

JEITA 日本語音声認識用語彙データベース 2013

1. 概要

最近の音声認識アプリケーションでは、地名、人名、楽曲名等を扱うものが出てきている。しかし、その開発に使用できる共通的な語彙データベースは存在しなかった。本データベースは、そのようなアプリケーションの開発に利用できる大規模な読み付き語彙データベースであり、音声認識開発を促進するためのものである。

2. 内容

JEITA で独自に集めたもので、約 4000 タイトルの楽曲データからなる。各タイトルには、楽曲名、アーティスト名、アルバム名の各項目について表記と読みが付与されている。楽曲データは CSV フォーマットにて提供する。

3. 読み表記

3.1 読み表記記号

「JEITA 日本語音声認識用読み記号(JEITA-IT-4003)」の、「表1 音声認識用読み記号」の範囲のかな、および、次節で述べる若干の記号を使用して読みを付与している。

3.2 読み付与

以下のガイドラインに従って読み付与している。

(1) 副題等の記述

副題等の部分を、半角両鉤括弧記号[]で囲む。[]のネストもありうる。

(2) 並列表記

年号表示等の場合で、複数の読み方が考えられる場合は、半角 pipe 記号(|)を用いて、読みを並列表記する。並列表記は、半角両小括弧記号()で囲む。

例 ' 0 5 (ぜろご | ぜろごねん)
 ' 9 8 (きゅーじゅーはち | きゅーじゅーはちねん)
 1 9 9 8 (せんきゅーひゃくきゅーじゅーはち | せんきゅーひゃくきゅー
 じゅーはちねん)
 7 4 (ななじゅーよん | ななじゅーよねん)

(3)外国語曲の外国語表記と日本語表記の並列表記

以下のように ,外国語曲に日本語タイトルが付与されている場合は ,半角二重 pipe 記号 (||) を用いて ,読みを並列表記する . 並列表記は ,半角両小括弧記号 () で囲む .

例 涙の乗車券 : Ticket To Ride ((なみだの じょーしゃけん) || (ちけつと とう らいど))

(4)長音の表記

日本語 , 外来語を問わず , すべて「おー」「えー」などの表記にする (但し , 形態素内に限る) .

(5)促音

促音「っ」は , 破裂音 (破擦音) ・ 摩擦音 , その有声音の直前に限り , それ以外では用いない .

(6)デリミタ

日本語表記は , 文節境界に半角スペースを入れる .

外国語表記は , 単語境界に半角スペースを入れる .

アーティスト名の姓と名の間にも半角スペースを入れる .

例 敏いとうとハッピーアンドブルー とし いとー と はっぴー あんど ぶ
るー

(7)英語に対する読みの付与

日本語の読みが原則 . 日本語として定着している表現があれば , それを用いる .

例

単語	読み	単語	読み
Two/2	つー	Twenty	とうえんていー
To	とー	Too	とー
what	ほあっと	Fat	ふあっと
when	ほえん	Why	ほあい
where	ほえあ	Anywhere	えにーうえあ
Who	ふー	Whom	ふーむ
how	はう		
white	ほわいと	Winter	ういんたー

備考

Mvt	Movement	むーぶめんと
pt.	Part	ぱーと
Livin'		りびん

(8)年号等の読み

年号の読みは、日本楽曲、外国楽曲にかかわらず、タイトル表記に合わせることにする。すなわち、タイトルが日本語表記の場合は日本語の読みを、外国語表記の場合は外国語の読みをそれぞれ付与する。ただし、日本語のタイトル中の数字においても、英単語との組み合わせなどで明らかに英語読みが浸透している場合は、この範囲外とする。

例 プレイバック Part2 ぶれいばっく ぱーとつー

4. 利用条件

本データは、各セットにおける語彙を全て網羅するものではなく、また、各語彙の表記および付与された読みについては、その正確性を保証するものではない。本データを利用する際には以下の利用条件を順守することとする。

1. 本データのすべてまたは一部について、有償・無償にかかわらず第三者に公開・配布・譲渡または貸与することはできません。
2. 本データは、商用を含め、上記 1.以外では自由に使えます。
3. 本データを使用して得られた知見に関して成果を発表する際には、「**JEITA 日本語音声認識用語彙データベース 2013 を使用**」と明記してください。
4. 本データについて、修正や拡充のご希望がありましたら、個別にお問い合わせください。
5. 本データの内容について、JEITA は一切保障いたしません。ここでいう保障にはデータベースの品質の他に、著作権など第三者の有する権利利益を侵害していないことの保障を含みます。
6. 内容の瑕疵により直接的・間接的に発生した損害について、使用者は JEITA に対して一切責任を求めないこととします。
7. 本データの利用を終了する場合には、すべてのデータを破棄してください。