

从语义网到知识图谱

——Web知识技术体系的变迁

鲍捷

邮箱: baojie@memect.co

微信 [baojie_memect](#)

2022-09-06 杭州 W3C TPAC China



北京文因互联科技有限公司



必也，正名乎

——“Web 3.0”的官司





Web 3.0 概念的诞生：2006年 by Tim Berners-Lee

The New York Times

A 'more revolutionary' Web

Give this article



By Victoria Shannon

May 23, 2006

EDINBURGH — Just when the ideas behind "Web 2.0" are starting to enter into the mainstream, the mass of brains behind the World Wide Web is introducing pieces of what may end up being called Web 3.0.

"Twenty years from now, we'll look back and say this was the embryonic period," said Tim Berners-Lee, 50, who established the programming language of the Web in 1989 with colleagues at CERN, the European science institute.



2007 W3C CEO的演讲 Web 3.0 Emerging

Web 3.0 Emerging

Steve Bratt

Chief Executive Officer

World Wide Web Consortium



January 2007

These slides: <http://www.w3.org/2007/Talks/0123-sb-W3CEmergingTech/>

PDF: <http://www.w3.org/2007/Talks/0123-sb-W3CEmergingTech/Overviewp.pdf>

2007 Jim Hendler的一封回信

Annotations again (was Re: RDFS compatibility information in OWL-DL documents)

This message: [[Message body](#)] [[Respond](#)] [[More options](#)]

Related messages: [[Next message](#)] [[Previous message](#)] [[In reply to](#)]

From: Jim Hendler <hendler@cs.umd.edu>

Date: Thu, 19 Apr 2007 15:13:02 -0400

Message-Id: <p06240613c24d6eaea0e2@[128.213.51.217]>

To: Dan Brickley <danbri@danbri.org>, Bijan Parsia <bp@cs.cmu.edu>

Cc: wangxiao@muscd.edu, Alan Ruttenberg <alanruttenberg@gmail.com>

changing the subject a bit, but riffing off Dan's email, I note that one of the things we'd love to do more of, be nice to have a good way to do in OWL 1.1, is to have a way to either subproperty these annotation properties or to have some new things with, perhaps, an operational (rather than formal) semantics that has some instruction for machine readable properties -- just as an example, we've started to put descriptors in some OWL ontologies that point to pages in places like flickr and wikipedia that describe the concept - problem is that SWOOP and other editors think they are just text strings (if we use the rdfs: properties) or think they are meaningful semantic properties if we use OWL data or object type properties. The implementation work around we use is to subproperty rdfs:seeAlso - which technically puts us in OWL full (although Pellet is smart enough to ignore them :-))

Anyway, I know this topic has come up before, but looking through the discussion I don't see a resolution, so I thought I'd add this important "Web 3.0" use case

-JH

--

How can you be in two places at once if you're not anywhere at all?

(Firesign Theatre, 1969)

Prof James Hendler

Tetherless World Constellation Chair

Computer Science Dept

Rensselaer Polytechnic Institute, Troy NY 12180

<http://www.cs.rpi.edu/~hendler>

301-405-2696 (work)

301-405-6707 (Fax)



2009 Web 3.0 Manifesto

Web 3.0 Manifesto

How Semantic Technologies in Products and Services
Will Drive Breakthroughs in Capability, User Experience,
Performance, and Life Cycle Value



Mills Davis
Principal Investigator
mdavis@project10x.com

Project10X
2853 Ontario Road NW #501
Washington, DC 20009 USA
1-202-667-6400
1-800-713-8049 fax
www.project10x.com



Table of Contents

- Overview, 3
- What is the Semantic Wave?, 4
- What Are Semantic Technologies?, 4
- What is Value in the Next Stage of the Internet?, 5
- What is the Value Space for Semantic Technologies?, 5
- How Do Semantic Technologies Tap New Sources of Value?, 6
- Why is User Experience a Key Value Dimension?, 7
- How Much Can Semantic Technologies Impact Performance?, 8
- How do Semantic Technologies Maximize Life Cycle Value?, 9
- How do Capabilities of Semantic Technology Drive Value?, 11
 - Knowledge Representation, 11
 - Automated Reasoning, 13
 - Semantic Content Tools, 13
 - From Search to Knowing, 14
 - Semantic User Experience, 15
 - Semantic Social Computing, 16
 - Semantic Collaboration, 17
 - Semantic Enterprise Processes, 18
 - Knowledge-based Applications, 18
 - Intelligent Systems, 19
 - Semantic Architectures, 20
 - Semantic Storage, 21
 - Semantic Transport, 21
 - Semantic Processors, 22
 - Semantic Software Development, 22
 - Semantic Ecosystem, 23
- Why is Business Concept Innovation Important in the Next Stage of the Internet?, 24
- Is the Market Ready for Semantic Technologies?, 25
- Summary, 26
- Semantic Wave Report, 28
- Semantic Exchange, 29

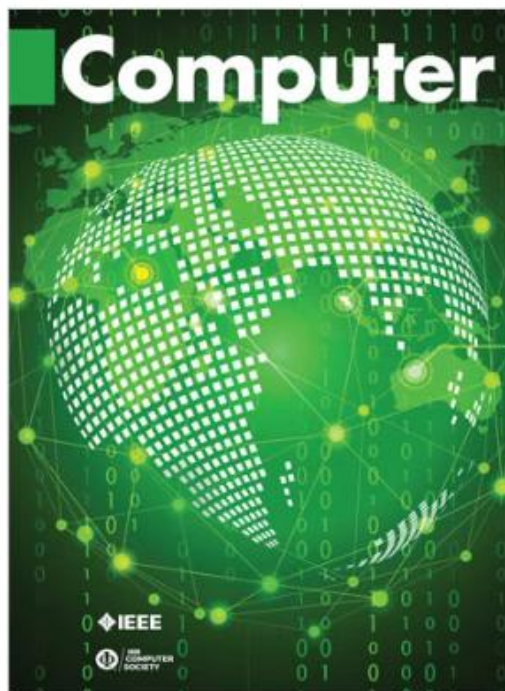
About the Author

Mills Davis is Founder and Managing Director of Project10X — a research consultancy specializing in next wave semantic technologies, solutions, and business models. The firm's clients include technology manufacturers, global 2000 corporations, government agencies, and next-generation web start-ups. Mills serves as principal investigator for the Semantic Wave research program. A noted consultant and industry analyst, he has authored more than 100 reports, white papers, articles, and industry studies.

Mills is active in both government and industry-wide technology initiatives that are advancing semantic technologies. He co-chairs SemanticCommunity.net, which carries on the mission the Federal Semantic Interoperability Community of Practice (SiCoP) in supporting Communities of Interest in both government and private industry. Mills is a founding member of the AllIM interoperable enterprise content management (IECM) working group, and a founding member of the National Center for Ontology Research (NCOR). Also, he serves on the advisory board of several new ventures in the semantic space.

mdavis@project10x.com
www.project10x.com/
www.semanticexchange.com/

2009 IEEE Computer 杂志特刊



[Home](#) / [Magazines](#) / [Computer](#) / 2009.01

Computer

Web 3.0 Emerging

January 2009, pp. 111-113, vol. 42

DOI Bookmark: [10.1109/MC.2009.30](#)

Authors

[Jim Hendler](#), Rensselaer Polytechnic Institute



DOWNLOAD PDF



SHARE ARTICLE



GENERATE CITATION

Keywords

Web 3 0, Web Technologies, Semantic Web, Linked Data

Abstract

While Web 3.0 technologies are difficult to define precisely, the outline of emerging applications has become clear over the past year.



Previous



Next



Table of Contents



Past Issues



Related Articles



■ 2009 AAAI 研讨会: Web 2.0 + Web 3.0 = ?

■ SOCIAL SEMANTIC WEB: WHERE WEB 2.0 MEETS WEB 3.0

Papers from the 2009 AAAI Spring Symposium

Mark Greaves, Li Ding, Jie Bao, and Uldis Bojars, *Program Cochairs*

Technical Report SS-09-08. Published by The AAAI Press, Menlo Park, California. [This technical report is also available in book and CD format.](#)

Please Note: Abstracts are linked to individual titles, and will appear in a separate browser window. Full-text versions of the papers are linked to the abstract text. Access to full text may be restricted to AAAI members. PDF file sizes may be large!

Contents

[Organizing Committee](#) / iii

Mark Greaves, Li Ding, Jie Bao, and Uldis Bojars

2014 Gavin Wood 重新定义了 Web 3.0

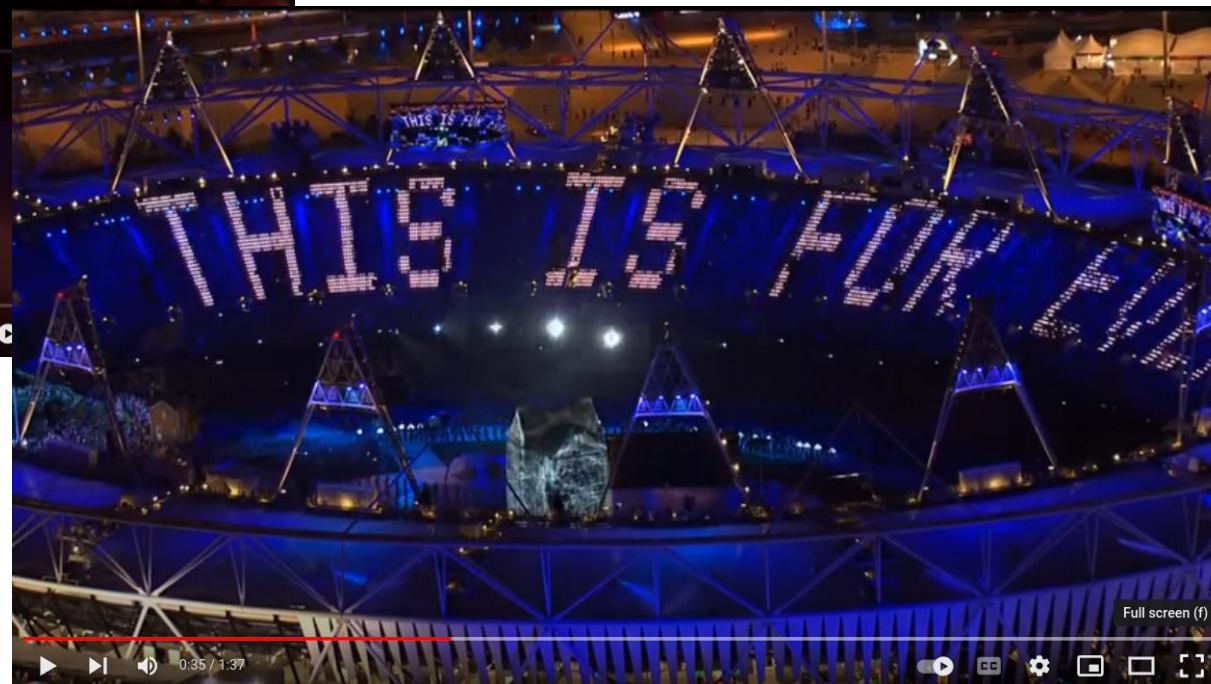
Note: originally posted Wednesday, 17 April 2014 on gavofyork's blog Insights into a Modern World.

DApps: What Web 3.0 Looks Like

As we move into the future, we find increasing need for a zero-trust interaction system. Even pre-Snowden, we had realised that entrusting our information to arbitrary entities on the internet was fraught with danger. However, post-Snowden the argument plainly falls in the hand of those who believe that large organisations and governments routinely attempt to stretch and overstep their authority. Thus we realise that entrusting our information to organisations in general is a fundamentally broken model. The chance of an organisation not meddling with our data is merely the effort required minus their expected gains. Given they tend to have an income model that requires they know as much about people as possible the realist will realise that the potential for convert misuse is difficult to overestimate.

The protocols and technologies on the Web, and even at large the Internet, served as a great technology preview. The workhorses of SMTP, FTP, HTTP(S), PHP, HTML, Javascript each helped contribute to the sort of rich cloud-based applications we see today such as Google's Drive, Facebook and Twitter, not to mention the countless other applications ranging through games, shopping, banking and dating. However, going into the future, much of these protocols and technologies will have to be re-engineered according to our new understandings of the interaction between society and technology.

■ Tim Berners-Lee 的愿景：This is for everyone





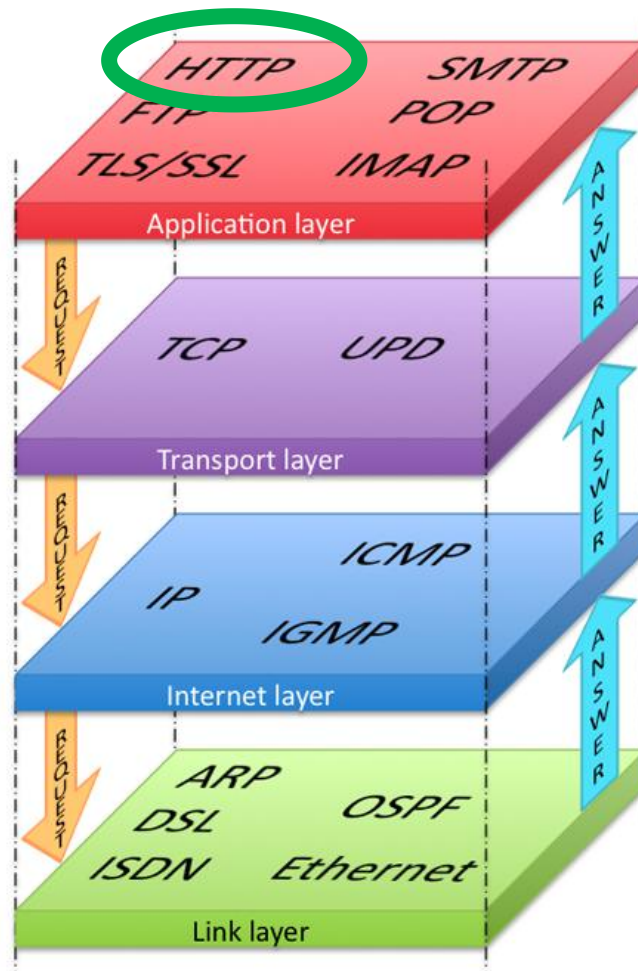
1

语义网的初衷

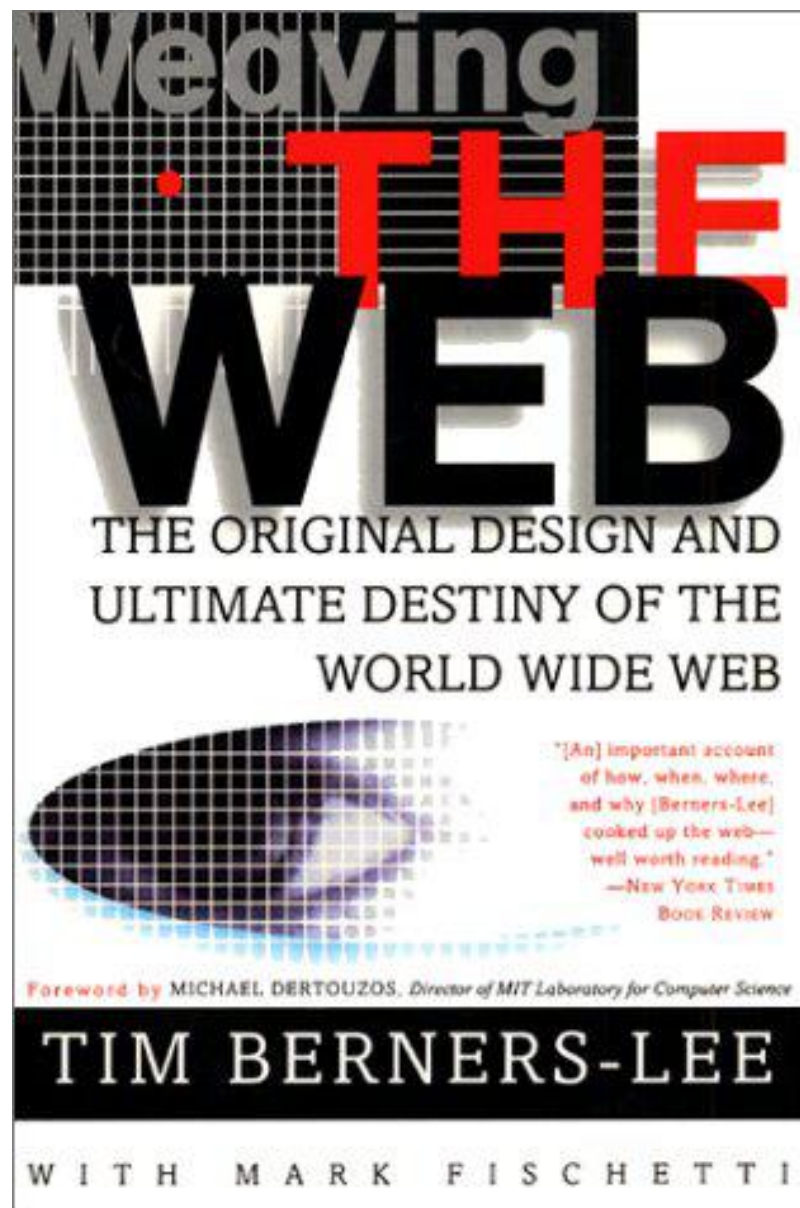


Web是什么？

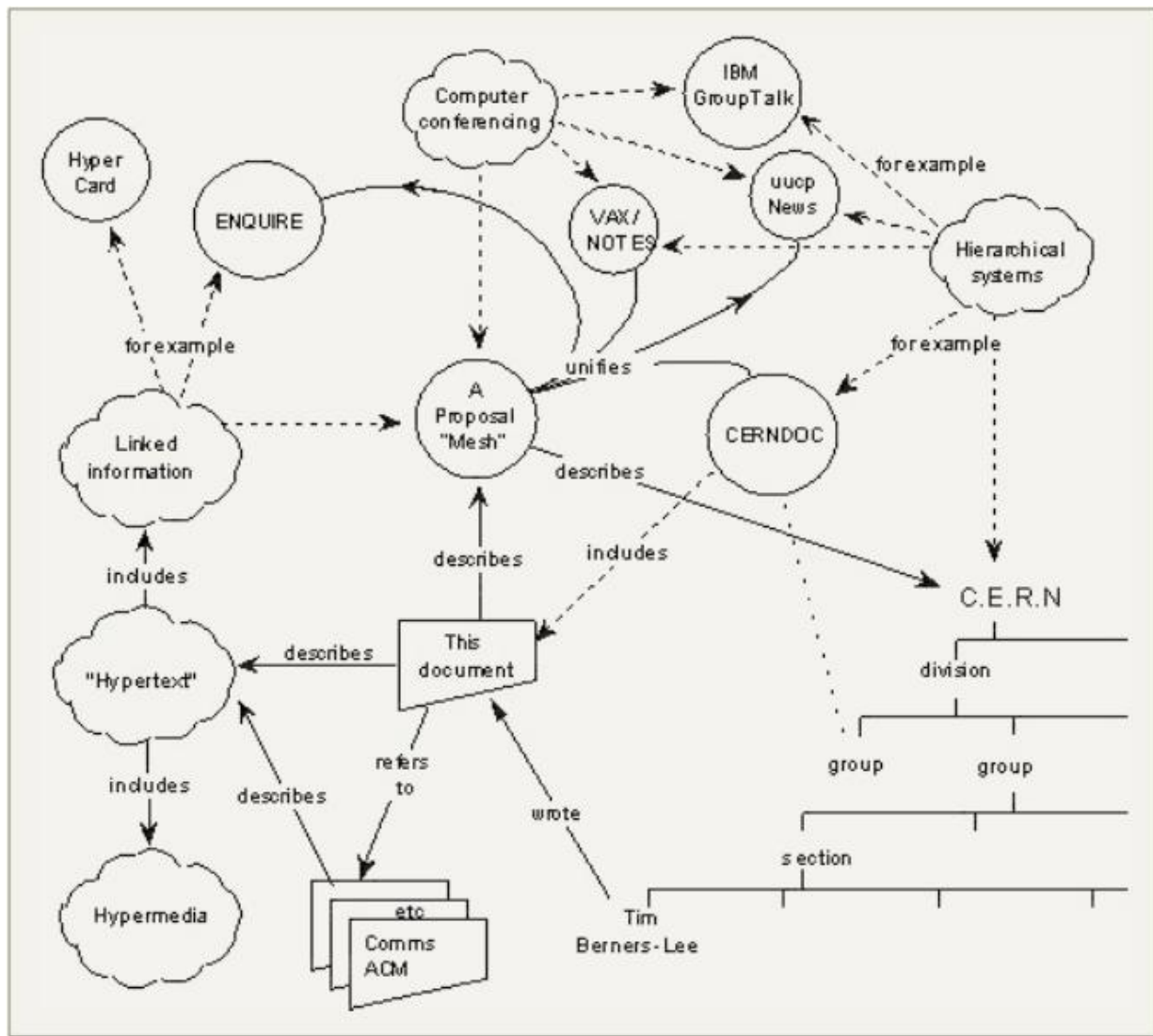
Web是因特网的一个应用



■ 一个宏伟规划

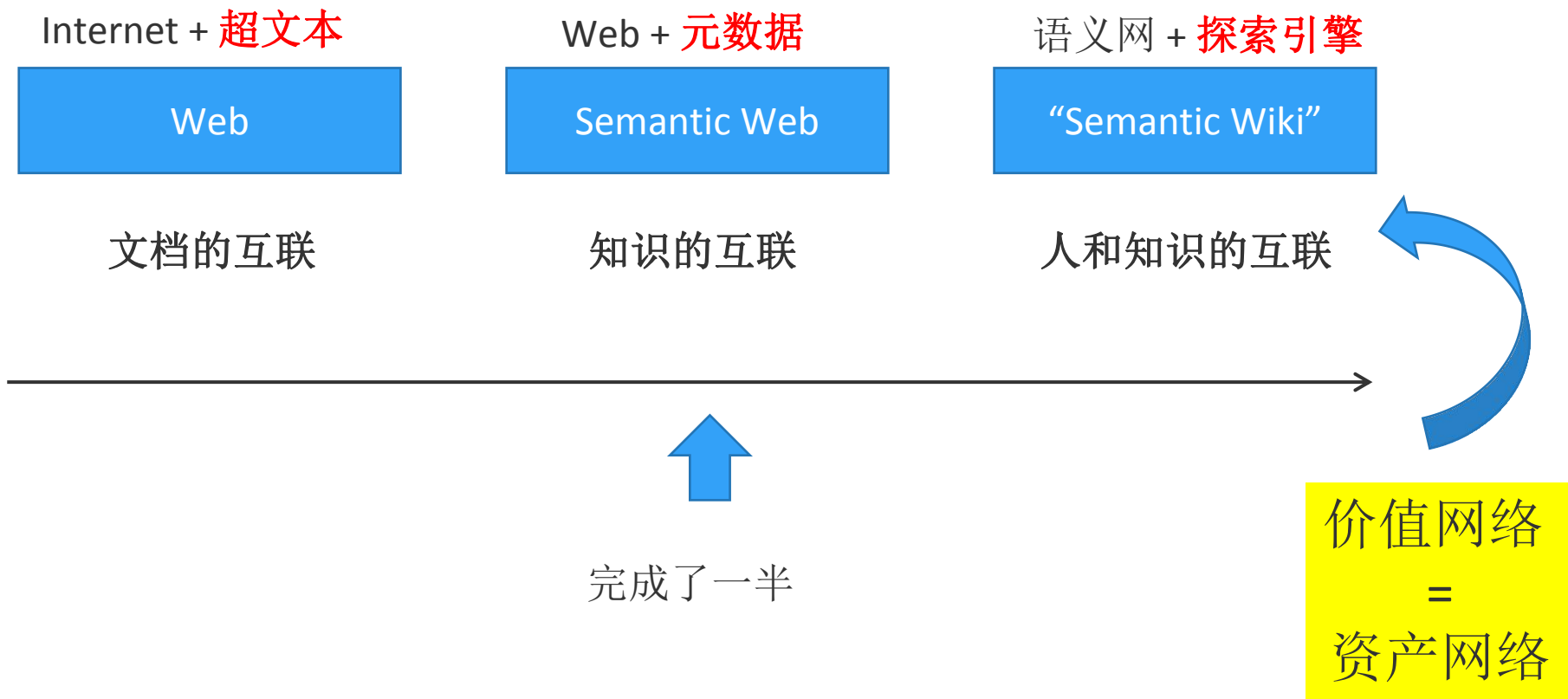


1989年Web规划即含"语义"



一个宏伟规划

Tim Berners-Lee在Web最初的规划中（1991），试图实现递进的三个网络





Web 1.0的本质

文档的互联

2004: 社会网络成为了Web 2.0



Social Network: 社会网络，而非社交网络

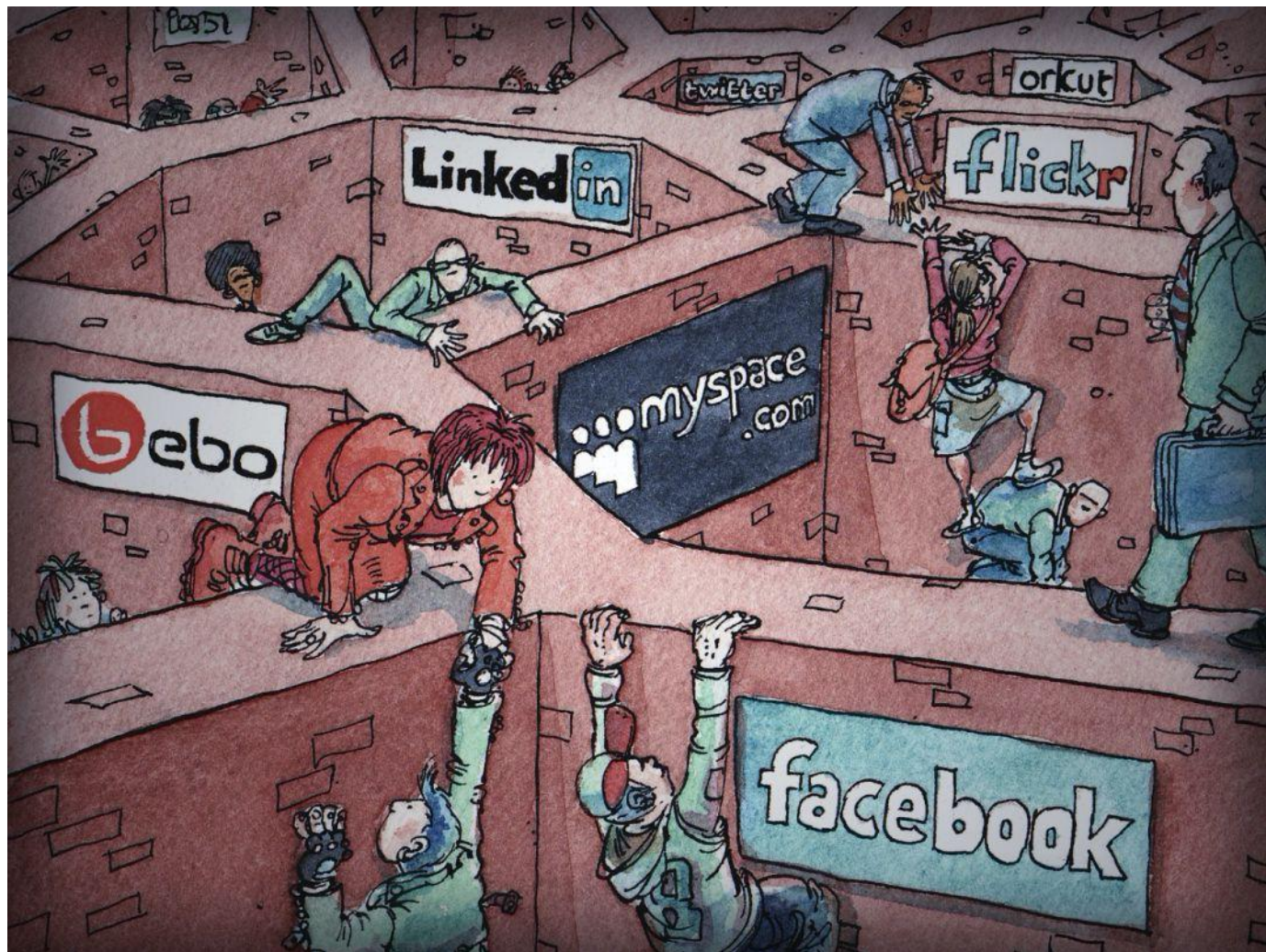


Web 2.0 的本质

实体的互联

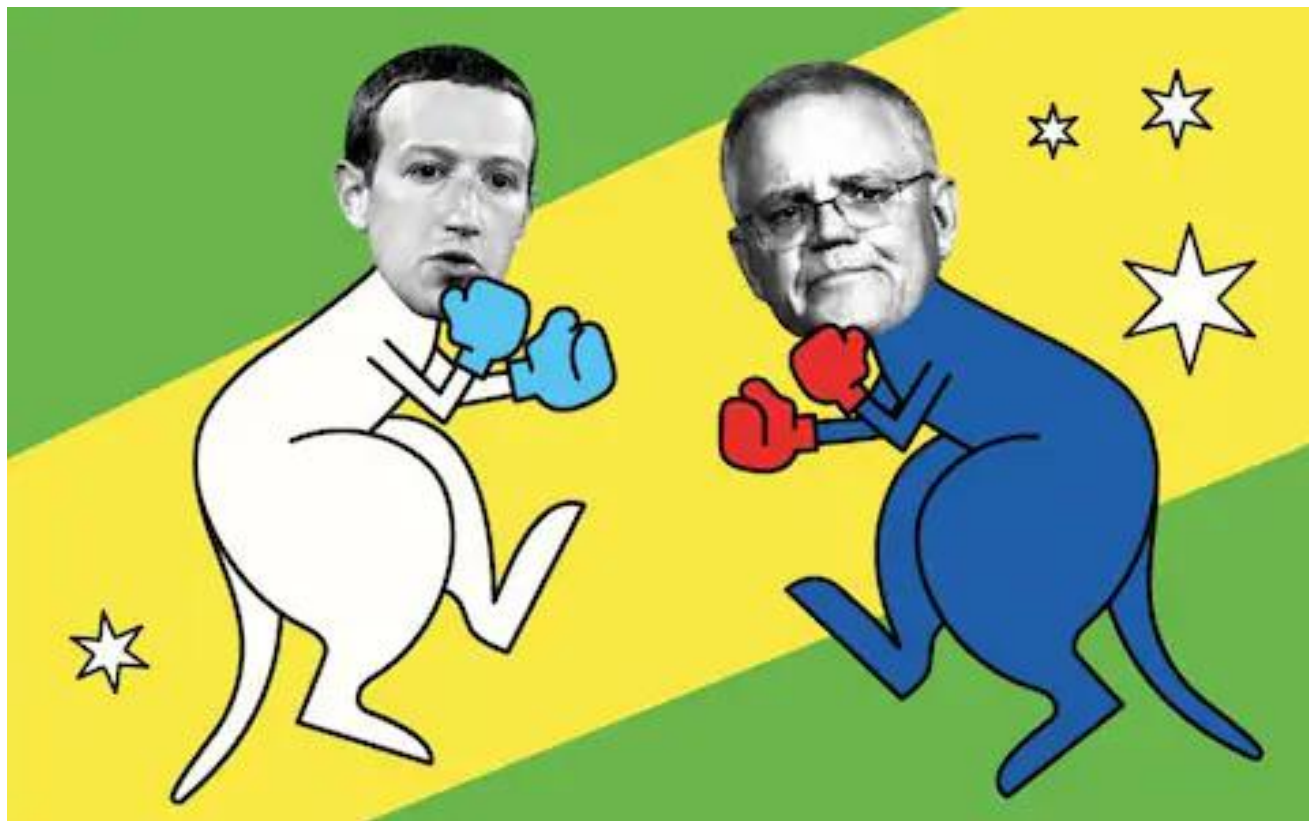
映射物理世界实体
实体间的关系即知识

但是社会网络走偏了：价值的封闭即价值的丧失

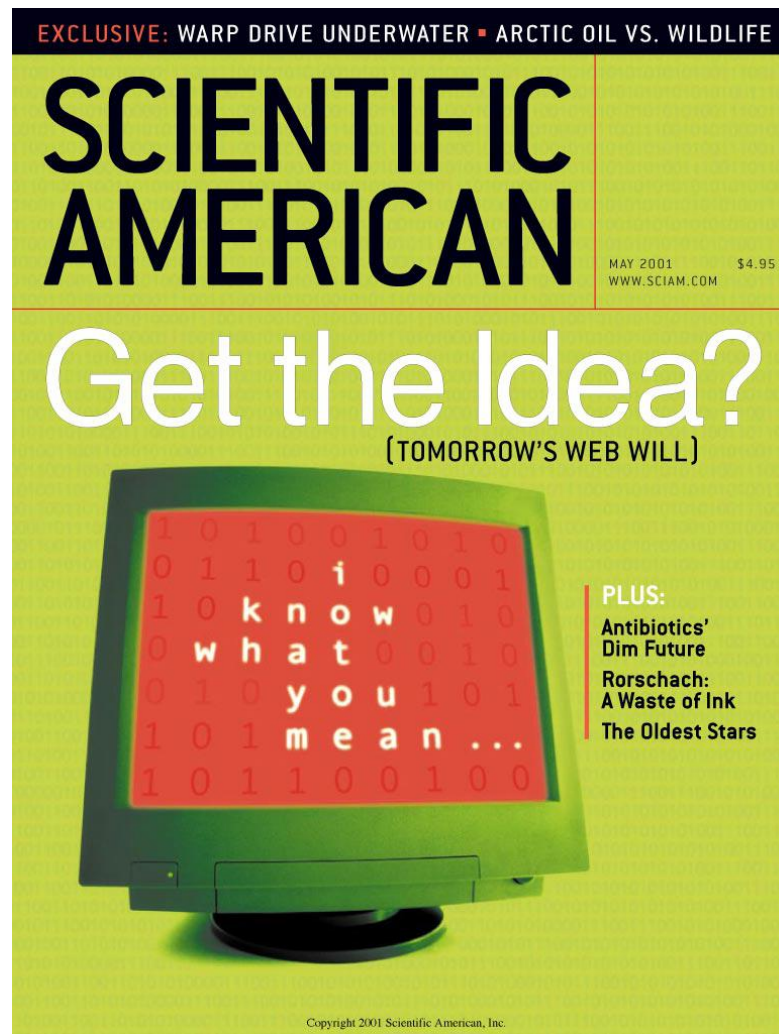




数据的权利是未来最大的权利



■ 语义网: Web3.0?



■ 语义网到底是什么？



机器而不是人去读网页？



语义网到底是什么？

语义网到底有什么用？

- 动态的数据建模，不需要什么schema或者shema可以快速演进（这是和数据库和XML比的好处）
- 让数据**流动**起来，而不是封固在一个个的网站或者应用里。
- 支持更多的**任务自动化**。(怎么做到？一些数据可机器处理，一些推理，一些查询)



- 语义网减少数据流动摩擦，加快数据流动速度的手段，也就是着眼减少人机界面的摩擦和机器与机器的摩擦这两个方面。
- 也就是：为了任务自动化
- 这两个方面都是依靠为网页添加可查询的注释。

以人为核心





语义网的本质

Web 任务自动化



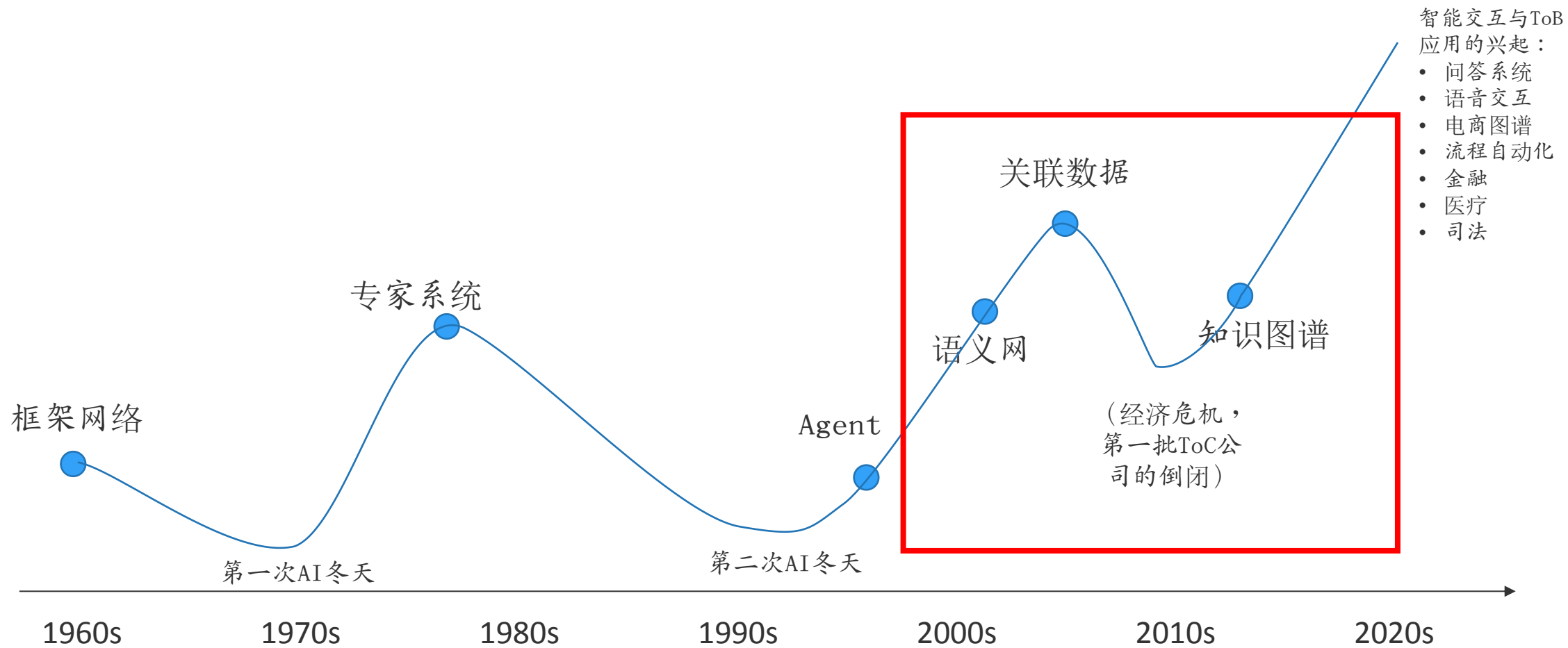
语义网的本质 (上一个世代的) DApp



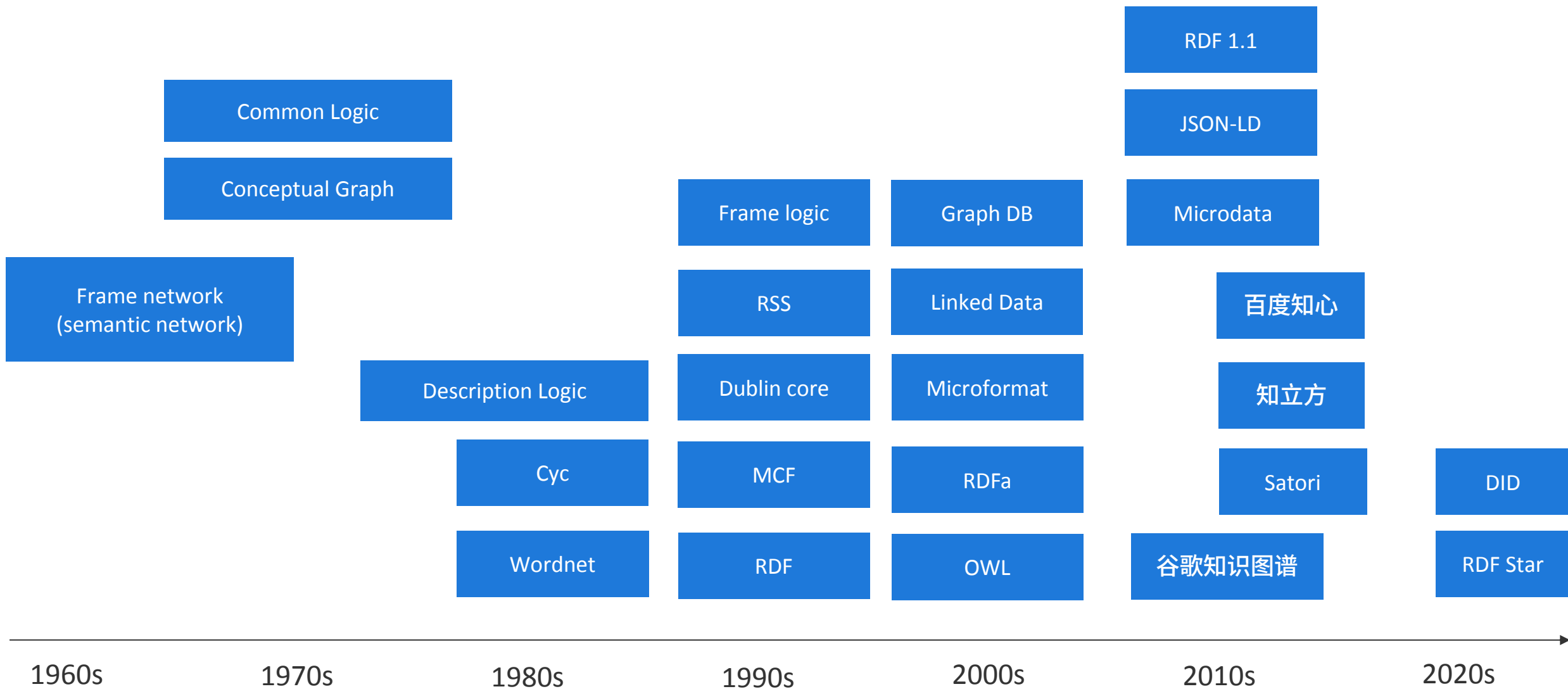
2

“语义”的兴衰

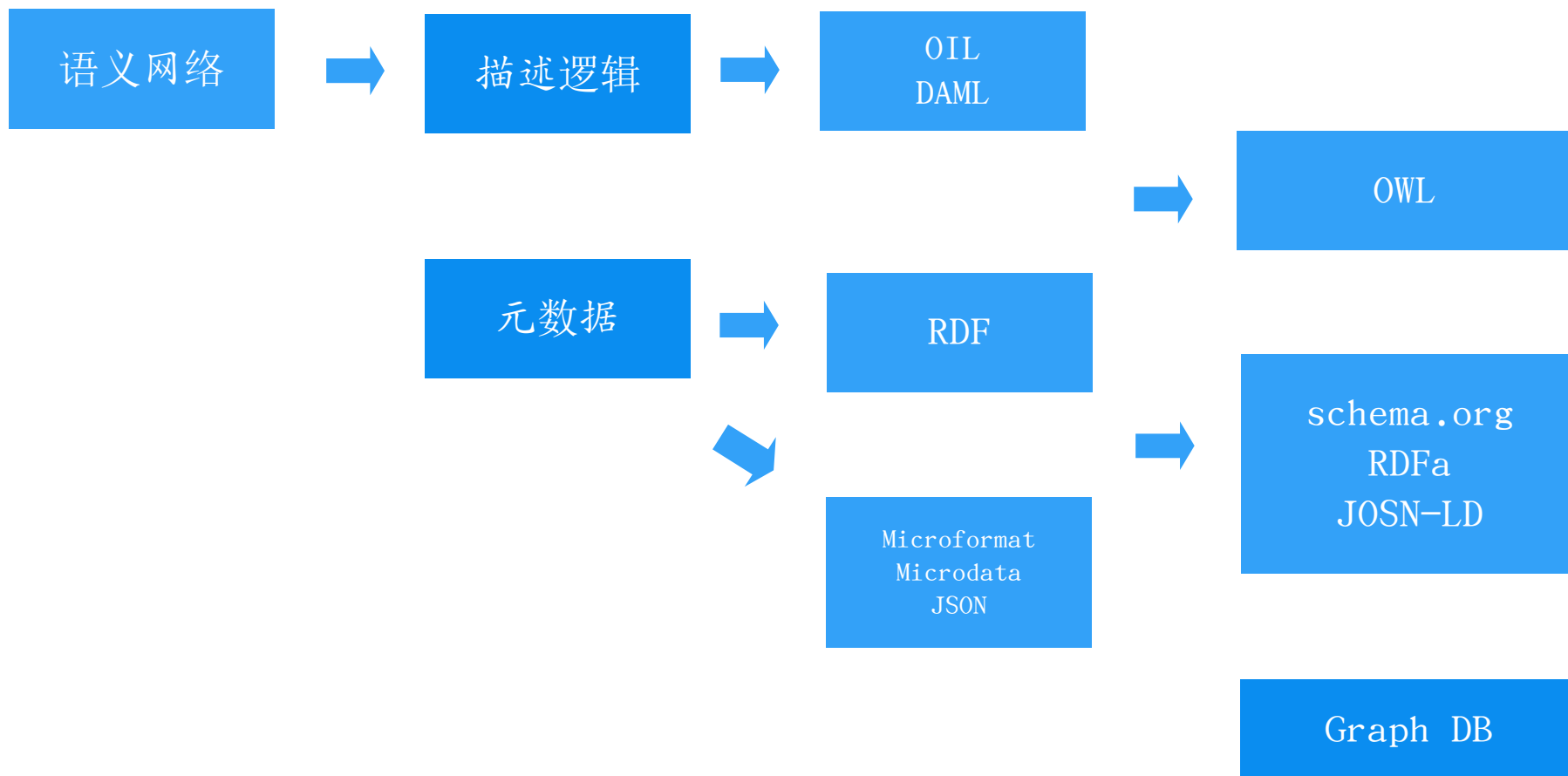




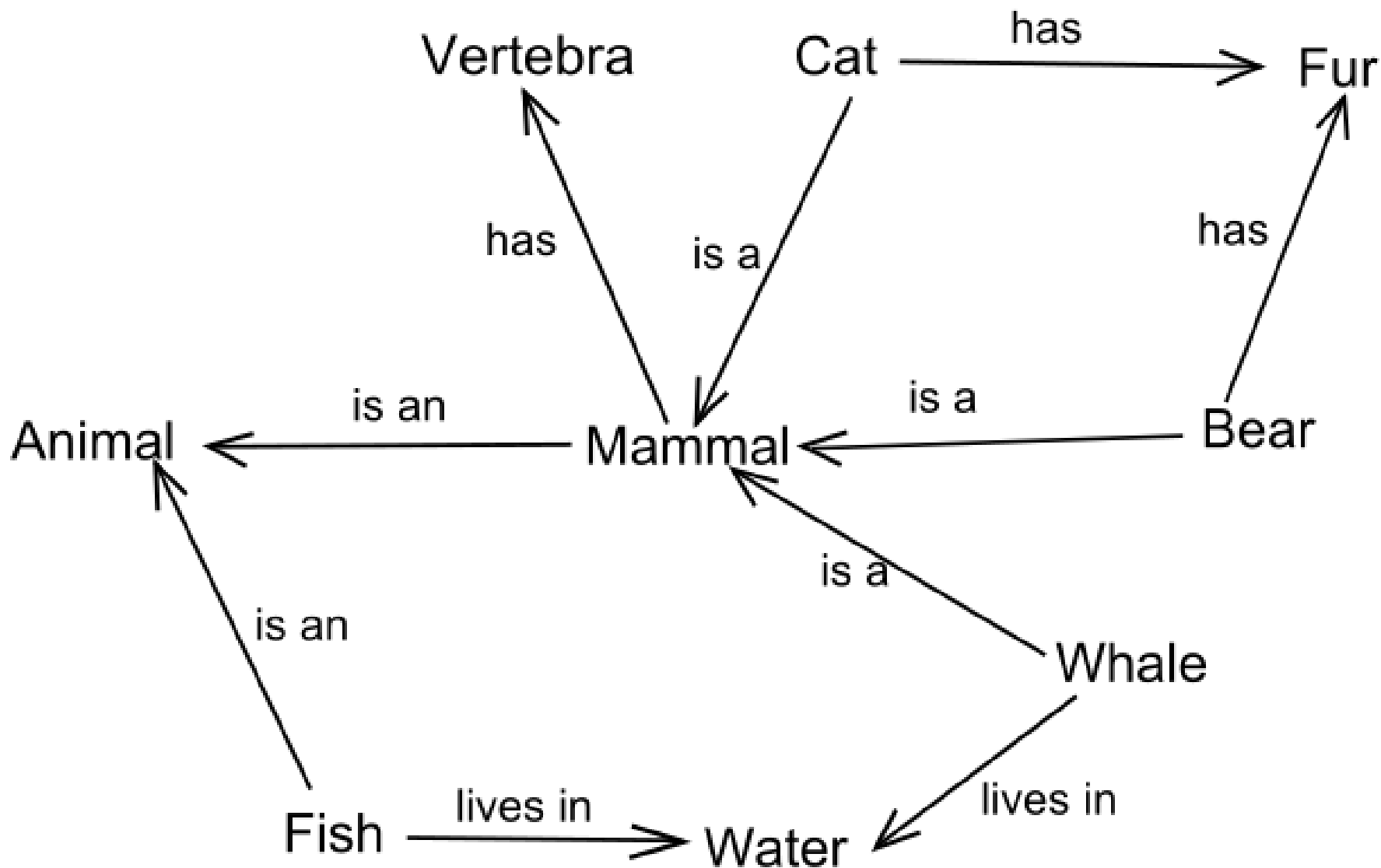
知识图谱的前世今身 (非完备列表)



知识图谱的演化 (简化)



语义网络 Semantic Network



■ 描述逻辑 Description Logic

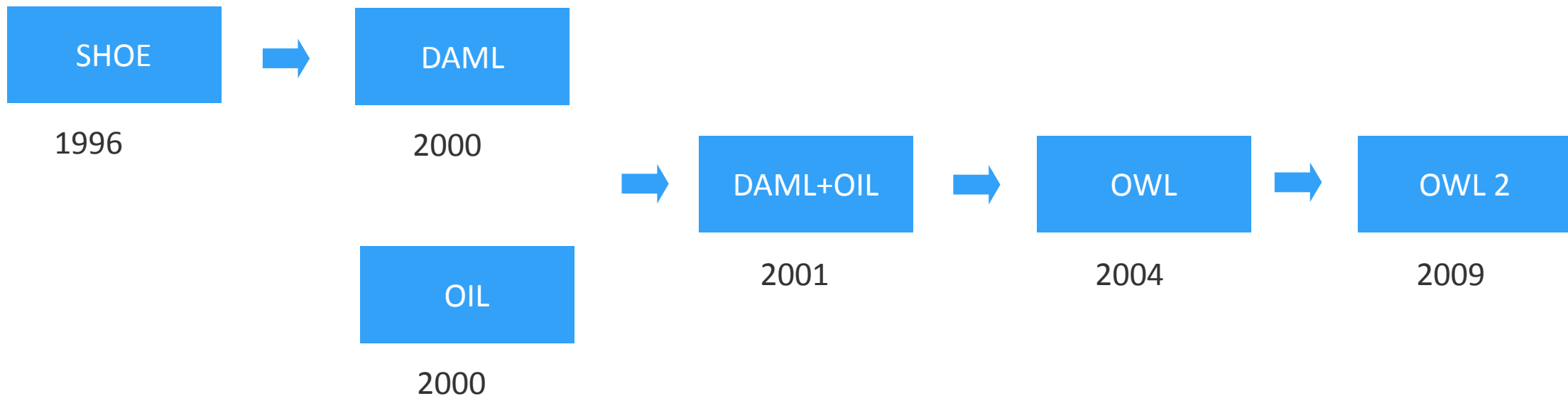
DL	FOL
$A \sqcup B$ $\neg A$ $A \sqcap B$	$A(x) \vee B(x)$ $\neg A(x)$ $A(x) \wedge B(x)$
$a : A \sqcap B$	$A(a) \wedge B(a)$
$\exists P.A$ $\forall P.A$	$\exists(y)(P(x, y) \vee A(y))$ $\forall(y)(P(x, y) \Rightarrow A(y))$
$(a, b) : P, P(a, b), aPb$	$P(a, b)$
$A \sqsubseteq B$ $A = B$	$\forall(x)(A(x) \Rightarrow B(x))$ $\forall(x)(A(x) \Leftrightarrow B(x))$
$P \sqsubseteq R$ $R = P^-, R = Inv(P)$	$\forall(x, y)(P(x, y) \Rightarrow R(x, y))$ $\forall(x, y)(P(x, y) \Rightarrow R(y, x))$

```
man::person.  
woman::person.  
brad:man.  
angelina:woman.  
person[hasSon=>man].  
brad[hasSon->>{maddox,pax}].  
married(brad,angelina).  
man(X) <- person(X) AND NOT woman(X).
```

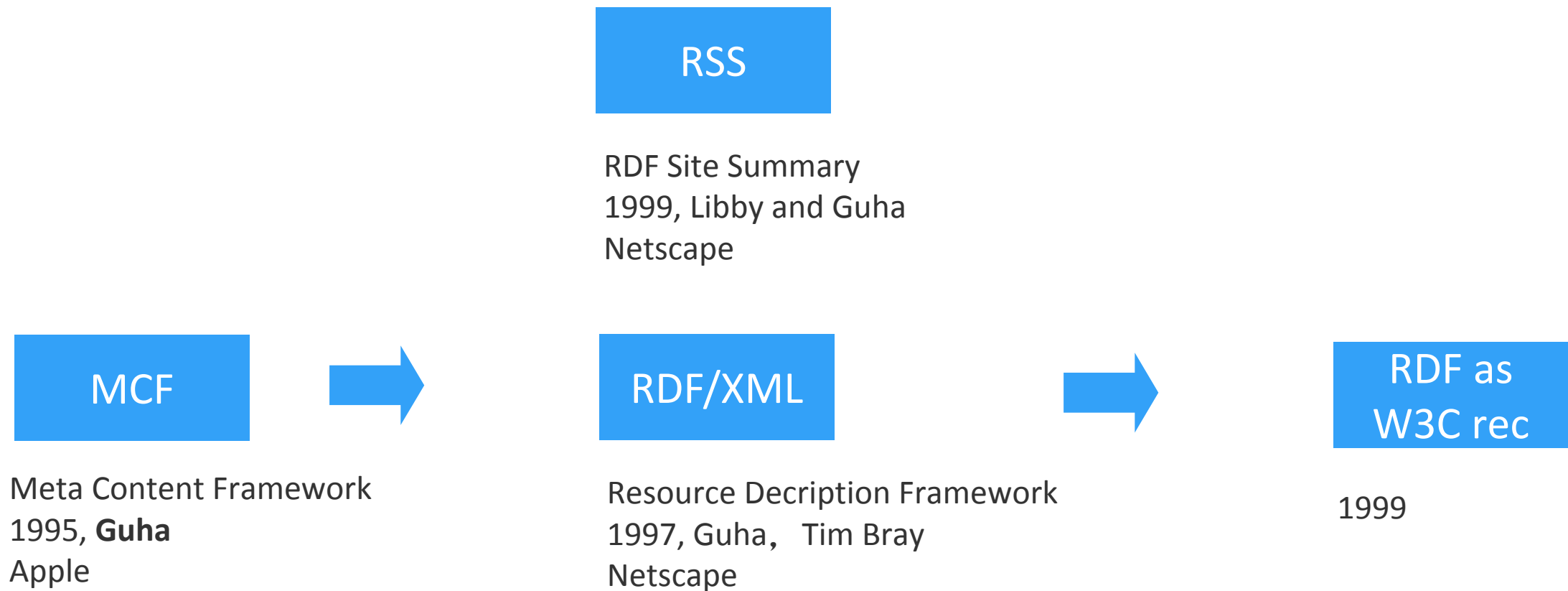
提供了另一种严格的“语义”（基于封闭世界假设）以支持推理。
可看作一阶逻辑的一个不可判定子集

后来在RIF中得到了实现

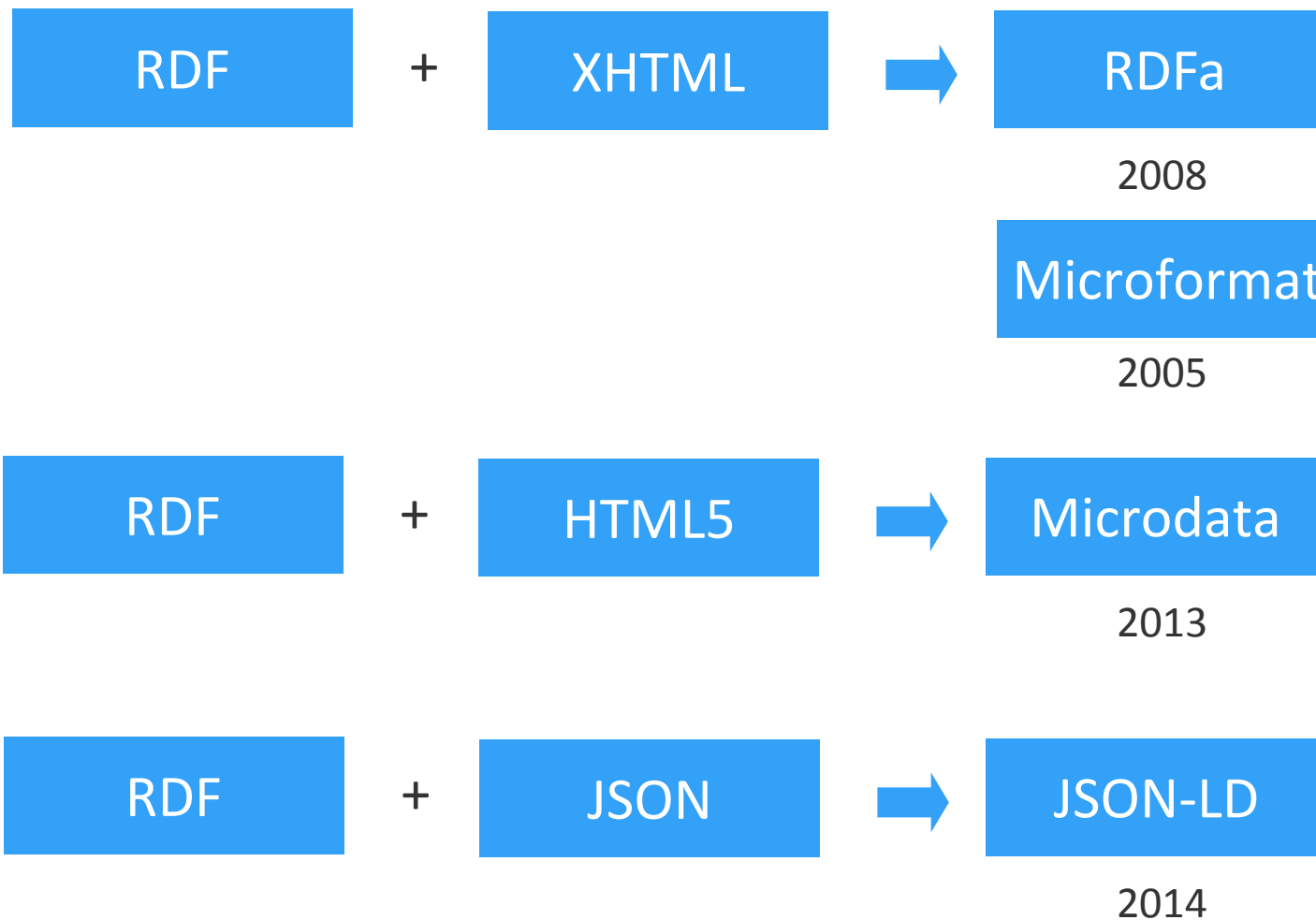
描述逻辑的HTML或XML语法



元数据框架到RDF

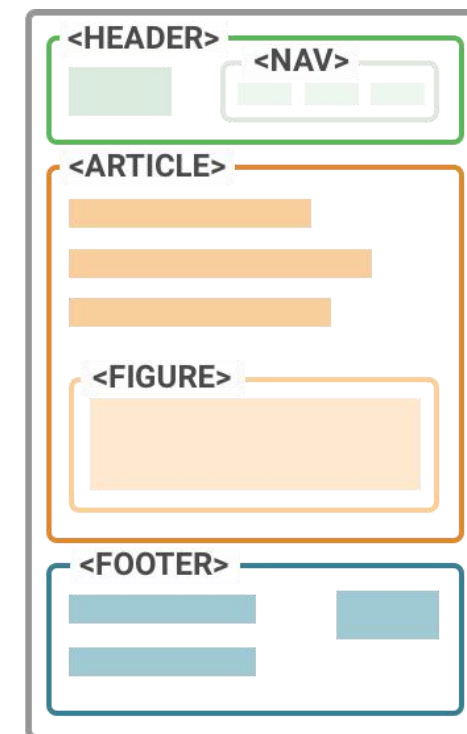


语法的发展

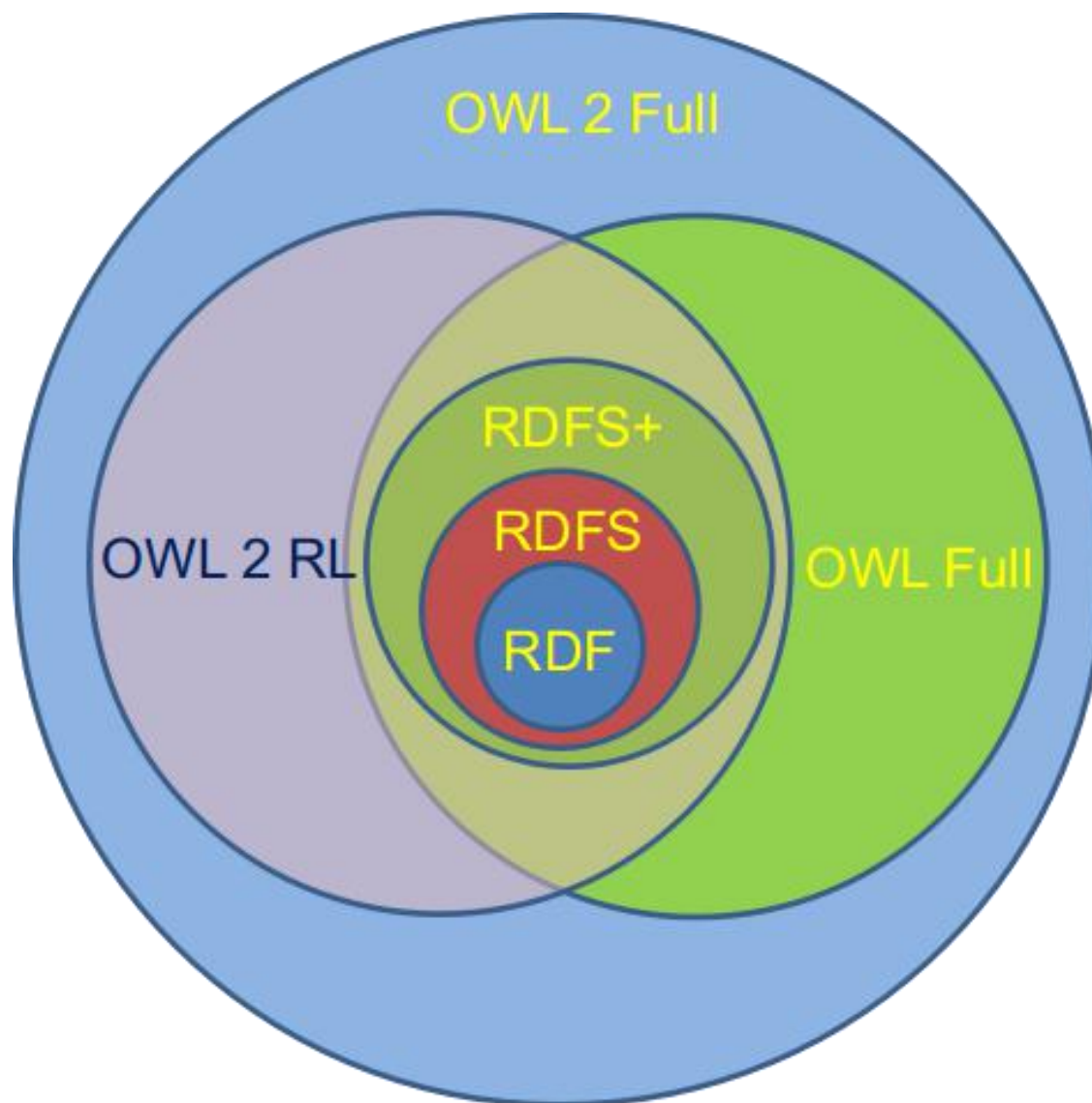


RDFa 到 HTML 5 Semantics

```
<div xmlns:v="http://rdf.data-vocabulary.org/#" typeof="v:Person">
  <span property="v:name">John Smith</span>
  <span property="v:nickname">Smithy</span>
  <span property="v:url">http://www.example.com</span>
  <span property="v:affiliation">ACME</span>
  <span rel="v:address">
    <span property="v:locality">Albuquerque</span>
  </span>
  <span property="v:title">Engineer</span>
  <a href="http://darryl-blog.example.com/" rel="v:friend">Darryl</a>
</div>
```

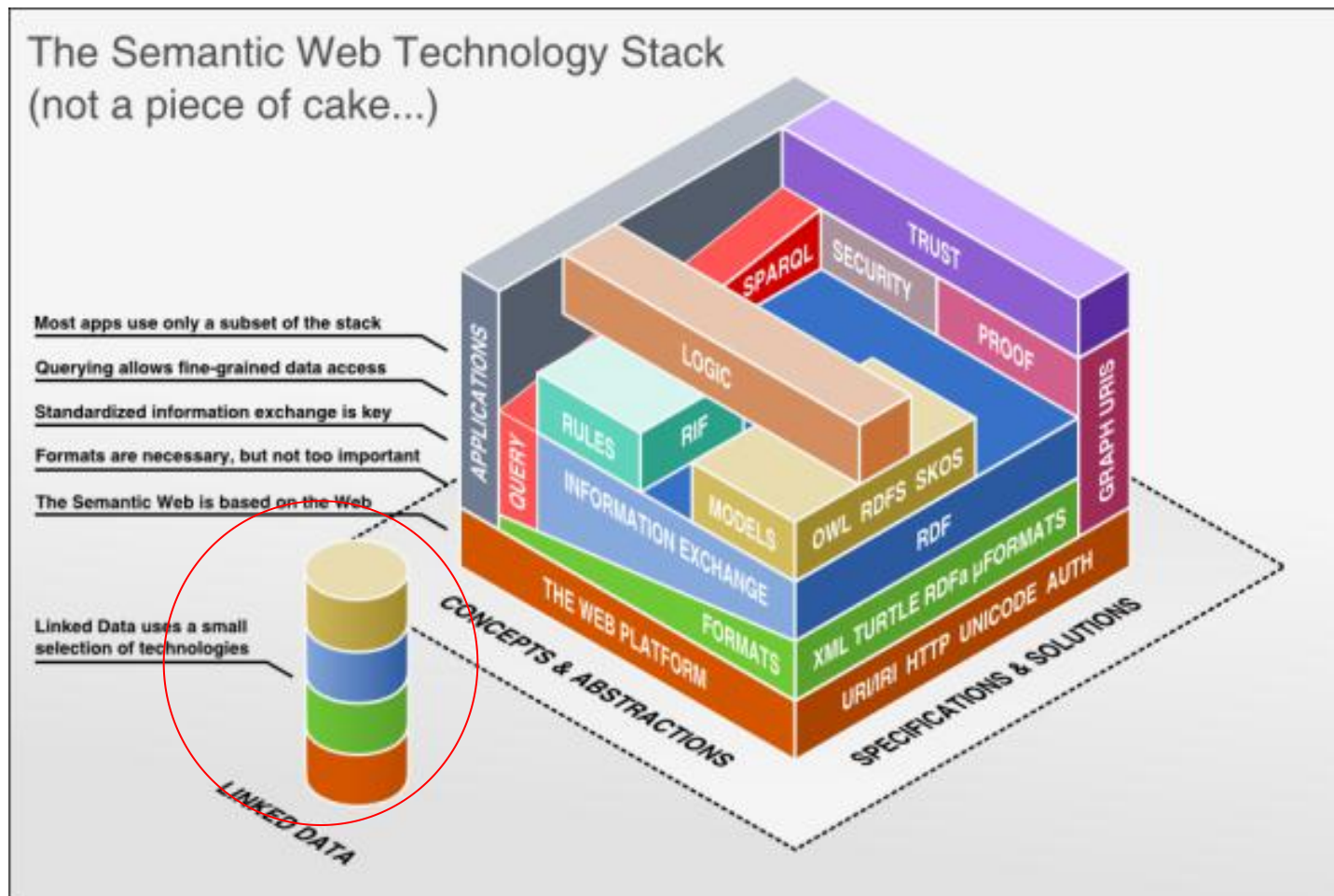


知识图谱语言：RDF and OWL



■ 关联数据 Linked Data

到了2006年，语义网技术堆栈又已经复杂到没多少人看得懂了，于是：



Tim Berners-Lee

Date: 2006-07-27, last change: \$Date: 2009/06/18 18:24:33 \$

Status: personal view only. Editing status: imperfect but published.

[Up to Design Issues](#)

Linked Data

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.

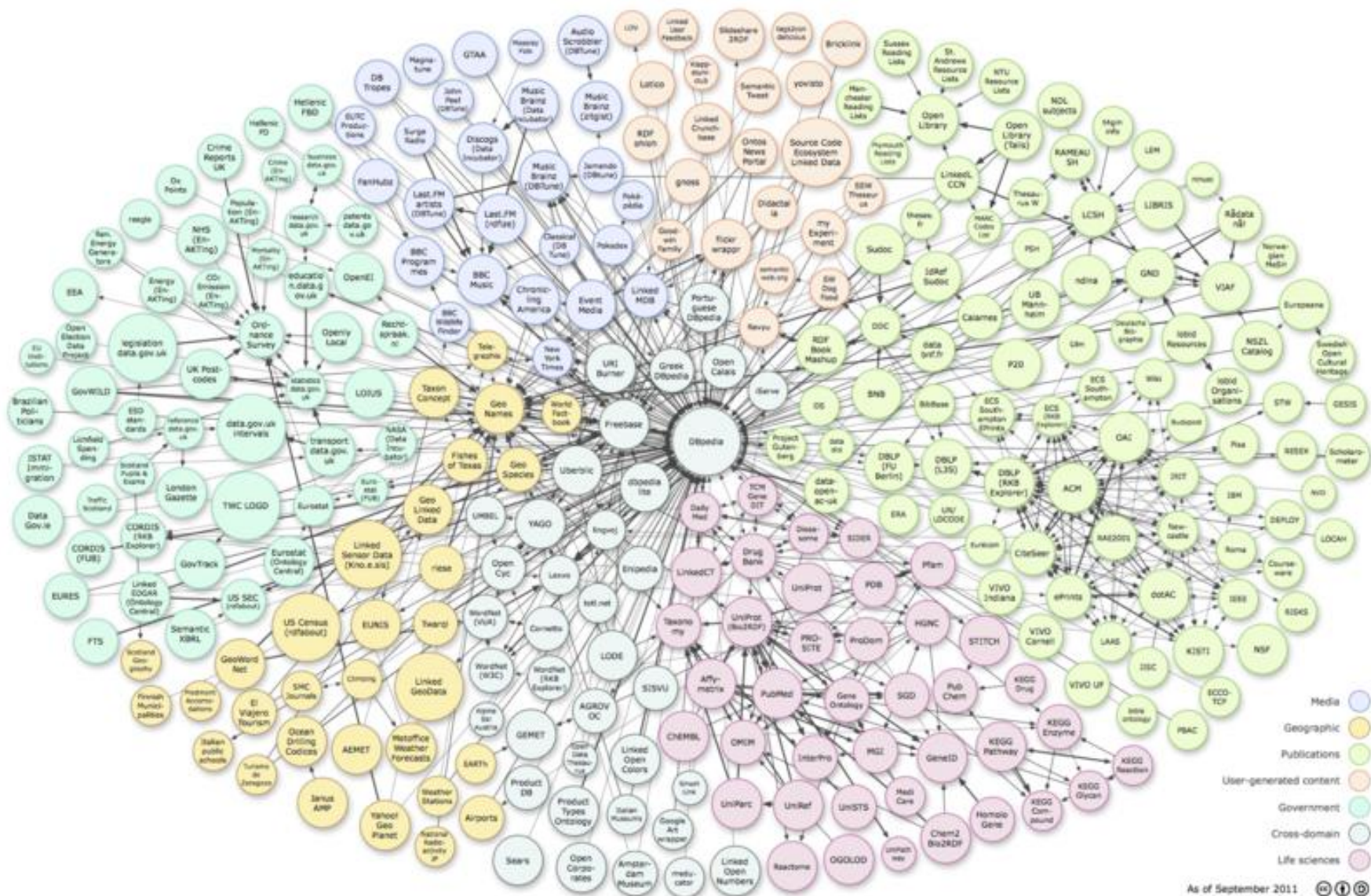
Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF,. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.



<http://www.w3.org/DesignIssues/LinkedData.html>

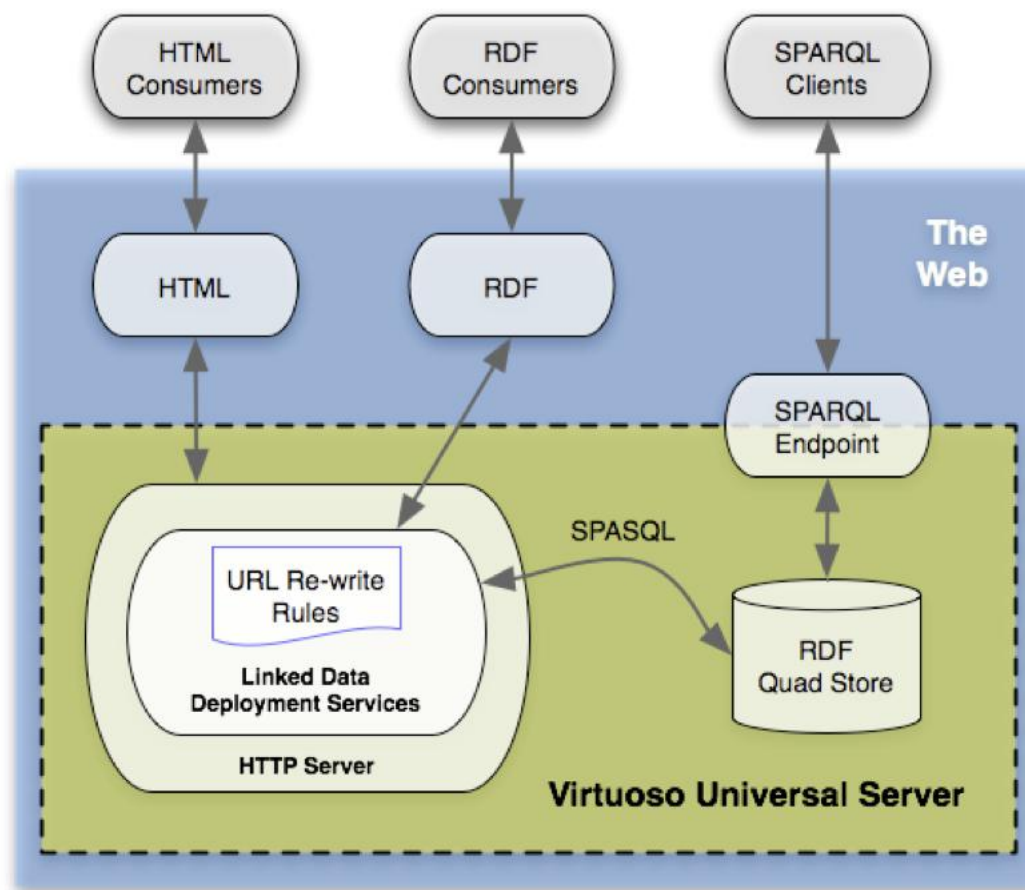
互联数据云 (2011)



DBpedia



- 从wikipedia抽取
- 英文版有4.58m个实体
- 基于RDF/SPARQL
- 广泛的成功应用

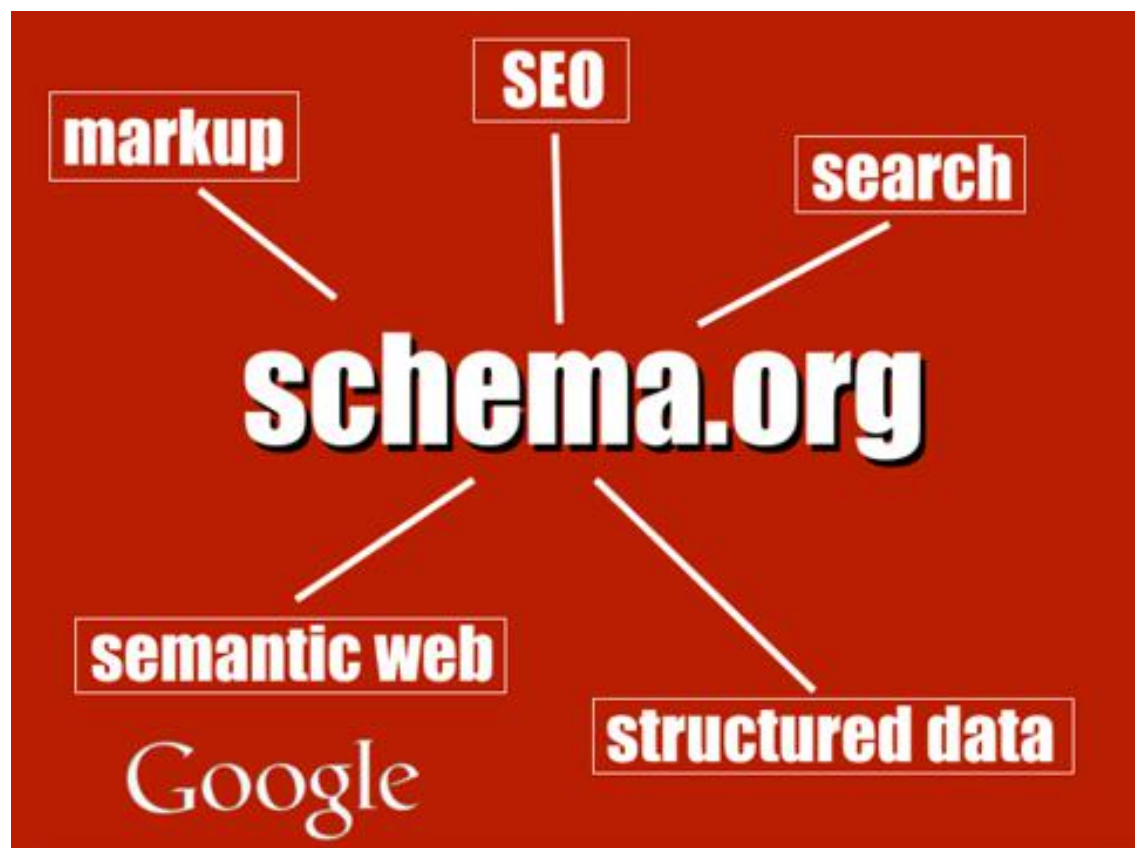


Freebase (谷歌知识图谱的前身)

The screenshot shows the Freebase website interface. At the top, there's a navigation bar with the Freebase logo, a search bar, and links for Browse, Query, Help, Sign In or Sign Up, and English. Below the navigation bar, a large orange banner reads "Important! Freebase is read-only and will be shut-down. More." In the center, a large black box displays the number "3,123,054,696" with the text "Facts (and counting)" next to it. Below this, a blue banner states "A community-curated database of well-known people, places, and things". A dark navigation bar contains links for Data, Schema, Queries, Apps, Loads, Review Tasks, and Users. The "Data" link is highlighted. Below the navigation bar, the "Explore Freebase Data" section features a table with columns for Domain, ID, Topics, and Facts. The table lists various domains like Music, Books, Media, People, Film, Location, and TV. To the right of the table, there's a sidebar with a yellow banner "How can you get started?", a section "Learn how it works" with a description of the database, and a section "Use Freebase data" with information about the open license and a list of actions like "Query Freebase using our Search". At the bottom, there are links for Terms of Service, How to Attribute to Freebase, and Feedback, along with the copyright notice "© 2015 Google" and links for View Source and Clear Cache.

Domain	ID	Topics	Facts
Music	/music	32M	232M
Books	/book	6M	15M
Media	/media_common	5M	17M
People	/people	4M	20M
Film	/film	2M	22M
Location	/location	2M	20M
TV	/tv	2M	19M

<http://www.freebase.com/>



+Dave Search Images Maps Play YouTube News Gmail Documents Calendar More

Google
marie curie

Search
8 personal results, 17,300,000 other results (0.32 seconds)

Everything
Images
Maps
Videos
News
Shopping
Books
More

Mountain View, CA
Change location

Any time
Past hour
Past 24 hours
Past 2 days
Past week
Past month
Past year
Custom range...

[Marie Curie - Wikipedia, the free encyclopedia](#)
[en.wikipedia.org/wiki/Marie_Curie](#)
Marie Skłodowska-Curie (7 November 1867 – 4 July 1934) was a French-Polish physicist and chemist famous for her pioneering research on radioactivity.
 ↳ Marie Curie Cancer Care - Pierre Curie - Irène Joliot-Curie - Radioactive decay

[Marie Curie - Biography - Nobelprize.org](#)
[www.nobelprize.org/nobel_prizes/physics/_/marie-curie-bio.html](#)
 Short profile from the foundation that awards the Nobel Prize.

[Marie Curie - Biography - Nobelprize.org](#)
[www.nobelprize.org/nobel_prizes/chemistry/_/marie-curie-bio.html](#)
Marie Curie, née Maria Skłodowska, was born in Warsaw on November 7, 1867, the daughter of a secondary-school teacher. She received a general education ...

[Images for marie curie](#) - Report images

[Marie Curie and The Science of Radioactivity](#)
[www.aip.org/history/curie/](#)
 The life of **Marie Curie**, from the AIP Center for History of Physics. Text by Naomi Pasachoff and many illustrations describe Curie's contributions to the science of ...

Marie Curie

Marie Skłodowska-Curie was a French-Polish physicist and chemist famous for her pioneering research on radioactivity. She was the first person honored with two Nobel Prizes—in physics and chemistry. Wikipedia

Born: November 7, 1867, Warsaw

Died: July 4, 1934, Sancellemoz

Spouse: Pierre Curie (m. 1895–1906)

Children: Irène Joliot-Curie, Ève Curie

Discovered: Radium, Polonium

Education: École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris, University of Paris

People also search for

Albert Einstein

Pierre Curie

Ernest Rutherford

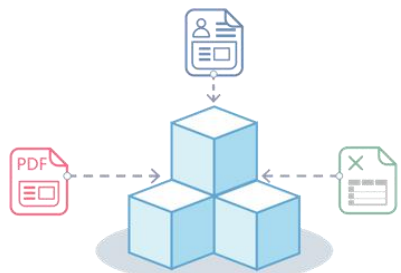
Louis Pasteur

John Dalton

[Report a problem](#)

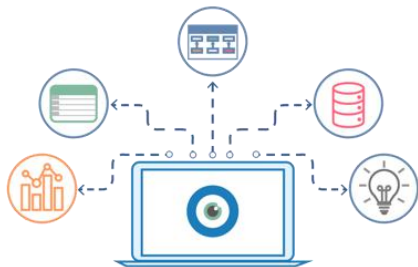
知识图谱：2012年的汇集

知识图谱（Knowledge Graph）是一种应用于诸多人工智能相关领域的关键技术，主要应用在数据结构化处理、解析、关联以及后续的分析与推理。



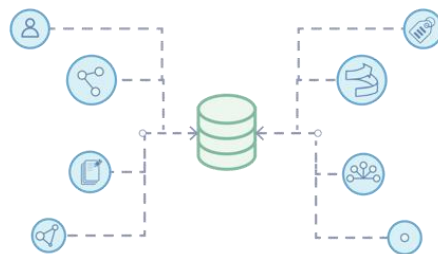
知识提取

利用自然语言处理、机器学习、模式识别等解决结构化数据生成问题。



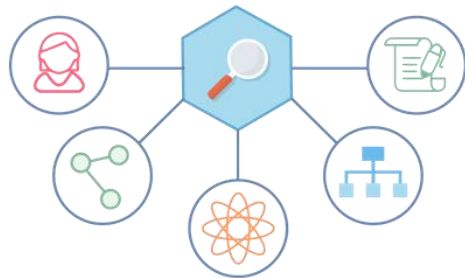
知识表现和推理

组织结构化数据，使得机器能够自动化处理。



知识存储

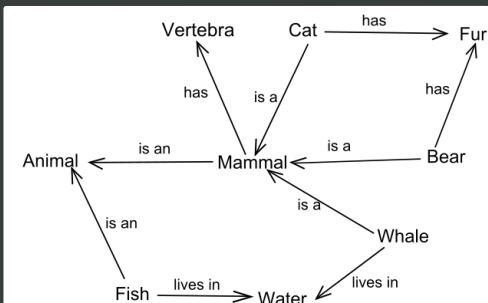
进行大量的结构化数据管理，同时混合管理结构化和非结构化数据。



知识检索

用语义技术提高搜索与查询的精准度，为用户展现最合适的信息。

1960s



语义网络

2001



语义网, OWL

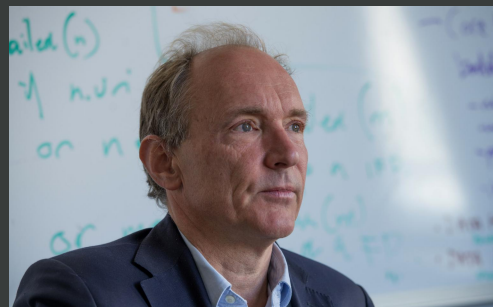
Jim Hendler

“语义网之父”

RPI教授

美国白宫Web顾问

2006



互联数据

Tim Berners-Lee

Web之父

W3C主席

MIT教授

2016图灵奖得主

2012



知识图谱

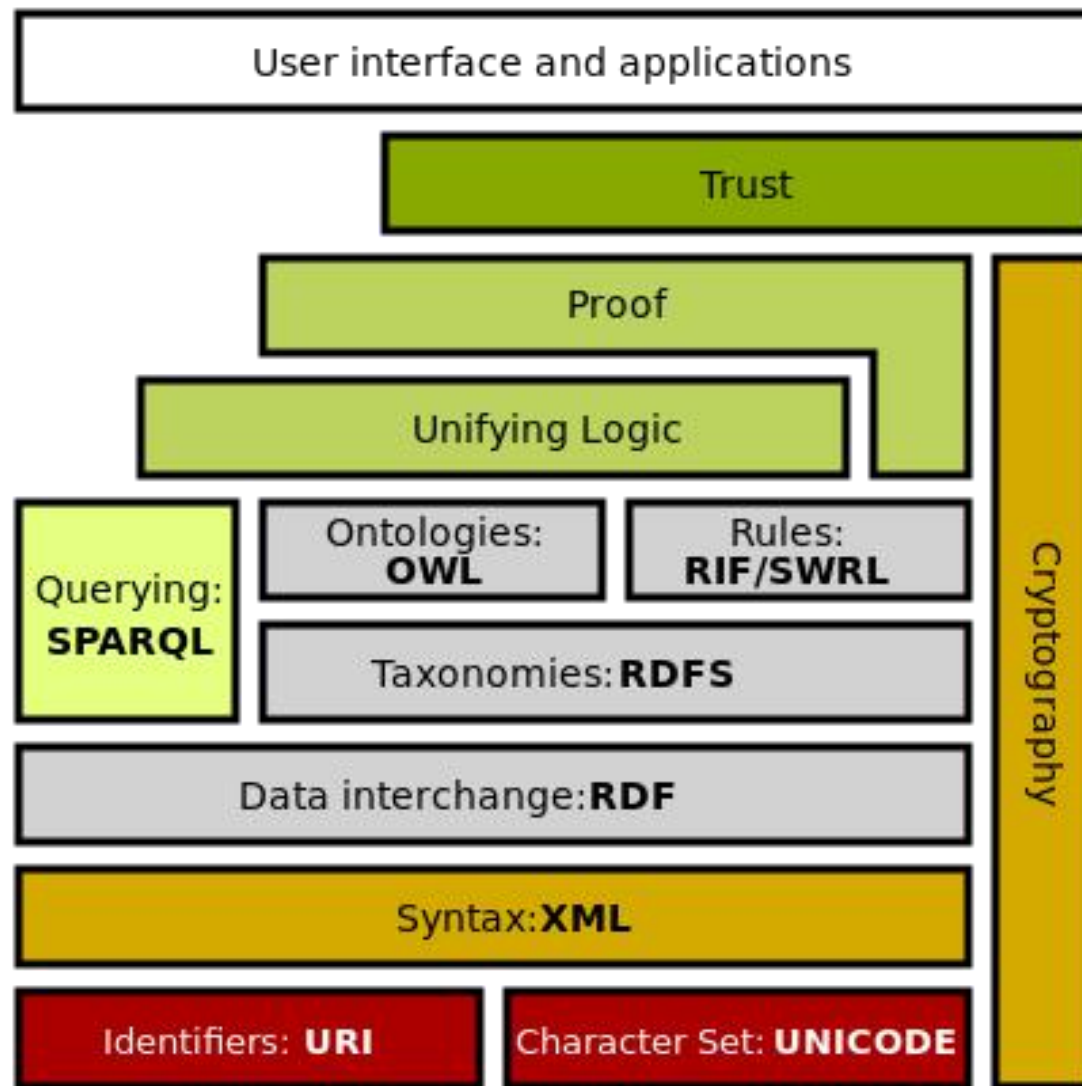
谷歌



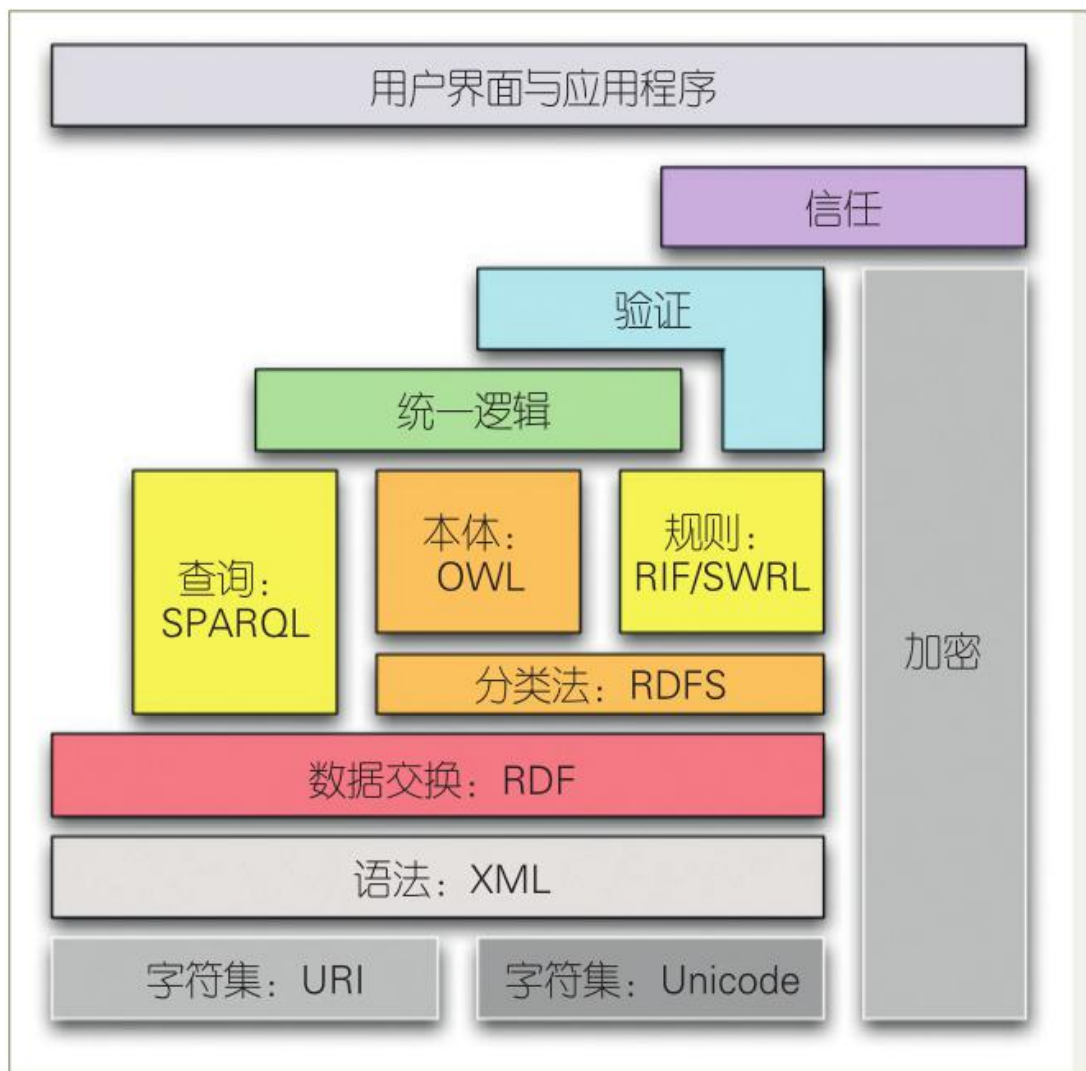
3

回顾语义网层次蛋糕





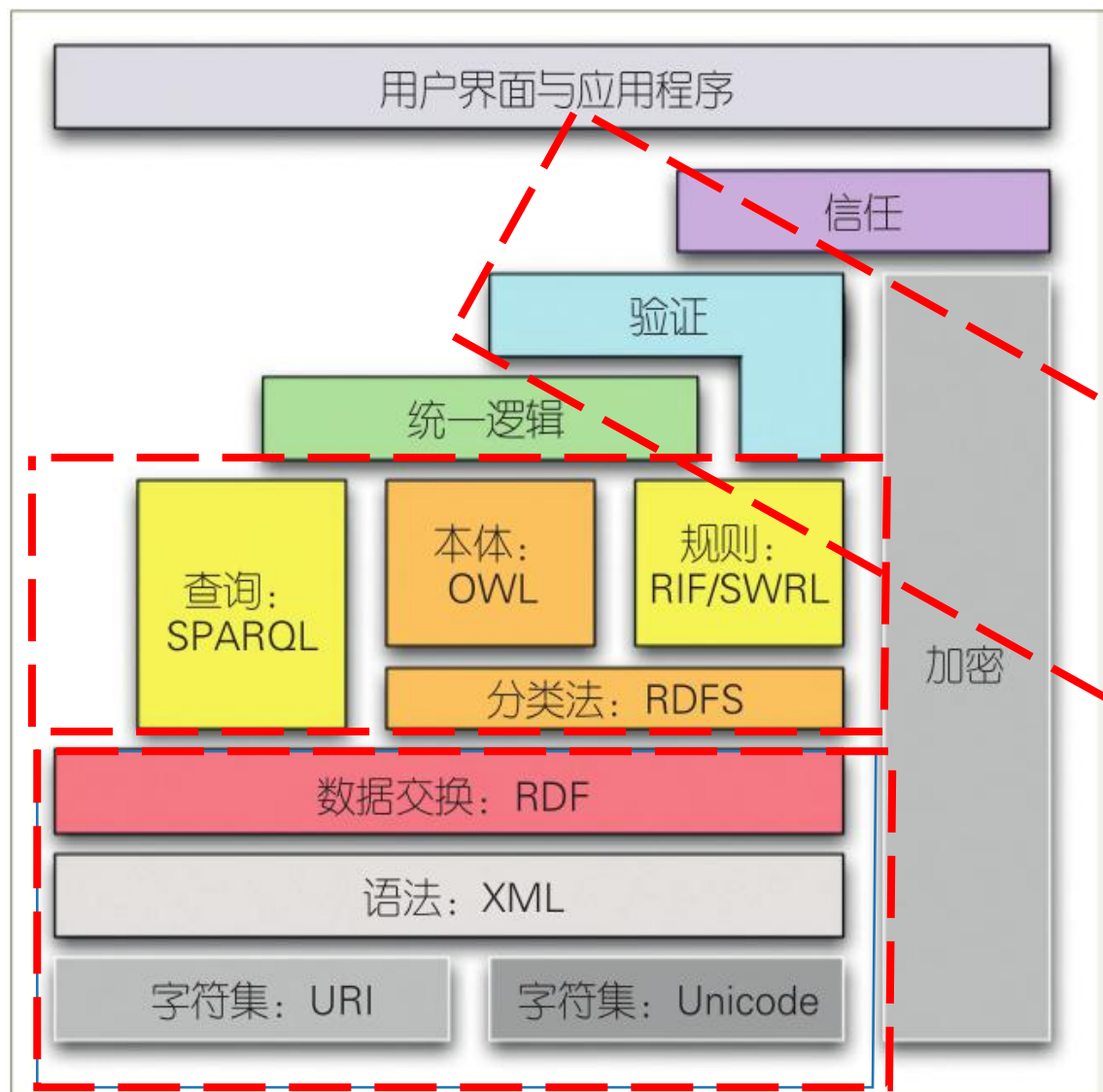
W3C语义网技术栈（中文版）



语义网技术被三个技术领域继承

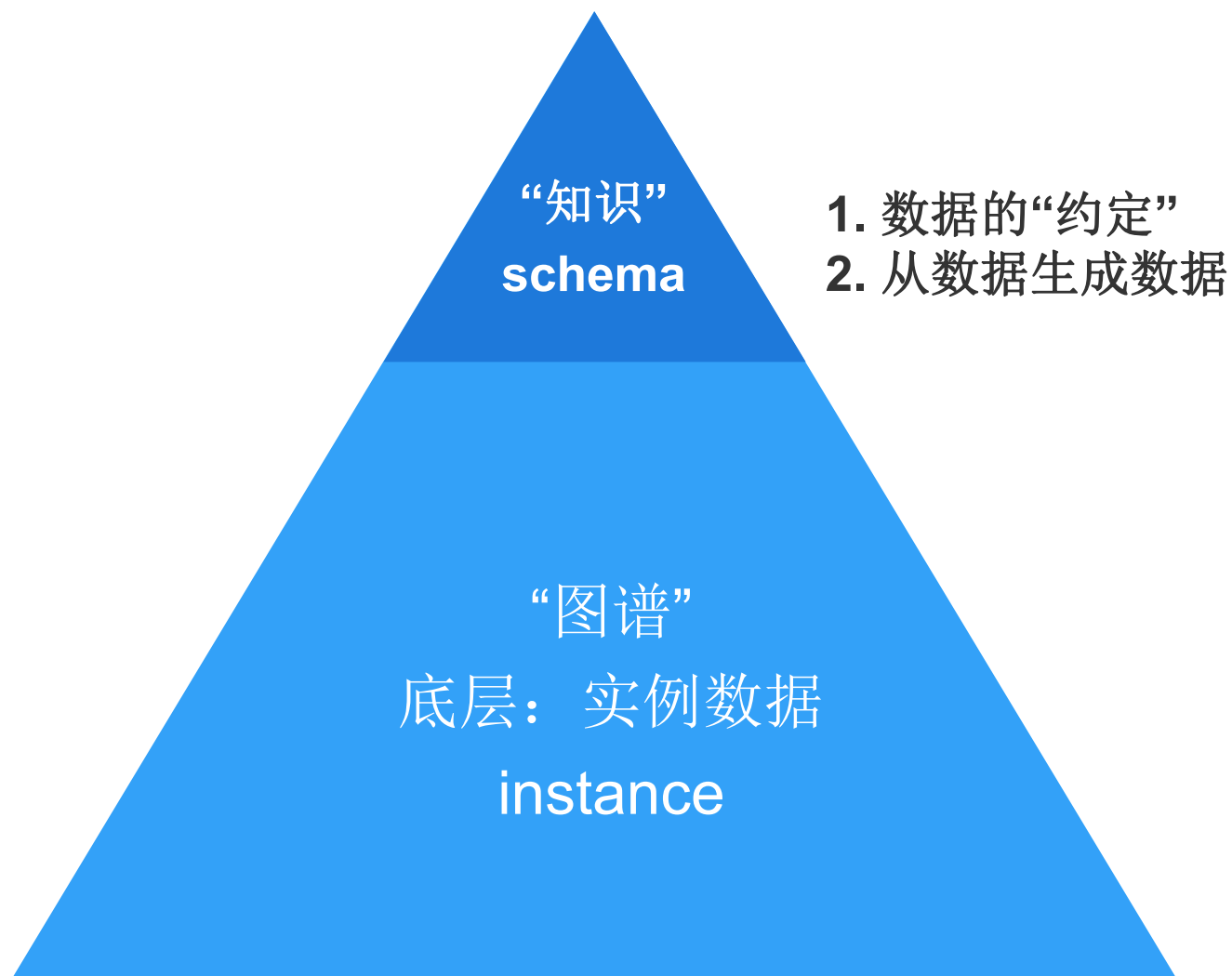
知识技术

图谱技术

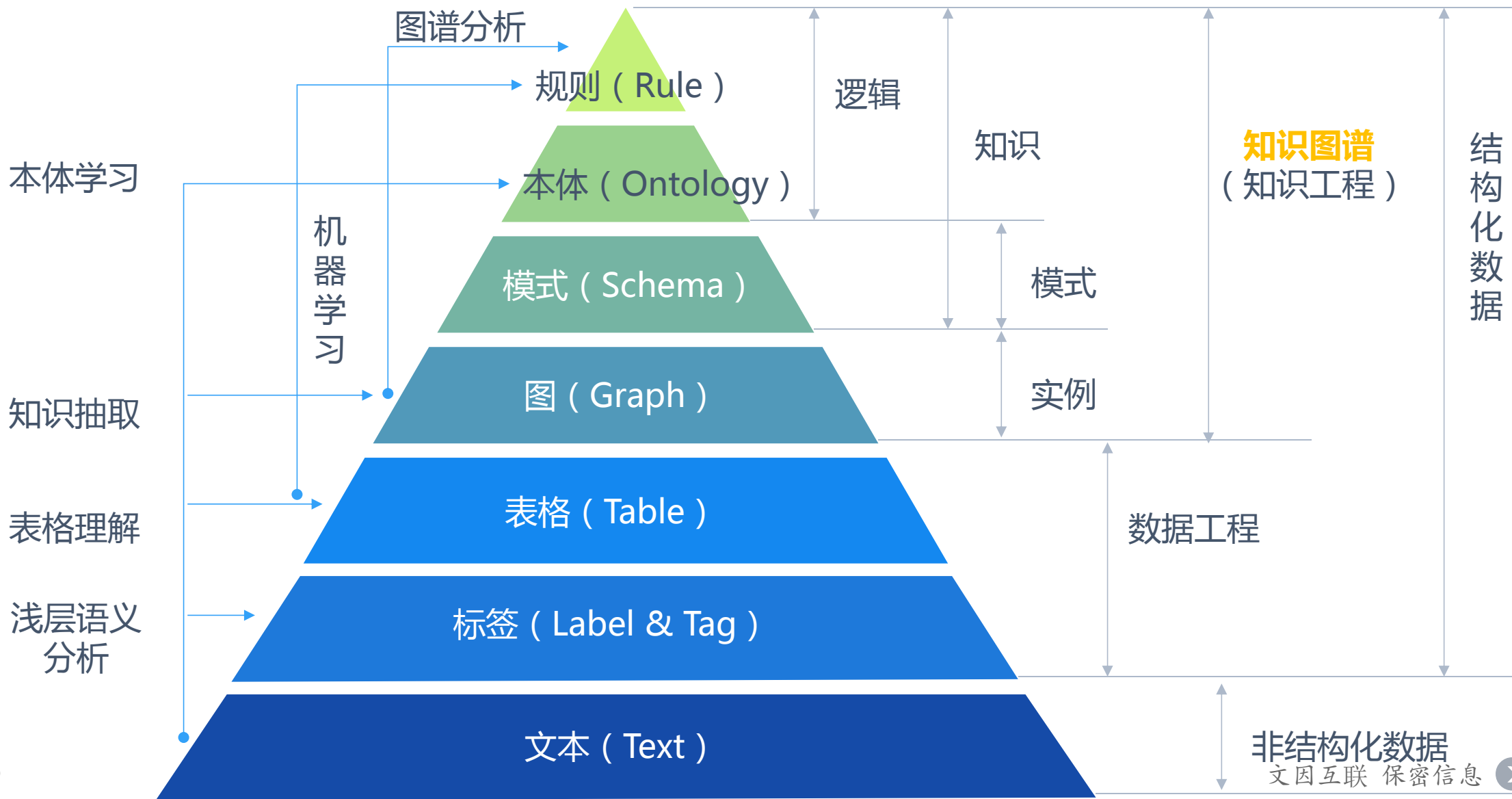


互信技术

知识图谱的层级 = 知识 + 图谱



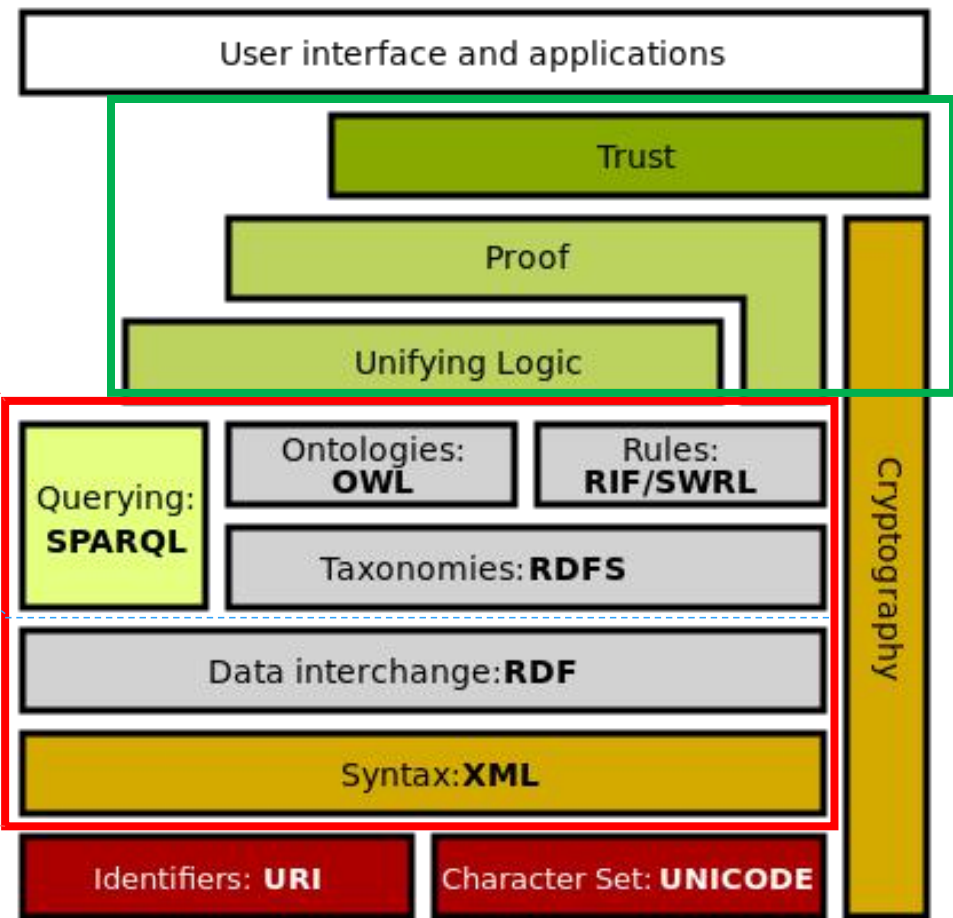
知识图谱技术金字塔



两种知识图谱技术栈范式之间的关系



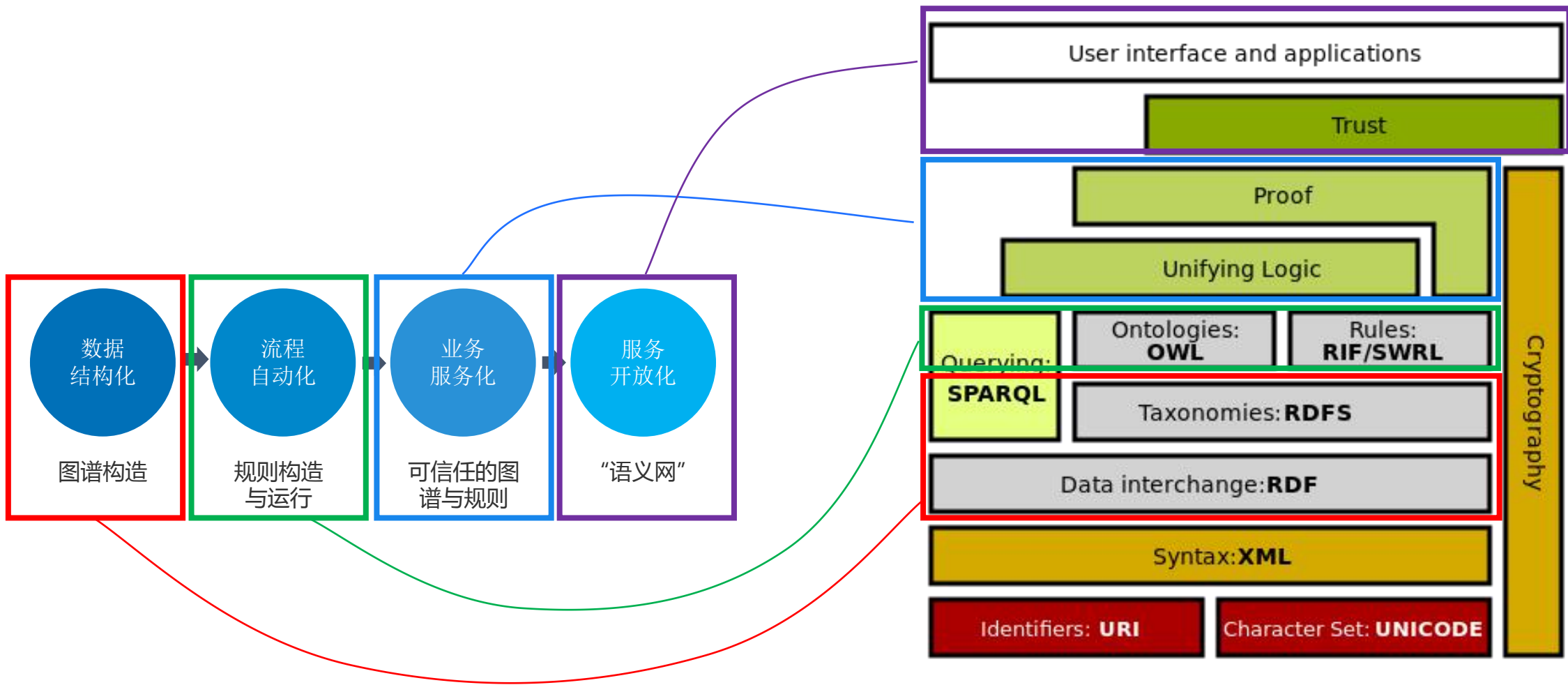
认知智能技术金字塔



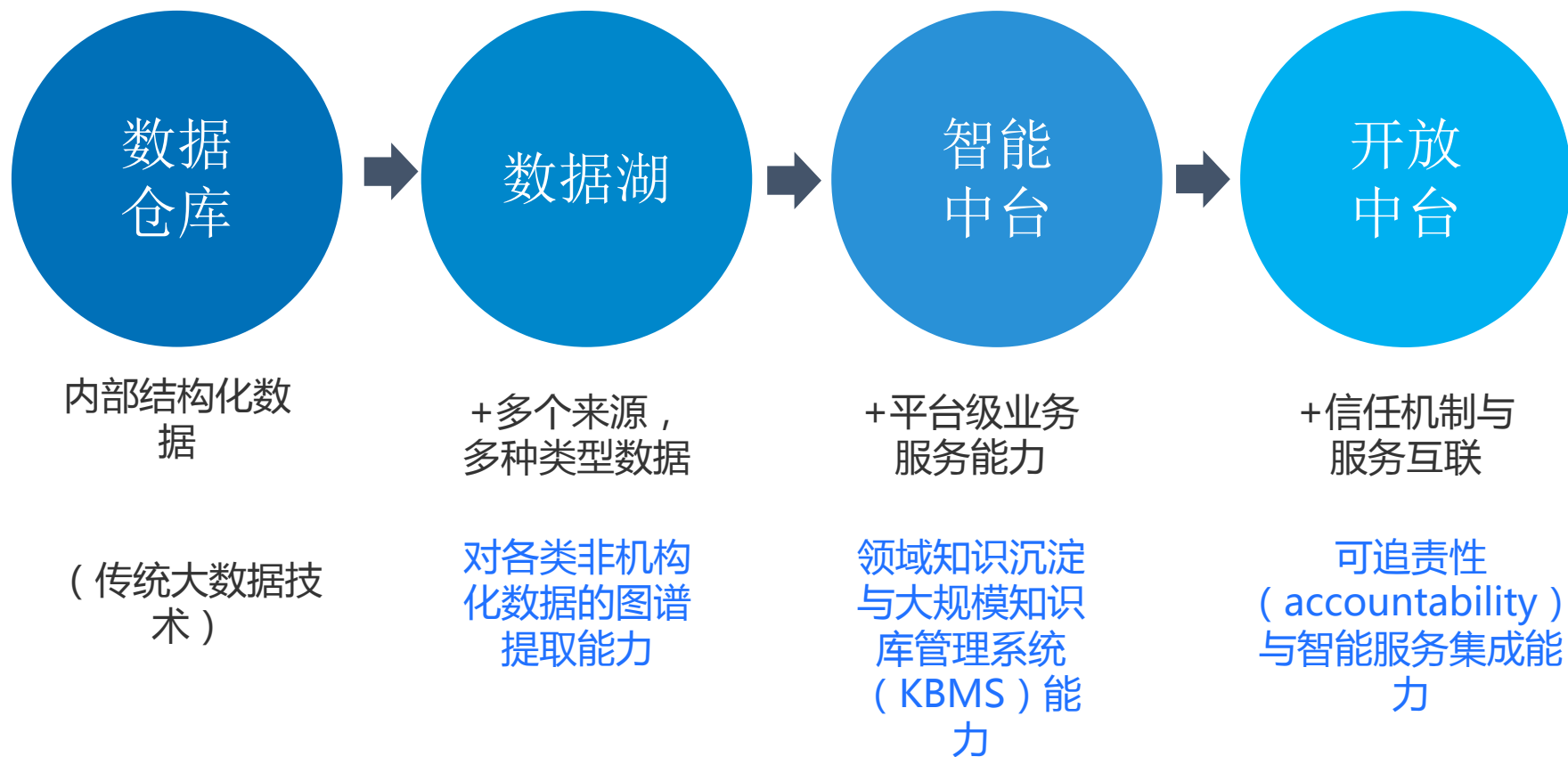
这部分融合了“区块链”

W3C语义网技术栈

知识图谱与企业数智化改造



知识图谱与大数据的关系





4

Web 3.0

——另一种“语义网”



- 货币（价值）将成为互联网的固有特征
- 去中心化应用（DApp）为用户提供新的功能
- 数字身份和数据资产权利

数字资产市场经济

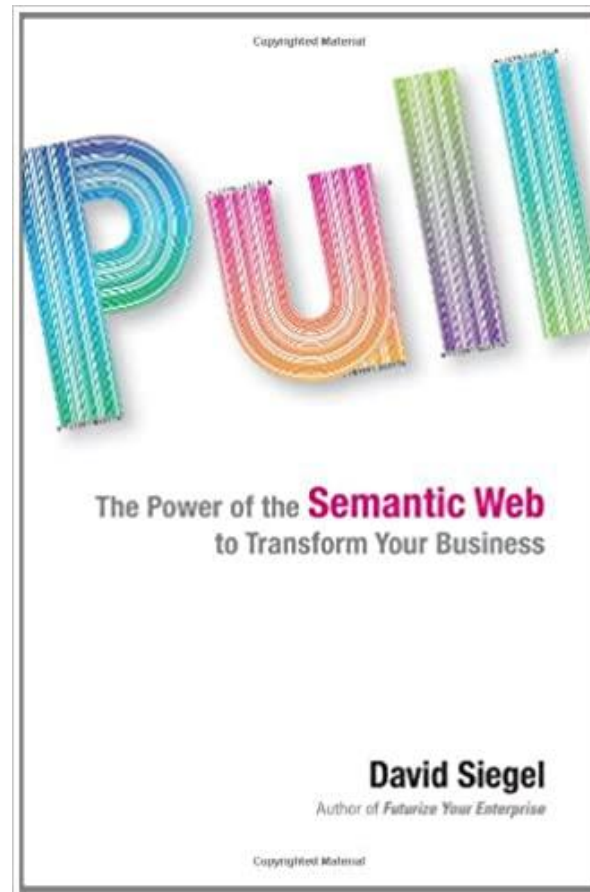
财产私有制度促进了产业的分工
知识产权保护制度促进了知识的传播
数字资产权利将催生全新的产业

■ Web 3.0的应有本质： 自由的知识交易



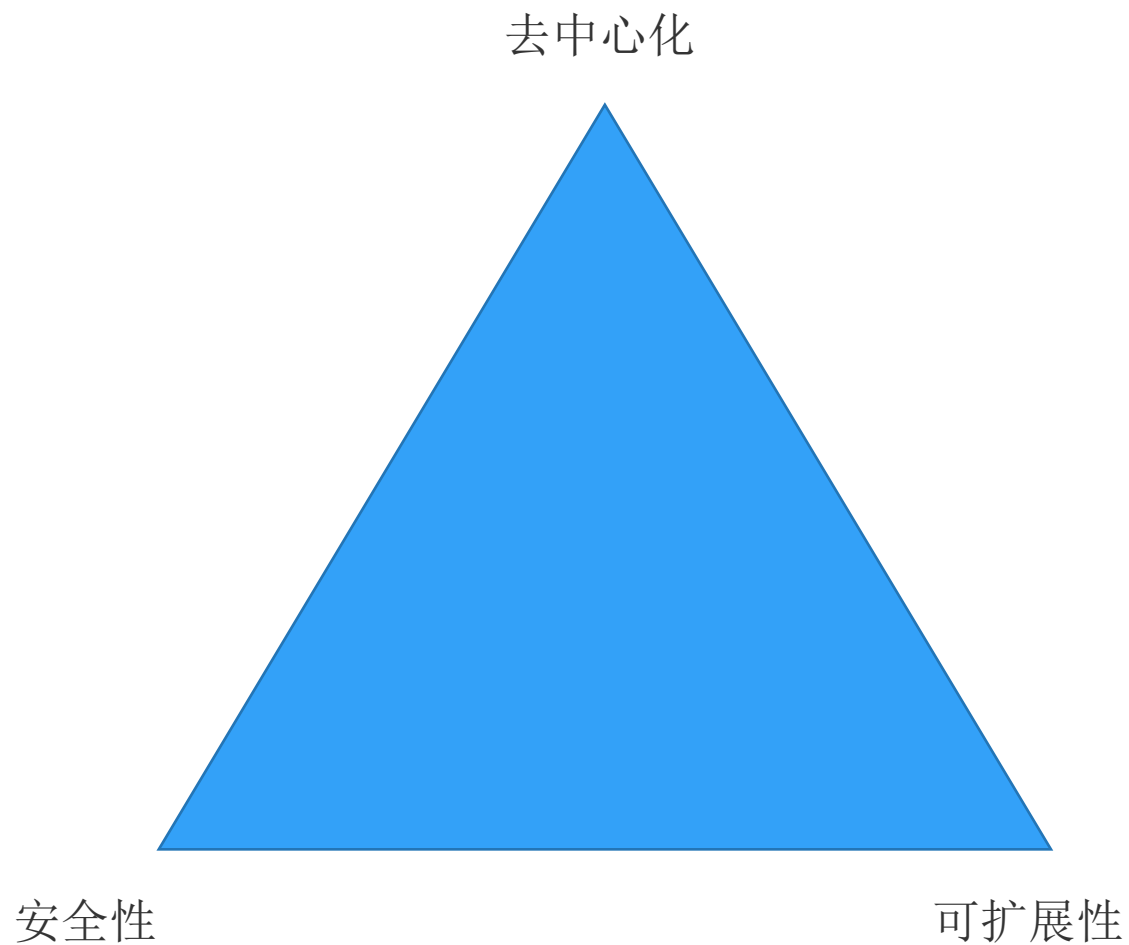
David Siegel

author of Pull: The Power of the Semantic Web to Transform Your Business



拉引擎：激活个人数据仓库，自动化数据的智能推送

Web 3.0的核心属性是去中心化吗？



强中心化往往表现为“去中心化”



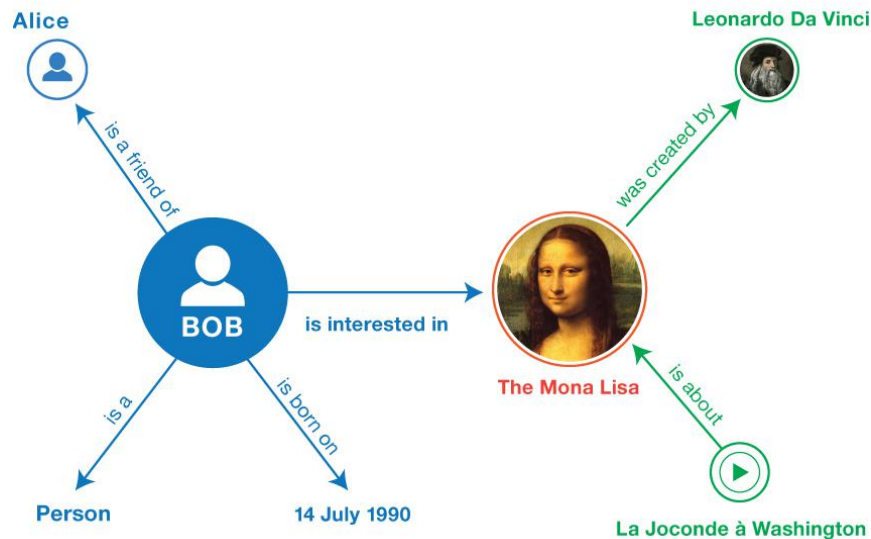
2008年：两种互信技术“同时”出现



区块链



Satoshi Nakamoto



AIR（基于RDF的可追责语言）



Tim Berners-Lee

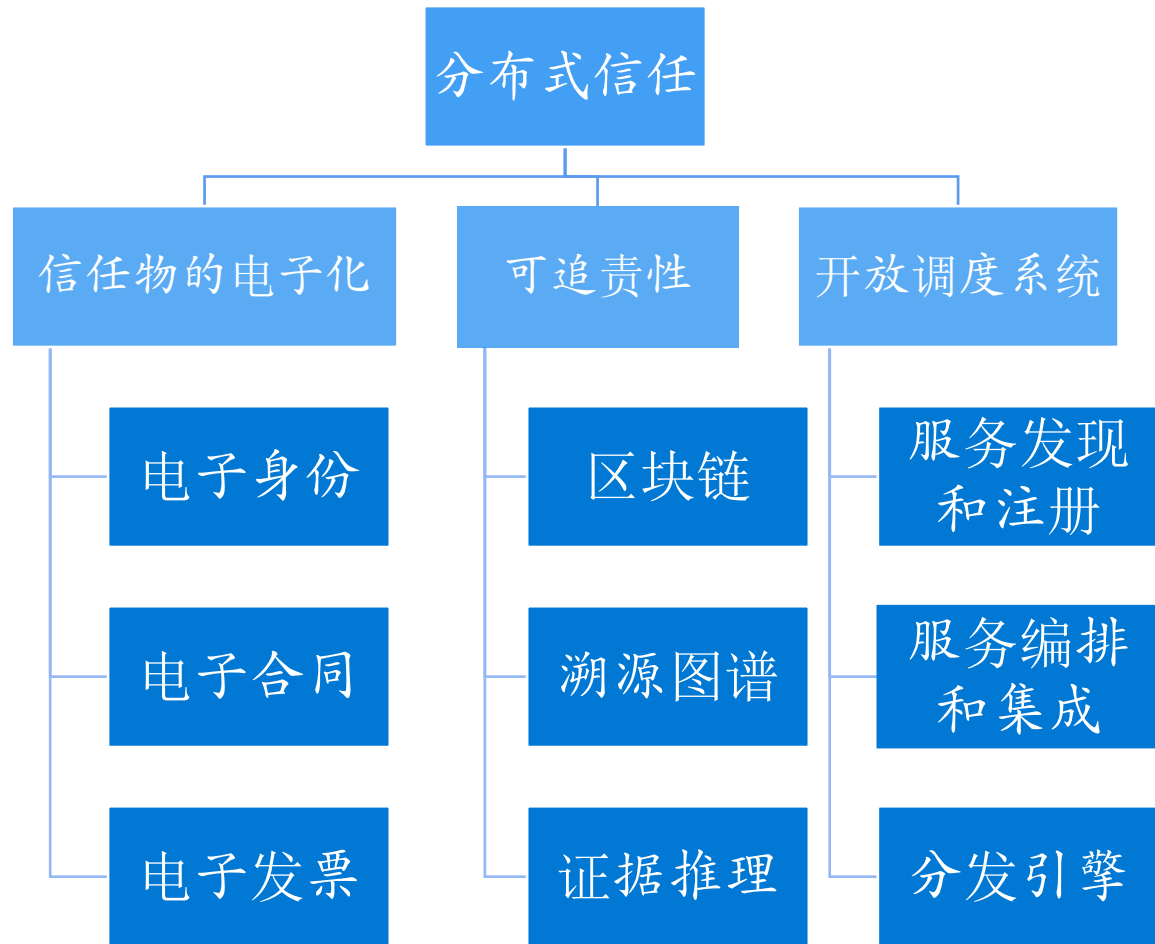
如何开放价值：可追责性 Accountability

不事前制止，而是记录行为并事后追责

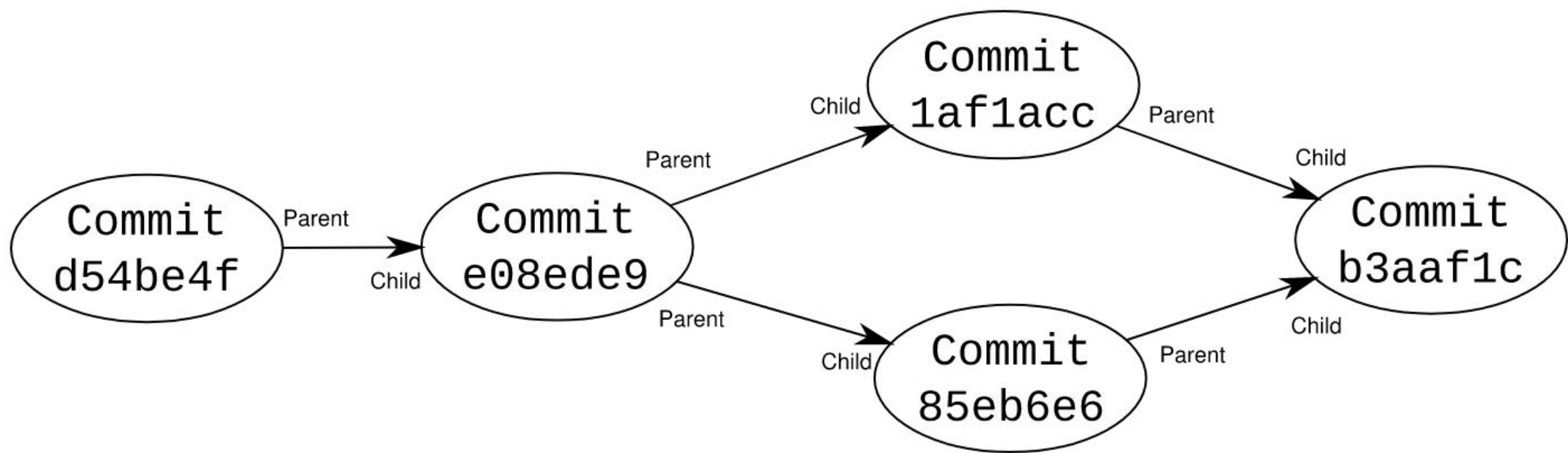


Tim Berners-Lee
2016 Turing Award

■ 分布式信任不仅仅是区块链



■ 链式追责是计算机领域很早就实践的方法



git, 2005 by Linus Torvalds

■ Web 3.0 是否必然建于区块链之上？

Web inventor Tim Berners-Lee: Screw Web3 — my decentralized internet doesn't need blockchain

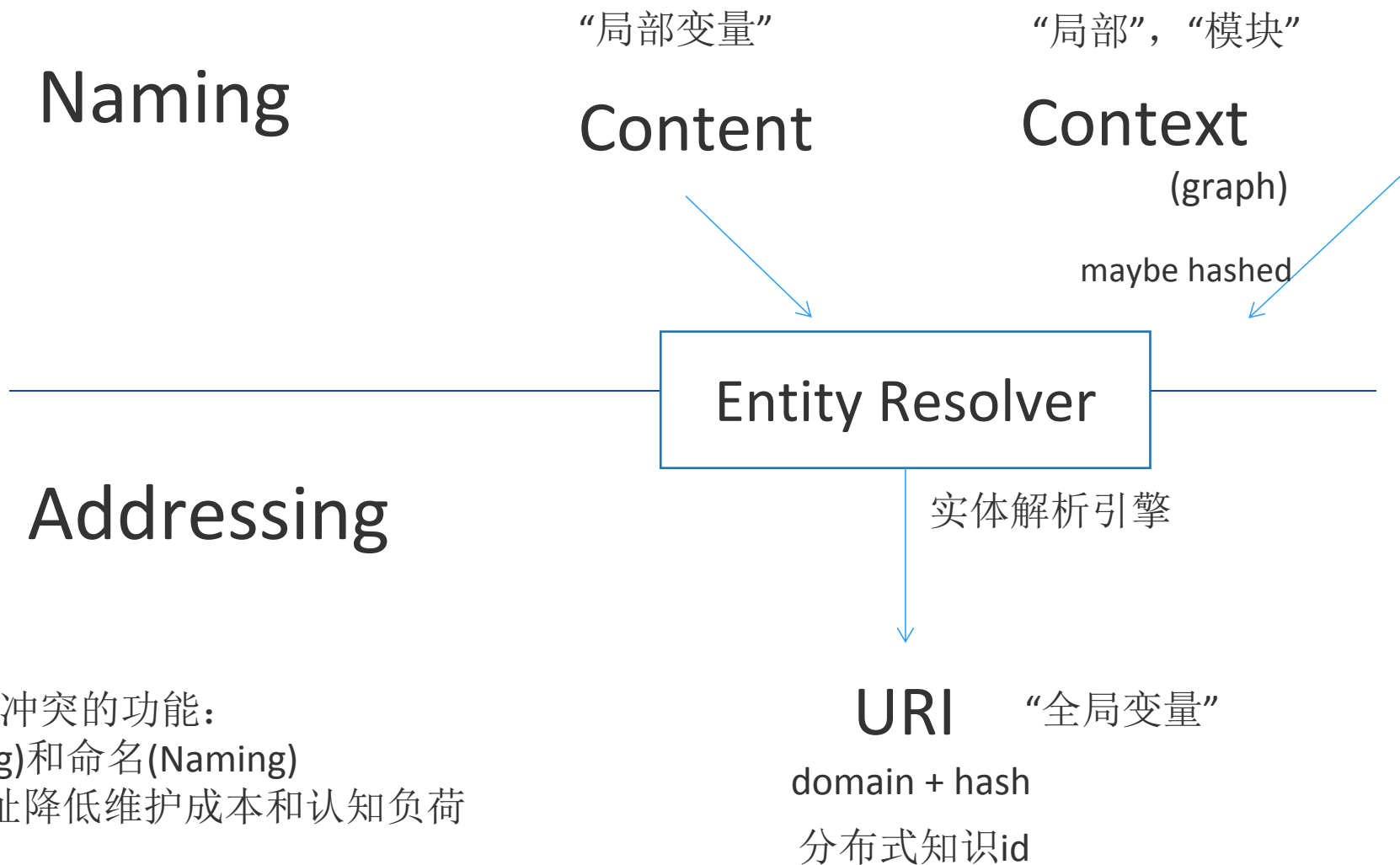
The cyberspace pioneer is skeptical about a blockchain-based internet

June 23, 2022 - 9:03 am

- 协作是目的，链式追责只是手段之一
- 共识或者激励机制是不是“可扩展性”的前提？
- 面对利维坦，最好还是合作而非对抗——寻找改良之道



“区块”到底是什么：基于内容的寻址

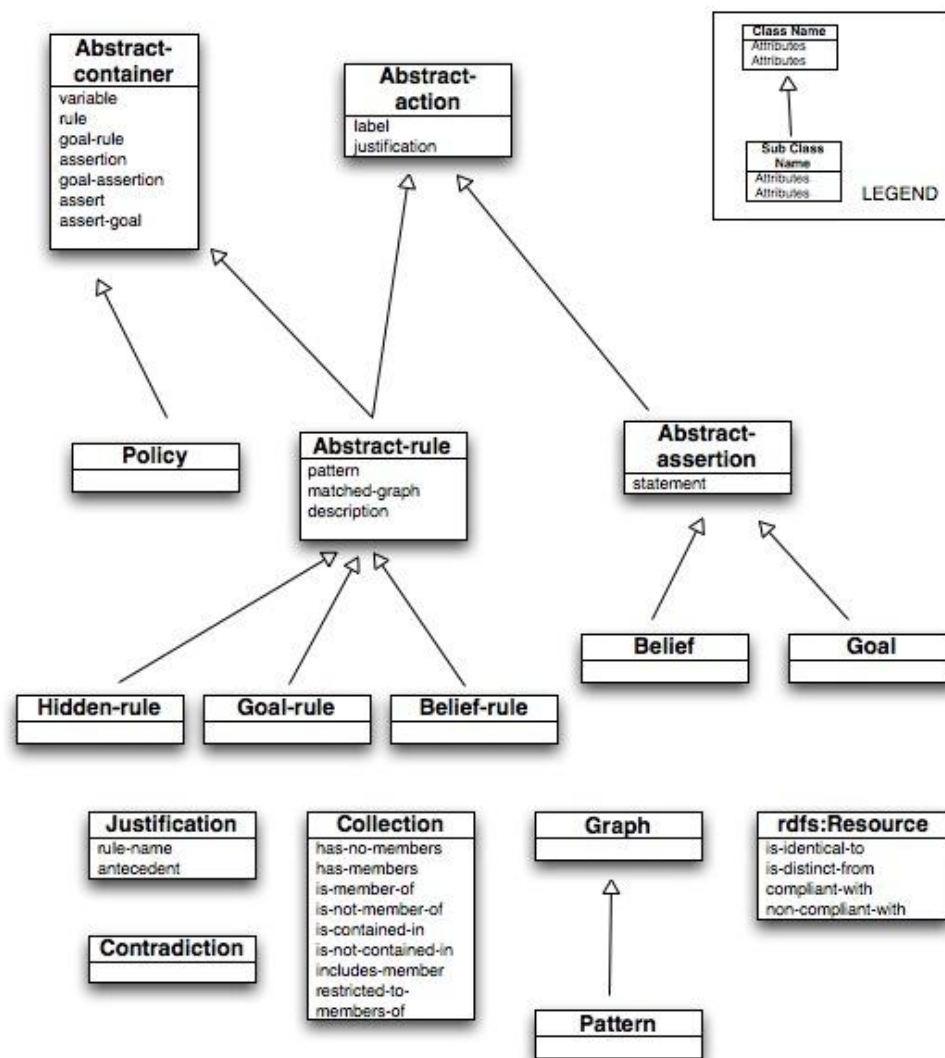


URI承载着两个冲突的功能：

寻址(**Addressing**)和命名(**Naming**)

基于内容的寻址降低维护成本和认知负荷

AIR: 基于RDF的追责规则语言 (2008)



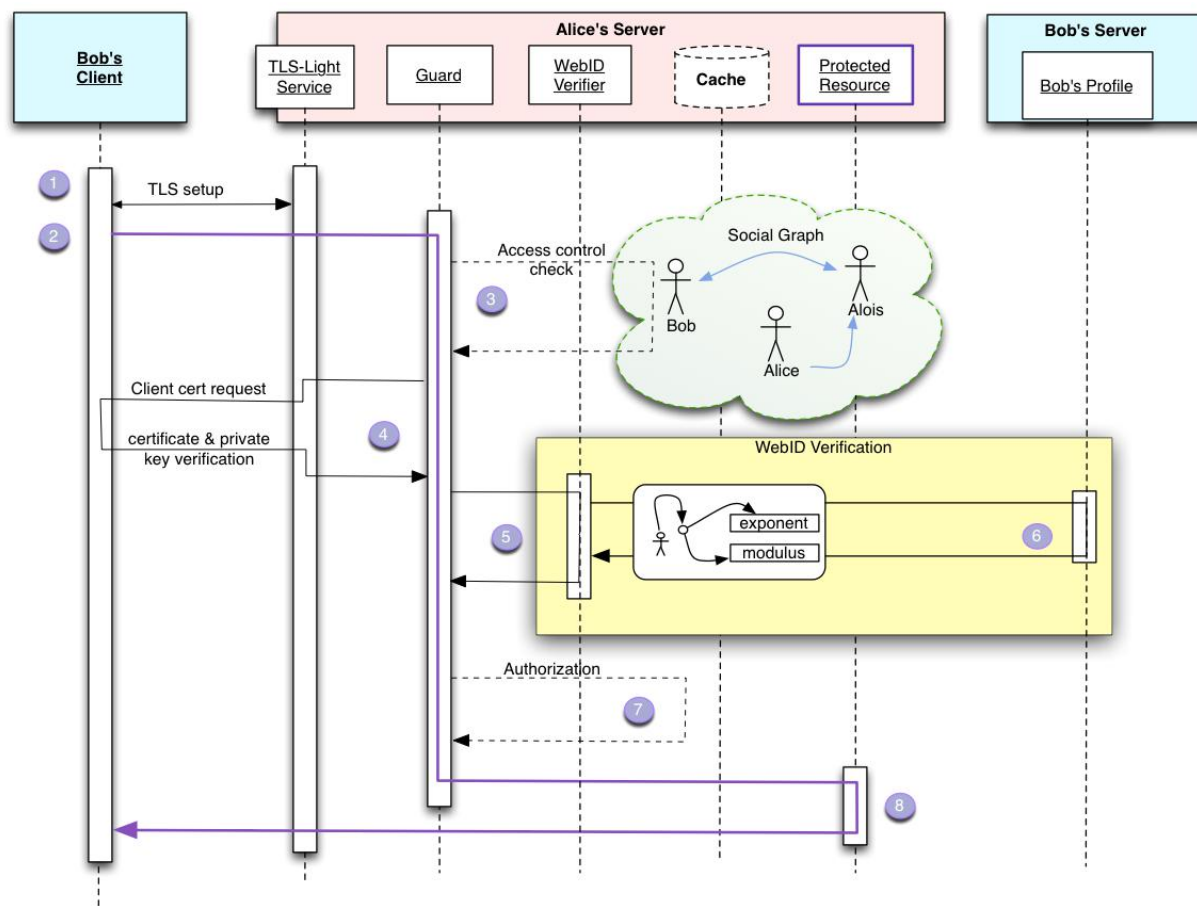
SOLID框架：身份是一种资产！



身份证是一种漏洞百出的设计
需要一种更安全、权利界定更清楚的数字身份认证

SOLID框架 (2016)

基于Linked Data Platform的分布式 WebID 认证



 Login with Facebook

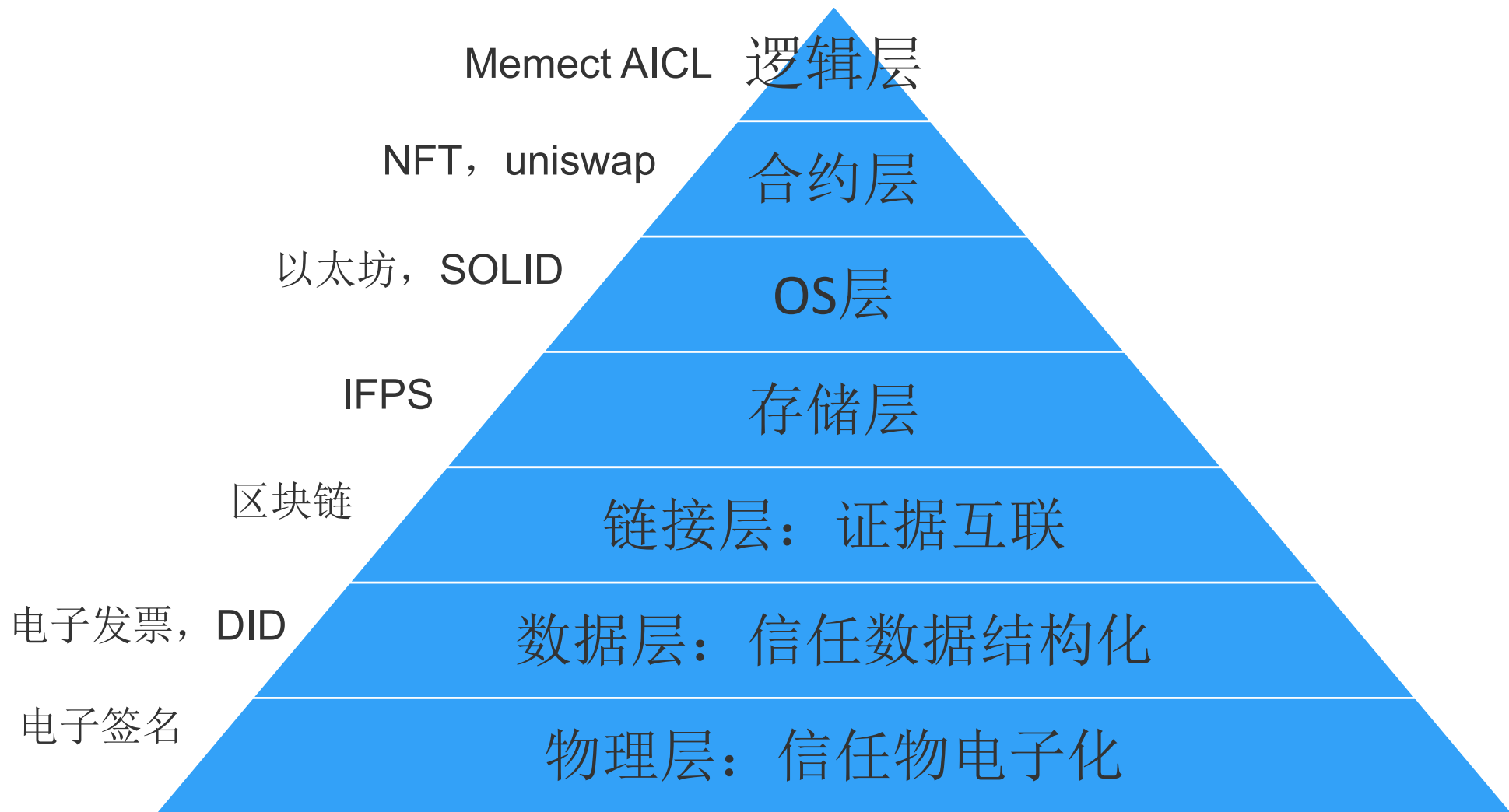
 Login with Google

 Login with Twitter



 Login with Solid

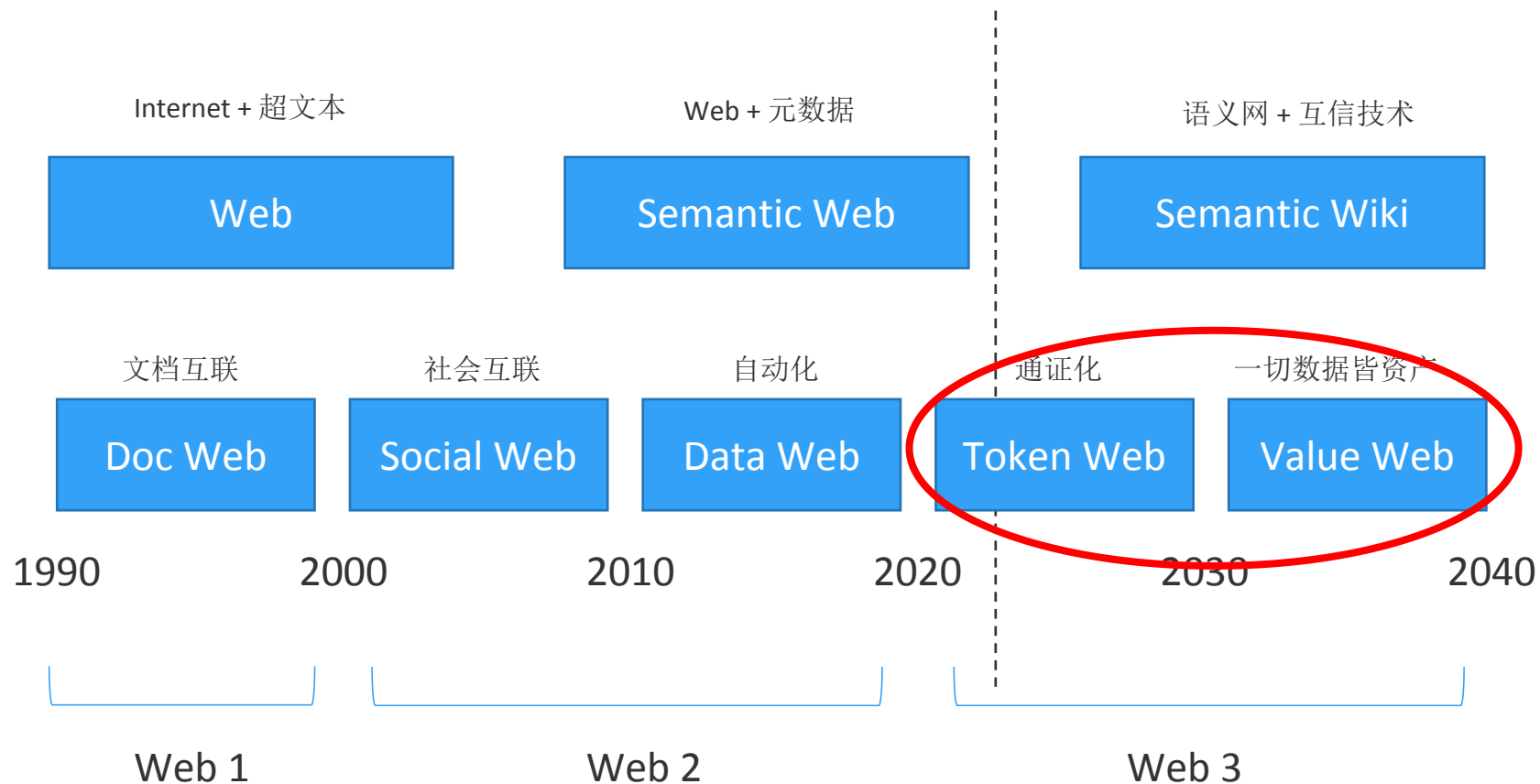
总结：信任技术的层级



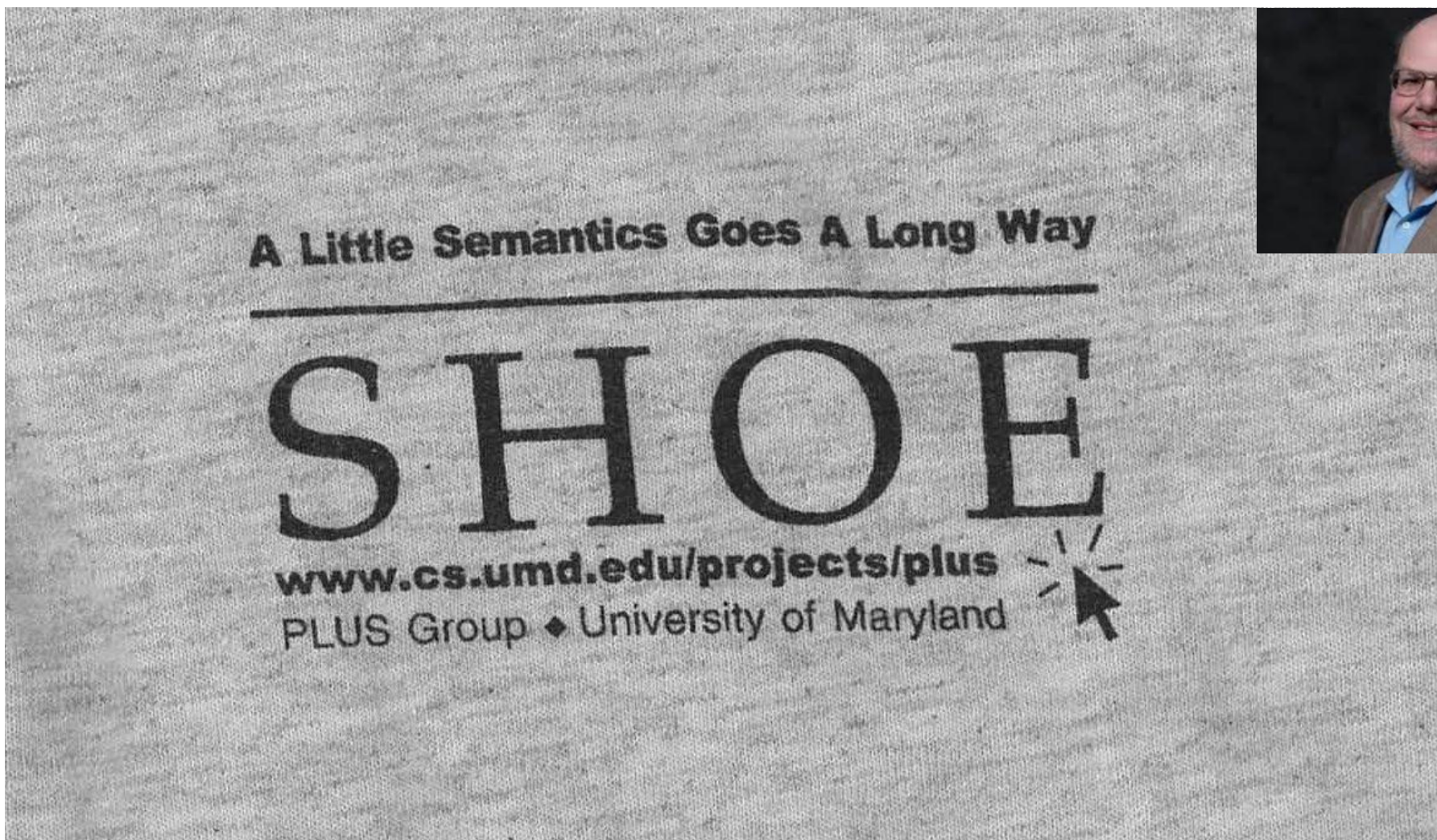
总结：Web 3的核心逻辑



新Web：资产的互联网络



Hendler假设: A little semantics goes a long way



文因互联



鲍捷 个人微信