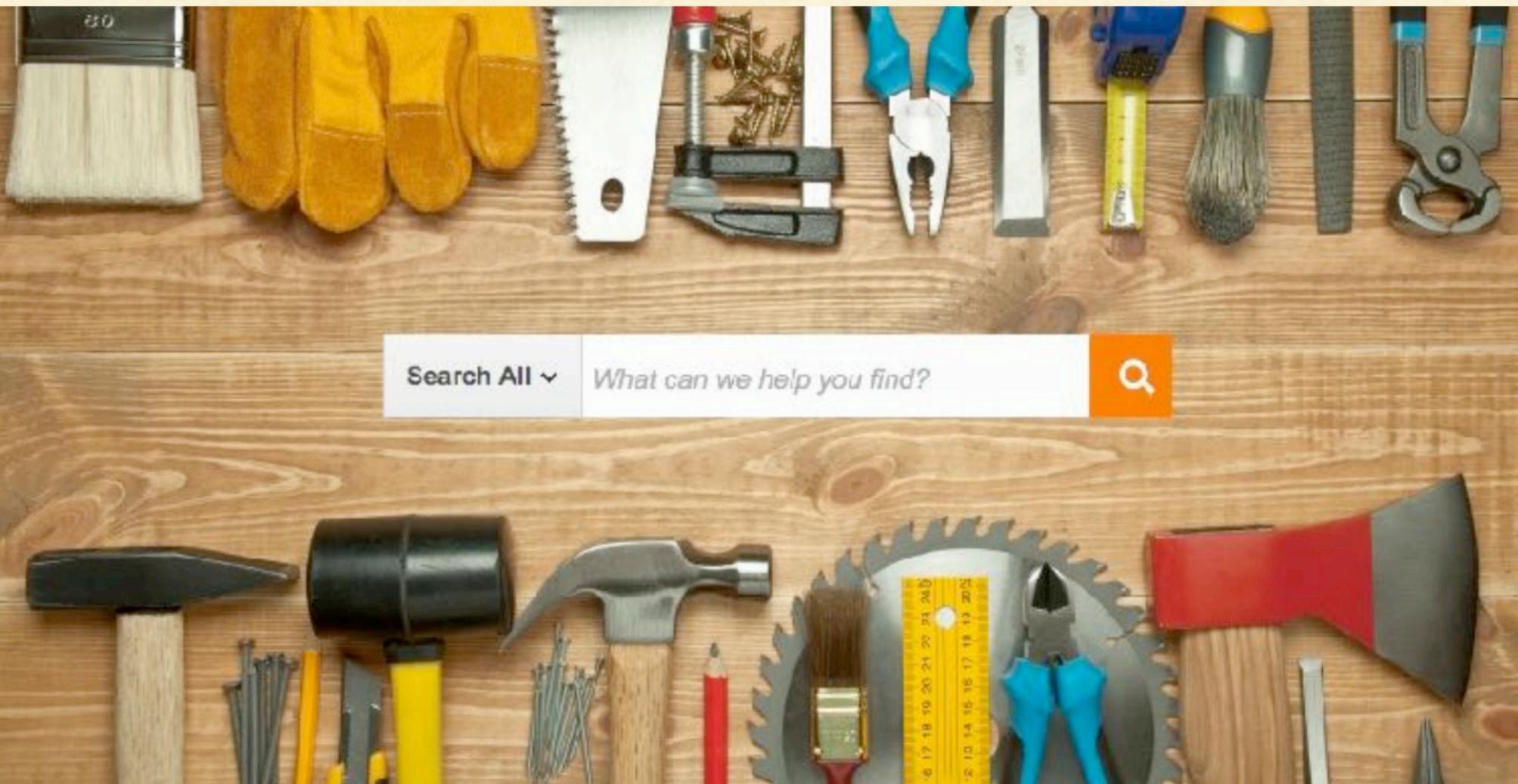

1082 WSM Project 1: Search and Rank via Vector Space Models

The Task

- 2,048 product descriptions



1. Calculate the **relevance** of each product by the given query.
2. Return the **top relevant** product IDs.

Processing Steps (1)

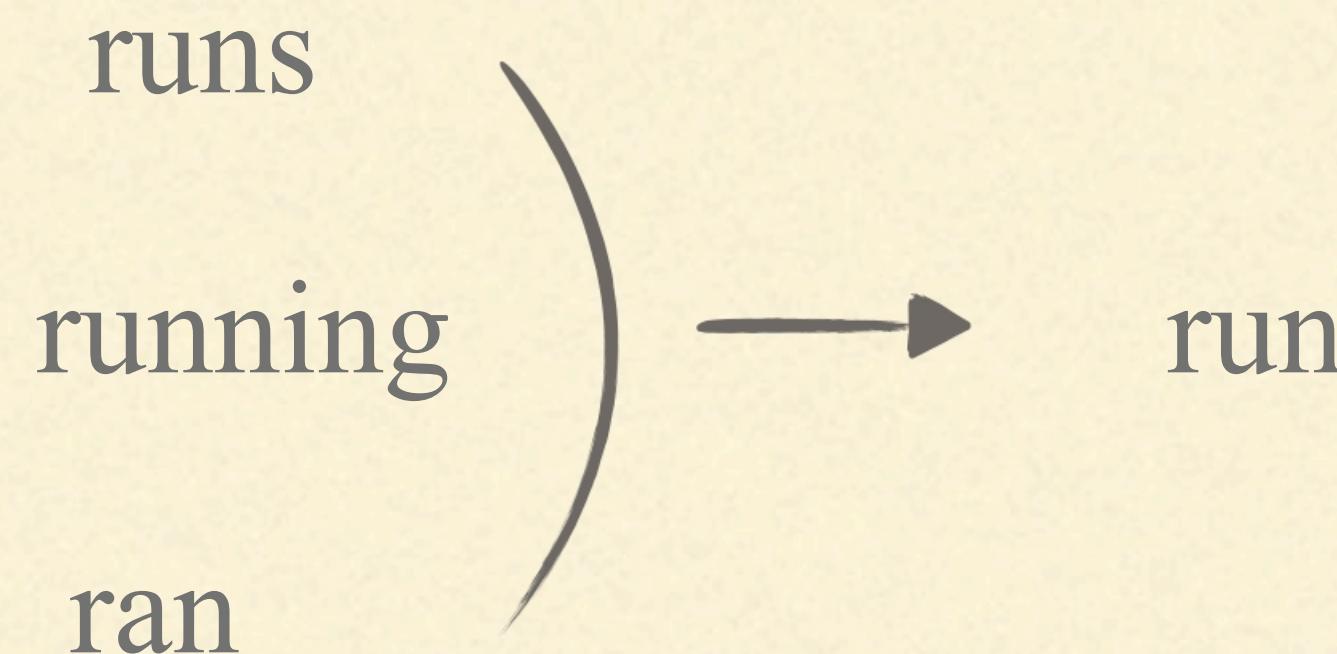
- (1). Stemming & Removing Stop Words & Indexing
- (2). Transfer Query into a Vector
- (3). Transfer Documents into Vectors
- (4). Calculate the similarity between the Query Vector and the Document Vectors
- (5). Rank the Documents according to the similarity scores

112394.product 124504.product 137320.product 148624.product
112442.product 124615.product 137404.product 148793.product
112457.product 124616.product 137438.product 148861.product
112487.product 124698.product 137543.product 148865.product
112514.product 124699.product 137625.product 149076.product
112551.product 124808.product 137631.product 149093.product
112553.product 124843.product 137636.product 149111.product
112578.product 124848.product 137705.product 149280.product
112601.product 124913.product 137754.product 149330.product
112647.product 124941.product 137776.product 149403.product
112674.product 124960.product 137941.product 149515.product

Whether for home, farm, builder or industrial customers, National has all the right hardware in the sizes and finishes you need. With over a century of service National is committed to maintaining the highest level of product quality, innovation and manufacturing technology. A top quality line of gate hardware designed for long life and maintenance free operation. Smooth Black finish with excellent corrosion resistance because they are triple-protected with zinc-plate, Black finish chromate dip and Black finish baked on enamel finish. Hot-rolled steel case and strike; steel rod barDesigned for out-swing gates, doors, and animal pensStrike can be mounted either on the gate or the postLatches automaticallyDesigned for right or left hand applicationsCan be padlocked for added securityHole punch in release lever is for rigging cord release from opposite sideMounting screws includedBlack finishLifeSpan limited lifetime warranty
124808.product (END)

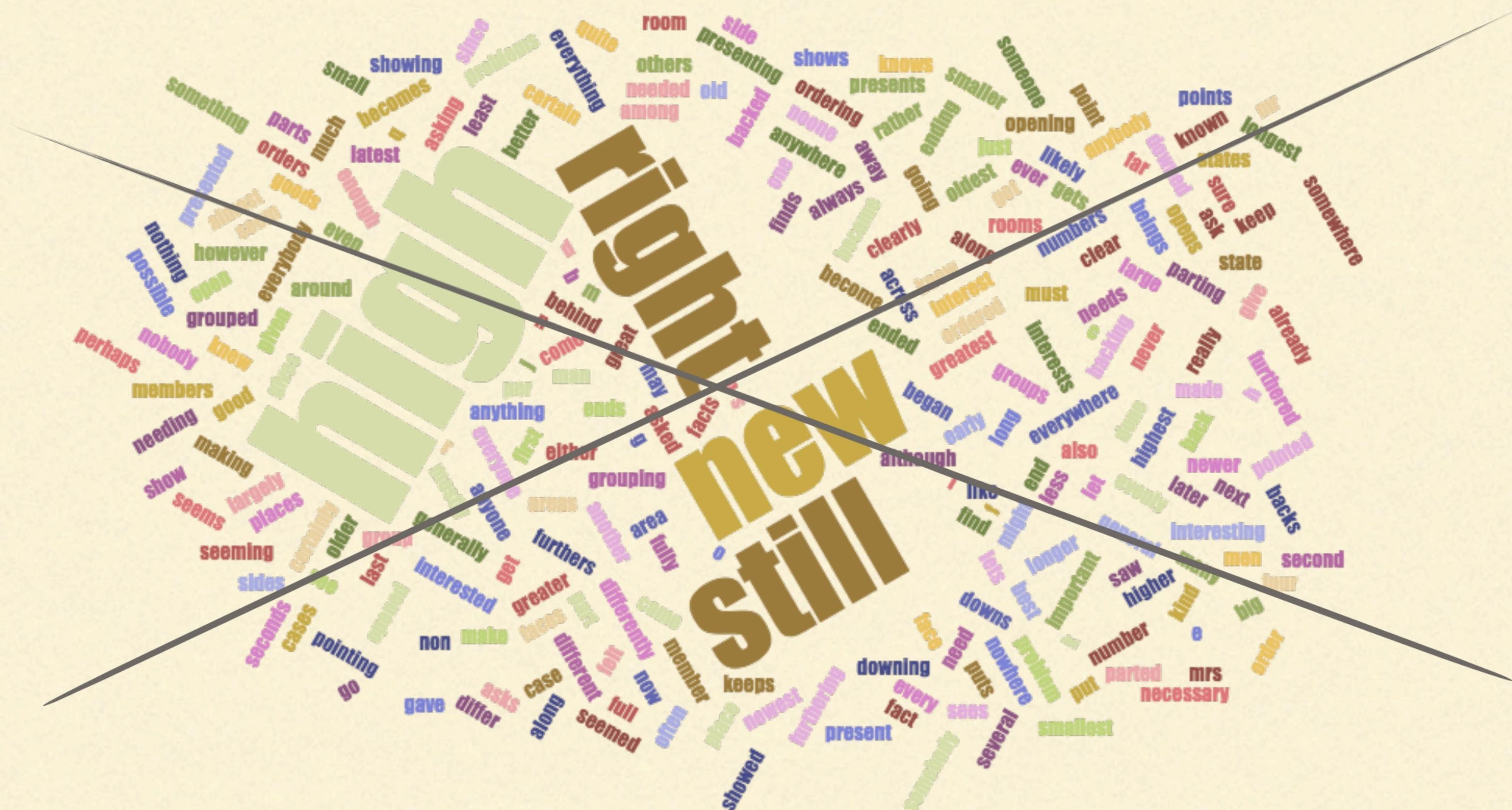
Processing Steps (2)

- (1). **Stemming** & Removing Stop Words & Indexing
- (2). Transfer Query into a Vector
- (3). Transfer Documents into Vectors
- (4). Calculate the similarity between the Query Vector and the Document Vectors
- (5). Rank the Documents according to the similarity scores



Processing Steps (3)

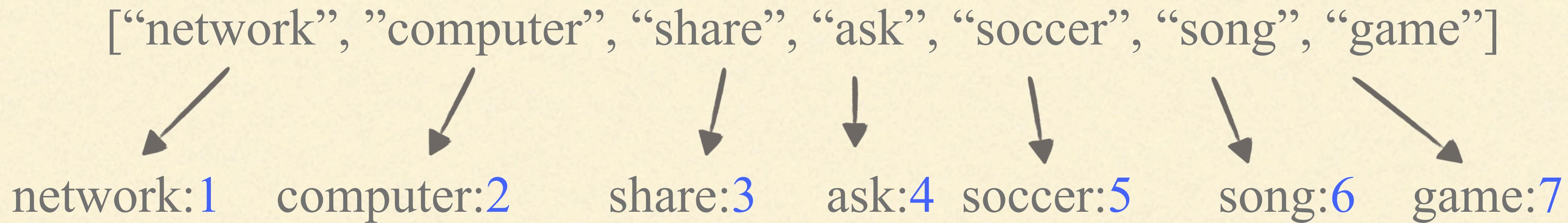
(1). Stemming & Removing Stop Words & Indexing



You can find stop words list from any place

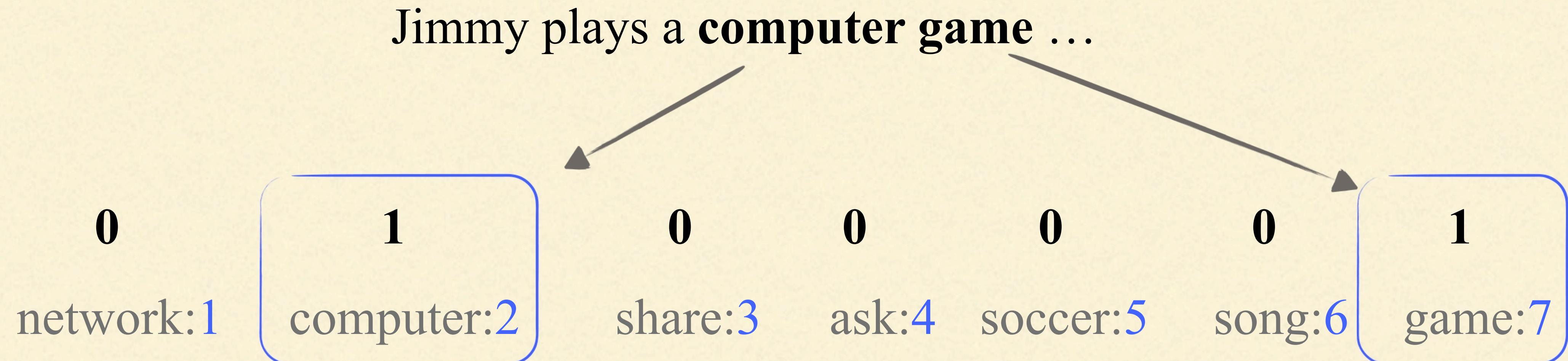
Processing Steps (4)

- (1). Stemming & Removing Stop Words & **Indexing**
- (2). Transfer Query into a Vector
- (3). Transfer Documents into Vectors
- (4). Calculate the similarity between the Query Vector and the Document Vectors
- (5). Rank the Documents according to the similarity scores



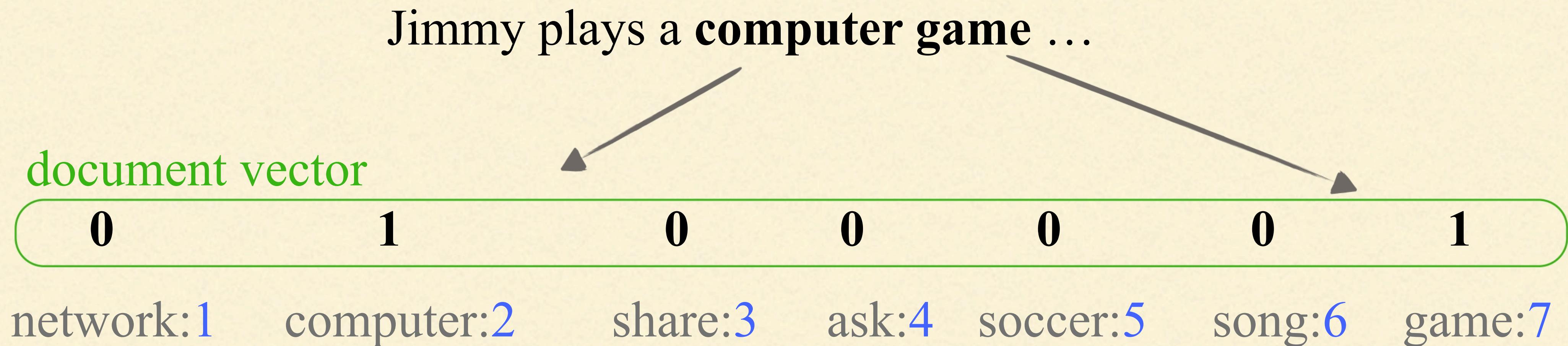
Processing Steps (5)

- (1). Stemming & Removing Stop Words & Indexing
- (2). Transfer Query into a Vector
- (3). Transfer Documents into Vectors
- (4). Calculate the similarity between the Query Vector and the Document Vectors
- (5). Rank the Documents according to the similarity scores

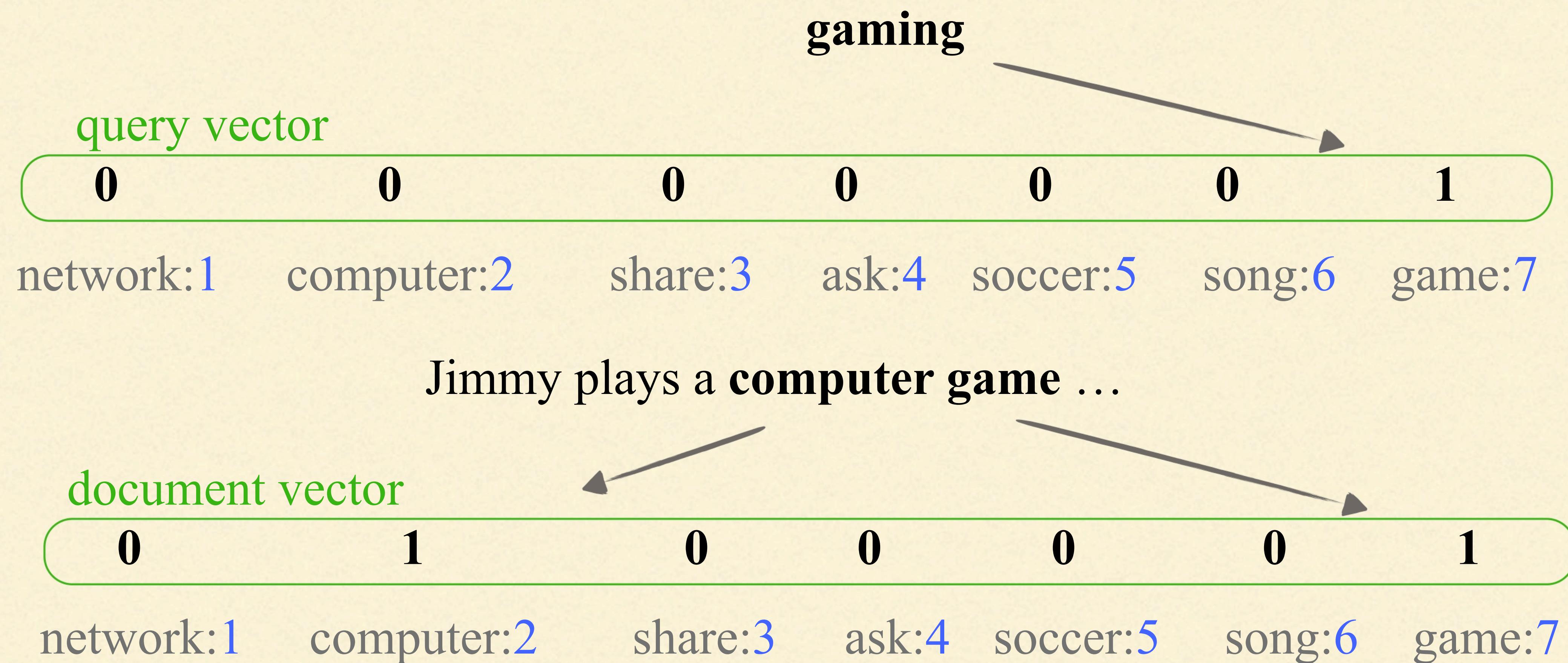


Processing Steps (5) cont'd

- (1). Stemming & Removing Stop Words & Indexing
- (2). Transfer Query into a Vector
- (3). Transfer Documents into Vectors
- (4). Calculate the similarity between the Query Vector and the Document Vectors
- (5). Rank the Documents according to the similarity scores



Processing Steps (5) cont'd

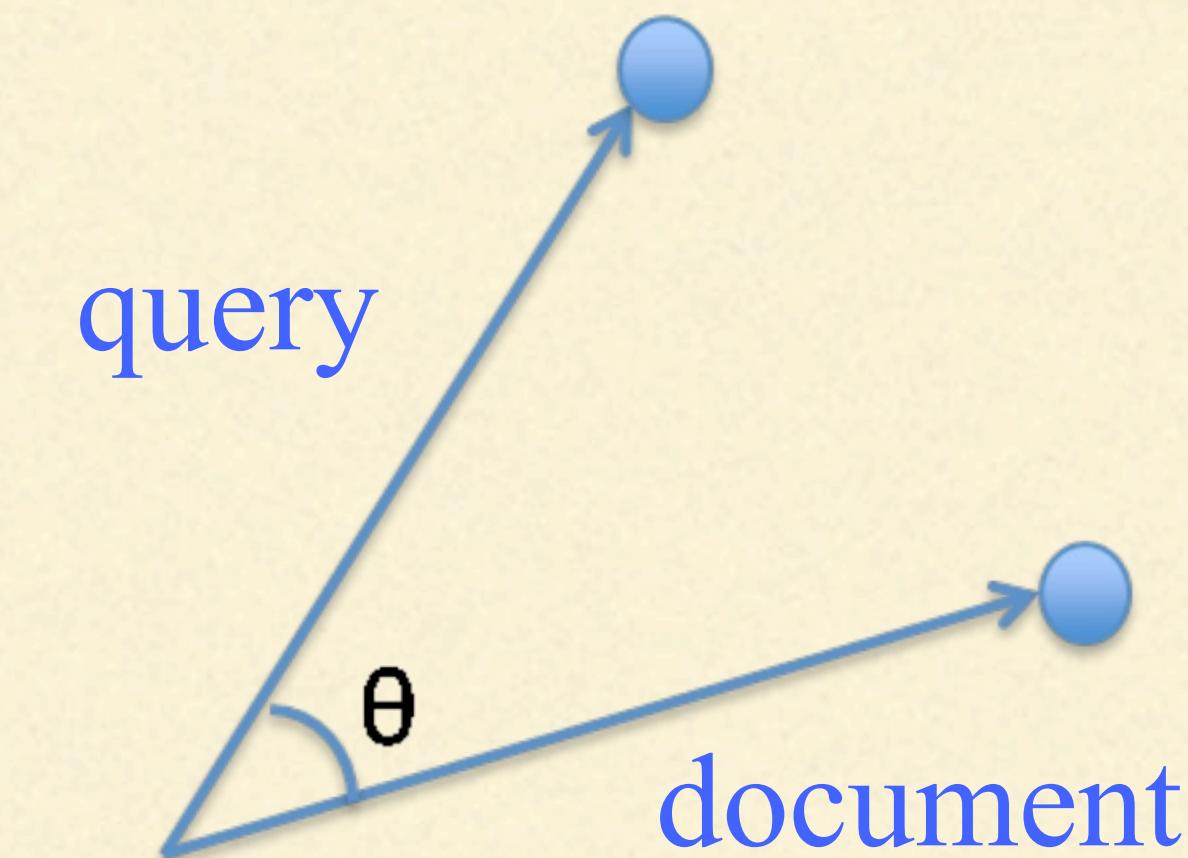


Processing Steps (6)

- (1). Stemming & Removing Stop Words & Indexing
- (2). Transfer Query into a Vector
- (3). Transfer Documents into Vectors
- (4). Calculate the similarity between the Query Vector and the Document Vectors
- (5). Rank the Documents according to the similarity scores

query vector document vector

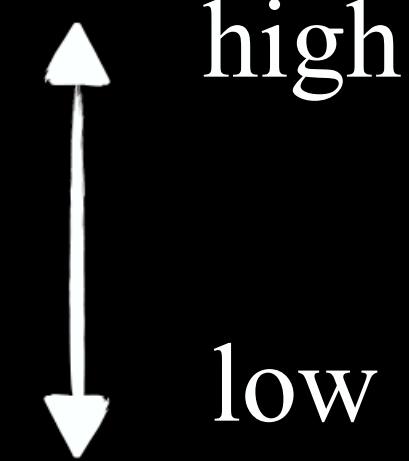
$$\text{sim}(\underline{A}, \underline{B}) = \cos(\theta) = \frac{\underline{A} \cdot \underline{B}}{\|\underline{A}\| \|\underline{B}\|}$$



Processing Steps (7)

- (1). Stemming & Removing Stop Words & Indexing
- (2). Transfer Query into a Vector
- (3). Transfer Documents into Vectors
- (4). Calculate the similarity between the Query Vector and the Document Vectors
- (5). Rank the Documents according to the similarity scores

TF-IDF Weighting + Cosine Similarity :	
DocID	Score
932	0.700321
248	0.383555
38	0.277447
234	0.234500
569	0.221815



The diagram shows a vertical double-headed arrow pointing upwards on the left and downwards on the right, positioned next to the table. The word "high" is placed above the upward-pointing arrow, and the word "low" is placed below the downward-pointing arrow, indicating that the scores decrease from top to bottom.

Relevance Feedback

TF-IDF Weighting + Cosine Similarity :	
DocID	Score
932	0.700321
248	0.383555
38	0.277447
234	0.234500
569	0.221815

Extract Feedback Vector

Relevance Feedback (cont'd)

TF-IDF Weighting + Cosine Similarity :	
DocID	Score
932	0.700321
248	0.383555
38	0.277447
234	0.234500
569	0.221815

Feedback Queries + TF-IDF Weighting + Cosine Similarity:	
DocID	Score
932	0.900000
624	0.191011
234	0.176471
336	0.165563
25	0.163121

Feedback Vector
Leave only Nouns & Verbs

Query Vector
+
Feedback Vector

Relevance Feedback (cont'd)

TF-IDF Weighting + Cosine Similarity :	
DocID	Score
932	0.700321
248	0.383555
38	0.277447
234	0.234500
569	0.221815

Feedback Vector

Leave only Nouns & Verbs

You are allowed to use any toolkit
for such grammatical tagging.

keywords:
part-of-speech tagging
POS-tagging
simply tagging

Submission Format:

1. You are asked to handle a zip file (named in <學號>.zip) to wm5 with every related flies (including programs and data) you need.
2. Please write a ReadMe to introduce your code.
3. For people who use Python:
 1. The main execute file should be named main.py
 2. My execute command will always be “python main.py --query <query>”, please make sure all your program can be run like this. You can use “argparser” or other packages to achieve this.
4. For people who “do not” use Python:
 1. Write the executing way and the packages you need in ReadMe.

Any Questions?
