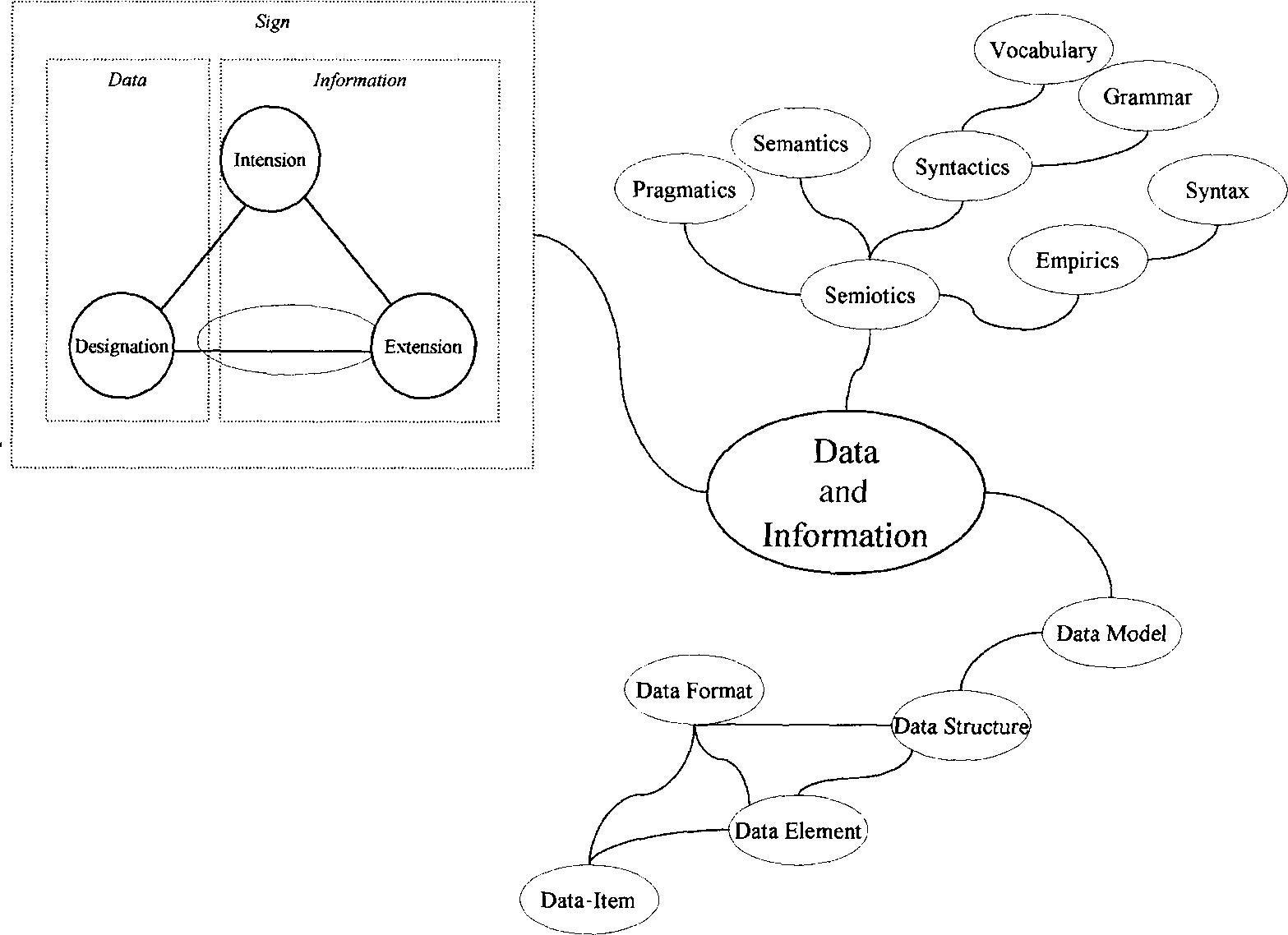
**数据和分析**

在有数据之前建立理论是一个重要错误。不知不觉地，人们开始歪曲事实以适应理论，而不是以事实为依据。



学习目标

在本章的最后，读者将能够:

根据符号的概念区分数据和信息。

描述一个符号的四个层次。

区分语法和语义，并解释这个区别与数据库模式概念的相关性。

**2.1 简介**

在第一章中，我们将数据定义为关于某些话语领域的事实或积极论断。根据这个定义，数据库是一个逻辑上有组织的关于某个域的事实集合。数据——数据的单位——是用来表示某物的符号或一组符号。符号和它们所代表的东西之间的关系就是我们所说的信息的本质。因此，信息是解释数据提供语义的数据。在本章中，我们使用符号的概念和研究符号的学科——符号学——来帮助我们更清楚地解释数据和信息之间的差异。大量的数据库工作，以及信息系统的工作，都是在没有检查数据和信息之间的关键区别的情况下进行的。因此，读者可以跳过本章，而不必担心无法开发有效的数据库系统。然而，对于这个问题的理解很重要，原因有很多:

它与数据库开发的本质有关。我们将讨论需求规格说明技术，如实体关系图(第16章)是符号系统。

它帮助我们更清楚地了解信息通信技术系统和信息系统中数据库系统的地位(第4章)

它帮助我们理解现代商业中数据的关键地位，以及有效规划(第21章)和管理数据(第23章)的需要。

**2.2 说明**

信息的概念是一个非常模糊的概念，开放到许多不同的解释(Stamper, 1989)。在计算文献中流行的一个概念是数据处理的结果:数据的汇编、分析或汇总。这种对化学蒸馏的信息的概念是有用的，但忽略了人类对信息的理解的重要地位。

我们认为数据和信息都体现在符号的概念中，因此符号和符号系统是数据库系统的基本内容。

符号是任何重要的东西。从某种意义上说，人类做的每件事在某种程度上都是有意义的。人类发现自己的世界与信号系统共振。符号系统是任何有组织的标志集合。日常口语可能是一个信号系统最容易接受和最复杂的例子。然而，在大多数其他形式的人类活动中都存在着符号，因为符号对于人类交流和理解的过程是至关重要的。

注意单词的符号与英语的显著性之间的联系。这些词显然有相同的词根。标志意义的概念不能与人脱离。不同的人会发现不同的事情有意义。许多这种解释上的差异是由于交流的语境和文化差异造成的。

**2.3 符号学**

符号学或符号学是研究符号系统的符号或更精确的符号。符号和符号系统可以被认为是四个相互依赖的水平(Stamper, 1989):语用学，语义，句法和经验。这些构成了符号学的四个主要分支。

**2.3.1 语言学**

语用学是对一般语境和交际文化的研究，或对人类理解的共同假设。为了在人类之间进行交流，符号必须存在于共同理解的语境中。正如我们将要看到的，在一群人之间必须达成一致的期望，关于符号和所指的概念或概念。语用学是对这种相互理解的研究。许多语用学可以被看作是对文化的研究——在特定语境下，人类交际行为的共同期望。

**2.3.2 语义论**

语义学是研究符号意义的研究——符号与行为之间的联系。语义学可以被认为是研究符号与它们的指称或概念之间的联系，尤其是符号与人类行为的关系。

**2.3.3 语法学**

句法是对符号系统的逻辑和语法的研究。语法是用来研究物理形态而不是符号的内容。

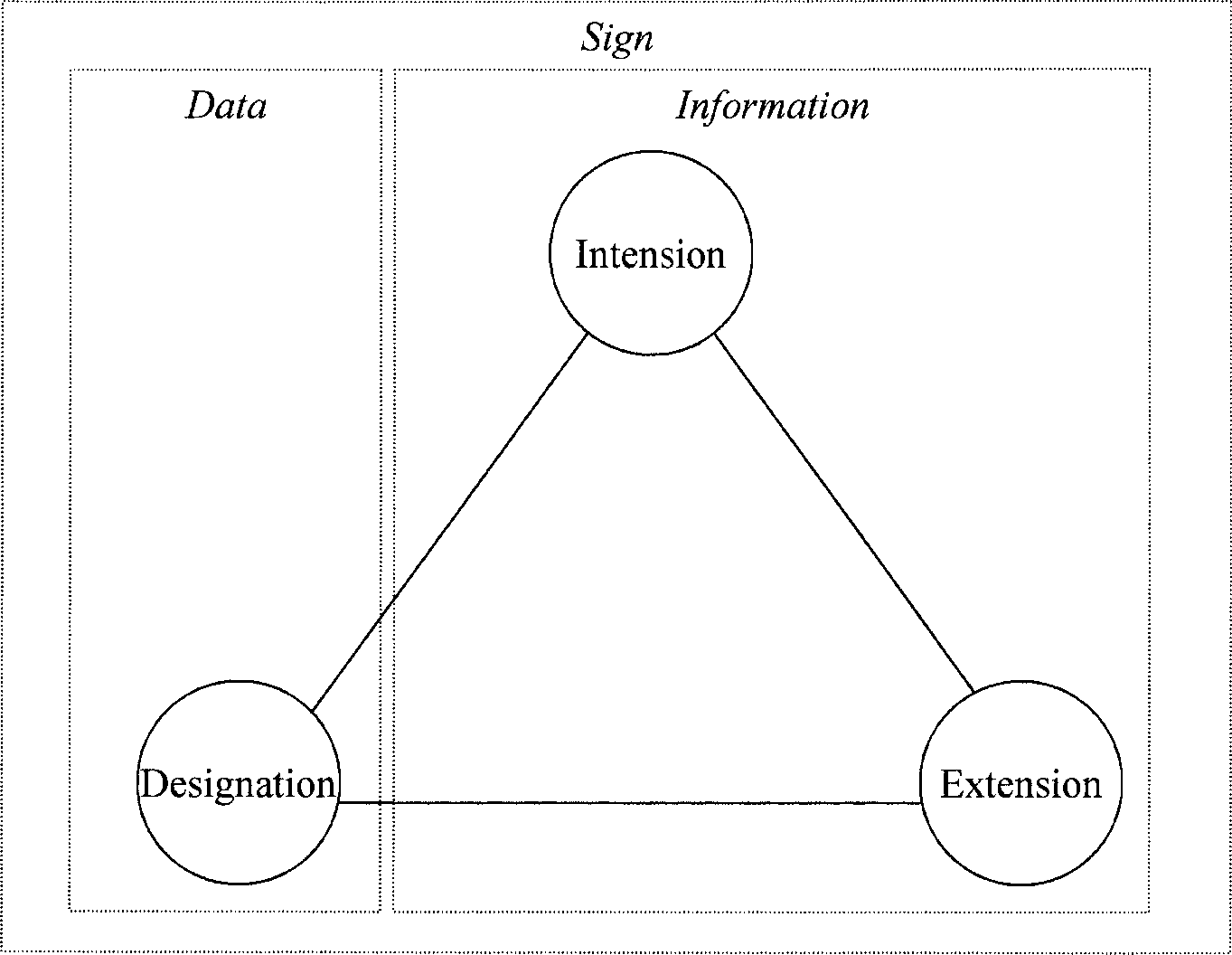
**2.3.4 经验**

经验是对交际媒介物性特征的研究。经验主义致力于研究通信信道及其特性，例如声音、光、电子传输等。

**2.4 语义。**

信息与数据之间的关系是在语义的意义上的。语义学是研究符号所指的东西。交流涉及到符号的使用和解释。当我们交流时，发件人必须将他或她的意图具体化。在面对面的交谈中，这将涉及到使用语言符号。信息的接收者必须解读信号。换句话说，他或她必须为信息的符号赋予一些意义。语义学关注的是这个赋予符号意义的过程。一个简单的语义模型是一个符号可以被分解成三个组成部分，通常被统称为“意义三角”(Sowa, 1984)(见图2.1):

•名称。符号的指定是指某些概念的符号(或符号的集合)。一个符号的指定有时被称为表示者，即表示某物的符号。



**图2.1意义三角形。**

扩展。符号的扩展是指概念在某些方面上覆盖的现象的范围。扩展有时被称为指示物或意指——指的是所指。

内涵。符号的内涵是属性的集合，在某种程度上表征了扩展中的现象。内涵是意义的想法。

标识或符号等同于信息系统的经典语言中的数据。一个数据，一个单一的数据项，是一组用来表示某物的符号。信息特别发生在“标准”中，即在指定和内涵之间，以及内涵及其延伸之间。

不应该用“三角形”来表示符号具有内在的意义。一个标志可以代表任何一个特定的社会群体选择它的意思。在不同的社会环境中，同样的标志可能意味着不同的东西。作为手语翻译，人类非常精通在特定的语境中为一个符号指定正确的解释。

**2.5 语法学**

人类交流中最重要的一种符号系统是语言。语法涉及到语言的所有这些表征方面。所有语言都包括以下要素:

词汇。一种语言术语的完整列表。

语法。正确表示术语的操作规则及其在语言句子结构中的使用。我们可以区分两大类语言:自然语言和正式语言:

自然语言。这些语言用于日常交流，包括英语、法语和威尔士语。自然语言随着时间的推移在特定的语言群体中不断发展，并且确实在不断地发展。因此，在词汇、语法和句法方面，它们往往极其丰富和复杂。

正式语言。这些是通常为某些特定目的而构建的人工语言。因此，他们往往有更简单的词汇，整句,语法和句法。在信息系统领域的工作有很多这样的正式语言。

正式语言中最发达的一些例子是那些在形式逻辑中发展起来的，包括命题和谓词逻辑。这些语言很有用，因为它们代表了对话语的限制，并对它们的属性进行了推理。

**2.6 模式。**

话语世界(UoD)，或话语领域，有时被用来描述一组符号(通常是语言学术语)被社会团体或团体持续使用的语境。对于信息系统领域的工作，重要的是要对这些术语的结构有一个详细的了解。在数据库工作中，这种结构称为模式。图式是对某些UoD进行抽象描述的尝试，通常是用一种正式的语言(如逻辑)来描述。

在经典的亚里士多德逻辑中，符号的内涵是一种适当的关系的集合，它可以被分成两组:定义的属性，所有的现象都必须有;以及phenom- ena可能有的或没有的特性。定义属性的性质是这样的，它是客观地决定一个属性是否有特定的现象。在经典逻辑中，这样的属性是由谓词(第9章)指定的。因此，在这个视图中定义属性决定了一个符号的成员和非成员之间的一个明显的边界。

数据库模式由一系列符号组成。因此，模式是一些现实的表示。尽管在传统意义上，内涵与外延之间的联系还有其他的解释，但图式中的符号遵循亚里士多德式的解释。在数据库中，我们必须能够提供一个明确的程序来判断一个现象是否是一个符号的成员。

**2.7 数据结构、数据项、数据元素和数据格式**

要构建一个模式，我们必须有一个符号系统或者一种我们可以使用的语言。这种语言必须有一致的词汇、语法和句法。用于定义模式的正式语言通常被描述为数据模型。数据模型建立了一套组织数据的原则。这组定义数据模型的原则可以分为三个主要部分:

数据定义：一组有关数据结构的原则。

数据操作：有关如何操作数据的一组原则。

一般而言，任何数据模型中数据定义的语法都倾向于使用数据项、数据元素和数据结构的层次结构(Tsitchizris和Lochovsky, 1982)。

数据项是数据组织的最低级别。数据项是任何数据模型中的原子结构。换句话说，就是这个构造，不能再分开了。

数据元素是数据项的逻辑集合。

数据结构是数据元素的逻辑集合。

数据模型中的数据项、数据元素和数据结构的一些语义是由数据格式的概念提供的。数据项的格式通常被称为数据类型。给定数据项的一些含义体现在它的数据类型中。数据类型声明在数据项上可能的有效操作的范围。

商业信息系统中需要的大部分数据都是我们所说的标准数据。标准数据是根据一些标准数据类型定义的。数据类型作为数据的分类，它不仅定义了某些数据项的格式，而且还定义了我们可以在该项目的数据上执行的允许的操作集。大多数信息系统应用程序都使用了许多标准数据类型。这些包括:

文本：由字母和其他字符组成的符号串。

数字：包括整数、十进制和实数。

时间单位：包括日期，秒，分钟和小时。

信息系统现在被用来捕获、存储和操作比上面讨论的标准数据类型更复杂的数据类型。这是为了使这些系统能够处理不同的媒体。这些复杂的数据类型包括:

图像：图片和照片

音频：各种形式的声音数据。

视频：各种形式的移动图像。

现在，信息系统也需要处理复杂的数据结构和复杂的数据项或元素。组织希望能够定义用于通信的文档结构。像发票或合同这样的文档是一个复杂的数据结构，它可以由文本、数字数据和图像组成。另外，不同的组织可能会使用不同的格式。因此，同样的工业部门也面临着对这些文件制定标准的压力。使用一种称为XML可扩展标记语言的正式语言来指定这些标准。XML是Internet语言HTML的一种变体，在第39章中有更详细的讨论。

**2.8 案例研究:大学模块描述**

信息对于大学的成功运作非常重要，它将包括学生、职员和模块的信息。在这个环境中，模块描述是一个关键的信号系统。作为一种组织沟通，模块描述可以在多个层次上进行分析:

语用学——对交流的一般语境和文化的研究，或对人类理解的共同假设。围绕使用模块描述的关键假设是，教育经验可以用模块来打包或分组，并在离散单元中交付和评估。有效模块成为高等教育经验的基石。它们是由大学提供的教育项目。

语义学——研究符号或符号系统的意义。在大学设置中，学生和工作人员倾向于将模块描述视为。非正式的合同。他们特别定义了教育经历的交付内容。他们还定义了可能用于评估学生的机制。

综合策略——研究符号系统的逻辑和语法。在句法级别上，模块描述可以被认为是一个数据结构。一个模块描述的标准格式通常会在特定的大学中制定，包括模块代码、模块名称、层次、学习成果、教学大纲、评估模式和建议阅读等数据元素。这些元素中的数据项很可能会存储以书面文本作为关键符号的字符。

经验——研究社区关系中媒介的物理特性。模块描述可以以物理形式提供。作为学生或模块手册的一部分。在现代大学，它也可以通过大学内联网电子化，并且可以存储在数据库系统中，便于检索和维护。

**2.9 总结**

一个数据库可以被认为是一个有组织的数据集合，它代表着一些话语的宇宙。

数据是事实。数据是数据的一种单位，是用来表示某物的一种符号或一组符号。

数据本身是没有意义的。要证明有用的数据必须被解释。信息是数据解释。信息是放置在有意义的上下文中的数据。信息是具有指定语义或意义的数据。

数据和信息的区别体现在符号的概念上。

符号学或符号学是对符号的研究，或者更准确地说是符号系统的研究。

从语用学、语义学、句法和经验等四个相互依存的层次可以考虑符号和符号系统。

话语世界(UoD)，或话语领域，有时被用来描述一组符号(通常是语言学术语)的上下文。常被社会团体或团体使用。对于信息系统领域的工作，重要的是要对这些术语的结构有一个详细的了解。在数据库工作中，这种结构称为模式。图式是对某些UoD(通常是正式语言)的抽象描述的尝试。

要构建一个模式，我们必须有一个符号系统或者一种我们可以使用的语言。这种语言必须有一致的词汇、语法和句法。正式的用于定义模式的语言通常被描述为数据模型。

一般而言，任何数据模型中的数据定义语法都倾向于使用数据项、数据元素和数据结构的层次结构。

**2.10 活动**

从符号学的角度分析你所知道的一些组织沟通。举个例子，在什么情况下，学生可以把评估作为一个标志?它在语用学、语义学、句法和经验方面有什么意义?

阅读第3部分后，尝试详细说明正式语言SQL的一些词汇、语法和语法。例如，选择状态的语法是什么?

使用“意义三角”来分析一项数据——订单号，从银行账户中提取——就一个符号的概念而言。例如，订单号的设计、内涵和延伸是什么?

**2.11 参考**

Sowa, J.F. (1984). Conceptual Structures: Information Processing in Mind and Machine. Reading, MA, Addison-Wesley.

Stamper, R.K. (1989). Information in Business and Administrative Systems. London, Batsford. Tsitchizris, D.C. and F.H. Lochovsky (1982). Data Models. Englewood Cliffs, NJ, Prentice-Hall.