

Goals

- Analyze Disney movie scripts and their original stories to find interesting relationships
- Create a book/movie recommender for kids and adults





Methodology

Gather & Clean Data

(17 movie scripts & 17 story texts)

Tokenize

(word level, Tf-idf, remove punctuation, stop-words lemmatize, POS)

Tokenize

(line/sentence level)

NMF

Word2Vec& PCA

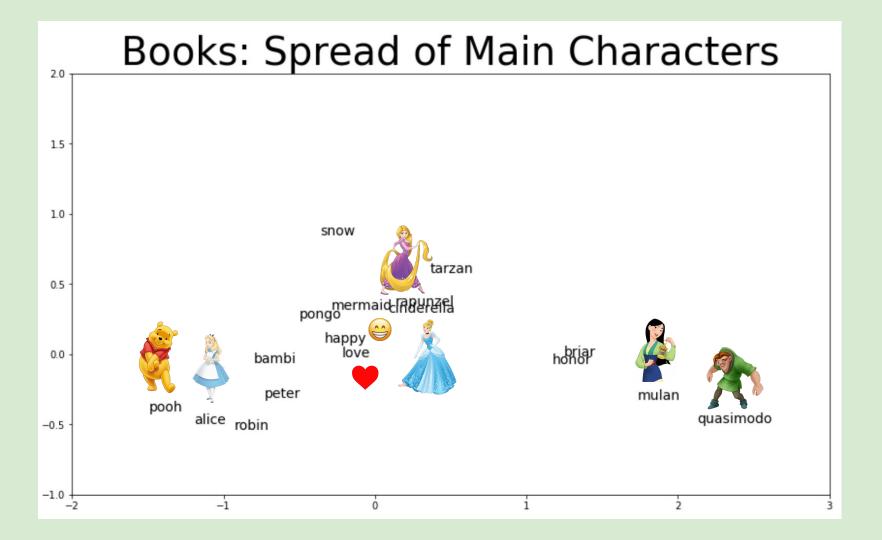
Sentiment Analysis

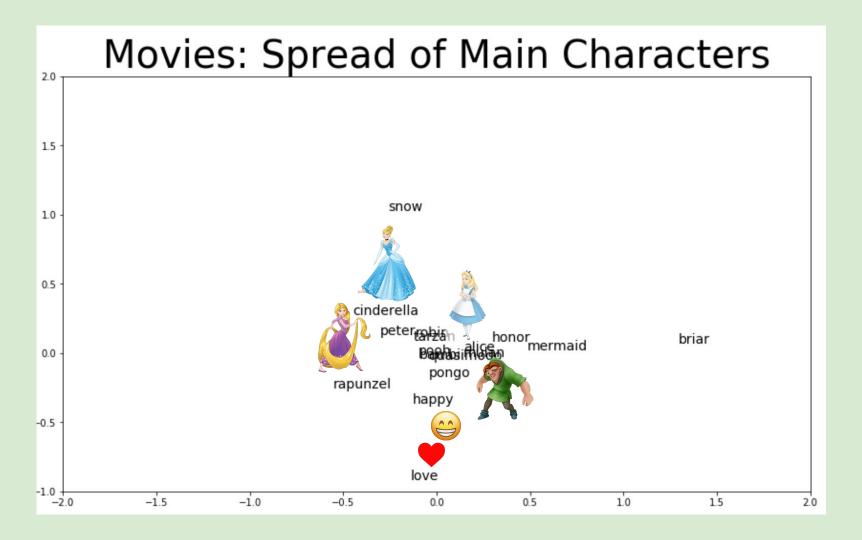
(vader with DTW)

Flask App: Book/Movie Recommender

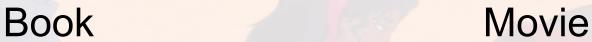
(Cosine similarity & DTW distance)

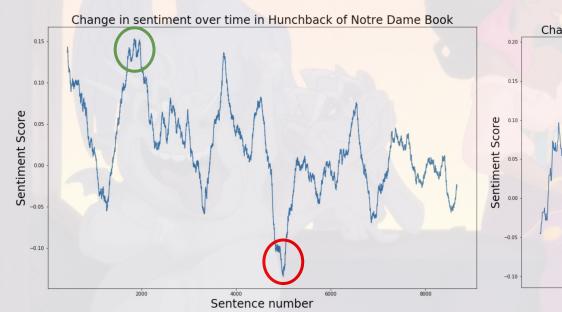


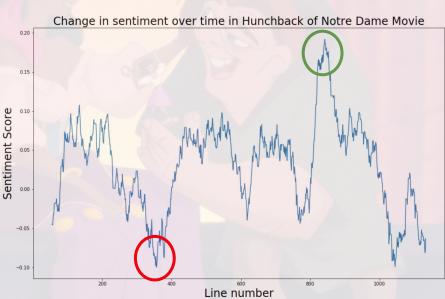




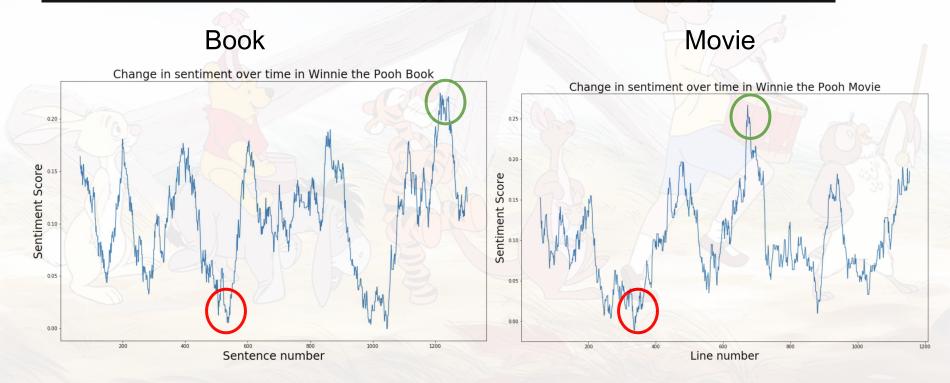
<u>Hunchback of Notre Dame Book/Movie DTW distance = 33</u>







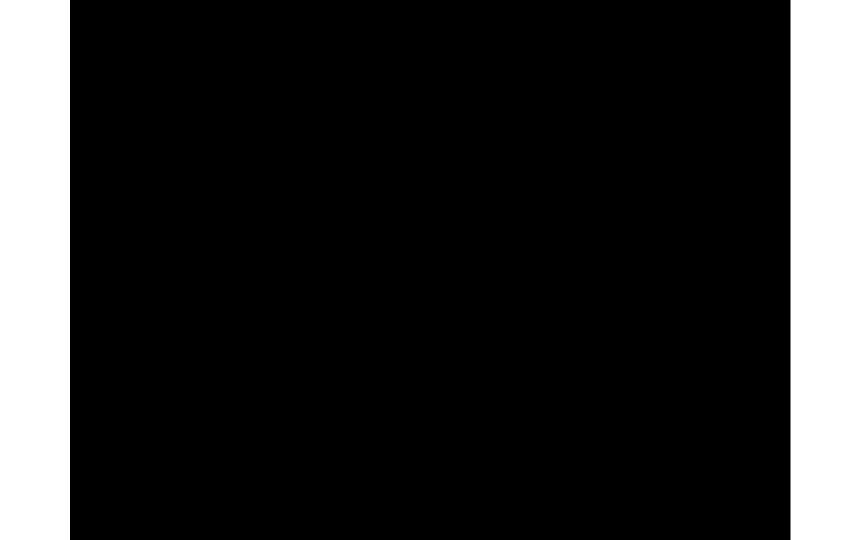
Winnie the Pooh Book/Movie DTW Distance = 7



Recommendation system

Takes into account:

- Dynamic time warping distance of each book/movie against book/movie
- Cosine similarity of each book/movie against book/movie using their topic weights across 34 topics (17 movies + 17 books)
- 3. How much user cares about topic similarity and sentiment similarity



"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."

Lewis Carroll, Alice in Wonderland



Appendix A: Topics

Topic 0: gypsy, notre, topsy, turvy, festival, guasimodo, dame, dy, bell, frollo, sanctuary, eleison, captain, irae, guasi Topic 1: pooh, piglet, robin, honey, rabbit, christopher, bear, balloon, bee, dum, oh, tigger, hum, rain, owl Topic 2: peter, wendy, michael, nana, mr. john, pirate, nursery, hook, boy, shadow, child, pan, mother, night Topic 3: pongo, roger, puppy, cruella, music, anita, chuckle, perdy, colonel, oh, horace, vil, barking, jasper, park Topic 4: bambi, mother, gobo, faline, meadow, stag, leaf, bush, time, air, little, thicket, deer, hare, ena Topic 5: tarzan, ape, clayton, jungle, cabin, great, tribe, little, man, black, philander, tree, body, captain, hand Topic 6: thou, robin, sheriff, thy, hood, thee, john, stout, good, merry, yeoman, man, ale, little, tinker Topic 7: cinderella, qu, jag, zuk, pick, drizella, ash, step, anastasia, ucifee, lentil, maiden, duke, hazel, pigeon

Appendix B: Topics continued...

Topic 8: pinocchio, marionette, geppetto, poor, mastro, fox, piece, cricket, fairy, little, pi, voice, hand, wig, good Topic 9: princess, maleficent, gift, phillip, uh, stefan, hail, sixteen, hubert, oh, merryweather, dear, majesty, flora, kingdom Topic 10: snow, white, glass, gueen, look, huntsman, little, dwarf, liver, thou, woman, bed, cottage, envy, mug Topic 11: ariel, sebastian, flounder, triton, eric, scuttle, grimsby, max, ursula, shark, human, jetsam, sea, flotsam, daddy Topic 12: aladdin, sultan, magician, genie, lamp, vizier, princess, mother, son, palace, fruit, slave, ring, stone, jewel Topic 13: alice, mouse, little, rabbit, hatter, caterpillar, cat, duchess, thing, way, large, time, dinah, dormouse, dodo Topic 14: rapunzel, mother, satchel, lantern, birthday, flynn, gothel, flower, dream, hair, tower, pascal, okay, blah, glow Topic 15: beauty, beast, merchant, monster, father, sister, good, palace, great, poor, oblige, daughter, horse, tear, endeavor Topic 16: mulan, fa, family, ancestor, dragon, mushu, emperor, aaah, hun, uh, imperial, honor, army, dishonor, luck

Appendix C: Topics continued...

Topic 17: gaston, belle, papa, girl, philippe, town, provincial, beast, maurice, good, monsieur, oh, master, castle, book Topic 18: sea, mermaid, water, ship, sister, prince, beautiful, little, wave, blue, flower, sun, fish, large, surface Topic 19: merchant, sister, fortune, beauty, country, house, great, pride, guality, young, humility, daughter, poor, town, father Topic 20: It, lamp, jasmine, prince, abu, jafar, princess, sultan, ali, don, suitor, street, If, diamond, rat Topic 21: gringoire, saint, paris, cardinal, archdeacon, jehan, edifice, rue, la, gypsy, poet, master, eye, century, architecture Topic 22: heigh, ho, ah, doc, dig, grumpy, dwarf, white, snow, twitter, sneezy, hum, tsk, whistle, dont Topic 23: enchantress, rapunzel, rampion, garden, wife, hair, braid, descend, window, tower, husband, king, son, tress, desire Topic 24: missis, pongo, puppy, nanny, dalmatian, dog, dearlys, pup, butler, cruella, mr, spaniel, perdita, kitchen, car Topic 25: robin, oo, hood, sheriff, sire, tax, john, hiss, prince, uh, marian, birthday, lally, oh, nottingham

Appendix D: Topics continued...

Topic 26: rabbit, late, oh, uh, mi, oyster, unbirthday, goodness, mr, cabbage, monster, ann, mary, world, curious Topic 27: pinocchio, conscience, uh, figaro, jiminy, little, actor, boy, diddle, dee, real, pinoke, string, ho, cricket Topic 28: spindle, king, hedge, old, year, briar, castle, woman, prick, thorn, sleep, gueen, room, unspoken, princess Topic 29: drip, april, bird, drop, prince, bambi, lt, little, meadow, hello, morning, thumper, good, shower, sound Topic 30: tarzan, gasp, terk, jane, growl, piranha, coo, grunt, gorilla, kala, kerchak, anh, chitter, sniff, snarl Topic 31: brilliance, return, hometown, khan, sister, battle, fleet, merit, spin, gong, warrior, camel, sentry, steel, yan Topic 32: peter, hana, ganda, smee, mana, pan, captain, tee, wendy, teedle, pirate, aye, ee, hook, oh Topic 33: market, river, parent, yellow, dusk, dawn, saddle, rumble, sound, bridle, yan, steed, blanket, whip, leave