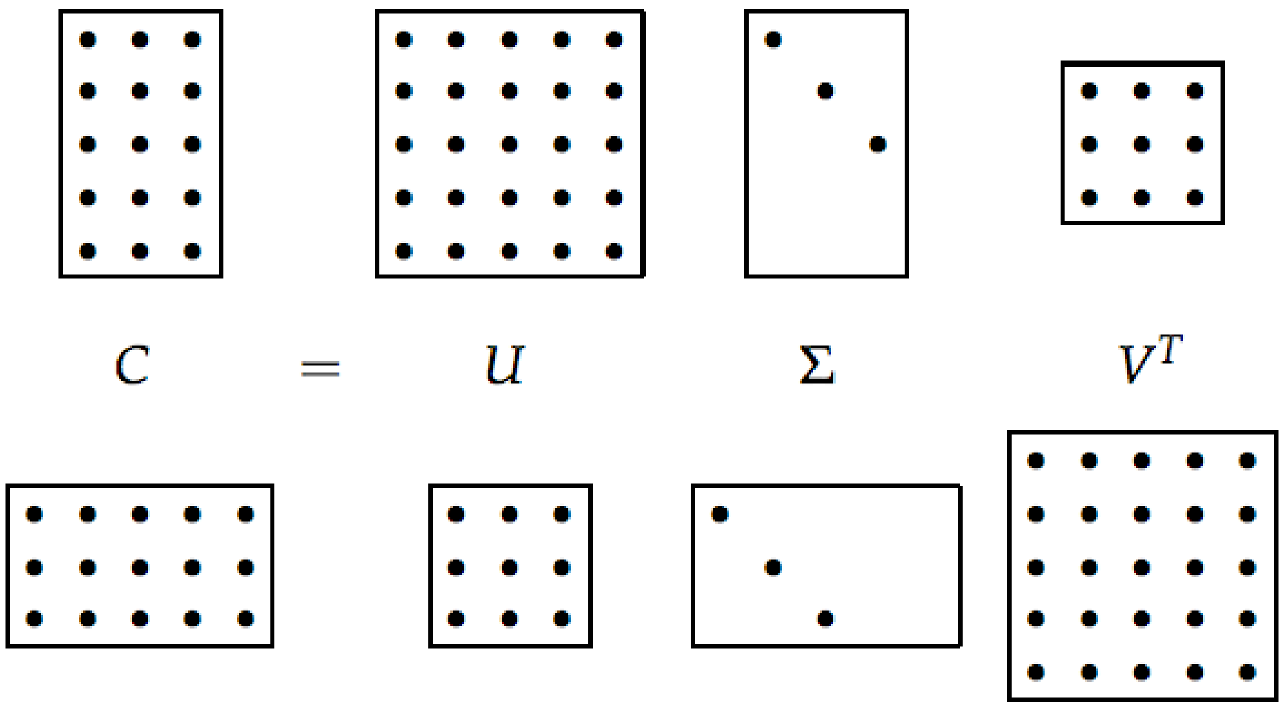
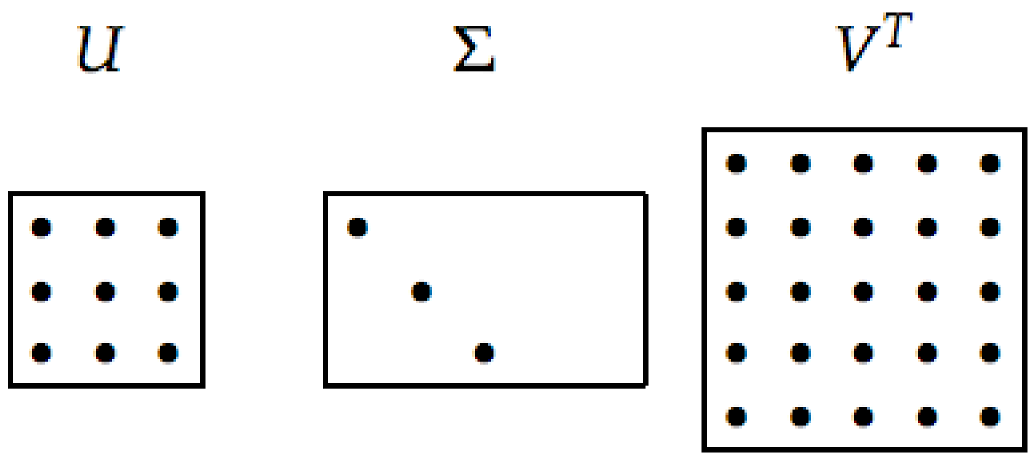
Matrix Decomposition

- Singular Value Decomposition
  - Is a technique that splits a matrix into three components with these properties.



Matrix Decomposition

- Singular Value Decomposition
  - SVD reduces the size of your term-document matrix
    - reduces the dimensionality or the rank
    - you can arbitrarily reduce the dimensionality by putting zeros in the bottom right of sigma
    - this is the mathematically optimal way of reducing dimensions



Matrix Decomposition

- Singular Value Decomposition
  - If the old dimensions where based on terms
    - after reducing the rank of the matrix the dimensionality is based on concepts or semantics
    - a concept is a linear combination of terms

$$SVD_{dimension_1} = a * td_{dim_1} + b * td_{dim_2} + c * td_{dim_3} + d * td_{dim_4}$$

$$SVD_{dimension_2} = a' * td_{dim_1} + b' * td_{dim_2} + c' * td_{dim_3} + d' * td_{dim_4}$$

$$SVD_{dimension_3} = a'' * td_{dim_1} + b'' * td_{dim_2} + c'' * td_{dim_3} + d'' * td_{dim_4}$$

Matrix Decomposition

- Singular Value Decomposition

$$SVD_{dimension_1} = a * td_{dim_1} + b * td_{dim_2} + c * td_{dim_3} + d * td_{dim_4}$$

$$SVD_{dimension_2} = a' * td_{dim_1} + b' * td_{dim_2} + c' * td_{dim_3} + d' * td_{dim_4}$$

$$SVD_{dimension_3} = a'' * td_{dim_1} + b'' * td_{dim_2} + c'' * td_{dim_3} + d'' * td_{dim_4}$$

$$\begin{vmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{vmatrix}$$

Matrix Decomposition

- Singular Value Decomposition

$$SVD_{dimension_1} = a * td_{dim_1} + b * td_{dim_2} + c * td_{dim_3} + d * td_{dim_4}$$

$$SVD_{dimension_2} = a' * td_{dim_1} + b' * td_{dim_2} + c' * td_{dim_3} + d' * td_{dim_4}$$

$$SVD_{dimension_3} = a'' * td_{dim_1} + b'' * td_{dim_2} + c'' * td_{dim_3} + d'' * td_{dim_4}$$

$$\begin{vmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{vmatrix} * \begin{vmatrix} td_{dim_1} \\ td_{dim_2} \\ td_{dim_3} \\ td_{dim_4} \end{vmatrix}$$

Matrix Decomposition

- Singular Value Decomposition

$$SVD_{dimension_1} = a * td_{dim_1} + b * td_{dim_2} + c * td_{dim_3} + d * td_{dim_4}$$

$$SVD_{dimension_2} = a' * td_{dim_1} + b' * td_{dim_2} + c' * td_{dim_3} + d' * td_{dim_4}$$

$$SVD_{dimension_3} = a'' * td_{dim_1} + b'' * td_{dim_2} + c'' * td_{dim_3} + d'' * td_{dim_4}$$

$$\begin{vmatrix} SVD_{dim_1} \\ SVD_{dim_2} \\ SVD_{dim_3} \end{vmatrix} = \begin{vmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{vmatrix} * \begin{vmatrix} td_{dim_1} \\ td_{dim_2} \\ td_{dim_3} \\ td_{dim_4} \end{vmatrix}$$

Matrix Decomposition

- Singular Value Decomposition

$$SVD_{dimension_1} = a * td_{dim_1} + b * td_{dim_2} + c * td_{dim_3} + d * td_{dim_4}$$

$$SVD_{dimension_2} = a' * td_{dim_1} + b' * td_{dim_2} + c' * td_{dim_3} + d' * td_{dim_4}$$

$$SVD_{dimension_3} = a'' * td_{dim_1} + b'' * td_{dim_2} + c'' * td_{dim_3} + d'' * td_{dim_4}$$

$$\begin{vmatrix} SVD_{dim_1} \\ SVD_{dim_2} \\ SVD_{dim_3} \end{vmatrix} = \begin{vmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{vmatrix} * \begin{vmatrix} td_{dim_1} \\ td_{dim_2} \\ td_{dim_3} \\ td_{dim_4} \end{vmatrix}$$

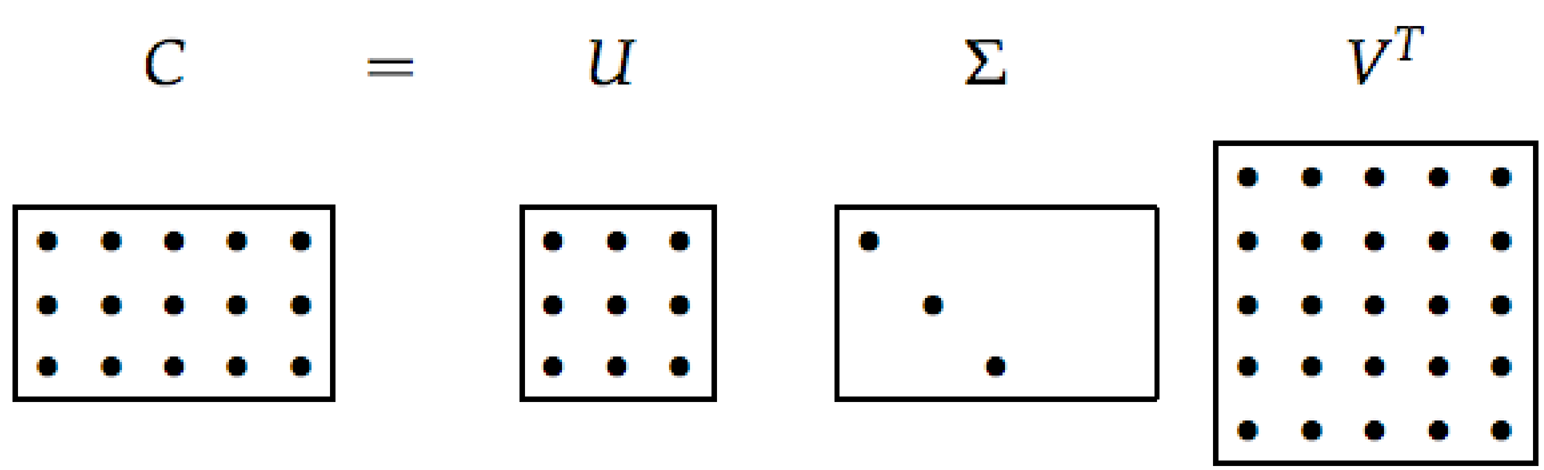
$$SVD_{ConceptSpace} = M * query_{TermSpace}$$

Matrix Decomposition

- Singular Value Decomposition

$$\begin{vmatrix} SVD_{dim_1} \\ SVD_{dim_2} \\ SVD_{dim_3} \end{vmatrix} = \begin{vmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{vmatrix} * \begin{vmatrix} td_{dim_1} \\ td_{dim_2} \\ td_{dim_3} \\ td_{dim_4} \end{vmatrix}$$

$$SVD_{ConceptSpace} = M * query_{TermSpace}$$



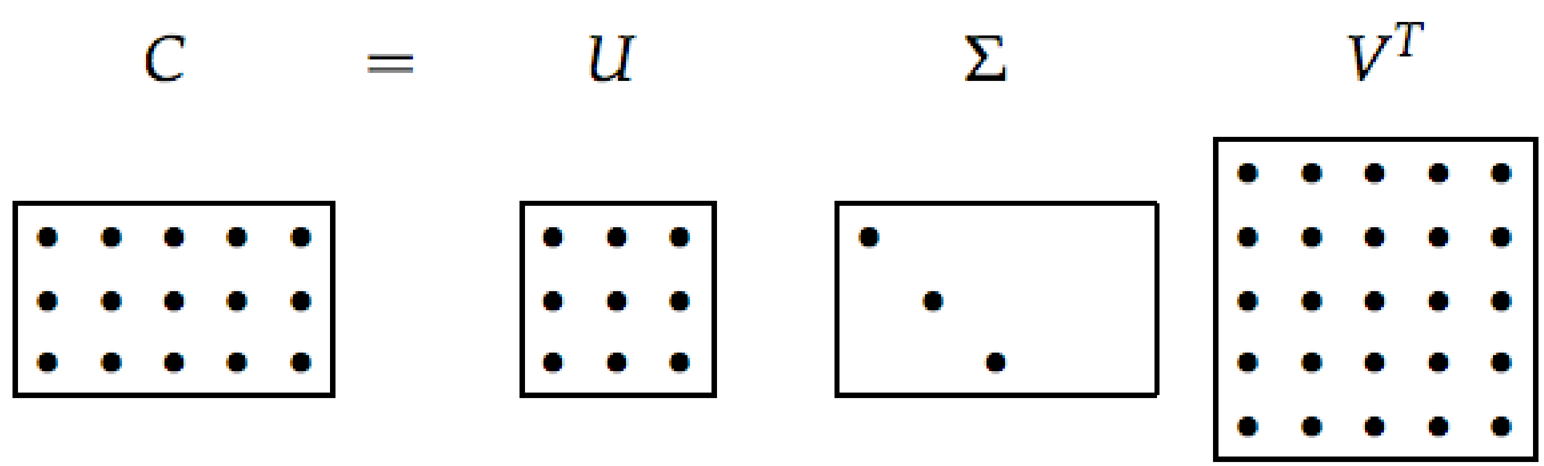
Matrix Decomposition

- Singular Value Decomposition

$$\begin{vmatrix} SVD_{dim_1} \\ SVD_{dim_2} \\ SVD_{dim_3} \end{vmatrix} = \begin{vmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{vmatrix} * \begin{vmatrix} td_{dim_1} \\ td_{dim_2} \\ td_{dim_3} \\ td_{dim_4} \end{vmatrix}$$

$$SVD_{ConceptSpace} = M * query_{TermSpace}$$

$$M = \Sigma_k^{-1} U_k^T$$



Matrix Decomposition

- Singular Value Decomposition

$$\left| \begin{matrix} SVD_{dim_1} \\ SVD_{dim_2} \\ SVD_{dim_3} \end{matrix} \right| = \left| \begin{matrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{matrix} \right| * \left| \begin{matrix} td_{dim_1} \\ td_{dim_2} \\ td_{dim_3} \\ td_{dim_4} \end{matrix} \right|$$

$$SVD_{ConceptSpace} = M * query_{TermSpace}$$

$$M = \Sigma_k^{-1} U_k^T$$

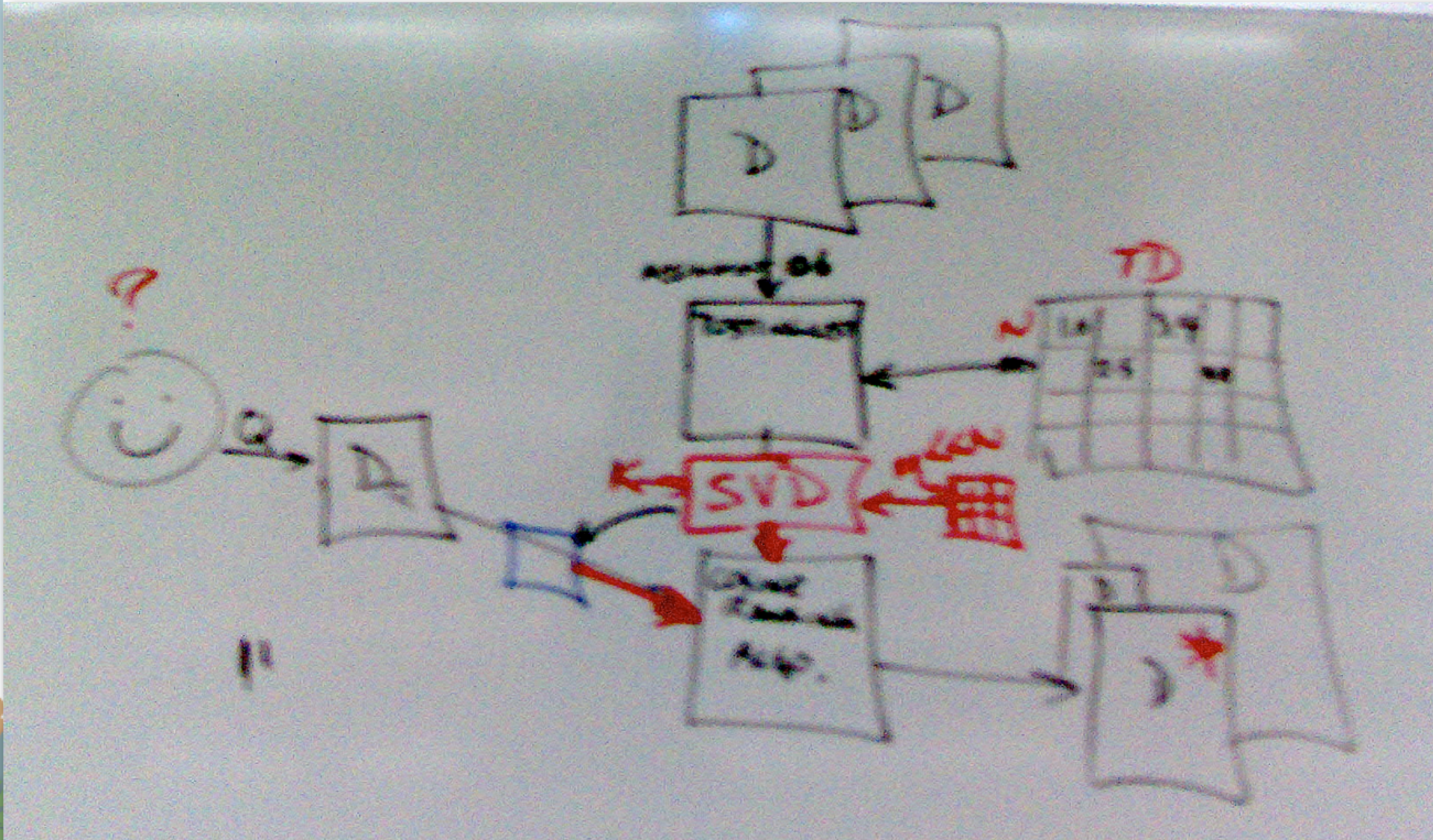
$$query_{ConceptSpace} = \Sigma_k^{-1} U_k^T query_{TermSpace}$$

Matrix Decomposition

- Singular Value Decomposition
  - SVD is an algorithm that gives us $\Sigma \; U \; V^T$
  - With these quantities we can reduce dimensionality
  - With reduced dimensionality
    - synonyms are mapped onto the same location
      - “bat” “chiroptera”
    - polysemies are mapped onto different locations
      - “bat” (baseball) vs. “bat” (small furry mammal)

Matrix Decomposition

- Big Picture



Matrix Decomposition

- “I am not crazy”
  - Netflix

Matrix Decomposition

- “I am not crazy”
  - Netflix
  - Machine translations
    - Just like “bat” and “chiroptera” map the same
    - “bat” and “murciélago” can map to the same thing

The math is hard but it's beautiful and powerful

La matemáticas es dura pero es hermosa y de gran alcance

Jene mathematisch ist hart, aber ist und an langer Reichweite schön

That one mathematically is hard, but is beautiful and at long range