Different Modes of Semantic Representation in Image Retrieval

By Rory Bennett
Advisor: Kristina Striegnitz

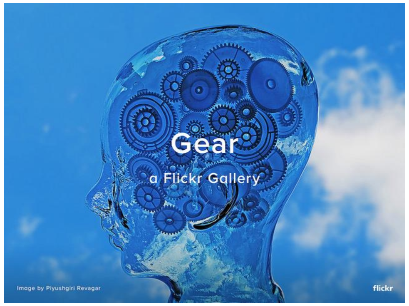
Image Retrieval

dog

war

Concreteness & Imageability

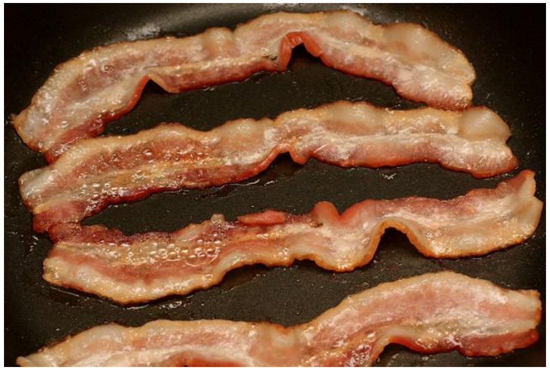
Abstract(less concrete), less imageable: *concept*



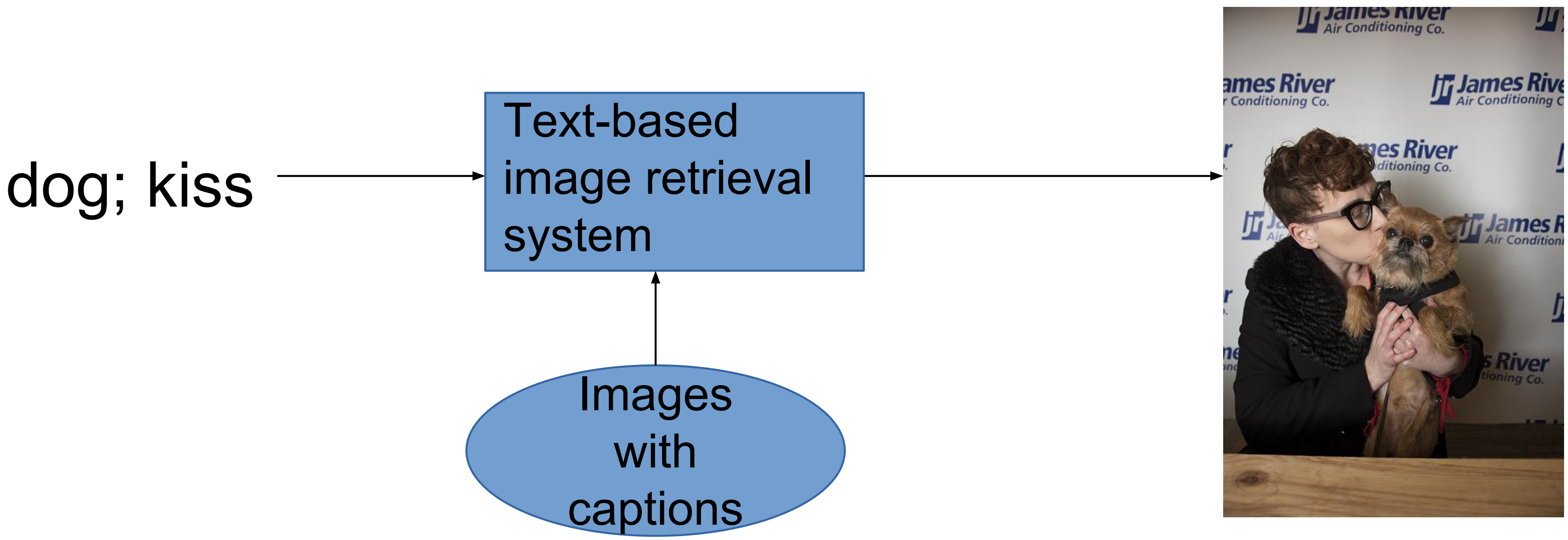
Concrete, less imageable: *argue*

Abstract, more imageable: *plead*

Concrete, more imageable:



Text-based Image Retrieval (TBIR)



This woman is giving her dog a kiss

Text-based Image Retrieval (TBIR)

dog; kiss

Text-based image retrieval system

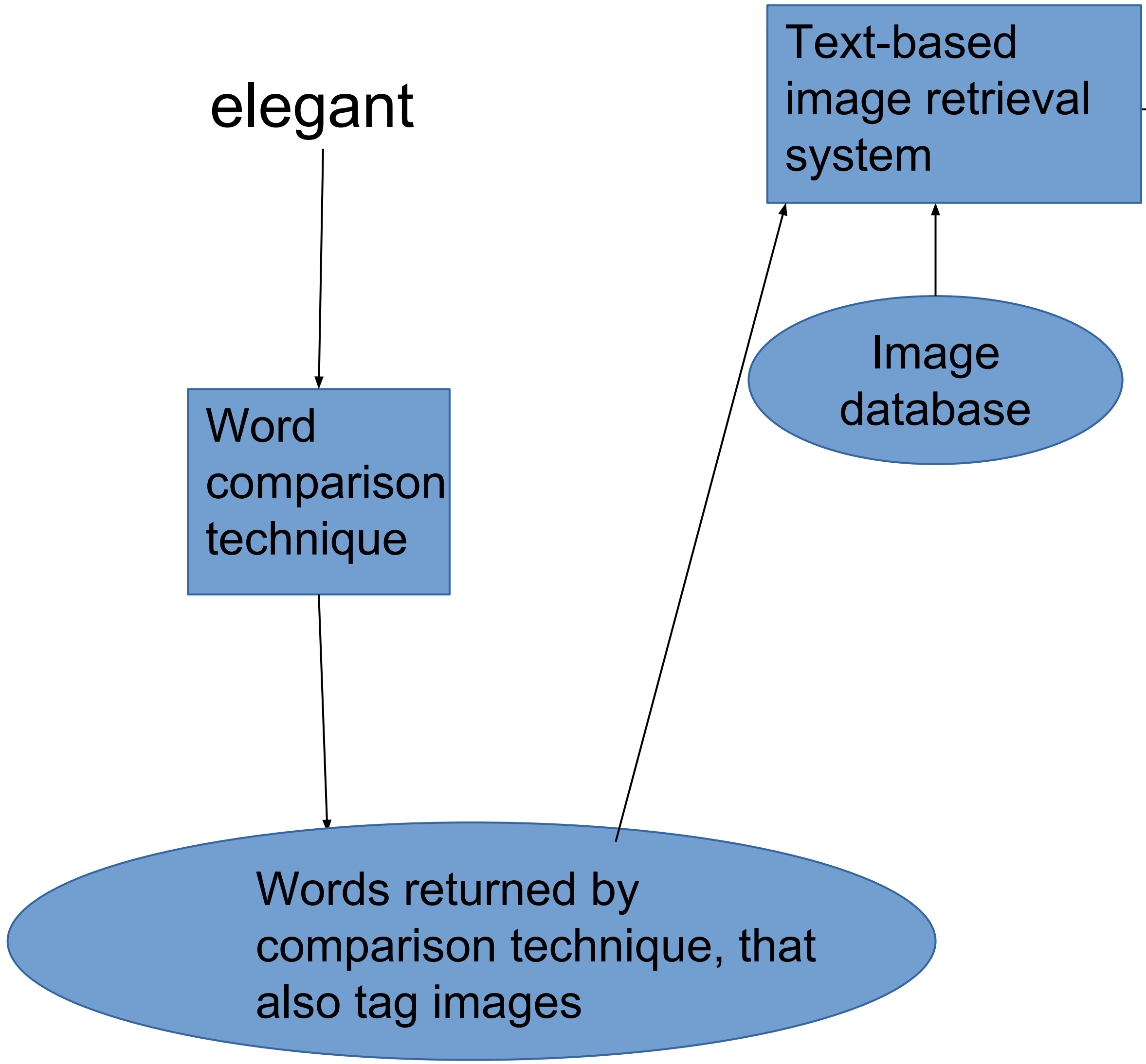
Images with captions

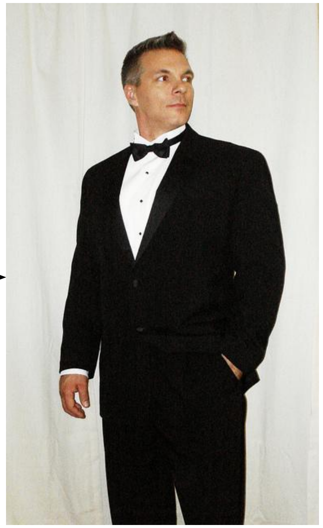
This woman is giving her dog a kiss

love; war

???

Retrieval Based on Word Similarity





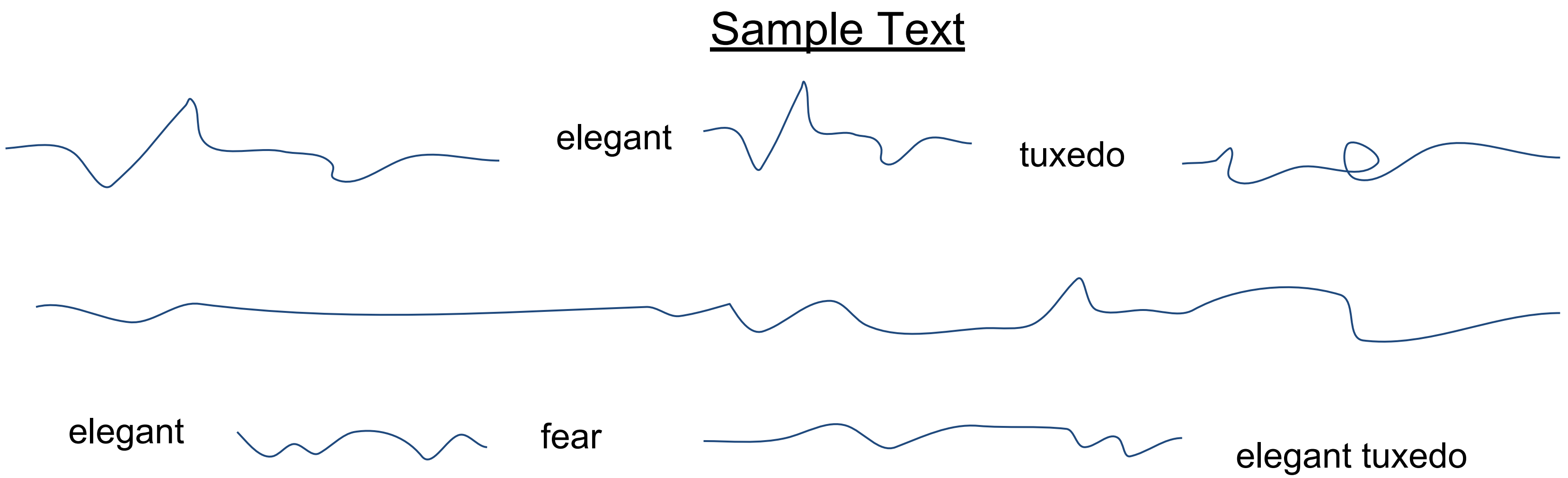
The tuxedo is the perfect formal garb.

Semantic Vector Representations

elegant: [-0.081428, 0.102486, -0.198815 , -0.145852 , -0.148051, …]

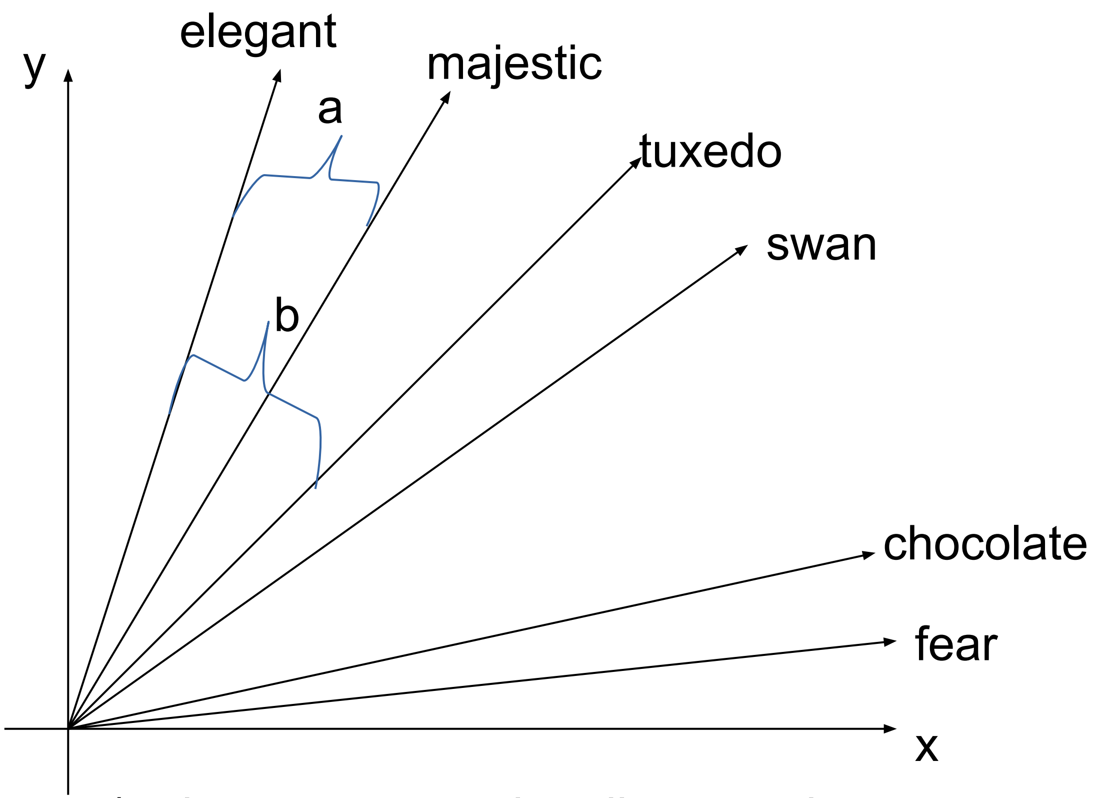
tuxedo: [-0.116671, -0.163012, -0.094523, -0.108007, 0.084851, …]

fear: [0.121500, -0.413079, -0.040310, 0.113604, -0.353846, …]



Semantic Vector Representations (cont.)

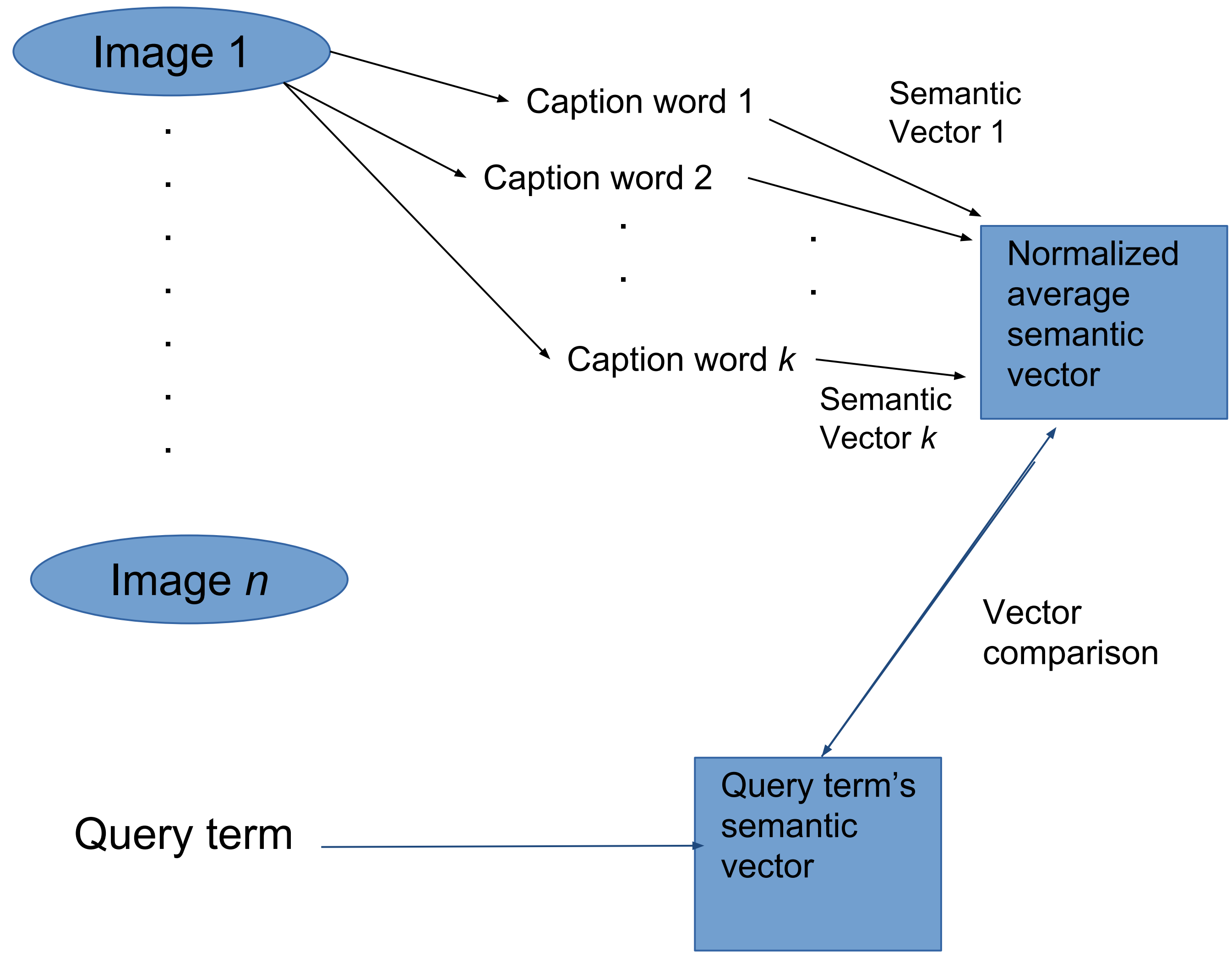
- All vectors are mapped to a common vector space, to compare vector cosines and thus find words with similar meanings



*a, b represent cosine distances between semantic vectors

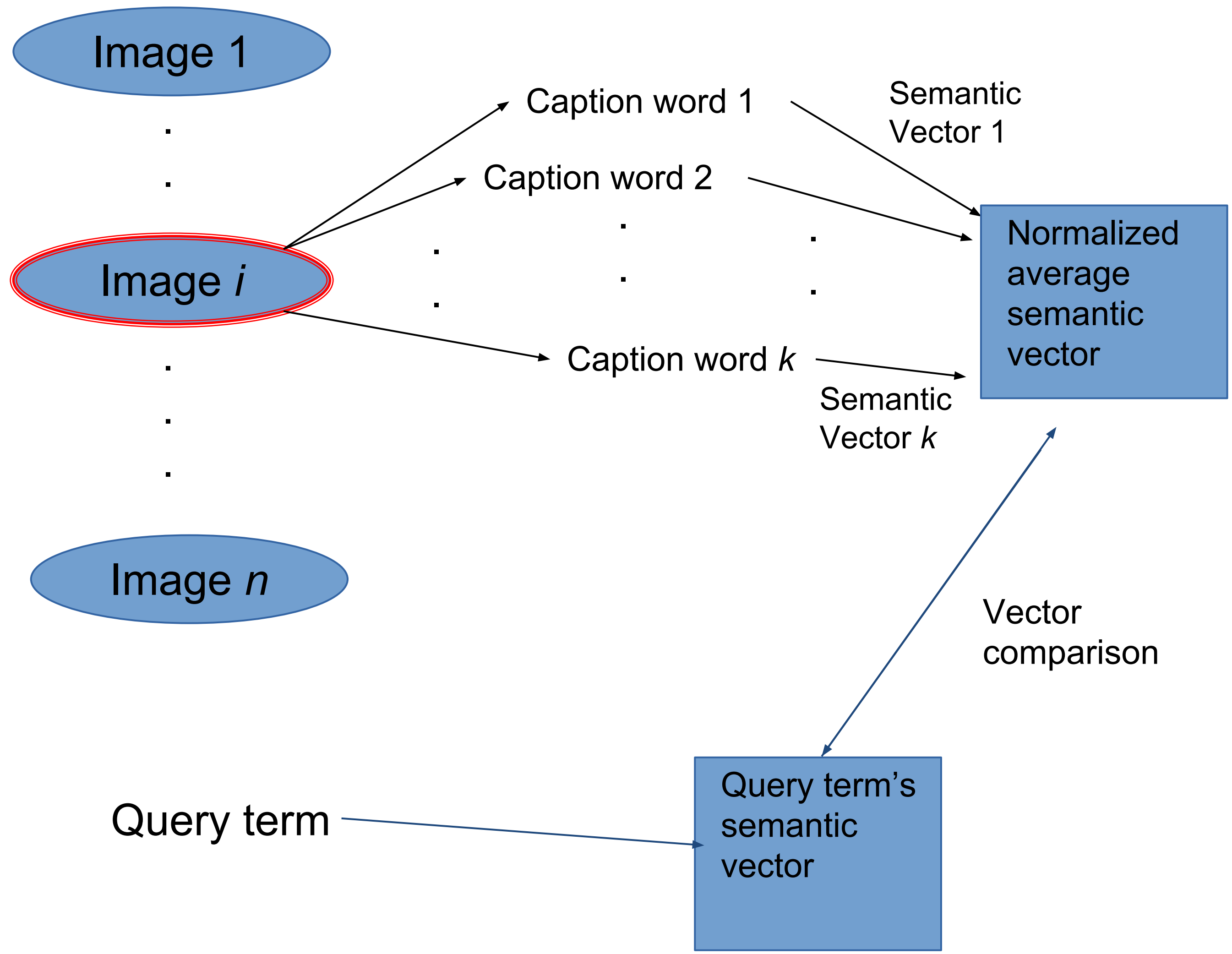
Vector Comparison, Approach A

<u>Entire Image Dataset</u>



Vector Comparison, Approach B

<u>Images directly tagged by words most similar to query term</u>



Abstract Words' Meanings Encapsulate Concrete Words' Meanings

- Lawrence W. Barsalou, Katja Wiemer-Hastings: abstract terms provide more general, overarching descriptions of images related to concrete terms
- Google query for abstract term, "love":

Augmenting Textual Data With Perceptual Information

- Felix Hill and Anna Korhonen used the Text8 textual corpus, and perceptual datasets comprising captioned images and feature-annotations of cue words.

<u>Text Corpus</u>

The **dog** sits happily on the porch **...**

.
.
.
.
.
.
.

Images with captions

dog, *fur*, *tail*, *kibble*, ...

Insert words
into text corpus

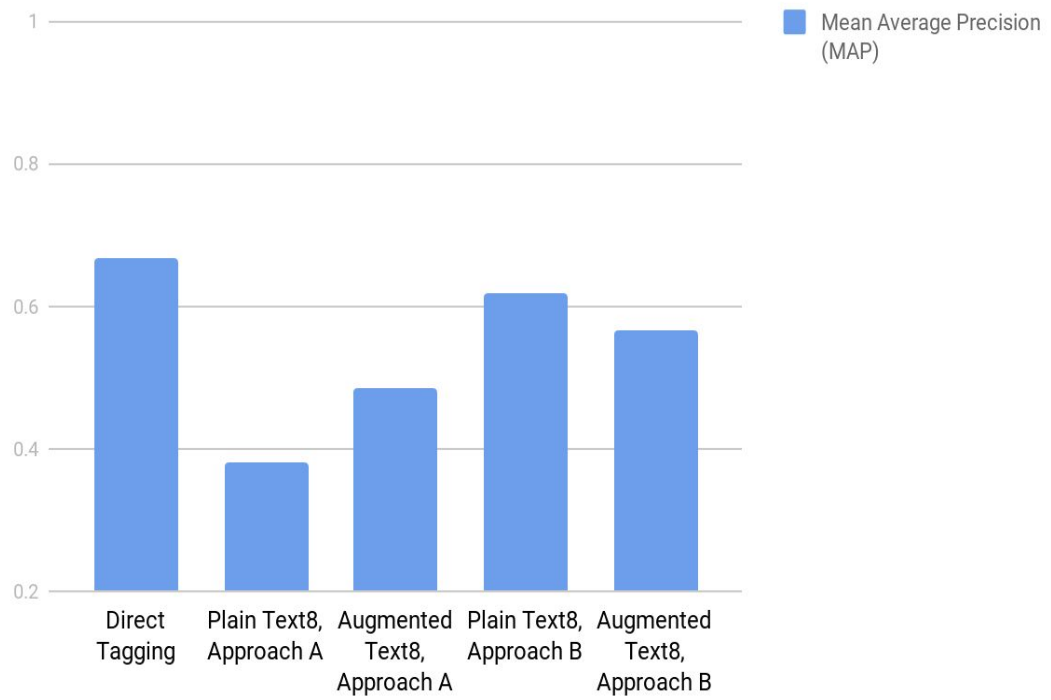
Experiment – Five Approaches

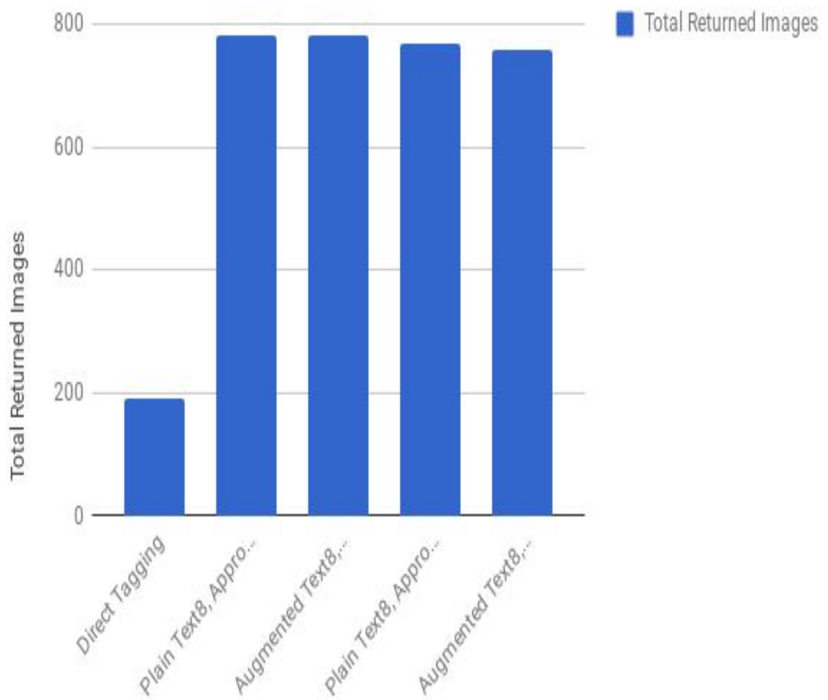
- Retrieve images directly tagged by query term
- Apply Approach A on plain Text8 corpus
- Apply Approach B on plain Text8
- Apply Approach A on augmented Text8
- Apply Approach B on augmented Text8

Experiment – Query Terms

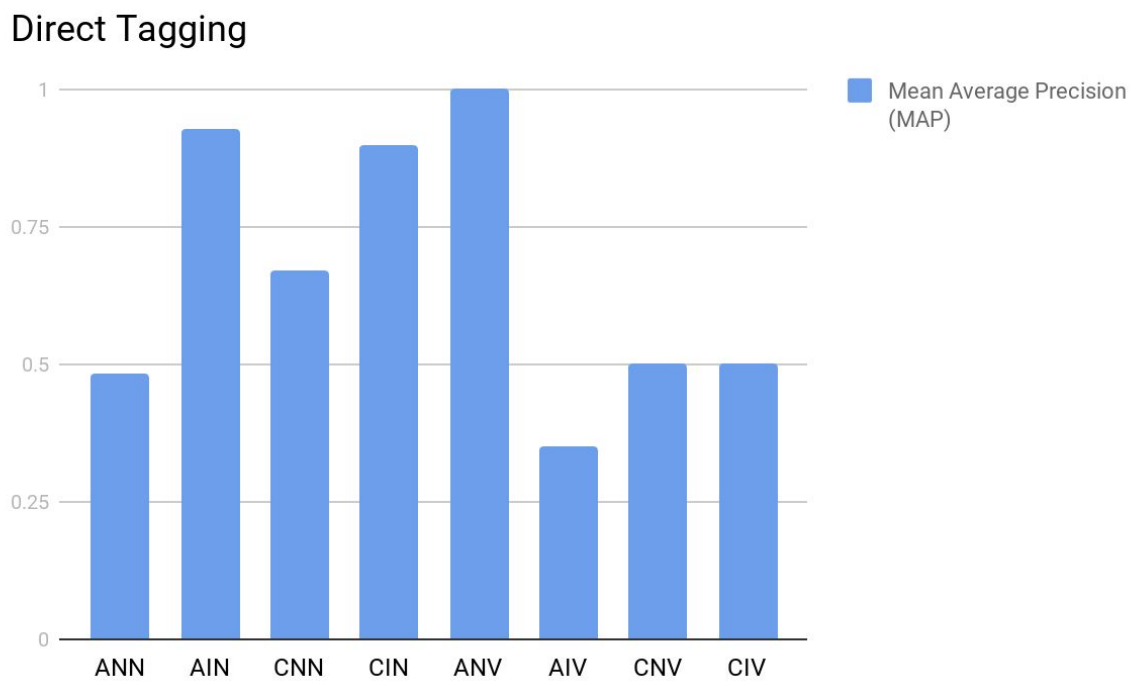
Term	Imageability	Concreteness	
norm	142	2.18	Less concrete, less imageable nouns
expense	160	2.77	
custom	166	2.99	
concept	197	1.97	
silence	413	3.09	Less concrete, more imageable nouns
chaos	426	2.50	
hazard	459	3.38	
demon	533	2.56	
roach	365	6.42	More concrete, less imageable nouns
creek	378	5.95	
nylon	415	6.16	
jury	426	6.17	
airport	650	6.31	More concrete, more imageable nouns
bacon	650	6.46	
tractor	655	5.86	
leaf	655	5.89	
become	105	2.66	Less concrete, less imageable verbs
allow	170	2.64	
restore	178	2.71	
prove	221	2.54	
choose	239	3.00	Less concrete, more imageable verbs
amuse	255	3.17	
plead	265	3.08	
send	274	3.08	
weigh	384	3.54	More concrete, less imageable verbs
grind	390	4.37	
argue	395	3.23	
spell	429	3.49	
tickle	450	4.69	More concrete, more imageable verbs
knock	460	5.09	
bake	481	4.76	
marry	498	3.41	

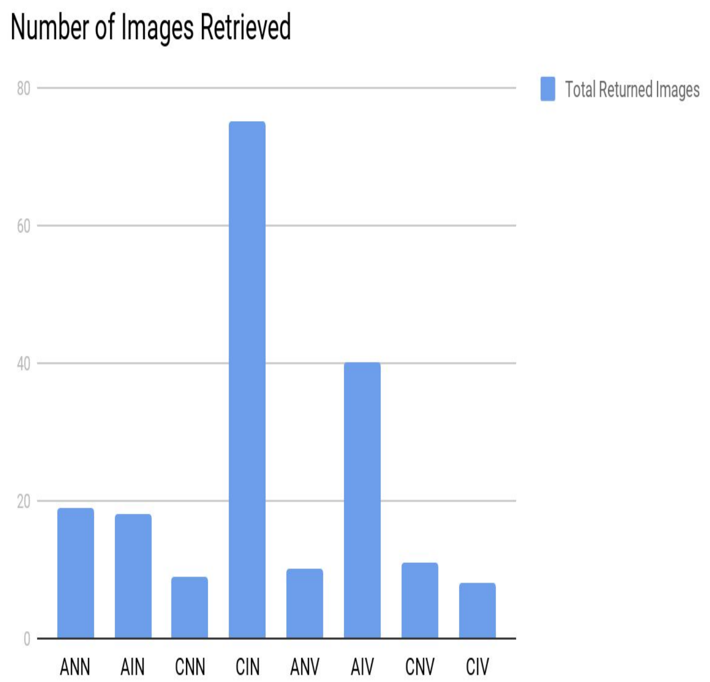
Experiment – Results, Part I



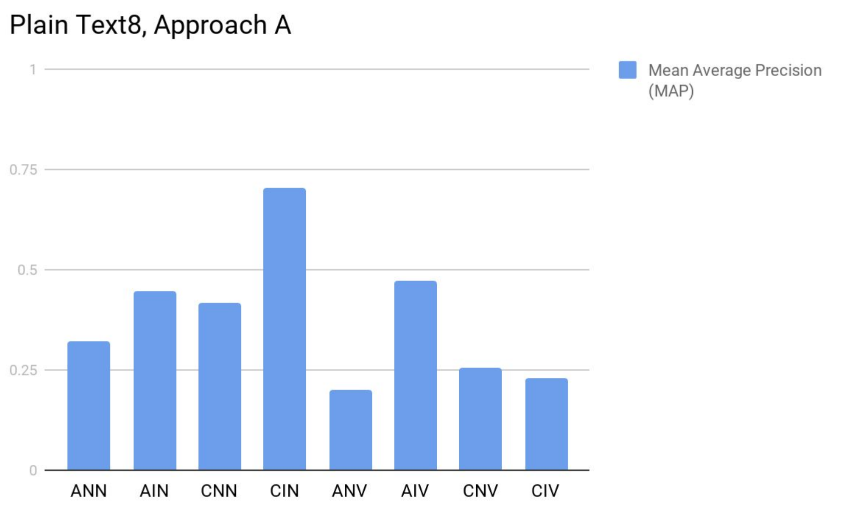


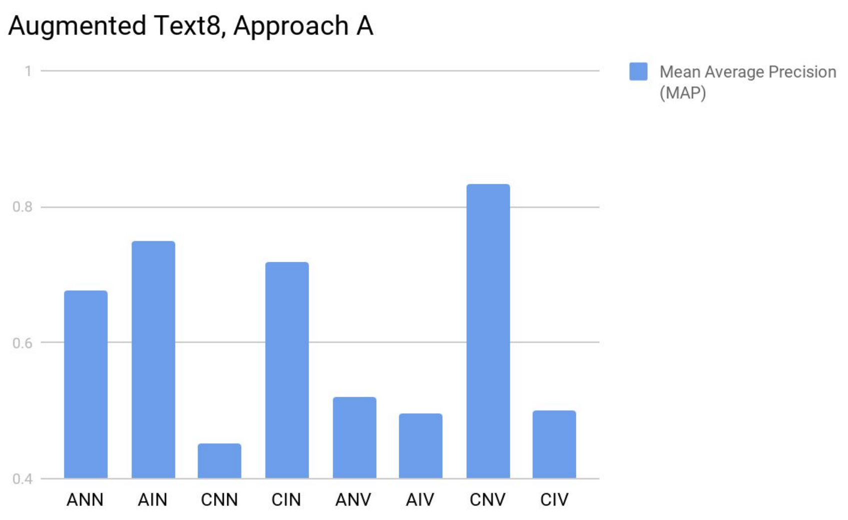
Results – Part II

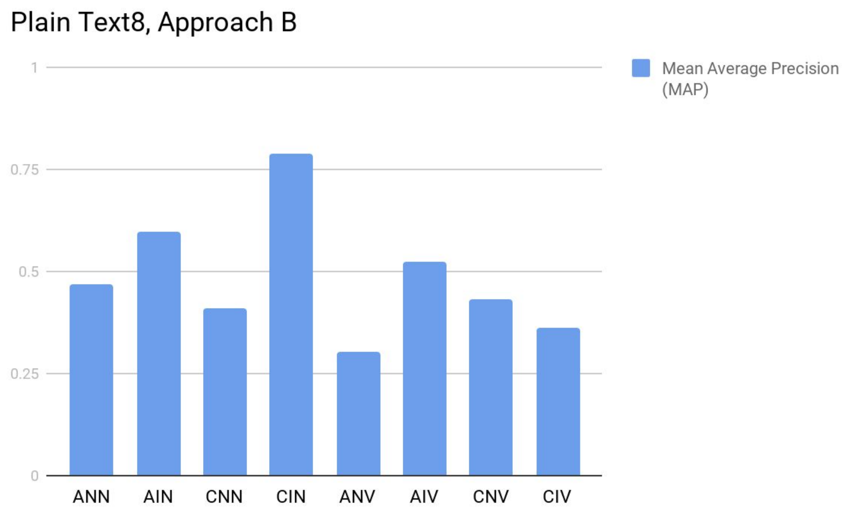


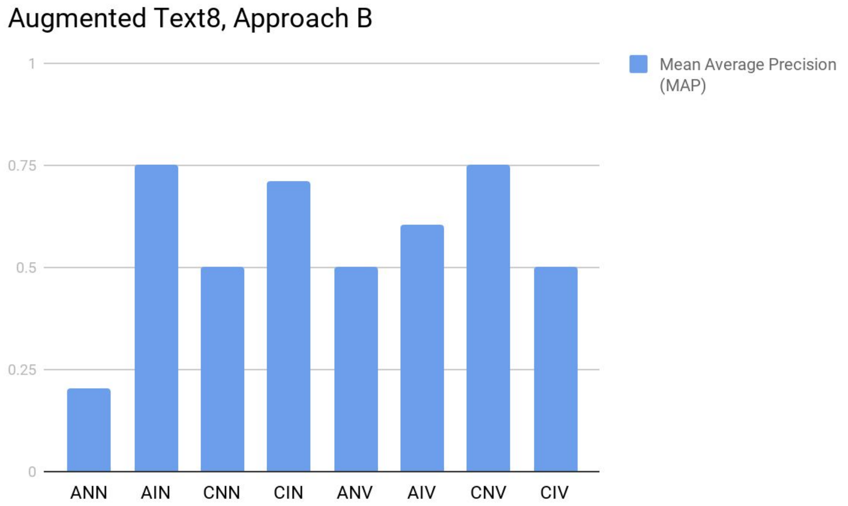


Results – Part III









Conclusions

- Utilizing perceptual information to form semantic vectors does not significantly inhibit, and can actually improve, the relevance of returned images.
- There is at least some (if insignificant) increase in the relevance of retrieved images when switching from applying Approach A to applying Approach B for a single textual corpus.
- If we assume that results from direct tagging are ideal, regardless of their paucity, then this indicates that including perceptual data brings retrieval closer to this ideal

Future Work

- Focus on vector representations for words whose part of speech is typically very abstract, *e.g.*, adverbs
- Better account for representation words with multiple diverse meanings