



中华人民共和国国家标准

GB/T 34083—2017

中文语音识别互联网服务接口规范

Specification of programming interface for Chinese speech recognition
internet service

2017-07-31 发布

2018-02-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

目 次

前言	I
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 语音识别服务系统能力范围	2
4.1 概述	2
4.2 基本功能	3
4.3 扩展功能	3
5 输入数据	4
5.1 语音识别音频数据格式要求	4
5.2 语音识别语法	4
5.3 热词集	5
5.4 语音识别参数	5
6 输出数据	11
6.1 语音识别结果格式	11
6.2 语音识别结果描述参数	11
7 服务接口	17
7.1 概述	17
7.2 语音识别服务接口	17
7.3 数据操作接口	18
8 接口返回值要求	19
附录 A (资料性附录) 语音识别结果消息示例	20
A.1 一般识别结果消息示例	20
A.2 语言、说话者信息消息示例	20
A.3 不同结果类型消息示例	21
A.4 进阶模式消息示例	22
A.5 识别结果多候选消息示例	28
A.6 自定义热词消息示例	29
A.7 命令字识别消息示例	30
图 1 语音识别一般流程示意图	3
表 1 音频数据格式	4
表 2 识别语法的类型	5
表 3 语音识别参数	5
表 4 result-info 取值允许的结果描述参数集合	10
表 5 语音识别结果描述参数	11

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

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这些专利的责任。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位:科大讯飞股份有限公司、中国电信集团公司、中国电子技术标准化研究院、苏州思必驰信息科技有限公司、北京慧听科技有限公司、阿里云计算有限公司、哈尔滨工业大学、中国科学院声学研究所。

本标准主要起草人:李洪亮、严峻、于继栋、朱恩德、程彩锋、张陈、胡修文、杨震、赵菁华、周伟达。

中文语音识别互联网服务接口规范

1 范围

本标准规定了中文语音识别服务系统在互联网环境下提供服务的能力范围、输入数据、输出数据、服务接口、接口返回值要求等。

本标准适用于基于互联网的中文语音识别服务系统接口的设计、开发和应用。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是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 13000 信息技术 通用多八位编码字符集(UCS)(GB 13000—2010,ISO/IEC 10646:2003,IDT)

GB 18030 信息技术 中文编码字符集

GB/T 21023—2007 中文语音识别系统通用技术规范

汉语拼音方案(中华人民共和国第一届全国人民代表大会第五次会议决议通过)

RFC 2045 多用途互联网邮件扩展 第1部分:互联网消息体格式(Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies)

RFC 2586 多用途互联网邮件扩展内容类型(The Audio/L16 MIME content type)

RFC 5234 语法规则的扩展巴科斯范式(Augmented BNF for Syntax Specifications: ABNF)

RFC 5646 语言标识标签(Tags for Identifying Languages)

SRGS 1.0 语音识别语法规则(Speech Recognition Grammar Specification Version 1.0)

国际语音学会手册:国际音标使用指南(Handbook of the International Phonetic Association: A Guide to the Use of International Phonetic Alphabet)

3 术语和定义

下列术语和定义适用于本文件。

3.1

语音识别 speech recognition

将人类的声音信号转化为文字或者指令的过程。

[GB/T 21023—2007,定义 3.1]

3.2

语音识别语法 speech recognition grammar

识别语法 recognition grammar

语法 grammar

语音识别引擎进行语音识别所需的语句形式和语法要素,语法要素包括词汇、词汇模式、词汇语言等。

3.3

命令字识别 command word recognition

一种基于语音识别语法的语音识别方式,是在语音识别语法规则限定的范围内,对于给定的语音输