

Result

This two columns are contrapositive

Permutation Name	P _{0.25}	P _{0.50}	P _{0.75}	P _{1.00}	P _{mean1}	P _{mean2}	R _{norm}	P _{norm}
Raw TF	0.0952	0.8803	0.5917	0.8192	0.5218	0.7543	0.9023	0.6310
Boolean	0.1001	0.5162	0.5581	0.5112	0.0579	0.3912	0.8991	0.5989
Dice	0.1136	1.2411	1.0392	0.5852	0.7982	0.5332	0.9205	0.6561
Unstemmed	0.1171	0.4879	0.8300	0.7094	0.4784	0.5078	0.8238	0.5758
Include Stopword	0.1572	1.4933	1.3518	1.7778	1.0008	0.8139	0.9273	0.7027
Equal weight	0.1171	0.4879	0.8300	0.7094	0.4784	0.5078	0.8238	0.5758
1.1.4.1 weight	0.1892	1.6460	0.5042	1.4687	0.7808	0.7253	0.9228	0.6657
Default	0.1572	1.4933	1.3522	1.177	1.0010	0.8142	0.9271	0.7028

Part 2

Question 1

Documents most similar to query number 6

Run: vector1 vector1

Show the terms that overlap between the query and retrieved docs (y/n):

Documents Most Similar To Query number 6

Similarity	Doc#	Author	Title
0.11802297	2753	Pfefferkorn	# A Heuristic Problem Solvin
0.11369955	2828	Clark	# Hierarchical Geometric Mod
0.10426575	3035	Wetherbe	# A Strategic Planning Metho
0.09487657	2087	Pager	# A Number System for the Pe
0.09209212	2389	Eastman	# Preliminary Report on a Sy
0.08325429	2187	Amarel	# Computer Science: A Concep
0.07948477	1543	Howard	# Computer Formulation of th
0.07556431	2721	Claudson	# The Digital Simulation of
0.07030709	2230	Bracchi	# A Language for Treating Ge
0.06964399	2671	Stone	# A Note on a Combinatorial
0.06911736	2836	Loui	# Weighted Derivation Trees
0.06678033	2505	Roy	# Reflection-Free Permutatio
0.06429983	2707	Bitner	# Backtrack Programming Tech
0.05703098	695	Carlson	# Use of the Disk File on St
0.05663329	2485	Nolan	# Managing the Computer Reso
0.05441230	3086	Fredman	# On the Complexity of Compu
0.05421074	2834	Bitner	# Efficient Generation of th
0.05388920	3040	Freuder	# Synthesizing Constraint Ex
0.05341327	2909	Wirth	# What Can We Do about the U
0.05303448	605	—	# Computer Simulation Of Cit

Documents most similar to query number 9

Show the terms that overlap between the query and retrieved docs (y/n):

Documents Most Similar To Query number 9

Similarity	Doc#	Author	Title
0.29967121	2621	Purdy	# A High Security Log-in Pro
0.29111007	2311	Benjamin	# A Generational Perspective
0.28601928	2949	Tajibnapis	# A Correctness Proof of a T
0.26890708	1685	Schurmann	# GAN, a System for Generati
0.26643270	2345	Ashenhurst	# Curriculum Recommendations
0.25615308	2317	Rosen	# Programming Systems and La
0.25521602	1750	Fuchel	# Considerations in the Desi
0.23819977	3150	Winograd	# Beyond Programming Languag
0.23172427	2372	Conway	# On the Implementation of S
0.23084057	1844	Baskin	# A Modular Computer Sharing
0.22859145	1728	Coffman	# Further Experimental Data
0.22272442	3141	Chang	# An Improved Algorithm for
0.21877208	2407	Couger	# Curriculum Recommendations
0.21780789	3082	Lamport	# Time, Clocks, and the Orde
0.20998310	3068	Popek	# A Model for Verification o
0.20798938	2522	Rodriguez	# The Design, Implementation
0.20070007	2151	Nemeth	# User Program Measurement i
0.19377160	2482	Howard	# Mixed Solutions for the De
0.18795481	2951	Mamrak	# Dynamic Response Time Pred
0.18538285	1747	Van	# Three Criteria for Designi

Documents most similar to query number 22

```
Show the terms that overlap between the query and retrieved docs (y/n):
*****
Documents Most Similar To Query number 22
*****
Similarity Doc# Author Title
=====
* 0.36680008 2678 Wright # Visible Surface Plotting P
* 0.34151904 2751 Phong # Illumination for Computer
* 0.29085269 2369 Matsushita # Hidden Lines Elimination f
0.27576401 2004 Bouknight # A Procedure for Generation
* 0.27189734 2473 Macleod # Hidden-Line Plotting Progr
* 0.25537623 2827 Levin # A Parametric Algorithm for
* 0.24314205 2384 Williamson # Hidden-Line Plotting Progr
* 0.24224051 2829 Blinn # Texture and Reflection in
0.24112065 2913 Crow # The Aliasing Problem in Co
* 0.23320085 2828 Clark # Hierarchical Geometric Mod
* 0.19665432 2692 Sutherland # Reentrant Polygon Clipping
0.19599872 1915 Galimberti # An Algorithm for Hidden Li
0.19013702 1978 Smith # The Use of Interactive Gra
0.14210263 2925 Fuchs # Optimal Surface Reconstruct
0.13362597 2924 Wu
0.11044683 2152 Newman # Display Procedures
0.10980543 2674 Barrett # Scan Conversion Algorithms
0.10705814 2687 Jordan # A Cell Organized Raster Di
0.10646645 2325 Traub # Numerical Mathematics and
0.10420109 3059 Towsley # Models for Parallel Proces
```

Question 2

6(1-5) vector overlap

```
=====
Vector Overlap 6 2753 Docfreq
=====
planning 2 67.363186 18
=====
Vector Overlap 6 2828 Docfreq
=====
motion 4 5.874931 9
geometric 2 41.124515 9
=====
Vector Overlap 6 3035 Docfreq
=====
planning 2 41.454268 18
=====
Vector Overlap 6 2087 Docfreq
=====
combinatorial 2 20.305692 20
=====
Vector Overlap 6 2389 Docfreq
=====
planning 2 25.908918 18
```

6(6-10) vector overlap

```
Continue (y/n)?: =====
Vector Overlap 6 2187 Docfreq
=====
planning 2 25.908918 18
=====
Vector Overlap 6 1543 Docfreq
=====
motion 4 11.749861 9
=====
Vector Overlap 6 2721 Docfreq
=====
planning 2 25.908918 18
dynamics 2 33.429305 4
=====
Vector Overlap 6 2230 Docfreq
=====
planning 2 25.908918 18
geometric 2 5.874931 9
=====
Vector Overlap 6 2671 Docfreq
=====
combinatorial 2 20.305692 20
```

9(1-5) vector overlap

```
Continue (y/n)?:
```

=====				
Vector Overlap	9	2621	Docfreq	
=====				
Security	2	18.378735	7	
operating	2	13.105459	121	
systems	4	20.083623	344	
=====				
Vector Overlap	9	2311	Docfreq	
=====				
systems	4	44.630273	344	
=====				
Vector Overlap	9	2949	Docfreq	
=====				
networks	2	34.303725	44	
network	2	43.045994	64	
operating	2	13.105459	121	
distributed	2	62.092140	27	
=====				
Vector Overlap	9	1685	Docfreq	
=====				
networks	2	21.439828	44	
network	2	117.398167	64	
=====				
Vector Overlap	9	2345	Docfreq	
=====				
systems	4	37.935732	344	

9(6-10) vector overlap

=====				
Vector Overlap	9	2317	Docfreq	
=====				
operating	2	19.658189	121	
systems	4	20.083623	344	
=====				
Vector Overlap	9	1750	Docfreq	
=====				
operating	2	39.316377	121	
systems	4	26.778164	344	
=====				
Vector Overlap	9	3150	Docfreq	
=====				
systems	4	29.009677	344	
=====				
Vector Overlap	9	2372	Docfreq	
=====				
Security	2	18.378735	7	
operating	2	16.381824	121	
systems	4	22.315137	344	
=====				
Vector Overlap	9	1844	Docfreq	
=====				
network	2	3.913272	64	
systems	4	64.713896	344	

22(1-5) vector overlap

=====			
Vector Overlap	22	2678	Docfreq
=====			
hidden	4	22.028824	13
line	2	14.710825	81
surface	2	20.110531	21
computer	2	7.281006	519
graphics	2	16.640529	50
=====			
Vector Overlap	22	2751	Docfreq
=====			
hidden	4	38.550442	13
surface	2	35.193430	21
algorithms	2	2.830408	189
computer	2	9.101257	519
graphics	2	16.640529	50
=====			
Vector Overlap	22	2369	Docfreq
=====			
hidden	4	27.536030	13
line	2	18.388531	81
This	2	1.919423	470
computer	2	14.562011	519
graphics	2	33.281058	50
=====			
Vector Overlap	22	2004	Docfreq
=====			
hidden	4	22.028824	13
line	2	14.710825	81
surface	2	45.248696	21
This	2	1.919423	470
computer	2	12.741760	519
graphics	2	16.640529	50
=====			
Vector Overlap	22	2473	Docfreq
=====			
I	2	8.705014	176
hidden	4	22.028824	13
line	2	14.710825	81
surface	2	20.110531	21

22(6-10) vector overlap

Continue (y/n)?:			
=====			
Vector Overlap	22	2827	Docfreq
=====			
hidden	4	27.536030	13
surface	2	55.303962	21
This	2	1.919423	470
computer	2	7.281006	519
graphics	2	16.640529	50
=====			
Vector Overlap	22	2384	Docfreq
=====			
hidden	4	22.028824	13
line	2	14.710825	81
surface	2	20.110531	21
=====			
Vector Overlap	22	2829	Docfreq
=====			
hidden	4	22.028824	13
surface	2	35.193430	21
This	2	1.919423	470
computer	2	7.281006	519
graphics	2	16.640529	50
=====			
Vector Overlap	22	2913	Docfreq
=====			
hidden	4	27.536030	13
surface	2	25.138164	21
algorithms	2	2.830408	189
This	2	1.919423	470
computer	2	7.281006	519
graphics	2	16.640529	50
=====			
Vector Overlap	22	2828	Docfreq
=====			
hidden	4	22.028824	13
surface	2	50.276329	21
algorithms	2	22.643266	189
computer	2	1.820251	519

Question 3

Documents most similar to document number 239

```
Show the terms that overlap between the query and retrieved docs (y/n):
*****
Documents Most Similar To Document number 239
*****
Similarity Doc# Author Title
=====
1.00000000 239 Verhoeff # Inefficiency of the Use of
0.47631477 1032 Belzer # Theoretical Considerations
0.20582286 651 Gremis # A Survey of Languages and
0.17846748 652 Sable # Use of Semantic Structure
0.16159511 634 Salton # Manipulation of Trees in I
0.15581755 1329 Mano # Simulation of Boolean Func
0.15498661 1207 Dodd # Remarks on Simulation of B
0.14172778 275 Sams # Dynamic Storage Allocation
0.12039028 3012 Lucas # The Use of an Interactive
0.11927326 635 Baker # A Note on Multiplying Bool
0.11921741 2070 Hsiao # A Formal System for Inform
0.11650282 2965 Hanani # An Optimal Evaluation of B
0.11069451 2160 Wong # Canonical Structure in Att
0.11035045 891 Whitley # Everyman's Information Ret
0.10730473 1233 --- # Conventions for the Use of
0.10484284 292 Kehl # An Information Retrieval L
0.10431240 1457 Salton # Data Manipulation and Prog
0.10133687 3168 Laird # Comment on "An Optimal Eva
0.10113658 2824 Duong # An Improvement to Martin's
0.09786275 948 Healy # Note on the Use of Procedu
0.09722253 2340 Martin # A Boolean Matrix Method fo
```

Documents most similar to document number 1236

```
*****
Documents Most Similar To Document number 1236
*****
Similarity Doc# Author Title
=====
1.00000000 1236 Salton # The SMART Automatic Docume
0.26313687 1457 Salton # Data Manipulation and Prog
0.25511309 634 Salton # Manipulation of Trees in I
0.21952842 2307 Salton # Dynamic Document Processin
0.18316917 2711 Salton # A Vector Space Model for A
0.15439814 2575 Van # The Best-Match Problem in
0.12829245 1536 Lesk # Dynamic Computation of Der
0.12640688 2990 Yu #
0.12398634 3012 Lucas # The Use of an Interactive
0.11581885 1699 Rubinoff # Experimental Evaluation of
0.11149786 3135 Lesk # Detection of Three-Dimensi
0.10703329 891 Whitley # Everyman's Information Ret
0.10701727 1271 Davis # Secondary Key Retrieval Us
0.10275866 2278 Tan # On Foster's Information St
0.10058508 1927 Salton # Information Science in a P
0.09802835 275 Sams # Dynamic Storage Allocation
0.09450376 1515 Levien # A Computer System for Infe
0.09191872 651 Gremis # A Survey of Languages and
0.09095246 1937 Day # CODAS: A Data Display Syst
0.08961770 798 Scheff # A Catalogue Entry Retrieva
```

Documents most similar to document number 2740

```
*****
Show the terms that overlap between the query and retrieved docs (y/n): Documents Most Similar To Document number 2740
*****
Similarity Doc# Author Title
=====
1.00000000 2740 Lauesen # A Large Semaphore Based Op
0.33995429 1749 Dijkstra # The Structure of the "THE"
0.25770863 2379 Liskov # The Design of the Venus Op
0.23047885 2920 Devillers # Game Interpretation of the
0.22895926 2378 Gaines # An Operating System Based
0.20851202 2228 Holt # Comments on Prevention of
0.20387781 2500 Frailey # A Practical Approach to Ma
0.19427369 2342 Gilbert # Interference Between Commu
0.19190956 3043 Hansen # Distributed Processes: A C
0.17824970 2280 Parnas # Comment on Deadlock Preven
0.16790969 2080 Hansen # The Nucleus of a Multiprog
0.15519363 2597 Hoare # Monitors: An Operating Sys
0.15349835 2865 Owicki # Verifying Properties of Pa
0.15327706 2320 Hansen # Structured Multiprogrammin
0.15266054 2777 Parnas # On a Solution to the Cigar
0.14506236 2618 Lampert # A New Solution of Dijkstra
0.14426861 2376 Habermann # Synchronization of Communi
0.14355143 2542 Graham # A Software Design and Eval
0.14126352 2541 Balzer # An Overview of the ISPL Co
0.14086196 2700 Lipton # Reduction: A Method of Pro
```

Part 3

I select the extension 9 using SVD to reduce the term vector dimensionality.

First, I transform the "dec_vector"(which is a list in the vector1.prl) to a matrix(A). The height is total number of documents and the wide is total number of terms in the corps and queries.

Then I use the SVD function in numpy to transform this big matrix to a small list(sigma). After reduce the 3 matrix and set the tolerance of the comparison function, I compare the matrix A and the reconstruction matrix and the return is true.