

Changes in the learning algorithm:

Triplie will learn associations from continuous blocks of texts written by another user and markov chains from lines written by the users.

Two special words will be added: the “beginning” and the “end”, to allow triplie to finish a sentence.

Changes in the response algorithm:

Normally, triplie extracts keywords, expands them by querying the association net then uses the expanded keywords to connect through the markov chains.

In the process, associated keywords are excited by the amount of mutual information that is available.

The modified algorithm will be a bit more complex (but a superset of the original algorithm):

Extract keywords

Instead of looking for the rarest of keywords, look for something that at least satisfies minimum frequency e.g. 1 or 2 occurrences. (configurable)

The value of course is still the accumulated mutual information.

Expand using a stemmer

Expand the initial set of keywords by finding words that have the same base form. The base form similarity checker can be plugged in, for example an english-specific stemmer can be used, or a levenstein comparison can be used (anything that compares two words and says true if the word is basically the same) or maybe a stemmer that always returns false.

Expand using n-gram context similarity.

Each of the results from the previous phase will be expanded with an n-gram context similarity search. This is how this algorithm works:

- 1) Find all the n-grams where the word is in the middle (n=3 or n=5 with a configurable parameter)
- 2) Find all the words which can also be placed in any of those n-grams along with the number K of n-grams where they were found. (minimum K is also configurable)

Do multiple-BFS on the markov graph

Now every word in the expanded set has its own set of alternative replacements which can be found - while doing the markov breadth first search. If any acceptable alternative is found we can continue with the search even though the “wrong” word was found, but we mark what the right word is in that position.

When the (configurable) minimum number of keywords is reached, the algorithm is allowed to find the “startword” or the “endword”.

After the search is complete we go through all the marked words and replace them with the right ones.