

自然语言与计算机程序设计语言对比分析及运用

张志霞

广东理工学院 信息技术学院, 广东 肇庆 526100

DOI: 10.61369/TACS.2025070017

摘 要 : 自然语言与计算机程序设计语言虽然形成过程不同,但是就其本质而言,都属于交流、沟通工具,具有一致性。从这一本质出发研究两者异同,探讨基于树状模型的复杂自然语言查询转 SQL 技术,为相关工作的开展提供理论基础与实践方法。从核心结构来看,两者目的与使用对象、语法与结构特性均有所不同,前者主要用于人与人相互交流,具有一定灵活性和容错性,后者用于人与计算机交流,是高度严格的,这决定了两者演化与标准化方式也有所不同。同时,两者在符号与结构基础、生成性能力等方面也存在一定共性。两者之间的差异与共性,为计算机自然语言处理技术发展带来更多可行性,比如自然语言查询转 SQL(NL2SQL)技术将用户以自然语言形式提出的查询请求转换为结构化查询语言(SQL)语句,以便数据库能够准确理解和执行这些查询。该技术通过深入分析自然语言的语法结构与语义信息,利用树状模型对查询语句进行解析与重构,从而生成符合 SQL 语法规则的查询语句。

关 键 词 : 自然语言; 计算机程序设计语言; 对比; 运用; 树状模型

Comparative Analysis and Application of Natural Language and Computer Programming Language

Zhang Zhixia

School of Information Technology, Guangdong University of Technology, Zhaoqing, Guangdong 526100

Abstract : Although natural language and computer programming language have different formation processes, they are essentially tools for communication and have consistency. Starting from this essence, this paper studies the similarities and differences between them, and discusses the technology of converting complex natural language queries into SQL based on tree models, so as to provide theoretical basis and practical methods for the development of related work. From the perspective of core structure, the two differ in purposes, user groups, grammar and structural characteristics. The former is mainly used for communication between people, with certain flexibility and fault tolerance; the latter is used for communication between people and computers, and is highly rigorous. This determines that their evolution and standardization methods are also different. At the same time, they share certain commonalities in symbol and structural foundations, generative capabilities, etc. The differences and commonalities between them bring more feasibility to the development of computer natural language processing technology. For example, the Natural Language to SQL (NL2SQL) technology converts query requests put forward by users in natural language into Structured Query Language (SQL) statements, so that databases can accurately understand and execute these queries. By in-depth analyzing the grammatical structure and semantic information of natural language, this technology uses tree models to parse and reconstruct query statements, thereby generating query statements that conform to SQL grammar rules.

Keywords : natural language; computer programming language; comparison; application; tree model

引言

自然语言指人与人相互交流运用的语言,它形成于人类生存发展过程之中,属于交流、沟通工具。计算机程序设计语言是人与计算机交流运用的工具,技术人员通过计算机程序设计语言设计计算机程序,命令计算机完成相应任务。从本质上来看,两者具有一定一致性,围绕其一致性进行研究,能够促进自然语言学习、教学方法在计算机程序设计语言课程中的应用,继而降低人们掌握计算机程序设计语言的难度与门槛。

基金项目: 广东理工学院校级科研项目(025YBZK019)。

作者简介: 张志霞(1979-),女,硕士,副教授,研究方向为计算机网络与数据通信。

一、核心差异对比

（一）目的与使用对象

自然语言的使用对象主要是人类，在不同文化、地域背景下，人们运用自然语言进行日常交流、文学创作、学术探讨等各类活动，以满足情感传递、知识共享、社交互动等需求。它具有高度的灵活性和适应性，能够根据不同的语境和交流对象进行动态调整^[1]。

程序设计语言是形式化、结构化的符号系统，服务于人机交互，需通过编译器转化为机器指令，逻辑必须精确无歧义。程序员借助程序设计语言来描述计算机的操作步骤和算法逻辑，从而让计算机按照预定的指令执行任务，实现各种软件和系统的功能。其使用注重严谨性、规范性和高效性，以确保程序能够准确无误地运行^[2]。

（二）语法与结构特性

自然语言允许灵活性和容错性，如同义词、省略句和语境依赖（如“我想打你”中主语位置决定语义）。语法形成于自然语言的长期使用过程，是一系列被大众所认同的，约定俗成的组合规则。它们的存在，能够保证通过相同的字词可以表达出大家都理解、认可的意思。语法规则规定了字（词）之间的排列顺序、组合方式以及相互关系，使得自然语言具有了系统性和条理性，对句子组成过程具有规范、约束作用^[3]。

程序设计语言语法高度严格，需要遵循固定符号、关键词和控制结构的规则，例如 if-else 语句必须闭合，否则无法编译运行。在计算机程序设计语言中，字（词）是指计算机指令和其他自然语言符号，主要可以划分为系统定义字（词）、用户自定义字（词）两大类。前者是计算机程序设计语言中预先设定好的、具有特定功能和含义的符号，例如常见的关键字如“if”“else”“for”“while”等，它们在程序中具有固定的语法作用和执行逻辑。后者则是由编程人员根据实际需求自行定义标识符，如变量名、函数名等，一般用于在程序中代表特定的数据或操作，所以其命名需要遵循一定规则，以确保程序的正确性、可读性。与自然语言中的字（词）类似，计算机程序设计语言中的字（词）也需要准确选择、按照一定规则进行组合，从而实现特定的编程目标。

（三）歧义性与确定性

自然语言通过上下文和隐喻表达复杂语义，如“房间很冷”可能指温度或氛围。语义是自然语言中字（词）和句子所承载的意义内容，它依赖于字（词）的组合方式以及语境因素，是自然语言使用的目的。它的表达需要基于语言文化、规则，以及人们对语言的基本认知。文章、说话是人们表达语义的两种主要方式，前者通过书面语言说明或者解决问题，多用于人与人之间相互沟通、增进了解，后者多通过表情达意语句进行交流、沟通，表达内心想法，不同语句之间的关系通常呈现松耦合关系^[4-5]。

程序设计语言需完全消除歧义，每条语句对应唯一操作（例如 $x = y + 1$ 仅表示变量赋值）。在计算机程序设计语言中，算法特指为解决某一问题拟定的系统解决方案，涉及实施过程、步

骤、方法、设计、逻辑等多方面内容，能够完整且准确地描述问题解决方案。算法是计算机程序设计语言的核心要素之一，同时也是计算机程序设计的灵魂，它决定了程序如何高效、准确地完成任务。由于计算机不具备主动思考能力，所以算法与自然语言中的语义表达相比，更加注重逻辑性和精确性，要求准确定义每一步操作，且每一步操作具备相应执行顺序。编程人员进行算法设计时，需要结合具体需求选择适宜的算法策略，以保证设计出的程序可以通过最优方式解决问题。当前，排序算法、搜索算法、递归算法等都是计算机程序设计中的常用算法^[6]。

（四）演化与标准化

自然语言随社会文化自然演变，如网络用语“绝绝子”快速流行。其演化过程缺乏统一规划，受地域、群体、时代等多种因素影响，呈现出多样性和动态性特征。不同地区、不同年龄段的人群可能使用不同的词汇和表达方式，而且随着时间推移，新的词汇和语法结构不断涌现，旧的则可能逐渐被淘汰。

程序设计语言由标准化组织定义版本（如 Python PEP 规范），更新需严格遵循兼容性规则。计算机程序设计语言的应用需要遵守一定规则，也就是说需要符合语法要求，否则将会导致程序无法正常运行。人有主动思考的能力，所以自然语言即便存在语法方面的瑕疵，也能完成交流，实现准确表达。但是，计算机没有主动思考的能力，只能结合既定规则理解语言，故而无法识别语法存在瑕疵的语言。计算机程序设计语言的语法规则极为严格，它详细规定了程序中各个元素（如关键字、操作符、标识符等）的排列顺序、组合方式以及它们之间的逻辑关系。这些规则确保了程序能够被计算机准确无误地解释和执行。与自然语言语法相比，可以说计算机程序设计语言的语法更加形式化、精确化，任何微小的语法错误都可能导致程序编译失败或运行异常。因此，编程人员在编写程序时，必须严格遵守语法规则，确保程序的正确性和稳定性^[7]。

二、关键共性分析

（一）符号与结构基础

两者均包含基本语言单位（字 / 词）和层级化规则（语法 / 句法）。字（词）、是构成自然语言的最小元素、最小音义结合体，每一个字（词）都有自己独立的意义、读音，而且单字成词、两个或者两个以上的字组合成词在语言交流中都存在，后者相对常见。与自然语言中的字（词）类似，在计算机程序设计语言中，字（词）是指计算机指令和其他自然语言符号，主要可以划分为系统定义字（词）、用户自定义字（词）两大类。

语法形成于自然语言的长期使用过程，是一系列被大众所认同的，约定俗成的组合规则。与自然语言类似，计算机程序设计语言的应用需要遵守一定规则，也就是说需要符合语法要求，否则将会导致程序无法正常运行^[8]。

（二）生成能力

1. 自然语言通过有限词汇生成无限句子

这一特性源于其词汇的灵活组合以及语法规则的多样性。自

然语言中的词汇就像是一块块积木，通过不同的排列组合方式，可以构建出各式各样、意义各异的句子。句子是人们表达语义的主要载体，用于人与人之间相互沟通、增进了解，且不同语句之间的关系通常呈现松耦合关系。

2. 程序设计语言通过组合代码模块实现功能扩展

这一能力得益于代码模块的复用性和结构化设计。在计算机程序设计语言中，代码模块是预先编写好的、具有特定功能的代码段，它们可以被重复调用和组合，以实现更复杂的功能。程序员通过将这些模块按照一定的逻辑和顺序进行组合，可以构建出功能强大、结构清晰的程序。

三、实际应用场景

（一）基于树状模型的复杂自然语言查询转 SQL 技术

自然语言查询转 SQL(NL2SQL) 技术旨在将用户以自然语言形式提出的查询请求转换为结构化查询语言 (SQL) 语句，以便数据库能够准确理解和执行这些查询。树状模型作为一种有效的自然语言处理工具，能够解析自然语言查询中的语法结构和语义关系，并将其映射到相应的 SQL 语句结构上^[9]。

（二）基于树状模型的复杂自然语言查询转 SQL 技术框架

本文对 DuSQL 数据集上出现的 SQL 形式进行了归纳和分解，创新性地提出了用于 SQL 结构预测的树状统一模型。

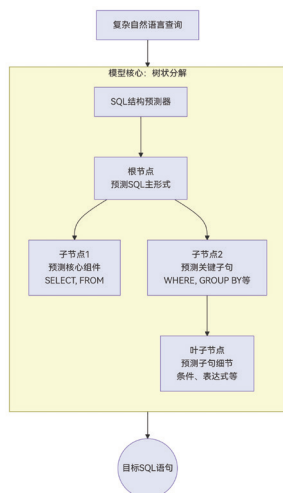


图 1 复杂 SQL 树状分解模型

（三）基于树状模型的复杂自然语言查询转 SQL 模型实现

在模型实现层面，基于树状模型的复杂自然语言查询转 SQL 技术有着独特的流程与要点。

1. 在数据预处理阶段：对自然语言查询文本进行细致的清洗与标注，以便后续模型能够精准识别和处理这些重要元素^[10]。

2. 模型构建环节：以树状结构为核心来设计模型架构，树的节点对应 SQL 语句的不同组成部分，如 SELECT 子句、FROM 子句、WHERE 子句等。

树状模型的节点概率计算可表示为：

$$P(n_i) = \sigma(W \cdot h_{parent} + U \cdot x_i + b)$$
其中， n_i 为当前节点， h_{parent} 为父节点隐藏状态， x_i 为节点特征向量， W 、 U 为权重矩阵， b 为偏置项， σ 为激活函数。

3. 模型训练环节：大量的标注数据来优化模型参数。这些标注数据包含自然语言查询及其对应的正确 SQL 语句。

SQL 语句优化可采用代价估算模型： $Cost(Q) = \sum_{i=1}^k C_i \cdot f_i(Q)$ 其中 C_i 为操作代价系数， $f_i(Q)$ 为查询 Q 中第 i 类操作的执行频率。

4. 实际应用部署：这一环节需要充分考虑模型的效率和可扩展性。由于数据库查询通常对响应时间有较高要求，模型需要能够在短时间内生成 SQL 语句，因此，可以采用模型压缩、量化等技术来减少模型的计算量和内存占用，提高推理速度。

四、总结与未来展望

综上所述，围绕“自然语言与计算机程序设计语言对比分析及运用”这一主题展开深入研究，从构成要素、应用等不同层面分析两者异同点，能够为人们研究相关领域，进行计算机程序设计语言应用、学习提供理论支持与方法借鉴。

未来，基于树状模型的复杂自然语言查询转 SQL 技术有望取得更大突破，实现更高效、更准确的自然语言到 SQL 的转换：

（1）深度学习算法的持续优化和创新将为模型提供更强大的语义理解和生成能力，使模型能够更好地处理自然语言中的歧义、隐喻等复杂语义现象。

（2）大数据和云计算技术的发展，能够丰富模型能够使用的标注数据和计算资源，进一步提升其性能和泛化能力。

参考文献

- [1] 李佳泓, 黄嘉, 古炬贤. 自然语言处理中的价值观渗透问题——以 ChatGPT 为例[J]. 情报杂志, 2024, 43(08): 66-71+65.
- [2] 顾彬彬. 自然语言处理技术在合同管理中的运用[J]. 电气自动化, 2024, 46(04): 116-118.
- [3] 刘冠麟, 赵志耘, 曾文. 一种基于神经网络的技术融合预测方法[J]. 情报杂志, 2024, 43(12): 117-124+185.
- [4] 刘雪晴. 声音识别技术在日语自然语言处理系统中的应用[J]. 电声技术, 2024, 48(07): 72-74.
- [5] 刘魂浩, 陈超. 基于自然语言处理技术的跨媒体感知与分析技术实施路径[J]. 中国信息化, 2024, (07): 58-60.
- [6] 解振兴. 基于人工智能技术的增强现实动画场景的设计与表现[J]. 佳木斯大学学报(自然科学版), 2024, 42(05): 132-135.
- [7] 刘丽莉, 李明, 罗晓兰, 等. 基于自然语言处理智能技术的中医术语研究文献计量分析[J]. 上海中医药杂志, 2024, 58(07): 1-6+14.
- [8] 张永伟, 范明威. 基于自然语言处理技术的老年人多模态语料库切分标注平台研制[J]. 语言学研究, 2024, (01): 49-61.
- [9] 卢世主, 闵子怡, 黄秋雨, 等. 自然语言处理模型与文本图像生成技术驱动下的包装设计[J]. 包装工程, 2024, 45(22): 232-241+258.
- [10] 黄冬玲, 刘源, 袁小帅, 等. 基于自然语言处理技术的中国氢能政策数据挖掘研究[J]. 中国科学院院刊, 2024, 39(06): 1032-1046.