

Different Modes of Semantic Representation in Image Retrieval

By Rory Bennett
Advisor: Kristina Striegnitz

Image Retrieval

dog

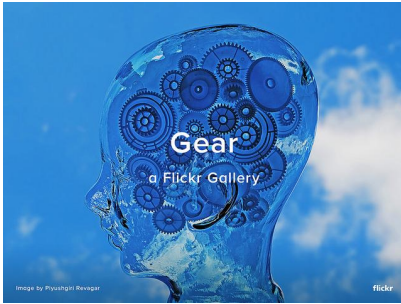


war



Concreteness & Imageability

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



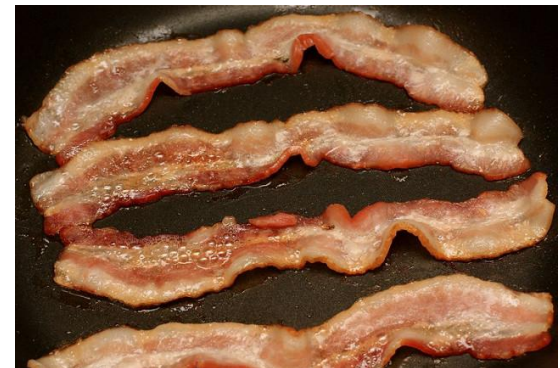
Concrete, less imageable: *argue*



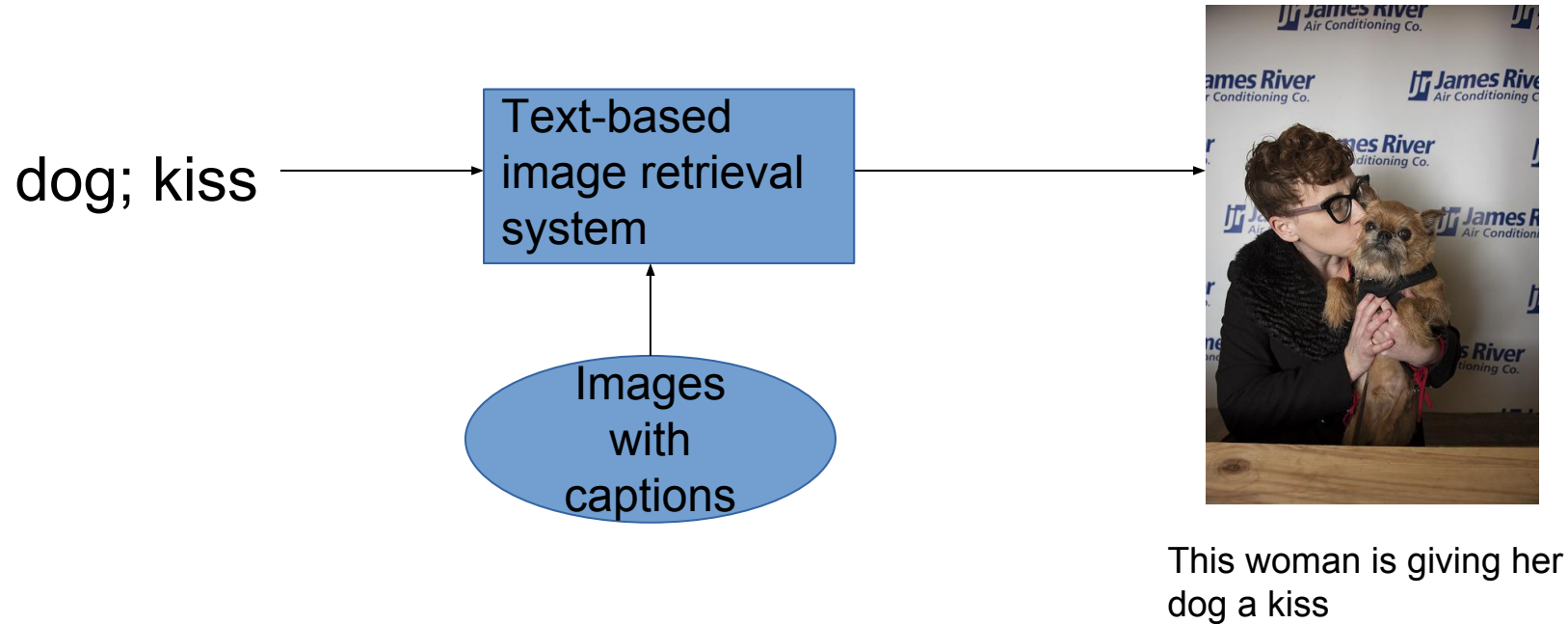
Abstract, more imageable: *plead*



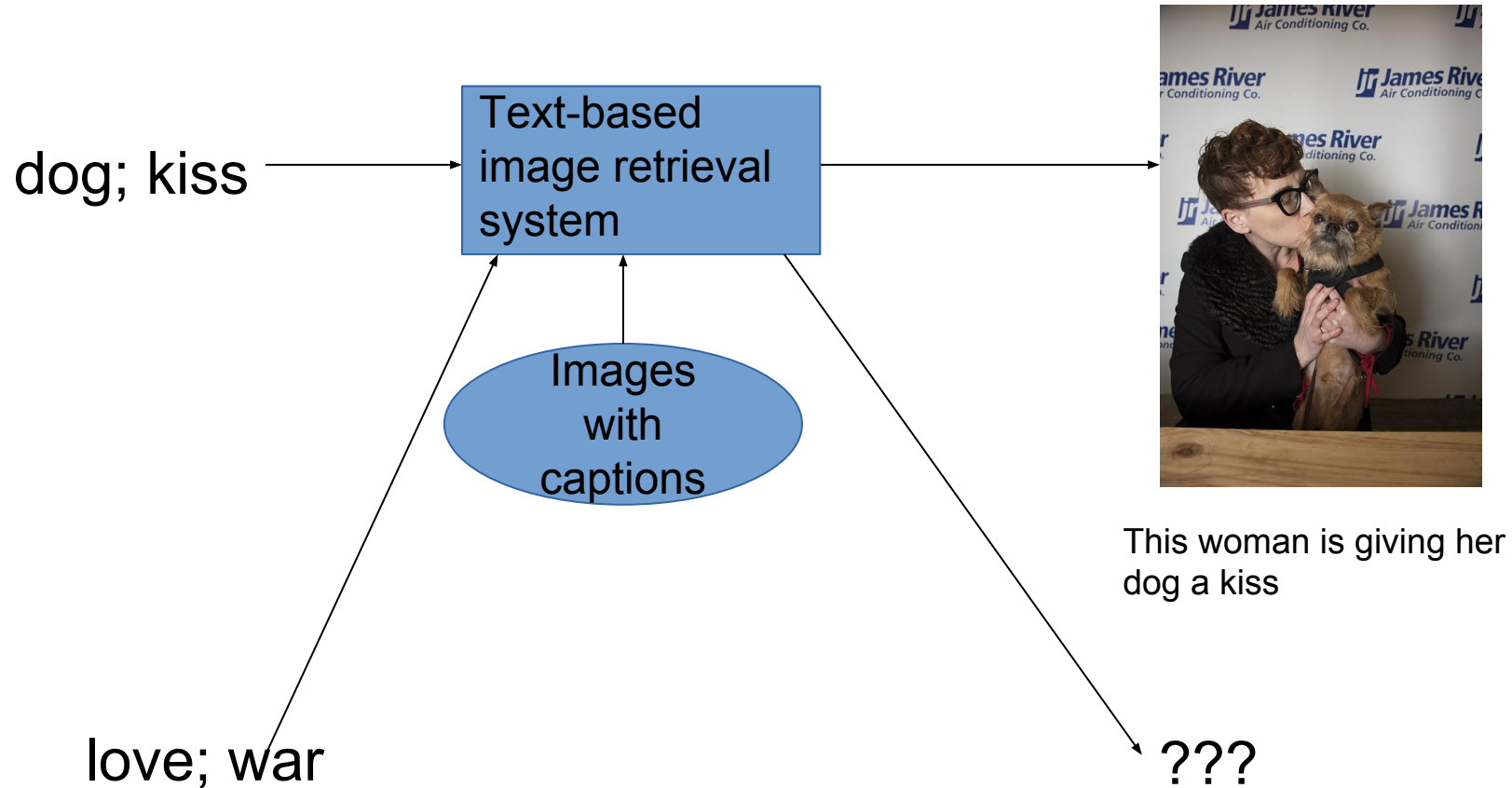
Concrete, more imageable: *bacon*



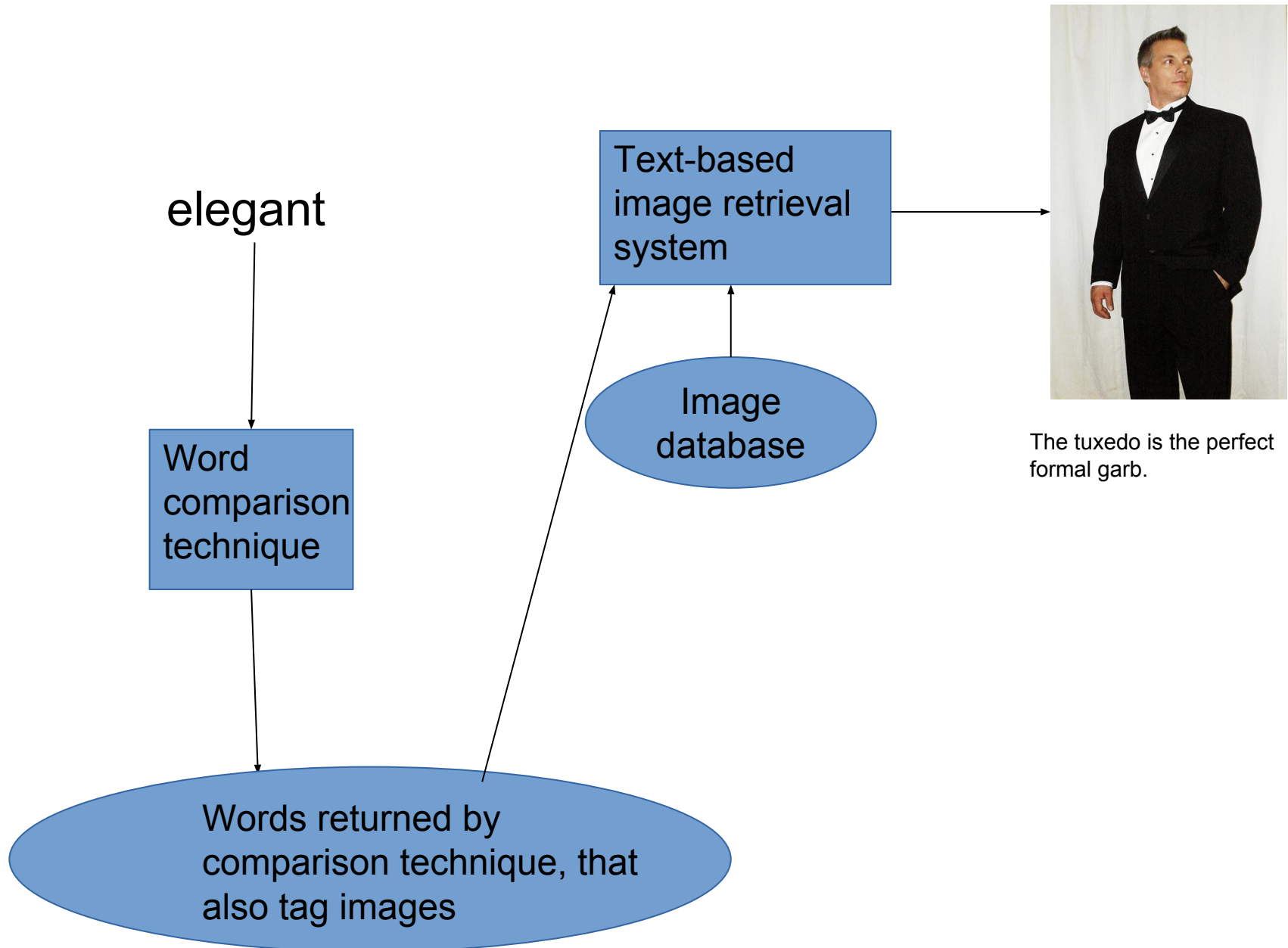
Text-based Image Retrieval (TBIR)



Text-based Image Retrieval (TBIR)



Retrieval Based on Word Similarity



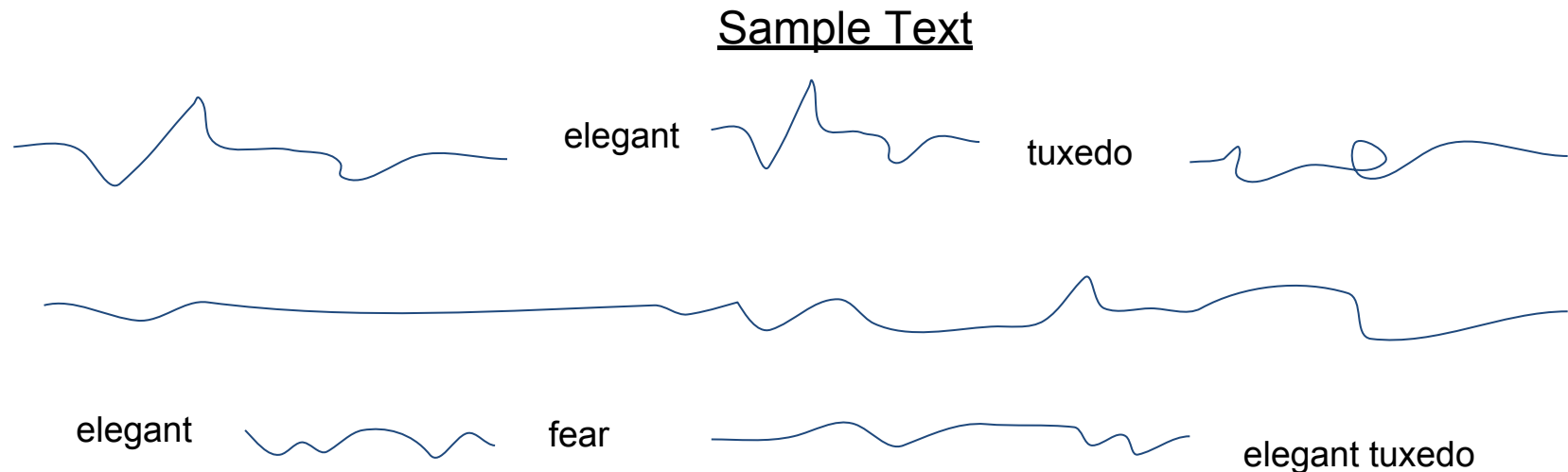
Semantic Vector Representations

-Each vector represents a word's distribution in a given text corpus

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

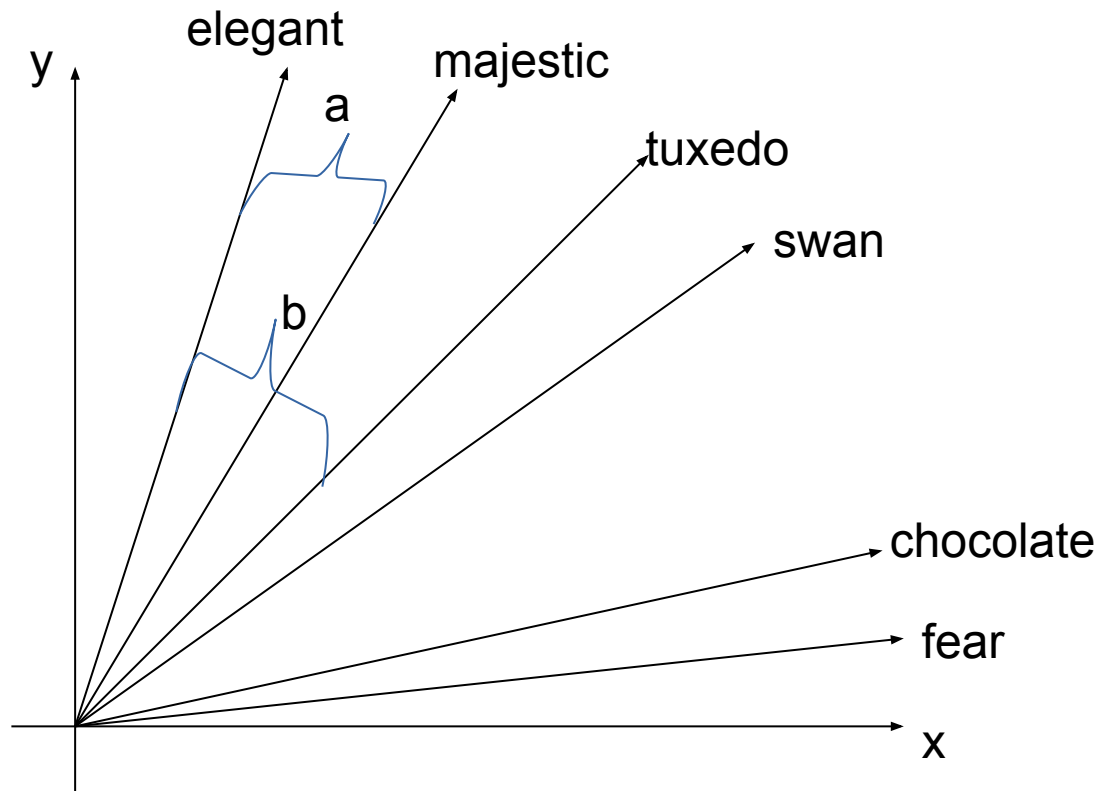
tuxedo: [-0.116671, -0.163012, -0.094523, -0.108007, 0.084851, ...]

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



Semantic Vector Representations (cont.)

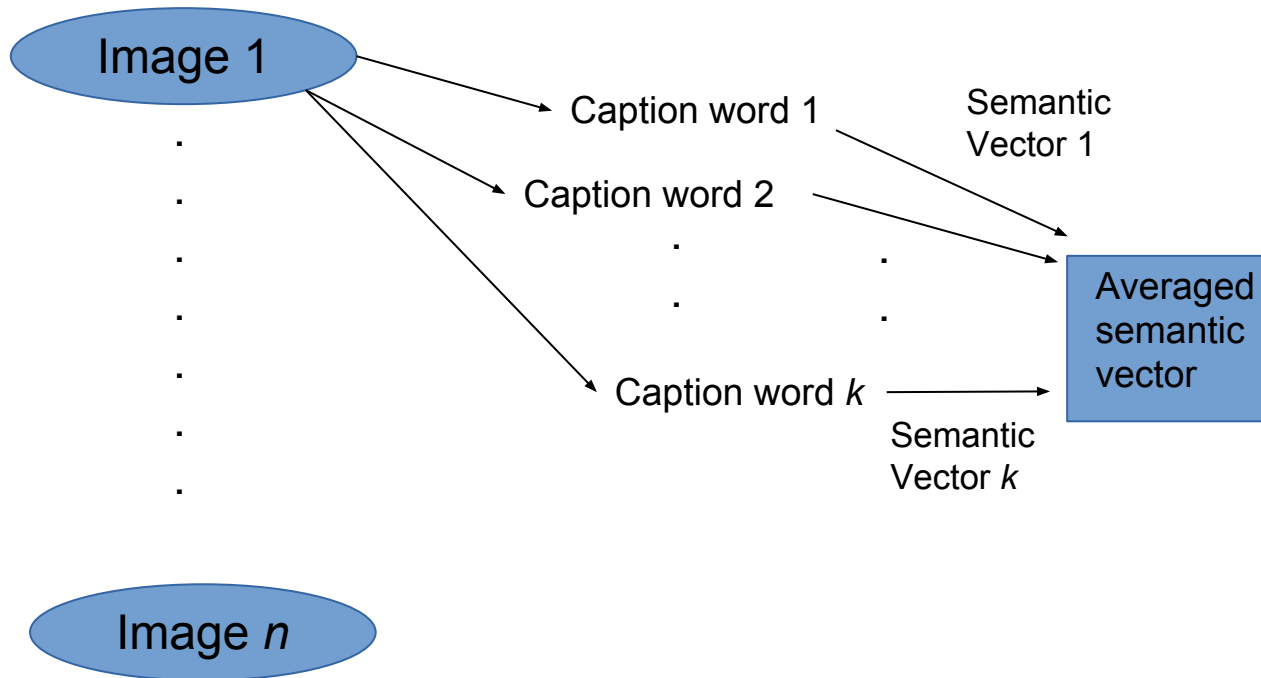
- All vectors are mapped to a common vector space, to compare vector cosines and thus find words with similar meanings



*a, b represent cosine distances between semantic vectors

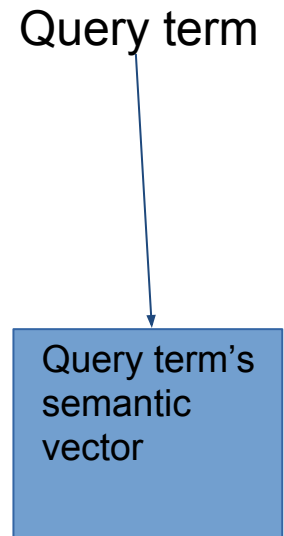
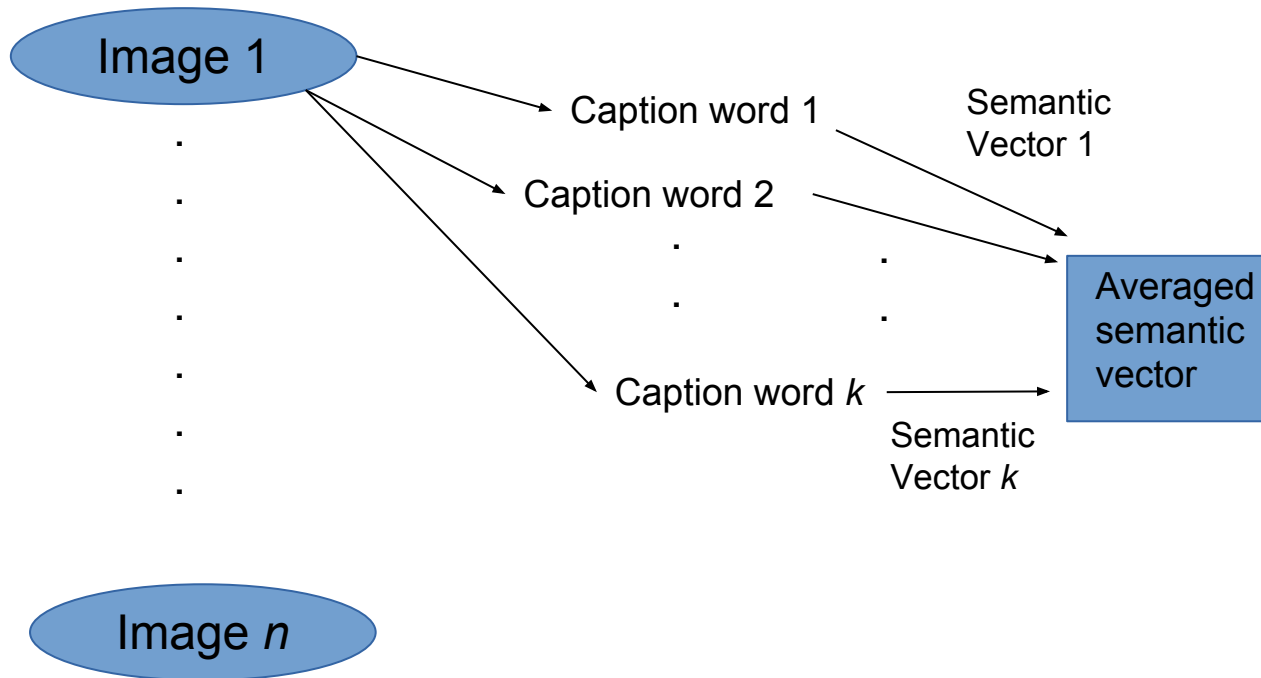
Vector Comparison Approach: Unfiltered

Entire Image Dataset



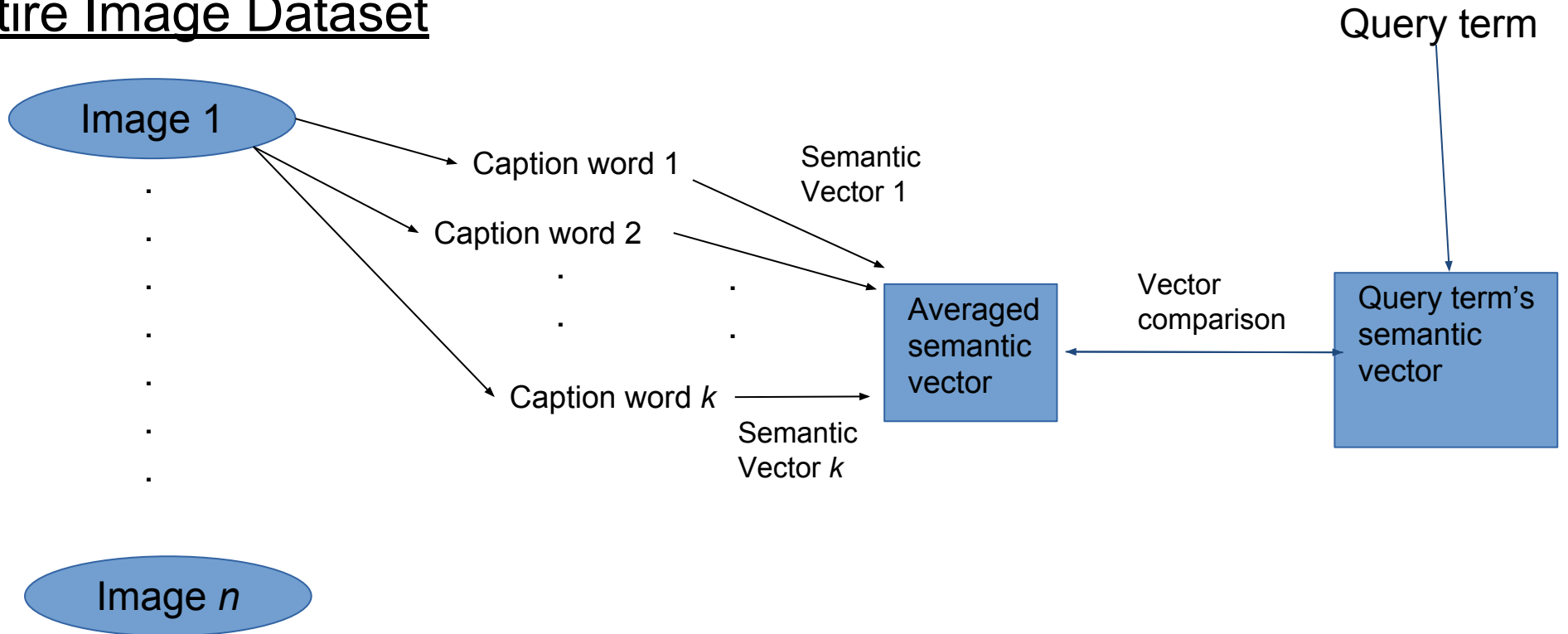
Vector Comparison Approach: Unfiltered

Entire Image Dataset



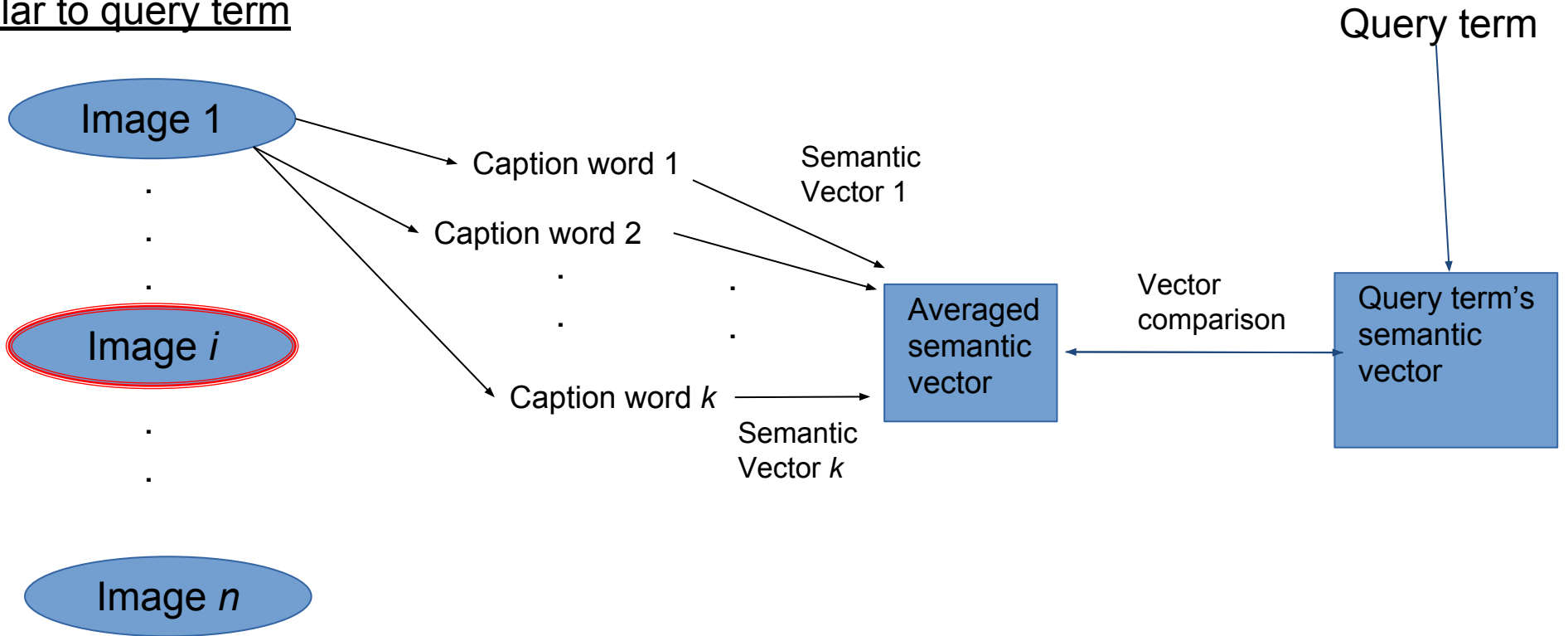
Vector Comparison Approach: Unfiltered

Entire Image Dataset



Vector Comparison Approach: Filtered

Images directly tagged by words most similar to query term



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”:



Experiment – Five Approaches

- Retrieve images directly tagged by query term
- Unfiltered vector comparison on plain word vectors
- Filtered vector comparison on plain word vectors
- Unfiltered vector comparison on augmented word vectors
- Filtered vector comparison on augmented word vectors

Examples of Query Terms in Experiment

Nouns

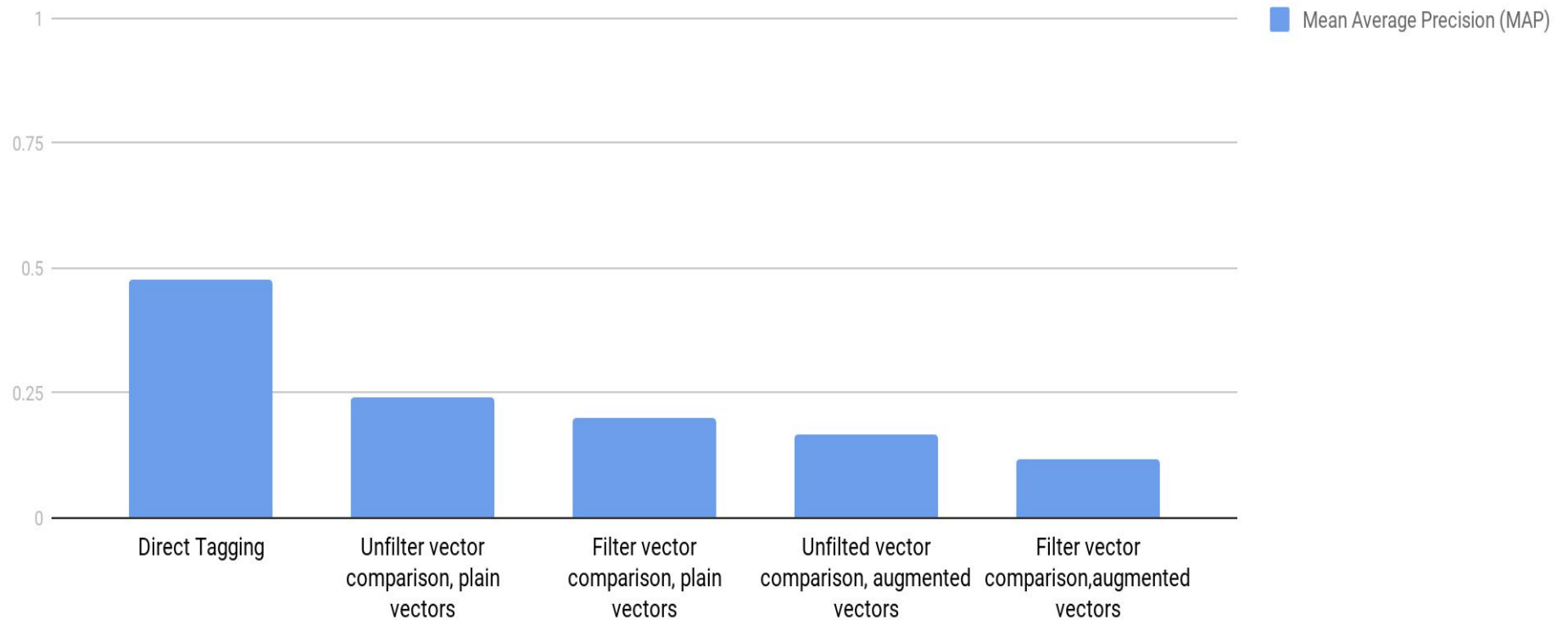
- Abstract, less imageable noun: *custom*
- Abstract, more imageable noun: *demon*
- Concrete, less imageable noun: *jury*
- Concrete, more imageable noun: *tractor*

Verbs

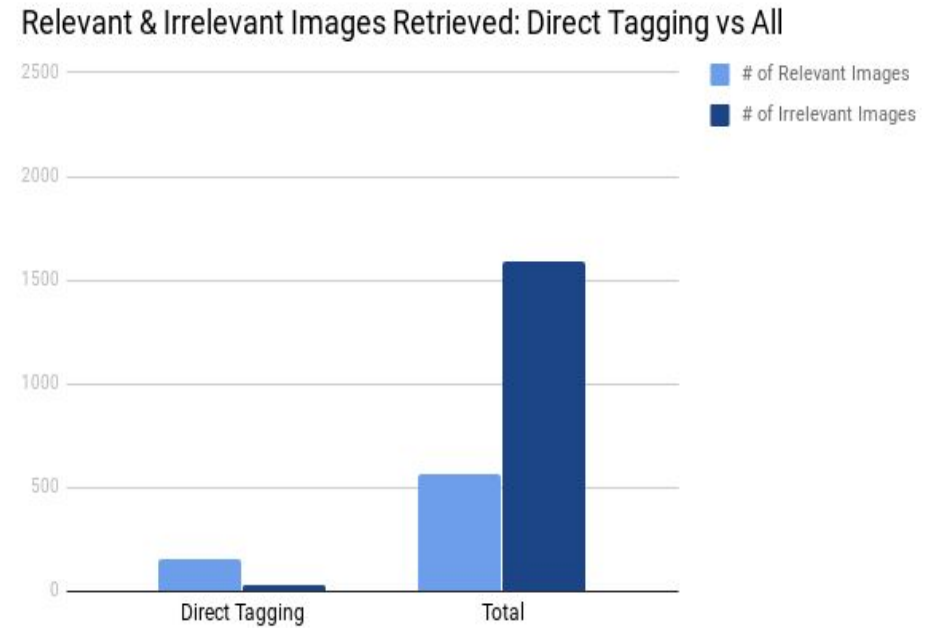
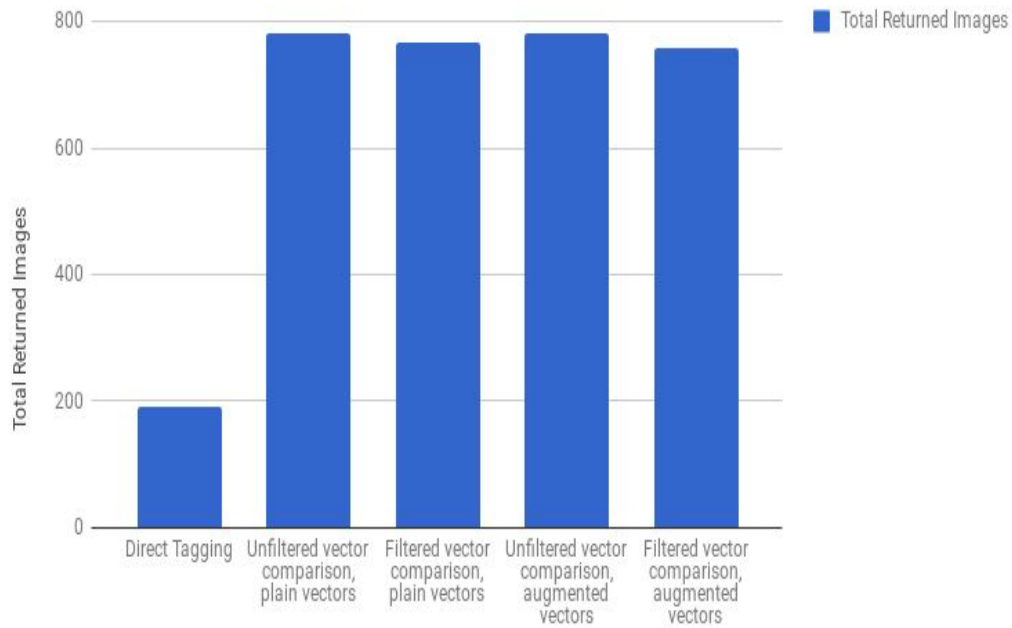
- Abstract, less imageable verb: *become*
- Abstract, more imageable verb: *plead*
- Concrete, less imageable verb: *grind*
- Concrete, more imageable verb: *tickle*

Experiment – Results, Part I

Overall Mean Average Precisions (MAPs), per Approach



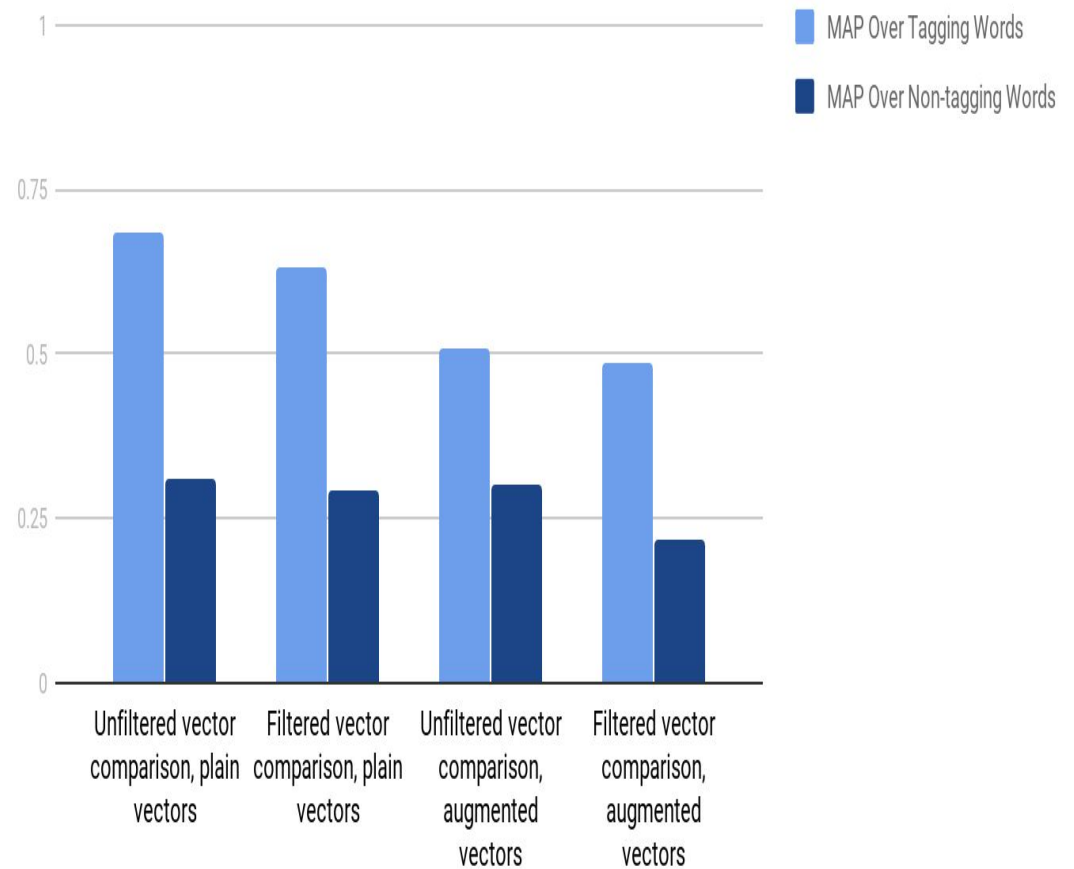
Experiment – Results, Part II



Results – Part III

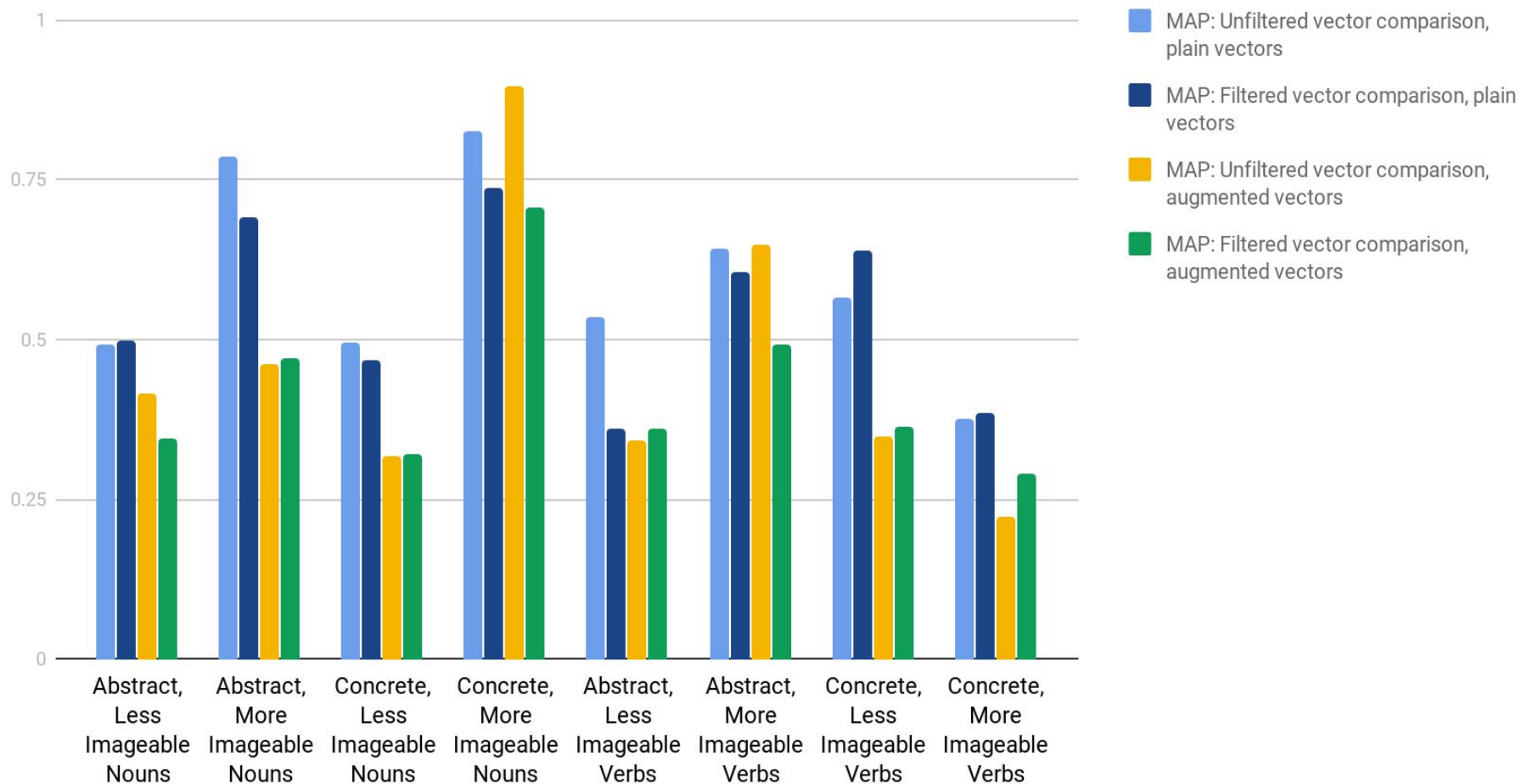
Term	Concreteness	Imageability
Abstract, Less Imageable Nouns		
norm	2.18	142
expense	2.77	160
custom	2.99	166
concept	1.97	197
Abstract, More Imageable Nouns		
silence	3.09	413
chaos	2.50	426
hazard	3.38	459
demon	2.56	533
Concrete, Less Imageable Nouns		
roach	6.42	365
creek	5.95	378
nylon	6.16	415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

MAP: Tagging Words vs. Non-tagging Words



Results – Part IV

MAP, per Group of Words



Conclusions

- In general, the vector comparison models retrieved more relevant results for words that tagged at least one image
- The semantic vector models were, overall, able to produce noticeable results for query terms not initially associated with images, i.e., not often used in captions
- Direct tagging should only be considered a benchmark method for those words for which it actually retrieves images.
- The vector comparison models that propagated perceptual information to the vectors only performed, at best, as well as those models that did not.

Future Work

- Improve precision of vector comparison models
- 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