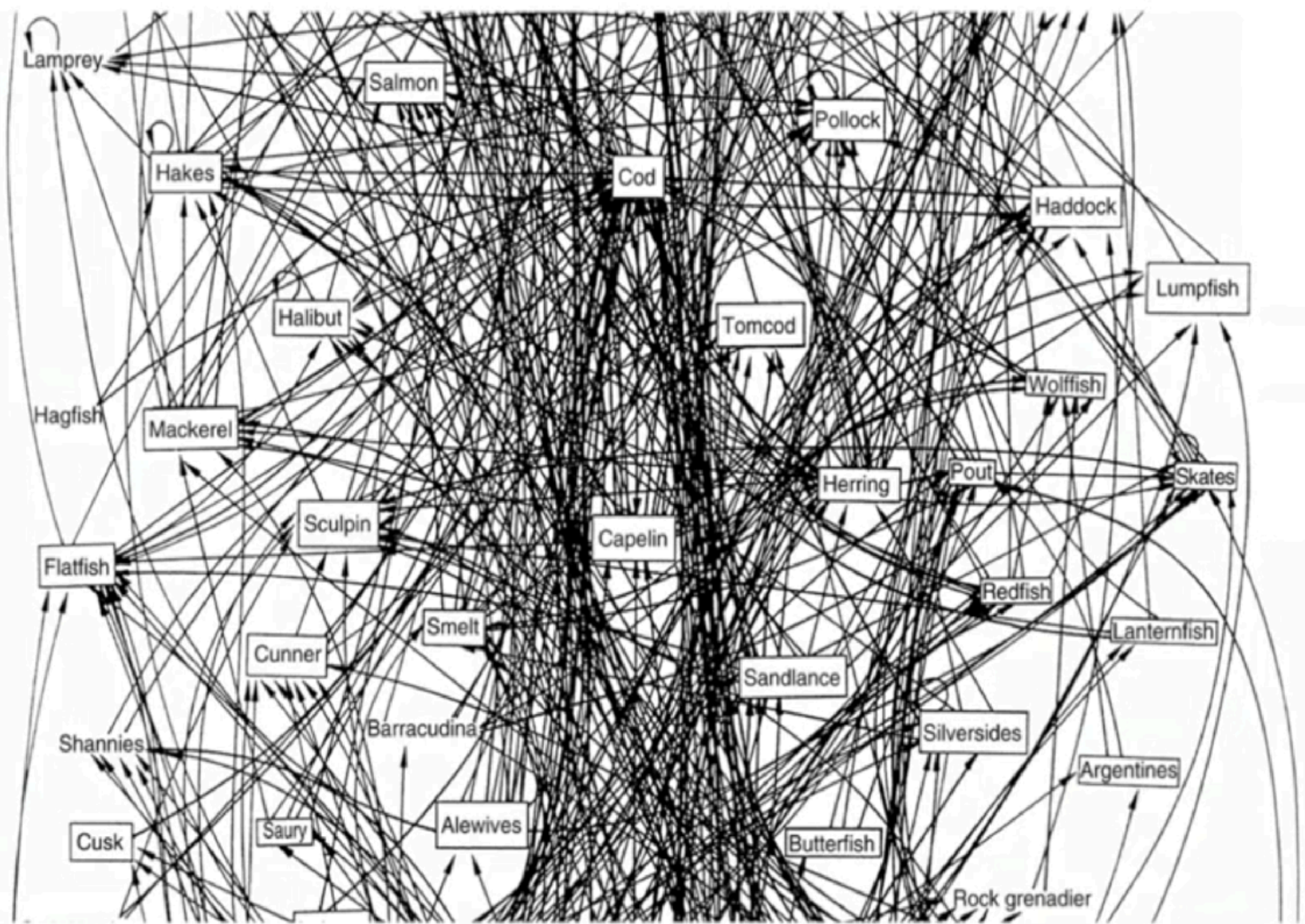
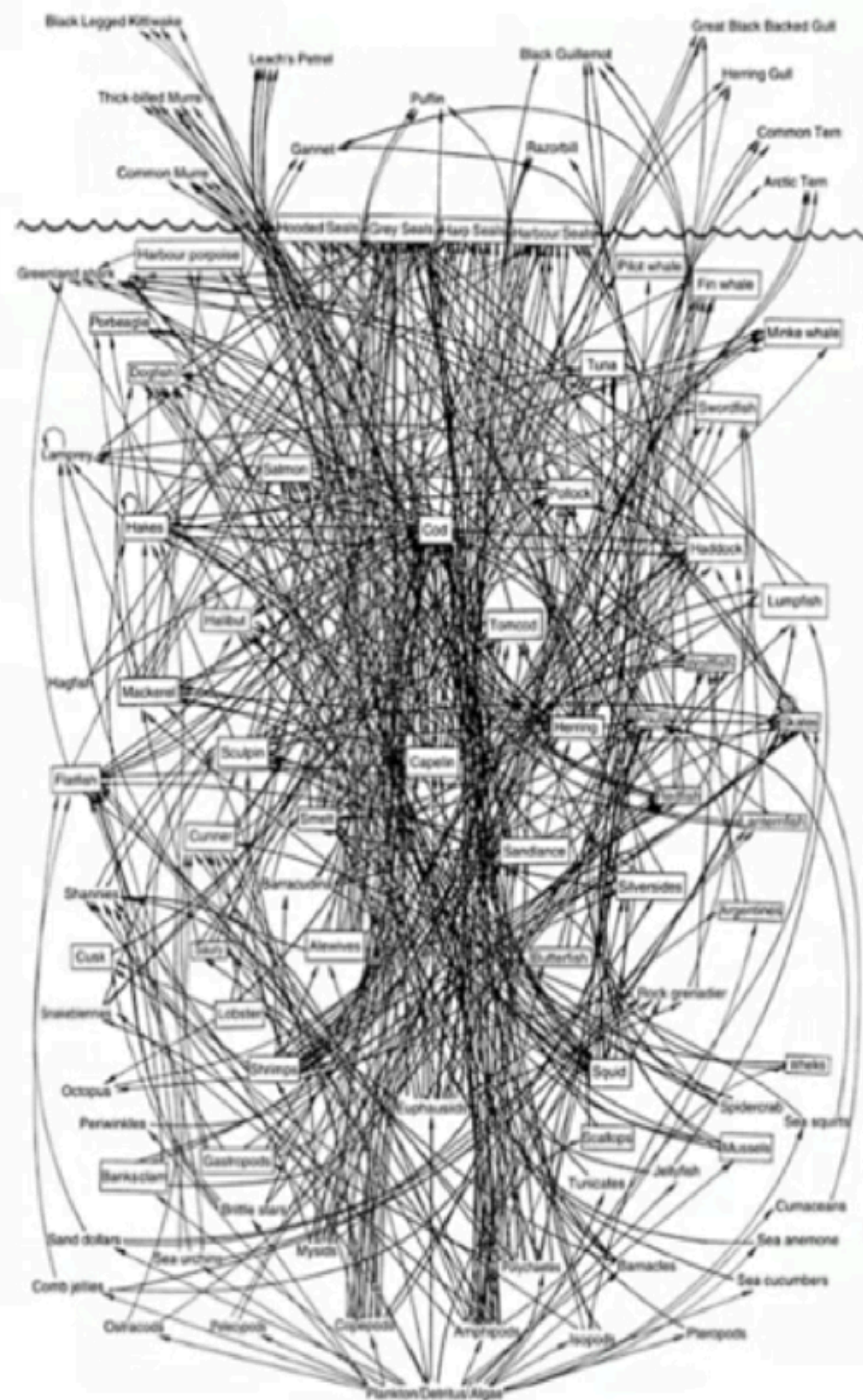


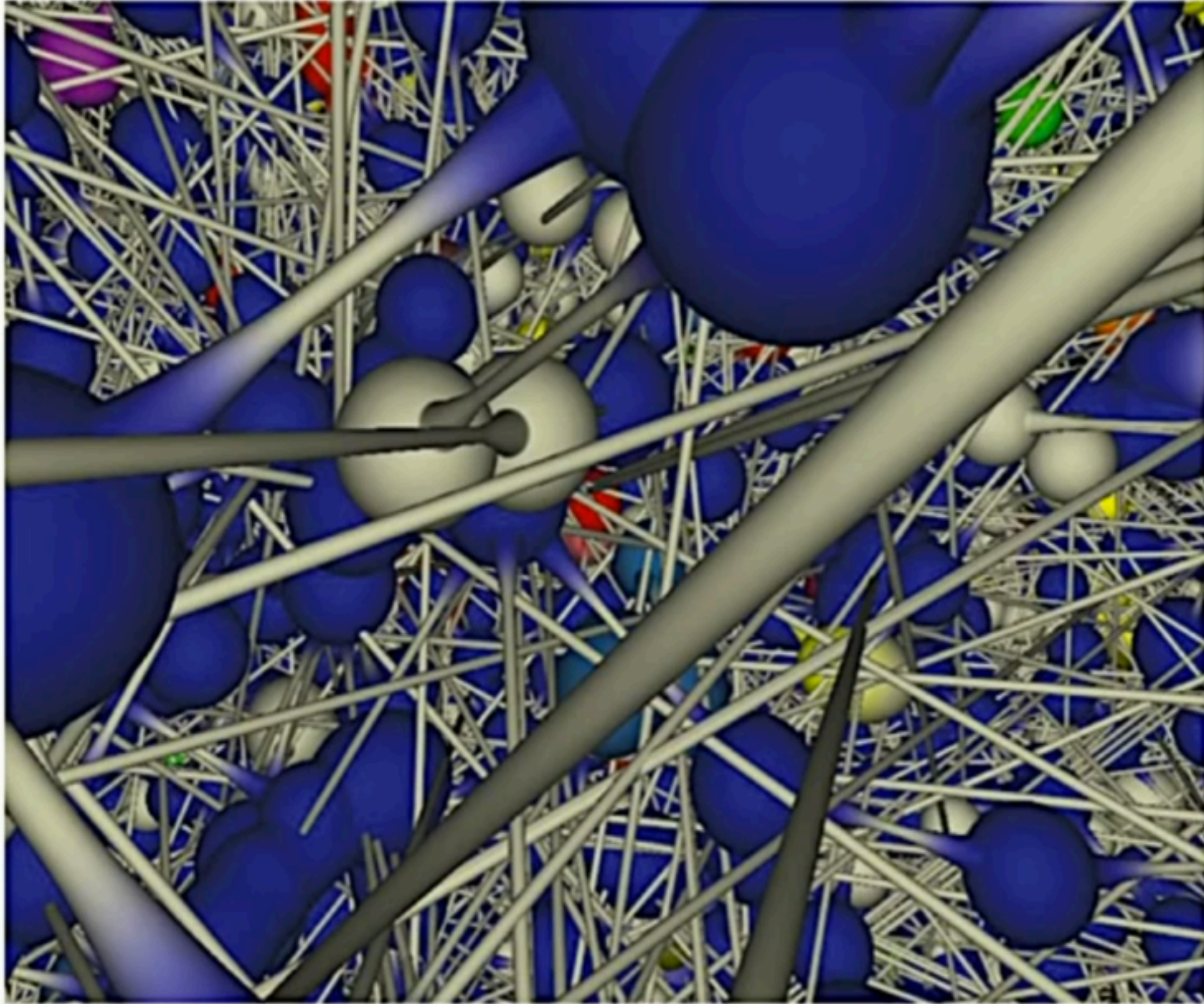
Species

The Net of Life (2005)
V. Kunin, L. Goldovsky, N. Darzentas, and C. A. Ouzounis



Ecosystems





IP Mapping (2001)

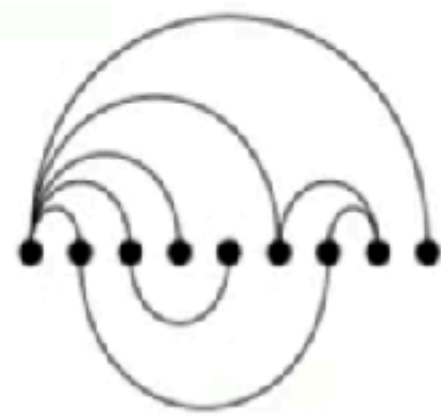
Stephen Coast



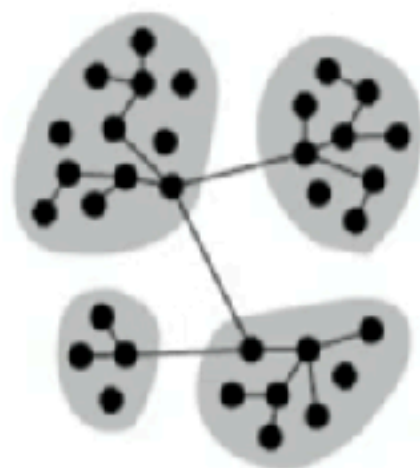
Transient Structures and Unstable Networks (2008)

Sharon Molloy
(oil and enamel on canvas)

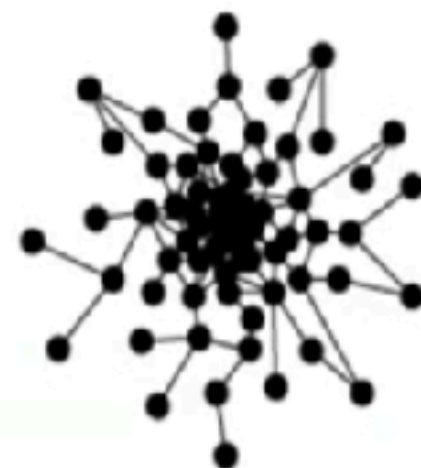
任何秩序都是片面的



Arc Diagram



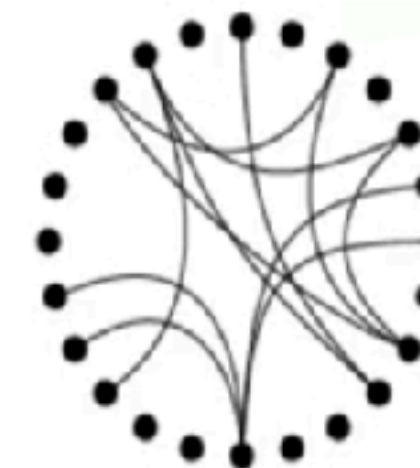
Area Grouping



Centralized Burst



Ramifications



Radial Convergence



Centralized Ring



Circled Globe



Circular Ties



Spheres



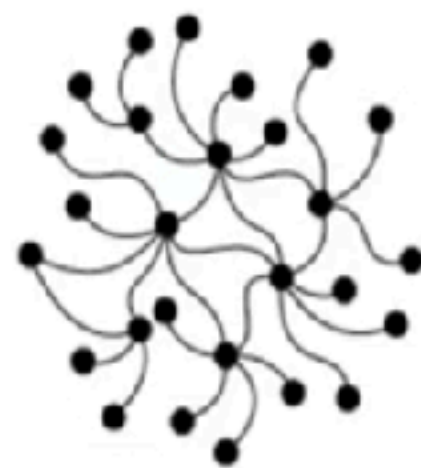
Scaling Circles



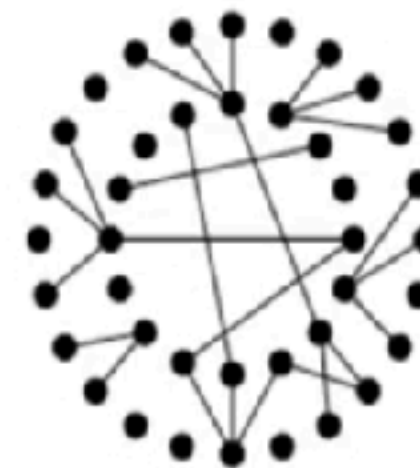
Elliptical Implosion



Flow Chart



Organic Rhizome



Segmented Radial Convergence



Radial Implosion

结论：只有回归本源才是最终秩序

Nonlinearity

Decentralization

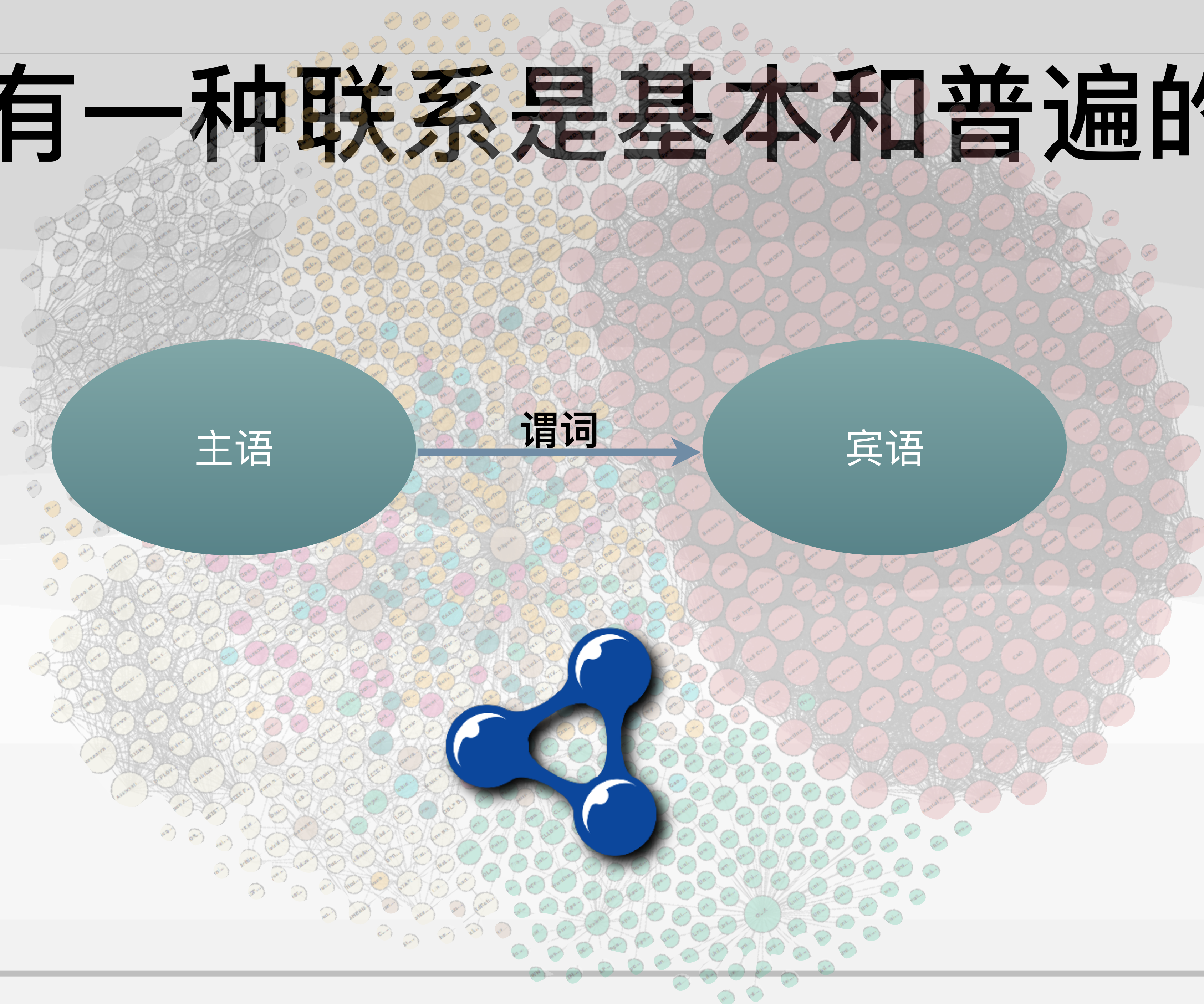
Interconnectedness

Interdependence

Multiplicity

© Chris Harrison

只有一种联系是基本和普遍的



众生平等

“When everything is connected to everything else, for better or for worse, everything matters.”

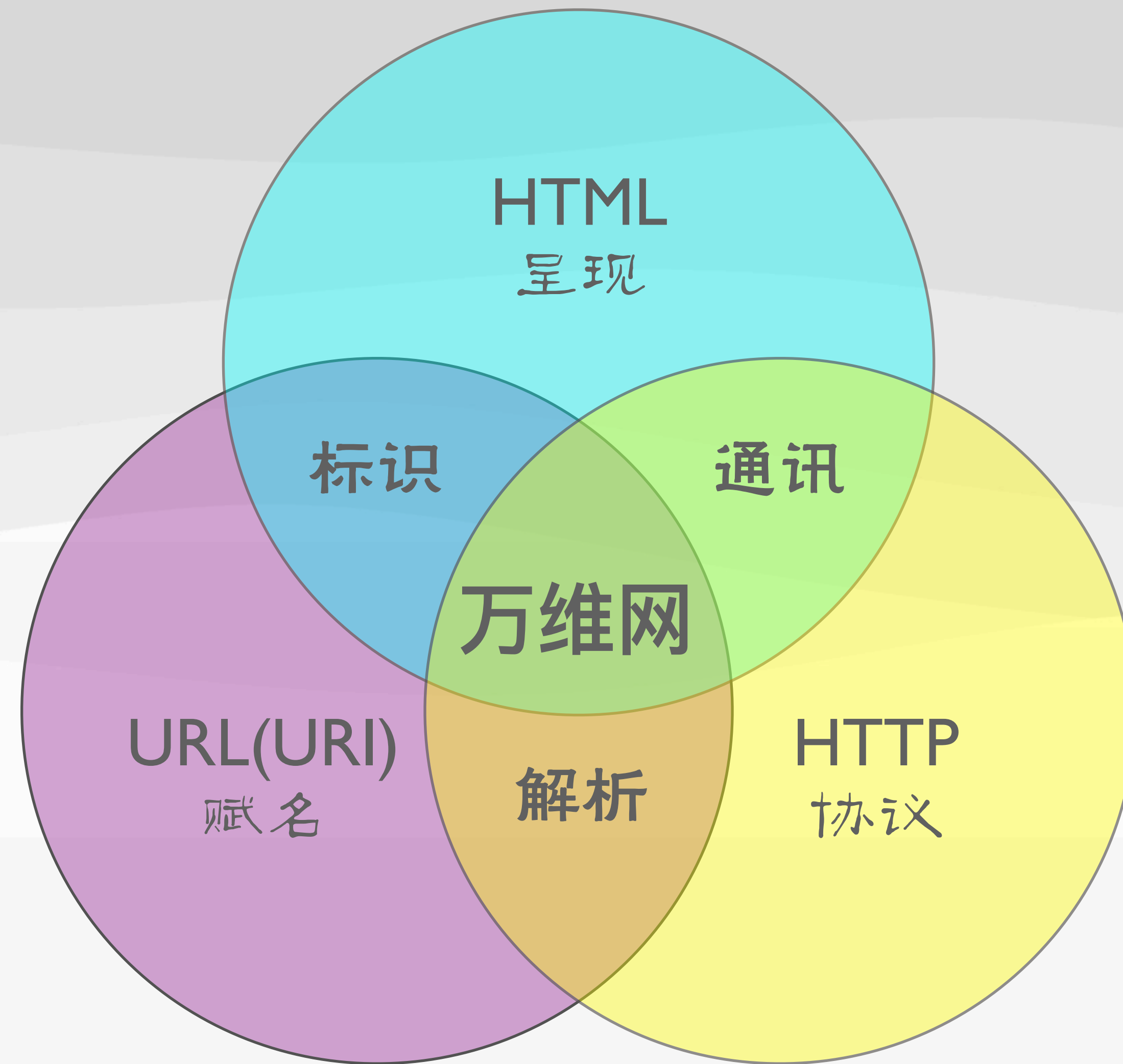
Bruce Mau, Massive Change

只有当计算机强大到今天，才有可能模拟所有秩序。

—— Keven, 2017年11月16日

从关联数据到智慧数据

所谓万维网

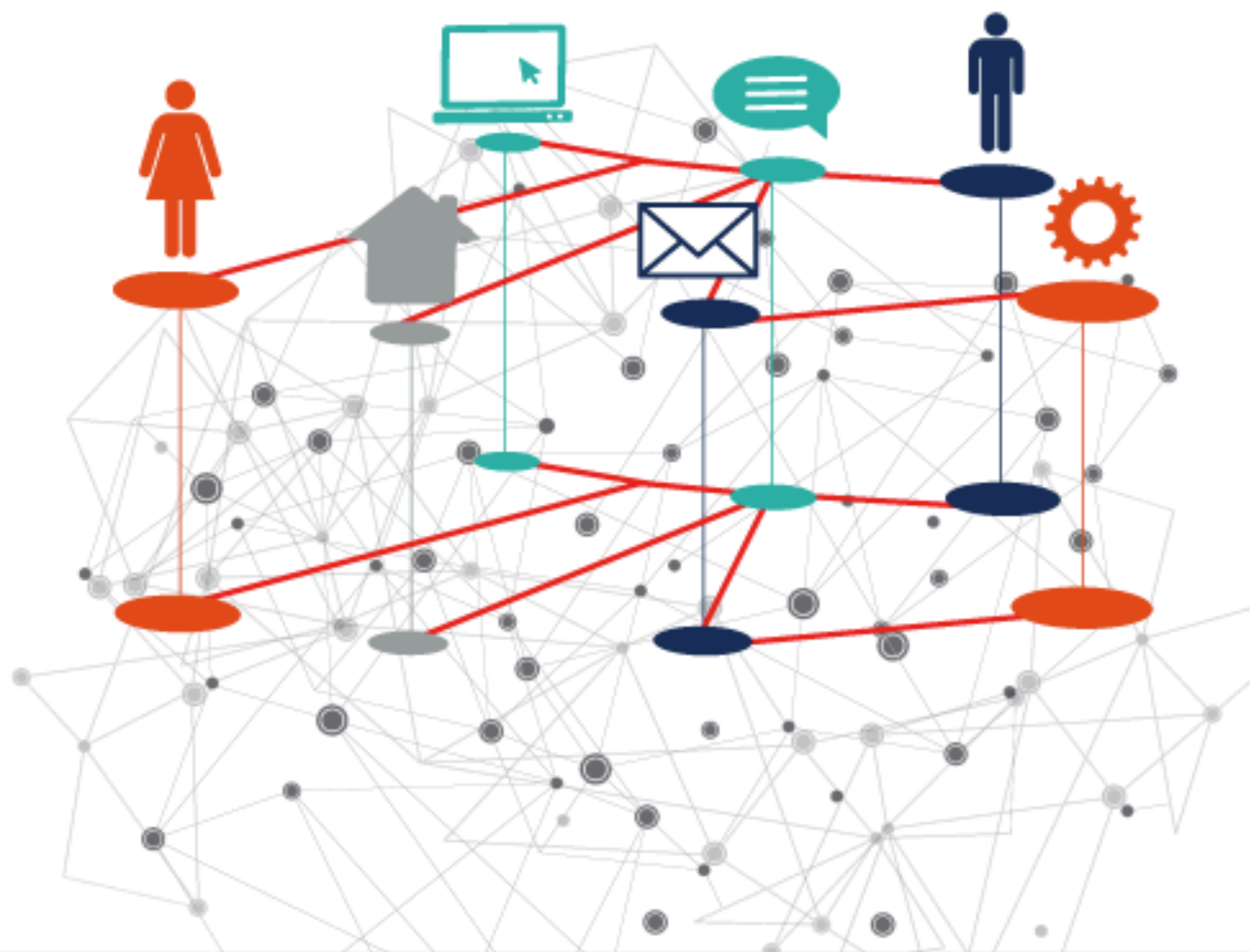


语义万维网：集体大脑1.0

Web



Semantic Web

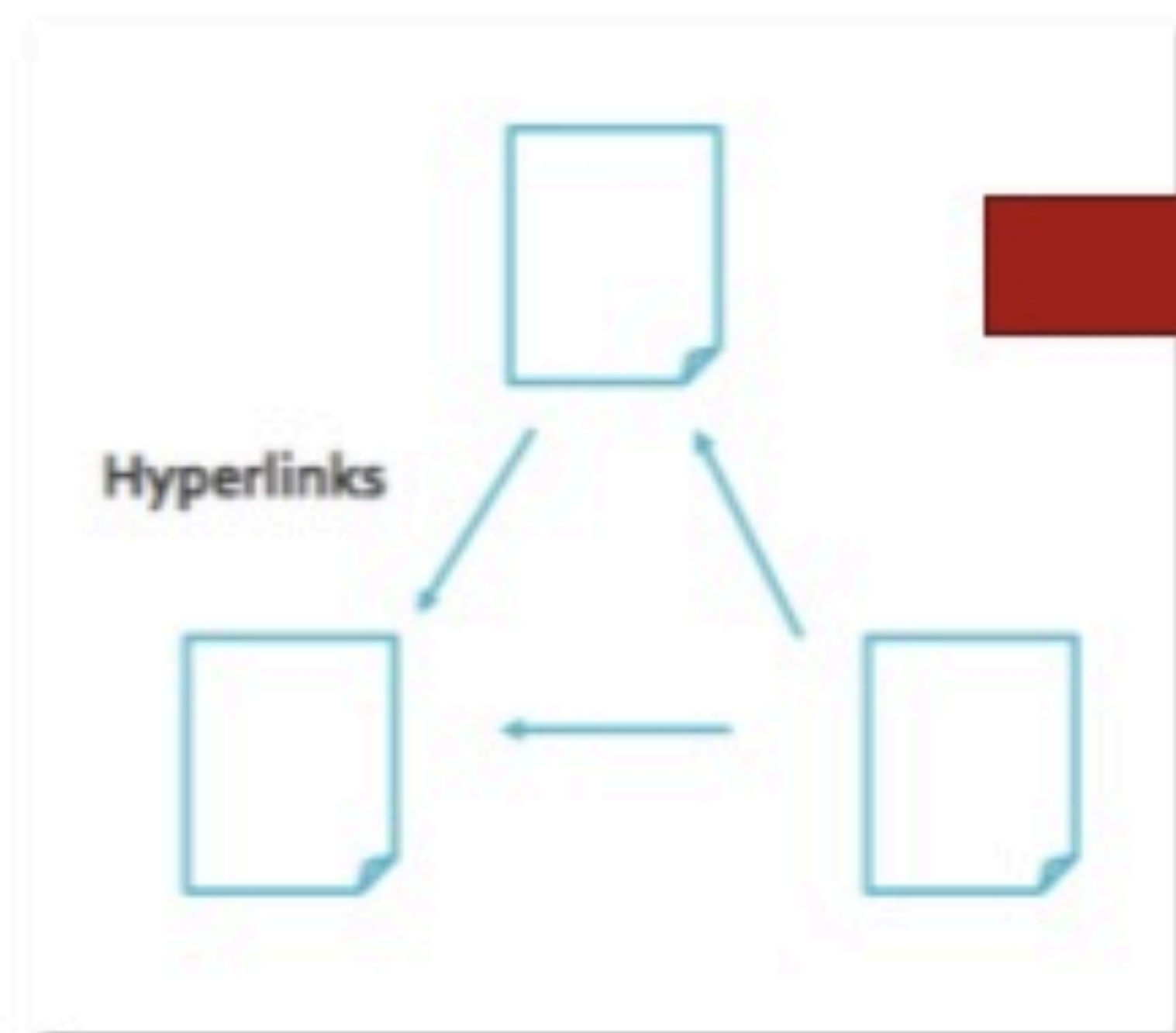


所谓关联数据

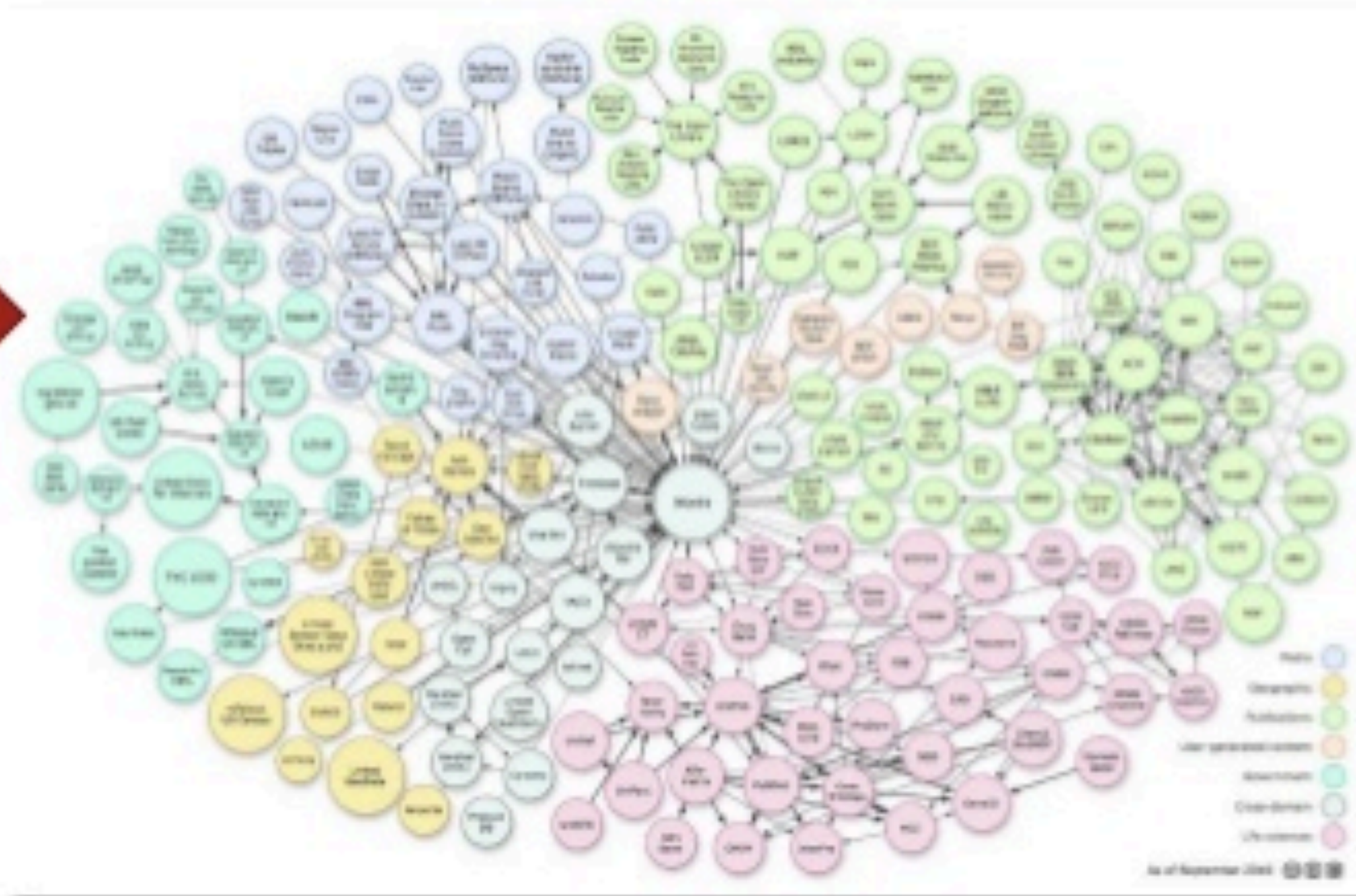
1. 使用URI作为任何事物的标识名称 【天生的规范控制】
Use URIs as names for things
 2. 使用 HTTP URI作为万维网上唯一的名称，使其可以被解引 (dereference)
【天生的面向网络】 Use HTTP URIs so that people can look up those names
 3. 被访问时以 RDF 形式提供有用的信息 【天生的关联关系】
When someone looks up a name, provide useful information
 4. 尽可能提供指向其它 URI 的链接，以使人们发现更多的相关信息 【天生的复杂网络】
Include links to other URIs so that they can discover more things
-

所谓关联数据

Web of documents...

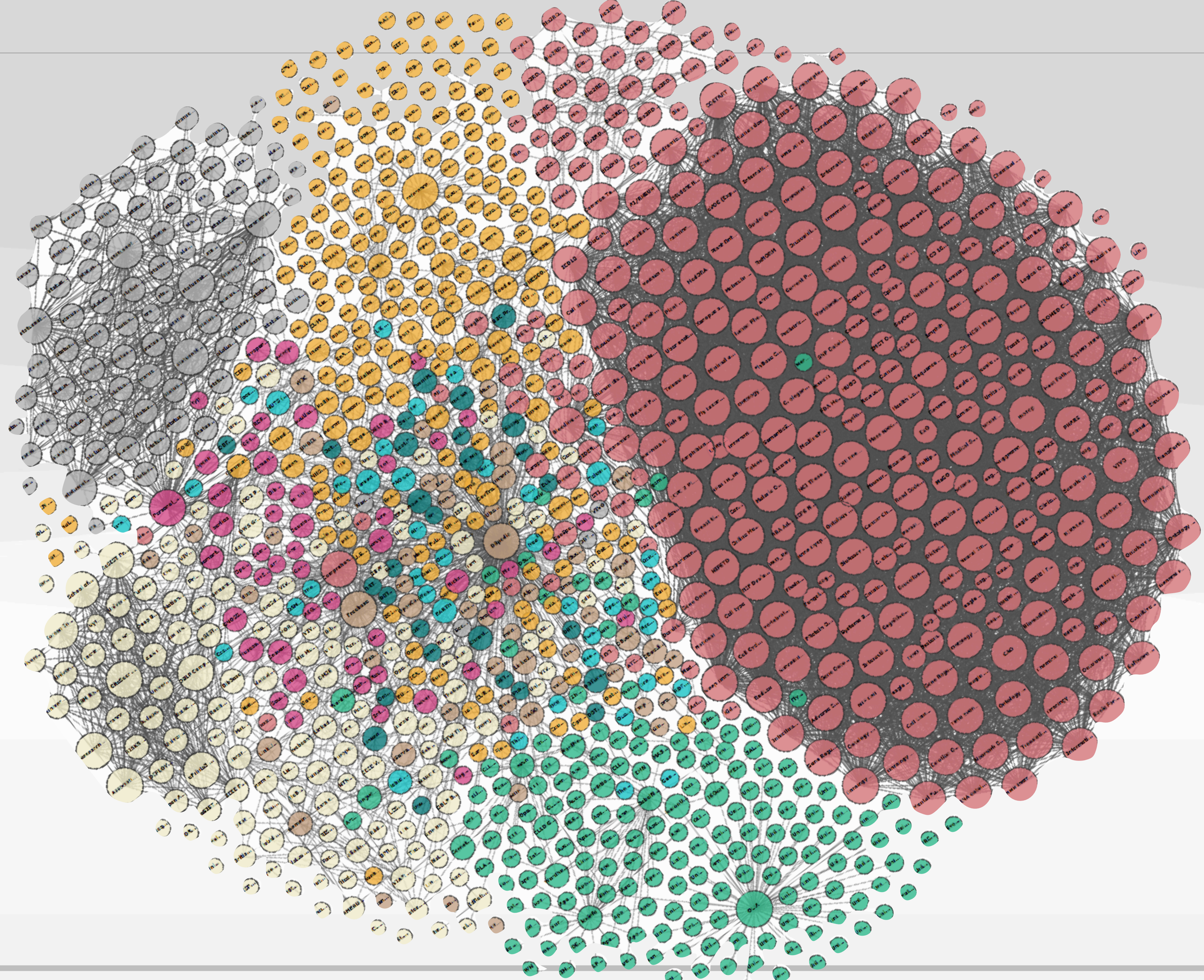


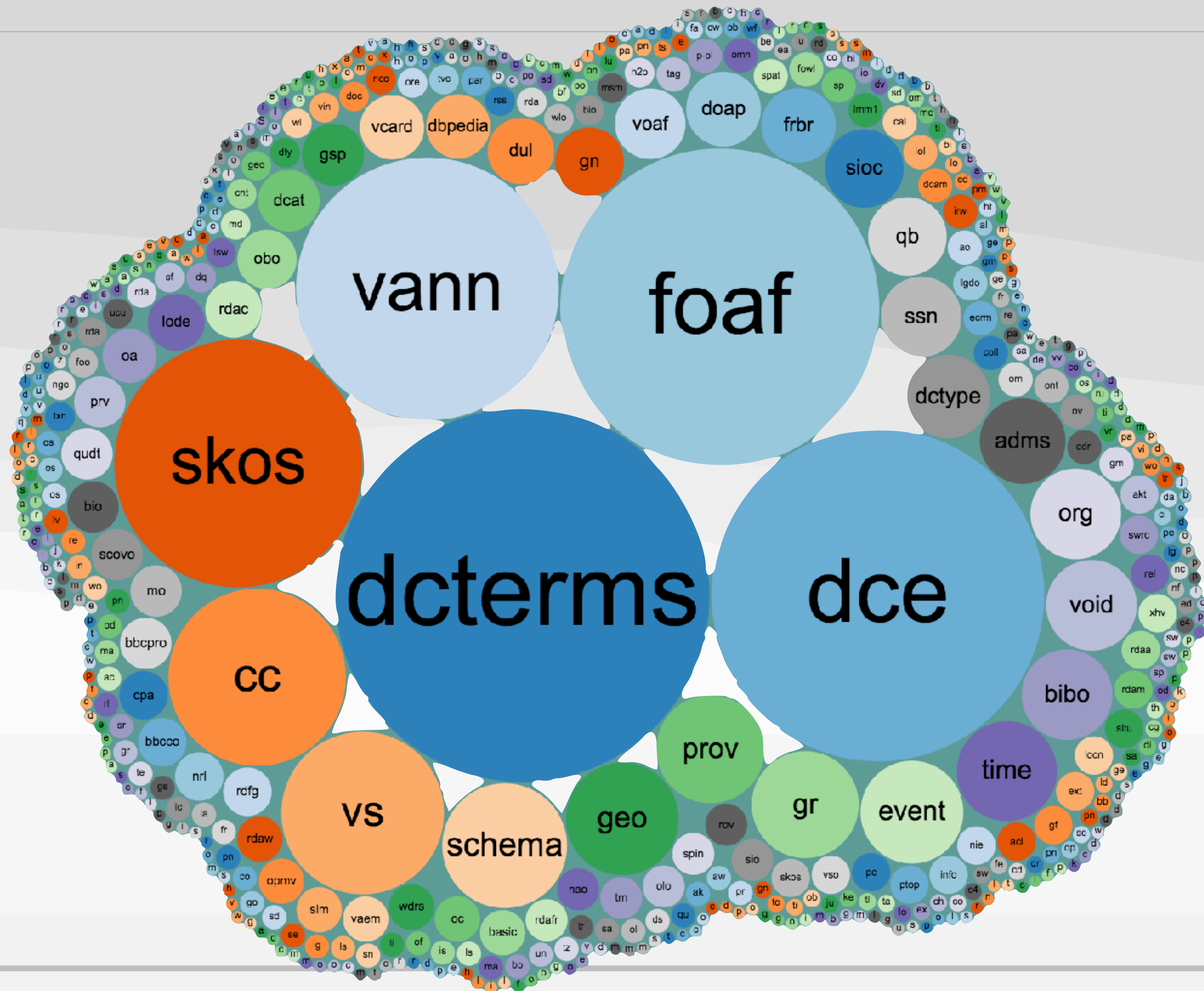
Web of linked data...



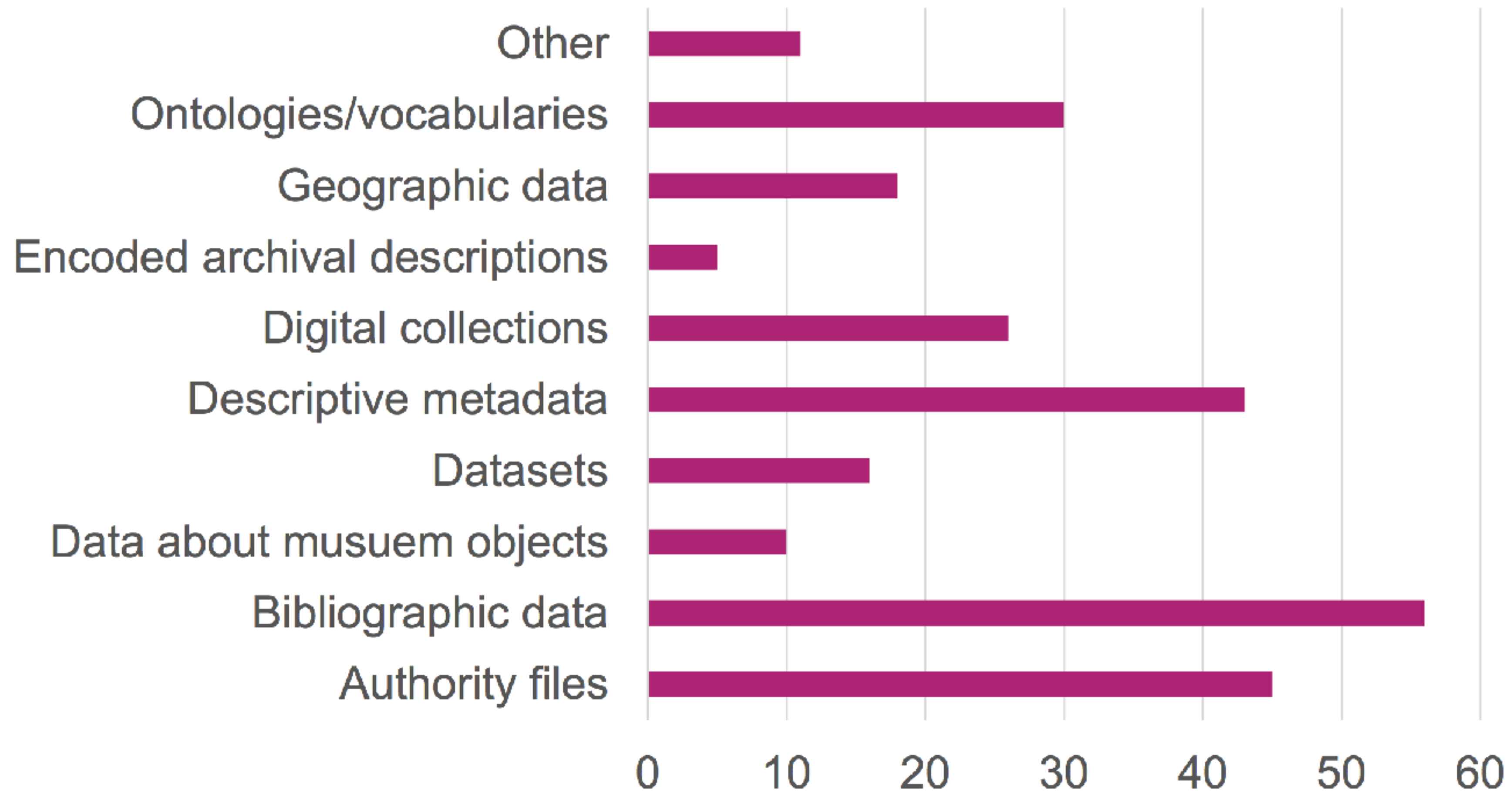
关联数据是一种自带智慧的数据

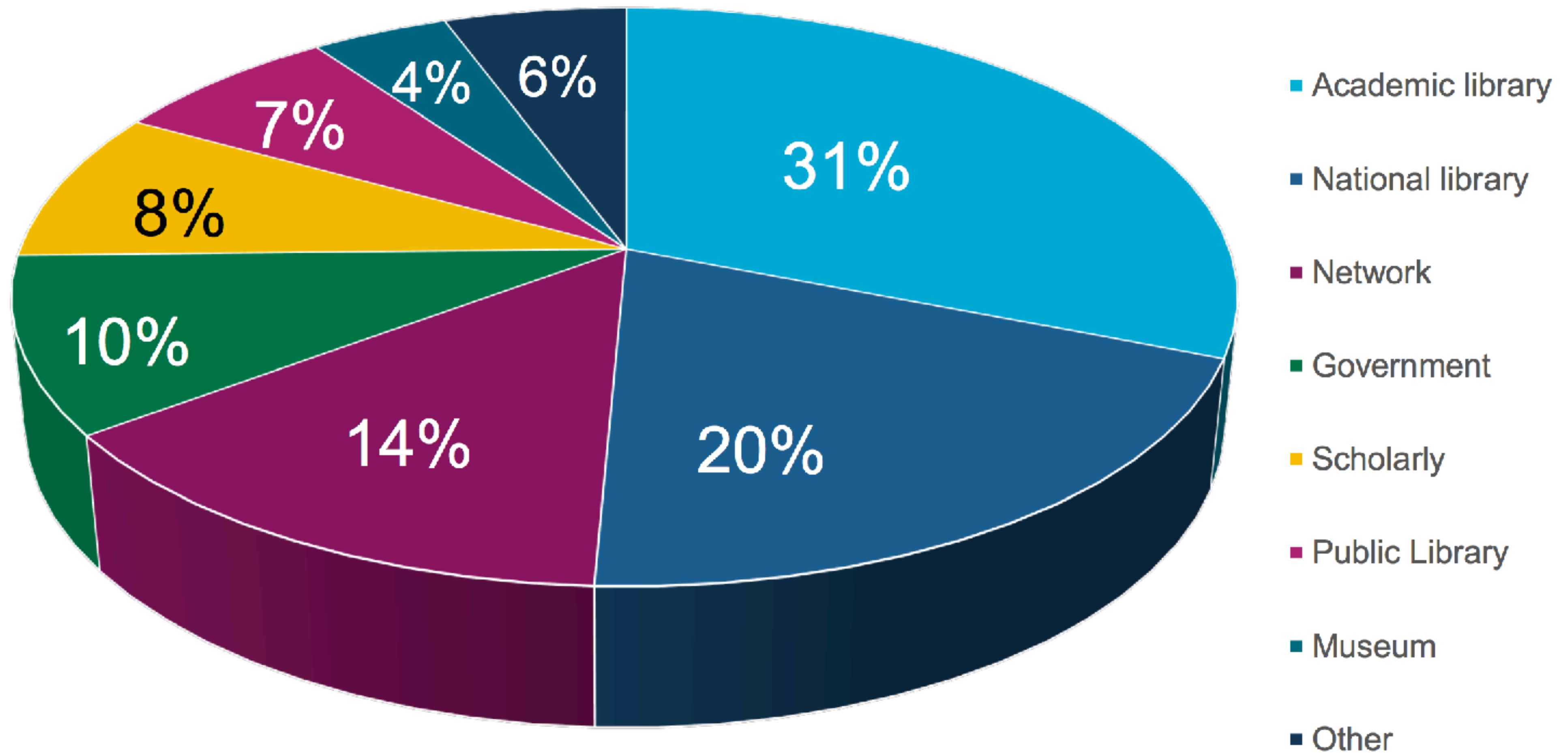






What is published as linked data





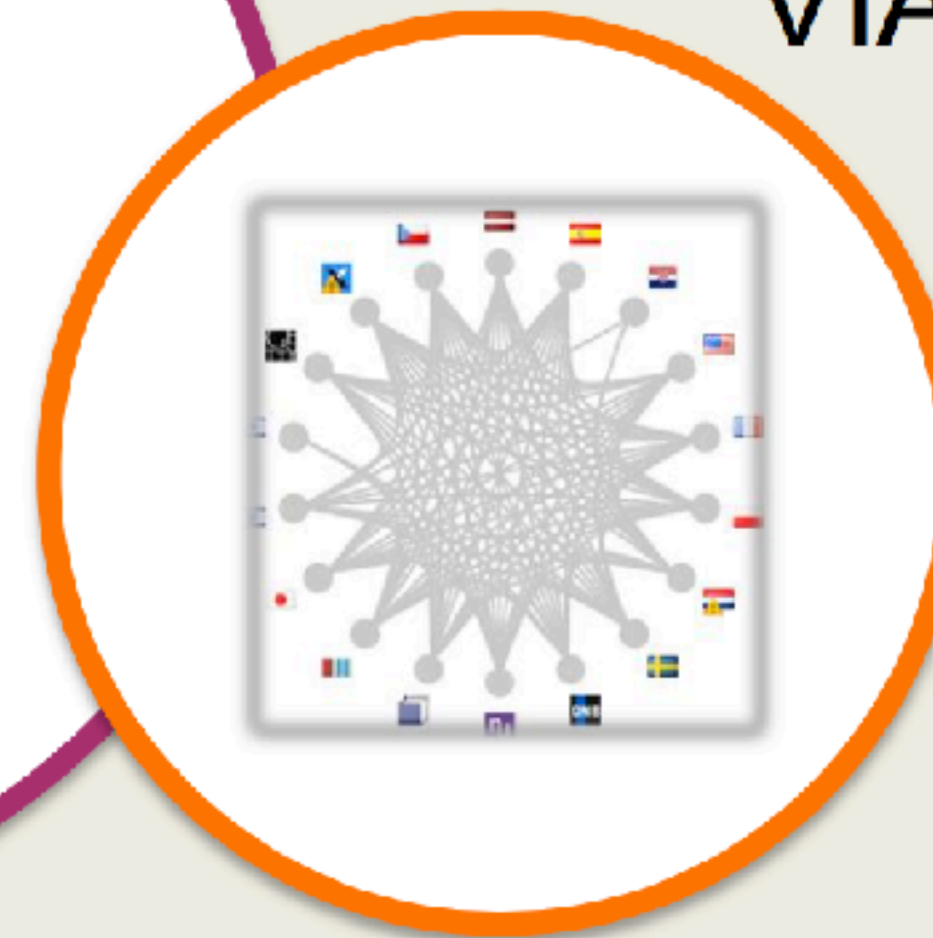
WorldCat Works

ISNI



WorldCat Catalog

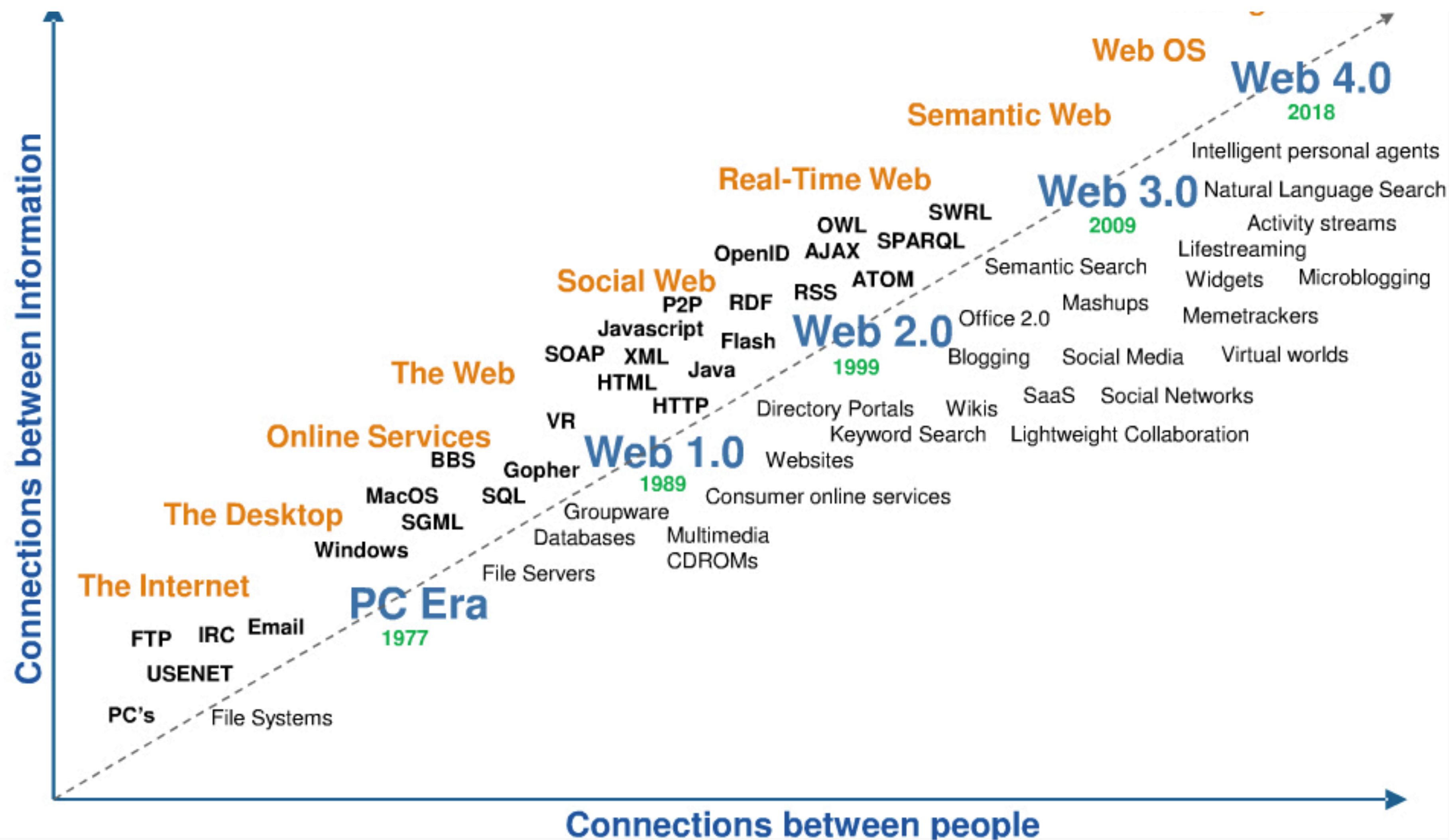
VIAF



FAST



智慧数据的达成



另一种智慧

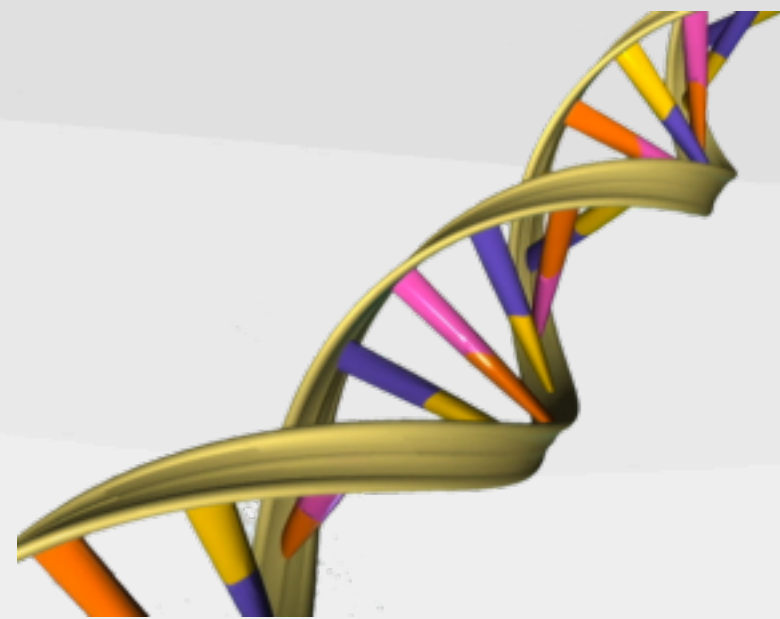
人工智能

机器学习

深度学习

什么是机器智能？

自然智能



人类智能



社会智能



机器智能



容量 30亿碱基对

150亿神经元

500TB

指数级增长

进化 千年

百年

十年

每天

智能 动物

认知

人类

“上帝”

影响 生物体

周遭环境

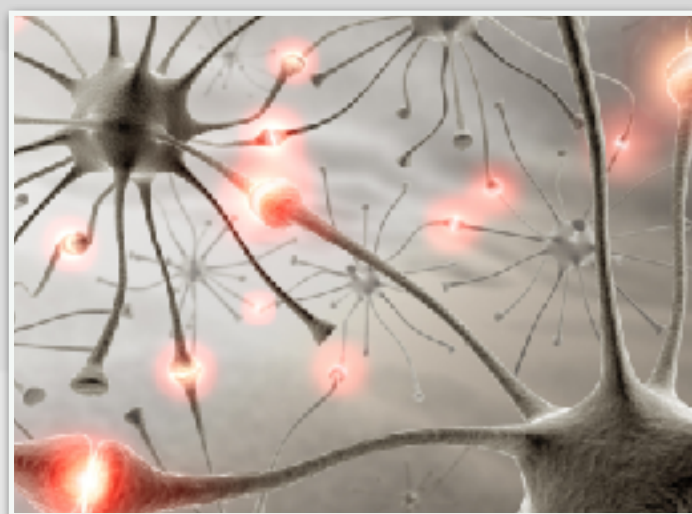
地球系统

宇宙？

符号学派



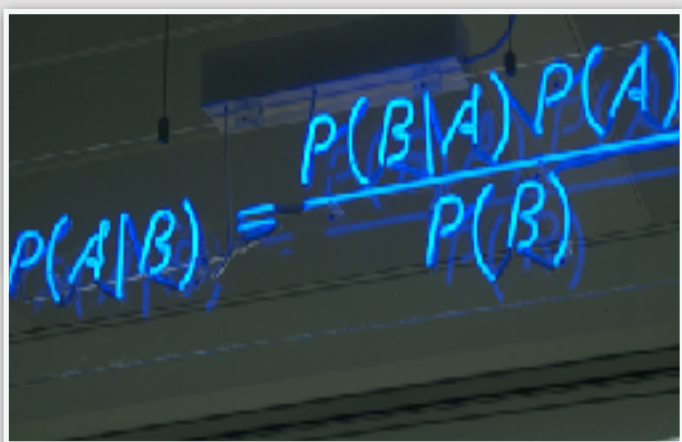
联接学派



进化学派



概率学派



类比学派



特点
来源
算法
应用
案例

发现未知
哲学逻辑
反向推演
Inverse deduction
数学证明
专家系统

仿真大脑
脑神经学
逆向传播
Back Propagation
阈值判别
图像识别

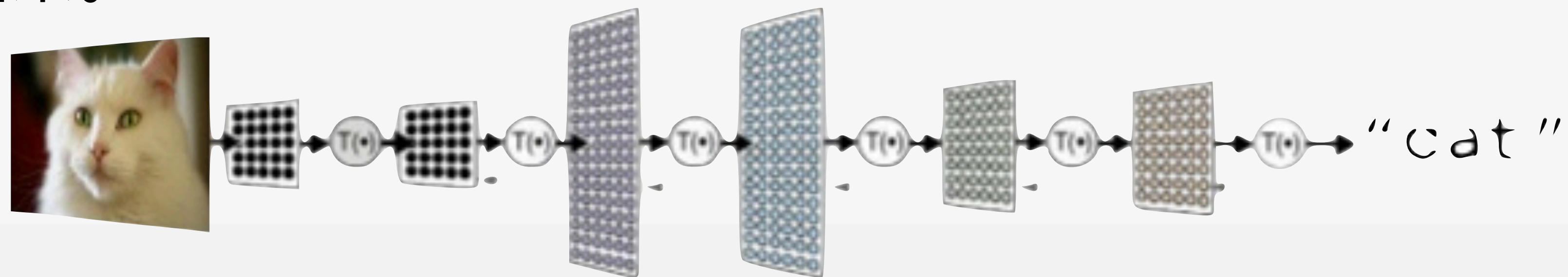
模拟进化
生物进化
遗传编程
Genetic Programming
结构发现
家用机器人

增加确定
统计学
概率推理
Probabilistic inference
不确定性
垃圾邮件

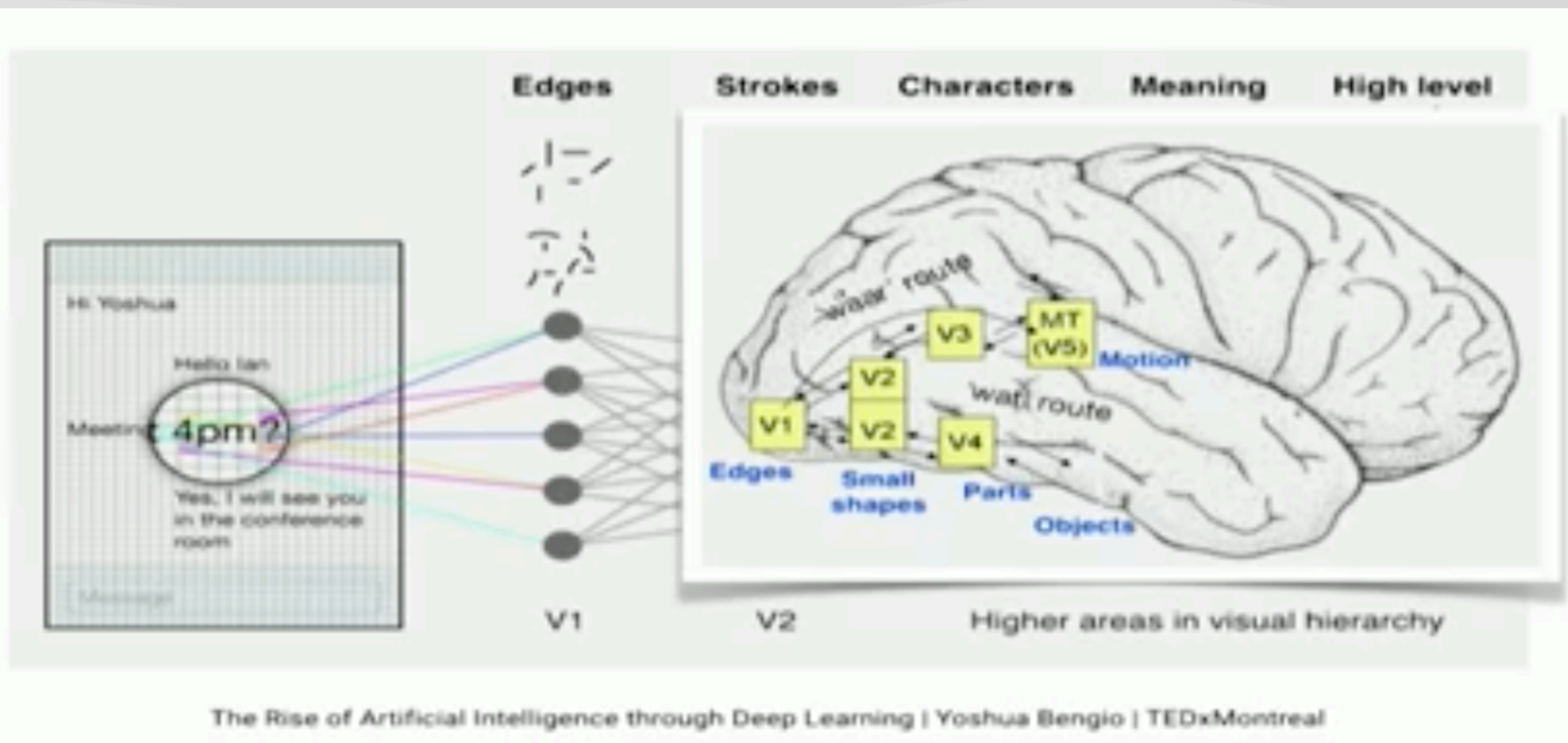
寻找相似
心理学
核机器
Kernel machines
相似性
推荐系统

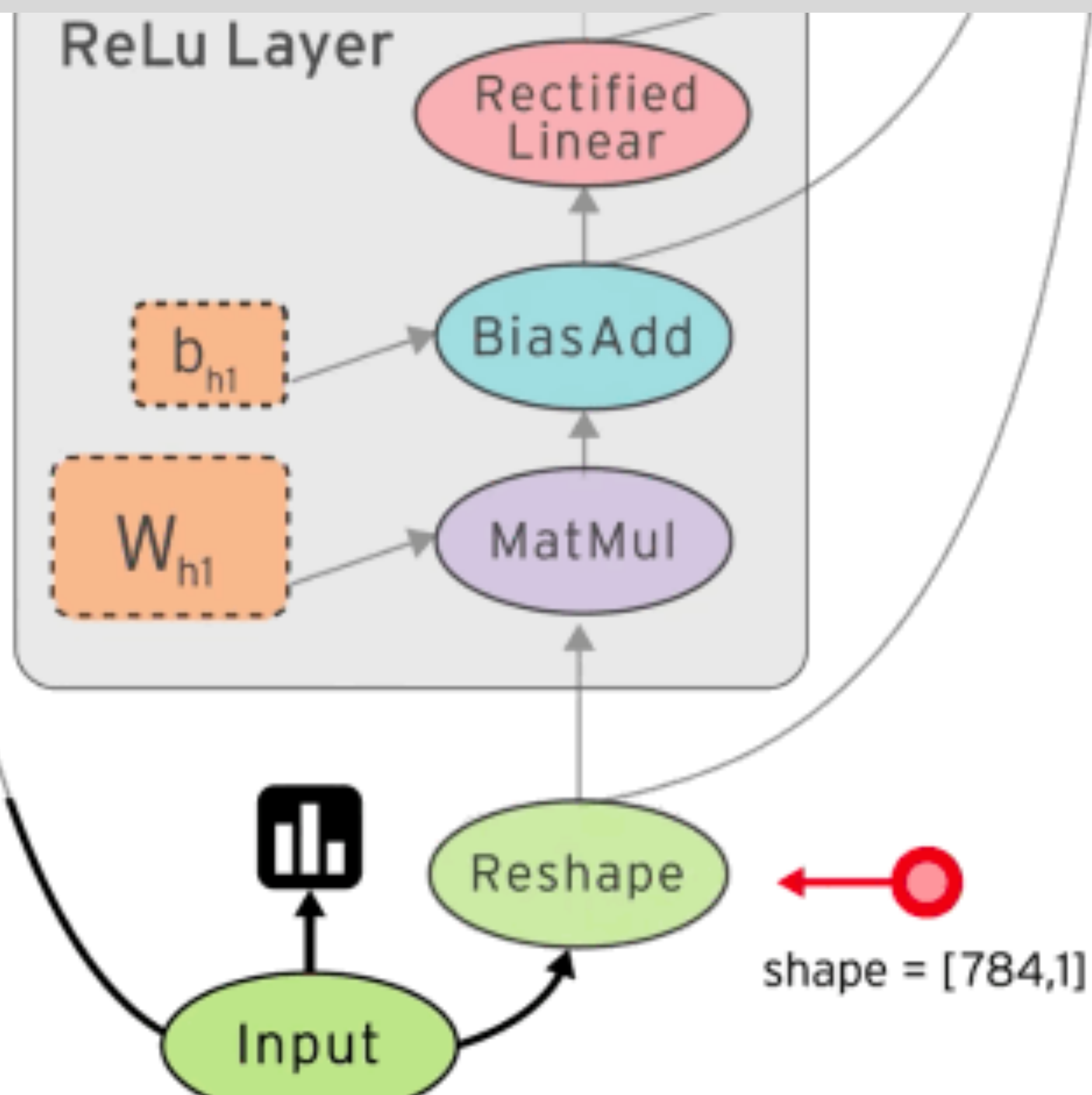
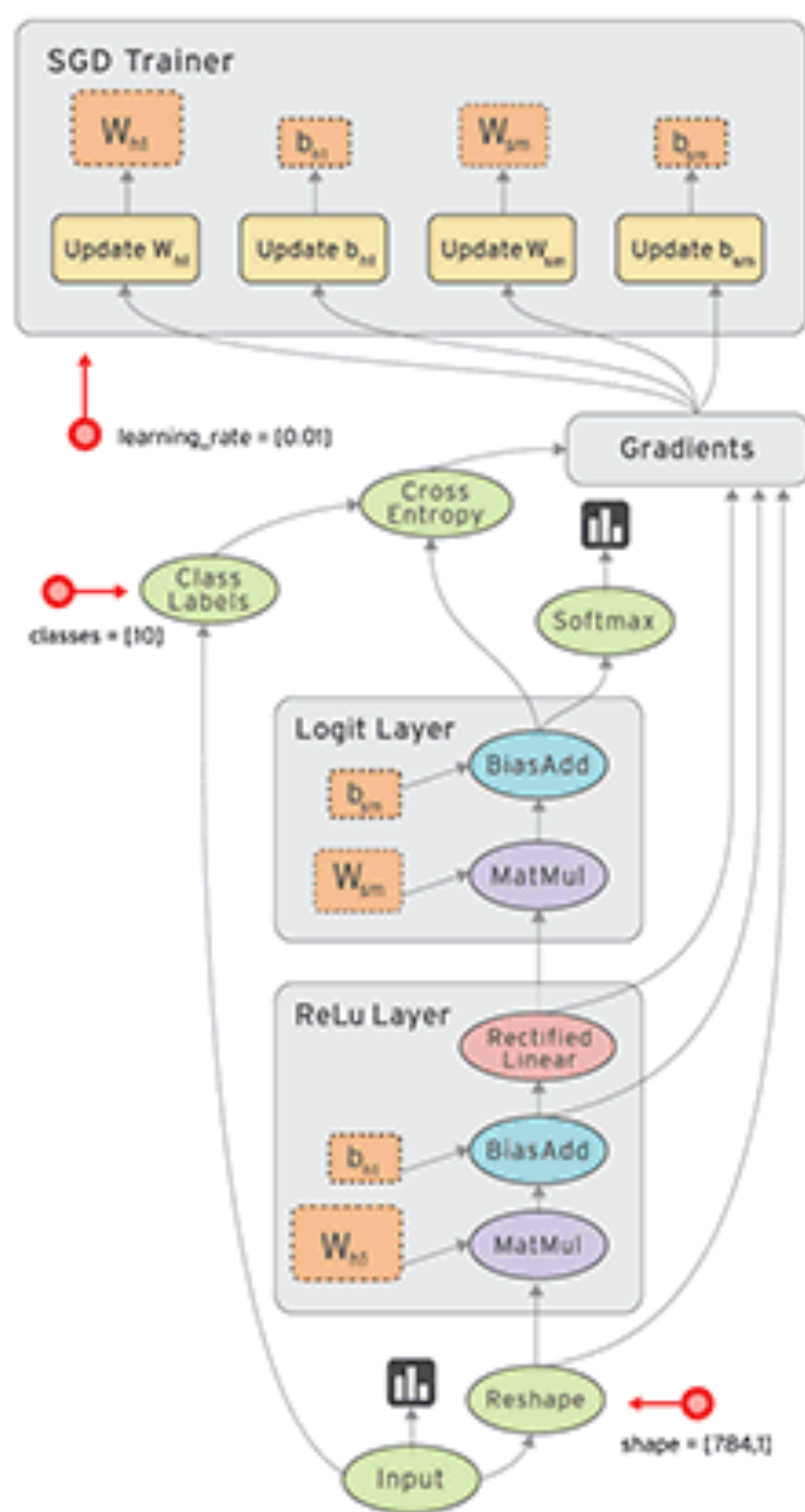
终极算法？

深度学习即人工神经网络，其灵感来自大脑皮层的结构。最基本的层次是感知器（perceptron），可以认为是用数学表示的神经元。与大脑中的层次结构一样，神经网络也有多层相互连接的感知器。第一层是输入层，该层中的每个节点传入一个输入，然后将节点的输出作为输入传递给下一层中的每个节点，在同一层中的节点之间通常没有连接。中间部分为隐藏层。这些神经元没有与外部的连接，并且只由上一层的节点激活，超过阈值的刺激经过简单的加权求和，传递到下一层的各个节点。最后一层输出处理后的结果。

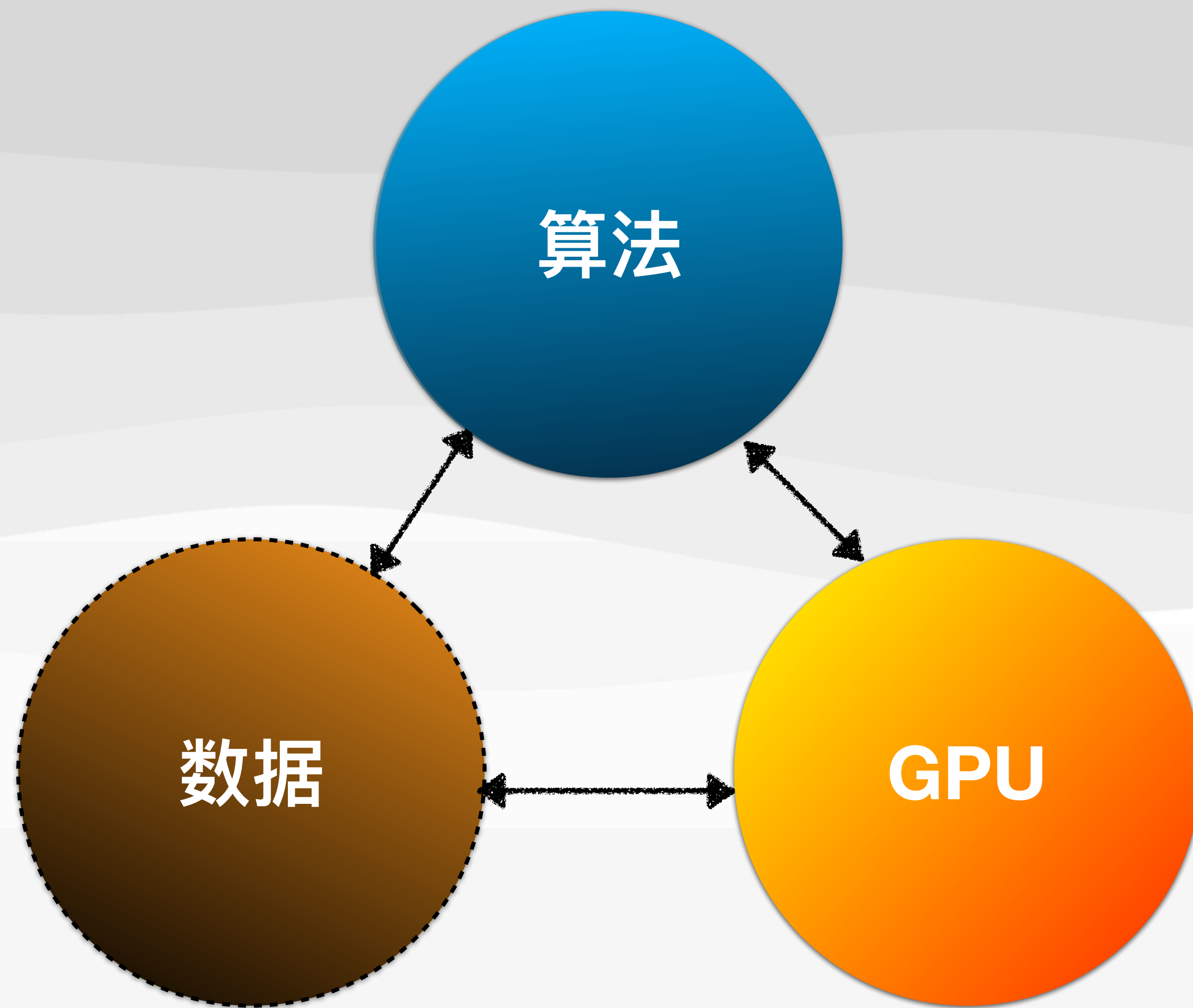


深度神经网络： 算法圣杯？





www.cybercontrols.org

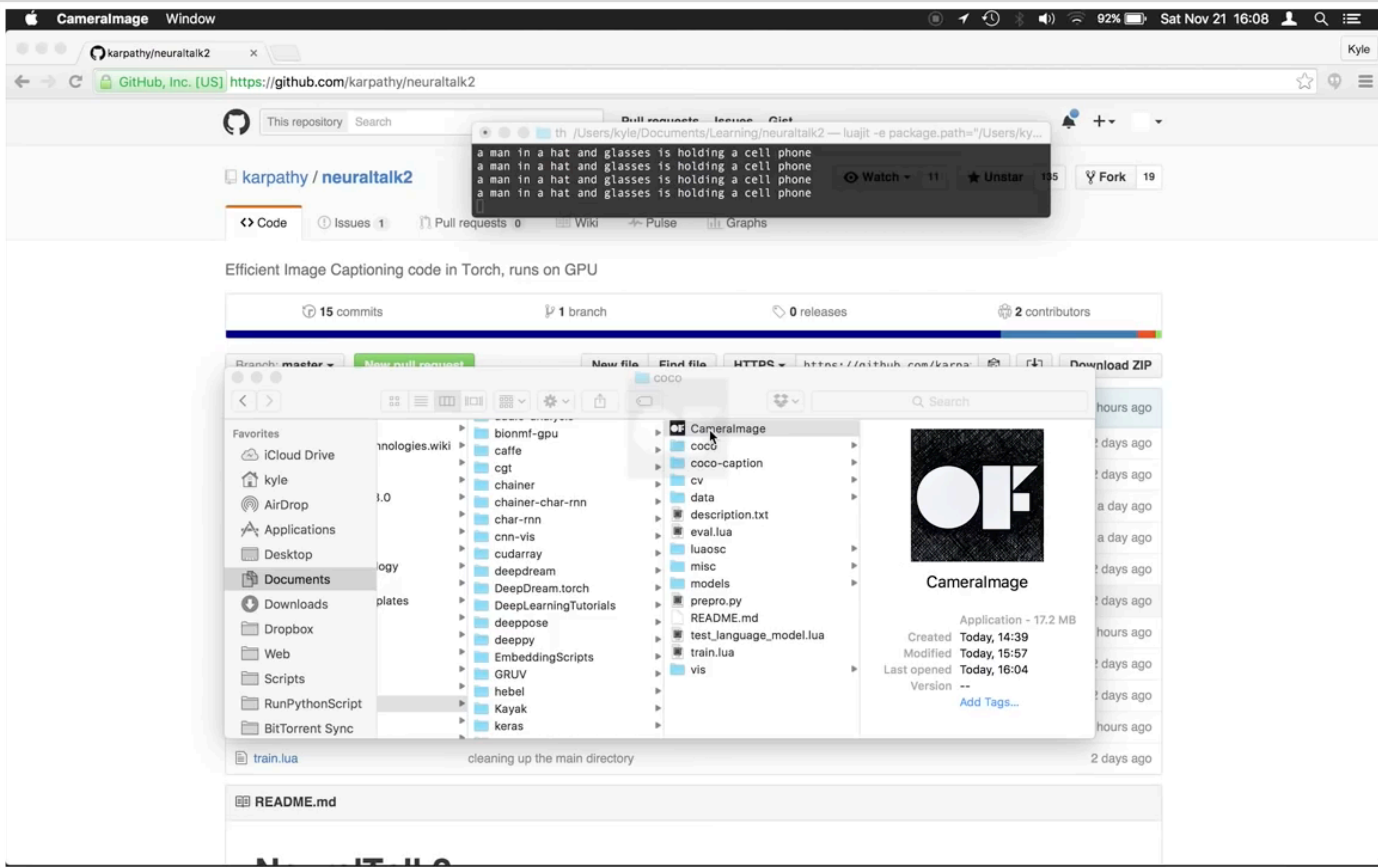




2000:
\$110 Million
6 Megawatts



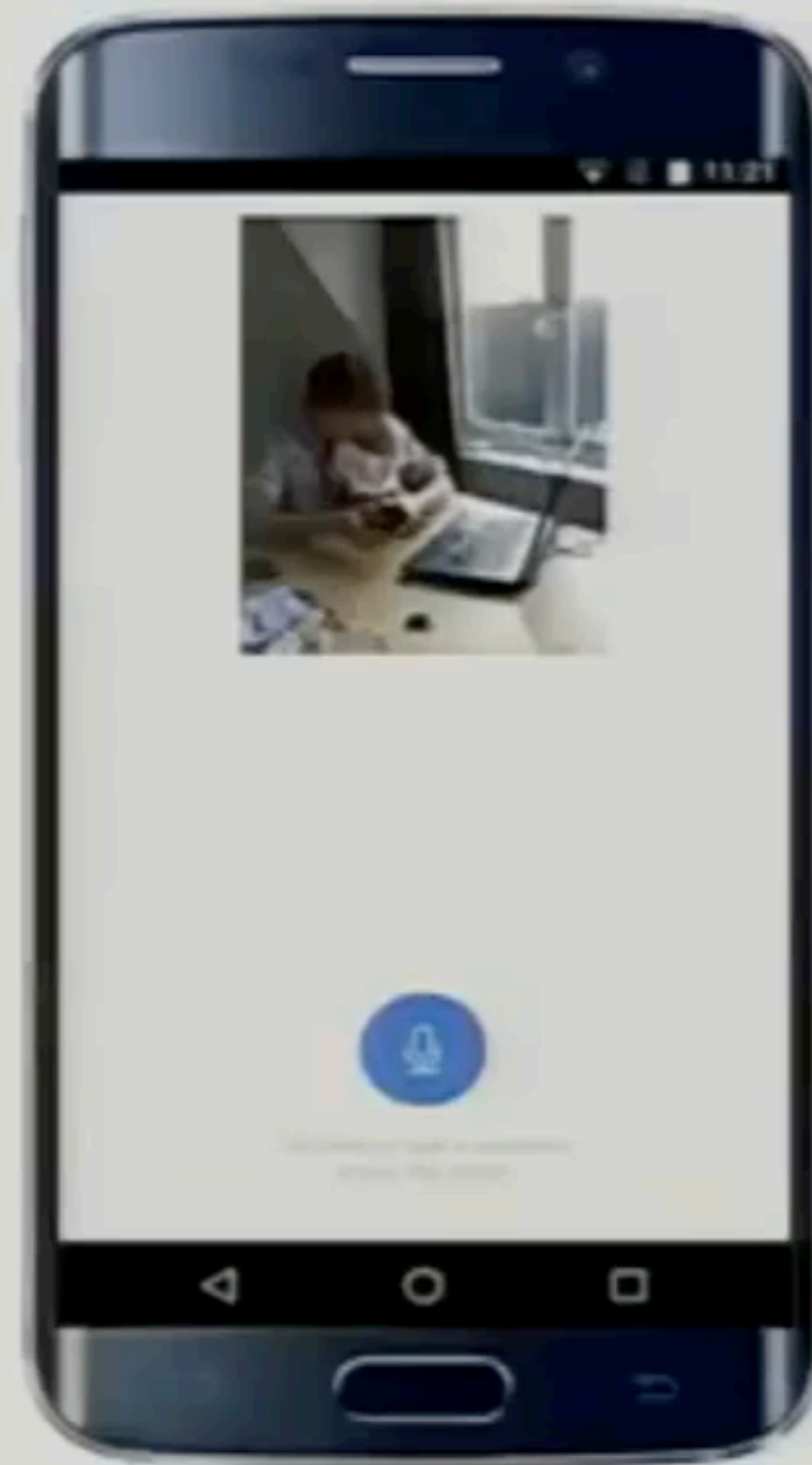
2015:
\$1000
300 watts



The background of the image is a dark teal gradient with a network diagram overlay. The diagram consists of numerous circular nodes connected by thin black lines, forming a complex web. Several nodes are highlighted with a bright cyan glow, creating a sense of depth and focus. The text 'YOLO v2' is centered in a large, white, sans-serif font.

YOLO v2

<http://pureddie.com/yolo>



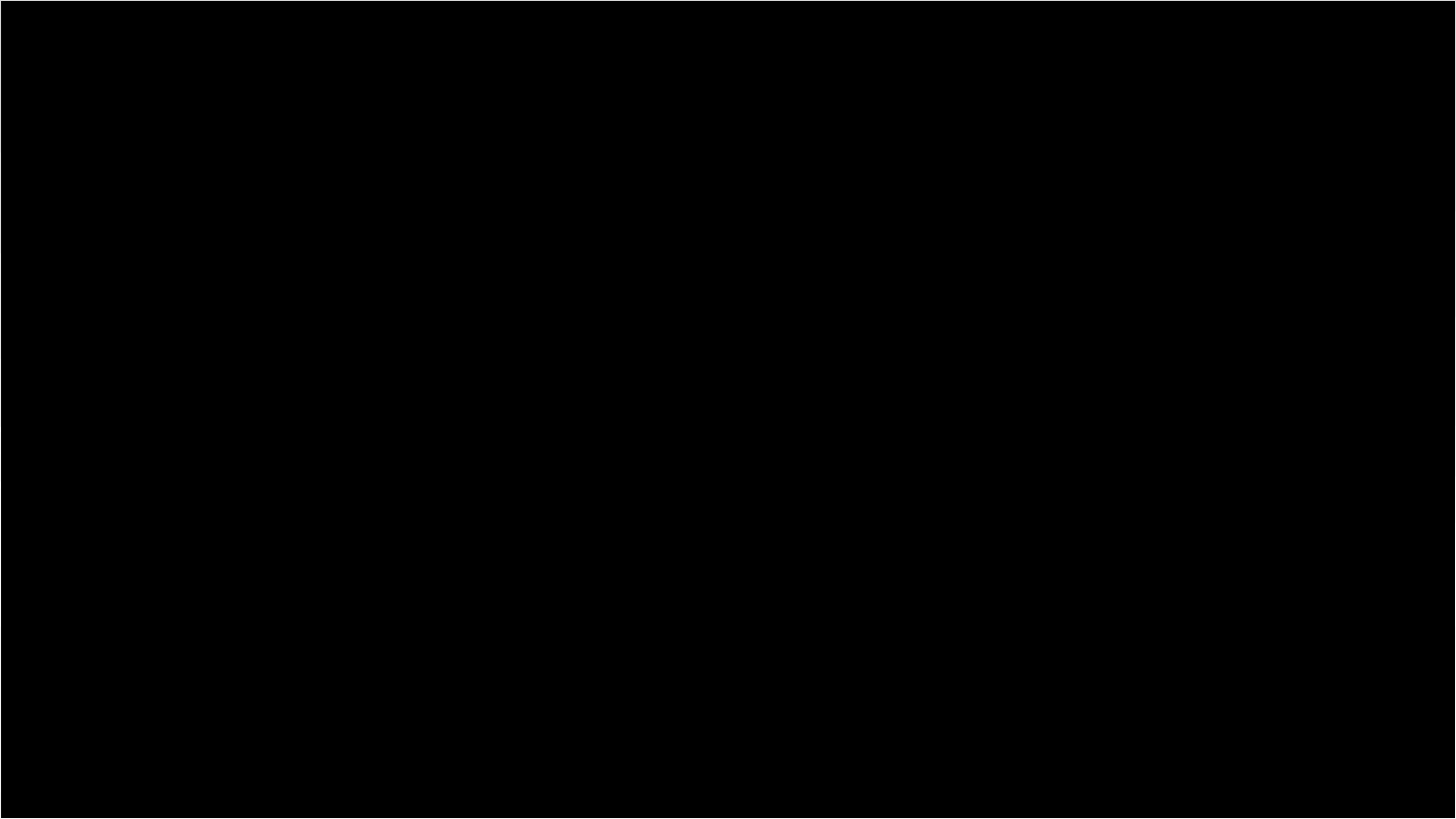


Generated story about image

Model: Romantic Novels

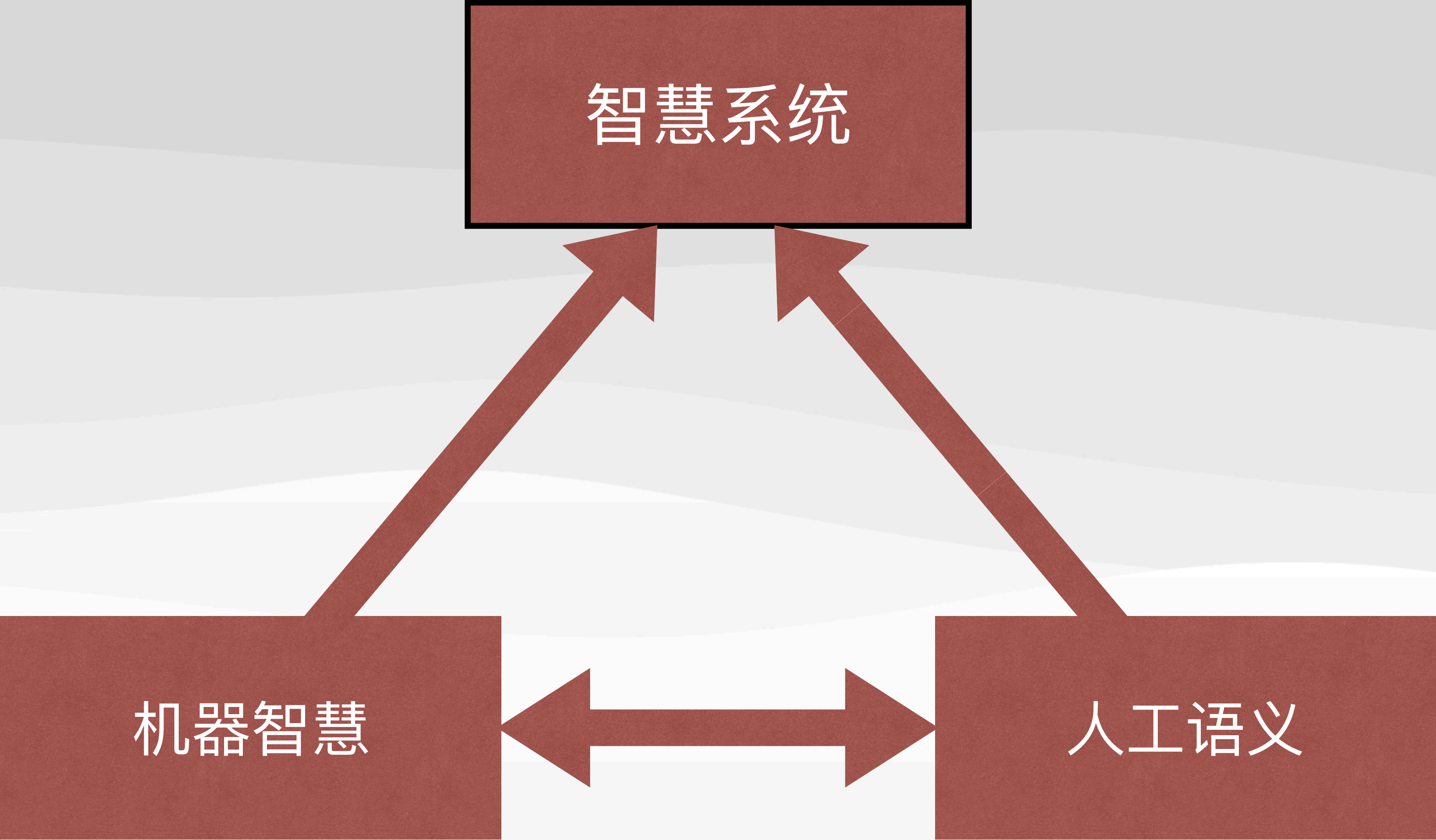
“He was a shirtless man in the back of his mind, and I let out a curse as he leaned over to kiss me on the shoulder.

He wanted to strangle me, considering the beautiful boy I’d become wearing his boxers.”

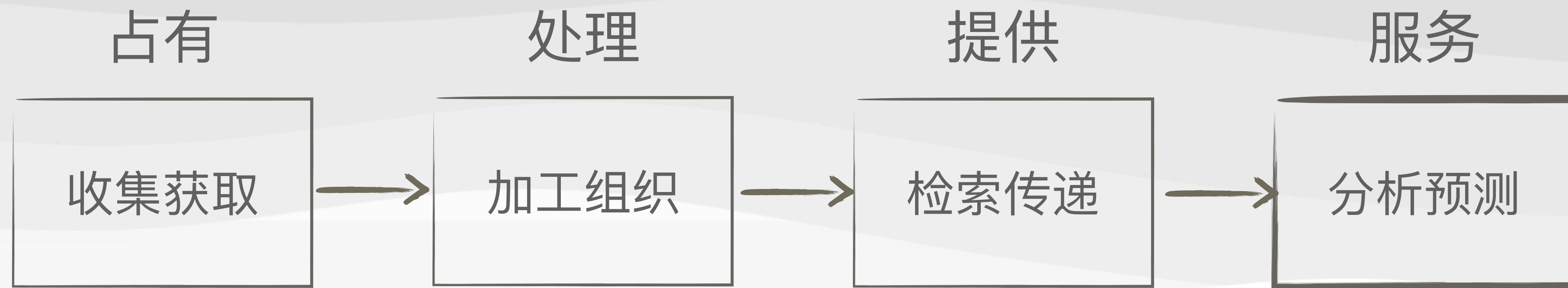




We were barely able to catch the breeze at the beach , and it felt as if someone stepped out of my mind . She was in love with him for the first time in months , so she had no intention of escaping . The sun had risen from the ocean , making her feel more alive than normal . She s beautiful , but the truth is that I don t know what to do . The sun was just starting to fade away , leaving people scattered around the Atlantic Ocean . I d seen the men in his life , who guided me at the beach once more .



新编目：数字资源的描述、揭示和组织



这些过程，其实就是计算机技术所追求的终极目标。

其中每个环节，没有人工智能达不到的

目前相关技术还不够成熟和普及，并且还不具备足够的基础设施条件

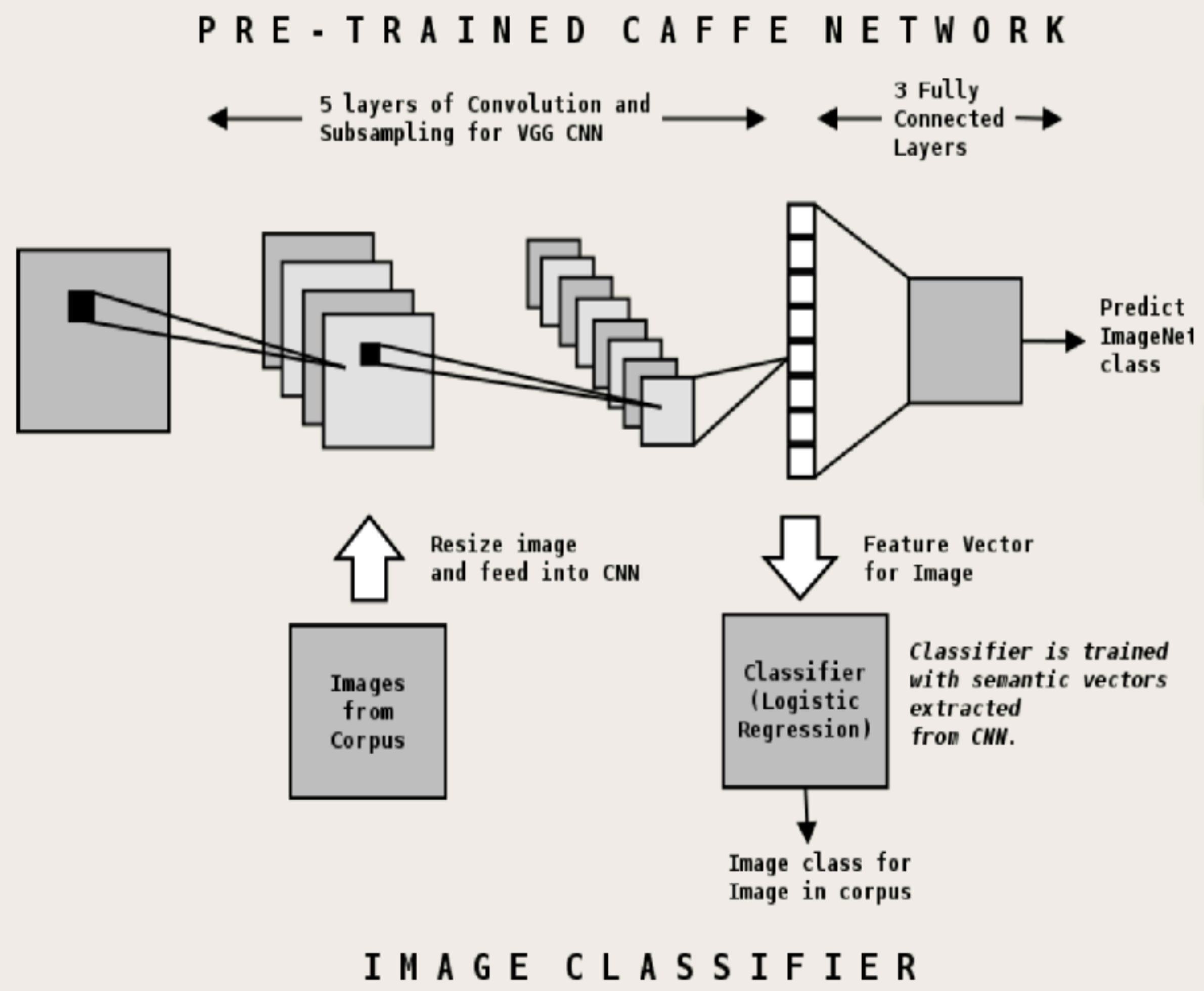
从数字化到数据化到智慧化



为什么我们不能搜索作者而得到关于作者的信息及其作品信息？包括他(她)的出生年月、简历和兴趣爱好？这都不难，关键在于我们对于书目信息的观念是否能改变……真正的挑战是改变我们目前对于数据的认识，以技术还原事物的关系，而不是对过去的孤芳自赏。

Coyle, K. (2016). FRBR, before and after: a look at our bibliographical models. Chicago: ALA Editions.

ELSEVIER: DEEP LEARNING FOR IMAGE ANNOTATION



ClinicalKey

English CME Login Register

All Types hypertension

Books Journals More

Filter By: Source Type Study Type Specialties Date

34991 results

Rate Results

Relevance Subscribed Content

CLINICAL OVERVIEW

Hypertension

Synopsis Terminology Diagnosis Treatment

Complications and Prognosis Screening and Prevention

FIRST CONSULT

Hypertension

Yanqiong Huan, MD, Assistant Professor of Medicine, Renal, Electrolyte and Hypertension Division, University of Pennsylvania Health System, Philadelphia, Pennsylvania; Ruben J. Nazario, MD, MA, Contributing Editor, First Consult. Published January 4, 2014. Last updated September 15, 2013.

FIRST CONSULT

Hypertension in children

Charles Kwon, MD, Director, Center for Pediatric Nephrology, Cleveland Clinic, Cleveland, Ohio; Ruben Nazario, MD, MA. Published February 19, 2014. Last updated September 17, 2013.

Searches related to hypertension

pulmonary hypertension pre-eclampsia essential hypertension portal hypertension

pulmonary hypertensive arterial disease intracranial hypertension

BOOK CHAPTER

Hypertension

Rosen's Emergency Medicine. Levy, Philip D. Published January 1, 2014. Pages 1113-1123 e3. © 2014.

Selected Content (6)

Comparison of Hypertension Guidelines

Mayo Clinic Proceedings • February 2015

Standardised effects of a 10 mm Hg reduction in systolic blood pressure stratified by blood pressure

The Lancet • March 2018

Evidence-based dosing of antihypertensive medications

Medical Clinics of North America • July 2015

Was this helpful? Yes or No

ELSEVIER: MACHINE LEARNING FOR LINK DISCOVERY

[illegible]

have been documented to cause
is assessed through
progresses more rapidly than
recommend
supports the assumption that
is the death of

contact dermatitis
evaluation
primary open-angle glaucoma
treatment
oxidative stress
retinal ganglion cells

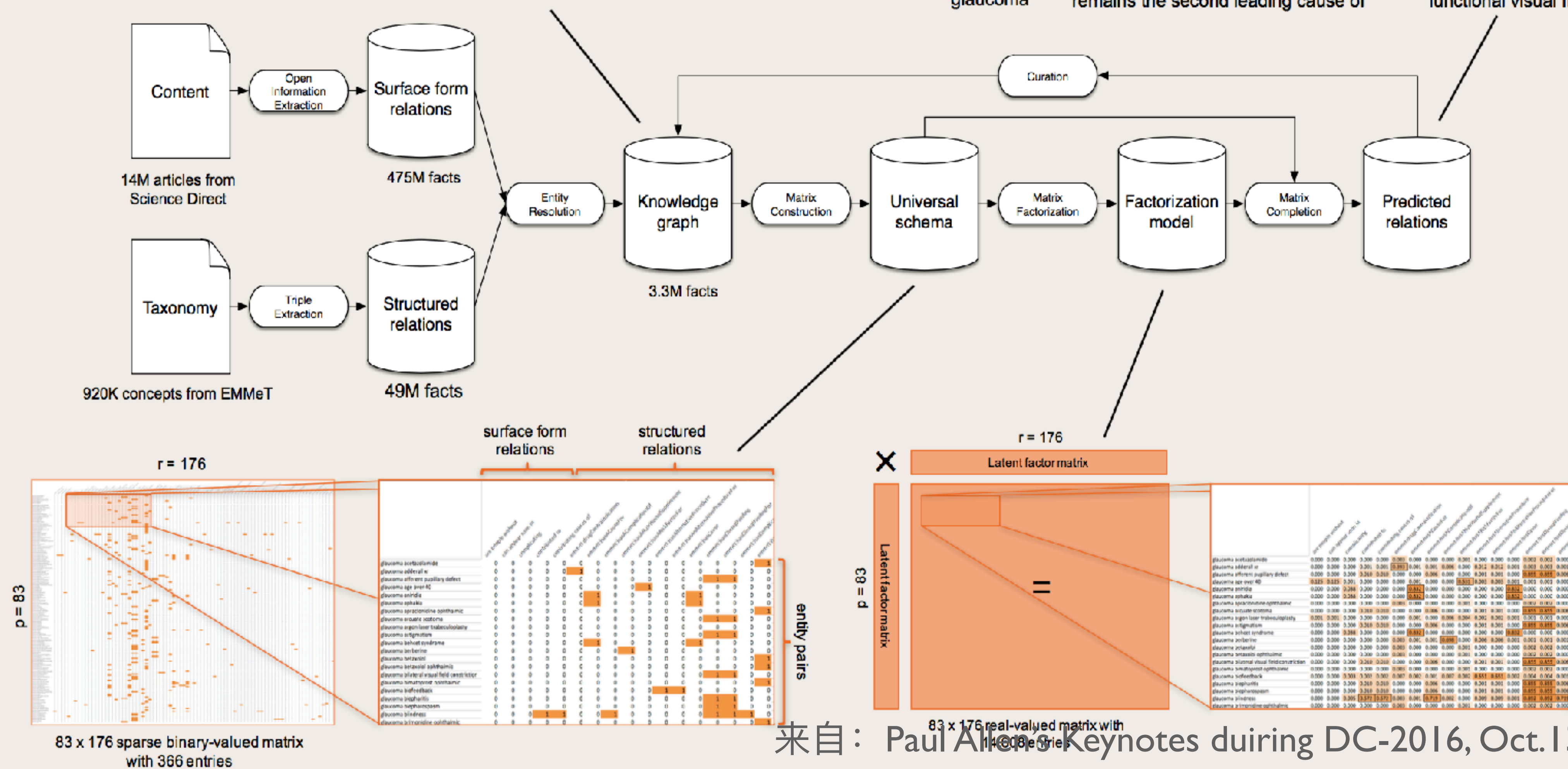
3815093
5415395
8247149
5216597
8184588
8002088

diseases
qualifier
diseases
procedures
diseases
anatomy

glaucoma
glaucoma
glaucoma
glaucoma
glaucoma
glaucoma
glaucoma
glaucoma

developed many years after
develop following
can appear soon in
can appear soon in
the risk of
contributing causes of
contributed to
is considered the second leading cause of
remains the second leading cause of

chronic inflammation of uveal tract
chronic inflammation of uveal tract
family history of glaucoma
age over 40
functional visual field loss
functional visual field loss
functional visual field loss
functional visual field loss
functional visual field loss



来自：Paul Allen's Keynotes duiring DC-2016, Oct.13, 2016

实现技术：人工智能与机器智能的结合

- 自然语言处理：NER，自动摘要、自动写作
- 图像处理：OCR、标注
- 音视频处理：文本化，翻译
- 社会关系分析：自动标注
- 关联数据：服务整合、语义互操作

新编目：在整个数字生态链中进行描述、揭示和组织

全：基于WEB的书目控制，BIB的LD化

深：数字文献的对象化、数据化：人地时事物

专：领域知识模拟：模型化、本体化、关联化

用：面向服务的文本/时空/关系的分析挖掘与可视化

广：服务平台化和社会化

基：促进社会知识交流新生态



谢谢!

刘炜 kevenlw@gmail.com

