

SSML使用说明

1简介

SSML (Speech Synthesis Markup Language) 是一种基于XML的语音合成标记语言。用户可以利用SSML精确控制数字符号的读法、词汇的发音、韵律的停顿等特征。本系统基于W3CSSMLv1.1实现, 支持SSMLv1.1中的一部分常用的标签类型。

2使用说明

XML的version字段必须为"1.0"。
XML的encoding字段表示编码, 目前只支持"utf-8"编码, 没有encoding字段的情况下默认为"utf-8"。
待合成的文本需要放在< speak >标签内。
其中, version字段必须为"1.1"。
无论待合成内容是中文还是英文, xml:lang字段必须为"cn"。
每个语音合成任务中有且仅有一个< speak >标签。

示例:

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <phoneme ph="xiao3"></phoneme>
  <break time="200ms"/>
  <say-as interpret-as="date">2018-01-22</say-as>
  <w></w>
</speak>
```

3标签说明

3.1 概述

本系统支持的SSML标签包括: < speak >、< w >、< say-as >、< phoneme >和< break >。

< speak >标签是SSML的根节点。
待合成的文本需要放在< speak >标签内。
每个语音合成任务中有且仅有一个< speak >标签。

< w >标签是词汇标志。
TTS算法会将< w >标签内部的文本当做一个完整词汇来处理。
使用< w >标签可以提高词汇朗读的流畅度, 避免词汇内部出现异常停顿。

< say-as >标签是读法标志。
待合成文本中常有一些需要根据特殊规则来朗读的内容, 如时间、数字、姓名、地址等等。
用户可以通过< say-as >标签指定本系统提供的特定读法规则。

< phoneme >标签是注音标志。
待合成文本中汉字的读音可以通过< phoneme >标签来指定。
本系统支持用汉语拼音对汉字注音, 不支持对英文单词注音。

< break >标签是停顿标志。
用户可以通过< break >标签在合成结果中插入停顿, 使合成语音的韵律更加自然。

3.2 标签嵌套关系

标签嵌套基于以下规则:

- < speak >是根节点, 其他标签全部在此标签内部;
- < speak >内部可以包含纯文本, 以及被其他标签修饰的文本;
- < w >、< say-as >、< phoneme >和< break >都是可选标签;
- < w >内部可以包含< say-as >和< phoneme >;
- < say-as >和< phoneme >不强制要求在< w >内部, 可以单独使用;
- < say-as >和< phoneme >不可以互相嵌套;
- < say-as >和< phoneme >内部不能包含任何标签;
- < break >不可出现在< w >、< say-as >和< phoneme >内部。

示例:

例1< speak >内部包含纯文本

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">

  </speak>
```

例2< speak>内部包含< w>

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <w></w>
</speak>
```

例3 < speak>内部包含< w>, < w>内部包含< phoneme>和< say-as>

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <w><say-as interpret-as="code">61</say-as><phoneme ph="e2"></phoneme></w>
</speak>
```

例4 < speak>内部包含< say-as>

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <say-as interpret-as="name"></say-as>
</speak>
```

例5 < speak>内部包含< phoneme>

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <phoneme ph="xiao3"></phoneme>
</speak>
```

例6 < speak>内部包含< break>和< phoneme>

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">

  <break time="200ms"/>
  <phoneme ph="xiao3"></phoneme>
</speak>
```

3.3 < speak>标签

3.3.1 说明

< speak>标签是SSML的根节点。
待合成的文本需要放在< speak>标签内。
每个语音合成任务中只能存在一个< speak>标签。

3.3.2 语法

```
<speak version="1.1" xml:lang="cn"></speak>
```

3.3.3 属性

字段名称: version
字段功能: 指定SSML的版本
是否必填: 是
字段说明: 字段的值必须为"1.1"

字段名称: xml:lang
字段功能: 指定合成的语言
是否必填: 是
字段说明: 无论待合成内容是中文还是英文, 字段的值必须为"cn"

3.4 < w>标签

3.4.1 说明

<w>标签是词汇标志。
TTS会将<w>标签内部的文本当做一个完整词汇来处理。
使用<w>标签可以提高词汇朗读的流畅度，避免词汇内部出现异常停顿。

3.4.2 语法

```
<w></w>
```

3.4.3 属性

无

3.4.4 示例

例1 <w>内部包含纯文本

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <w></w>
</speak>
```

例2同时出现多个<w>

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <w></w><w></w>
</speak>
```

例3 <w>内部包含<phoneme>和纯文本

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <w><phoneme ph="chang2"></phoneme></w>
</speak>
```

例4 <w>内部包含<say-as>

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <w><say-as interpret-as="code">61</w>
</speak>
```

例5 <w>内部包含<phoneme>和<say-as>

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">
  <w><say-as interpret-as="code">61</say-as><phoneme ph="e2"></phoneme></w>
</speak>
```

3.5 <say-as>标签

3.5.1 说明

<say-as>标签是读法标志。
待合成文本中常有一些需要根据特殊规则来朗读的内容，如时间、数字、姓名、地址等等。
用户可以通过<say-as>标签指定本系统提供的特定读法规则。

3.5.2 语法

```
<say-as interpret-as=""></say-as>
```

3.5.3 属性

字段名称: interpret-as
字段功能: 指定读法的类型
是否必填: 是

3.5.4 读法类型

数值类

读法类型: 数读
读法代码: num
说明: 按数值方式朗读的整数或小数
示例:
123 (一百二十三)
123.33 (一百二十三点三三)

读法类型: 数读 (两)
读法代码: num_liang
说明: 按数值方式朗读的整数或小数, 其中“2”读作“两”, 用于后接量词的情况
示例:
2 (两)
123 (一百二十三)
123.33 (一百二十三点三三)

编码类

读法类型: 码读
读法代码: code
说明: 按编码方式朗读的数字和英文字母的组合
示例:
123X (一二三X)
ISO9001 (ISO九零零一)

读法类型: 码读 (幺)
读法代码: code_yao
说明: 按编码方式朗读的数字和英文字母的组合, 其中“1”读作“幺”
示例:
123X (幺二三X)
ISO9001 (ISO九零零幺)

时间类

读法类型: 日期
读法代码: date
说明: 支持由“-”或“/”分隔的年月日, 年要求用4位数字表示, 月日要求用2位数字表示
示例:
2001-01-01 (二零零一年一月一日)
2001/01/01 (二零零一年一月一日)

读法类型: 日期 (年月)
读法代码: date_ym
说明: 支持由“-”或“/”分隔的年月, 年要求用4位或2位数字表示, 月要求用2位或1位数字表示
示例:
2001-01 (二零零一年一月)
2001/01 (二零零一年一月)
2001-1 (二零零一年一月)
2001/1 (二零零一年一月)
20-01 (二零零一年一月)
20/01 (二零零一年一月)
20-1 (二零零一年一月)
20/1 (二零零一年一月)

读法类型: 时间
读法代码: time
说明: 支持由“:”分隔的时间, 时分秒都要求用2位数字表示
示例:
21:01:30 (二十一点零一分三十秒)
21:01 (二十一点零一分)

其他类

读法类型: 汉语人名
读法代码: name
说明: 汉语人名或姓氏, 其中姓氏采用作为姓氏时的常见读音, 名字中的多音字采用高频读音
示例:
单田芳 (shan4tian2fang1)
查 (zha1)

读法类型：电话号码
读法代码：tel
说明：支持常见电话号码格式
示例：
13812345678（幺三八，幺二三四，五六七八）
010-6606100（零幺零，六六零，六幺零零）
010-66061001（零幺零，六六零六，幺零零幺）
(010) 6606100（零幺零，六六零，六幺零零）
(010) 66061001（零幺零，六六零六，幺零零幺）
95588（九五五八八）
4008801234（四零零，八八零，幺二三四）

读法类型：地址
读法代码：address
说明：支持常见地址格式
示例：
通州区孙各庄303#（北京市通州区孙各庄三百零三号）
通州区孙各庄3-2-1（北京市通州区孙各庄三杠二杠幺）

读法类型：车牌
读法代码：plate
说明：支持国内车牌，停顿规则为2-2-3或2-3-3
示例：
京A12345（京A，幺二，三四五）
京A12345E（京A，幺二三，四五E）

读法类型：航班号
读法代码：flight_number
说明：支持国内外航班号，停顿规则为2-3或2-4
示例：
CA901（CA，九零幺）
MU5173（MU，五幺七三）

3.5.5 示例略

3.6 <phoneme>标签

3.6.1 说明

<phoneme>标签是注音标志。
待合成文本中汉字的读音可以通过<phoneme>标签来指定。
本系统支持用汉语拼音对汉字注音，不支持对英文单词注音。

3.6.2 语法

<phoneme ph="dai4 he2 cheng2 wen2 ben3"></phoneme>

3.6.3 属性

字段名称：ph
字段功能：指定待合成文本的拼音
是否必填：是
特殊说明：
拼音由发音和声调组成，拼音之间由空格隔开，如pin1yin1；
声调包括1~4声，5代表轻声；
女的拼音是nv3，铝的拼音是lv3，虐的拼音是nve4；
拼音的数量和待合成文本中的字数必须相同；
<phoneme>标签内部最多包含20个汉字。

3.6.4 示例

略

3.7 <break>标签

3.7.1 说明

<break>标签是停顿标志。
用户可以通过<break>标签在合成结果中插入停顿，使合成语音的韵律更加自然。

3.7.2 语法

```
<break time="ms"/>
```

3.7.3 属性

字段名称: time

字段功能: 指定停顿的时长, 取值范围为25ms~5000ms

是否必填: 是

特殊说明:

此功能主要面向语音流中的短时间的停顿, 建议停顿时长保持在300ms以内。

3.7.4 示例

```
<?xml version="1.0" encoding="utf-8"?>
<speak version="1.1" xml:lang="cn">

    <break time="200ms"/>

</speak>
```