1. 为什么要知识图谱

计算机一直面临着这样的困境——无法获取网络文本的语义信息。尽管近些年人工智能得到了长足的发展，在某些任务上取得超越人类的成绩，但离一台机器拥有一个两三岁小孩的智力这样一个目标还有一段距离。这距离的背后很大一部分原因是机器缺少知识。为了让机器能够理解文本背后的含义，我们需要对可描述的事物(实体)进行建模，填充它的属性，拓展它和其他事物的联系，即，构建机器的先验知识。

知识图谱是使用图形化的方式展现知识的内部结构以及外部关系的知识表示形式，是语义网的重要基础设施。当今的互联网正从仅包含网页和网页之间超链接的文档万维网转变成包含大量描述各种实体和实体之间丰富关系的数据万维网。

利用知识图谱为查询词赋予丰富的语义信息，建立与现实世界实体的联系，从而帮助用户更快找到所需的信息。

1. 数据库和深度学习有什么关系（数据库天然和人工智能在一起）

数据库对深度学习：大量的训练集和数据模型提高了深度学习的准确性，但是副作用却是训练过程的代价的增大。正因如此数据库社区累积的大量的系统优化的经验可以用于深度学习训练集的优化。例如分布式的计算和存储在深度学习的应用，①优化调度：通过优化数据库的事务执行和查询方法，来提高深度学习中相互独立问题的计算速度。例如数据库通过代价模型来评估查询计划的好坏。我们也可以建立一个深度学习代价模型来查找最优的计算次序。②储存管理：深度学习需要大量的存储空间，可以通过优化页面和cache来优化调度；通过添加数据库日志取代静态的进行深度学习的运行分析，通过垃圾回收机制来优化GPU存储管理。③并发性和一致性：数据库技术可以提供灵活的一致性用于深度学习的分布式训练。

深度学习对数据库中的数据进行分析，可以更好的发现哪些数据更重要，哪些数据的作用没有特别大，这样对于数据库的进一步优化会起到十分关键的作用。数据库系统设计的好坏也直接影响着深度学习训练集的正确性和效率，深度学习中的数据量庞大，正确性要求高，故一个好的数据库系统对于一个深度学习分析过程影响极为重要。

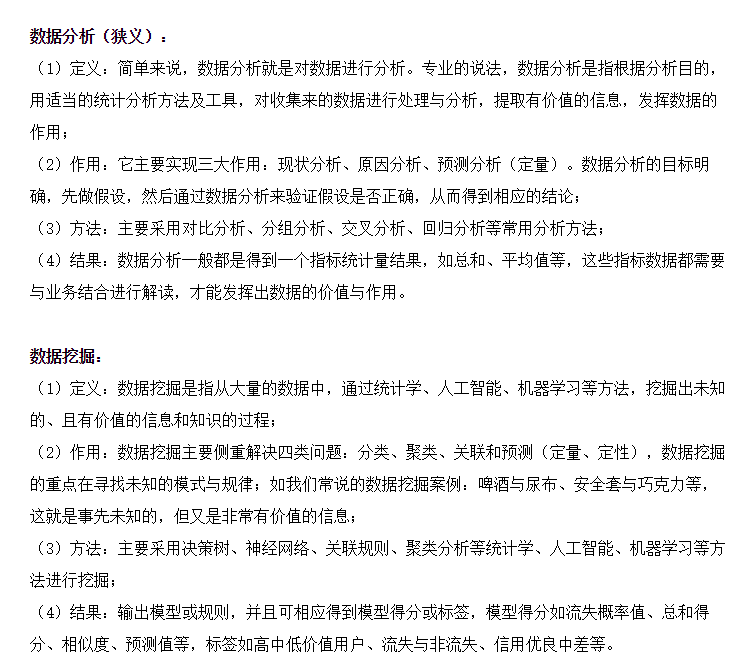
深度学习对数据库：深度学习结合了深度学习和计算机视觉领域，对于数据库中的某些非确定性问题，如知识融合，众包，查询推导等可能优化的较好：①自然语言查询接口：自然语言查询接口因为其在非专家系统中的美好应用前景和大量的需求而一直被讨论，但难点在于如何让数据库系统来理解自然语言查询的语义。而深度学习的RNN模型在解析自然语言领域有很好的表现，因此可以将RNN模型用于将自然语言查询语句解析为SQL查询语句，加以提炼用于数据库查询②查询执行计划：对于QueryPlan 的优化是一个传统的数据库难题，我们可以训练一个Query Planner可以通过学习大量的SQL查询语句及其最优的执行顺序对新输入的SQL查询语句也可以产生相似的最优解。

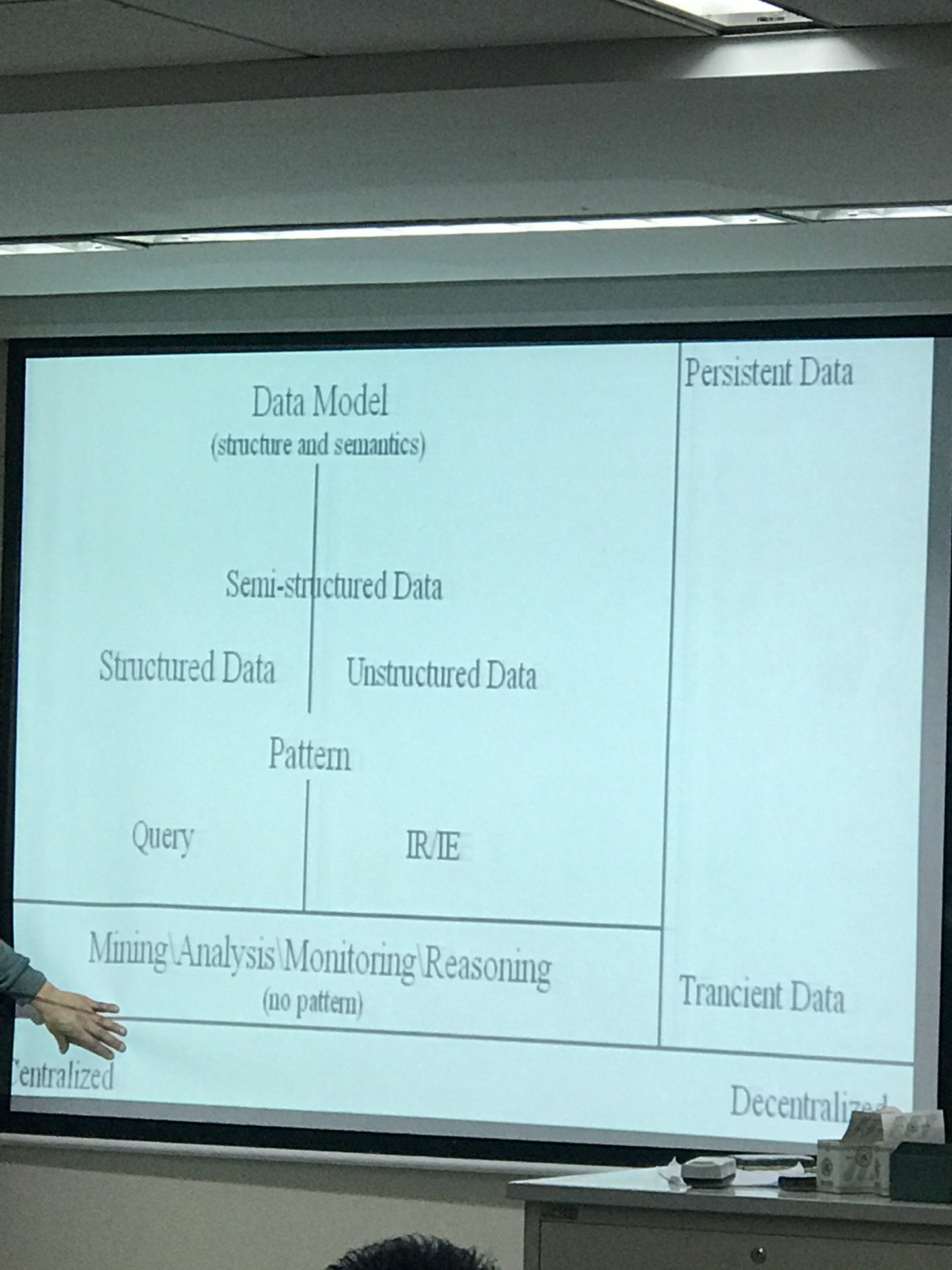
1. 数据库的逻辑模型（从已知到未知）（搜了一下三层模型，不知道已知到未知什么鬼。。。）

根据数据的不同抽象层次，数据库有三级模式：物理模式（内模式）在物理层描述数据库中全体存储结构和存取方法，而逻辑模式（概念模式）则在逻辑层描述数据库中全体数据的逻辑结构和特征。在视图层也可分为若干模式，称为子模式（外模式），它描述了数据库用户能够看见和使用的局部数据的逻辑结构和特征。通常一个数据库只有一个物理模式和一个逻辑模式，但是子模式有若干个。

1. 数据库三四十年取得的成就
2. 网状数据库
3. 关系数据库
4. 事务处理，数据库系统实现
5. 关系数据库系统的成就
6. 关系数据模型：数据的语法和语义
7. 索引：化解大海捞针的问题，查询导向相关数据
8. 查询优化，代价较小的执行计划
9. 并发调度：吞吐量
10. 分布的可扩展性：垂直扩展性和水平扩展性
11. 发展历程：
12. 逻辑数据库：Datalog(谓词=关系，事实=关系数据）
13. 面向对象数据库：数据类型和操作的定义
14. XML,RDF数据库：半结构数据
15. NoSQL：大数据存储（键值模式）
16. 封闭假设，中心化和一致性
17. 全部事实存储在数据库表格中
18. DBA拥有全部权限
19. 多个副本一致性（相同值），逻辑一致性（断言约束）
20. 什么叫query，mining，analysis

查询得到的是数据库存在的，最直接的数据





数据模型是数据特征的抽象，数据模型从抽象层次上描述了系统的静态特征，动态行为和约束条件，为数据库系统的信息表示与操作提供了一个抽象的框架。数据模型描述的内容有三部分：数据结构，数据操作和数据约束。

结构化数据简单来说就是数据库是由二维表结构来逻辑表达和实现的数据，严格地遵循数据格式与长度规范，主要通过关系型数据库进行存储和管理。

半结构化数据就是介于完全结构化数据和完全无结构的数据之间的数据，XML、HTML文档就属于半结构化数据。它一般是自描述的，数据的结构和内容混合在一起，没有明显的区分。

非结构化数据库是指其子段长度可变，并且每个字段的记录又可以由可重复或不可重复的子字段构成的数据库，用它不仅可以处理结构化数据，而且更适合处理非结构化数据，包括所有格式的办公文档、文本、图片、xml、html，各类报表、图像和音频/视频信息等等。

1. 人工智能和区块链

保持数据一致性是系统实现基本要求：一是关系依赖的数据保证某种约束条件，二是数据的多个副本保持一致性。随着互联网技术进一步发展，互联网的去中心化宗旨逐渐得到重现。在去中心化，数据的一致性有重新的定义和诠释，需要在参与主体间保持数据的一致性形成共识，这就需要解决故障攻击外的恶意攻击，即拜占庭协议和sybil攻击。从形式上看，数据的共识性要求更多副本的一致性，从而在参与主体上实现存在性证明，而这种证明可以应用于数字货币和交易以及支付等具有价值属性的人类活动。这是目前比较成功的应用案例。数据的一致性和共识性是在去中心化环境下对数据一致性的不同语义，目前多共识性的基本实现方法通过在每个节点的完全冗余备份实现数据的共识性，也有一种实现方法是在通过paxos协议实现备份的一致性，再通过拜占庭协议实现共识性，就可以回归到分布式数据存储系统上实现共识性，从而实现数据共识性，这个过程实现从数据一致性到数据共识性。

因此也就解释了数据库遇到区块链要解决的根本问题是：将中心化环境中的数据一致性拓展到去中心化环境中的数据共识性，从数据库发展的历史中，当前数据库遇到区块链：一方面，数据库需要解决的问题是从一致性拓展到共识性的技术方法；另一方面，数据的技术方法可以促进当前的区块链技术发展，数据对共识性的理解是对一致性的拓展，一是副本数目的飙升和恶意攻击；二是要扩充区块链中仅仅关于存在性共识的语义，需要实现在关系模型语义下的一致性、共识性，这种思路有希望解决通过完全冗余副本来实现存在性共识的低效机制。