This exercise is taken from the article “An Introduction to Latent Semantic Analysis” by T. Landauer et al..

Nine documents are given by:

Example of text data: Titles of Some Technical Memos

c1: *Human* machine *interface* for ABC *computer* applications
c2: A *survey* of *user* opinion of *computer system response time*
c3: The *EPS user interface* management *system*
c4: *System* and *human system* engineering testing of *EPS*
c5: Relation of *user* perceived *response time* to error measurement

m1: The generation of random, binary, ordered *trees*
m2: The intersection *graph* of paths in *trees*
m3: *Graph minors* IV: Widths of *trees* and well-quasi-ordering
m4: *Graph minors*: A *survey*

This results in the following word-document matrix:

| | c1 | c2 | c3 | c4 | c5 | m1 | m2 | m3 | m4 |
|---|---|---|---|---|---|---|---|---|---|
| **human** | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| **interface** | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **computer** | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **user** | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| **system** | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 |
| **response** | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| **time** | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| **EPS** | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| **survey** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **trees** | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| **graph** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| **minors** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

Please calculate the SVD using numpy (or a comparable library in your favorite programming language). Calculate the document-document correlations for a two dimensional subspace and compare it to the document-document correlations from the original word-document matrix.