



中华人民共和国国家标准

GB/T 45949—2025

中文新闻语义结构化标注

Specification for semantic structuring annotation of Chinese news

2025-06-30 发布

2025-06-30 实施

国家市场监督管理总局
国家标准管理委员会 发布

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 新闻语义结构化标注概述	2
4.1 标注流程	2
4.2 标注范式	3
5 预处理	4
5.1 标注对象标识符设置	4
5.2 新闻稿件元数据采集	4
6 关键词标注	5
6.1 标注内容	5
6.2 标注模板	5
6.3 标注方法	5
7 语义模板标注	6
7.1 标注内容	6
7.2 标注模板	14
7.3 标注方法	15
8 语义单元标注	16
8.1 标注内容	16
8.2 标注模板	17
8.3 标注方法	18
附录 A (资料性) 应用场景举例	23
A.1 新闻内容结构化管理	23
A.2 生成稿件内容知识库	23
A.3 建立高质量知识图谱	23
A.4 低耗高效训练人工神经网络模型	23
附录 B (资料性) 新闻语义结构化标注完整示例	24
B.1 原文	24
B.2 预处理	24
B.3 关键词标注	26
B.4 语义模板标注	27
B.5 语义单元标注	29
参考文献	32

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。

本文件由全国中文新闻信息标准化技术委员会(SAC/TC 352)提出并归口。

本文件起草单位：文灵科技(北京)有限公司、北京语言大学、东南大学、新华通讯社、中国互联网新闻中心、中国信息通信研究院、浙江日报报业集团、保定市数据局(国家数据标注基地)。

本文件主要起草人：王楠、饶高崎、杨鹏、丁德胜、王熠、付蓉、王贵荣、王建平、黄菁、李荪、贾成喜、张闯、李冰、杨冬梅、周天外、宋永生。

引　　言

随着人工智能技术在新闻领域的广泛应用,建立规范化的中文新闻数据加工标准已成为行业发展的
重要基础。本文件的制定旨在构建统一的中文新闻语义标注体系,通过标准化的数据加工流程,实现
以下两大目标:一是为新闻知识库与知识图谱的体系化建设提供支撑;二是提升人工神经网络模型训练
的数据质量与效率,推动新闻行业的智能化转型。

中文新闻语义结构化标注

1 范围

本文件规定了中文新闻语义结构化标注的预处理、关键词标注、语义模板标注和语义单元标注的方法。

本文件适用于新闻内容提供商、媒体应用与研究机构对新闻稿件的人工标注，也能用于机器自动化标注。

本文件不适用于跨稿件的语义结构化标注。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 20092.2—2022 中文新闻信息内容 第2部分：新闻元数据

GB/T 20093—2022 中文新闻信息分类与代码

3 术语和定义

下列术语和定义适用于本文件。

3.1

关键词 keyword

代表新闻稿件中特定语义的词汇。

注：“特定语义”指标注者认为重要的新闻稿件内容，涵盖稿件主题、重要人物、重要事件及重要常识等。

3.2

语义单元 semantic unit

新闻稿件中表达完整语义的基本单元。

注：语义单元主要分为实体、事物、事件等类型，是组成包括知识图谱在内的复杂语义结构的基础。

3.3

语义单元参数 parameter of semantic unit

描述或限定语义单元的特征量。

注：语义单元参数包括属性、性质、方面、能力（用途）、类别、限定和状态等类型。

3.4

语义单元状态 state of semantic unit

语义单元参数在特定时空中的取值集合。

注：语义单元状态用于描述事物发展变化的动态性，标注时，作为语义单元的一个参数来标注。

3.5

语义模板 semantic template

表达语义的固定结构。

注：由核心要素、限定要素及其相互关系组成。