

Web Search & Mining - Presentations of examples of results -

Phase 1

José Monteiro - 43211

Francisco Rodrigues - 42727

RESULT 1 - Query with good semantics and context under Bag of Entities:

Bag of Entities detects relevant entities in the query text, due to it's good semantics, but the dataset doesn't contain the same entities.

Original Query: The most eye-catching nominee is the Fife comic Richard Gadd, overlooked for his extraordinary stunt-comedy show Waiting for Gaddot last year

Entities Extracted: ['Fife', 'Richard Gadd', 'Gaddot']

Result #1

Distance: 1.0

Original Text: Here Are All The Coolest Things Happening In Edinburgh This August

#EdFringe #EdinburghFest

[https://www.buzzfeed.com/hilarywardle/awesome-things-to-do-in-edinburgh-this-august?](https://www.buzzfeed.com/hilarywardle/awesome-things-to-do-in-edinburgh-this-august?utm_term=.fbopG8IO6#.vrpoXpbg4)

utm_term=.fbopG8IO6#.vrpoXpbg4 ...pic.twitter.com/U8h5jTJ50J

Entities Extracted: ['Are All', 'Coolest Things', 'Edinburgh']

Result #2

Distance: 1.0

Original Text: Congratulations @JavaDanceCo on the fantastic smiley cheeky and utterly engaging #backofthebus #mustsee #edfringe pic.twitter.com/ysEFETfIK6

Entities Extracted: []

Result #3

Distance: 1.0

Original Text: And it's goodbye #edfringe for #ARoyalFlush what a month it's been... Thanks to everyone that's supported us!pic.twitter.com/KRmdTE6Eht

Entities Extracted: ['ARoyalFlush', 'Thanks']

Result #4

Distance: 1.0

Original Text: Fireworks are ready for tonight #edfringe @EdinburghTattoopic.twitter.com/qyxCtV7zsw

Entities Extracted: []

Result #5

Distance: 1.0

Original Text: Moody sky tonight as the @edfringe end looms. #Edinburgh #edfringe pic.twitter.com/weL0Swe2r1

Entities Extracted: ['Moody', 'Edinburgh']

RESULT 2 - Bag of Words performs better than Bag of Entities when query entity extraction is not optimal

Example where Bag of Words performance in terms of relevance of results is much higher of Bag of Entities due to poor semantics of query and dataset texts.

Bag Of Words

Original query: Deep time Show

Tokens: ['deep', 'time', 'show']

Result #1

Original: Edinburgh Festival Review: Deep Time <http://bit.ly/2b7q7Ex> pic.twitter.com/eZKYMP3Pvd

Tokens: ['edinburgh', 'festival', 'review', ':', 'deep', 'time', 'pic.twitter.com/ezkym3pvd']

Result #2

Original: Have you got your tickets for Deep Time yet? #EdFringe- <http://ow.ly/8VTq302PUVO> pic.twitter.com/cT0rmUeVPZ

Tokens: ['have', 'you', 'got', 'your', 'tickets', 'for', 'deep', 'time', 'yet', '?', 'edfringe', '-', 'pic.twitter.com/ct0rmuevpz']

Result #3

Original: time for #edfringe adventures round five with @_lisaduncan - @MalibuUK cocktail time <https://instagram.com/p/BlxoEgcj7Jm/> pic.twitter.com/ohKI5WNmTW

Tokens: ['time', 'for', 'edfringe', 'adventures', 'round', 'five', 'with', '-', 'cocktail', 'time', 'pic.twitter.com/ohkl5wnmtw']

Result #4

Original: Deep Time tickets now available from Edinburgh Festival, will be a stunning event @EventScotNewspic.twitter.com/TOXUHvEHbz

Tokens: ['deep', 'time', 'tickets', 'now', 'available', 'from', 'edinburgh', 'festival', ',', 'will', 'be', 'a', 'stunning', 'event', ':.twitter.com/toxuhvehbz']

Result #5

Original: Capturing the castle: Edinburgh festival's Deep Time spectacular - in pictures <http://dlvr.it/Lz427L> pic.twitter.com/3rEphPfMsQ

Tokens: ['capturing', 'the', 'castle', ':', 'edinburgh', "festival's", 'deep', 'time', 'spectacular', '-', 'in', 'pictures', 'pic.twitter.com/3rephpfmsq']

Bag Of Entities

Original query: Deep time Show

Entities Extracted: ['Deep']

Result #1

Distance: 1.0

Original: Here Are All The Coolest Things Happening In Edinburgh This August

#EdFringe #EdinburghFest

https://www.buzzfeed.com/hilarywardle/awesome-things-to-do-in-edinburgh-this-august?utm_term=.fbopG8IO6#.vrpoXpbg4 ...pic.twitter.com/U8h5jTJ50J

Entities Extracted: ['Are All', 'Coolest Things', 'Edinburgh']

Result #2

Distance: 1.0

Original: Congratulations @JavaDanceCo on the fantastic smiley cheeky and utterly engaging #backofthebus #mustsee #edfringe pic.twitter.com/ysEFETfIK6

Entities Extracted: []

Result #3

Distance: 1.0

Original: And it's goodbye #edfringe for #ARoyalFlush what a month it's been... Thanks to

everyone that's supported us!pic.twitter.com/KRmdTE6Eht

Entities Extracted: ['ARoyalFlush', 'Thanks']

Result #4

Distance: 1.0

Original: Fireworks are ready for tonight #edfringe @EdinburghTattoopic.twitter.com/qyxCtV7zsw

Entities Extracted: []

Result #5

Distance: 1.0

Original: Moody sky tonight as the @edfringe end looms. #Edinburgh #edfringe pic.twitter.com/weL0Swe2r1

Entities Extracted: ['Moody', 'Edinburgh']

RESULT 3 - HISTOGRAM OF GRADIENTS AND HISTOGRAM OF COLORS ALWAYS FINDS IDENTICAL IMAGES FIRST *Good to find duplicates*

Histogram of Colors

Query image: 770382667279896580.jpg



Result #1 - 0.0 - CrDy8EdXEAAbus2.jpg



Result #2 - 0.2835920810632604 - CrEC2VuWAAEq3oC.jpg



Result #3 - 0.4183080181095323 - CmCzUZaXIAyn3O.jpg



Result #4 - 0.4214805890702398 - CpC3nwMWgAAle2w.jpg



Result #5 - 0.4639592263639929 - CplBVObW8AA0wal.jpg



Histogram of Gradients

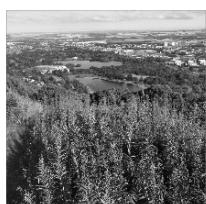
Query image: 759764821415956480.jpg



Result #1 - 3.650024149988857e-08 - Cos6E8YXgAA8X87.jpg



Result #2 - 0.39819640980285137 - Cp6pllgyYAAztql.jpg



Result #3 - 0.41008082717926453 - CpVQk71WgAAnDVW.jpg



Result #4 - 0.41049542817700313 - Cq4ZStAWIAQapwp.jpg



Result #5 - 0.41049542817700313 - Cq4ZStAWIAQapwp.jpg



RESULT 4 - CNN finds identical images because they will be tagged with the same concepts

Query image: 760513155030220800.jpg

Tags: [[('n09472597', 'volcano', 0.99551976), ('n04296562', 'stage', 0.0007322411), ('n04606251', 'wreck', 0.00049785577), ('n09256479', 'coral_reef', 0.0004787832), ('n03388043', 'fountain', 0.000475113)]]



Result #1 - 0.29289321881345254 - Co7WJuyXgAAZ3zQ.jpg

Tags: [('n09472597', 'volcano', 0.99487877), ('n04606251', 'wreck', 0.0006714705), ('n03388043', 'fountain', 0.00044587324), ('n04296562', 'stage', 0.0003450219), ('n02814860', 'beacon', 0.00033390033)]



Result #2 - 0.29289321881345254 - Co3irs4WgAAuMxR.jpg

Tags: [('n09472597', 'volcano', 0.99551976), ('n04296562', 'stage', 0.0007322439), ('n04606251', 'wreck', 0.00049785624), ('n09256479', 'coral_reef', 0.00047878275), ('n03388043', 'fountain', 0.00047511345)]



Result #3 - 0.29289321881345254 - Cq77OMCWAAAR871.jpg

Tags: [('n09472597', 'volcano', 0.9895354), ('n04606251', 'wreck', 0.0026111903), ('n04296562', 'stage', 0.00086728), ('n03388043', 'fountain', 0.00065283116), ('n01910747', 'jellyfish', 0.00057882385)]



Result #4 - 0.29289321881345254 - CnEve9LWcAEcrI5.jpg

Tags: [('n09472597', 'volcano', 0.990101), ('n04606251', 'wreck', 0.0009416508), ('n03388043', 'fountain', 0.00093915133), ('n04296562', 'stage', 0.00075947086), ('n09256479', 'coral_reef', 0.0005724213)]



Result #5 - 0.29289321881345254 - Cp2G0CcXEAAC6Ux.jpg

Tags: [('n09472597', 'volcano', 0.990101), ('n04606251', 'wreck', 0.0009416508), ('n03388043', 'fountain', 0.00093915133), ('n04296562', 'stage', 0.00075947086), ('n09256479', 'coral_reef', 0.0005724213)]



RESULT 5 - CNN shows more relevant results because it relies on image tagging instead of image features (hoc and hog)

Query image: 816795436816089089.jpg



Histogram of Colors

Result #1 - 0.5278705642557942 - CrOVAd0WAAA4vxx.jpg



Result #2 - 0.5278705642557942 - CqI_OUwWYAADed8.jpg



Result #3 - 0.5988444496331322 - Cx9HKQiWIAA8hl5.jpg



Result #4 - 0.6059599128268541 - CpPrsfTWcAAu0pp.jpg



Result #5 - 0.6059599128268541 - CpvcZS0WcAE0qET.jpg



Histogram of Gradients

Result #1 - 0.6288777559872413 - Cq9riSUWEAAQwxR.jpg



Result #2 - 0.6290020974691838 - CpSTNHLXgAETEoA.jpg



Result #3 - 0.6296009203785072 - CpvQk71WgAAnDVW.jpg



Result #4 - 0.634002986366626 - CqP-g3XWEAApfbp.jpg



Result #5 - 0.6431219251065767 - Cp_LYywW8AA0Zs_.jpg



CNN - VGG16

```
[['n03724870', 'mask', 0.2663365), ('n04229816', 'ski_mask', 0.072559334), ('n04296562',  
'stage', 0.047687907), ('n03424325', 'gasmask', 0.04697289), ('n03045698', 'cloak',  
0.04641951)]]
```



Result #1 - 0.5757359312880714 - CqqLAc9WgAEhYu3.jpg

[('n03720891', 'maraca', 0.19862232), ('n04296562', 'stage', 0.10316918), ('n09229709', 'bubble', 0.043001898), ('n04259630', 'sombrero', 0.0374089), ('n03724870', 'mask', 0.037276063)]



Result #2 - 0.717157287525381 - Cp-NzUOWYAAj8NP.jpg

[('n02669723', 'academic_gown', 0.24143954), ('n03787032', 'mortarboard', 0.122109644), ('n04296562', 'stage', 0.12184953), ('n04418357', 'theater_curtain', 0.0536856), ('n03045698', 'cloak', 0.027808886)]



Result #3 - 0.717157287525381 - CpAxKy3WgAAAxrQ.jpg

[('n04296562', 'stage', 0.3357392), ('n03325584', 'feather_boa', 0.25848117), ('n04584207', 'wig', 0.07847353), ('n09229709', 'bubble', 0.048865814), ('n02787622', 'banjo', 0.043072842)]



Result #4 - 0.717157287525381 - CoduKraWAAEo_c7.jpg

[('n04487394', 'trombone', 0.50139934), ('n03110669', 'cornet', 0.12052086), ('n04296562', 'stage', 0.05157364), ('n09229709', 'bubble', 0.051457062), ('n04141076', 'sax', 0.048207816)]



Result #5 - 0.717157287525381 - Co4QKQsXgAA1MOm.jpg

[('n04418357', 'theater_curtain', 0.33725467), ('n04296562', 'stage', 0.13604), ('n10148035', 'groom', 0.10124001), ('n03529860', 'home_theater', 0.08219733), ('n09229709', 'bubble', 0.028398389)]



RESULTS 6 - HoG detected shapes and thus has more probability of finding similar/relevant images

Query image: 766179905533083648.jpg



Result #1 - 0.0 - CqIEjWKWAAA-ekg.jpg



Result #2 - 0.5142918061834583 - Cp_fKvoXgAEFJzU.jpg



Result #3 - 0.5358629838936944 - Cp-Y4xWXEAApWW.jpg



Result #4 - 0.582183210184426 - CrWXUbQWAAAOfK.jpg



Result #5 - 0.6022628588888223 - CpLmLhcWEAA-r0R.jpg



RESULTS 7 - CNN incorrect tagging affects the relevance of the retrieved images

Query image: 759764821415956480.jpg

Tags: [[('n02894605', 'breakwater', 0.11192308), ('n03933933', 'pier', 0.09661752), ('n04366367', 'suspension_bridge', 0.07240466), ('n03216828', 'dock', 0.049584355), ('n03598930', 'jigsaw_puzzle', 0.0472002)]]



Result #1 - 0.29289321881345254 - Cos6E8YXgAA8X87.jpg

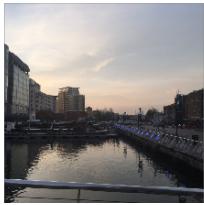
Tags: [[('n02894605', 'breakwater', 0.111922555), ('n03933933', 'pier', 0.09661752), ('n04366367', 'suspension_bridge', 0.07240488), ('n03216828', 'dock', 0.0495845), ('n03598930', 'jigsaw_puzzle', 0.04720029)]]



Result #2 - 0.434314575050762 - CyMtsjAW8AA9A3O.jpg

Tags: [[('n04366367', 'suspension_bridge', 0.13570675), ('n03933933', 'pier', 0.10936327),

('n04311004', 'steel_arch_bridge', 0.106035315), ('n03216828', 'dock', 0.087984785), ('n02980441', 'castle', 0.06759983)]



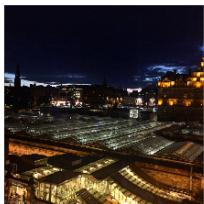
Result #3 - 0.434314575050762 - Cq3NxTTWAAzVEm.jpg

Tags: [('n04366367', 'suspension_bridge', 0.40200678), ('n03933933', 'pier', 0.12606986), ('n04311004', 'steel_arch_bridge', 0.11733803), ('n02980441', 'castle', 0.07452345), ('n03877845', 'palace', 0.064013086)]



Result #4 - 0.434314575050762 - CpINbyIWcAASlbz.jpg

Tags: [('n03933933', 'pier', 0.28328508), ('n03877845', 'palace', 0.10495462), ('n03216828', 'dock', 0.06533371), ('n02894605', 'breakwater', 0.046331145), ('n09332890', 'lakeside', 0.034038115)]



Result #5 - 0.434314575050762 - Cp7SwwYXYAAL0jl.jpg

Tags: [('n04606251', 'wreck', 0.128022), ('n02980441', 'castle', 0.10313934), ('n03216828', 'dock', 0.077452466), ('n04366367', 'suspension_bridge', 0.075387076), ('n02894605', 'breakwater', 0.050696254)]

