

中国计算机用户协会团体标准

《版本典藏资源智慧展陈 语音交互技术要求》

（征集意见稿）编制说明

一、标准编制的背景

近年来，互联网技术和信息技术的发展一直保持着强劲的势头，人们在社交、交通出行、购物消费、饮食娱乐等日常活动中也总能见到互联网技术的应用。在文化遗产馆方管理、保护、传承方面，博物馆是以提升自身服务能力，发挥博物馆社会教育、文物收藏、保管、修复功能，培养群众文化自信为目标的，正在通过信息化管理平台、人工智能、VR 技术等实现人和人、人和建筑、人和展品、人和业务、业务和业务、展品和藏品、展品和展品之间形成更好的智能交互和情景体验，让整个馆成为一座智能体，实现了高效运营，减少了资源耗费，这是文博产业得以持续发展的强大助力。尤其是在文物展览引导方面，展馆设置了智慧导览系统，可以在手机上即可查看地图，获得文物信息和语音讲解，也可以现场进行各种交互体验，近距离的感受文化遗产自身的魅力。让参观者对文物有更深了解，触动其内心，使其受到相应的教育。

国家版本馆的核心目标是建设成为国家版本典藏、展示、研究、交流中心，作为馆藏内容的版本典藏资源，是在中国版本图书馆馆藏上千万册现代版本的基础上，整合全国乃至海外版本资源而成，其中数字化资源包括国家版本典藏实物的数字版本资源、原生数字版本资源，以及互联网信息资源等。版本典藏数字资源具有多源异构的数据特点，如何利用数字化技术最大程度的发挥版本典藏资源的公共服务作用，是文化科技所面临的挑战。

开发沉浸式交互虚拟现实智慧展陈内容组件对于虚拟现实智慧展陈场景创作、版本典藏数字资源开发利用与传承保护均具有重要的意义。随着虚拟仿真沉浸式交互技术的成熟，沉浸式可视化的智慧展陈环境亟需融合视觉、听觉和触觉等多个感知通道，提供全新的多模态人机智能交互叙事方式。语音识别、手势隐喻动作设计、姿态估计与追踪等是沉浸式交互虚拟现实智慧展陈内容组件构建的关键技术。

文博展陈中的智能交互可以分为文物管理过程中管理员和馆藏交互、博物馆物理构件间交互、参观过程参观者和馆藏交互、社会公众和博物馆交互几个方面。人机智能交互当中包含语音交互、手势交互、位置交互、眼动追踪交互、色彩交互、触控交互、文字命令交互、嗅觉交互、味觉交互以及多模态交互等众多技术。本标准主要是针对参观过程中参观者和馆藏交互中的语音交互，通过语音交互的方式，将机器学习这一计算机技术与馆藏文物这一展馆展项融合到一起。作为创新展项，基于语音识别的展示系统，不光具有展馆传统展品的展示功能，而且可以让参观者有更高的自主性，在另一方面还可以普及人工智能的概念。传统的展馆建筑展览系统已经无法满足人们与日俱增的科技需求，在人工智能技术的发展历程中，语音识别的出现极大程度的满足了人们在日常生活中方便、快捷获取信息的需求。展馆在展示陈列方面运用了人工智能和虚拟现实技术，通过友好的人机交互提高了参观者的用户体验，既能更好完成国家版本馆展陈内容的介绍工作，又能丰富国家版本馆社教功能更好地服务大众。通过制订本标准，可以进一步规范馆方在设计数字文物展览系统过程中语音交互技术的统一性，对于实现整个展馆展示内容的数字化、场景化、体验化等方面具有十分重要的意义。通过对语音交互技术的定义，进一步规范了馆藏数字文物在内容设计中交互接口的标准设计，为后续进一步扩大利用数字资源内容奠定了基础，同时提升了资源的可复用性，大大提升了数字资源的利用价值和提升了馆藏文物的传播空间。

本标准适用于有版本典藏数字资源智能展陈语音交互技术需求的国家版本馆及其分馆，以及其它博物馆、纪念馆以及各种智慧化展陈设计制作内容的企事业单位。

二、任务来源

根据中国计算机用户协会下达的 2023 年下半年第一批团体标准制修订计划，中国科学院自动化研究所作为主要牵头单位筹建了标准起草组，承担《版本典藏资源智慧展陈 语音交互技术要求》标准的研制任务。该标准的立项计划号为 T/CCUA LX008 2023，技术归口单位为中国计算机用户协会。

三、编制过程

在下达计划号前，2023 年 8~9 月期间多次召开起草组大纲讨论会。

2023年11月1日召开标准立项评审会通过项目立项。

2023年3月完成标准工作组稿。

2024年3月，标准起草组完成了标准草案的初稿。

2023年4月底开始在起草组评审委员会内进行评审。

2023年5月初，将标准征集意见稿提交标委会进行意见征集。

2023年5月中旬，收到标委会反馈的征集意见稿修改意见。

2023年6月中旬，按照征集意见稿进行修订并提交标委会审核。

共经历了1次组内评审1次标委会专家审核，期间收到了12条意见。

四、编制原则

标准的用语、格式按照 GB/T1.1-2020 给出的规则起草。

标准内容的编制坚持以下原则：

1. 简洁性与清晰性

语音指令应简单易懂，避免使用复杂或专业术语，使用日常语言，确保所有用户都能理解信息和指令。

2. 反馈及时性

用户的每条指令或询问应获得快速且明确的反馈，设计即时的声音或视觉反馈机制，确认用户指令已被接收和处理。

3. 容错性与指导

系统应能妥善处理用户的错误输入，并提供修正的途径，当语音识别错误时，系统应提供明确的错误信息和帮助用户正确地重新输入。

4. 一致性

语音交互的行为、提示语和反馈应保持一致，使用统一的语言风格和指令集，让用户能够快速学习并适应系统。

5. 隐私保护

确保用户在使用语音交互时的隐私安全，为敏感对话提供隐私选项，例如处理支付信息时转为文字输入或提供安全环境。

6. 可扩展性

设计时应考虑未来的功能扩展和技术升级，构建模块化的设计，方便未来添加新功能或整合新技术。

五、标准主要内容

本文件确立了版本典藏资源智慧展陈语音交互技术总体框架,规定了语音交互界面设计、版本典藏资源库、语音处理、服务接口和安全技术要求。

本文件适用于版本典藏资源智慧展陈语音交互系统或产品的设计、开发、应用和维护。

1. 总体框架设计要求。
2. 语音交互界面设计,包括显示界面、采集界面、播报界面和输入输出界面。
3. 版本典藏资源库中语音和文本语料资源的要求。
4. 语音处理中指令识别、语音合成、语音理解与答复、多模态支持。
5. 同其他应用系统或模块之间对接的服务接口要求。
6. 对系统运行环境、交互逻辑、接口之间的安全要求。

六、有关技术的说明

有关本标准起草过程中的一些技术问题说明如下:

1. 智能语音技术作为人工智能时代的核心技术,是通过声音信号的前端处理、语音识别、自然语言处理、语音合成等技术,实现人机语音交互,从而让人与机器交流更为方便快捷。在标准制定方面需要结合用户行为分析、位置追踪和人工智能,提供个性化的语音交互服务。

2. 博物馆内部环境可能存在噪声或回声问题,这要求语音交互系统具有高度的声学适应性。可以使用高级的声音处理技术,如回声消除和噪声抑制,保证语音输入和输出的清晰度。

参考的主要标准:

1. GB/T36339 智能客服语义库技术要求
2. GB/T34083 中文语音识别互联网服务接口规范
3. GB/T34145 中文语音合成互联网服务接口规范
4. GB/T 36464.1-2020 信息技术 智能语音交互系统 第1部分:通用规范
5. GB/T 41813.1-2022 信息技术 智能语音交互测试方法 第1部分:语音识别

七、关于标准的性质

鉴于本标准作为团体标准发布，属于推荐性标准。由本团体成员约定采用或者按照本团体的规定供社会自愿采用。

八、有关专利的说明

本标准不涉及专利问题。

标准起草组

2024 年 6 月 15 日