

FEATURE EXTRACTION FROM POEMS

R.Pandian
Computer Science and
Engineering,
College of Engineering, Guindy

M.Manikandan
Computer Science and
Engineering,
College of Engineering, Guindy

Guide : Dr Mahalakshmi G S
Assistant Professor, Computer
Science and Engineering,
College of Engineering, Guindy

Abstract :

Poems are actually very different from a normal prose and processing a poem can be quite different from processing a prose. Various researches are being done extensively to develop on the ideas of processing a poem. Basically, a reader who reads a poem needs to get the feel of various poetical features used in the poem and the actual meaning or context of the poem. This project basically takes a poem as input and finds out various poetical features that are used in the poem. Because to find out the actual meaning of context of the poem, first of all, the various poetical features used in the poem should be found out and taken into account. The features that are covered in the project are : Similes, whether they are used, if used what things are compared etc, compound words, idioms, stylistic forms and the type of the poem.

Introduction :

The project aims at bringing out the various poetic features of a poem. A poem is given as input to the software and it gives the features used in the poem as output. Various techniques are used to

extract the features of the poem. The detailed implementation and methodology for the project are as described.

Related works :

Poetic Features for Poem Recognition: A Comparative Study
JOURNAL OF PATTERN
RECOGNITION RESEARCH (2008)

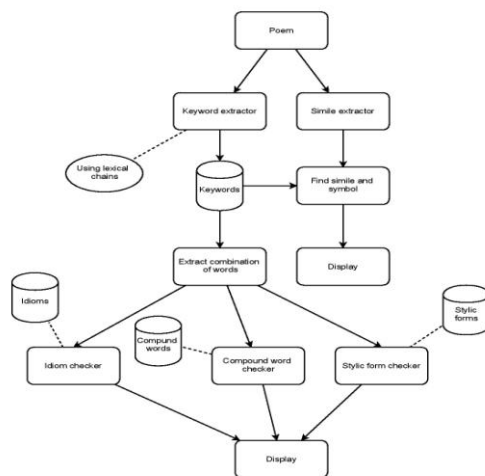
This paper describes the method used to distinguish a poem from a prose. This project basically gets a text as an input and tells whether the text is a poem or a prose. If it is a poem, then it extracts the various features used in the poem. Based on the poetic features extracted, the classification of poem is made and it is distinguished from other normal poems or prose.

Methodology :

The working of the project can be divided into different modules. The features which are considered as a part of the project are :

- Simile
- Compound words
- Idioms
- Stylistic forms
- Type of poem

Module diagram :



Keyword extractor :

This module is the fundamental module which actually extracts all the keywords from the poem. This uses lexical chains to extract the keywords. While extracting the keywords, all unwanted words have to be removed like articles, prepositions, connectors etc. Hence, this part basically provides the input for all other modules which are the keywords used in the poem.

Simile extractor :

Similes are said to occur when two things are explicitly compared using words like 'like' or 'as'. These such words are identified in the given poem and the things which are compared are found out. The finding of the things which are compared involves finding out whether the things are actually compared.

Extract combination of words :

Combination of consecutive words are extracted in order to check for idioms,

compound words and stylistic forms.

Database is maintained to keep track of all possible combination of words to check for other features.

Idioms, compound words and stylistic forms :

In order to find the idioms, compound words and stylistic forms that are used in the poem, the combination of words extracted in the previous step is checked with the database for idioms, compound words and stylistic forms. Separate data sets are maintained each for idioms, compound words and stylistic forms. The combination which matches with one of the data sets is identified to be a feature and is displayed.

Type of poem :

The type of poem can be either narrative or descriptive. The type of poem can be identified from words such as 'I', 'me', 'you' etc. If some of these words are present, then the poem is identified to be a narrative poem. Otherwise, the poem is considered to be a descriptive poem. A narrative poem is said in the poet's point of view ie, in first person view. A descriptive poem is one which actually describes something which is also narrated by the poet but is not addressed to the readers in first person point of view.

Results :

The project is evaluated against a set of poems and the different poetic features are extracted and tested. The results have found to be effective in finding the

features mentioned above from the poem. For example, the following shows the output obtained from two sample poems :

Poem - 1 :

INPUT:

Indian Dancer

*EYES ravished with rapture, celestially
panting, what passionate bosoms
aflaming with fire
Drink deep of the hush of the hyacinth
heavens that glimmer around them in
fountains of light;
O wild and entrancing the strain of keen
music that cleaveth the stars a wail of
desire,
And beautiful dancers with houri-like faces
bewitch the voluptuous watches of night.*

*The scents of red roses and sandalwood
flutter and die in the maze of their gem-
tangled hair,
And smiles are entwining like magical
serpents the poppies of lips that are
opiate-sweet;
Their glittering garments of purple are
burning like tremulous dawns in the
quivering air,
And exquisite, subtle and slow are the
tinkle and tread of their rhythmical,
slumber-soft feet like blue moon.*

*Now silent, now singing and swaying and
swinging, blossoms that bend to the
breezes or showers,
Now wantonly winding, they flash, now
they falter, and, lingering, languish in
radiant choir;
Their jewel-girt arms and warm, wavering,
lily-long fingers enchant through
melodious hours,
Eyes ravished with rapture, celestially
panting, what passionate bosoms
aflaming with fire!*

OUTPUT:

Similes

*symbol : smiles simile : serpents
symbol : garments simile : dawns
symbol : feet simile : moon*

Compound words :

around

Stylic forms

Idioms

blue moon

Poem type : Descriptive

Poem -2 :

INPUT:

Indian Weavers

*WEAVERS, weaving at break of day,
Why do you weave a garment so gay? . . .
Blue as the wing of a halcyon wild,
We weave the robes of a new-born child.*

*Weavers, weaving at fall of night,
Why do you weave a garment so bright? .*

..

*Like the plumes of a peacock, purple and
green,
We weave the marriage-veils of a queen.*

*Weavers, weaving solemn and still,
What do you weave in the moonlight chill?*

...

*White as a feather and white as a cloud,
We weave a dead man's funeral shroud.*

OUTPUT:

Similes

symbol : garment simile : wing
symbol : moonlight simile : feather
symbol : moonlight simile : cloud

Compound words :

moonlight

Stylic forms

Idioms

Poem type : Narrative

Output	Evaluation Parameters (Recall) %
Output for poem 1	96
Output for poem 2	94.5
Output for poem 3	93.6
Output for poem 4	89.3
Output	Evaluation Parameters (Precision) %
Output for poem 1	98

Output for poem 2	99
Output for poem 3	96
Output for poem 4	92

Limitations :

- It finds out only few features of the poem. But to identify the actual meaning of context of the poem, all possible poetic features have to be found out.
- The input needs to be distinguished whether it is a poem or a prose. This project assumes that the input is always a valid poem and gives the poem type as 'narrative' or 'descriptive'. It doesn't consider the case when normal prose is given as input.

Future direction :

The above limitations can be taken into account and the project can be extended to take care of those limitations. More poetic features can be identified from the poem and displayed so that it will be easy to identify the actual meaning or context of the poem.

Also, the input should be validated for a valid poem and invalid prose should be identified. But, identifying this is very much difficult. One basis on which this could be done is based on the number of words in each line. Poems usually follow the same number of words in every line. But, this cannot be taken as the only distinguishing attribute, because some poems don't follow this rule and may

contain different number of words in different lines. Another attribute that can be used to distinguish is the number of other poetic features used in the poem. If there are no other poetic features, the input can be considered to be a prose. But then, many poems which are written by modern poets do not contain any poetic features. Hence, this can also not be considered to distinguish between a poem and a prose.

Research is still being done to actually distinguish between a prose and poem. When this distinguishing is done and incorporated in the project and many other features are found, the project becomes a very much useful component in analysis of poems.

Conclusion :

Hence the project can be considered to be an important part in the future analysis of poems because analysis of poems is gaining lot of importance and it also has a huge research component. When the mentioned limitations are rectified, this can be used to bring out what actually a poem tries to convey. This might help readers in understanding the poem.

Future references :

Pandian -
pandian4mail@gmail.com
Manikandan -
manikandancse12@gmail.com