知识图谱不是图谱

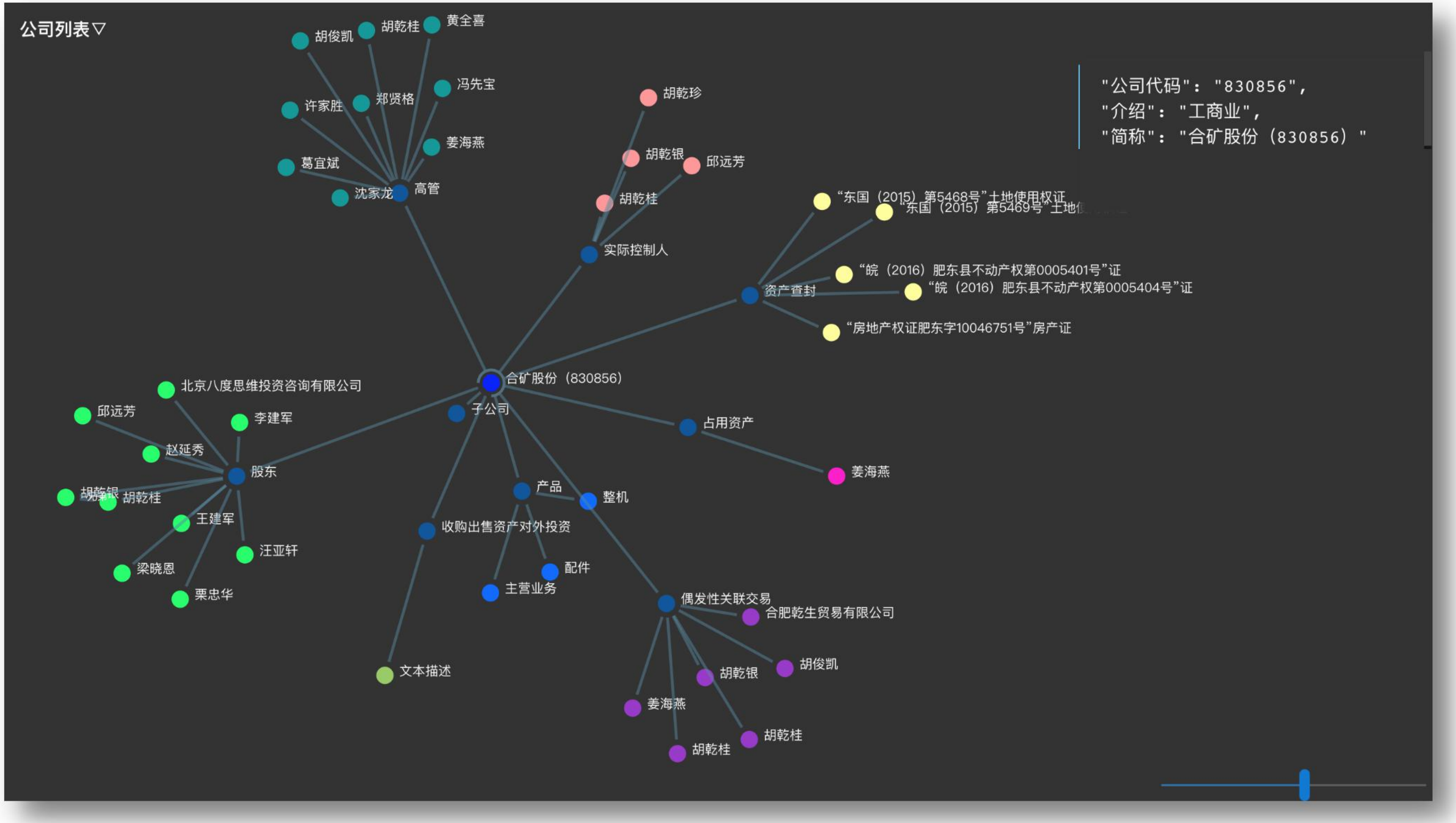
——把智能放在数据而非代码里

鲍捷

2019-03-27

北京文因互联科技有限公司

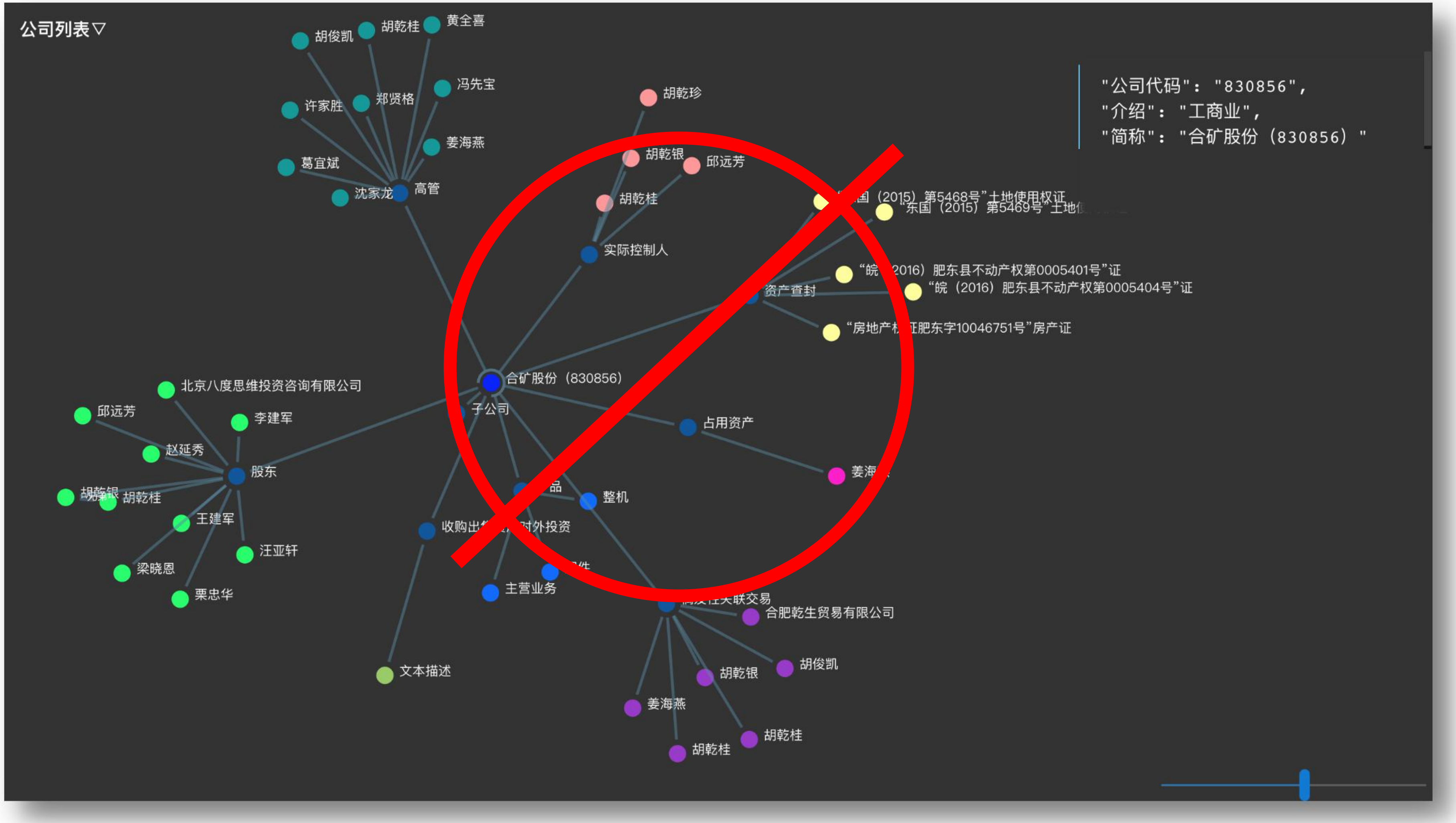
图谱



公司列表▽

```
"公司代码": "830856",
"介绍": "工商业",
"简称": "合矿股份 (830856) "
```

知识图谱 ≠ 图谱



```
"公司代码": "830856",
"介绍": "工商业",
"简称": "合矿股份 (830856) "
```

Semantics

Meaning

Relation

Structure

Relation gives meaning

Semantics ≈ Structure

Knowledge
≈
Structure that may generate structure

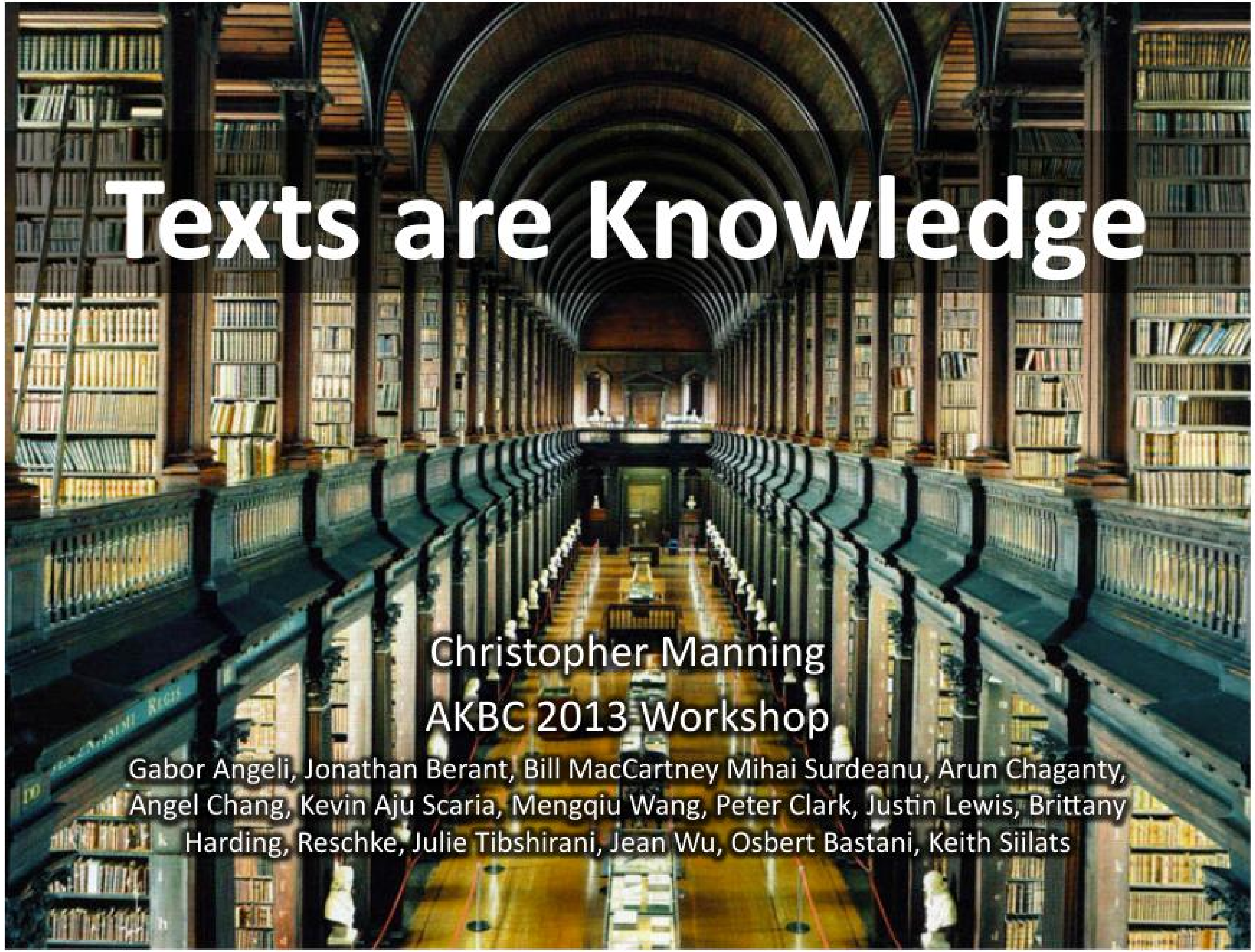
可以产生结构的结构

Knowledge graph
≈
Machine readable structure that gives "meaning"

机器可读的定义"意义"的数据结构

Knowledge graph
≈
Machine readable "dictionary"

机器可读的“字典”



Texts are Knowledge

Christopher Manning

AKBC 2013 Workshop

Gabor Angeli, Jonathan Berant, Bill MacCartney Mihai Surdeanu, Arun Chaganty, Angel Chang, Kevin Aju Scaria, Mengqiu Wang, Peter Clark, Justin Lewis, Brittany Harding, Reschke, Julie Tibshirani, Jean Wu, Osbert Bastani, Keith Siilats

知识结构怎么来？

结构化数据的处理方式

图书馆学：标签

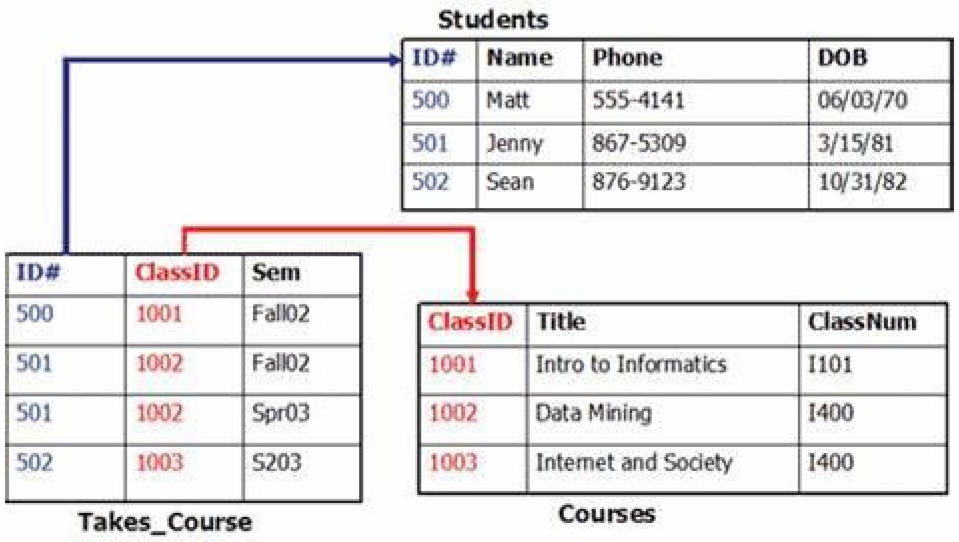
软件工程：一维表、哈希表、树等

数据工程：二维表

知识工程：剩下的（那些难的）

知识工程处理那些能产生数据的数据

“数据库”和知识图谱有啥不同？



Students

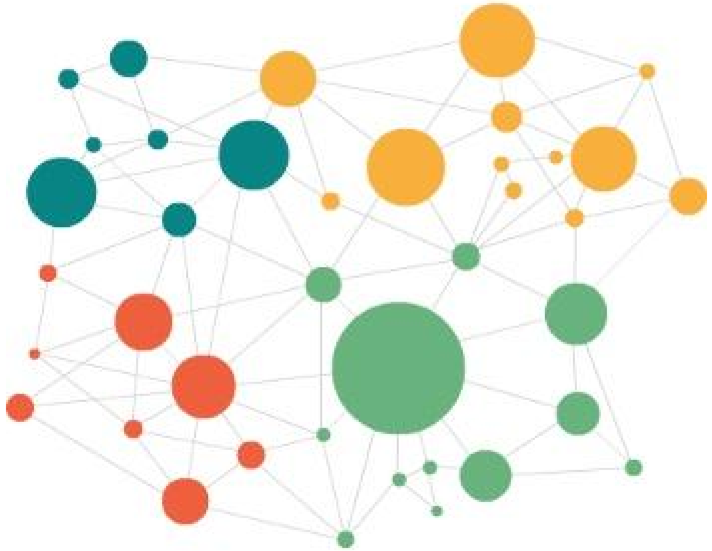
ID#	Name	Phone	DOB
500	Matt	555-4141	06/03/70
501	Jenny	867-5309	3/15/81
502	Sean	876-9123	10/31/82

ID#	ClassID	Sem
500	1001	Fall02
501	1002	Fall02
501	1002	Spr03
502	1003	S203

Takes_Course

ClassID	Title	ClassNum
1001	Intro to Informatics	I101
1002	Data Mining	I400
1003	Internet and Society	I400

Courses



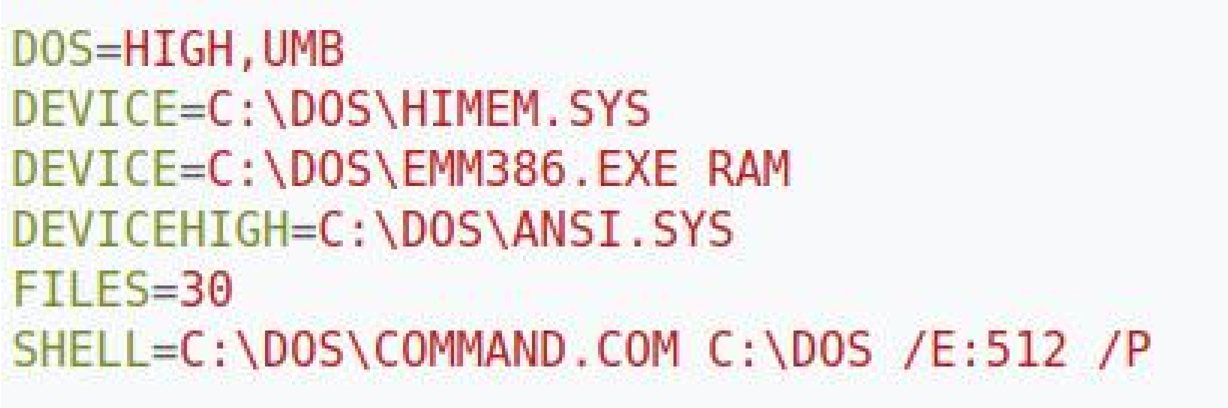
更多"语义"

- 表间关系
- 列间关系
- 行间关系
- 单元格值关系
- 完整性约束
- 条件触发

知识图谱的眞正威力：

把智能从代码里挪到数据里

把智能从代码里挪到数据里



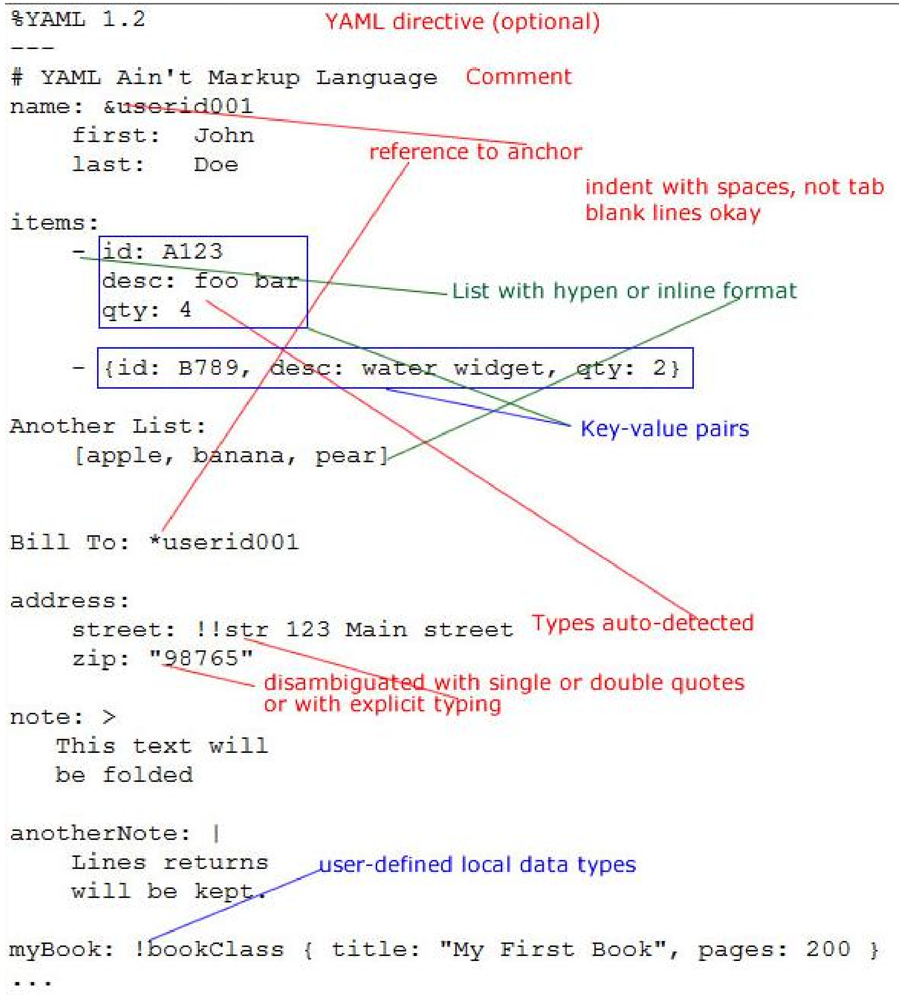
```
DOS=HIGH,UMB
DEVICE=C:\DOS\HIMEM.SYS
DEVICE=C:\DOS\EMM386.EXE RAM
DEVICEHIGH=C:\DOS\ANSI.SYS
FILES=30
SHELL=C:\DOS\COMMAND.COM C:\DOS /E:512 /P
```

所有的知识表现语言开始都像是个配置文件

把智能从代码里挪到数据里

从配置文件到知识语言

- 配置的配置
- 增强模块化和重用性
- 增强节点命名
- 增强链接
- 规则的需求自然出现

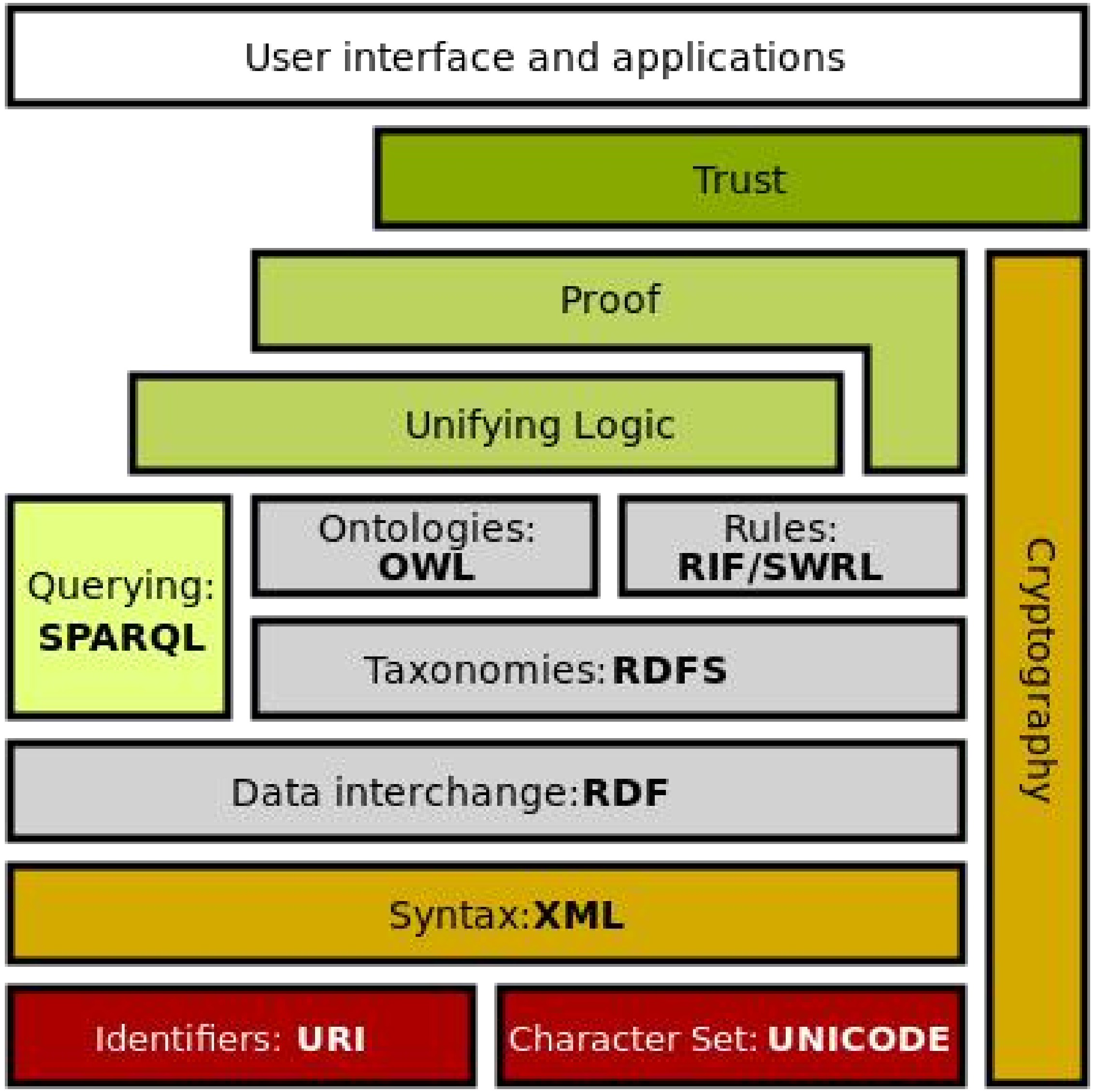


```
%YAML 1.2
---
# YAML Ain't Markup Language
name: &userid001
    first:  John
    last:   Doe

items:
    - id: A123
      desc: foo bar
      qty: 4

    - {id: B789, desc: water widget, qty: 2}

Another List:
    [apple, banana, pear]


Bill To: *userid001

address:
    street: !!str 123 Main street
    zip: "98765"

note: >
   This text will
   be folded

anotherNote: |
    Lines returns
    will be kept.

myBook: !bookClass { title: "My First Book", pages: 200 }
...
```

YAML directive (optional)
Comment
reference to anchor
indent with spaces, not tab
blank lines okay
List with hypen or inline format
Key-value pairs
Types auto-detected
disambiguated with single or double quotes
or with explicit typing
user-defined local data types

所有的知识表现语言开始都像是个配置文件

知识图谱的（官方）层级



User interface and applications

Trust

Proof

Unifying Logic

Querying: **SPARQL**

Ontologies: **OWL**

Rules: **RIF/SWRL**

Taxonomies: **RDFS**

Data interchange: **RDF**

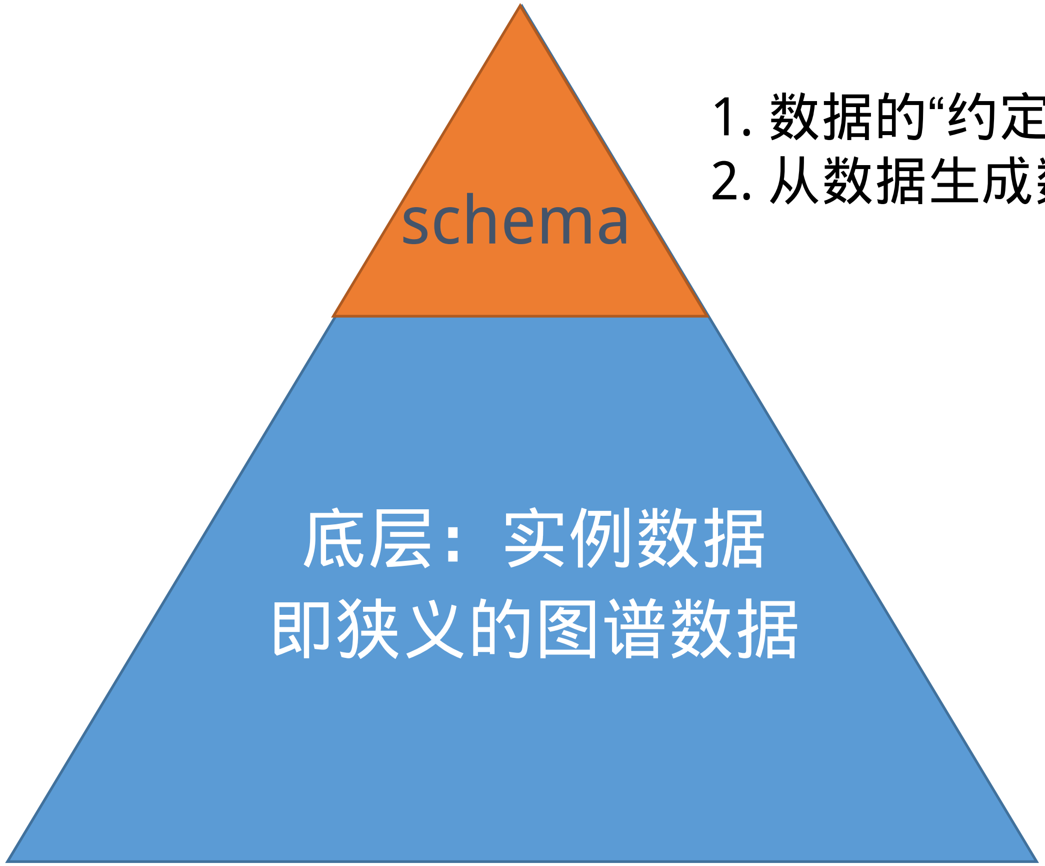
Syntax: **XML**

Identifiers: **URI**

Character Set: **UNICODE**

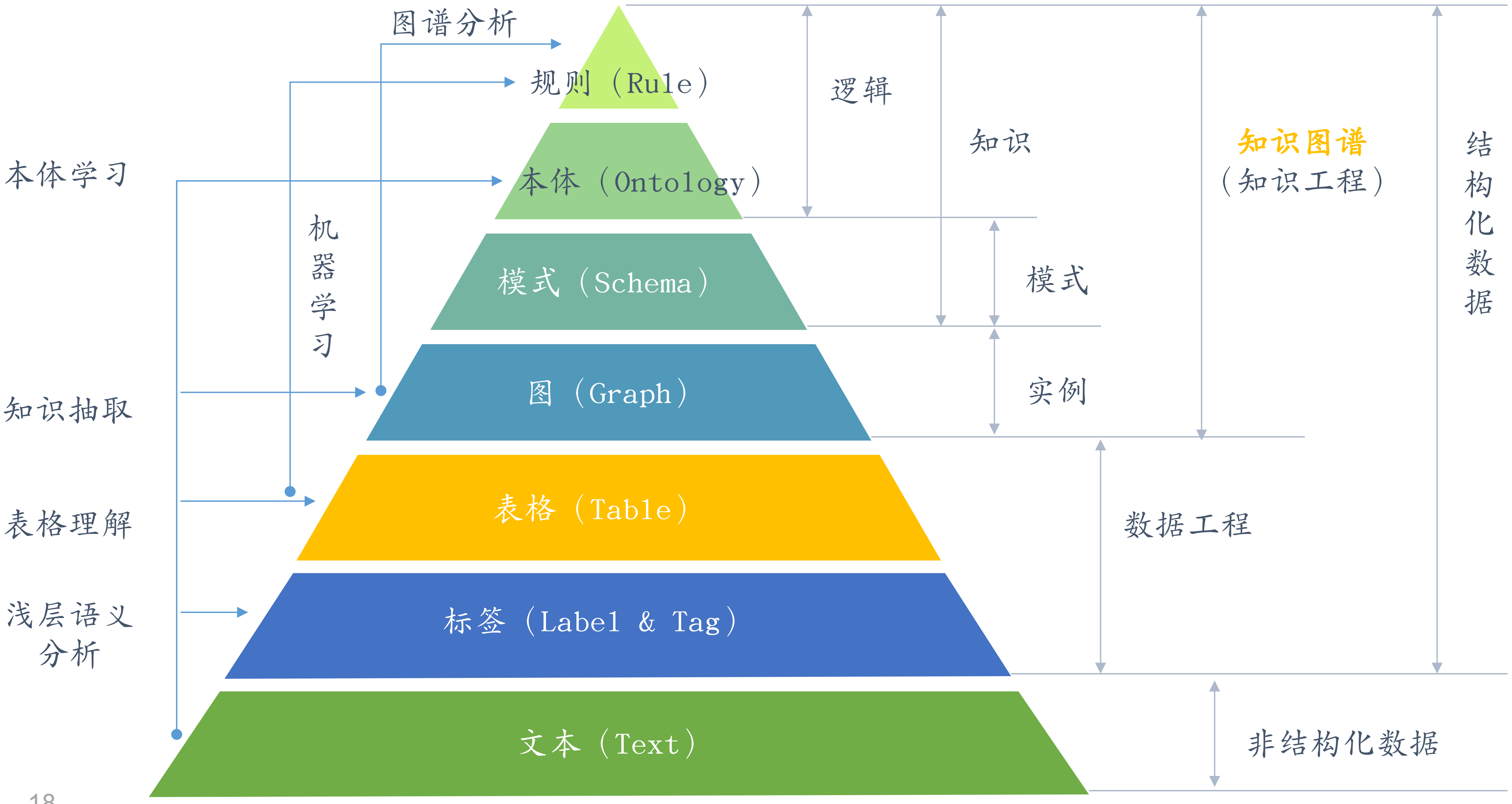
Cryptography

知识图谱的（抽象简化）层级



1. 数据的“约定”
2. 从数据生成数据

知识图谱技术金字塔


图谱分析
规则（Rule）
本体学习
本体（Ontology）
机器学习
模式（Schema）
知识抽取
图（Graph）
表格理解
表格（Table）
浅层语义分析
标签（Label & Tag）
文本（Text）
逻辑
知识
模式
实例
知识图谱
（知识工程）
结构化数据
数据工程
非结构化数据

知识图谱的“基本原则”

1． 数据与应用的分离

2． 数据的自描述性

3． 数据的弱／无模式性

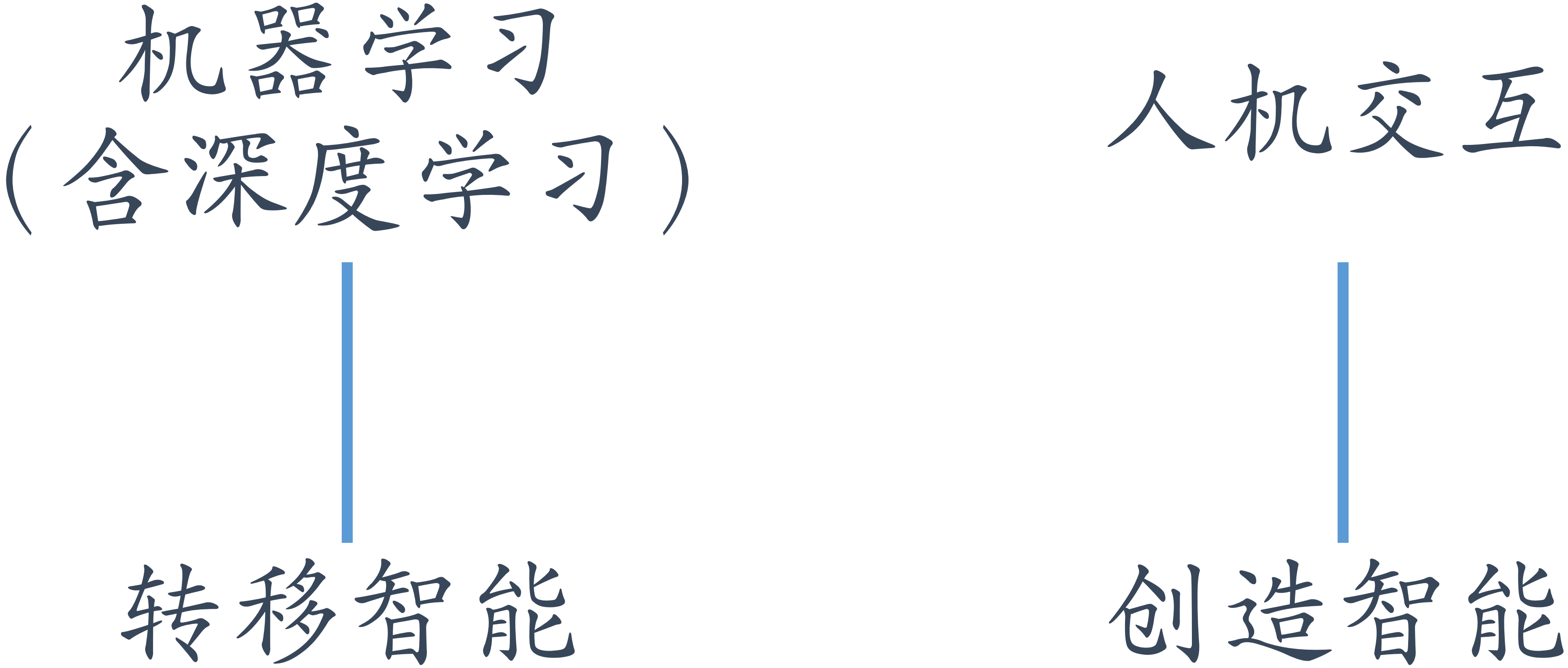
4． 数据的可演进性

“Artificial” Intelligence

×

Augmented Intelligence

√



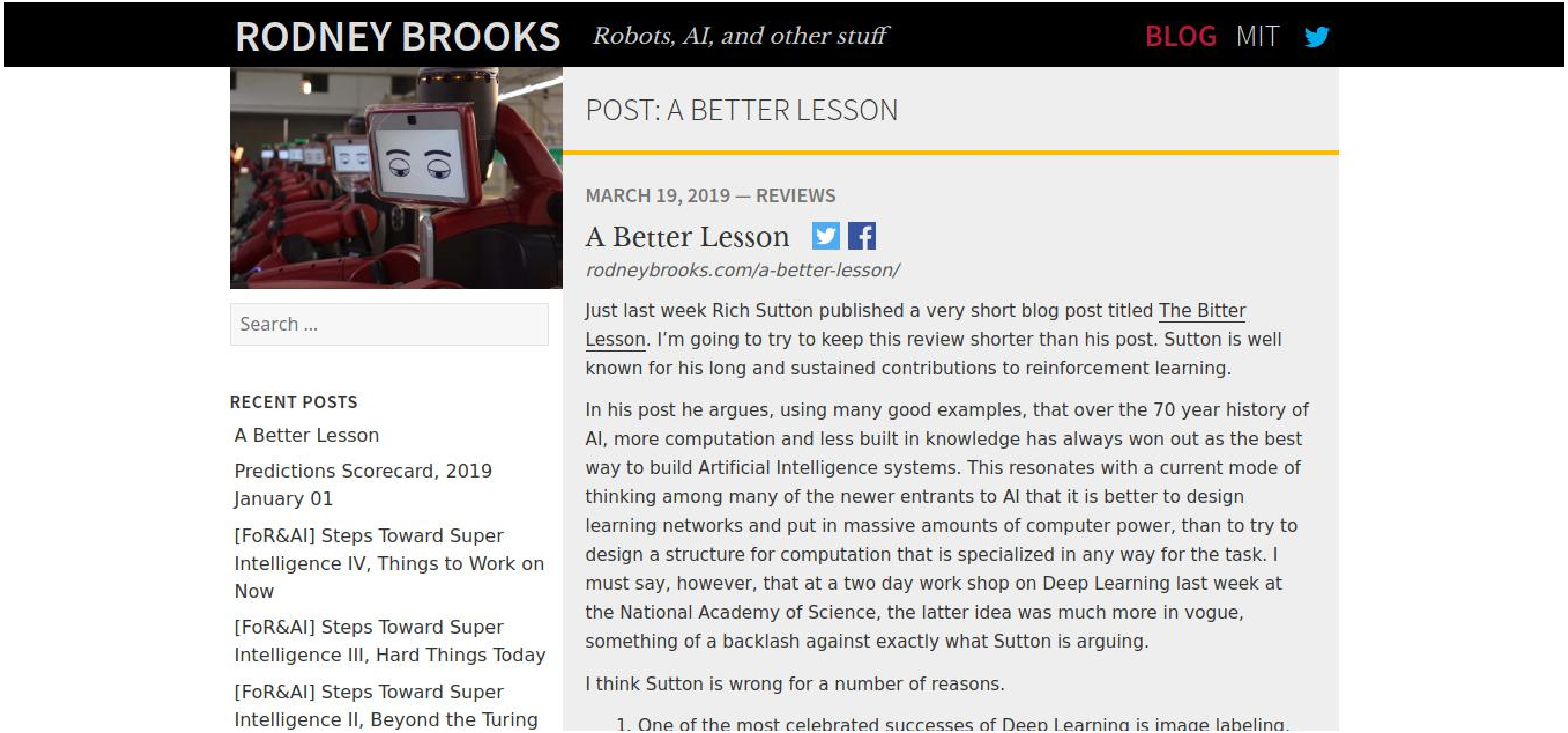
机器学习
（含深度学习）

转移智能

人机交互

创造智能

社会要富裕，光靠金融不行，要工业农业，把财富创造出来。
智能系统也一样，不管是训练样本还是知识规则，都要靠人来产生。



RODNEY BROOKS *Robots, AI, and other stuff*

BLOG MIT

Search …

RECENT POSTS

- A Better Lesson
- Predictions Scorecard, 2019 January 01
- [FoR&AI] Steps Toward Super Intelligence IV, Things to Work on Now
- [FoR&AI] Steps Toward Super Intelligence III, Hard Things Today
- [FoR&AI] Steps Toward Super Intelligence II, Beyond the Turing

POST: A BETTER LESSON

MARCH 19, 2019 — REVIEWS

A Better Lesson

rodneybrooks.com/a-better-lesson/

Just last week Rich Sutton published a very short blog post titled The Bitter Lesson. I'm going to try to keep this review shorter than his post. Sutton is well known for his long and sustained contributions to reinforcement learning.

In his post he argues, using many good examples, that over the 70 year history of AI, more computation and less built in knowledge has always won out as the best way to build Artificial Intelligence systems. This resonates with a current mode of thinking among many of the newer entrants to AI that it is better to design learning networks and put in massive amounts of computer power, than to try to design a structure for computation that is specialized in any way for the task. I must say, however, that at a two day work shop on Deep Learning last week at the National Academy of Science, the latter idea was much more in vogue, something of a backlash against exactly what Sutton is arguing.

I think Sutton is wrong for a number of reasons.

1. One of the most celebrated successes of Deep Learning is image labeling.

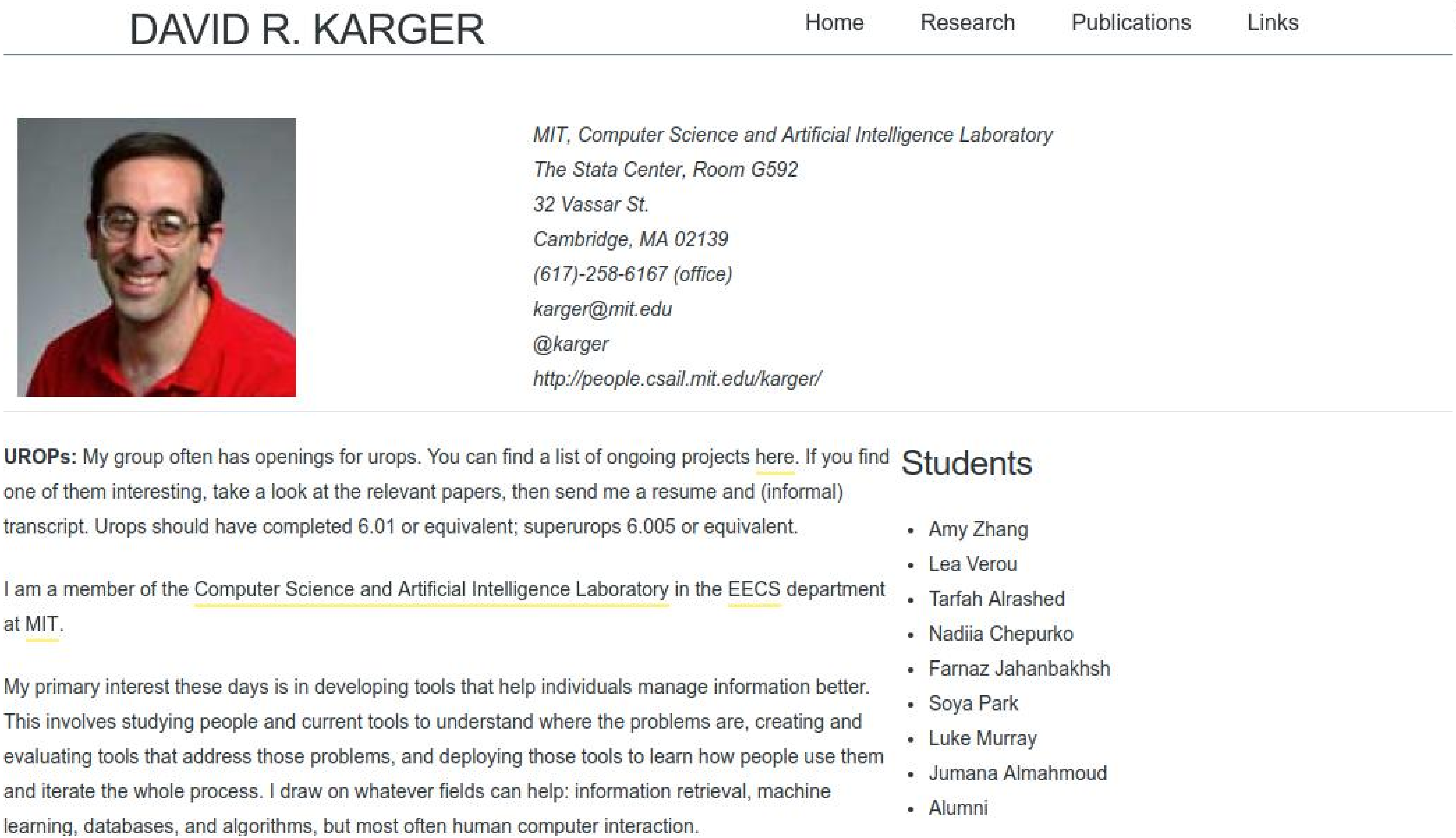
image by neilsethlevine.com

Intelligence at the Interface

Semantic Technology and the Consumer Internet Experience

Tom Gruber

Stealth-Company.com



DAVID R. KARGER

Home Research Publications Links

MIT, Computer Science and Artificial Intelligence Laboratory
The Stata Center, Room G592
32 Vassar St.
Cambridge, MA 02139
(617)-258-6167 (office)
karger@mit.edu
@karger
http://people.csail.mit.edu/karger/

UROPs: My group often has openings for urops. You can find a list of ongoing projects here. If you find one of them interesting, take a look at the relevant papers, then send me a resume and (informal) transcript. Urops should have completed 6.01 or equivalent; superurops 6.005 or equivalent.

I am a member of the Computer Science and Artificial Intelligence Laboratory in the EECS department at MIT.

My primary interest these days is in developing tools that help individuals manage information better. This involves studying people and current tools to understand where the problems are, creating and evaluating tools that address those problems, and deploying those tools to learn how people use them and iterate the whole process. I draw on whatever fields can help: information retrieval, machine learning, databases, and algorithms, but most often human computer interaction.

Students

- Amy Zhang
- Lea Verou
- Tarfah Alrashed
- Nadiia Chepurko
- Farnaz Jahanbakhsh
- Soya Park
- Luke Murray
- Jumana Almahmoud
- Alumni

Web3.0的核心是一场HCI的革命

结构化数据（知识图谱）是这场革命的催化剂

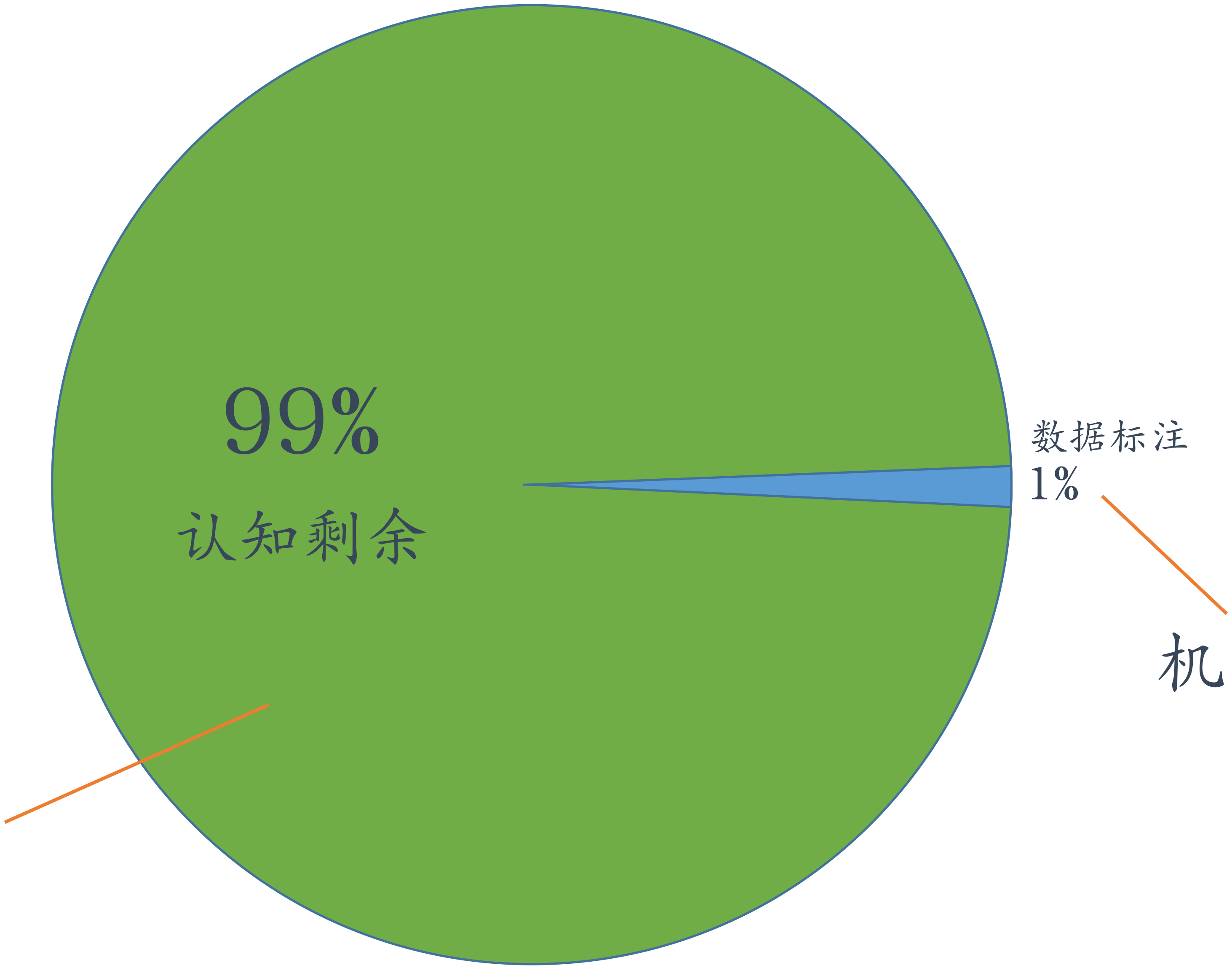
知识图谱的突破点可能在HCI

不在于知识提取或者知识表现

~~知识图谱~~人工智能的突破点可能在HCI

不在于知识提取或者知识表现

“低枝上的桃子”



HCI来摘

机器学习
来摘

知识图谱的交互层

- **检索式** （结构查找）
  - 查询 （计算机语言检索）
  - 搜索 （关键词检索）
  - 问答 （自然语言检索）
  - 探索 （人机结合检索，如分面浏览和电子表格） ◀ 机会点
- **响应式** （规则激发）
  - 推送 （条件激发内容）
  - 约束检查 （如合规检查）
  - 验证执行（如智能合约）
- **图形式** （具象化）
  - 节点-边型图可视化
  - BI （图表与自动化报告）
  - 卡片式总结（如infobox，factbox）
  - 增强现实（如智能眼镜）

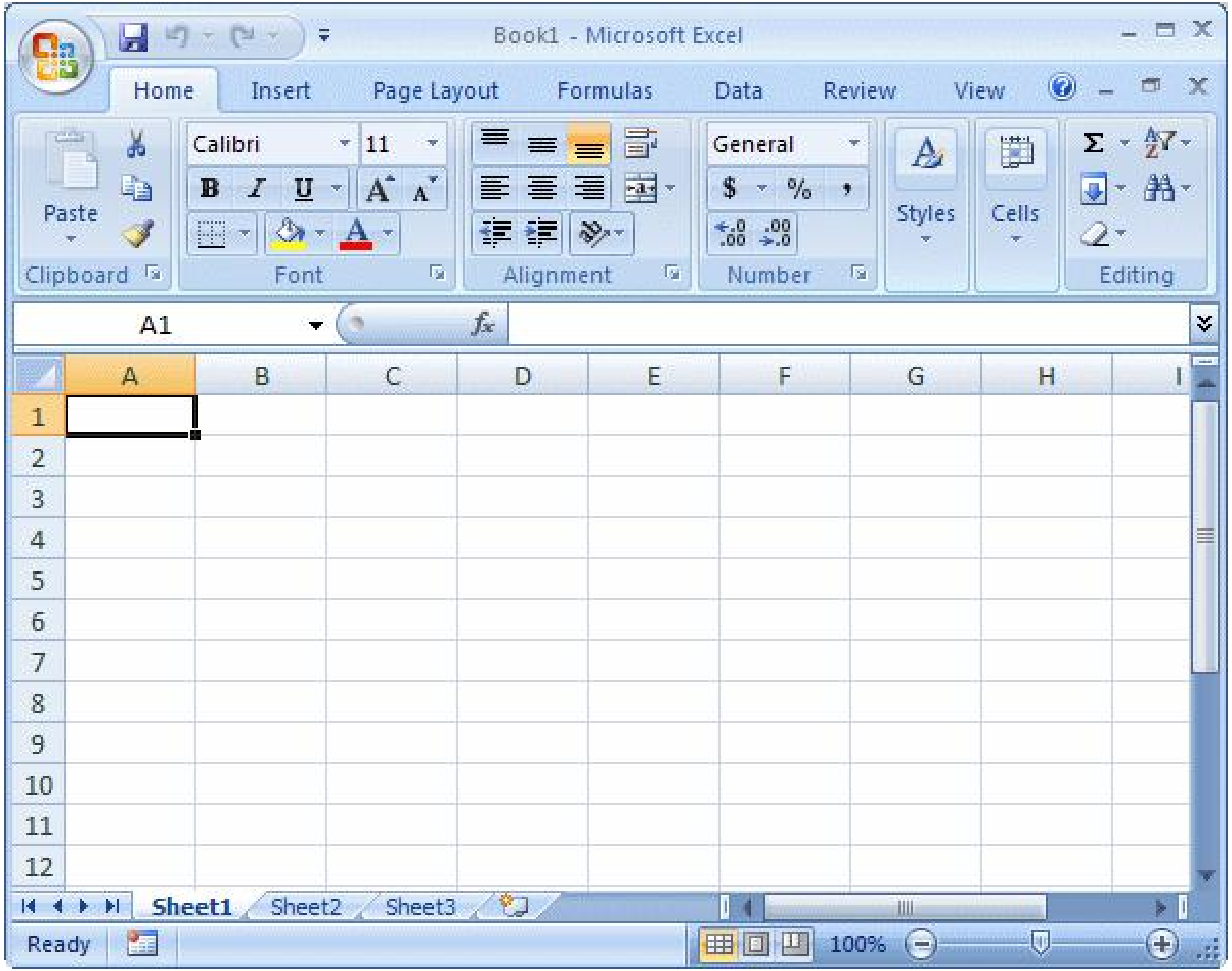
问答引擎

✗

探索引擎

√

最成功的探索引擎：电子表格



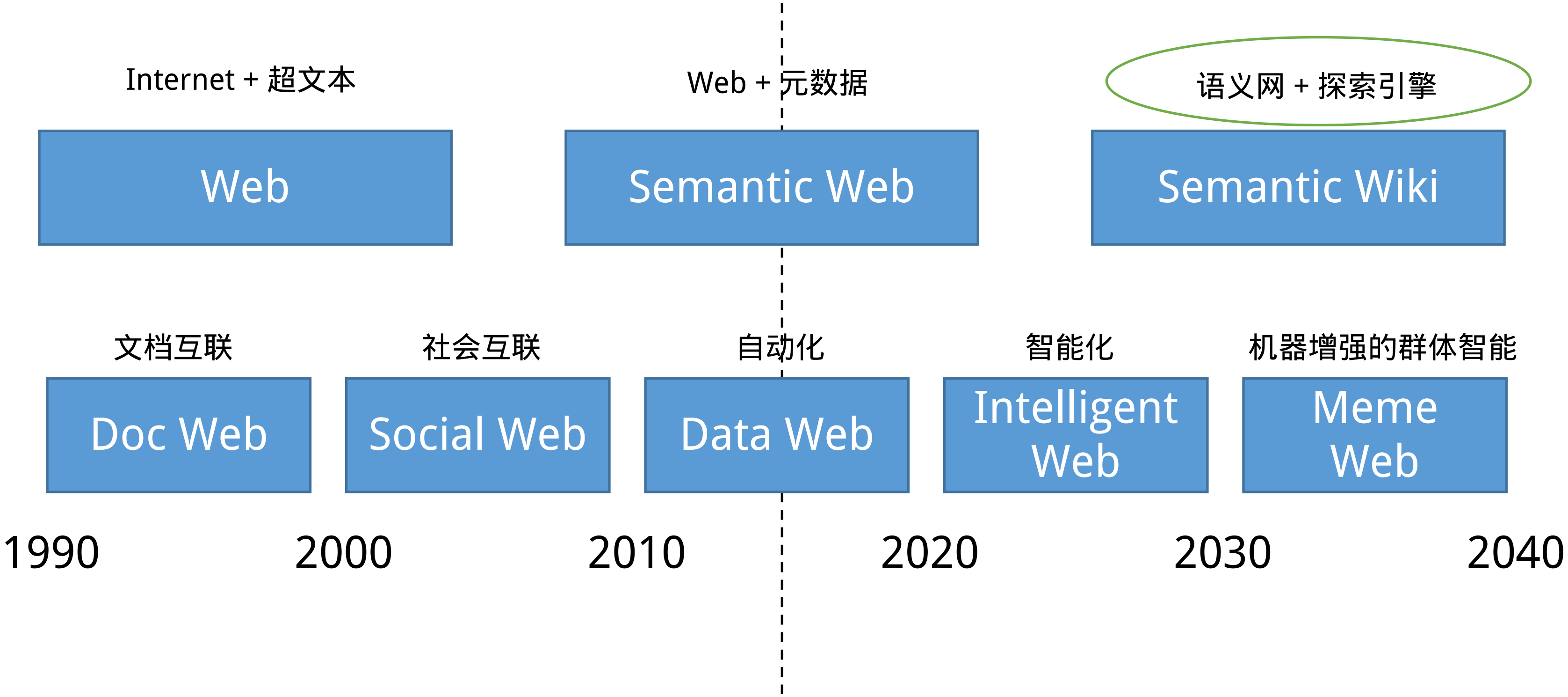
探索引擎

- 人机结合解决问题
- 提升数据的可发现性
- 和用户的交互中on-demand地更新和提取知识
- **把智能放在数据而非代码中**
  - 智能可以积累
  - 界面可以演化

需要重点关注的未来探索技术

- 分面浏览/编辑器 ◀ 机会点
- 本体编辑器 ◀ Palantir的发家技术
- 语义wiki
- VR中的信息流

Web的下半场


Internet + 超文本
Web
Web + 元数据
Semantic Web
语义网 + 探索引擎
Semantic Wiki
文档互联
Doc Web
社会互联
Social Web
自动化
Data Web
智能化
Intelligent Web
机器增强的群体智能
Meme Web
1990
2000
2010
2020
2030
2040

谢谢！

个人微信：baojie_memect

2019年3月

文因互联