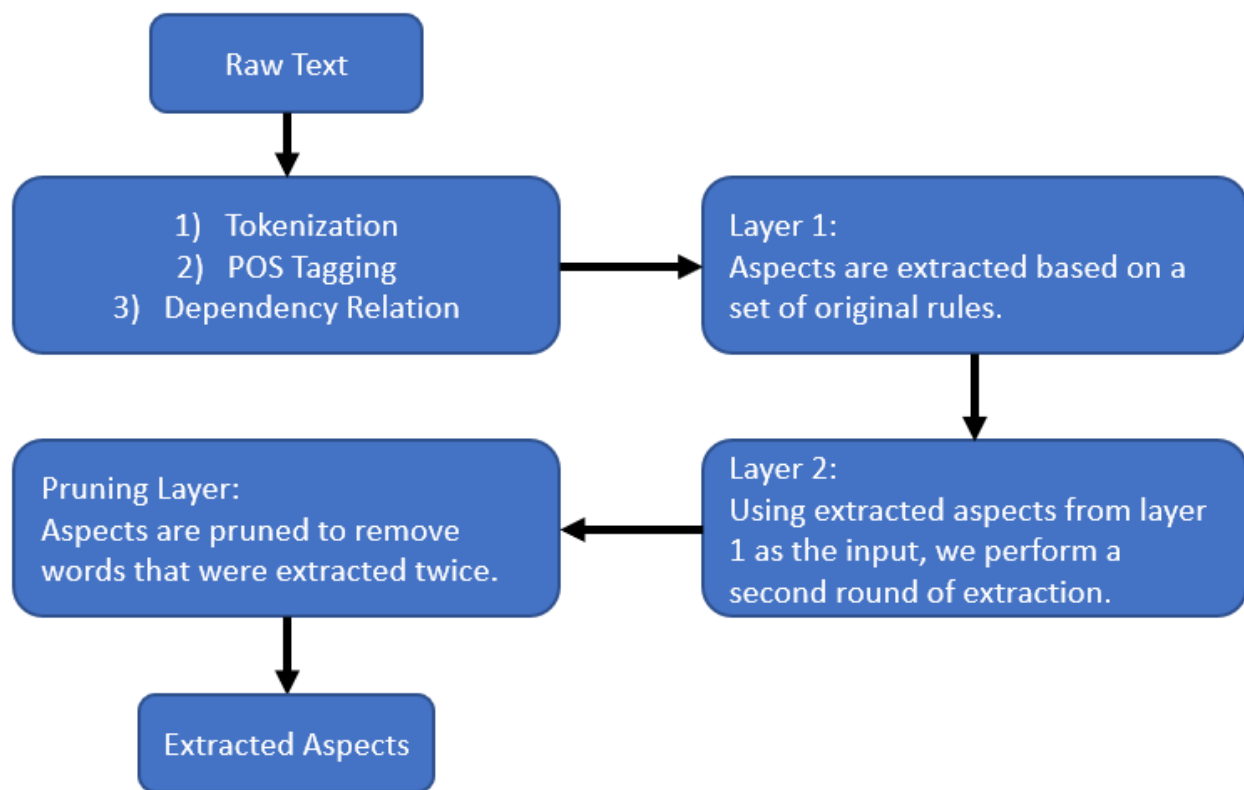# Unsupervised Aspect Term Extractions

Despite having the ability to train supervised machine learning models, utilizing unsupervised learning methods would be more applicable in the real world. Supervised models tend to be domain-specific. For example, we are unable to use a model trained using restaurant reviews to extract aspects from a contrasting domain such as book reviews. In many instances, there would be a lack of training data in specific domains. Hence, we recognize the importance of unsupervised learning models in this instance. Therefore, we have proposed an unsupervised approach which exploit the observed pattern of aspect terms to the fullest.

As shown in the figure above, the unsupervised model consists of a 2-layer aspect term extraction module which exploits the rich linguistic patterns and distill them into rules. The unsupervised model does not make use of the aforementioned preprocessing steps. Instead, the Stanford dependency parser, Stanza, will be used for both dependency parsing and part-of-

speech tagging. Stop word and punctuation removal was omitted as it could have adverse effects on the dependencies of our text corpus.

In the first layer, five rules will be implemented to extract aspect terms. For each input text, we iterate through all the words in the input text and apply the following rules:

1) If the target word is a singular noun, and the word before it is an adjective and the adjective is not an opinion word, concatenate both words and append it to the list of aspect terms.

2) If the target word is a singular noun and the word before it is a singular noun, concatenate both words and append it to the list of aspect terms.

3) If the target word is a singular or plural noun and the word before it is an adjective, and the adjective is an opinion word, append the word to the list of aspect terms.

4) If the target word is a singular or plural noun, and is in a 'dobj' dependency relation with a base form verb or a past participle verb, append the word to the list of aspect terms.

5) If the target word is a singular or plural noun, and is in a 'nsubj' dependency relation with an adjective, append the word to the list of aspect terms.

In the second layer, 1 rule will be implemented to further extract aspect terms. For each input text, we iterate through all the words in the input text and apply the following rule:

1) If the target word is a singular or plural noun, and has a direct dependency relation with any of the single word aspects that were already extracted, append the word to the list of aspect terms.

In the pruning layer, we remove all words that have appeared in any of the two-word aspects. For example, 'pizza' will be removed if 'pepperoni pizza' has been extracted as well.