www.menti.com

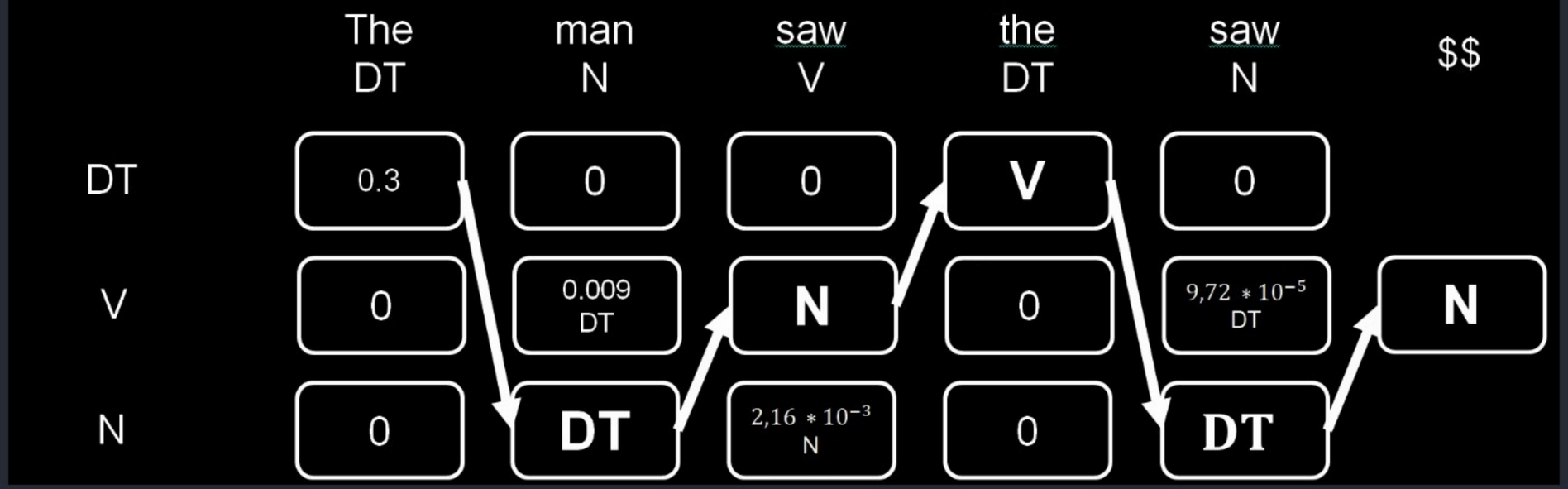
Questions?

0 questions
0 upvotes

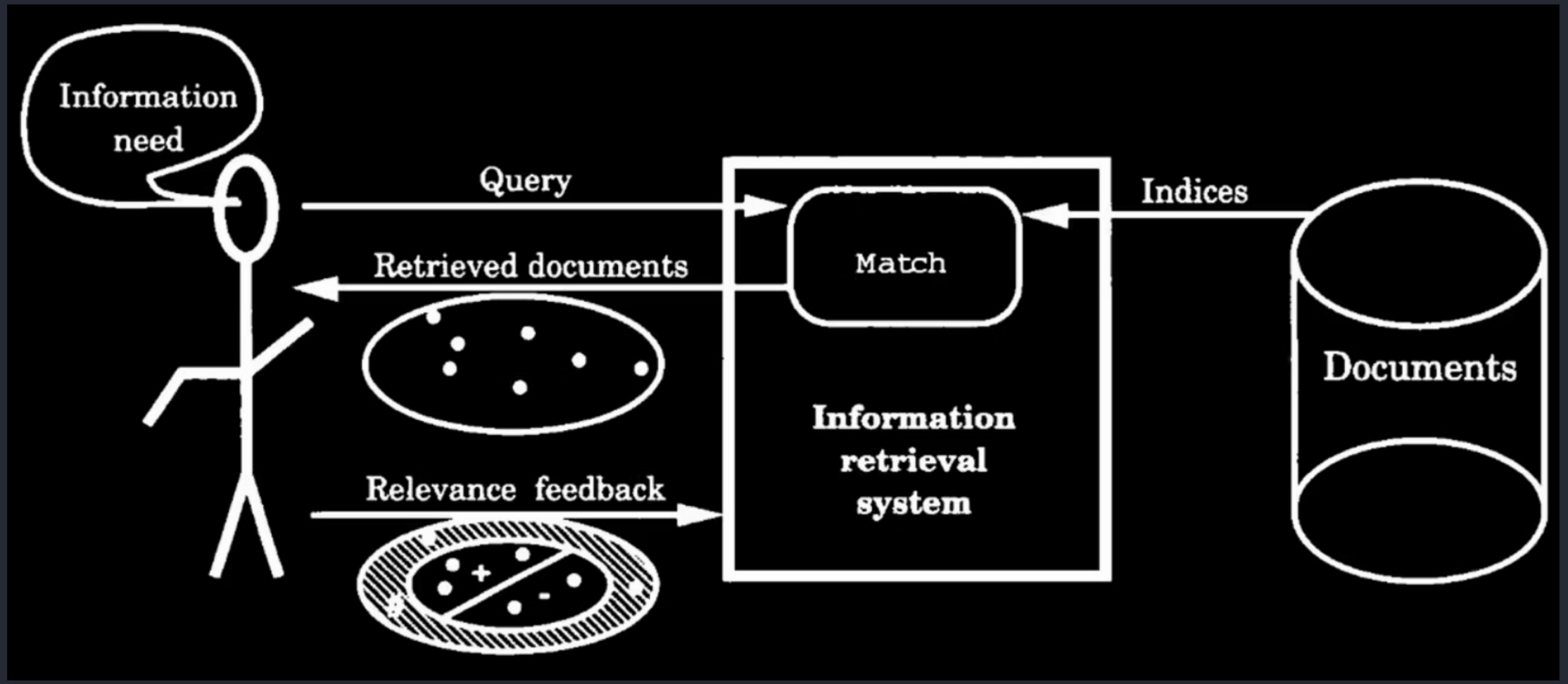
Complexity of the Viterbi algorithm: $O(ms^2)$,

- where m is the length of the input and s is the number of states in the model.

For every token (m) we we have to evaluate every POS (s) in combination with every possible predeccessor POS (s): m * s * s operations = ms^2



Viterbi



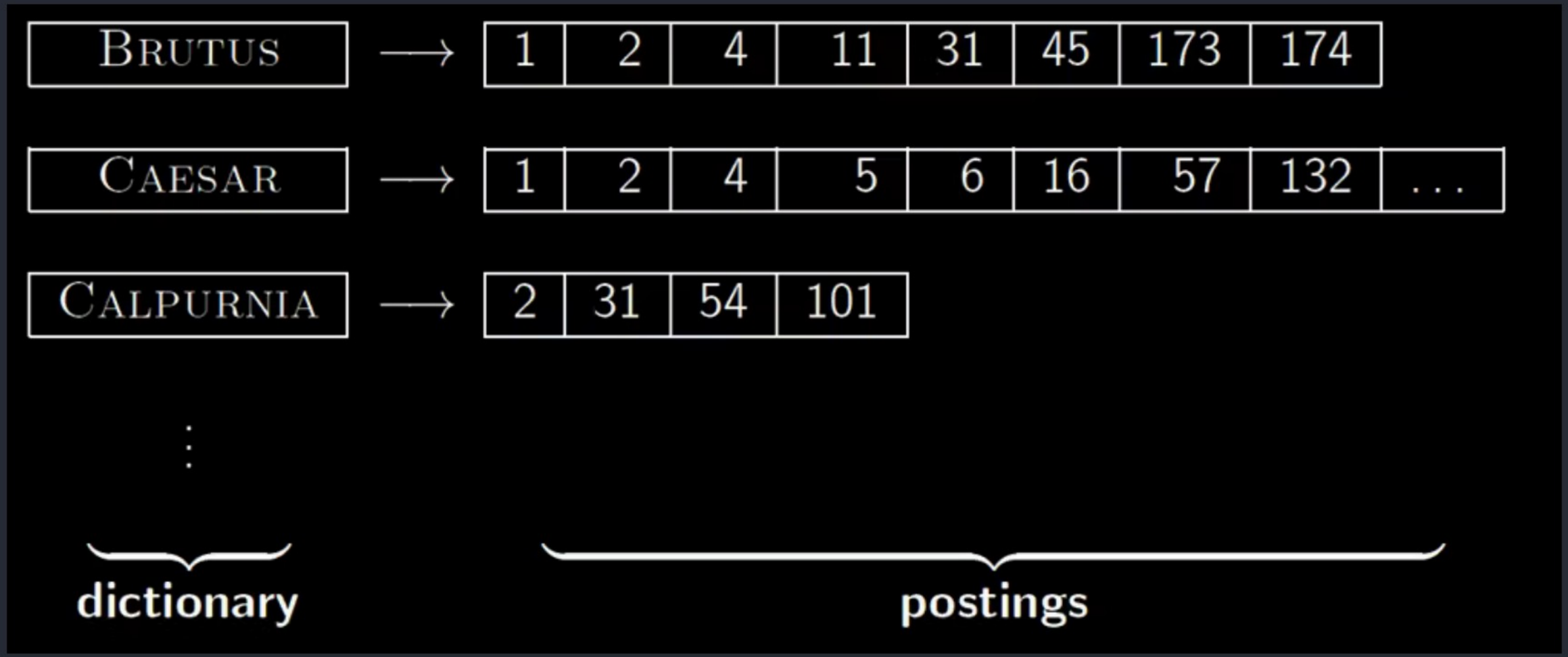
Information Retrieval

`Brutus AND Caesar` and `not Calpurnia`:

- **Complement** the vector of `Calpurnia`
- Do a (bitwise) AND on the three vectors
- 110100 and 110111 and 101111 = 100100

	Anthony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth	...
ANTHONY	1	1	0	0	0	1	
BRUTUS	1	1	0	1	0	0	
CAESAR	1	1	0	1	1	1	
CALPURNIA	0	1	0	0	0	0	
CLEOPATRA	1	0	0	0	0	0	
MERCY	1	0	1	1	1	1	
WORSER	1	0	1	1	1	0	
...							
result:	1	0	0	1	0	0	

Boolean Retrieval



Inverted Index

- The $\mathrm{tf}-\mathrm{idf}$ weight of a term is the product of its tf weight and its idf weight

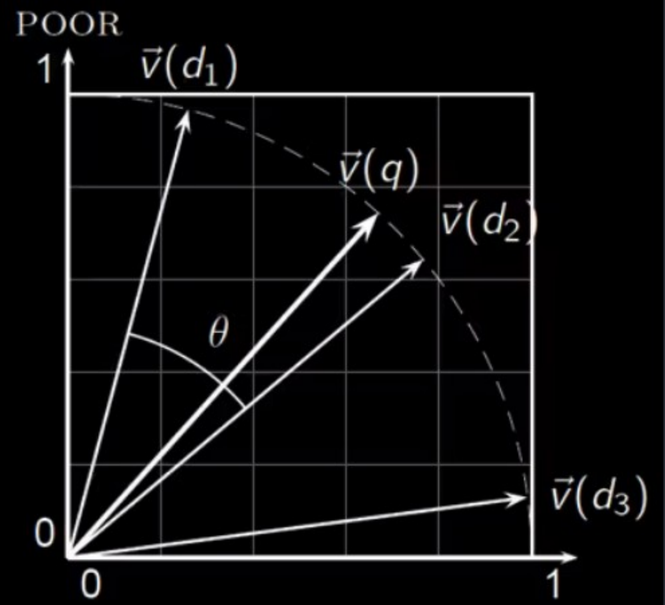
$$w_{t,d} = \left(1 + \log tf_{t,d}\right).\log\frac{N}{df_t}$$

tf-weight, idf-weight

IF-IDF Weightning

$$\cos(\vec{q}, \vec{d}) = \text{SIM}(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}||\vec{d}|} = \frac{\sum_{i=1}^{|V|} q_i d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \sqrt{\sum_{i=1}^{|V|} d_i^2}}$$

- q_i is the tf-idf weight of term i in the query
- d_i is the tf-idf weight of term i in the document
- $|\vec{q}|$ and $|\vec{d}|$ are the lengths of $\vec{q}$ and $\vec{d}$
- Cosine similarity of $\vec{q}$ and $\vec{d}$ = the cosine of the angle between $\vec{q}$ and $\vec{d}$



Cosine Similarity

Questions?

0 questions
0 upvotes

Mentimeter

Please enter the quiz on www.menti.com

Question 1

You are the DJ "Lost Term Frequencies", and you run the music in one of the coolest NLP bars in Darmstadt. It is saturday night, the crowd is going wild, and one of the guests comes to you with a strange song request: "Play the song about Conditional Probability!"

Which song do you throw in the mix?

[Which of these songs contains an algorithm that works primarily on conditional probabilities?]

Mentimeter

Which song is about Conditional Probability?

Leaderboard

No results yet

Top Quiz participants will be displayed here once there are results!

Mentimeter

Question 2

Humanity has just recovered from the attack of the evil LEMMATIZER - but already the Earth is threatened again. A monstrous mutation, a terrible terror, a vile villain: the PREPROCESSOR!

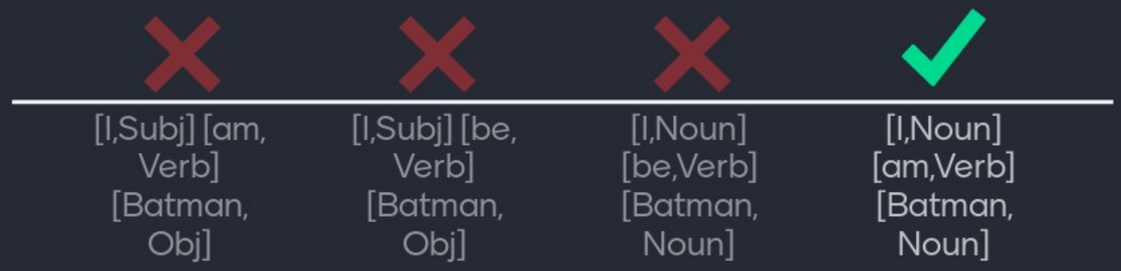
Immediately Batman is on the scene, shouting "I am Batman" at the terrible henchman.

However, the PREPROCESSOR is not impressed by this call, and applies an impressive combination of tokenisation, stemming and POS tagging! How does the PREPROCESSOR distort Batman's catch phrase?

(What is the result of "I am Batman" after tokenisation, stemming and POS tagging?)

Mentimeter

What is the result of "I am Batman" after tokenisation, stemming and POS tagging?



Leaderboard

No results yet

Top Quiz participants will be displayed here once there are results!

Question 3

In the forbidden library, young Harry Potter finds a new spell: "Boolean Retriviosa". As expected, the underage wannabe magician does not think twice and tries out the spell immediately.

With a loud "Bool-a-bang!", a black and white ghost appears in front of the sorcerer.

"I am Bool'ette, the Boolean ghost from faraway Bool'garia. Prove to me that you are powerful of Boolean expressions, or I will turn you into a Bool'dozer!

Immediately, a series of binary vectors appear to Harry, which must be linked bit by bit with Boolean operations.

What is the result? (Zeros and ones without spaces)

Mentimeter

$(1010 \wedge \neg 1100) \vee 1000$

The correct answer is: 1010

Leaderboard

No results yet

Top Quiz participants will be displayed here once there are results!

Question 4

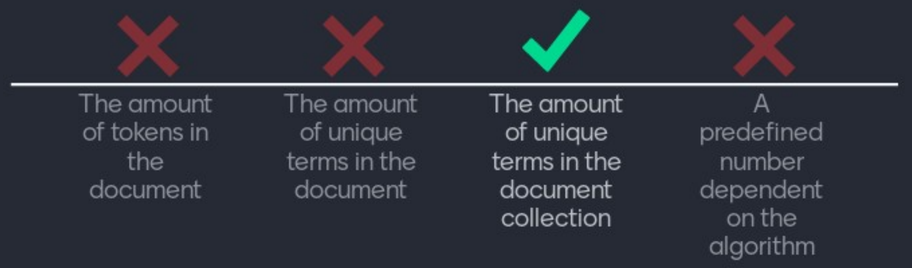
The hyper-dimensional cinema is back! It has as many dimensions as every movie needs! Today we will see another awesome Sci-Fi blockbuster:

"The Document Vector of the Vector Space Model"

What determines how many dimensions this movie needs?

Mentimeter

What determines the dimensions of a document vector in the Vector Space Model?



Leaderboard

No results yet

Top Quiz participants will be displayed here once there are results!

Mentimeter

Any questions?

0 questions
0 upvotes

Some discussion...?

Mentimeter

Any questions?

0 questions
0 upvotes