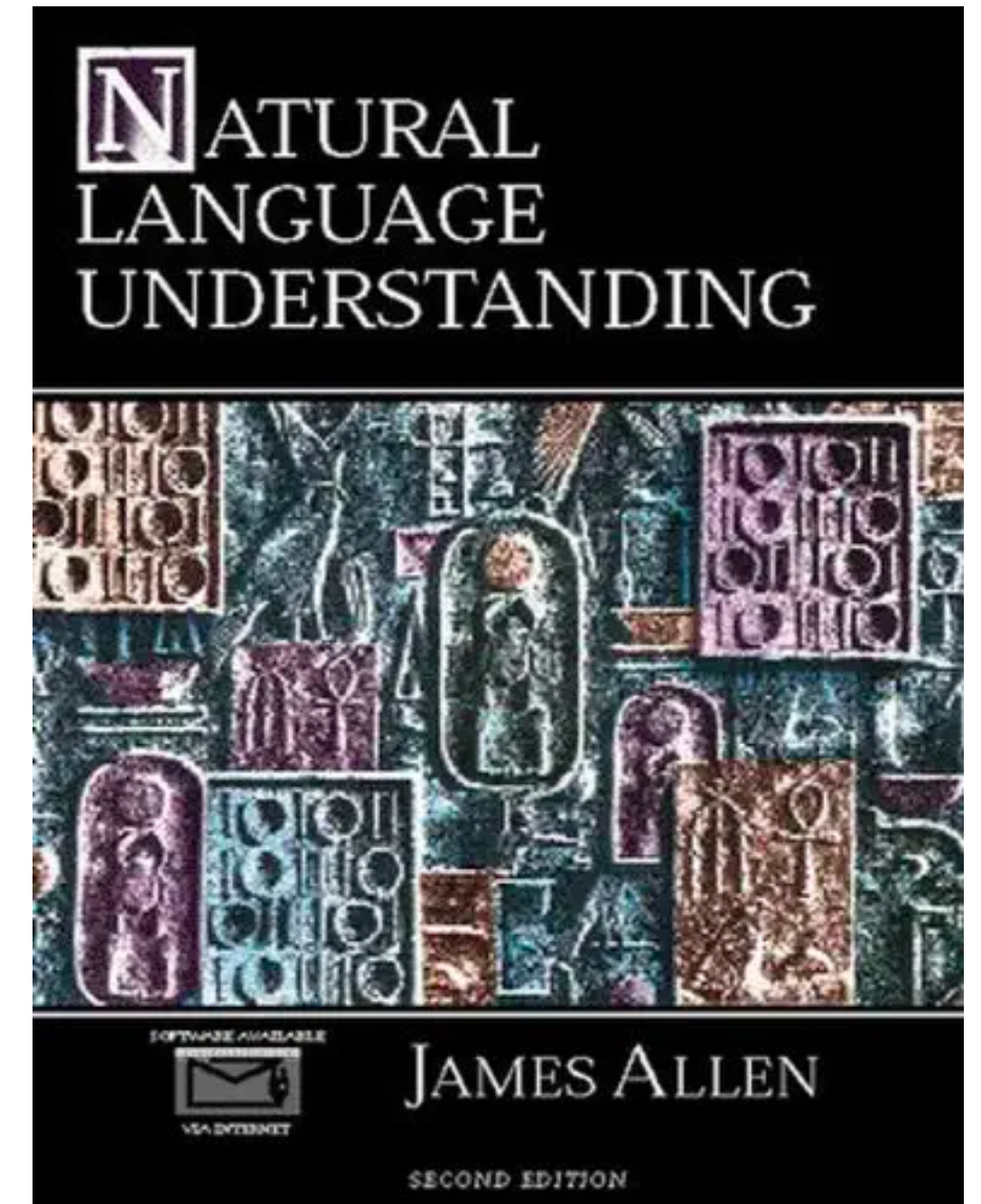


从编程语言到自然语言

@蔡学镛

我和自然语言处理的渊源

- 36 年前（1983年）买电脑，开始学习写程序。我学英文是从学 BASIC 语言的指令开始，背的第一批英文单词是 PRINT、DEFINE、RUN、IF、ELSE ...。
- 19 年前买了 James Allen 教授的 Natural Language Understanding, 2nd, 开始我的自然语言研究之路。并开始写 REBOL 代码实现书中的算法，尝试做一个英译中的 Pet Project。
- 9 年前我在创新工场时，开复一度对我的翻译程序感兴趣，但在知道我是规则式的翻译引擎后，他就失去兴趣了（开复是统计派不是规则派）。我后来主要是做互联网架构师的工作，也就没有继续做这个 Pet Project。
- 三年前，我设计出了一个编程框架（UniFrame），随着对这个框架的研究越来越深入，发现它和自然语言有许多共同的地方，甚至可以达到某种程度的融合。



本质上是希望做到人机沟通容易

人类

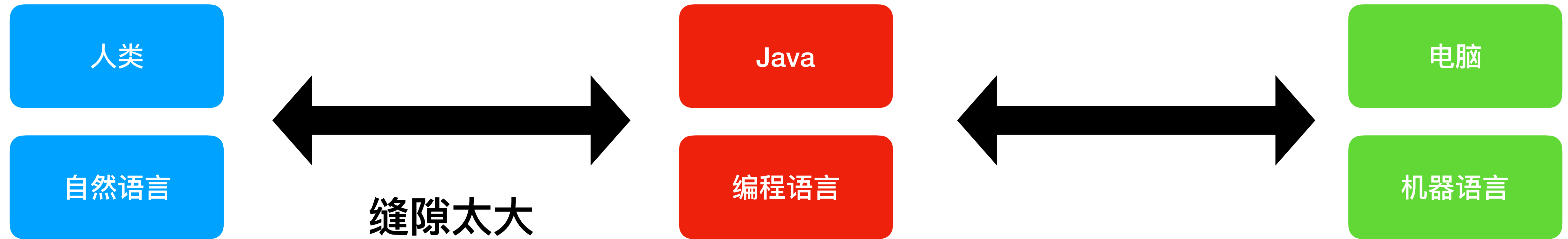
自然语言



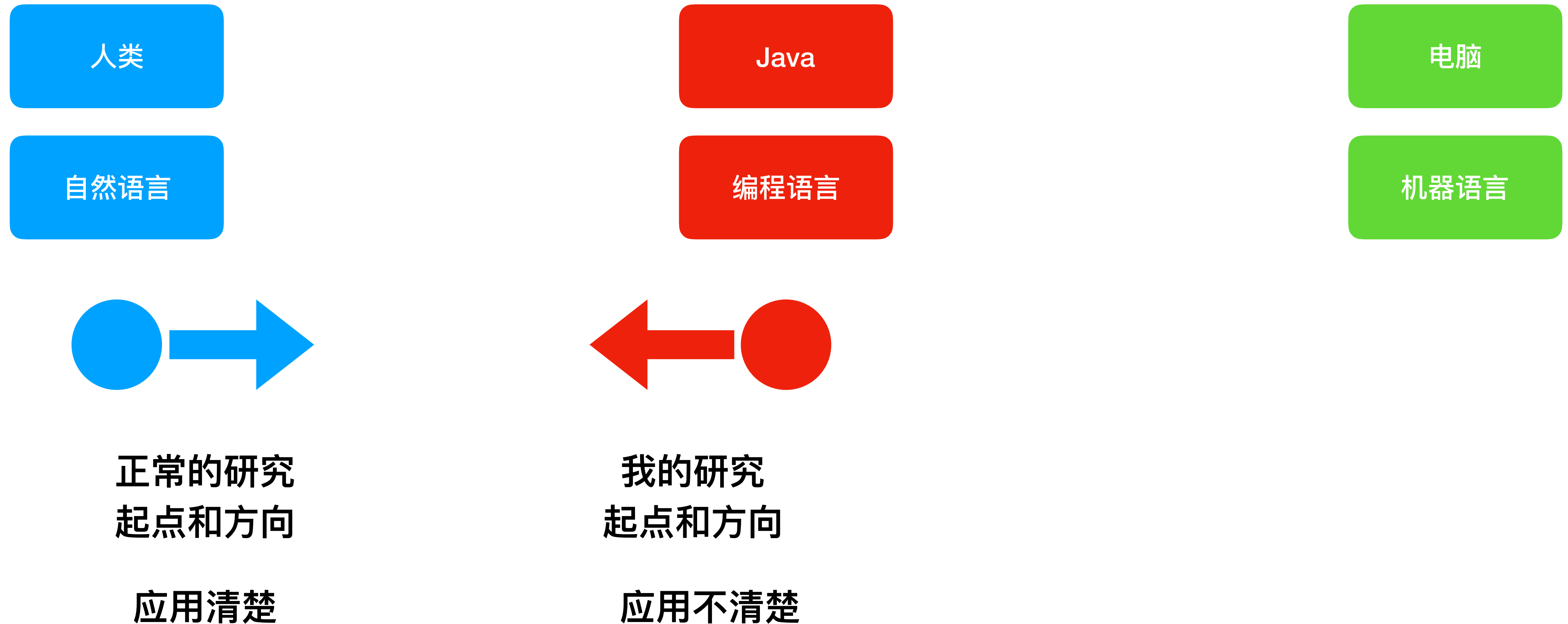
电脑

机器语言

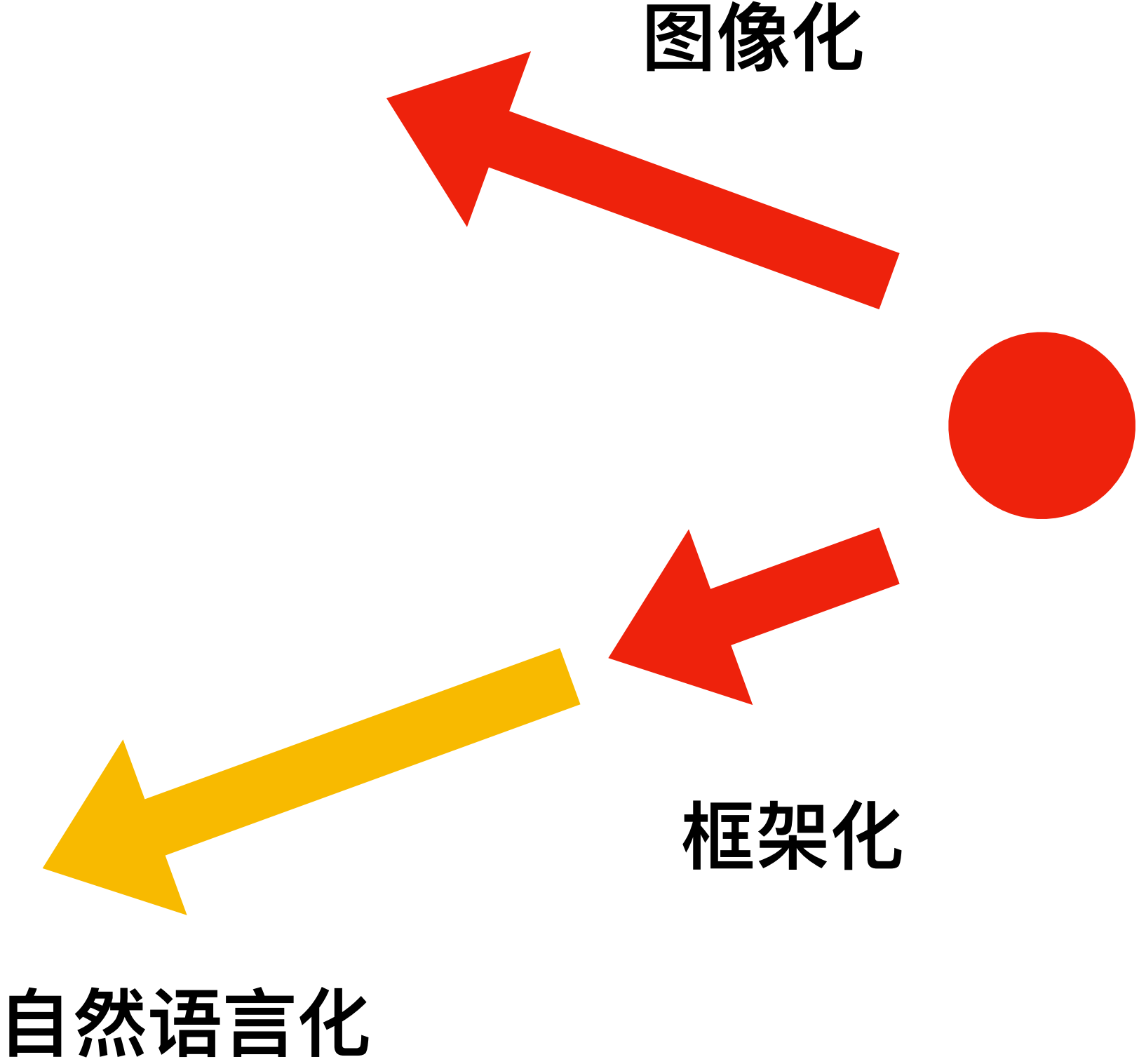
编程语言简化了人机沟通， 但远远不够

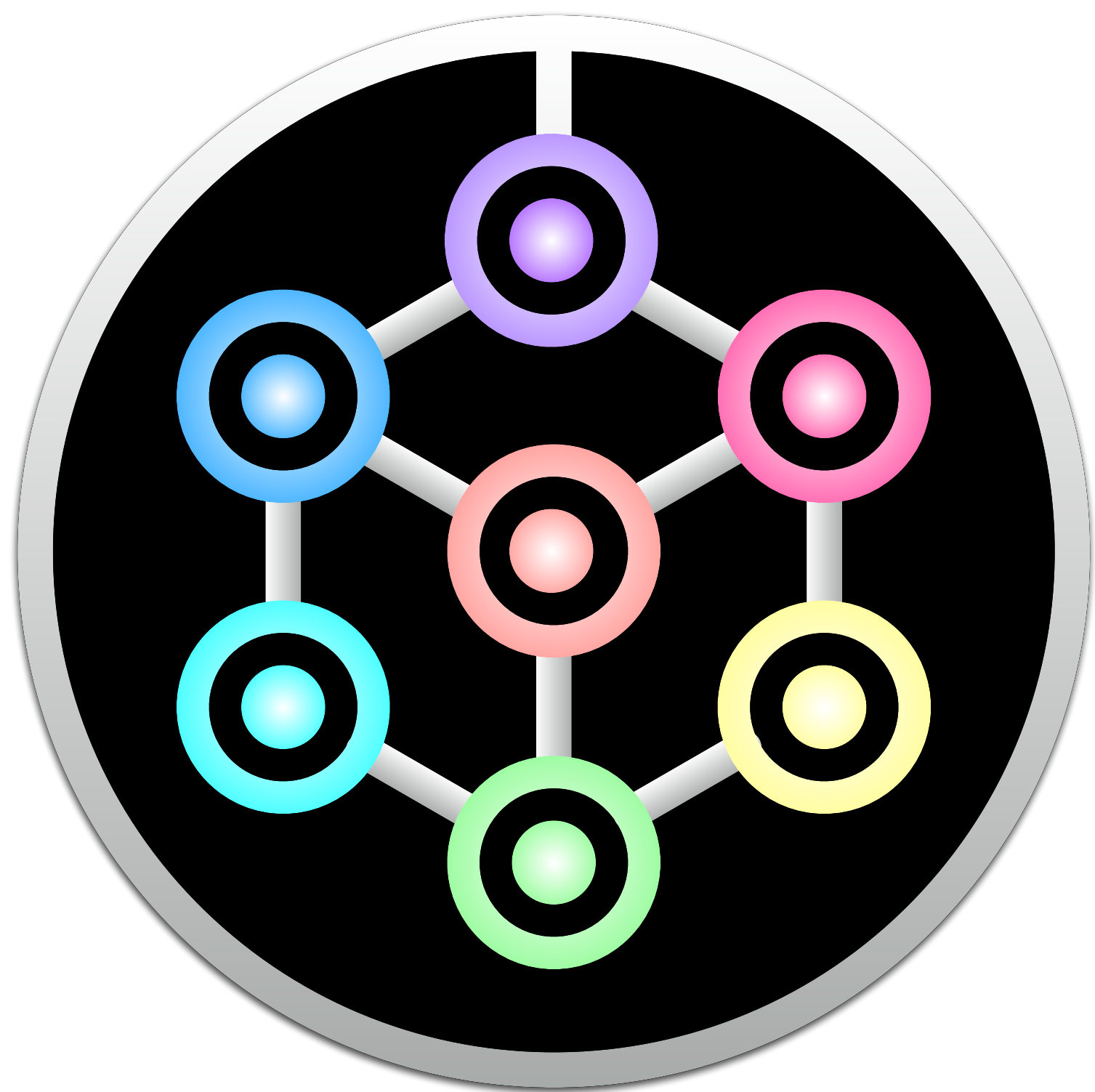


努力的起点和方向



但目标是明确的：Zero Code, Low Code





我设计的编程框架

这就是我微信和微博的头像。
其实这头像是个架构图。

自然语言引擎

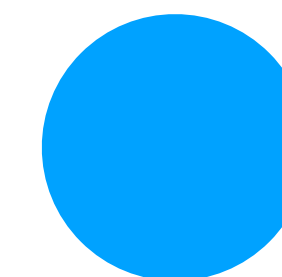
编程工具（具备部分自然语言能力，稍后展示）

GUI 框架 价值跟踪框架 高级框架

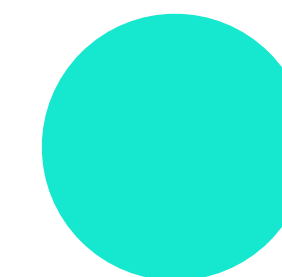
编程框架 数据框架 基础框架

虚拟机

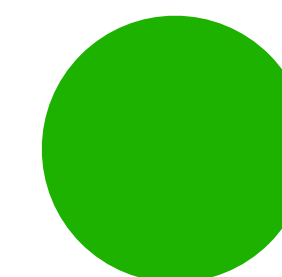
存储



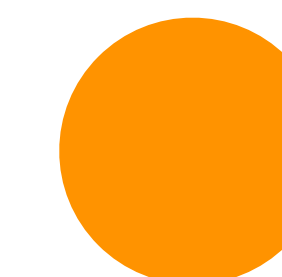
尚未开始



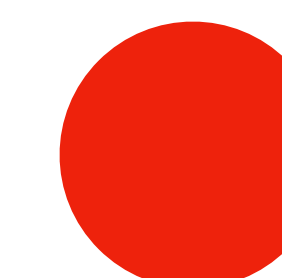
开始设计



完成设计



完成Prototype



开始用 Rust 实现

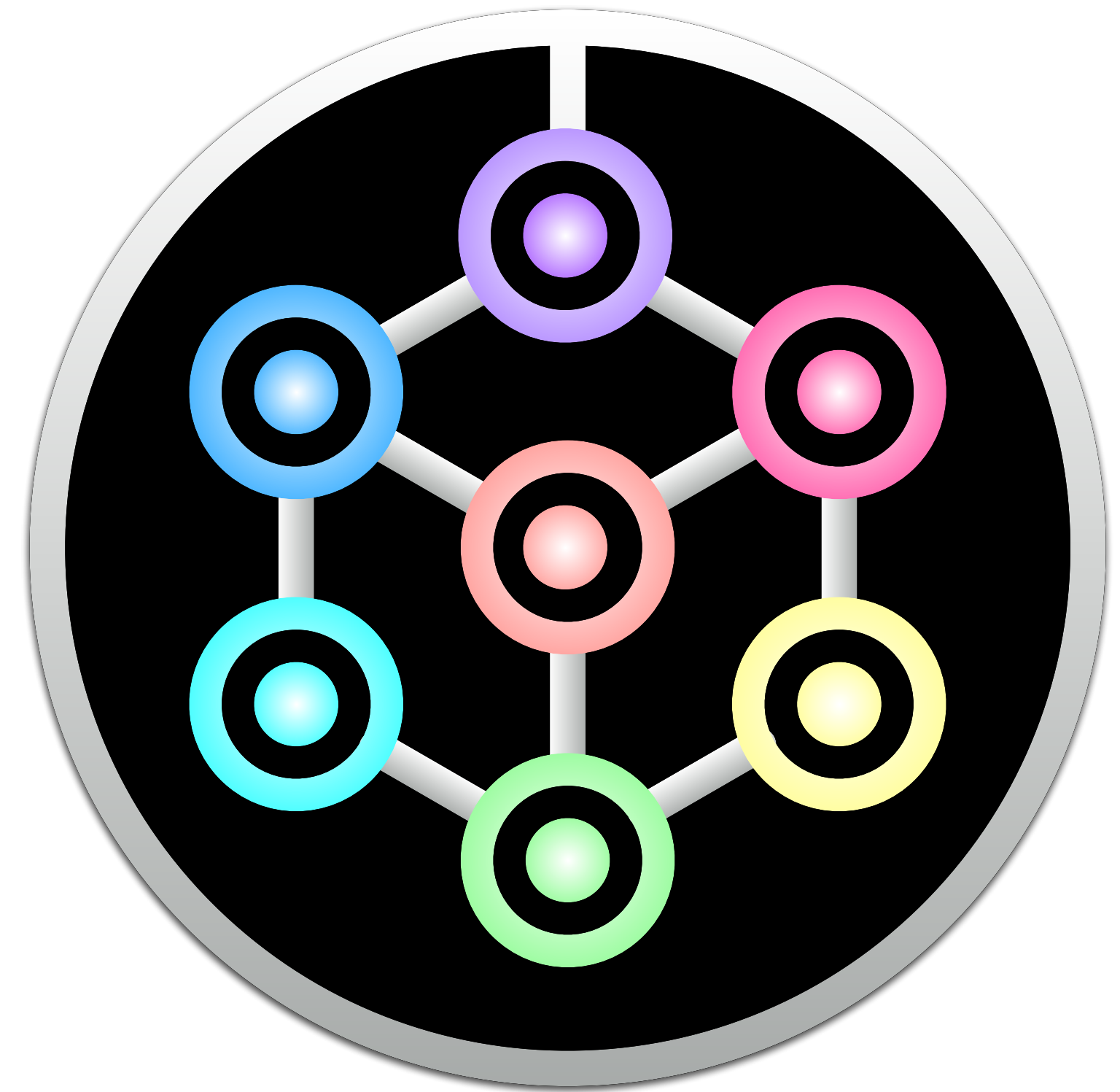
简单演示

初级的自然语言编程工具 Prototype

Discoder

应用场景?

- 区块链 + 智能合约
- 量化交易/程序化交易
- IoT
- 少儿编程教育
- AI 机器人 (自然语言 => 机器语言)
- 企业系统重构
- 在线孵化器

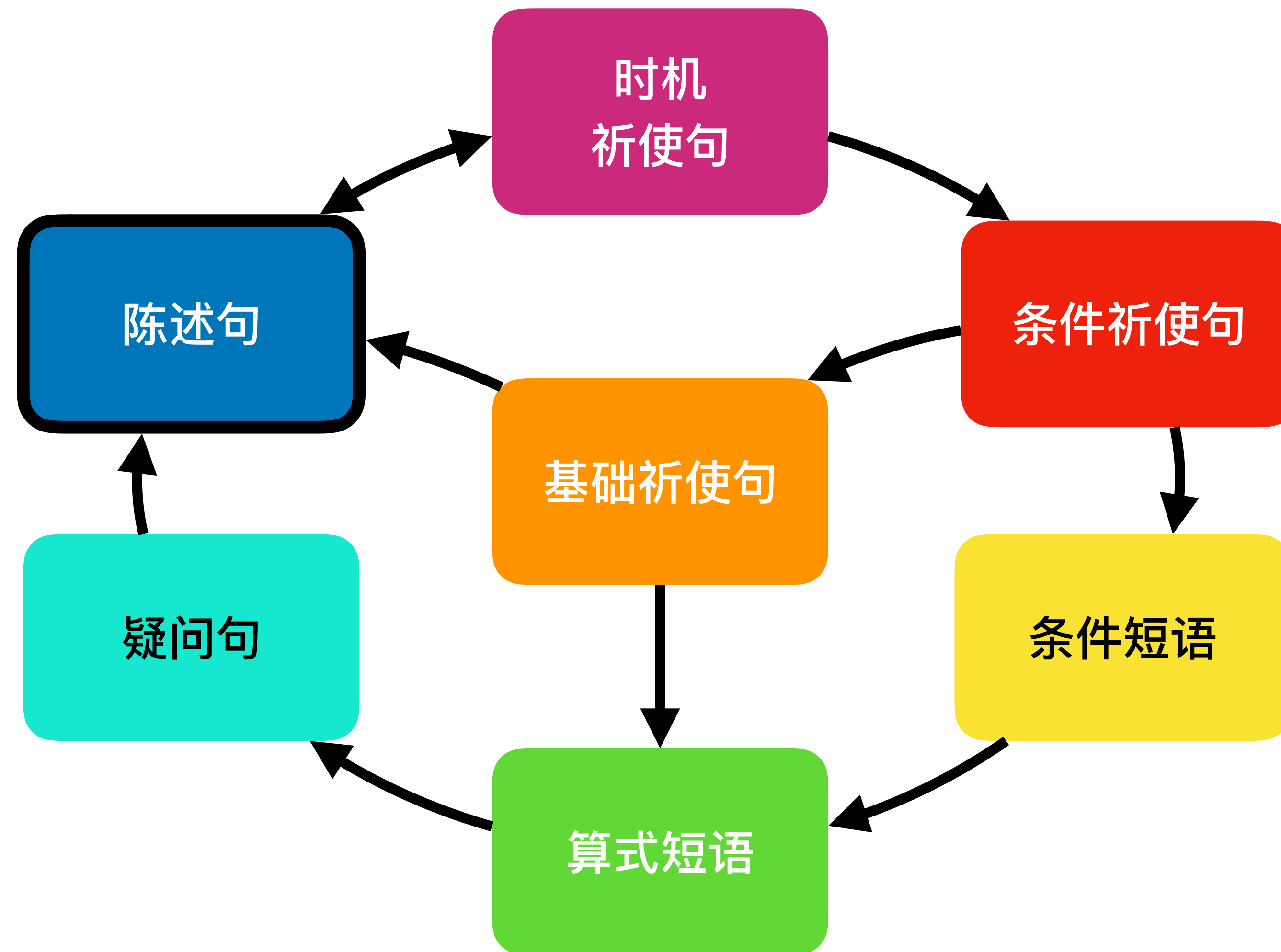


对我的价值？

- 写学术论文？
- 申请专利？
- 自己开公司做项目？
- 加入一家公司把项目实现？

UniFrame框架和自然语言的对应

后面的内容都是陈述句的部分，其他句式不描述。



陈述句的基本句式

主语

谓语

宾语

Flight BA1234

arrived at

Heathrow Airport

陈述句的嵌套



嵌套的例子

(((Flight BA1234 arrived at Heathrow Airport) on 12-Aug-98) at 10:25am)

第一个陈述句

第二个陈述句

第三个陈述句

一个应用的例子

An Internet retail bookseller operates through legal entities in various countries. Any legal entity may sell books to anyone. People are required to register with the legal entity before they can purchase. For copyright and legal reasons not all books are sold in all countries, so the books that each legal entity can offer a customer depend on the customer's country of residence.

Each legal entity sets its own prices in local currency according to the customer's country of residence. Price increases may be recorded ahead of the date that they become effective. Customers are awarded points when they buy, which may be traded in against the price of a purchase. The number of points awarded for a given book by a legal entity does not vary with the currency in which it is priced.

Legal entity sells Book

... worth Points

... in Country

... ... from Date

... at Price

Person lives in Country

Person customer of Legal entity

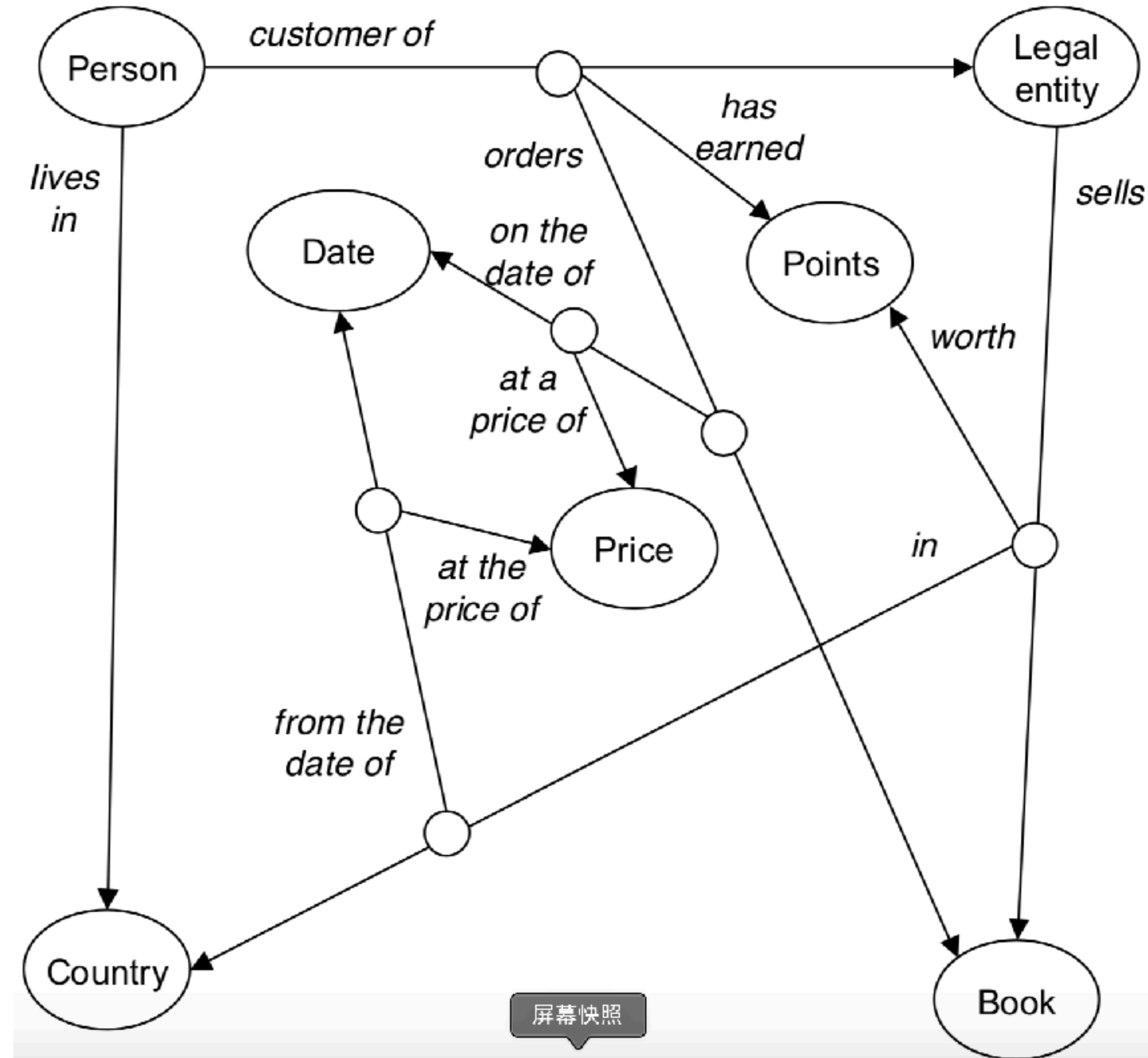
... has earned Points

... orders Book

... ... on Date

... at Price

不重复，所以不是树，是图，但不同于图数据库



数据结构： 两张表

Items	
Identifier	Name
77	Flight BA1234
08	Heathrow Airport
32	12-Aug-1998
48	10:25am
12	arrived at
67	on
09	at

Links			
Identifier	Source	Verb	Target
74	77	12	08
03	74	67	32
64	03	09	48

推荐阅读

书名：The Associative
Model of Data, 2nd Ed.

作者：SIMON
WILLIAMS

1. LIMITATIONS AND OPPORTUNITIES
2. DATA, DATABASES AND DATA MODELS
3. THE EARLY DAYS
4. THE OBJECT MODEL
5. CRITERIA FOR A NEW DATA MODEL
6. INTRODUCING THE ASSOCIATIVE MODEL
7. CONCEPTUAL LAYER
8. LOGICAL LAYER
9. METACODE
10. LANGUAGE FOR THE ASSOCIATIVE MODEL
11. BUSINESS RULES
12. BACKGROUND TO THE ASSOCIATIVE MODEL
13. CHALLENGES OF THE ASSOCIATIVE MODEL
14. THE RELATIONAL MODEL REVISITED
15. BENEFITS OF THE ASSOCIATIVE MODEL