
IT-309
EXPERIMENT- 09 REPORT

Indian Institute of Information Technology, Vadodara

Group-1

Contents

0.1	GloVe	2
0.2	Results	2

0.1 GLoVe

- GloVe is an unsupervised learning algorithm for obtaining vector representations for words.
- Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.
- In this experiment, we have used glove_python library for GloVe implementation.
- For training data, we have used 100 MB of cleaned text from <http://mattmahoney.net/>

0.2 RESULTS

Results for some of the queries are shown below:

- 10 words similar to 'dog'
 - ['laekenois', 0.8501926604345059),
 - ('tervueren', 0.8265487259476645),
 - ('dogs', 0.7430731150922689),
 - ('shepherd', 0.7420534035160864),
 - ('cat', 0.734242741275883),
 - ('malinois', 0.7194000746727076),
 - ('breed', 0.7042221745297993),
 - ('horse', 0.6864767930831852),
 - ('hound', 0.6769279887757707),
 - ('dead', 0.6765494209730863)]
- 10 Words similar to 'color'
 - ['internegative', 0.88609807390083),
 - ('colour', 0.8709920265673742),
 - ('blindness', 0.8038853044557297),
 - ('colors', 0.7411262460826074),
 - ('blue', 0.7328939590459189),

('yellow', 0.6998491110297039),
('palette', 0.6934651456733099),
('eeeeee', 0.6820967318784291),
('dddddd', 0.6705975298945843),
('screen', 0.6629410609385425)]

- 10 Words similar to 'computer'

[('confetti', 0.8254355716409304),
('computers', 0.8020718290370966),
('hardware', 0.7878534316364587),
('laptop', 0.751452943701868),
('graphics', 0.7301269196266198),
('pioneers', 0.724419181350407),
('desknote', 0.7202339966978349),
('programmer', 0.719783172116936),
('aided', 0.7149608577202161),
('simulation', 0.7104328576533571)]