

DataReaderScript

February 10, 2020

1 DataReader

1.1 Reads the data from downloaded <https://u.cs.biu.ac.il/~koppel/BlogCorpus.htm>, parses and saves to csv

```
[10]: import os
import pandas as pd
from xml.dom import minidom
from bs4 import BeautifulSoup

#reading data

path= 'Data/blogs/'
#This reader requires the blogs to be downloaded from
#https://u.cs.biu.ac.il/~koppel/BlogCorpus.htm and placed the the stated path
data=[]

with os.scandir(path) as entries:
    for e in entries:
        Filename = e.name

        if(Filename.__contains__(".male")):
            label = "male"
        else:
            label = "female"

        text = []
        # parse an xml file by name
        f= open(path+Filename, 'rb', encoding=None).read().decode('utf8',
↪ 'ignore')

        soup = BeautifulSoup(f, 'html')
        #reading blogs and adding them to a list
        posts = soup.find_all('post')
```


Darkstalkers, King of Fighters, Art of Fighting, and Fatal Fury series\' all playable, you just know that this fighter\'s gameplay is going to be varied like that of no other. Street Fighter Anniversary Collection - September 28, 2004: A compilation of the overly-hyped(?) Street Fighter III: Third Strike, as well as the also overly-hyped(?) Hyper Street Fighter II which is if I\'m not mistaken, a mix of Street Fighter II, Street Fighter II Champion Edition, Street Fighter II Hyper Fighting, Super Street Fighter II: The New Challengers, and Super Street Fighter II Turbo. Dead or Alive Ultimate - October 4, 2004: I know that many people feel that the Dead or Alive series isn\'t that great, at least compared to the almighty Virtua Fighter (or if you\'re into that Tekken crap), and that it\'s mainly about the "graphics," but this is the only 3-d fighter (ok, Soul Calibur is pretty good too I suppose) that I\'ve really gotten in to, although I\'ll admit, I haven\'t given Virtua Fighter a fair chance to this day. Dead or Alive Ultimate is a compilation of both Dead or Alive (originally released on the SEGA Saturn) and Dead or Alive 2 (originally released on the SEGA Dreamcast), which is supposed to be upgraded with improved graphics, 8 new fighting environments of which are supposed to have more focus on interactivity, as well as newly added cut-scenes about each characters\' past, that go deeper into the Dead or Alive storyline. And of course, what would a DOA game be without tons of new costumes to unlock? To quote IGN\'s X-Box division: The first disc will features an online version of 1997\'s original Dead or Alive while Disc 2 will include the enhanced version of Dead or Alive 2, also with online capabilities. But the enhancements here are what compel us to look at this DOA2 as the logical progression of the fighting series and a leap from DOA3. There are all new tag team moves in DOAU and the same ally-enemy system that makes team moves between friends more dramatic than team moves between enemies. There will be more of an emphasis on rolling and dodging and new countermoves and reversals in DOAU as well. While DOA3 had us fascinated with multilevel combat as we got tossed from rooftops and never ending plateaus, DOAU will have multilevel fighting as well as what Team Ninja refers to as "slope technique". In the trailers, you\'ll recognize these as the sequences when Gen-Fu is sent rolling down the stairs after being tossed by Lei Fang in the official trailer for the game. Whether it\'s a short flight of stairs or a little hill that leads to a tiny pond in the wilderness, the slope technique should be a nice twist on the multilevel combat concept. All characters can use throws to launch enemies down slopes but each character will also have a special slope technique move that should be especially humiliating. You\'ll recognize this as the sequence where Zack uses Jann Lee as a snowboard and rides him down an icy slope.

Ayane! :D ...Yeah ok, I went a little overboard on the DOA coverage... Ultra Bust-A-Move - October 4, 2004: It\'s Puzzle Bobble Online, what\'s there to say except for "frickin\' awesome!" All of the above are to have X-Box Live support. The reason why I\'m even mentioning these games is because that I feel that they will justify purchasing an X-Box Live account in the future even more-so if you\'re not interested in much of the current line-up (which seems to understandably be an issue with fighter and RPG fans), which in itself is quite congenial, just lacking in the fighters/RPG\'s department, especially with the cancellation of True Fantasy Online. I would have made a lot of the above into links, but it\'s late and I\'m tired, so

expect another update tomorrow with more information.\r\n \r\n \r\n',
\r\n\r\n \r\n can shove it.\r\n \r\n \r\n', '\r\n\r\n\r\n\r\n
\r\n As you can see to your right, I have added a new urlLink hit counter
to the page. It times out after one hour so that a non-unique user will
register as another hit after one hour since their last visit is up. My IP
address is blocked as well so that my views of this page do not register. The
counter has been officially in place since approximately 5:32 P.M. Eastern time,
Sunday, June 13, 2004.\r\n \r\n \r\n', '\r\n\r\n\r\n \r\n Here's a
little urlLink post that I just wrote up for a thread in urlLink Neo-
Geo.com's Unrelated Topics urlLink forum , discussing the stupidity of
today's gamers and consumers (the topic at the time however was about Nintendo
GameCube's poor sales statistics and why people prefer other consoles): Well
at least nobody here said "OMG ASL GAMECUBE HAHA KIDDY GAMEZ
LOLLOOLOLOLOLOLLZZZZZZZZZZZZ!!!!!!11111" I hate that mentality, and every
kid my age (and really all other age-groups as well) uses that as an excuse to
make it sound like it's the "cool," trendy" thing to own a Playstation 2. I
don't get it. If you even mention that you own anything that's not a
Playstation 2, from the current era that we live in or another one, they proceed
to fight with you even if you were just trying to have a normal conversation.
Their argument is always about the "graphix!" Yet I prove to them that GameCube
and X-Box are superior hardware-wise (games are opinion-based, but there's a
limit on that too), but it doesn't matter. One kid actually believed that the
Playstation 2 was "magical" and the reason why games don't look as good on it
as they do on Dreamcast (this was a couple years back) was because it was so
powerful that developers just couldn't handle the insane amounts of power it
could output at once! This however, did not matter, since even though the games
didn't "look" as good, they still have "better graphix0rz!" To this day I
don't get it, I just don't get it. A couple months ago I mentioned Dreamcast
to not_sonic when we were talking about I don't know what. Immediately some
ass-hat behind us (this was in school) chimed in and ranted about how much
Dreamcast sucks. This would happen everytime the word "Dreamcast" came into
play in a conversation and he overheard it. It's like these people have a
fucking radar and whenever they hear the word "Dreamcast," (it's not as bad as
GameCube or X-Box, but they're up there to on the hatred chart) they take it
offensively and start bitching at you. It's like liberals and President George
W. Bush. I mean seriously, this guy wasn't even in on the conversation, just
stay the fuck away if you don't like what we're talking about. It's almost
like a poor attempt at making a current, mainstream system into a "cult" system
if you look at it from the perspective that I've been seeing it at, which is
pretty pathetic. However, this does not seem to be far from the truth,
really... Like I said before, all the "cool," "trendy" kids have Playstation
2's and maybe their little brother/sister owns a GameCube since they're
younger than 7 years old. It's like a fad that never dies... Even Pokemon
died, geeze. Another interesting tidbit that I'd like to point out is that
most of the Sony fan-base consists of the most gullible, misinformed shit-heads
that I've ever had the misfortune to be in a conversation with. You can tell
them [i]anything[/i] and they'll believe it: Asshat 1: "Dude, did you know
that the Playstation 3 is coming out tomorrow, followed by the Playstation 9

next week!? It\'s going to be soooooooooo siiiiiiiiiiiiiiiiiiick!!!!!!!!!!!!11".

Asshat 2: "D00d are you serious!? OMG ASL I\'ve been all of my money for this day! Did you know that the Playstation 3 will take advantage of other Playstation 3s that aren\'t in use over the internet to make it more powerful!?", Asshat 1: "Dude, nooooooooooooo way!!!!!!!!!!!!111 OMG!!!!!!!!11"

Asshat 2: "It also can pump out 2390482390489234902380423420923423904 polygons, whatever that means!!!!!!!!!!!!11 It also uses *says really slow* C...E...E...L...L...L...T...E...C...H...N...O...L...O...G...Y...

WHATEVER THAT IS, I DUNNO, BUT IT SOUNDS FUCKING AWESOME SO IT\'S BETTER THAN EVERY OTHER SYSTEM, ROXORZZZZZZ!!!!!!!!!!!!11" Asshat 1: "OMG I just read that It\'s going to have Unreal Tournament 2007, Final Fantasy 13 and Super Mario Eclipse on launch, which is tomorrow! I can\'t wait!!!!!!!!111" Asshat 2: "HOLY SHIT YOU MEAN NINTENDO WAS BOUGHT OUT BY SONY!!!!!!!!1111?" Asshat 1: "YESSSSSSSSSSSS THAT\'S BECAUSE NINTENDO AND [i]ALL[/i] THEIR GAMES SUCKED SO MUCH THAT THEY NEEDED SONY TO LIVE ON!!!!!!!!11 SO NOW WE GET TO PLAY MARIO GAMES ON PLAYSTATION!!!!!!!!11 [ok, so why do they want to play Super Mario Eclipse if [i]all[/i] Nintendo games suck so much?] And the best part is, since Sony bought out Nintendo, Playstation 3 is backwards compatible with GAMECUBE, NES, GAMEBOY ADVANCE, TURBO DUO AND SEGA MASTER SYSTEM!!!!!!!!11" Asshat 2: "HOLY FUCKING SHIT, AWESOME!!!!!!!!111 Wait, what are Master System and Turbo Duo?"

Asshat 1: "I don\'t know, somebody was talking about them though so PS3 has got to be backwards compatible with it since nobody would talk about anything but Playstation 3 today." Asshat 1 and 2 together: "...AWESOME!!!!... SEGA SUXORZ!!!!!!!!11" Informed Person: "You\'re both retards, everything that\'s come out of your mouths for the past 5 minutes is about as accurate as GameFAQs\' rumor mill." Asshat 1 and 2 together: "OMG YOU POST AT GAMEFAQS TOO!?" Informed Person: "No, god-dammit!" Asshat 1 and 2 together: "FUCK YOU SONY HATER!!!!11 Go away now, before we beat you with our old PSONES!"

Informed Person: "...I\'ll be more than happy to..." I could go on with that conversation, but I think you get the idea. But to continue with how mis-informed people are, you can take my Dad\'s store for example. He works in a camera store and he has trouble with people not buying certain products for stupid reasons. Someone one might want a camera with a certain feature. He has a camera with that feature, but they insist that he sells them the Sony version of the camera which either doesn\'t exist, or costs more and does less. The customer will then refuse to buy the Samsung version of the camera which is more cost efficient and has more features/is of a higher quality. The customer will then either purchase the Sony product or nothing at all if a Sony version doesn\'t exist... or does it exist? Supposedly they just saw an ad in the newspaper for a Sony camera that some store is clearly selling even though it\'s not due out until 2005. Moving back to the Playstation 2 topic, let\'s not forget that all these people own their PS2\'s for is Grand Theft Auto 3/Vice City and Gran Turismo 3, nothing else (there\'s the occasional Final Fantasy X/X-2). What I don\'t understand is, is why everyone always raves about how great the "graphix0rz" are in the GTA series. GTA is probably the worst looking game on the Playstation 2... And the game is so shitty that it\'s practically a fact that it sucks (sorry, I had to say that, maybe the element that I\'m surrounded by everyday in school is rubbing off on me, but with a different

opinion). It's awesome for comedy though, running around blowing shit up and adding narration to it is definitely great (with other people that is). Sorry I went off on a rant there, I know this has been re-iterated a thousand times over, but the truth is people just plain suck at life, facts, and purchases. I really believe that not only is this mentality the cause of shitty sales for Nintendo, but the cause of a lot of problems in this world. I don't know about the rest of the world, but from what I've seen with the morons around me, they'll become biased to a particular company (usually Sony) and then claim that their products are the best, so they only buy camcorders, cameras, CD players, etc., that are Sony products, and Sony products only, and in the case of Sony, their products mysteriously stop working after 6 months, which happened to a friend of mine when she got her Sony CD player... She wasn't biased though, the CD player just happened to be made by Sony... It's this biased and misinformed society that we live in that causes the downfall of many products/companies/businesses these days. Disclaimer: I'm using Sony as the company that people are biased towards since it seems to be the number one company for people to cling to, not because of my hatred for Sony... that I don't have... . Keep in mind that this happens with all companies and that Sony fans aren't the only people to blame. And remember, "kiddy" games suck and PS2 Ownz j00 all fo' fr33! Disclaimer 2: I did not proof-read this post and I skipped around as I wrote it. I apologize if this was the most inconsistent post that you've read since Japhei was a member here. Edit: I forgot to talk about the "new is better" philosophy. But basically to show the hypocrisy behind that, at least with the Sony fan-based. Just ask yourself (and them) why they're so fond of Playstation 2 if X-Box and GameCube are both approximately 2 years younger? The "graphics mean everything" mentality I don't understand either. I've actually had people (mainly Playstation 2 owners) tell me that "graphics" are all matters/impresses wthem with games. If that's the case, why are they wasting their time playing video games and not just watching tech demos/CG movies instead? I've only scratched the surface, but I'm done for today. I apologize if that was kinda crappy, but the issue with stupid people in today's society needs to be addressed... Or does it really? We can't change people and their flaws that mess up the world, so why waste time talking about it? I really don't know.\r\n \r\n \r\n'])

```
[11]: #Putting the data into a data frame
dataFrame = pd.DataFrame(data)
dataFrame.head()
dataFrame.info
```

```
[11]: <bound method DataFrame.info of                                0
1
0      male  [\r\n\r\n      \r\n      Well this is my first...
1      female  [\r\n\r\n\t \r\n      Yes !      Finally, ...
2      male  [\r\n\r\n      \r\n      Yes my overworked body...
3      male  [\r\n\r\n      \r\n      tennis has always been...
4      female  [\r\n\r\n\t \r\n      hello. How are you? I'm ...
```

```

...
19315    male  [\r\n\r\n  \r\n      They ask her; "what ty...
19316    male  [\r\n\r\n\t \r\n      urlLink Modern Disciple...
19317   female  [\r\n\r\n  \r\n      The last day of the ye...
19318   female  [\r\n\r\n  \r\n      \r\n      LiveJournal ...
19319   female  [\r\n\r\n\t \r\n      Those of you who have ne...

```

```
[19320 rows x 2 columns]>
```

```
[12]: dataframe.shape
```

```
[12]: (19320, 2)
```

```
[63]: data.shape
      #saving the data to csv
      data.to_csv(r'Data\Gender_Blog.csv')
```

```
[ ]:
```

```
[ ]:
```