







(LSI).
SVD (LSI).
hatent Semantic indexing has two advantages
over the vector space worker.
Synonymy: Synonymy Refers to the case where two different words have the same meaning. Csay car and automobile)
) Synonymy. Synonymy have the same meaning.
two different words
Chay car and the
Polyenny reless to the case where term
has multiple meanings.
Summer and polysemy are handled in LSI.
Synonymy and polysemy are handled in LSI. but not in vector space model.
to the work of a state of the s
even for a collection of modest size, the
Even for a collection of modest size, the term document matrin C is likely to have
The Milander of The Milander
and a rank in tens of thousands as well.
The state of the s
In LSI, we use the SVD, to construct a lone
lank approximation by to the lemm - could the
mathen, for a value of 1 far smoother
original Rank of C.
Their each how / when is mapped to
K-dimensional space, this space is defined
to the largest eigen values) of cTC and CCT
The state of the s

Note that matrin cx is itself still an matrin Errespective of k.	gt
Note that matrin Cx is itself some con	+
more muchin orrespective of to	1
1 De la serie songe desire sons	1
A guery vector is mapped to its Representation	33
in the LSI space by the lansformation	
A query vector is mapped to its Representation in the LSI space by the transformation	
2x = 5x 4x 9	
PR R	
Now, we may use cosine similarities to	
compute the Similarity between a guery and	
a document, between two documents be	
between two terms manded has my many	
that water we vector applies and a model	
If a green is close to a document in the	
Supreme Supreme to Africa Color to the supremental	
an me no aumentional sound	
THE PERSON CANADA CANADA LA TAMBA LA TAMBBA LA TAMBA LA TAMBA LA TAMBA LA TAMBA LA TAMBA LA TAMBA LA TAMB	
document making	
ship 1 0 d3 d4 d5 d6	
ship 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
ocean 1 1 1 1 0 0 0 0	
voyage 1 0 0 0 0	
tip 0 0	•
to be a survey in all a second to	
the dismorphia eight manifeld correspondence	•
they the K truncipe of value of all budget	
	-

	9to SVD is ,
	U as below otherwood on with most
2.5	-0.44 -0.30 +0.57 0.58 0.25
5'9	-0.13 -0.33 -0.5.5 0.00 0.73
	-0.48 -0.5V -0.37 D -0.61
	0.10
	-0.26 0.65 -0.41 0.58 -0.09
	The state of the s
	E as below, Harris 65 (V) & status XI
-8.26	2.16.0-0.0-0.0-0
57.0	20001159-000-0040-0
	0 1/20
1000	O otte Reviewal Orality Oracle
	6. C. O. Marion Olita 80 metalla 0. 39
F	VT
	- 6.75 - 6.28 - 0.20 - 6.45 - 0.33 - 6.12
	-0.29 -0.53 -0.19 0.63 0.22 0.41
e) M	0.28 -0.75 0.45 -0.26 0.12 -0.33
eropela	0.00 0.00 0.58 0.00 -6.58 0.58
	-0.53 0.29 0.63 0.19 0.41 -0.22
17 ch	FOR K=12, 10 Westbrough offices Unity 1 10000
3/5	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
	Della a Orten Orte dimensional
	Business to volume of the space will
. 91)	20 facto that 10 and lable 0 data becomes A

From this we compute 62 From this we compute 62 -1.62 -0.60 -0.44 -0.97 -0.70 -0.26 -0.46 -0.84 -0.30 1.00 0.35 0.65 -0.46 -0.84 -0.30 1.00 0.0 0 0 0 0 0 0 0 Truncated (V) is -1.62 -0.60 -0.44 -0.97 -0.70 -0.26 -0.46 -0.84 -0.36 1.00 0.35 0.65 The relieval quality may improve by the dimensionality Refuse to and organizing data in high dimensional dimensional spaces (effen with hundred of thousands of dimensional settings such a not occur in love the problems is that when the dimensionality such in love the space incleases, the volume of the space incleases the volume of the space incleases to fast that the available data becomes space		
From this wo compute C2 -1.62 -0.60 -0.44 -0.97 -0.70 -0.26 -0.46 -0.84 -0.30 1.00 6.35 0.65 -0.46 -0.84 -0.30 1.00 6.35 0.65 -0.46 -0.84 -0.30 1.00 0.35 0.65 Truncated (V) is -1.62 -0.60 -0.44 -0.97 -0.76 826 -0.46 -0.84 -0.36 1.00 0.35 0.65 The Edieval quality may implay by the dimensionality Seduction. The Mse of dimensionality Various phenomena that arise when analyzing spaces (often with hundred of thousands of dimensional settings such a not occur in love the dimensional settings such and occur in love the dimensional settings such and occur in love		Bave de
Thuncated (V) is Thuncated (V) is -1.62 -0.60 -0.44 -0.30 1.00 6.35 6.65 -0.46 -0.84 -0.30 1.00 0.0 Thuncated (V) is -1.62 -0.60 -0.44 -0.97 -0.76 826 -0.46 -0.84 -0.36 1.00 0.35 0.65 The Kelieval grality may improve by the dimensionality Reduction. The Mse of dimensionality Reduction. The wise of dimensionality Refers to and organizing data in high dimensional spaces (often toith hundled of thousands of dimensional settings, such occur in lone The deblems in the common theme at these		to cond to U
Thuncated (V) is Thuncated (V) is -1.62 -0.60 -0.44 -0.30 1.00 6.35 6.65 -0.46 -0.84 -0.30 1.00 0.0 Thuncated (V) is -1.62 -0.60 -0.44 -0.97 -0.76 826 -0.46 -0.84 -0.36 1.00 0.35 0.65 The Kelieval grality may improve by the dimensionality Reduction. The Mse of dimensionality Reduction. The wise of dimensionality Refers to and organizing data in high dimensional spaces (often toith hundled of thousands of dimensional settings, such occur in lone The deblems in the common theme at these		From this wa compute 2.
Truncated (V) is Truncated (V) is -162 -066 -044 -0 97 -0.76 -0.26 -162 -066 -0.44 -0 97 -0.76 -0.26 -162 -066 -0.44 -0.35 0.65 The Elieval quality may improve by the dimensionality Reduction. The wase of dimensionality Refers to and organizing data in high dimensional spaces (often touth hundred of thousands of dimensional Settings, such them and states them.		111 -0.97 -0.70 -0.20
Thuncated (V) is Thuncated (V) is -1.62 -0.60 -0.44 -0.97 -0.76 824 -0.46 -0.84 -0.36 1.00 0.35 0.65 The Kelieval quality may improve by the dimensionality Reduction. The cuse of dimensionality Refueltan. Various phenomena that arise when analysis spaces (often with hundred ex thousands of dimensional settings such. The common theme at these there are there are the set of dimensional settings. Such.		-1.62 -0.60 -0.35 0.65
Truncated (V) is -162 -060 -044 -0 97 -0.76 -0.26 -046 -0.84 -0.36 1.00 0.35 065 The Kelieval quality may improve by the dimensionality Reduction. The cuse of dimensionality Various phenomena that arise when analyzing spaces (often with hundred of thousands of dimensional settings such and occur in love high leaves is the common theme at these		Er. 0-0460 -0.89
Truncated (V) is -1.62 -0.60 -0.44 -0.97 -0.70 -0.26 -0.46 -0.84 -0.36 1.00 0.35 0.65 The Kelieval quality may improve by the dimensionality Reduction. The wase of dimensionality Refers to and organizing data in high dimensional dimensions) that do not occur in love highlens is the common theme at these		1300 0 0 0 0 0 0 0
Thuncated (V) is -1.62 -0.60 -0.44 -0.97 -0.76 -0.26 -0.46 -0.84 -0.36 1.00 0.35 0.65 The Ketrieval quality may improve by the dimensionality Reduction. The curse of dimensionality Sarious phenomena that arise when analysis spaces (often with hundred of thousands of dimensional settings. Auch. Applement is the common theme at these		A DATE OF THE PROPERTY OF THE
The Lettieval quality may improve by the dimensionality Reduction. The use of dimensionality Various phenomena that arise when analysis spaces (often both hundred of thousands of dimensional settings. Such occur in love The common theme at these of dimensional of dimensional settings.		629-0351 0
The Lettieval quality may improve by the dimensionality Reduction. The use of dimensionality Various phenomena that arise when analysis spaces (often both hundred of thousands of dimensional settings. Such occur in love The common theme at these of dimensional of dimensional settings.		
The Keltieval quality may improve by the dimensionality Reduction. The cuse of dimensionality Various phenomena that arise when analysis spaces (often with hundred of thousands of dimensional settings. Such. The common theme at these		Isuncated (V) is
The retrieval quality may improve by the dimensionality Reduction. The cuse of dimensionality Various phenomena that arise when analysis spaces (often with hundred of thousands of dimensional settings. Such. The common theme at these		100 100 100 -0 97 - 76 97
The Kelieval quality may improve by the dimensionality Reduction. The wise of dimensionality Refers to various phenomena that arise when analysis spaces (often with hundreds of thousands of dimensions) that do not occur in love they lead to the common theme at these		
The Kelieval quality may improve by the dimensionality Reduction. The times of dimensionality Refers to various phenomena that arise when analysis spaces (often with hundred of thousands of dimensions) that do not occur in love they common theme is there		
The curse of dimensionality Refers to various phenomena that arise when analysing spaces (often with hundred of thousands of dimensional settings. Inch. The common theme of these		The solvieur gradity many more
The wese of dimensionality Refers to various phenomena that arise when analysing spaces (often with hundred of thousands of dimensional settings. Inch. State of dimensional settings. Inch. Curse of dimensionality Refers to dimensional when analysing spaces (often with hundred of thousands of dimensional settings. Inch. The common theme at these		by the dimensionality sodies
The wise of dimensionality Refers to various phenomena that arise when analyzing and organizing data in high dimensional spaces (often loith hundred of thousands of dimensional settings such occur in love the dimensional settings such theme of these		January Manuelan.
The wise of dimensionality Refers to various phenomena that arise when analyzing and organizing data in high dimensional spaces (often with hundred of thousands of dimensions) that do not occur in love the dimensional settings such.		The state of the s
The wise of dimensionality Refers to various phenomena that arise when analyzing and organizing data in high dimensional spaces (often loith hundred of thousands of dimensions) that do not occur in love the dimensional settings such.	213-	Curse of dimensionality
spaces (often loith hundreds of thousands of dimensional settings. Such. Shallens is that do not occur in love The common theme of these	1 #1	250 890 610- (23)
spaces (often with hundreds of thousands of dimensional settings. Such. Shoplems is that do not occur in love The common theme of these	25.8	The wise of dimension of the calcuto
dimensions) that do not occur in love highlenss is that common theme is there		Various phenomena that ariso soll a hour
dimensions) that do not occur in love the common theme is there		and organizing data in high discourses
hoplems is that common theme of these		spaces (often both hundred of thousands of
headlens is that common theme of there		dimensions) that do not occus.
problems is that when the dimensionality increases, the volume of the space mixeaser so fast that the available data second source		sellings. Such. in 100
incleases the volume of the space wiceases data second source		highlens is that common themo of store
So fast that the available the space mixeases		incleased the volume the dimensionality
data secon of source.		to fast that the available of the space will all
		data secon of space.

This spassity is problematic for any method that requires statistical significance. Organizing and seatching data often selves on detecting areas where objects form groups with similar properties, in high dimensional data however objects appear to be sparse and diminisher in many ways which prevents common data organization strategies from being efficient. Reduction. (Description spangered 1) mas First relations between teams (synonymu, and linger a amely of termine themelate it into the Odinerisional Gode and find matchin decements (intermation stries Find the best similarly between hand single of character a surranted was lie was trated of Knowledge corpus) eq. Vous a MCC million weeze question amovering I

PV.	Application of LSI.
Q.e	Lastrata transfer was to come to
66,83	The new low dimensional space can be typically used to:
barro	Typically used to :
linos	C to the in long dimensional space
Mark Count	(data (lustesing document classification).
N CON	Compare the documents in low dimensional space (data Clustering, document classification).
2.	Find similar documents across languages, after
المدلق	analyzing a base set of translated documents
V	Find similar documents across languages, after analyzing a base set of translated documents (cross language selieval).
3.	Find relations between terms (synonymy, and
	Find relations between terms (synonymy, and polysemy).
4.	gwen a guely of terms, translate it into low
	Juven a quely of terms, translate it into low dimensional space and find matching documents (information Setrieval).
	Joseph American Ameri
5.	Find the best similarity between small groups of
	terms in a servantic way (ie in a context
	terms in a servantic way (ie in a Context of Knowledge corpus). eg: in a MCB multiple choice questions answering model.
	answering o model.