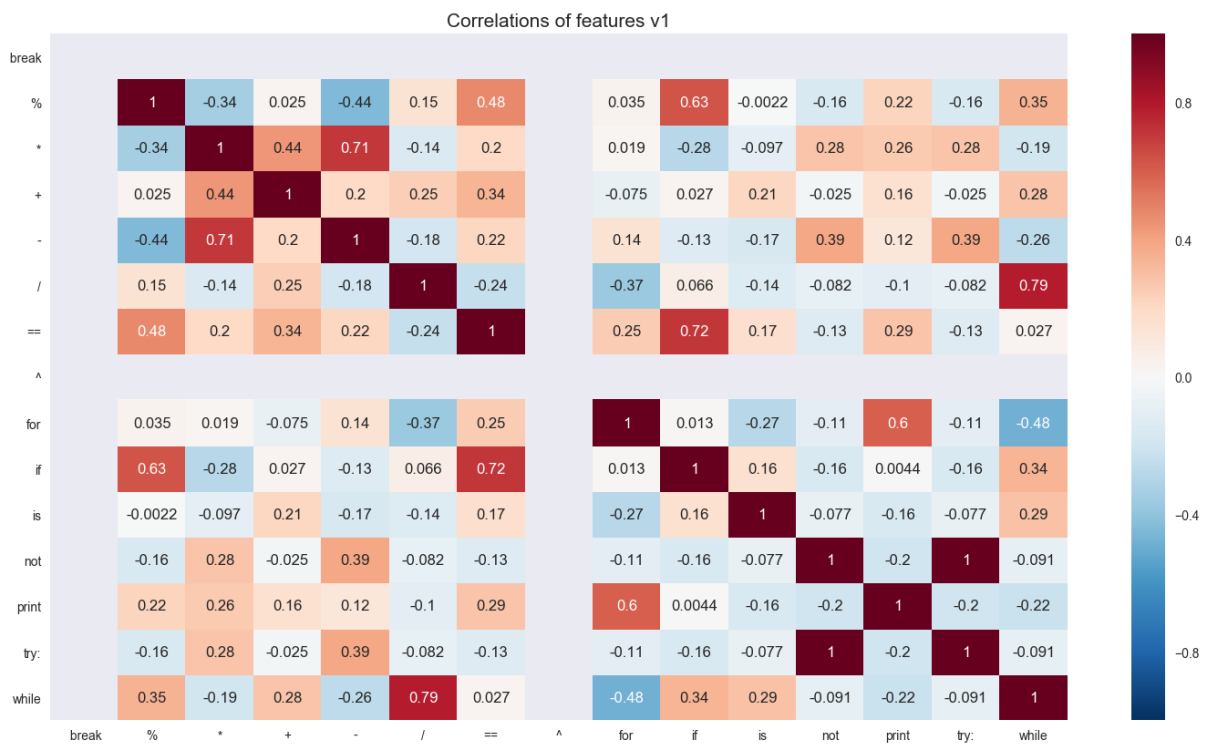
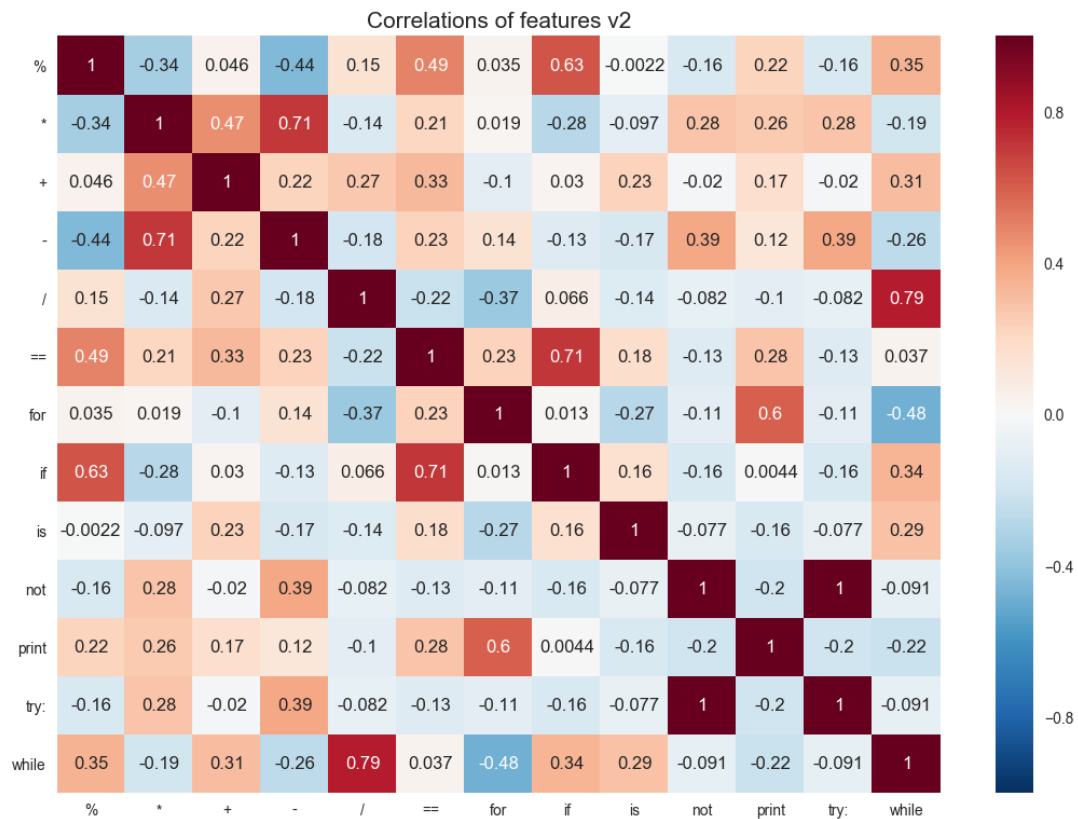


Decision of what features to use

First try:



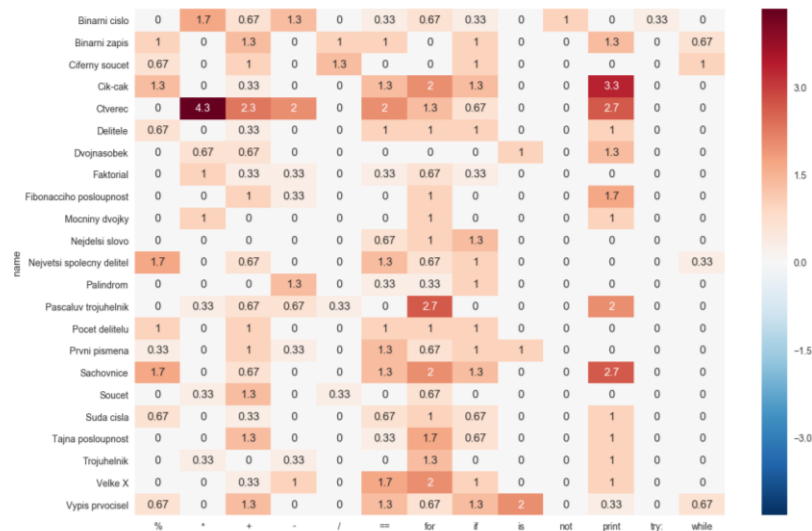
Results: the 'break' and '^' features were not found at all so I decided to remove them.



Next try with removed 'break' and '^' features:

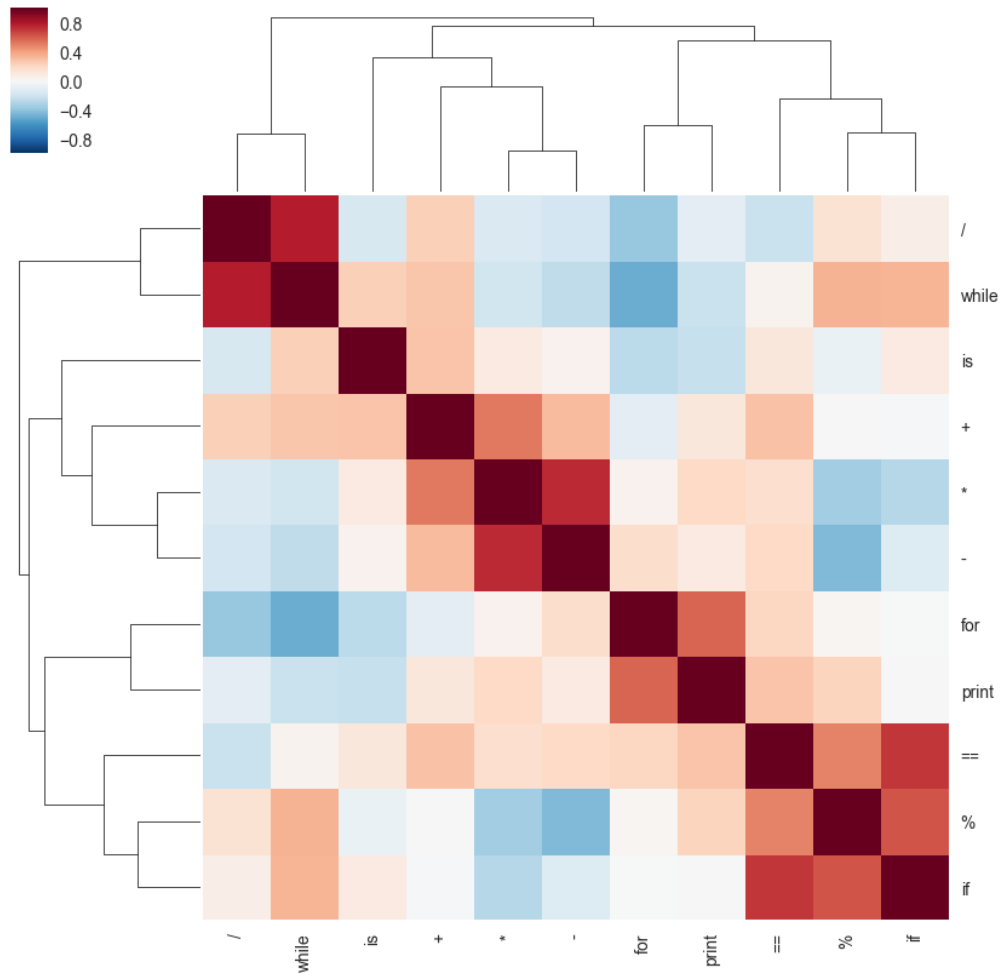
I had noticed that the correlation of 'try' and 'not' is too high so I decided to take a closer look at it.

This is the feature matrix used in the comparison above:



I have noticed that the 'try' and 'not' features have been used only in one item so they have a high correlation. I have decided to remove them as well.

The next correlation looks like this:



Results: We can see that some of the features has a high correlations but I think it is not that high that it must be removed. I think this features are the ones, that I would use.