

# Japanese Linguistic Inquiry & Word Count (JIWC) Manual

Social Computing Lab.  
Nara Institute of Science and Technology

July 31st 2020

## 1 What is JIWC?

Japanese Linguistic Inquiry and Word Count (JIWC) dictionary is used for scoring words with the emotion related to it and has been constructed using EPISODE BANK (EP) data set. The main purpose of construction of the JIWC dictionary is the usage in the analysis of the emotion related to texts. An example of JIWC use case is the analysis of speech belonging to people suffering from dementia [2].

## 2 Construction of JIWC dictionary

### 2.1 EPISODE BANK

EPISODE BANK (referred to as EB hereafter) is a service used to collect and store episodes relating to a fixed set of emotions (joy, sadness, etc) from various people (referred to as episode data set hereafter) through crowdsourcing. The process of collection of the episode data set is explained below.

### 2.1.1 Episode Data Set

EB consists of episodes collected through Yahoo! crowdsourcing\*<sup>1</sup> and includes actual expressions related to specific emotions obtained from a large number of people. The episodes were collected in 7 categories of "Sadness", "Anxiety", "Anger", "Disgust", "Trust", "Surprise", and "Joy" based on Plutchik's circle of emotions [1]. Each participant in the crowdsourcing was asked to contribute a total of 7 episodes, each corresponding to an emotion described in the previous sentence. Data collection is carried out four times a year to give about 4,000 episodes per emotion and about 28,000 episodes in total every year. Data collection in a given year is used to create the JIWC dictionary for that year. Data is collected four times a year to consider the seasonal trends in word usage. Please check the website (<http://github.com/sociocom/Episode-Bank>) to learn more about the EPISODE BANK.

### 2.1.2 Sample Episodes

Data set collected in 2.1.1 contains episodes similar to the examples shown in Table 1.

## 2.2 Construction of JIWC dictionary

The JIWC dictionary was constructed from the following process using the episode data set.

#### Step 1: Morphological Analysis

Morphological analysis of all the texts in the episode data set is conducted and the vocabulary list is obtained. Japanese morphological analyzer (Juman++)\*<sup>2</sup> is used for the morphological analysis and extracted vocabulary is limited to nouns, verbs, adjectives, and adverbs. Similarly, we delete all the numbers from the vocabulary list.

---

<sup>1</sup><https://crowdsourcing.yahoo.co.jp/>

<sup>2</sup><http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

Table 1: Examples of episodes in episode data set

悲しい	両親が立て続けに亡くなったこと 好きな女の子に告白してふられたとき
不安	大地震を体験したこと 就職の面接の時
怒り	浮気をされたこと 小学生のとき初デートのタイミングで雨が降って中止になったとき，天候にイカリ
嫌悪感	浮気をされたこと 大嫌いな上司に嫌味を言われたとき
信頼感	いい人と結婚できたこと困っている自分を俺が助けてやる，と言ってくれた 小学生の時の担任の先生
驚き	子供が受験に合格したこと 大学に合格した時
楽しい	自立できたこと バスケットボールの試合をしているとき

Step 2: Counting frequency of words

For the vocabulary obtained in Step 1, we find the frequency of each word. We delete the words with a frequency of less than 10 from the vocabulary list. For each word left in the vocabulary, the frequency of occurrence in each emotion category is counted.

Step 3: Scoring of words

In this step, we calculate the score  $S_{ij}$  of every word in the vocabulary list with respect to each emotion. Scoring is done as shown in equation 1.

$$S_{ij} = \frac{W_{ij}}{W_i^*} \log_2(W_{ij} + 1) \quad (1)$$

where,

$W_{ij}$  is the frequency of the word  $w_i$  in emotion category  $j$

$W_i^*$  is the total frequency of word  $w_i$  in all categories

Step 4: Classification

For every word, we find the emotion category for which the score

	A	B	C	D	E	F	G	H
1	Words	Sadness	Anxiety	Anger	Disgust	Trust	Surprise	Joy
2	アーティスト	0.3917	0.0233	0.0233	0	0.216	0.3917	2.7328
3	アイス	1.2957	0	0.0769	0	0	0.0769	0.9942
4	アイドル	0.6738	0.04	0.1268	0.24	0.04	0.6738	0.6738
5	あおり	0.1689	0.0182	3.0525	1.1891	0	0.0182	0
6	あげ	1.175	0.0909	0.5455	0	0.0909	0	0.0909
7	あたった	0	0	0.0833	0	0.774	1.75	0
8	あった	0.9016	0.9335	0.8384	0.7452	0.4785	0.7761	0.3144
9	あって	0.6502	0.3314	0.1538	0.0813	1.4025	0.2381	0.1538
10	あと	0.5619	0.7324	0.5619	0.2609	0.0435	0	0.2609
11	アドバイス	0	0	0.0961	0	4.697	0	0

Figure 1: JIWC-A

is the highest and add the word to the dictionary as the expression of that emotion. Words for which two or more categories have the highest scores, are added to the dictionary of all of such emotions. We additionally created a dictionary with words as keys and the emotion/s with the highest score for the words as the values.

### 3 Distribution of JIWC dictionary

The latest version of the dictionary is made available on the GitHub page<sup>\*3</sup> of the social computing laboratory of Nara Institute of Science and Technology. JIWC dictionary is available in three different formats as described below.

#### 1. JIWC-A

JIWC-A contains the score of words with respect to each emotion as obtained in Step 3 of section 2.2. An example of JIWC-A 2018 version is shown in Figure 1.

#### 2. JIWC-B

JIWC-B contains each word and the emotion category or categories it is associated with. An example of JIWC-B 2018 version is shown in Figure 2.

<sup>\*3</sup><https://github.com/sociocom/JIWC-Dictionary>

	A	B	C	D	E	F
1	Words	Emotion-1	Emotion-2	Emotion-3	Emotion-4	Emotion-5
2	ペット	悲しい				
3	死んだ	悲しい				
4	友人	信頼感				
5	亡くなった	悲しい				
6	ない	信頼感				
7	何も	怒り				
8	して	信頼感				
9	おっさん	嫌悪感				
10	嫌な	嫌悪感				
11	顔	嫌悪感				

Figure 2: JIWC-B

	A	B	C	D	E	F	G
1	Sadness	Anxiety	Anger	Disgust	Trust	Surprise	Joy
2	アイス	あった	あおり	いい	あって	アイドル	アーティスト
3	アイドル	あと	アパート	いじめ	アドバイス	あたった	アイドル
4	あげ	いか	あまりに	いた	あなた	あまり	アニメ
5	オークション	いく	いい加減な	いつまでも	アメリカ	いう	いった
6	おじい	いけない	いくら	いる	あり	いきなり	イベント
7	かかった	いつ	いけない	いわ	ある	いつの間にか	いま
8	がん	いつまで	イジメ	ウソ	アルバイト	インターネット	インターネット
9	ガン	うつ病	いつまでも	エレベーター	い	かかった	おいしい
10	キャンセル	うまく	イライラ	おかしい	いって	かなり	おしゃべり
11	クラウドソーシング	お腹	インターネット	おじさん	いつでも	くじ	お出かけ

Figure 3: JIWC-C

### 3. JIWC-C

JIWC-C holds the dictionary of each emotion category and all words associated with that emotion. An example of JIWC-C 2018 version is shown in Figure 3.

### 参考文献

- [1] Robert Plutchik. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4):344-350, 2001.
- [2] 柴田 大作, 若宮 翔子, 木下 彩栄, and 荒牧 英治. 音声発話による認知

症スクリーニング技術の開発 感情表現辞書を用いた発話内容の質的  
分析. 医療情報学, 37(6):303–311, 2017.