Some more results:

I've implemented:

4 versions of term frequency (the tf bit in tf-idf). Let f be the raw term occurrence count:

- Sqrt(f), as used in Lucene
- f, Wikipedia
- log(f + 1), Wikipedia
- 0.5 + 0.5 * f / max(f, over all terms in the doc), Wikipedia

4 versions of idf. Let d be the raw document occurrence count and numDocs be the total num of docs in the corpus:

- 1 + log(numDocs / (d + 1)), Lucene (d + 1 to handle a query term not in the corpus)
- log(numDocs / d), Wikipedia
- numDocs / d, non-log version of above suggested by Scott
- 1, don't use idf term, suggested by Scott

The quality metric Scott suggested below.

Refs:
Wikipedia: http://en.wikipedia.org/wiki/Tf%E2%80%93idf
Lucene:
http://lucene.apache.org/core/4_0_0/core/org/apache/lucene/search/similarities/TFIDFSimilarity.html

I've used 2000 docs the same 3 news-groups, with distinct subject areas, from the 20Newsgroups data set used in
"Exemplar-based Visualization of Large Document Corpus", Yanhua Chen et. al.
http://www.cs.wayne.edu/~mdong/tvcg09.pdf
They are:

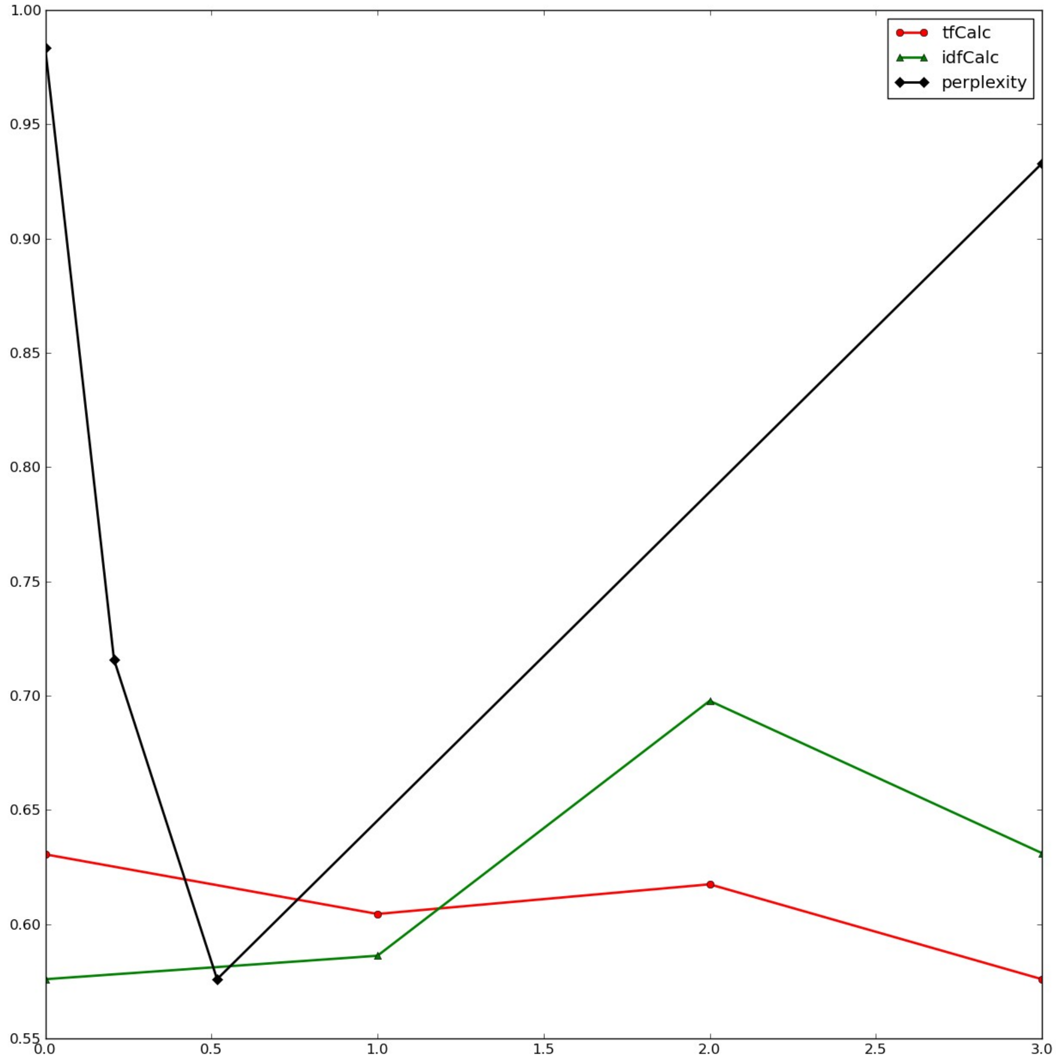
- sci.med
- rec.sport.baseball
- comp.sys.ibm.pc.hardware

I've generated term matrices with all 16 tf-idf versions and processed them with bh_tsne using 4 values for perplexity (http://homepage.tudelft.nl/19j49/t-SNE.html: " It is comparable with the number of nearest neighbors*k* that is employed in many manifold learners."): 1, 3, 6 and 30, stopping after 500 iterations.

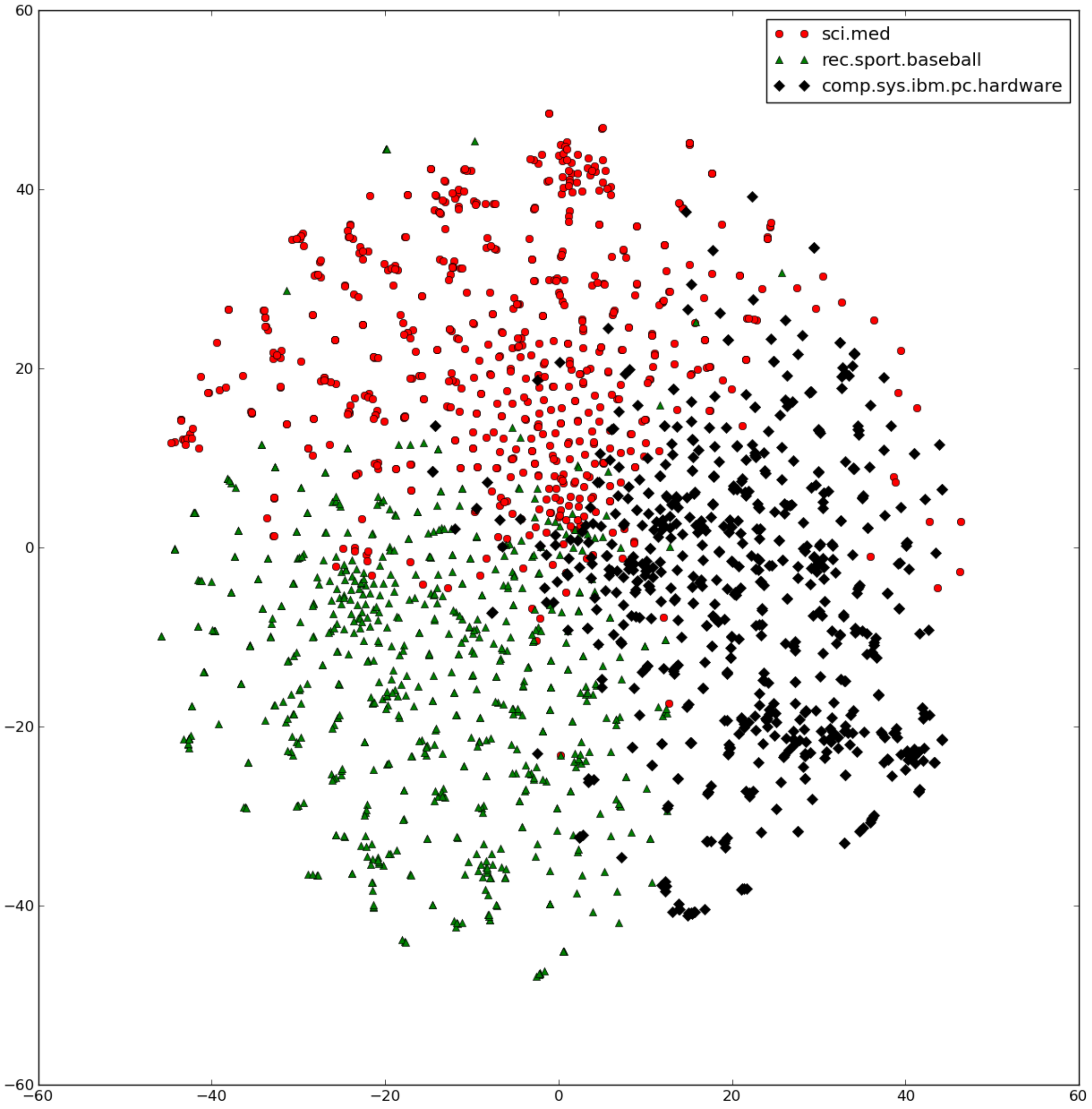
Here is the quality metric (smaller is better - bearing in mind that this is a clustering metric and we're not really clustering) plotted against one of the 3 variables with the other two fixed.
The best value was with:

- tfCalc = 3 (last of the versions described above),
- idfCalc = 0 (first of the versions described above); and
- perplexity = 6 (I've scaled the plotted values to fit the 0..3 range of the other variables so this appears around 0.5 on the plot).

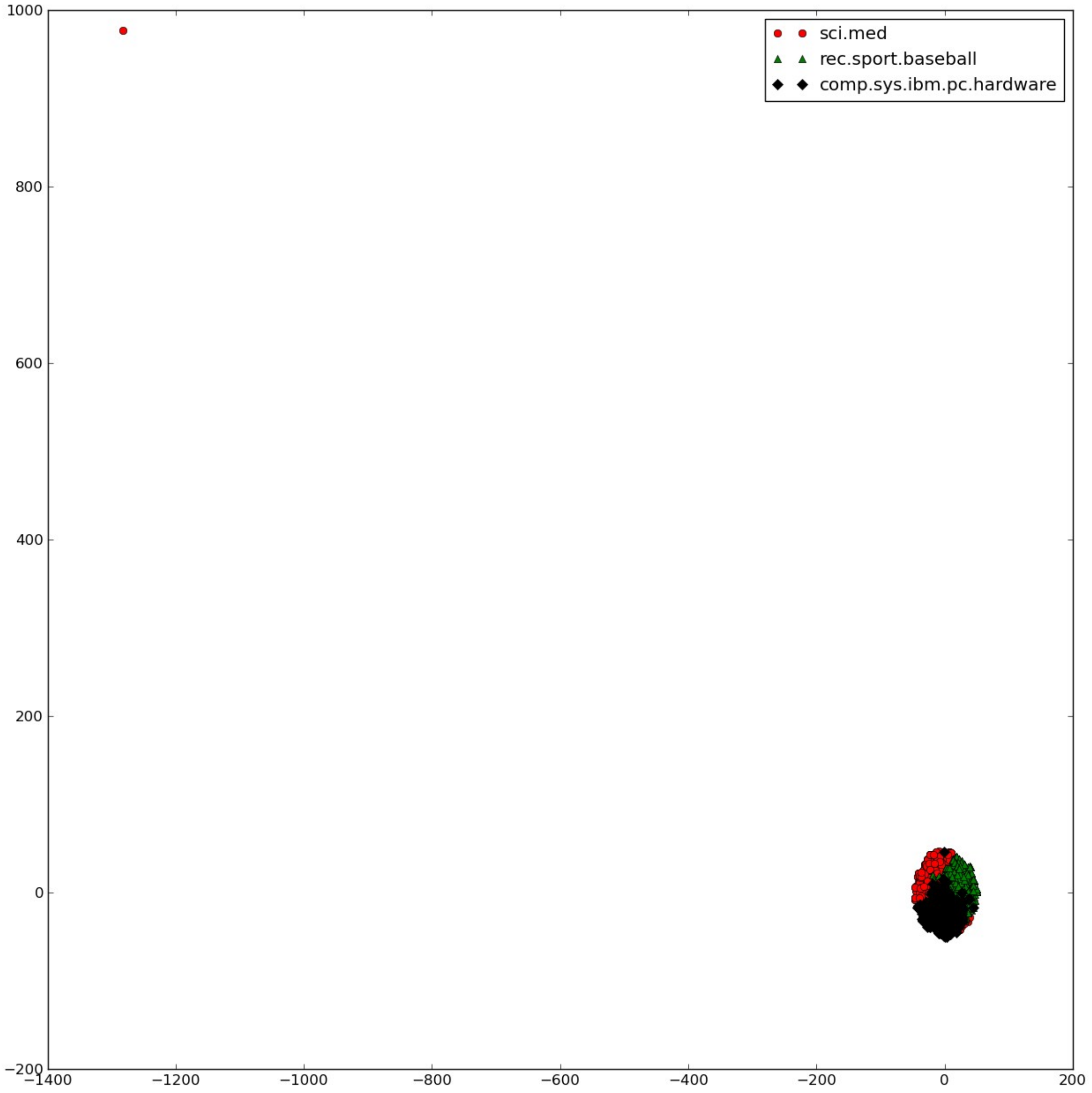
So these are the values used for the fixed values. It looks like we could further improve the result by adjusting perplexity around 6.



and here is the winning result:



Some of the other results with similar quality scores have a single distant outlier, e.g. tfCalc=3, idfCalc=1, perplexity=6; which may indicate an issue with the bh_tsne implementation.


● ● sci.med
▲ ▲ rec.sport.baseball
◆ ◆ comp.sys.ibm.pc.hardware