|  |  |
| --- | --- |
| ICS  | 97.220 |
| CCS  |

|  |
| --- |
| D:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T.pngD:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T后面的反斜杠.png CSSS |

Y55 |

中国体育科学学会团体标准

T/CSSS XXXX—XXXX

运动生理参数知识图谱构建指南

Technical guidelines for knowledge graph construction of physiological parameters for sports

XXXX - XX - XX发布

XXXX - XX - XX实施

中国体育科学学会  发布

目次

[前言 II](#_Toc164440213)

[1 范围 1](#_Toc164440214)

[2 规范性引用文件 1](#_Toc164440215)

[3 术语和定义 1](#_Toc164440216)

[4 知识图谱构建 1](#_Toc164440217)

[4.1 构建步骤 1](#_Toc164440218)

[4.2 知识抽取加工 2](#_Toc164440219)

[4.3 知识存储 3](#_Toc164440220)

[4.4 图分析与图挖掘 3](#_Toc164440221)

[参 考 文 献 4](#_Toc164440222)

前言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由河北体育学院提出。

本文件由中国体育科学学会归口。

本文件起草单位：河北体育学院、清华大学、中国科学院大学、中国人民大学、福州大学、清华海峡研究院（厦门）、南京润楠医疗电子研究院有限公司、数聚工研（北京）科技有限公司。

本文件主要起草人：史东林、杨吉江、赵越、雷毅、赵荣建、刘凯、裴梓彤、陈永锐、王忠瑞、李玉榕。

运动生理参数知识图谱构建指南

* 1. 范围

本文件给出了构建运动生理参数知识图谱的基本概念，提供了构建知识图谱的步骤、知识抽取加工、知识存储、图分析与图挖掘等方面的建议。

本文件适用于运动生理参数知识图谱的构建。

* 1. 规范性引用文件

本文件没有规范性引用文件。

* 1. 术语和定义

下列术语和定义适用于本文件。

 知识图谱 knowledge graph

一种以结构化的形式描述客观世界中概念、实体及其关系的方式。

注：知识图谱将互联网的海量信息表达成更接近人类认知世界的形式，提供了一种更好地组织、管理和理解互联网海量信息的能力。

实体 entity

具有独立逻辑含义的一类概念、人、事物。

关系 relationship

实体与实体之间特定时间、特定行为下产生的联系。

属性 attribute

实体或关系所具有的特性。

知识推理 knowledge reasoning

在计算机中进行机器思维、求解问题的过程。

心电图 Electrocardiogram

心脏活动所产生的电活动变化的记录。计量单位：mV。

心率 heart rate

心脏每分钟跳动的次数。计量单位：次/分。

呼吸率 respiratory rate

每分钟呼吸的次数。计量单位：次/分。

血氧饱和度 oxygen saturation

动脉血氧与血红蛋白结合的程度，是单位血红蛋白含氧百分数。计量单位：百分比，符号：%。

* 1. 知识图谱构建
		1. 构建步骤
			1. 采用自顶向下的方式构建运动生理参数知识图谱，具体步骤如图1所示。



**图1 知识图谱构建步骤**

* + - 1. 知识抽取加工的目标是将结构化、半结构化和非结构化数据转化形成满足知识图谱存储要求的标准形式，主要包括数据甄别、数据清洗、数据转换和数据融合等步骤。
			2. 知识存储的目标是有效管理知识图谱依赖的图数据、属性数据。
			3. 图分析与图挖掘是在知识推理技术之上构建分析模型和算法挖掘模型，对知识图谱进行深入挖掘，生成分析挖掘结果，支撑表1中所列运动健康应用场景的构建。

**表1 运动健康应用场景**

|  |  |  |  |
| --- | --- | --- | --- |
| 应用场景 | 输入项 | 计算逻辑 | 输出结果 |
| 运动适应能力评估 | 生理参数，包括但不限于心率、心电图、呼吸频率、血压、血氧饱和度等 | 根据运动过程中的生理参数变化趋势，通过图谱分析和挖掘，计算适应能力指数或适应能力等级 | 输出运动适应能力评估，评估用户对运动的适应程度 |
| 运动疲劳预警 | 根据运动过程中的生理参数变化趋势，通过图谱分析和挖掘，计算疲劳指数或疲劳等级 | 输出运动疲劳预警，提示用户可能已接近或达到疲劳状态 |
| 运动风险评估 | 根据运动过程中的生理参数变化趋势，通过图谱分析和挖掘，计算风险指数或风险等级 | 输出运动风险评估，提示用户可能存在的健康风险 |

* + 1. 知识抽取加工
			1. 知识抽取

知识抽取主要包括以下4个步骤。

——数据甄别：明确建立知识图谱的数据来源。包含但不限于心率、心电图、呼吸频率、血压、血氧饱和度等运动生理参数的数据源。

——数据清洗：对数据中的干扰项，特别是来自互联网的错误信息、虚假信息进行清洗，对表示不规范的数据进行统一与规范。

——数据转换：将不同形式、不同格式的数据转换成统一的表达形式。

——数据融合：针对不同来源的数据在字段、元组等层次的融合，不涉及实体、属性、关系的融合。

* + - 1. 知识加工

知识加工以知识抽取阶段形成的高质量基础数据作为输入，输出高质量的知识图谱构建所依赖的实体、关系数据并形成运动生理参数知识库。

知识加工主要包括如下步骤：

——本体设计；

——实例获取；

——知识融合。

本体设计的内容主要包括以下4个方面。

——概念层定义：明确定义知识图谱构建需要的实体类型与关系类型，涉及到运动健康生理参数领域概念、概念属性以及概念之间的关系。

——属性定义：明确知识图谱构建的各类实体需要的属性名称。

——约束/规则定义：明确需要约定的限定条件与规则。

——模式精简：对同样范围的实体、关系，选择尽量精简的关联与表达方式。

实例获取的内容主要包括以下4个方面。

——实体识别：在完成知识抽取后的数据中识别上述已经明确的实体类型对应数据；

——关系识别：在完成知识抽取后的数据中识别上述已经明确的关系类型对应数据；

——实体链指：将完成文本解析后的事件关联方链接至对应实体；

——文本解析：利用自然语言处理技术对运动健康生理参数相关文本（心率、心电图、呼吸频率、血氧饱和度）进行分析，实现事件关联方的提取、实体/关系属性的槽填充，以及评估的量化结果。

知识融合的主要内容包括以下3个方面。

——共指消解：将具有相同标识的两个实体合并为同一实体。

——属性融合：将同一标识实体下，相同名称的属性所对应的属性值进行融合。

——规范化一：将不同实体属性值与不同关系属性值的表达方式进行统一。

* + 1. 知识存储
			1. 概述

知识存储阶段将知识图谱的概念层和数据层以计算机可识别的数据格式进行物理保存，针对知识图谱的知识表示形式设计匹配的底层存储方式，完成各类知识的存储，以支持对大规模图数据的有效管理和计算。

* + - 1. 图数据存储

图数据存储内容宜包括实体数据和关系数据，分别以实体表和关系表的形式体现。图数据存储方式宜根据数据规模决定采用分布式存储或单机存储。

* + - 1. 属性数据存储

属性数据存储内容宜包含实体属性和关系属性。属性数据存储方式宜根据数据规模决定采用分布式存储或单机存储，主要包括单个和批量属性数据的增加、更新、删除操作。

* + 1. 图分析与图挖掘
			1. 概述

图分析与图挖掘基于知识推理。

知识推理的方法主要包括下列内容。

——基于概率的知识推理：根据知识存在准确性等不确定因素，通过概率理论进行推理判断。

——基于机器学习的知识推理：处理具有集约特征，即知识的获取、表示和推理合为一体，均通过机器学习训练实现，从而可以充分利用样本性知识，而样本性知识相对来说是最容易获得的。

* + - 1. 图分析

图分析以知识图谱为基础，构造具有相对固定模式化的分析结论。图分析主要包括统计汇总和时序分析。

* + - 1. 图算法

图算法以知识图谱为基础，构造具有相对固定模式化的分析结论。图算法主要包括实时图算法与离线图算法。

参 考 文 献

1. GB/T 42777—2023 基于文本数据的金融风险防控 知识图谱构建技术框架指南
2. YD/T 4044—2022 基于人工智能的知识图谱构建技术要求

