Michiko

Haiku is a famous genre of Japanese poetry. One haiku poem is a stanza consisting of 3 lines. Each line has fixed count of sound units (morae, similar to syllables in English).

In the beginning of the 20th century, haiku was adapted to English and later also into other languages. Foreign adaptation consists of 3 lines of 5-7-5 syllables, since morae does not have an equivalent in other languages. Foreign haiku does not always follow strict rules of traditional haiku poem. Therefore, topics are not limited to nature and often haiku in other language does not conform to the formal haiku rules.

The proposed approach creates English haiku poems about nature. Poems are generated in 3 lines and conform to traditional 5-7-5 syllable pattern.

Poem model defines basic structure of haiku poem to be generated. It contains knowledge on how to create the output poem by filling these models with words from dictionary based on their part of speech and syllable count.

Example model

*Numeral – noun*

*Preposition – adjective – noun – noun*

*Noun – noun –noun*

*1 – 4*

*1 – 2 – 2 – 2*

*2 – 1 – 2*

Example of generated poem from the above stated model:

*each January*

*in frozen sunshine brightness*

*cutting pine branches*

The corpus is a large set that consists of haiku poems written by human authors. It is used for two main goals:

* creating dictionary of words - splitting poems into words, gathering metadata for each word
* extracting poem models - list of parts of speech, each with certain syllable count

It is important to note, that the algorithm implementing the proposed approach is not using nor quoting whole haiku poems nor the verses from the corpus. It only uses the poems to extract information that is further used to create new poems in haiku generation process. Haiku corpus is not saved in database, only models and dictionary, extracted from the poems, are saved.

Haiku poem has a characteristic topic – the topic of nature. To achieve that generated poem is thematically consistent with haiku genre, dictionary is created by using words from the corpus. The corpus is a large set of haiku poems written by human authors from online haiku portals. This dictionary provides the word set for selection into the output poem.

Proposed approach creates poems by filling poem models with words based on their syllable count and part of speech. These properties of words are considered as metadata, providing basic information on when and where can be the word used in the resulting poem. Thus, dictionary of words with these metadata available is a precondition to successfully generate haikus.

Metadata determination was implemented using Words API. Database of Words API contains 150,000 words and, besides part of speech and syllable count, it also provides more features, such as list of synonyms, antonyms or rhymes.

The proposed approach and application are work in progress. Its goal is to create haiku poems that would be considered by users as aesthetically pleasing.