

Determination of Idiomatic Sentences in Paragraphs Using Statement Classification and Generalization of Grammar Rules

Naziya Shaikh, Government College of Arts, Science and Commerce, Quepem-Goa

Description

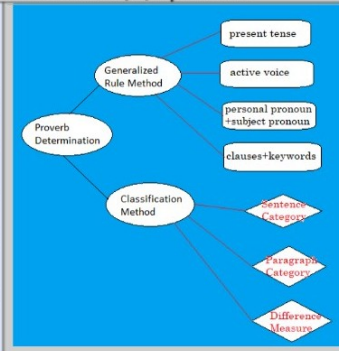
The translation systems are often not able to determine the presence of an idiom in a given paragraph. Due to this many systems tend to return the word-for-word translation of such statements leading to loss in the flavor of the idioms in the paragraph.

Approach to efficiently determine probability of any statement in a given English paragraph to be an idiom is suggested here.

This approach combines the rule-based generalization of idioms in English language and classification of statements based on the context to determine the idioms in the sentence.

The grammar-based detection involves parsing the sentences using parts of speech tagger and analyzing the tokens for various grammar rules that are mostly specific to the proverbs.

Context-based detection categorizes the keywords in the sentence and compared the difference between the categories of the given sentence with the categories of the rest of the sentences in the paragraph.



Classification Method

- Based on the fact that most proverbs do not relate semantically to the rest of the context that is present in the paragraph.
- Comparing the category of the sentence to the category of the entire paragraph without that sentence.
- Grading the difference in the general categories and the similarities between these categories.
- The higher the difference more is the probability of any given sentence being an idiom.

Generalized Rules Method

- Certain rules of the English language usually apply for most of the idioms in a paragraph.
- Certain rules can be generalized and used as criterias to indicate the presence of an idiom in a sentence.

The Results

Example Paragraph:

There is this boy in my neighborhood. He is Very strangely dressed and remains quite aloof. He has tattoos all over his body. But you know what they say, don't judge a book by its cover. So I went ahead and tried to converse with this boy. And I was right, this boy was indeed had a very interesting nature. He was just a teenager in his growing phase. That is when I realized to understand others we need to put ourselves in other people's shoes.

Context-Based Approach Example

Sentence	Sentence Category	Remaining paragraph without sentence	Paragraph Category	Score based on categories: Completely unmatched - 4.5 Matching to some extent: 3.5 matching well: 2 complete match: 1.5
There is this boy in my neighborhood.	real estate / low income housing	Very strangely dressed and remains quite aloof. He has tattoos all over his body. But you know what they say, don't judge a book by its cover. So I went ahead and tried to converse with this boy. And I was right, this boy was indeed had a very interesting nature. He was just a teenager in his growing phase. That is when I realized to understand others we need to put ourselves in other people's shoes.	/ style and fashion / footwear / shoes	3.5
	/ home and garden / gardening and landscaping / yard and patio		0.77	
	/ food and drink / food and grocery retailers / bakeries		0.25	
			0.44	
He is very strangely dressed and remains quite aloof.	/ law, govt and politics / politics	There is this boy in my neighborhood. He has tattoos all over his body. But you know what they say, don't judge a book by its cover. So I went ahead and tried to converse with this boy. And I was right, this boy was indeed had a very interesting nature. He was just a teenager in his growing phase. That is when I realized to understand others we need to put ourselves in other people's shoes.	/ style and fashion / footwear / shoes	3.5
	/ pets / cats		0.43	
	/ business and industrial / company / merger and acquisition		0.36	
			0.22	1.0

○ Natural Language Rule-Based Approach Example

Sentence	Clause Present?	Tense	Tense Score Present	Clause + that	Words: Realize +, told, said, knew, say -	Voice Active - 2.5 Passive - 0	He + Who (+1) Only He (-1)	Total Score
There is this boy in my neighborhood.	0	Simple Present	2.5	0	0	1.5	0	3.5
He is Very strangely dressed and remains quite aloof	0	Present	2.5	0	0	0	-2	0.5
He has tattoos all over his body.	0	Simple Present	2.5	0	0	1.5	-2	2
But you know what they say, don't judge a book by its cover.	0.5	Simple Present	2.5	0	0.5	1.5	0	4.5
I went ahead and tried to converse with this boy.	0	Past	0	0	0	1.5	0	1.5
And I was right, this boy indeed had a very interesting nature.	0	Simple Past	0	0	0	1.5	0	1.5
He was just a teenager in his growing phase.	0	Past Continuous	0	0	0	1.5	-2	-0.5
That is when I realized, to understand others we need to put ourselves in other people's shoes.	0.5	Simple Present	2.5	1	0.5	1.5	0	6