

Nagyadat Coolmini

Gesztesy Tamás, Nemkin Viktória

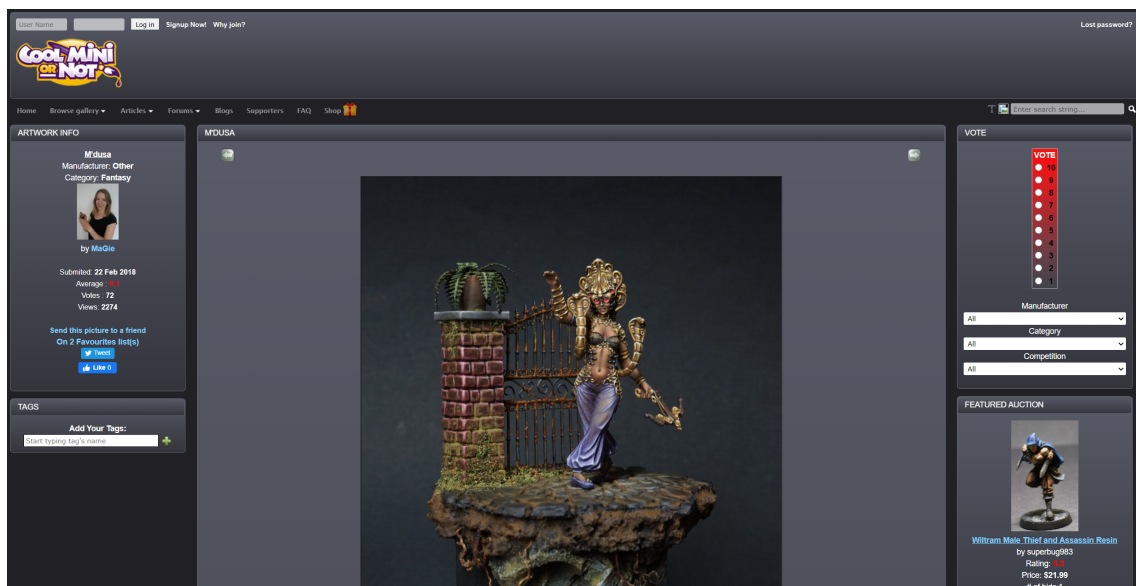
1. Bevezetés

Az elkészült programkódok és adathalmazok elérhetőek a <https://www.github.com/nemkin/cool-mini-or-not> webcímen.

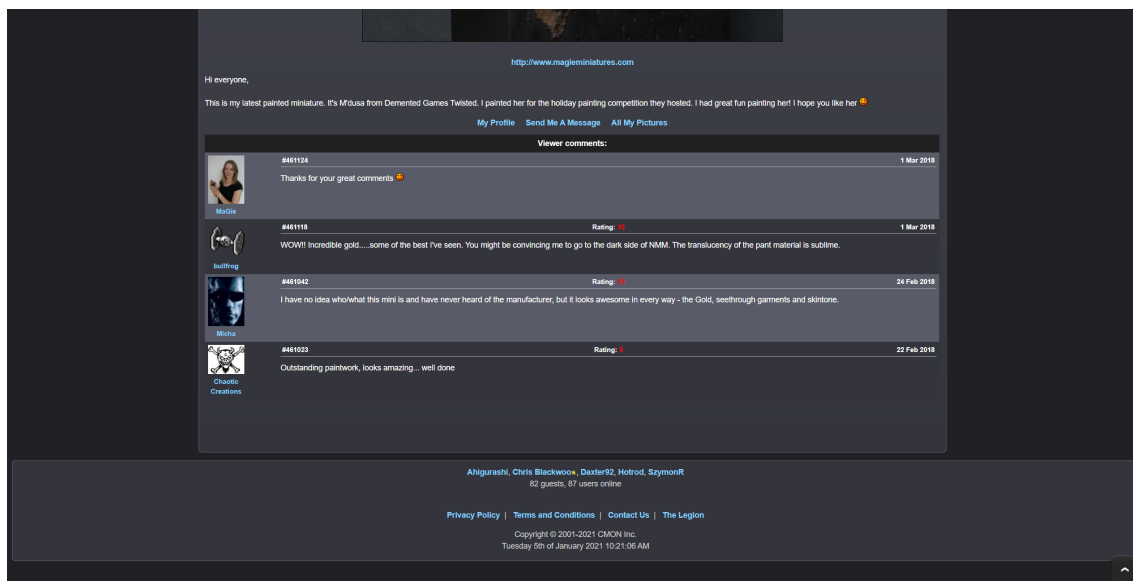
2. Feladat

A félév során a Cool Mini Or Not weboldal (<http://www.coolminiornot.com>) adatelemzésével foglalkoztunk.

Az Cool Mini Or Not weboldal egy figurafestésekre specializálódott közösségi oldal, ahova a kifestett miniket lehet feltölteni, a többi felhasználó pedig szavazhat és kommentelhet a fotóinkra.



1. ábra. Coolmini feltöltés



2. ábra. Coolmini feltöltés

3. Adathalmaz letöltése

Az első feladat a weboldal adatainak a kinyerése volt. Ez általában egy nehézkes és körülményes feladat. Nagy méretű adathalmazról van szó, több 10 GB-nyi adat, melyet napokon keresztül tudunk csak letölteni. Általában a weboldalak üzemeltetői nem szeretik az ilyen jellegű letöltéseket, ezért előfordulhat, hogy blokkolják az IP címünket (például DDOS támadásnak hiszik a letöltést).

Szerencsére azonban a weboldal tartalma archiválva lett a Wayback Machine által (https://web.archive.org/web/2020*/coolminiornot.com), ehhez pedig már létezik letöltő script (<https://github.com/hartator/wayback-machine-downloader>).

A letöltő scriptet Raspberry Pi-n keresztül futtattam, mivel összesen 2 héten keresztül tartott a letöltés.

Sajnos problémát jelentett, hogy bizonyos régebbi bejegyzések rossz formátumban töltődtek le az oldalról, ezért egyéb letöltési lehetőségekkel is próbálkoztam (wget), azonban ezekben az esetekben a felhasználók kommentjeit nem lehetett letölteni, ezért végül az eredeti adathalmazt használtam fel, a hibás fájlok kiszűrésével.

4. Transzformáció

A következő feladat az adatok kinyerése volt a HTML fájlokból. Ehhez az lxml nevű Python modult használtam, melynek segítségével XML fájlokat lehet parsolni. A következő python program segítségével készítettem el az adathalmazt:

```

1 import sys
2 import os
3 import pandas as pd
4 import helper
5 import re
6 import traceback
7 from datetime import datetime as dt
8
9 from lxml import html
10 from lxml import etree
11
```

```

12 location = "/home/nemkin/Mount/drive/Wayback"
13 #location = "/home/nemkin/Mount/drive/Wget/www.coolminiornot.com"
14 index_file = "index.html"
15
16 submissions = []
17 ebay = []
18 comments = []
19
20 def strip(l):
21     return list(map(lambda s: s.strip(), l))
22
23 def get_entries():
24     global location
25     dirs = os.listdir(location)
26     entries = list(filter(lambda dir: dir.isnumeric(), dirs))
27     return entries
28
29 def process_comments(file_path, entry_id, tree):
30     global comments
31
32     comment_rows = tree.xpath('//table[@class="comments_list"]/tbody/tr')
33
34     for comment_row in comment_rows:
35         [commenter_td, comment_text_td] = comment_row.xpath('td')
36
37         [commenter_url, *_] = commenter_td.xpath('.//a/@href')
38
39         [comment_id] = comment_text_td.xpath('.//td[@class="comment_header_left"]/b/text()')
40         try:
41             [vote] = comment_text_td.xpath('.//td[@class="comment_header_center"]/b/span/text()')
42         except ValueError:
43             vote = ''
44         [comment_date] = comment_text_td.xpath('.//td[@class="comment_header_right"]/b/text()')
45         comment_date_parsed = dt.strptime(comment_date, "%d %b %Y").strftime("%Y-%m-%d")
46
47         comment = " ".join(strip(comment_text_td.xpath('text()')))
48         comment_sanitized = re.sub(' {2,}', ' ', comment.replace('\n', ' '))
49
50         _, commenter_user = commenter_url.split('?')
51
52         [commenter_user_id, *_] = commenter_user.split('-')
53         commenter_user_name = commenter_user[len(commenter_user_id)+1:]
54
55         comment_columns = [
56             entry_id, comment_id[1:], comment_date_parsed, commenter_user_id,
57             commenter_user_name, vote, comment_sanitized]
58
59         comments.append(strip(comment_columns))
60
61 def process_entry(entry_id):
62     global location
63     global index_file
64     global submissions
65     global ebay
66
67     dir_path = f"{location}/{entry_id}"
68     main_file_path = f"{dir_path}/{index_file}"
69
70     try:
71         with open(main_file_path, "r") as main_file:
72             main_page = main_file.read()
73     except IOError as e:
74         print(f"IoError during opening main file path: {e}")
75         print()
76         return
77
78     try:
79         tree = html.fromstring(main_page)
80     except:
81         print(f"Can't get HTML tree for {entry_id} with page {main_page}")

```

```

82     print()
83     return
84
85 boxes = tree.xpath('//div[@class="box"]')
86 [artwork_info, _, submission, *_] = tree.xpath('//div[@class="box"]/div[@class="block"]')
87
88 try:
89     [entry_name] = artwork_info.xpath('b/u/text()')
90 except Exception as e:
91     print(f"Error during name parse: {e}")
92     entry_name = ""
93 entry_name = entry_name.replace('; ', ',')
94
95 [manufacturer, category] = artwork_info.xpath('b/text()')
96 [user_id, _] = artwork_info.xpath('a[@class="userProfile"]/@uid')
97 user_name = artwork_info.xpath('a[@class="userProfile"]/text()')[-1]
98
99 [entry_date, vote_count, view_count] = artwork_info.xpath('span[@class="socialLink"]/b/text()')
100 entry_date_parsed = dt.strptime(entry_date, "%d %b %Y").strftime("%Y-%m-%d")
101
102 [vote_average] = artwork_info.xpath('span[@class="socialLink"]/b/font/text()')
103
104 [entry_image_url] = submission.xpath('//img[@id="artworkIMG"]/@src')
105
106 entry_image = entry_image_url
107
108 ebay_forms = submission.xpath('./form')
109 if len(ebay_forms) == 1:
110     [ ebay_form ] = ebay_forms
111     [ ebay_id ] = ebay_form.xpath('./a/text()')
112     ebay_data = ebay_form.xpath('./b/text()')
113
114     if len(ebay_data) == 4:
115         [ ebay_current_price, ebay_first_bid, ebay_number_of_bids, ebay_end_date ] = ebay_data
116         ebay_end_date_parsed = dt.strptime(ebay_end_date, "%Y- %m-%d %H:%M:%S").strftime("%Y-%m-%d
↵ %H:%M:%S")
117     else:
118         [ ebay_current_price, ebay_first_bid, ebay_number_of_bids, ebay_end_date ] = [""]*4
119         ebay_end_date_parsed = ""
120
121     ebay_columns = [ entry_id, ebay_id, ebay_current_price, ebay_first_bid, ebay_number_of_bids,
↵ ebay_end_date_parsed ]
122     ebay.append(strip(ebay_columns))
123
124 else:
125     if 1 < len(ebay_forms):
126         print(entry_id, " multiple ebay forms?")
127
128 columns = [
129     entry_id, entry_date_parsed, entry_name, entry_image, user_id, user_name, manufacturer,
↵ category,
130     view_count, vote_count, vote_average
131 ]
132
133 submissions.append(strip(columns))
134 process_comments(f"{entry_id}/{index_file}", entry_id, tree)
135
136 dir_contents = sorted(os.listdir(dir_path))
137 for dir_content in dir_contents:
138     sub_dir_path = f"{dir_path}/{dir_content}"
139     if os.path.isdir(sub_dir_path):
140         sub_file_path = f"{sub_dir_path}/{index_file}"
141         try:
142             with open(sub_file_path, "r") as sub_file:
143                 sub_page = sub_file.read()
144         except IOError as e:
145             print(f"IOError during subpage html file opening: {e}")
146             print()
147             continue
148         sub_tree = html.fromstring(sub_page)

```

```

149         process_comments(f"{entry_id}/{dir_content}/{index_file}", entry_id, sub_tree)
150
151     entries = get_entries()
152     results = []
153     bad_entries = open('../data/bad_entries.txt', 'a')
154
155     l = len(entries)
156     print(entries)
157     print(f"Total: {l}")
158     helper.printProgressBar(0, l, prefix='Progress:', suffix='Complete', length=50)
159     for i, entry in enumerate(entries):
160         try:
161             process_entry(entry)
162         except :
163             bad_entries.write(f"BADENTRY: {entry} ({i})")
164             print()
165             print()
166             print(f"BADENTRY: {entry} ({i})")
167             print()
168             traceback.print_exc()
169             print()
170     helper.printProgressBar(i+1, l, prefix='Progress:', suffix=f"Current: {entry}", length=50)
171
172     bad_entries.close()
173
174     submissions_df = pd.DataFrame(
175         data = submissions,
176         columns = [
177             "entry_id", "entry_date", "entry_name", "entry_image", "user_id", "user_name", "manufacturer",
178             ↪ "category",
179             "view_count", "vote_count", "vote_average"
180         ])
181     ebay_df = pd.DataFrame(
182         data = ebay,
183         columns = [
184             "entry_id", "ebay_id", "ebay_current_price", "ebay_first_bid", "ebay_number_of_bids",
185             ↪ "ebay_end_date"
186         ])
187     comments_df = pd.DataFrame(
188         data = comments,
189         columns = [
190             "entry_id", "comment_id", "comment_date", "commenter_user_id",
191             "commenter_user_name", "vote", "comment"
192         ])
193
194     submissions_df.to_csv("../data/cool_mini_or_not_submissions.csv")
195     ebay_df.to_csv("../data/cool_mini_or_not_ebay.csv")
196     comments_df.to_csv("../data/cool_mini_or_not_comments.csv")

```

A következő csv fájlokat hoztam létre:

Submissions A submissions fájl tartalmazza soronként 1-1 feltöltött mini adatait:

- entry_id
- entry_date
- entry_name
- entry_image: A képhez tartozó url cím.
- user_id
- user_name
- manufacturer
- category

- view_count
- vote_count
- vote_average

	A	B	C	D	E	F	G	H	I	J	K	L
	entry_id	entry_date	entry_name	entry_image	user_id	user_name	manufacturer	category	view_count	vote_count	vote_average	
1	0	385238	2015-08-30 Lands Shuttle X-Wing re-paint	http://www.coolminior.net/pics/pics16/img5e32b6d23b2f.jpg	7735	boyzie	Other	SF	3614	38	8.2	
2	1	363638	2014-08-20 B wing	http://www.coolminior.net/pics/pics16/img53f4c7d1b0707.jpg	3524	caiman	Other	SF	5537	88	8.9	
3	2	318068	2012-11-07 Twinblade	http://www.coolminior.net/pics/pics16/img509ab392d36cf.jpg	12372	Thorek	Pegaso	Fantasy	4061	122	8.8	
4	3	335291	2013-06-21 Mjagur, Jotunn Mierce Miniatures (Closeups)	http://www.coolminior.net/pics/pics16/img51c492b1cbb8.jpg	18462	Bohun	Other	Fantasy	38155	775	9.7	
5	4	340962	2013-09-11 Elder Falcon	http://www.coolminior.net/pics/pics16/img3206680a1897.jpg	25811	rene	Games Workshop	SF	5705	146	9	
6	5	268057	2011-01-01 The Last Stand of the Crimson Fists	http://www.coolminior.net/pics/pics16/img4d41d7907b4f0.jpg	24239	volomir	Games Workshop	SF	31514	1252	9.6	
7	6	323285	2013-05-12 Chaos Decimator Engine [Forge World] - more close-ups	http://www.coolminior.net/pics/pics16/img519055c862d29.jpg	16585	Ana	Games Workshop	SF	50026	854	9.7	
8	7	322082	2013-01-07 Mr. Hugo Tumblepot	http://www.coolminior.net/pics/pics16/img50ea811a8804.jpg	56626	Toolinger	Other	Fantasy	8644	245	9.3	
9	8	261927	2010-05-14 Drop of Wrath, Elder Wraithlord	http://www.coolminior.net/pics/pics16/img4c66cc751bcf5.jpg	7572	Gunamar Ironwill	Games Workshop	SF	10698	176	9.6	
10	9	321454	2012-12-28 Dark Elder dragon	http://www.coolminior.net/pics/pics16/img5d0d8e7a6b3a2.jpg	64470	Puzi	Games Workshop	SF	5109	96	8.6	
11	10	334903	2013-06-16 High Elf	http://www.coolminior.net/pics/pics16/img51be03136c3fe.jpg	57362	medows	Games Workshop	Fantasy	6167	100	8.7	
12	11	339705	2013-08-26 High Elf Dragonlord	http://www.coolminior.net/pics/pics16/img521b14b1d9767.jpg	24239	volomir	Games Workshop	Fantasy	11698	329	9.7	
13	12	301355	2012-03-19 KVLV or SLAANESH demonprince	http://www.coolminior.net/pics/pics16/img4966fe7d28711.jpg	28657	In The Middle	SF	Other	19127	311	9.3	
14	13	344580	2013-11-05 Beast of Teentech - Green-	http://www.coolminior.net/pics/pics16/img5276c114a07b3.jpg	21858	Raffaele Stumpo	Other	Fantasy	14903	395	9.6	
15	14	332609	2013-05-16 Run, you fool!	http://www.coolminior.net/pics/pics16/img5194f4d75f7d4.jpg	48315	Picster	Studio McVey	Fantasy	14177	169	9.4	
16	15	345453	2013-11-17 Undead King Bust	http://www.coolminior.net/pics/pics16/img52891f5f4bd19.jpg	9412	Mohand	Other	Fantasy	4899	145	9.2	
17	16	340434	2013-09-04 Meadow of Doom (Wraithknight and Plague Drone of Nurgle)	http://www.coolminior.net/pics/pics16/img5226d5cde6cct.jpg	72201	fantasygames.com.pl	Games Workshop	SF	40571	1899	9.7	
18	17	348886	2014-01-12 Tyrant Exorcist	http://www.coolminior.net/pics/pics16/img52d2bbcc4b5b8.jpg	58115	thefigurecorner	Games Workshop	SF	7506	107	8.6	
19	18	350821	2014-02-07 Filbot FDV4 Male with Railgun	http://www.coolminior.net/pics/pics16/img52f514d1c176c5.jpg	30784	Countersunk81	Other	SF	13935	171	9.4	
20	19	397472	2016-04-30 Hecaton Alakos Minotaur Contemptor Dreadnought	http://www.coolminior.net/pics/pics16/img57254a88a48ec.jpg	73957	We7	Games Workshop	SF	5767	181	9.5	
21	20	173155	2007-11-01 DWARF LORD WITH STANDARD BEARER.	http://www.coolminior.net/pics/pics16/img472a43b600acc.jpg	32535	ElvenH	Games Workshop	Fantasy	1274	95	7.9	
22	21	449807	2020-07-26 Saurus Warrior	http://www.coolminior.net/pics/pics16/img51fd1d61d61ad.jpg	172372	Szerk	Games Workshop	Fantasy	24	1	5.5	
23	22	173129	2007-11-01 Templar Knight XIII century	http://www.coolminior.net/pics/pics16/img472f9f233230f.jpg	30093	MAP	Other	Historical	2149	112	8.5	
24	23	173118	2007-11-01 Aragorn	http://www.coolminior.net/pics/pics16/img4729a2d61b0b7.jpg	25648	Skinskinner	Games Workshop	Fantasy	1139	88	7.5	
25	24	173112	2007-11-01 Conan the barbarian 54 mm (Schwarzenegger version)	http://www.coolminior.net/pics/pics16/img4729a9016b861.jpg	26709	Peppepe	Andrea	Fantasy	1730	189	9	
26	25	47110	2004-02-29 widowmaker	http://www.coolminior.net/pics/pics16/img404238bd950d.jpg	449	cyril	Privateer Press	Steampunk	22166	1162	9.6	
27	26	355242	2013-06-21 The Guy With The Killin' Stick: Horc Limbopper Low Life Minis	http://www.coolminior.net/pics/pics16/img51c3d77b77f09.jpg	15059	xredmenacex	Other	Fantasy	664	49	6.1	
28	27	348036	2013-12-30 Angry Nigunser: Worm of the Happy Plate Club Low Life Minis	http://www.coolminior.net/pics/pics16/img52b13c2a344a.jpg	15059	xredmenacex	Other	Fantasy	479	40	5.2	
29	28	350036	2014-01-26 One-eyed Croach: Croach Smellcaster Low Life Minis	http://www.coolminior.net/pics/pics16/img5265c13832d3c.jpg	15059	xredmenacex	Other	Fantasy	558	40	5.9	
30	29	337708	2013-07-25 Uncious Pooof: Smelven Smellentalist Low Life Minis	http://www.coolminior.net/pics/pics16/img51f0b4283251c.jpg	15059	xredmenacex	Other	Fantasy	564	50	6.5	
31	30	336373	2013-07-06 Trozgoxx the Lobe: Oofu Dementalist Low Life Minis	http://www.coolminior.net/pics/pics16/img51d7a77937f46.jpg	15059	xredmenacex	Other	Fantasy	713	48	6	
32	31	333437	2013-05-27 Grenadier Ruster Beast / Rust Monster	http://www.coolminior.net/pics/pics16/img51a370b3c0253.jpg	15059	xredmenacex	Grenadier	Fantasy	775	66	7.3	
33	32	315153	2012-09-29 Aerial Servant with Pizza	http://www.coolminior.net/pics/pics16/img506793f06e768.jpg	15059	xredmenacex	Ral Partha	Fantasy	601	53	5.8	
34	33	334272	2013-06-08 Black Hole Robots	http://www.coolminior.net/pics/pics16/img51b39e24e4def3.jpg	15059	xredmenacex	Other	SF	615	50	5.8	
35	34	311725	2012-08-09 Doctor Who - Peter Davison the Fifth Doctor	http://www.coolminior.net/pics/pics16/img5024134f3883a.jpg	15059	xredmenacex	Other	SF	1017	53	6.2	
36	35	304360	2012-04-25 Doctor Who - Patrick Troughton the Second Doctor	http://www.coolminior.net/pics/pics16/img498948e86aa4.jpg	15059	xredmenacex	RAFM	Fantasy	822	57	6.4	
37	36	299229	2012-02-20 Giant Tick - Asgard FM2 Vintage Mini	http://www.coolminior.net/pics/pics16/img44283186772b.jpg	15059	xredmenacex	Other	Fantasy	942	53	4.7	
38	37	329666	2013-03-29 Asgard Unicorn FM16	http://www.coolminior.net/pics/pics16/img51560a61457fb.jpg	15059	xredmenacex	Other	Fantasy	615	51	5.6	
39	38	299230	2012-02-20 Doctor Who - William Hartnell the first doctor	http://www.coolminior.net/pics/pics16/img4428318549540.jpg	15059	xredmenacex	RAFM	SF	915	58	5.4	
40	39	308885	2012-06-29 Doctor Who - Tom Baker the Fourth Doctor	http://www.coolminior.net/pics/pics16/img4428318549540.jpg	15059	xredmenacex	Other	SF	829	53	6.3	
41	40	389810	2015-11-27 Unbelliferan Carrot Man	http://www.coolminior.net/pics/pics16/img5658e1b3db9cc.jpg	15059	xredmenacex	Other	SF	464	35	6.1	
42	41	293268	2011-12-05 Wererat - Asgard PM4 Vintage Mini	http://www.coolminior.net/pics/pics16/img4ed46624c1395.jpg	15059	xredmenacex	Other	Fantasy	1441	53	5.9	
43	42	310219	2012-07-17 Grenadier Basilisk	http://www.coolminior.net/pics/pics16/img5005a638c84f5.jpg	15059	xredmenacex	Grenadier	Fantasy	981	63	6.2	
44	43	308397	2012-06-22 Asgard Wight FM51	http://www.coolminior.net/pics/pics16/img4ef4d6f83f588.jpg	15059	xredmenacex	Other	Fantasy	543	51	6.1	
45	44	374440	2015-02-18 Thundarr the Barbarian - Interloper Miniatures Wasteland Warriors	http://www.coolminior.net/pics/pics16/img54e53279fdefc.jpg	15059	xredmenacex	Other	SF	1478	37	5.8	
46	45	372707	2013-07-17 Heritage/Custom Cast Lesser Goblin with Bow	http://www.coolminior.net/pics/pics16/img51e74d3428351.jpg	15059	xredmenacex	Heritage	Fantasy	560	51	5.3	
47	46	370966	2014-12-28 Sleestak - Hissing Lizards from Interloper Miniatures	http://www.coolminior.net/pics/pics16/img4a0d838a6f72.jpg	15059	xredmenacex	Other	Fantasy	773	36	5.2	
48	47	360393	2014-06-28 Sangan Herodicta Repaint	http://www.coolminior.net/pics/pics16/img53aa3e34f4c0ab.jpg	15059	xredmenacex	Other	Fantasy	735	37	6.2	
49	48	360323	2014-06-25 Pandalope the Panderer - Low Life Miniatures	http://www.coolminior.net/pics/pics16/img53aa3e34f4c0ab.jpg	15059	xredmenacex	Other	Fantasy	804	38	6.4	
50	49	356824	2014-04-28 Byulunculus the Vigilant - Low Life Minis	http://www.coolminior.net/pics/pics16/img535eed487ab7c.jpg	15059	xredmenacex	Other	Fantasy	498	39	5.7	
51	50	362473	2014-08-01 Shadow Soul Herodicta Repaint	http://www.coolminior.net/pics/pics16/img53d0271690ba.jpg	15059	xredmenacex	Other	Fantasy	829	36	5.7	
52	51	371798	2014-05-04 Cancor from Tarantula superhero miniature	http://www.coolminior.net/pics/pics16/img3667d3600135.jpg	15059	xredmenacex	Other	SF	516	39	5.4	
53	52	354263	2014-03-27 Ocfoer Gleech: Bodul Hoink Low Life Minis	http://www.coolminior.net/pics/pics16/img533444f6b953e.jpg	15059	xredmenacex	Other	Fantasy	618	40	6.2	
54	53	363363	2014-08-16 Alpha Centauri from Doctor Who	http://www.coolminior.net/pics/pics16/img53efdc0b11171.jpg	15059	xredmenacex	Other	SF	930	36	6.4	
55	54	361523	2014-07-15 Walloping Krong - Low Life Miniatures	http://www.coolminior.net/pics/pics16/img53c5f890a0a89.jpg	15059	xredmenacex	Other	Fantasy	438	36	6.1	

3. ábra. submissions.csv

Comments A comments fájl tartalmazza soronként a minikhez adott kommenteket. Az entry_id-val azonosítom, hogy a komment melyik feltöltéshez érkezett:

- entry_id
- comment_id
- comment_date
- commenter_user_id
- commenter_user_name
- vote
- comment

J	B	C	D	E	F	G	H
entry_id	comment_id	comment_date	commenter_user_id	commenter_user_name	vote	comment	
1	361658	426154	2014-08-11 14:5446	Jamie-Scott		Very nice and in a challenging scale, excellent balance of effects and restraint!!!	
2	361658	426072	2014-08-08 08:3861	Dragonmarch	9	Father truly. Nice work of a strange ship.	
4	361658	426061	2014-08-07 00:007	Marcos_VF1	9	I thought I was good at Star Wars re-paints. Obviously I still have a lot to learn, master.	
5	318068	362928	2012-11-17 8:5986	Jedgar	9	Great work, love how the reds turned out and the skin tones.	
6	318068	362921	2012-11-17 04:171	Wingtip	10	Really like this presentation! Congrats on a really nice job!	
7	318068	362893	2012-11-14 12:077	TyranMagda	6	Spentacole!!! O_O	
8	318068	362229	2012-11-08 03:06	avio	10	Great!	
9	318068	362120	2012-11-07 24:19	Gorvaki	10	For ME, one of best painters!! Amazing Thorset!	
10	318201	402040	2016-07-13 15:0986	Barrak	10	PRO	
11	318201	402031	2016-06-07 07:0001	Serge-Torder	9	Great! really nicely!!	
12	318201	401808	2016-02-12 11:0334	Grosvenor	10	Must be hard handing over commission work... all that superb work. Amazing painting!!!	
13	318201	401846	2016-02-04 14:0886	MarcusDC	9	Great job!	
14	318201	401845	2016-02-01 25 14:5798	mart_dino	6	Mind blowing detail.	
15	318201	401844	2016-01-16 14:2736	simonryoung80	10	Very impressive	
16	318201	401843	2016-12-09 13:0386	skidkidd	10	Awsome detail!	
17	318201	401802	2015-12-09 48:096	petran	10	amazing work!	
18	318201	401823	2015-11-19 13:0277	Mister JM	9	Impressive...	
19	318201	401809	2015-09-11 9:0949	Atrill	10	I love the armor & the hammer ... impress me...	
20	318201	401807	2015-08-29 12:0075	Ramirez23	10	well! awesome arg!!	
21	318201	401776	2015-08-25 10:0008	hounm103	10	Brillwork on the left brezer is incredible!	
22	318201	401840	2015-08-23 8:7057	M5meagle	10	Spectacular!	
23	318201	401871	2015-08-11 11:0312	Leetokoe	10	Awsome job	
24	318201	401791	2015-07-30 9:9743	Loon	10	Another 12.	
25	318201	391767	2015-06-30 7:4661	Albentartarus	10	Fantastic work	
26	318201	391766	2015-06-29 9:8367	Wachtel2013	10	Cases Bogus! Astonishing, you keep raising the bar, you must be slightly mental with some of the paint jobs you have been producing recently! Followed it on FB, you are the matter. I am not going to give a critique, you are beyond that. Congratulations!!!	
27	318201	391760	2015-06-18 11:0609	Ullriggman	6	amazing in the word I want to use, but it wouldn't do it justice! This is painting mastery.	
28	318201	391710	2015-06-08 8:6161	BloodsMedium	10	Well done again in your texture bringing Illusion to the surface! Model master! Must def look into your cabinet and smile upon your work time and time again! Remarkable work both!!! 10 points plus!!	
29	318201	391698	2015-06-08 7:1880	Ferian	10	Wow, this is a piece of art... its just amazing. As Matt said, one of the best texture Illusions.	
30	318201	391689	2015-06-08 7:1551	LO hall	10	Bravo, painting at the highest level.	
31	318201	391683	2015-06-08 14:421	Matt Cewahh	10	Yeah... Welcome to Texture Painting Heavens, Bohun... One of the Best Texture Illusions I have ever seen in Miniatures! ... 10 Points +...	
32	318201	391664	2015-06-08 3:5658	Rogind	10	The best... No words... The best... no words... the best!	
33	318201	391664	2015-06-08 7:2139	ng9rk	10	Fantastic work like always!!	
34	318201	391637	2015-06-04 13:3462	Bohun	10	Thanks everyone for your comments! Eurya, I don't understand, you say that the scarring would have been amazing if it was sculpted on the model, but not if it is painted?!! Sorry, but I don't see any sense in that. They are not brains, but dreadlocks, the pattern has been sculpted, it's not a fresh	
35	318201	391612	2015-06-03 12:1716	dermis	10	Worth every second, I just couldn't let a commission like that go!! Change your phone number	
36	318201	391611	2015-06-03 13:3962	Harvel	10	PIA~"ma robota, Bogus! Nie jestem wielkim fanem Twojego stylu, ale Aupa~" masz wyrobion~" i jestak! A wstety w tym, co robisz Gratuluj~"!! Bu~"nia, s~"a, s~"a!!	
37	318201	391629	2015-06-03 13:4848	flunktion	10	"You not human!!" You and Yellow One somehow can just implement so much detail into models it's really impressive. The indentation in the armour is really convincing. Every inch of the model has its own unique texture. Bravo sir!	
38	318201	391619	2015-06-02 12:2371	mankon	10	"Just lovely" and u give 9!! Its not even 10, its 12/10! Absolutely astonishing!	
39	318201	391617	2015-06-02 8:3466	Matic	9	amazing features	
40	318201	391620	2015-06-02 7:2649	Lynn Moon	9	Just lovely	
41	340962	404209	2015-09-19 11:0312	Leetokoe	6	OMG amazing	
42	288057	436814	2014-02-10 01:3190	HeadBattered?	10	A wonderful meditation of one of the most iconic pieces of B&W artwork ever.	
43	288057	436827	2014-04-11 14:1625	Major Bob	10	Best diorama I have ever seen!	
44	288057	436874	2014-04-14 14:1474	Thao	10	Stuuuuuuuu	
45	288057	408884	2013-11-30 19:97	Duffi	10	WOW	
46	288057	361635	2013-08-04 3:7059	KARLFRANZ1	10	How dare they make it a silver in a category... should have been a Slayer Sword!	
47	288057	344780	2013-04-12 7:2014	lefoxy-watt@btinternet.com	10	2nd to 10th in 2 days. This should definitely be in top 3 2011.	
48	288057	343387	2013-04-01 13:4539	hopi	10	Fantastic piece!! loved the pic and you totally captured the feel, well done!	
49	288057	343336	2013-03-29 9:5649	werthe	9	I still look for the right words to describe this masterpiece without being overzealous...	
50	288057	343135	2013-03-29 5:2907	M&Nlter	10	I LOVE the artwork this is based on and I really feel that you've done it justice. Amazing stuff right there.	
51	288057	343087	2013-03-28 16:3559	M&Nlter	10	incredible work...	
52	288057	343135	2013-02-26 7:2795	mrdan	10	Just like the picture	
53	288057	340725	2013-02-11 4:2438	AsroDg	10	Me gusta!! really have no advice as to how I could make it any better!	
54	288057	340680	2013-02-09 0:0862	mksacornis	10	Man this is great! Always wanted to do something like this but never took the time, really excellent work!	
55	288057	337474	2013-01-21 0:3703	KARLFRANZ1	10	This is a lifetime me friend.	

4. ábra. comments.csv

Ebay A feltöltött minik egy részéhez tartozik ebay link, ahol licitálni lehet azok megvásárlására. Ezt egy külön adathalmazban rögzítettem.

- entry_id
- ebay_id
- ebay_current_price
- ebay_first_bid
- ebay_number_of_bids
- ebay_end_date

	A	B	C	D	E	F	G
1		entry_id	ebay_id	ebay_current_price	ebay_first_bid	ebay_number_of_bids	ebay_end_date
2	0	448589	324187265056				
3	1	445350	233497247028	GBP 24.99	\$0.00	1	2020-02-21 13:44
4	2	445349	233497244276	GBP 67.50	\$0.00	4	2020-02-21 13:42
5	3	445342	223913966448	\$89.99	\$0.00	0	2020-02-26 23:00
6	4	445331	264632828401	\$10.00	\$0.00	1	2020-02-23 14:59
7	5	445330	264632828750	\$10.00	\$0.00	1	2020-02-23 14:59
8	6	445321	223912227588	\$21.05	\$0.00	2	2020-02-22 11:02
9	7	445320	223912181398	\$49.99	\$0.00	0	2020-02-20 10:28
10	8	445319	223912172740	\$29.99	\$0.00	0	2020-02-22 10:22
11	9	445348	233497242042				
12	10	445343	223914839521	\$35.00	\$0.00	0	2020-02-23 18:54
13	11	445329	264632829585	\$10.00	\$0.00	1	2020-02-20 14:59
14	12	445326	174192782355	\$19.99	\$0.00	1	2020-02-22 19:19
15	13	445322	223912231220	\$65.00	\$0.00	0	2020-02-25 11:05
16	14	445340	223913837856	\$29.99	\$0.00	0	2020-02-21 23:00
17	15	445334	124086309283	GBP 55.88	\$0.00	5	2020-02-23 16:32
18	16	445315	124086311767	GBP 40.88	\$0.00	2	2020-02-23 16:35
19	17	445290	264633696407				
20	18	445269	223908417453	\$69.99	\$0.00	0	2020-02-20 09:18
21	19	445202	223903100747	\$29.99	\$0.00	1	2020-02-15 09:31
22	20	445289	264633682653				
23	21	445280	264633192696	\$0.01	\$0.00	1	2020-02-23 14:59
24	22	445275	223908422928	\$99.99	\$0.00	0	2020-02-18 09:23
25	23	445268	223908409951	\$65.99	\$0.00	0	2020-02-23 09:12
26	24	445245	274263319804	EUR 75.00	\$0.00	0	2020-02-19 08:49
27	25	445194	174186778001	\$19.99	\$0.00	0	2020-02-16 20:37
28	26	438473	254255738689				
29	27	443835	174124637309				
30	28	443751	153757196439				
31	29	445255	164079985189	\$36.00	\$0.00	4	2020-02-20 00:19
32	30	445244	274263331519	EUR 75.00	\$0.00	0	2020-02-19 08:58
33	31	445238	164080863701	\$137.00	\$0.00	0	2020-02-23 14:55
34	32	445237	164080859738	\$84.00	\$0.00	0	2020-02-23 14:52
35	33	445228	223905075714	\$18.99	\$0.00	0	2020-02-16 10:22
36	34	445226	164080853029	\$63.00	\$0.00	0	2020-02-23 14:44
37	35	445243	274263349272	EUR 110.00	\$0.00	0	2020-02-19 09:07
38	36	445236	164080857752	\$81.00	\$0.00	0	2020-02-23 14:50
39	37	445227	164080855870	\$79.00	\$0.00	0	2020-02-23 14:48
40	38	445225	164080850168	\$76.00	\$0.00	0	2020-02-23 14:40
41	39	445224	223904917491	\$35.99	\$0.00	0	2020-02-16 08:10
42	40	445223	223904914141	\$35.00	\$0.00	0	2020-02-21 08:07
43	41	445043	223893428209	\$59.99	\$0.00	0	2020-02-14 08:34
44	42	445064	202897911876	\$15.99	\$0.00	0	2020-02-11 19:07
45	43	445066	124072365118	\$19.99	\$0.00	0	2020-02-11 21:19
46	44	443872	420e580198:g:hjYAAOSwR8hd8rPO				
47	45	445028	233484778997	GBP 59.00	\$0.00	5	2020-02-08 15:24
48	46	445027	233484776605	GBP 51.00	\$0.00	2	2020-02-08 15:20
49	47	444953	193325614478	\$280.00	\$0.00	0	2020-02-06 17:58
50	48	444935	223886552827	\$159.99	\$0.00	0	2020-02-06 08:50
51	49	444930	124069789232	GBP 25.77	\$0.00	2	2020-02-09 16:31
52	50	444920	202892313812	\$425.00	\$0.00	0	2020-02-05 12:05
53	51	444956	124069786283	GBP 56.55	\$0.00	3	2020-02-09 16:27
54	52	335980	251297681129	GBP 0.00	\$0.00	0	2013-07-05 14:30
55	53	430981	263987265891	\$89.00	\$0.00	0	2019-03-12 21:10
56	54	419335	202892318097	\$250.00	\$0.00	0	2020-02-05 12:09

5. ábra. ebay.csv

5. Online ajánlórendszer

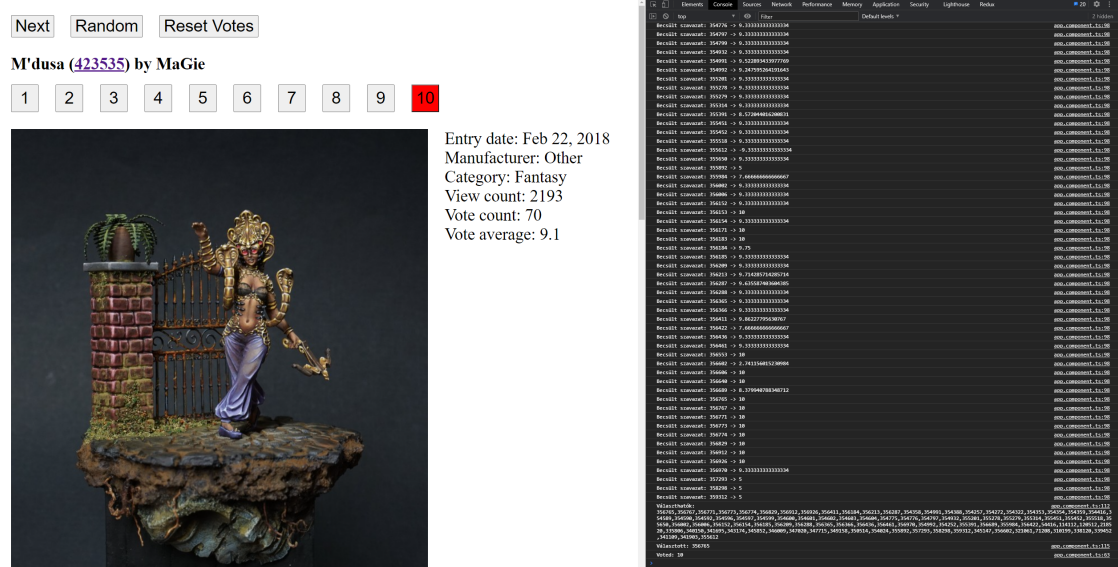
Az ajánlórendszer Item-based Collaborative Filtering alapon működik az ajánlórendszerek házihoz megosztott Recommender Systems, [Sammut and Webb \(2010\)](#) könyvből vett fejezet és az ott hivatkozott [Linden et al. \(2003\)](#) cikk alapján.

Ennek célja, hogy online, real-time képesek legyünk miniket ajánlani egy felhasználónak, miközben a miniket értékeli. A felhasználók összehasonlításán alapuló rendszerek esetében az aktív felhasználó

korrelációit minden szavazat után újra kellene számolni, ami online módon nagyon erőforrásigényes lenne. Ezzel szemben, az item-alapú ajánlórendszert használva előre ki tudtam számolni a minik közötti korrelációkat. Ezt egy nagy méretű json fájlban tároltam el, mely [ezen a linken](#) tekinthető meg.

Látható, hogy sok mininek 1.0, illetve -1.0 a korrelációja, ez abból adódik, hogy egyetlen felhasználó a kapcsolat közöttük, ez a felhasználó pedig ugyanazt a szavazatot adta mindkét minire.

Az ajánlórendszerhez készült egy egyszerű Angular-os weboldal, mely így néz ki:



6. ábra. Ajánlórendszer

A Next gombra kattintással lehet kérni a következő ajánlott minit, a Random gombra kattintással véletlen minit kérhetünk (ajánlórendszer nélkül), a Reset Votes gomb pedig törli az összes leadott szavazatunkat.

Amini neve melletti entry_id-ra kattintva az eredeti weboldalra navigálva megtekinthető az eredeti beküldés. A számolások részeredményeit a konzolban lehet nyomonkövetni.

6. Tanulságok

A félév során nagyon sokat tanultam a feladat elkészítése során.

A letöltés sokkal bonyolultabb lett mint amire számítottam: A letöltött weboldalak néha teljesen másképp néztek ki mint az élesek, például a kommentszekciót a wget teljesen kihagyta. A törölt linkek nagyon sok problémát jelentettek, amiket nagyon körülményes volt kijavítani, van ahol nem is sikerült (valószínűleg elmozgatták azokat a képeket az éles website-on). Az adathalmaz méretéből fakadóan meg kellett oldani, hogy 2 héten keresztül zavartalanul fusson a letöltő, problémát okozott például egy kiment biztosíték, ami miatt előről kellett kezdeni a letöltést. Az adatok mérete szerencsére körülbelül 100 GB volt, de más adathalmaz esetében ebből is lehetett volna probléma. Emellett szerencsére nem tiltották le az IP címetet, még mindig elérem a weboldalt, pedig más weboldalak esetében ez is megtörténhetett volna.

Az adatelemzés közben próbáltam különféle összefüggéseket találni az adatokban, rajzoltam grafikonokat, szavak darabszámát elemeztem, ezen kódok egy része a <https://github.com/nemkin/cool-mini-or-not/blob/main/python/analysis.py> linken megtalálható. A problémát az jelentette, hogy nem igazán találtam érdekes összefüggést, amelynek mentén el lehetett volna indulni. Nem találtam olyan attribútumot az adathalmazban ami jelentősen meghatározta volna a szava-

zatok alakulását, ez várható is volt hiszen a festés minősége a fő döntő elem.

Ezért végül egy ajánlórendszer elkészítése mellett döntöttem a szavazatok és a képek alapján. Itt szerettem volna valamilyen interaktív megoldást adni, ezért választottam az Item-based Collaborative Filtering módszert, mert azzal előre kiszámolt korrelációk alapján lehetett ajánlatokat adni. A korrelációk kiszámolása több órát vett igénybe, ezt szavazatonként biztosan nem lehetett volna megvárni.

A weboldal használata közben észrevettem, hogy ha jó szavazatot adok egy mininek, akkor előfordul, hogy ahhoz hasonló méretű, kinézetű (más felhasználó által feltöltött) miniket kapok legközelebb, ezért összességében valamilyen összefüggést sikerült kihasználni a collaborative filtering módszerével.

Hivatkozások

Greg Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *Internet Computing, IEEE*, 7:76 – 80, 02 2003. doi: 10.1109/MIC.2003.1167344.

Claude Sammut and Geoffrey I. Webb, editors. *Encyclopedia of machine learning*. Springer, New York; London, 2010. ISBN 978-0-387-30768-8.