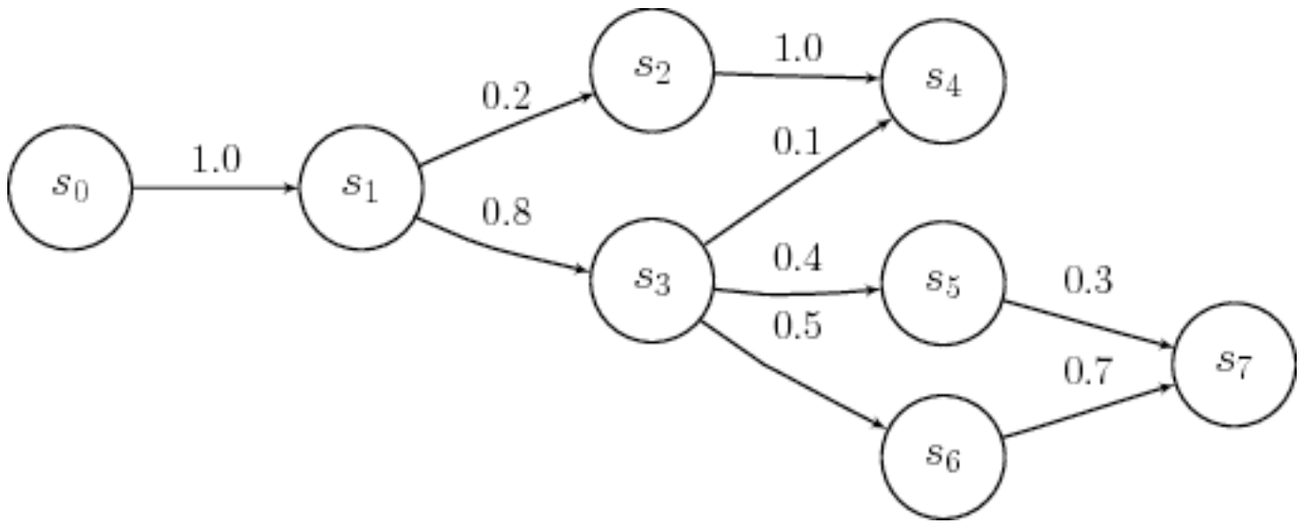
Finding the Essence with

Natural Language Processing

Me

Why?

Markov Chains


s_0
s_1
s_2
s_3
s_4
s_5
s_6
s_7
1.0
0.2
0.8
1.0
0.1
0.4
0.5
0.3
0.7

Bookshrink

Find the Essence

Linear Algebra

Two Assumptions

- **Every sentence can be represented by a set of words** (order doesn't matter)

- Similar words show up near by each other in natural text

Document-Term Matrix

	I	like	hate	databases
D1	1	1	0	1
D2	1	0	1	1

"I like databases" = [1, 1, 0, 1]
"I hate databases" = [1, 0, 1, 1]

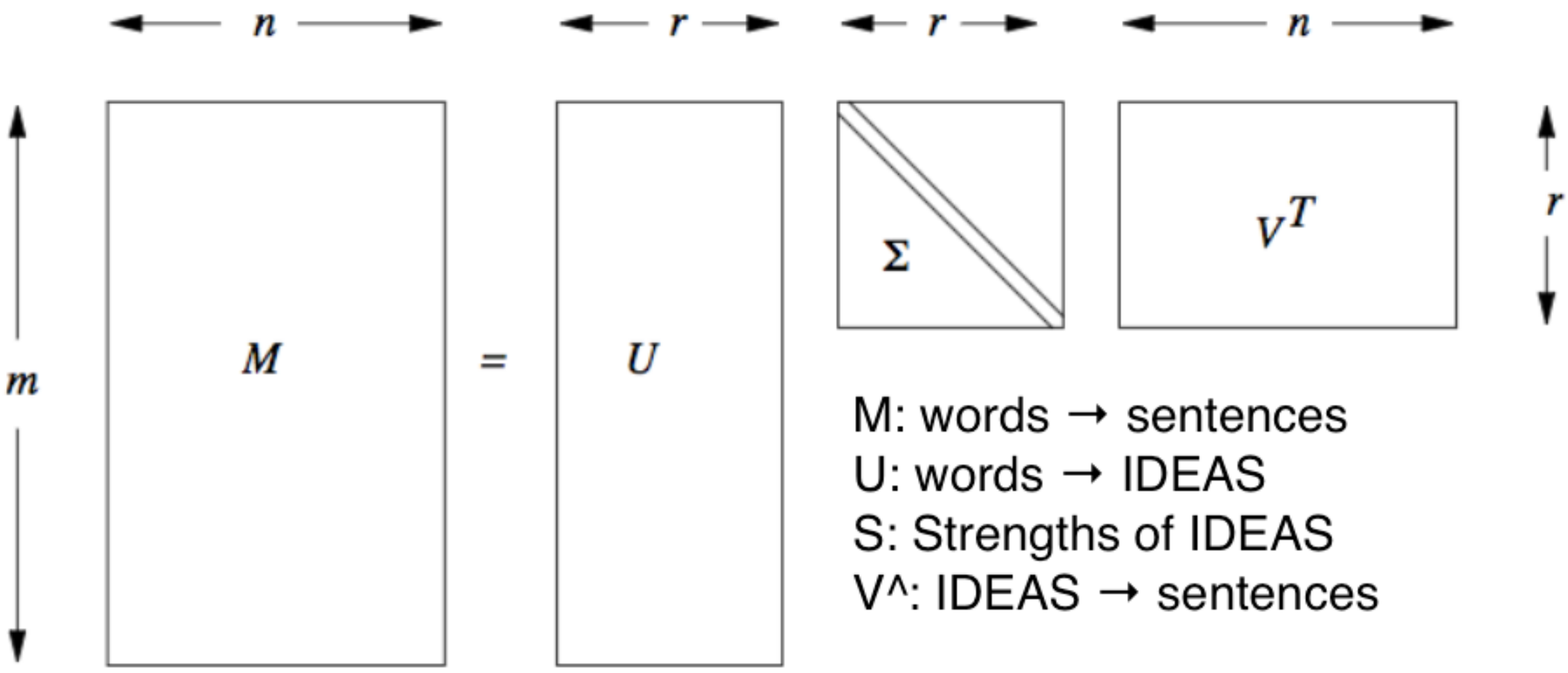


Figure 11.5: The form of a singular-value decomposition

https://github.com/peterldowns/essence-talk