

# 计算机专业课程体系知识图谱构建及分析

傅湘玲\* 唐轶 侯党 徐朗朗

北京邮电大学计算机学院, 北京, 100876 北京邮电大学教务处, 北京, 100876

**摘要** 通过收集国内外代表性高校的计算机专业课程相关数据,通过对课程属性以及彼此间的前置要求进行分析,提取先修课程与后续课程之间的关系,构建了一个计算机专业的知识图谱。基于该图谱分析了国内外典型高校计算机专业课程之间的脉络关系以及不同学校的课程设置差异,系统直观的展示了计算机专业课程体系完整结构,对于各个学校的计算机专业体系的科学有效设置提供了借鉴参考意义。本文构建的知识图谱也提供了公开的访问查询路径。

**关键字** 课程知识图谱, 计算机专业, 课程体系

## Construction and Analysis of Knowledge Graph in Computer Science Curriculum System

Fu Xiangling Tang Yi Hou Dang Xu Langlang

School of Computer Science (National Pilot Software Engineering School)  
Beijing University of Posts and Telecommunications  
Beijing 100876, China;  
fuxiangling@bupt.edu.cn

**Abstract**—A knowledge graph of computer science major is constructed by collecting the data of computer major courses in representative universities, and analyzing the courses' attributes and the pre requirements between them, and extracting the relationships between prerequisite courses and follow-up courses. Based on this graph, we analyze the relationships between computer professional courses and the differences between curriculum system settings in typical universities at home and abroad. This graph intuitively displays the complete structure of computer science curriculum system, and provides a reference of the scientific and effective setting of computer science curriculum system for universities. The knowledge graph constructed in this paper also provides an open access query path.

**Keywords**—Curriculum knowledge graph, Computer science major, Curriculum system

## 1 引言

课程体系是指在一定的教育价值理念指导下,将课程的各个构成要素加以排列组合,使各个课程要素在动态过程中统一指向课程体系目标实现的系统。步入全球化进程后,中国的工程教育建设面临来自国内工业界的新要求和来自国际工程教育的挑战,深化课程体系改革和建设是全面提高教育质量的基本内容和重要举措。高校课程培养目标是对学生的知识、能力和素质提出的理想预期,而课程体系在很大程度上决定了学生能呈现的知识、能力和素质结构,因此课程体系的构建问题是高校教育的核心问题。

课程体系构建过程中的关键问题包括哪些课程是主干核心课程、体系中各课程的地位如何,课程和课

程之间的关系如何,不同学校相同专业课程设置的差异性是什么。这些内容在目前各个高校各专业的培养方案中都有说明,但是都是文字描述性的,未能有直观可视化的联系能展示出来,并很难从中快速的得出不同学校课程体系之间的差异。若通过人工整理对课程体系的关系进行梳理和统计,实现数据的统计、分析及展示,不仅需要耗费大量人力,而且难以挖掘课程之间的深层次联系以及高校之间的课程体系设计差异,而知识图谱可以有效的解决上述问题:

(1)利用可视化图形界面可以有效展示专业学科的课程体系

目前许多高校的课程大纲几乎都以表格形式进行简单列举,这种方式虽然便于搜索查阅,但是难以直观地展示整个专业的课程体系结构,让人难以发现课程之间的关联关系。

(2)可以有效提高查询课程之间的关联关系的效率

\*基金资助: 本文得到{北京邮电大学 2019 年教育教学改革项目立项资助,项目编号: 2019JY-F02}资助。

课程体系涉及到大量的课程以及它们之间的关联关系,知识图谱可以有效地将这些信息组合在一起,从而把混乱的信息转换为有序且易于使用的知识,通过知识图谱中语义关系的描述,将课程大纲中散乱的信息进行整合,这样可以有效地对课程信息和课程之间的关联关系进行检索,从而提高后续任务的分析效率。

(3)可以挖掘分析不同高校之间构建课程体系的设计差异

由于受到社会发展、技术发展、人才需求等多种原因,不同高校之间的课程体系存在一定的设计差异。通过构建学科知识图谱,对多个高校的课程体系进行比对,从而有效分析其中差异,也有助于完善本校专业的课程体系。

计算机科学(Computer Science,简称CS)专业是现代高等教育中发展最为活跃的学科之一,也一直是国内外高校的热门学科,本文将以此专业为例,构建计算机专业的知识图谱进行研究。随着现代计算机技术的更新发展,CS专业也在不断发展变化。在20世纪中期,包括中国在内的许多国家采用的课程体系基本上都是参考美国。但是由于社会发展、技术发展、人才需求等原因,国内外高校CS专业的课程体系之间逐渐产生了一些差异。在1995年,中国教育部实施了新的课程体系改革措施,对CS专业的课程体系进行研究和改革,形成了分类、分层次培养CS专业学生的思路。

本文提出一种基于知识图谱的学科知识图谱构建方案,以北京大学,清华大学,浙江大学,北京邮电大学,卡耐基梅隆大学和牛津大学6所国内外高校的CS专业作为研究对象,通过收集处理各所高校的课程大纲和培养方案,利用实验数据研究学科知识图谱的构建与应用,对课程体系结构进行展示,并对课程关联关系和课程体系设计差异进行挖掘分析,为完善课程体系的构建给出借鉴意义。

## 2 CS课程知识图谱的构建

### 2.1 课程知识图谱的构建流程

知识图谱的构建主要有自顶向下(Top-Down)和自底向上(Bottom-Up)两种构建方式。自底向上的方式是指根据底层数据提取出的实体、关系、属性等概念,归纳设计出上层知识图谱的框架。本文数据来自各所高校的课程大纲和培养方案,这些数据大部分都是结构化数据,因此采用了自底向上的方法构建知识图谱,构建的知识图谱归纳为三个步骤,如图1所示。



图1 知识图谱的构建方案

在知识图谱的构建中,重要的是模式层和数据层的构建,模式层是指知识图谱本体在概念上的表示,是用于规范各类实体、关系和属性的模板。数据层是通过实际数据构建的一个包含事实数据的知识库。本课程体系知识图谱属于垂直知识图谱,对知识的专业性要求较高,所以需要先构建知识图谱的模式层,即本体建模。

### 2.2 模式层构建

知识图谱的模式层可以用  $G = (V, E)$ ,  $V = \text{vertex}(\text{节点})$ ,  $E = \text{edge}(\text{边})$  的形式表达,其具有以下特点:

- ① 包含节点和边;
- ② 节点和边具有属性(键值对);
- ③ 边具有方向,总是从一个节点指向另一个节点。

为了构建知识图谱本体,使用W3C推荐的资源描述框架(RDF),以三元组的形式表示实体、属性和关系等知识,通过对数据表的字段定义和它们之间关系的分析,从而确定了实体、关系和属性,其中包含高校和课程两种实体,2种关系(高校与课程的关系,课程与课程之间的关系),如表1和表2所示。

表1 实体类型及关键属性

实体类型	关键属性
高校(university)	高校名称
课程(course)	课程名称、课程类别、必修/选修

表2 关系类型

关系两边的实体	关系类型
高校-课程	包含关系
课程-课程	先修关系

图2展示了知识图谱本体模式层的一部分,包含代表高校和课程的节点,以及它们之间的关系。

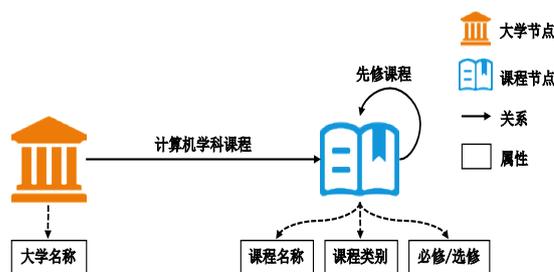


图2 知识图谱本体

### 2.3 数据层构建

#### (1) 数据处理

为使得构建的知识图谱的课程体系具有一定的借鉴意义,本文选择从国内外选择了有代表性的六所高校,分别是北京大学、清华大学、浙江大学、北京邮

电大学、卡耐基梅隆大学和牛津大学。数据采集选择从上述高校的官网搜集整理该校的计算机科学专业的课程大纲和培养方案。由于各所高校的课程大纲和培养方案的文件格式、内容形式都有所差异，通过前期处理加工将数据导入到数据库中。由于知识图谱中的数据是以三元组的形式存储，关系数据库中的这些数据并不能直接导入，所以使用 D2R 工具，将数据转换成 RDF 形式。D2R 通过描述关系数据库规则与 RDF 图之间的关系，可以将关系数据库的数据映射成 RDF 格式。最终导出了共 6 所学校，总计约 1800 个课程的相关数据，数据预处理的流程如图 3 所示。



图 3 数据预处理流程

(2) 知识存储

为更好的存储上述知识节点，我们需要一个存储的图数据库，本知识图谱选择用 Neo4j 图数据库实现知识的存储。Neo4j 是一个高性能的 NOSQL 图形数据库，有成熟的图引擎支撑，自带易于学习的查询语言 (Cypher)，可以满足任何数据模式的需求。相比关系型数据库，Neo4j 对于多级联关系的查询速度要快上许多。

最终构建好的知识图谱包含 6 所高校节点、384 个课程节点和 585 条关系。

算法 1: 关联路径查询算法

```

输入: 节点 A,B
输出: A,B 之间的关联路径
1: function ASSOCIATIONPATHSEARCH(A,B):
2:   A 进入队列 Q
3:   while Q 非空 do
4:     Q 中取出节点 S
5:     for S 的邻接节点集 do
6:       记录路径
7:       if S 的邻接节点 N == 节点 B then
8:         标记为关联路径
9:         goto final
10:      else
11:        节点 N 进入 Q
12:      end if
13:    end for
14:  end while
15:  final
16:  return 关联路径
17: end function
    
```

(3) 课程关联路径构建

有效的知识图谱需要能直观展示出课程之间的关联关系。本图谱中存在的关系主要是课程与它的先修课

程之间的关系，这代表计算机科学专业的课程存在一定的前后关系。课程关联关系挖掘的算法如算法 1 所示。

通过课程关联路径挖掘算法，可以得出数据结构课程与大数据技术基础课程之间存在的课程节点，如图 4 所示，从而能直观的理解课程之间的脉络关系。

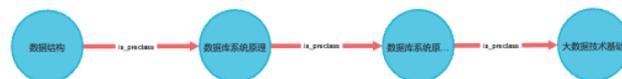


图 4 课程关联路径查询效果图

3 基于可视化图谱的 CS 课程体系分析

课程总体设置、课程之间的关联关系、以及各门课程的培养要求是课程体系的重要构成因素。本文构建的知识图谱也实现了上述因数的查询和对比功能。在图谱基础上，我们可以系统直观的了解每个高校的计算机专业的课程体系，同时做一些对比分析。

不同高校的计算机专业课程体系由于受到多种因素的影响，在长期的发展中产生了一些差异。利用学科知识图谱可以对不同高校间的课程体系进行对比，从而得到差异产生的原因以及可能的影响。如图 5、6 所示，以卡内基梅隆高校的课程体系和清华高校的课程体系为例，图 5 中深蓝色的节点为基础课程，紫色节点为其他课程。可以发现在卡内基梅隆大学的计算机专业课程体系中，从图 5 所示的图谱中可以发现计算机类课程的特点有：

- ① 必修课程的不多，不超过十门；
- ② 课程多为讲解基础知识和介绍理论，基本都是其他课程的前置课程，各课程之间的联系十分紧密；
- ③ 理论与实践相结合，几乎所有课程中都有各种形式的实验。

从图 6 可以发现，清华大学的计算机专业课程体系中的核心基础课程数量较卡耐基梅隆更多，同时专业课程数量也较多，这一方面说明清华大学更加重视学生的基础培养，同时也根据社会需求细分了专业方向，大大增加了“学”的力度。这也符合清华大学提出的“以实践为牵引，深化课程体系改革，实现培养方式转型，促进创新能力培养”的课程体系设计思路。

从如图 7 所示的北京邮电大学课程体系中可以看出，除了涵盖软硬件的计算机专业核心课程如计算机网络、操作系统、计算机组织原理、数据结构与算法等课程外，还包含了以通信和现代交换原理为核心的网络通信类课程，以培养具有网络通信背景的专业人才，这是区别于其他高校的特点。此外，专业课程主要以就业为导向，通过安排创新创业培训课程和专业



改造, 打造工程学科专业的升级版。我们上述分析的北京邮电大学课程设置中体现了北邮的通信特色和优势。行业特色型大学, 为我国工程建设的各方面提供人才的培养和储备, 为此行业特色型大学在课程设置中可以根据行业特色, 面向社会需求, 设置一些本学校的特色课程。

## 5 结束语

本文以 6 所国内外高校计算机专业课程的课程大纲和培养方案为实验数据, 按照前述定义和规则, 设计了课程体系知识图谱, 可视化展示了不同学校的计算机专业课程设置、课程关联路径查询、不同高校的课程设置对比等功能。在基于学科知识图谱的课程体系中, 各不同学校的课程设置, 课程之间的关联关系及其差异分析让我们对国内外目前的计算机学科专业课程体系设置有个全貌的理解, 从而给各高校如何更加科学有效的课程设置提供借鉴意义。

本文的工作主要是针对计算机专业课程之间的关系进行研究, 存在一定的局限性, 如学科知识图谱的数据量较小, 信息不够全面, 难以发现更多的计算机专业课程体系差异, 未来可以增添更多高校、乃至更多学科的数据, 从而实现更加全面的专业课程体系分析和挖掘。

## 参考文献

- [1] 阮彤, 王梦婕, 王昊奋, 等. 垂直知识图谱的构建与应用研究[J]. 知识管理论坛, 2016, 1(3): 226-234.
- [2] Richens R H. Preprogramming for Mechanical Translation[J]. Mechanical Translation, 1956, 3(1): 20-25.
- [3] Berners-Lee T, Hendler J, Lassila O. The Semantic Web[J]. Scientific American, 2001, 284(5): 34-43.
- [4] Bizer C, Lehmann J, Kobilarov G, et al. DBpedia - A Crystallization Point for the Web of Data[J]. Journal of Web Semantics, 2009, 7(3): 154-165.
- [5] Biega J, Kuzey E, Suchanek F M. Inside YAGO2s: A Transparent Information Extraction Architecture[C]. Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion. New York: ACM Press, 2013: 325-328.
- [6] Bollacker K, Evans C, Paritosh P, et al. Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge[C]. Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. 2008: 1247-1250.
- [7] 奥德玛, 杨云飞, 穗志方, 等. 中文医学知识图谱 CMeKG 构建初探[J]. 中文信息学报, 2019, 33(10): 1-9.
- [8] Paolo A, Luigi B, Michela I, Emanuel S, et al. Weaving Enterprise Knowledge Graphs: The Case of Company Ownership Graphs[C]. Proceedings of the 23rd International Conference on Extending Database Technology, Copenhagen, Denmark. 2020: 555-566.
- [9] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.
- [10] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.
- [11] 王鑫, 邹磊, 王朝坤, 等. 知识图谱数据管理研究综述[J]. 软件学报, 2019, 30(7): 2139-2174.
- [12] 鲁佳文, 严丽. 对象关系数据库到 RDF(S) 的映射方法[J]. 计算机科学, 2021, 48(10): 145-151.
- [13] 于慧琳, 陈炜, 王琪, 等. 使用子图推理实现知识图谱关系预测[J]. 计算机科学与探索, 2021: 1-9.
- [14] Ruan T, Xue L J, Wang H F, et al. Building and Exploring an Enterprise Knowledge Graph for Investment Analysis[C]. Proceedings of the 15th International Semantic Web Conference. Cham: Springer International Publishing, 2016: 418-436.
- [15] 张勇, 杨进才. 基于学科知识图谱的高校教学模式研究. 计算机教育, 2021(6): 141-144.